**Natural Hazards
and Earth System
Sciences**

Open Access

Influence of targeted observations on short-term forecasts of high-impact weather events in the Mediterranean

J. Campins[1]**, B. Navascués**[2]**, C. Santos**[2]**, and A. Amo-Baladrón**[2]

[1]Agencia Estatal de Meteorología, Moll de Ponent, 07015, Palma, Spain
[2]Agencia Estatal de Meteorología, Leonardo Prieto Castro 8, 28040, Madrid, Spain

Correspondence to: J. Campins (jcampinsp@aemet.es)

Abstract. The influence of targeted observations on short-range forecasts is tested over two different periods of PRE-VIEW (2008) and MEDEX (2009) data targeting field campaigns for a set of Mediterranean high-impact weather events. As targeted observations we have used not only extra radiosondes, but also enhanced satellite data observed in singular vector (SV)-based sensitive regions. Three parallel observing system experiments, based on the High-Resolution Limited-Area Model (HIRLAM) data assimilation and forecast system, have been conducted. Forecasts of the three experiments have been assessed using both verifying analyses for upper-air fields, and surface observations for several meteorological parameters. Furthermore, quantitative precipitation forecasts (QPF) have been objectively verified using the novel feature oriented Structure–Amplitude–Location (SAL) method.

The results obtained show that extra radiosondes have an overall positive impact on the forecasts (average improvement of all upper-air variables and vertical levels studied is 3.6 %). When in addition to extra radiosonde data also enhanced satellite data are assimilated, the overall forecast skill is almost doubled. However, a distinct behaviour is found between the PREVIEW and MEDEX cases. While for MEDEX cases the improvement is slight, for PREVIEW cases the improvement is significant (average improvements of 1.7 % and 8.9 %, respectively, for the experiment with enhanced satellite data). It is suggested that this is due to the location of the target areas and the spatial distribution of the composite observing system and to the different atmospheric predictability in these two periods.

1 Introduction

Numerical weather prediction (NWP) has steadily improved forecast skill over the last decades (Navascués et al., 2013). These advances arise from progress on numerical models and data assimilation (DA) methods as well as the increased quality and availability of data. However, there still exists significant space for improvement, especially in the forecast of high-impact weather. Two key sources underlie forecast errors: uncertainties in the initial conditions and the presence of inaccuracies in the forecast models (García-Moya et al., 2011).

One way to tackle this problem is to complement deterministic forecasts with a probabilistic approach, trying to sample atmospheric uncertainties and estimate the evolution of a probability density function (PDF) that would represent the atmospheric state in the phase space. Ensemble forecasting, a technique that runs a finite number of deterministic models to build such a PDF, has shown to be a valuable tool used in most operational centres (García-Moya et al., 2011).

Another way to reduce forecast error is to describe the initial state of the atmosphere as precisely as possible. Errors in the initial conditions are a result of inaccurate observations and background fields, or inadequacies in the assimilation scheme. Additionally, the observations distribution in space is far from uniform, with some land regions with a high data density, such as in Europe, and wide sparse areas located not only in oceanic but also in continental regions, such as northern Africa. Errors in initial conditions, even small ones, amplify in time and can lead to significant forecast deficiencies. In this context, data targeting (or targeted observing) has become a promising strategy to improve forecasts, especially

those of events with large societal and economic impact. Data targeting refers to the selection of additional observations in particular regions to reduce the initial condition errors.

Over the past years several research experiments have included a targeting component in their field experiments. These field experiments were organized in different regions of the world, focused on assessing the impact on forecast accuracy of adding extra observations in specific target areas identified using objective methods. Many field campaigns deploy special observations to add an extra component to the observing system (e.g. dropsonde data from research aircraft), but others make use of adapted observational resources that are routinely available (e.g. additional radiosondes). This approach is far more cost-effective and a practical way to do data targeting. The general conclusion of these campaigns was that, on average, targeted observations increased the short-range forecast accuracy, but the improvement found in the extratropics was quite small (Langland, 2005; Majumdar et al., 2011).

Enhanced regular satellite observations that may normally be excluded from data assimilation due to routine data thinning can be also considered as targeted observations (Langland, 2005; Majumdar et al., 2011). Dando et al. (2007) investigated the impact on forecast skill of using additional ATOVS (Advanced TIROS Operational Vertical Sounders) data in targeted regions for some case studies. Selective satellite data thinning was applied by Bauer et al. (2011) to reduce analysis uncertainty and forecast error in the Southern Hemisphere extratropics, increasing data density in singular vector (SV)-based sensitive regions.

The Mediterranean basin is frequently affected by different meteorological natural hazards such as heavy precipitation and strong winds. Heavy precipitation events are not uncommon phenomena over the Mediterranean, especially in autumn for the western basin and between December and February in the eastern basin. In some cases rainfall reaches several hundreds of millimetres in a few hours, producing flash-floods (Ramis et al., 1998); in other cases, large amounts of precipitation can be accumulated over several days (Homar et al., 2002). Heavy precipitation events are multiscale atmospheric phenomena as they result from a complex interaction of synoptic upper-level steering flows and local topographic barriers (Rudari et al., 2004): where synoptic and mesoscale ingredients interact (Nuissier et al., 2008).

Strong winds are amongst high-impact weather events that frequently affect some Mediterranean regions, including sea areas. A great part of the observed strong winds are local winds, such as Mistral, Tramontane and Bora (Reiter, 1975). On the other hand, intense cyclones can produce windstorms, and occasionally other hazards such as high wind waves or storm surges (Homar and Stensrud, 2004).

As most part of heavy precipitation and windstorm events are closely related to cyclones, the Mediterranean Experiment MEDEX, under the auspices of the World Meteo-

rological Organization (WMO), focused on cyclones that produce high-impact weather in the Mediterranean (http://medex.aemet.uib.es). The first phase of MEDEX was mostly devoted to better understand the different types of cyclones related to high-impact weather, and the factors that might influence the genesis and evolution of such cyclones. The second phase of MEDEX concentrated on the development and testing of targeting strategies oriented to the improvement of the forecasting of those cyclones. This objective was planned to be embodied in a field campaign. A first trial was framed into the PREVIEW Data Targeting System (DTS, Prates et al., 2009), and fully carried out for MEDEX-DTS (Jansà et al., 2011).

This paper presents new results concerning different aspects of data targeting experimentation: the impact of targeted observations on short-range (1–2 days) forecasts of high-impact weather events in the Mediterranean, the usage of adaptive observations screening strategies within the DA system to ameliorate the satellite data sampling, and the assessment of the forecast skill of surface meteorological parameters of high social impact (rainfall, wind) in addition to that of upper-air variables. These three issues are on the list of recommendations concerning data targeting made by the Working Group on Data Assimilation and Observing Systems of the WMO program THORPEX (Majumdar et al., 2011).

A set of parallel observing system experiments has been conducted using the HIRLAM (High-Resolution Limited-Area Model) data assimilation and forecast system. This NWP system is operational at the Spanish Meteorological Service, AEMET, (Navascués et al., 2013) and many other National Meteorological Services (NMSs) in Western Europe. The experiments have been carried out for two-week periods of PREVIEW-DTS and MEDEX-DTS field campaigns, which included a large set of case studies. As targeted observations we have used not only additional radiosondes, but also enhanced raw radiance data from satellite sounders located in SV-based sensitive regions. Forecasts have been evaluated against verifying analyses and surface observations from SYNOP meteorological reports. Performance assessment of quantitative precipitation forecasts (QPF) has also been addressed by means of the novel feature oriented Structure–Amplitude–Location method (SAL; Wernli et al., 2008) using rain gauge data from the high-resolution climate stations network operated by the European NMSs.

This document is structured as follows. First, the set-up of the targeting experiments and the methodologies used to assess the impact on forecasts are presented in Sect. 2. Section 3 provides a synoptic overview, as well as the location of the sensitive regions for two particular events, and a description of the targeted and routine observations deployed during the PREVIEW-DTS and MEDEX-DTS field campaigns. The results of the different objective verification methods applied are given in Sect. 4, while Sect. 5 outlines the main conclusions and final remarks.

2 Methodology

The impact experiments presented here have been carried out using the HIRLAM NWP system. The HIRLAM hydrostatic model is described in Undén et al. (2002). The parameterization for clouds and condensation processes was reformulated and the new scheme implemented is based on Kain-Fritsch for convection (Kain, 2004) and Rasch-Kristjansson for large-scale microphysics (Zhang et al., 2003). In these experiments we have used the incremental HIRLAM 4-D variational data assimilation (DA) system, 4-D-Var (Gustafsson et al., 2012), as upper-air analysis with a 6 h time window. Physical processes are represented in the tangent linear and adjoint models. Operational forecast fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) global model are used as boundary conditions. During the MEDEX and PREVIEW DTS campaigns all the additional observations were also disseminated through the Global Telecommunications System (GTS) and assimilated by the ECMWF global model a few hours later than HIRLAM. A rotated grid at 0.15° horizontal resolution with 60 vertical levels is used in this study. The model domain covers most of the Northern Atlantic Ocean, Europe and North Africa and its boundaries are far away from the different targeting areas where the additional observations were deployed.

2.1 Experiment description

Three experiments have been conducted over the two different periods of 2008 (PREVIEW) and 2009 (MEDEX) field campaigns. They are carried out in an operational context taking into account the existing composite observation system and the observation usage in the HIRLAM data assimilation system. The assimilation cycle started one week before for spin-up reasons.

The first experiment, EXP-RS, is the experiment designed to test the impact of the targeted radiosondes (RS). Data from surface, aircraft, radiosondes and other upper-air conventional observations were assimilated (SYNOP, SHIP, BUOY, AMDAR, TEMP, PILOT meteorological reports). It also assimilated raw radiances from different sounder instruments as the Advanced Microwave Sounding Unit-A and Unit-B (AMSU-A and AMSU-B), and the Microwave Humidity Sounder (MHS) for all the available space platforms over the sea areas. AMSU-A mainly contributes to the temperature analysis, whereas AMSU-B and MHS help to determine the atmospheric moisture initial state. The set of sounders is known as ATOVS, Advanced TIROS Operational Vertical Sounders.

The second experiment, EXP, is the control experiment. It receives exactly the same observations that EXP-RS, but the targeted radiosondes are not assimilated because they are blacklisted and then screened out of the minimization.

Thinning is needed in DA for observation types with high volume of data, mainly for improving computing efficiency and to reduce the effect of spatially correlated observation errors, and is normally applied to satellite DA. In HIRLAM, a spatial thinning is carried out during the screening step. Non-rejected data in previous quality control checks are spatially sampled in order to reduce the observation density for this particular data type. In EXP and EXP-RS, AMSU-A, AMSU-B and MHS data were thinned according to the standard procedure implemented in the screening part of the HIRLAM DA (Schyberg et al., 2003), and in the thinning step these data were sampled allowing a minimum distance of 0.9° between them.

The number and location of extra radiosondes were rather different from case to case during the field campaigns, and very often, they only partially sampled the sensitive regions. The third experiment, EXP-ATOVS*2, aims to investigate other data targeting strategies in HIRLAM DA, in particular a non-uniform data thinning for the satellite data located in the target region, following the work of Bauer et al. (2011) with ECMWF global model. With EXP-ATOVS2, we seek to test the potential additional improvement obtained with an enhanced sampling of ATOVS data located in the sensitive regions over the ocean and sea areas to complement the extra radiosondes mostly launched in Europe. EXP-ATOVS*2 experiment assimilated the same observations than EXP-RS but, in the sensitive areas, it allowed a double density of satellite data to influence the analysis (a minimum distance of 0.45°, close to the original AMSU-A data resolution). Some changes in the screening algorithms of the DA system were implemented accordingly.

2.2 Assessment of the forecast impact

Different metrics have been applied to thoroughly evaluate the impact of the targeted observations assimilated on the short-range forecast models. The goal is to assess both upper-air atmospheric fields and some surface meteorological parameters with a social impact: rainfall and wind. All results have been obtained only for those forecasts started at the targeting times. The verification domain used in this assessment is exactly the same verification area used to determine the target areas in the DTS, i.e. the region where the high-impact weather was expected to take place. It changed from one meteorological case to another. Next subsections describe the verification methods employed.

2.2.1 Verification against analyses

NWP model analyses are often used to assess the model representation of the large-scale flow. As far as analyses are produced by combination of observations and a short-range forecast used as first guess, typical issues related to raw observations (see following section) are better tackled. In principle, verification results may depend strongly on the analysis

selected. Some tests have been carried out to check how this may affect this study. The results obtained show that the main conclusions remain valid regardless the verifying analysis used. Therefore, the EXP-RS analysis has been chosen, as it is expected to be more accurate than the one from EXP, because the former has assimilated a larger number of radiosonde data.

Classical scoring rules have been used to assess this large-scale model performance in the different experiments set-up. Root mean square error (RMSE; Wilks, 2011) of forecast fields with respect to analysis fields for all the grid points included in the verification area (VA) at the verification time (VT) is used.

Skill scores are often used as a relative measure of performance with respect to a reference forecast (Wilks, 2011). Particularly, RMSE skill score (RMSESS) is given by 1-RMSE/RMSE$_{ref}$. Alternatively, the sign-opposite is used here, given in %, using EXP experiment (control) as reference and calling it relative RMSE (REL_RMSE):

$$\text{REL\_RMSE(DTE)} = 100 \cdot \frac{\text{RMSE(DTE)} - \text{RMSE(EXP)}}{\text{RMSE(EXP)}} \quad (1)$$

Its interpretation is straightforward: when REL_RMSE < 0, DTE (Data Targeting Experiment) forecast error is lower than EXP one and therefore there is an improvement, while for REL_RMSE > 0 it is the opposite. Both RMSE and REL_RMSE can be calculated separately for each data targeting case and also for all the data targeting cases at the same time (or stratifying 2008 events on one group and 2009 ones on the other group). A more simple way to evaluate the impact of data targeting on the forecasts is to compute the number of improvements over the total number of cases.

In this study the number of improvements over the total number of cases, as well as the REL_RMSE for both experiments, EXP-RS and EXP-ATOVS*2, are shown for several variables (geopotential height, temperature, wind speed and relative humidity) in a number of vertical levels (925, 850, 700, 500 and 300 hPa).

2.2.2 Verification against surface observations

Performance assessment of weather parameters is often carried out with observations. Compared with that one using analyses, it needs to address difficult issues related with homogeneity, uniformity, larger errors, representativeness, spatial scales, etc. (Jolliffe and Stephenson, 2011). Here, the assessment of the forecast impact experiments for surface parameters has been carried out making use of observations using standard point-wise verification methods (Wilks, 2011). Mean sea level pressure, wind speed, and accumulated precipitation data, contained in all available SYNOP reports from land meteorological stations located in the mobile verification region, have been used to verify only those forecasts started from the different targeting time analyses. Standard scores have been calculated (see Wilks, 2011; Jolliffe and

Stephenson, 2011 for a comprehensive description): BIAS, RMSE and equitable threat score (ETS) from corresponding contingency tables for different thresholds of accumulated precipitation and wind speed. In the data sets used, the observation density is far from uniform within the different verification areas and, for simplicity, the impact of this lack of uniformity has not been considered in this part of the study. Spatial methods are also used and they deal with this aspect (see following section).

2.2.3 Quantitative precipitation forecast (QPF) verification by means of the feature oriented method SAL

In the context of forecast evaluation, spatial scale of models and observations turn out to be a key point when model resolution becomes higher. Given a high-resolution forecast that is actually accurate but not perfect, classical scoring rules like RMSE can give better scores to a coarser grid forecast which is not better, due to a problem called double penalty (Davis et al., 2006). To address these issues, new spatial verification methods are under development in the recent years, with promising results (Gilleland et al., 2009). Along with the families of scale decomposition, neighbourhood, and field deformation, the feature oriented family has shown to be able to fairly assess forecast performance. The SAL method (Wernli et al., 2008), belonging to the family of feature oriented methods, provides explicit and quantitative information about three different features of forecast performance: structure, amplitude and location.

The method starts finding objects as regions of precipitation exceeding a given threshold that depends on the highest values (95th percentile) found in the domain. Then some properties are assigned to each object. After, some properties are assigned to the whole field as a set of objects. Finally, three numbers (S, A, L) compare properties of the two fields. A short explanation of these parameters follows here.

The structure component, S, indicates the size and/or shape of precipitation model structures with respect to observation structures. The range of possible values is $[-2, +2]$. Negative values mean that forecast objects are too small or too peaked compared to those of the observational field, while positive ones mean too large or too flat, with $S = 0$ corresponding to perfect structures. The amplitude component, A, provides a sort of model normalized bias for QPF amount. It takes values in the range $[-2, +2]$, $A = 0$ corresponds to a perfect amount, while positive (negative) values indicates overestimation (underestimation). The location parameter, L can range in $[0, +2]$ and is split in two components. The first one corresponds to the location of the model precipitation distribution with respect to the observation one. The second provides information about the relative position of objects in the fields. As a summary, for the three components, the closer to zero the better the forecast field compared to the observation field; for a perfect forecast $S = A = L = 0$. The

full description of these parameters can be found in Wernli et al. (2008).

The SAL method compares forecasts (quantitative precipitation forecast, QPF) and observations (quantitative precipitation estimate, QPE) in the same grid, thus some kind of gridding method is required for the QPE (Cherubini et al., 2002). Moreover, SAL method should be applied in limited size geographical areas (about 500 km × 500 km, see Wernli et al., 2009) in order to obtain clear conclusions. Because the verification areas (VAs) used in these DTS field campaigns were larger than this limit, all VAs at each VT were split in several sub-areas, considering each one in the SAL method as a different case. For the QPE, High-Resolution Rain Gauge Data (HRRGD) from the climate stations network operated by the different European NMSs have been upscaled by ECMWF to the model resolution (around 16 km) and provided for this study. Spatial and temporal HRRGD variability can have an impact on the QPE consistency, and hence in the results obtained, but this is beyond the scope of this study (see, for instance, PaiMazumder and Molders, 2009). However, some indication is provided here. In some European regions HRRGD covers most of the land domain including mountains (e.g. the Alps), while in some others it is rather sparse: HRRGD density can vary in Europe from 10^{-2} to 10^{-3} km^{-2}. Concerning daily coverage variation, some countries keep an almost constant number of quality-controlled daily data, but this is not generally the case (the variation in number of available daily data can range from 1 % to 30 %). Therefore, neither the coverage is homogeneous, nor the availability of each station data can be considered constant. In fact, HRRGD were available for most, but not for all, the weather cases studied in this work. In particular, the data were sparse for some Eastern Mediterranean countries over the MEDEX-DTS period. On average, around 1300–1400 data were daily available on each 500 km × 500 km verification domain. Within, and for daily precipitation, only those grid-boxes containing at least 2 observations were considered for the up-scaling process (the rest were neglected), improving the QPE consistency.

3 DTS cases in PREVIEW and MEDEX

The aim of data targeting is to add observational data into specific regions in order to improve the forecast of a particular weather event. For this purpose, in 2008 ECMWF, in partnership with the UK Met Office, developed an interactive web-based Data Targeting System (DTS) to efficiently manage the data targeting process (Prates et al., 2009). The work was jointly funded by the EUCOS (EUMETNET Composite Observing System) operational program of the European Consortium of NMSs, EUMETNET, and the European Commission as part of the PREvention Information and Early Warning (PREVIEW) Integrated Project of 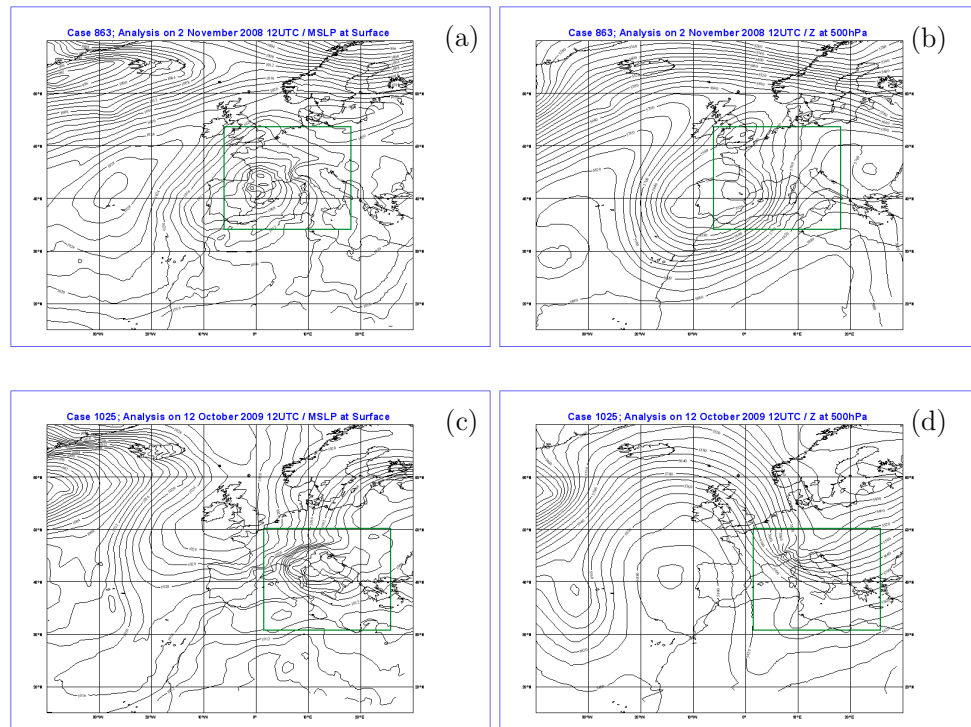the 6th Framework Programme. The PREVIEW-DTS spanned from March to December 2008 and the main objective was the improvement of short-range forecasts of potentially high-impact or high-uncertainty weather events in Europe. For autumn months (from 15 September to 19 December) data targeting was focused also on the Mediterranean basin. On autumn 2009 (from 11 September to 20 December), EUCOS supported the research field campaign of the Mediterranean experiment on cyclones that produces high-impact weather in the Mediterranean, in the so-called MEDEX-DTS (Jansà et al., 2011).

The process of data targeting follows three steps. The first step is the case selection: when (verification time, VT) and where (verification area, VA) that event might occur. The second step is the prediction of the sensitive areas (sensitive area predictions, SAPs). In this step the region where a more accurate definition of the initial state of the atmosphere might improve the forecast over the VA and at the VT are calculated. This region is called targeted area (TA), and the time at which observations must be deployed is the targeting time (TT). Focused on improving short-range forecasts, the difference between the verification time and the targeting time (the optimization time) ranged between 18 and 42 h.

There are several objective methods to calculate sensitive regions. Probably the most used techniques in field experiments are the singular vector (SVs, Buizza and Montani, 1999) and the ensemble transform Kalman filter (ETKF, Bishop et al., 2001). In PREVIEW-DTS and MEDEX-DTS, ECMWF provided sensitive area predictions (SAPs) based on a SV method, while the UK Met Office (only for PREVIEW) and Météo-France provided SAPs generated using ETKF. Details of the techniques used to identify the sensitive regions can be found in Prates et al. (2009).

Once the SAPs were calculated, in the third step the "lead user" evaluated the different SAPs and decided which of the available observations would be requested. From the operational point of view in PREVIEW-DTS, SVs and ETKF from the UK Met Office were used to guide the selection of targeted observations. Both methods often agreed, but in some cases they pointed out different sensitive regions. The request of targeted radio soundings was mostly driven by SVs. Only in a few cases did the "lead user" select a large targeted area to include additional radiosondes located not only in SVs but also in ETKF sensitive regions. On the contrary, in MEDEX-DTS only a SV method was used.

To investigate the forecast impact of targeted observations, we have selected all the cases included in a two-week period for both PREVIEW-DTS and MEDEX-DTS campaigns. For PREVIEW-DTS, the study period comprises from 22 October to 5 November 2008, and for MEDEX-DTS from 7 October to 23 October 2009. That means a total of 24 cases, but as some of the cases refer to the same weather event, they have been gathered in 7 events (Table 1) for meteorological consistency.

Fig. 1. Top: MSLP (left panel) and 500 hPa geopotential height (right panel) analyses for case 863 (event c). Bottom: like top panels but for case 1025 (event f). The analyses are derived from EXP-RS. The green box is the verification area for the corresponding cases.

3.1 Synoptic overview and sensitive region predictions for two events

Mean large-scale flow was rather different in MEDEX-DTS and PREVIEW-DTS periods. In the first period, large-scale flow was characterized by wave-breaking of amplified Rossby waves, which gave raise to successive cut-off lows in the south-western side of the verification areas. In the second period, amplified Rossby waves also affected the verification areas, with intense western jet streams but further north and normally without wave-breaking. Although each event features its own characteristics, events for each period share some similarities. For this reason, only an event for each period is described with some detail in the following paragraphs: event c (PREVIEW) and event f (MEDEX).

Event c: 1 and 2 November 2008

This event is characterized by a deep large-scale upper-level cut-off low with an intense southwesterly jet over the Western Mediterranean region. During the 1st November the cut-off, initially centred at 00:00 UTC over the gulf of Biscay, elongated along the northeast–southwest axis. On the 2nd November there are two main maximum vorticity centres inside the cut-off, one affecting southern France and the other the south-western part of Spain (Fig. 1b). At surface level
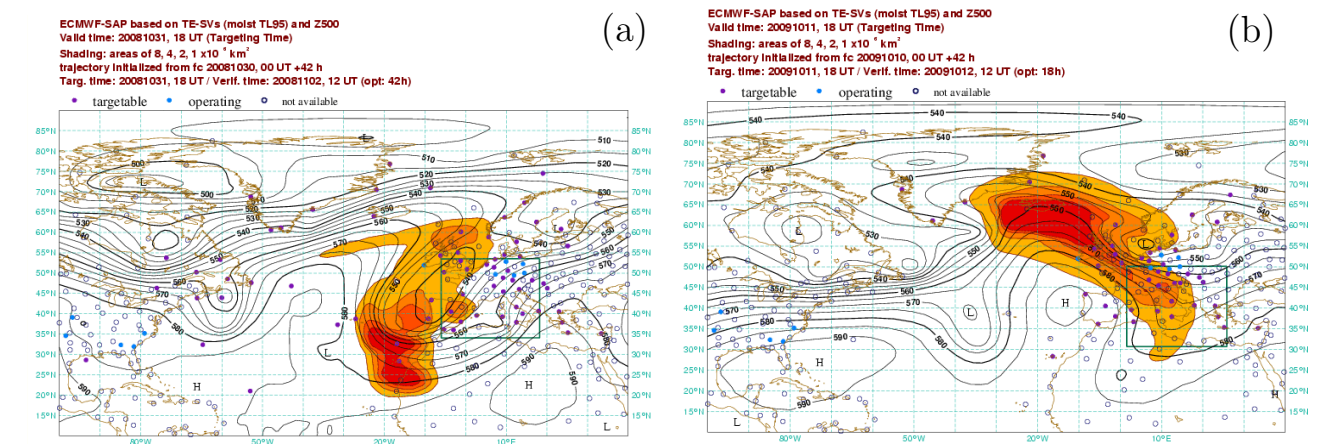
there appear two low pressure centres, the deepest one over the Pyrenees (Fig. 1a). This configuration produces a warm and wet maritime low-level flow over the Mediterranean coastal regions. Heavy precipitation affected southern France (the Cevennes region), with values over 300 mm 24 h^{-1} (Vié et al., 2011; Fresnay et al., 2012) and north-eastern Spain (Catalonia), with values larger than 100 mm 24 h^{-1} (Bech et al., 2011).

As a result of the quasi-stationary upper-level system, sensitive regions are near the same in all six cases of the event, and they locate around the cut-off, with maxima over the Canary Islands and the Iberian Peninsula (Fig. 2a). Verification areas are also very similar in all six cases and they contain the regions actually affected by heavy precipitations.

Event f: 11–13 October 2009

During this long event, two consecutive Genoa cyclones affected the central and eastern Mediterranean. On the 10th October, a strong northwesterly jet, related to a North Atlantic trough, impinges on the Alps, with an upper-level trough over central Mediterranean. As this first Genoa cyclone moves eastwards during the 11th October, another intense upper-level jet affected the Alps. As a result, on the 12th October a new and deeper Genoa cyclone develops (Fig. 1c and d). This cyclone moves towards central Europe during the 13th

Fig. 2. Sensitive area prediction from ECMWF-SAP (shaded) along with 500 hPa geopotential height ECMWF forecast (contour) at the target time for case 863 **(a)** and 1025 **(b)**. The verification area is marked with a green rectangle and circles identify radiosonde stations that were either available for targeting (purple), operating routinely at target time (blue) or not available at that time.

Table 1. Description of high-impact weather events for two-week periods of PREVIEW-DTS (2008) and MEDEX-DTS (2009) field campaigns. The optimization time (OPT) is the difference between verification time (VT) and target time (TT), and NRS is the number of targeted radiosondes deployed for each case.

Event	Case	TT	VT	OPT	NRS
a	830	2008102206	2008102312	30	10
b	847	2008102818	2008102912	18	28
	852	2008102918	2008103012	18	17
	853	2008102918	2008103100	30	17
c	861	2008103118	2008110200	30	13
	863	2008103118	2008110212	42	13
	862	2008110106	2008110200	18	11
	865	2008110106	2008110212	30	11
	867	2008110118	2008110212	18	9
	870	2008110118	2008110300	30	9
d	878	2008110318	2008110412	18	11
	879	2008110318	2008110500	30	11
e	1013	2009100806	2009100900	18	5
	1014	2009100718	2009100812	18	6
f	1019	2009100918	2009101100	30	8
	1024	2009101018	2009101200	30	15
	1025	2009101118	2009101212	18	18
	1028	2009101218	2009101312	18	21
g	1045	2009101918	2009102100	30	3
	1047	2009102006	2009102100	18	4
	1048	2009102006	2009102112	30	4
	1052	2009102018	2009102200	30	3
	1055	2009102106	2009102200	18	7
	1056	2009102106	2009102212	30	7

October. Intense precipitations affected the eastern borders of those intense cyclones, while strong winds the western ones.

For this event, the verification area moves eastwards, from the gulf of Genoa to the eastern Mediterranean, as the main weather system does. The sensitive areas reflect this displacement, although the general shape is rather similar. As the major driving factors of this event are successive northwesterly jet streams, in each case the sensitive area is in agreement with the position of the jet, but also far upstream, over the building ridge (Fig. 2b). For this reason, sensitive regions are located over central Europe (north of the Alps) and over the North Atlantic.
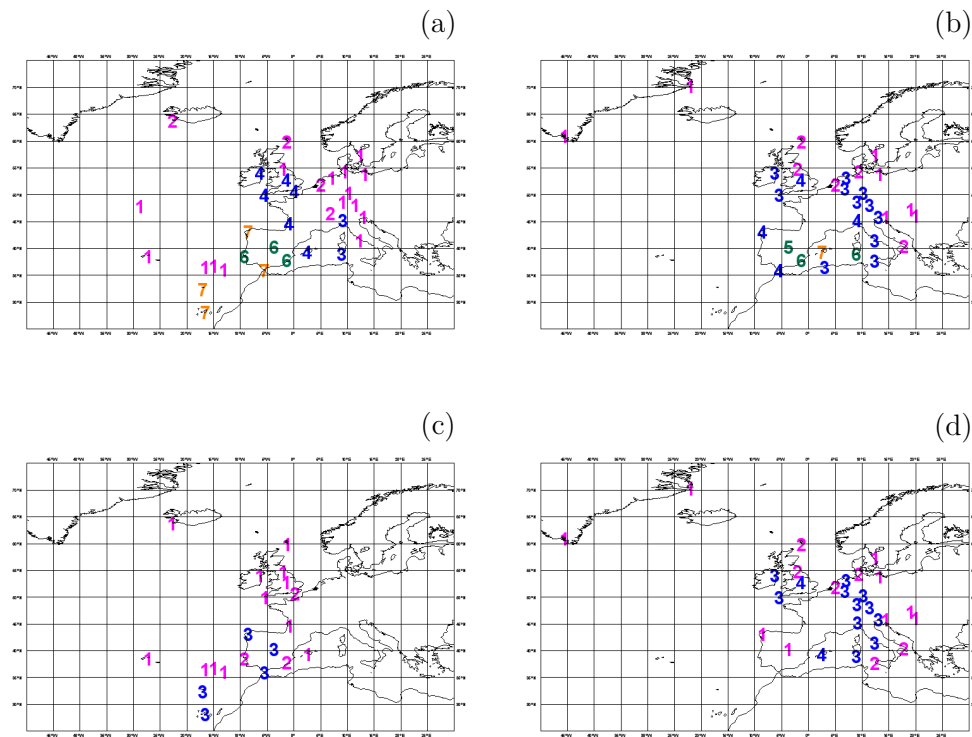
3.2 Distribution and density of upper-air observations in control and data targeting experiments

The mean upper-air observations that feed the HIRLAM data assimilation system come from three information sources: radiosondes, satellite sounders (ATOVS), and aircraft (AMDAR meteorological reports). In this subsection we briefly review the availability of these observation types in the sensitive areas during PREVIEW and MEDEX DTS campaigns.

3.2.1 Radiosondes

Throughout the PREVIEW-DTS and MEDEX-DTS, observations from radiosonde ascents at 06:00 or 18:00 UTC from several operational stations operated by many European NMSs (plus Canada and Bermuda for PREVIEW or Algeria for MEDEX), and radiosondes from the EUMETNET Automatic Ship Aerological Program (ASAP) ships operating in the Atlantic (only for PREVIEW) were available for targeting.

Figure 3a and b show the number and location of the targeted radiosondes deployed for 2008 and 2009 events respectively. Although the total number of observations for 2008

Fig. 3. Number of additional radiosondes deployed for all 2008 **(a)** and 2009 **(b)** events, and for event c **(c)** and event f **(d)**. Coloured number: 1–2 (violet), 3–4 (blue), 5–6 (green), and 7 or more (orange) radiosondes.

and 2009 is rather similar (99 and 90 radiosondes), some significant differences in their geographical distribution can be highlighted. Thus, for 2008 events, most of the targeted radiosondes were deployed over continental Spain and Portugal as well as southwest of Iberian Peninsula, including Madeira and Canary Islands. In that region, four ASAPs were deployed (three of them north of Madeira and the other farther north Azores islands). Some targeted radiosondes were also launched from the British Islands, Central Europe and Italy. This distribution obeys to the predicted sensitive regions for such events, that were located over the most important weather systems as North Atlantic troughs or cut-off lows (over the Gulf of Cadiz or the Iberian Peninsula). It is worth mentioning that sensitive regions also included western North Africa, but unfortunately no North African countries took part in this field campaign. In Fig. 3c, the distribution of targeted radiosondes for event c is displayed. One third of the total 2008 targeted observations were deployed during this event, the most intense of that period. For this event, most of the radiosondes concentrated around the Gulf of Cadiz, over a cut-off low, and only a few observations were deployed over the British Islands (but neither over Central Europe nor over Italy).

For 2009 events, most of the targeted observations were distributed over the British islands, Central Europe, Italy, and the Iberian Peninsula (see Fig. 3b). However, targeted ob-
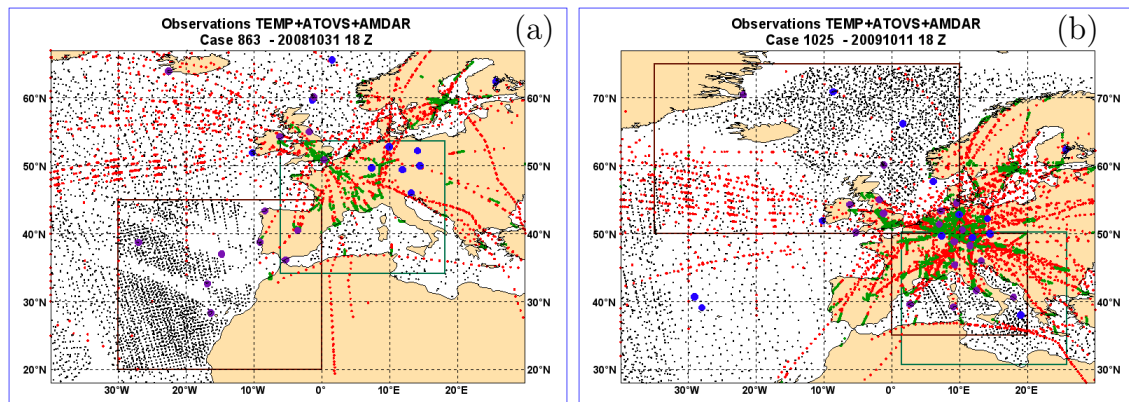
servations were not uniformly distributed in time as most of the additional radiosondes deployed for that period were requested for event f (62 of 90 radiosondes). Sensitive regions for event f are related to the evolution of the upper-level jets that impinge over the Alps and the later Genoa cyclogenesis. For this reason targeted radiosondes come from stations located in many places in Europe except for Iberian Peninsula (see Fig. 3d).

It must be pointed out that at 06:00 and 18:00 UTC, a few radiosonde stations were routinely operating at the target time, and for this reason they are not targetable. These radiosondes are assimilated in both control and data targeting experiments. They are not usually located in sensitive regions (blue dotted points in panels of Fig. 2). However, for some cases, these operating radiosonde stations fit the targeting areas, like for instance in case 1025, with some stations located over Germany (blue dotted points in Fig. 2b). In this case, targeted observations were deployed over Central Europe, an already high observational density area.

3.2.2 ATOVS data

Enhanced regular satellite observations can be considered as targeted observations. In the present paper an adaptive observations screening has been tested. A reduced (0.45°) satellite data thinning is applied in some regions, derived from the sensitivity maps, and a normal thinning (0.90°) in

Fig. 4. Different data assimilated for case 863 **(a)** and for case 1025 **(b)**. ATOVS data are displayed in black points, AMDAR cruise altitude (above 400 hPa) in red points, AMDAR profiles (below 400 hPa) in light green points, operational radiosondes in blue circles, and additional radiosondes in purple circles. The verification area is marked with a green rectangle and the brown rectangles mark the targeted area used for enhanced ATOVS data.

non-sensitive regions. In this experiment, SVs are used as a SAP method to be consistent with the "lead user" decisions to select the extra radiosondes in the operational targeting environment. For 2008 events, target regions are rather similar, and one single rectangular box with enhanced satellite data sampling was enough to cover all sensitive regions. For these events, this box is located around the Canary and Azores Islands (see for instance Figs. 2a and 4a). Instead, for 2009 events, and due to the position and shape of the sensitive regions, two rectangular boxes are needed to cover the target areas, which also substantially differ from one event to the other, and even in the same event. As an example, for case 1025, two different areas capture the SV-based sensitive regions at that target time (see Figs. 2b and 4b): one box covers the Norwegian Sea and the areas around Iceland, and the other one covers the Western and Central Mediterranean.

3.2.3 AMDAR data

EUMETNET manages a specific observation program for providing AMDAR observations over Europe, North Atlantic and northern Africa. During both PREVIEW-DTS and MEDEX-DTS field campaigns some additional AMDAR observations were requested, but unfortunately the detailed information about which extra AMDAR observations were deployed is not available. For this reason, all AMDAR reports have been assimilated in control and data targeting experiments.

Aircraft included in the AMDAR program collect upper-level observations as well as vertical profiles (on ascent or descent phases) of air temperature and wind. Vertical profiles have similar accuracy to radiosondes and can be used in same way. In some regions AMDAR data provide the only available information on the vertical structure of the atmosphere. AMDAR data from en-route aircraft provide upper-level ob-

servations, which are valuable information, especially over oceanic regions and data-void regions.

Figure 4 shows that in Central Europe the density of AMDAR profiles is rather high, co-existing with a large number of radio sounding stations. In southern Europe the density of AMDAR vertical profiles is much smaller and in northern Africa those data are very sparse. Concerning upper-level data, the route from Europe to North America, along 50–60° N, is very well covered, but wide regions over the North Atlantic (south of 50° N) and North Africa lack of these data.

4 Results

This section presents the results obtained concerning the objective verification for the three experiments carried out with the different metrics described in Sect. 2.2. Upper-air forecasts are assessed using verifying analysis of EXP-RS experiment. Surface observations are used to obtain standard verification scores for surface pressure and several meteorological parameters. High-resolution rain gauge data are finally used to obtain a better assessment of QPF with the SAL method.

4.1 Upper-air forecasts performance assessment using EXP-RS verifying analysis

All the results obtained for those forecasts started at the targeting times which valid time corresponds to the verification time. Figures 5 and 6 and Table 2 show the main results obtained.

First, the large-scale flow forecast performance is assessed looking at 500 hPa geopotential height (Z500). The RMSE of Z500 forecasts of EXP and of the two data targeting experiments (EXP-RS and EXP-ATOVS*2) for all the studied cases is displayed in Fig. 5a. In some cases, the RMSE is very

Table 2. Number of improved cases and relative RMSE (in %, see text for details) for experiments EXP-RS and EXP-ATOVS*2 (relative to CONTROL) for different variables (Z = geopotential height, T = temperature, SPD = wind speed and RH = relative humidity) at 500 and 850 hPa, and for the cases of PREVIEW-DTS-2008, MEDEX-DTS-2009, and both.

	EXP	No. Improvements			Relative RMSE (%)		
		2008	2009	Both	2008	2009	Both
Z500	RS	9	5	14	−10.7	0.1	−7.9
	ATOVS*2	9	5	14	−22.0	−1.9	−16.6
T500	RS	7	7	14	−5.7	−1.5	−4.4
	ATOVS*2	10	7	17	−14.7	0.3	−9.6
SPD500	RS	6	9	15	−4.3	−3.6	−4.0
	ATOVS*2	8	7	15	−8.1	−3.8	−6.4
RH500	RS	8	9	17	−2.2	−2.4	−2.3
	ATOVS*2	9	5	14	−7.0	−3.0	−5.3
Z850	RS	5	3	8	−7.6	2.1	−4.4
	ATOVS*2	8	8	16	−15.4	−2.3	−10.9
T850	RS	8	8	16	−1.0	−1.6	−1.3
	ATOVS*2	8	8	16	−2.5	−1.0	−1.8
SPD850	RS	6	9	15	−4.6	−2.1	−3.6
	ATOVS*2	7	9	16	−6.0	−2.4	−3.6
RH850	RS	10	8	18	−3.0	−1.8	−2.5
	ATOVS*2	7	7	14	−1.1	−1.2	−1.2

large (as for example in cases 863 and 865 of event c), but in others RMSE is much lower (e.g. cases 1013 and 1014 of event e). REL_RMSE (relative to EXP, in %) for both EXP-RS and EXP-ATOVS*2 is shown in Fig. 5b. For Z500 the number of improvements (negative values) is slightly larger than that of deteriorations (positive values), as improvements account for 14 of 24 cases (58.3 %) for both DTEs. However, the overall magnitude of the improvements is larger. For this reason, the total RMSE of Z500 (that is when the forecast error is calculated for all the cases), is clearly reduced when targeted observations are used: −7.9 % for EXP-RS and −16.6 % for EXP-ATOVS*2 (see Table 2).

Figure 5c and d and Table 2 display the results for geopotential height in the vertical stratified for the PREVIEW and MEDEX cases. If we look separately at results for Z500 in 2008 and 2009, there is a clear distinction between the balance of improvements and degradations. For the 2008 PREVIEW cases, improvements dominate (9 of 12 cases, 75 %) and are large for both data targeting experiments (−10.7 % for EXP-RS and −22.0 % for EXP-ATOVS*2). But, for the 2009 MEDEX cases, the number of improvements is lower than that of deteriorations (5 of 12 cases, 41.7 %). Consequently, the REL_RMSE reflects neither improvement nor degradation for EXP-RS (0.1 %) and a small improvement for EXP-ATOVS*2 (−1.9 %). These results can be appreciated in Table 2 and in Fig. 5c and d, where the REL_RMSE for the geopotential height at different levels for the 2008 PREVIEW and the 2009 MEDEX cases are displayed. Figure 5c also shows a significant reduction of the forecast error
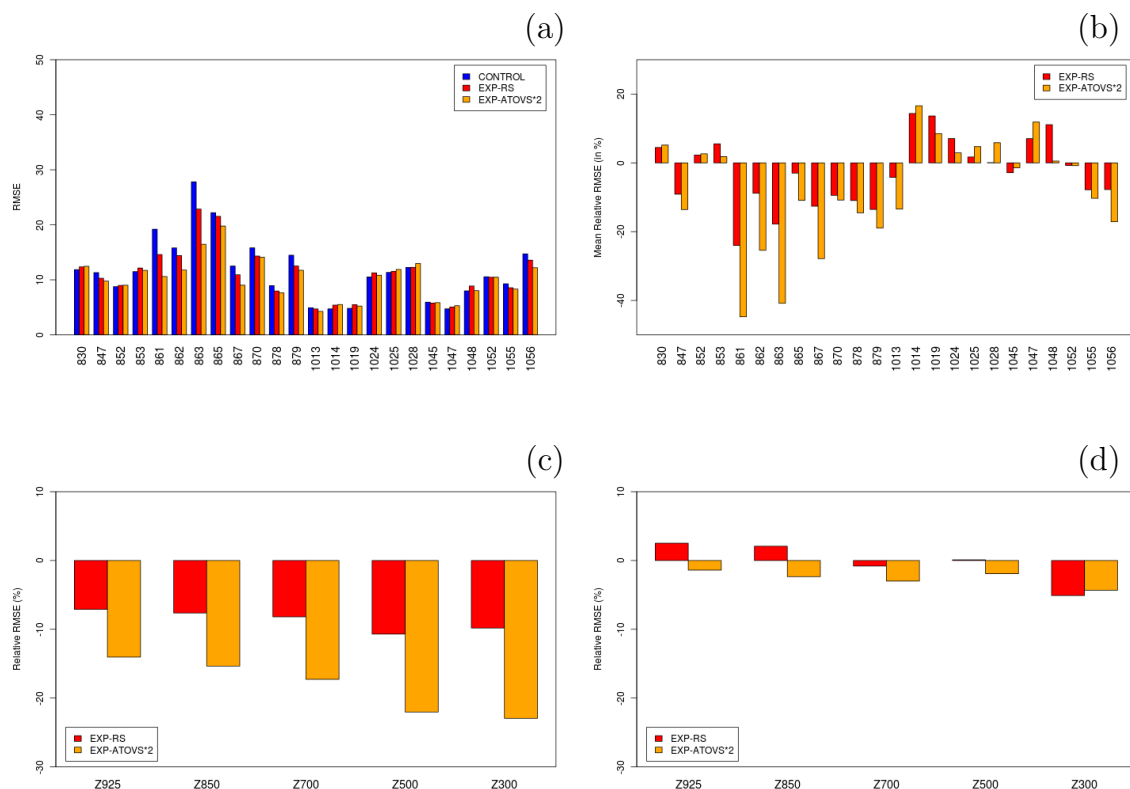
due to the assimilation of extra observations at all the vertical levels, especially at the upper ones, for the 2008 PREVIEW cases. Moreover, the improvement is larger when not only targeted radiosondes but also selective targeted satellite data are used. Oppositely, for 2009 cases, the assimilation of targeted radiosondes in the analyses slightly deteriorates the forecasts at all vertical levels, except at 300 hPa, while targeted satellite data is able to slightly improve the forecasts at all levels.

Looking in more detail, the results for Z500 at Fig. 5a and b for the two events described in Sect. 3.1: event c (PREVIEW) and event f (MEDEX), it is observed that both correspond to the largest RMSE of their respective field campaigns. However, targeted observations only benefited the short-range forecasts quality in 2008. RMSE of Z500 during event c was also the highest of all the cases considered in this study. The last column of Table 1 shows the number of extra radiosondes deployed for each case. In all cases belonging to event c, the number of targeted radiosondes was lower than that of those of event f. Overall it is difficult to appreciate any relationship between the number of extra observations deployed at each case and the impact or improvement obtained. It also seems that the optimization time is not a determinant factor for it (see Table 1 and Fig. 5a and b).

Forecast skill for the rest of variables is presented in Fig. 6 and Table 2. For 500 hPa temperature (T500) results are similar to those obtained for Z500 (see Table 2). The number of improvements is slightly larger than that of deteriorations (14 of 24, 58.3 %) for EXP-RS, but clearly larger (17 of 24, 70.8 %) for EXP-ATOVS*2. For T500 the forecast error is reduced −4.4 % when targeted radiosondes are assimilated and −9.6 % if, the selective satellite data are also used. Like for geopotential height, for T500 forecast error reduction is more effective for 2008 cases, with a reduction of −5.7 and −14.7 % for EXP-RS and EXP-ATOVS*2, respectively. Instead, for 2009 cases, the forecast error is slightly reduced (−1.5 %) for EXP-RS and slightly increased (0.3 %) for EXP-ATOVS*2.

For the same 500 hPa level, other variables such as wind speed (SPD500) and relative humidity (RH500) are also investigated (see Table 2). For SPD500, there are more improvements than deteriorations (15 of 24, 62.5 %) and the forecast error is significantly reduced for both data targeting experiments (−4.0 for EXP-RS and −6.4 % for EXP-ATOVS*2). For RH500, although the total number of improvements is similar or even larger than for other variables at the same level, the forecasts error reduction is lower: −2.3 and −5.3 % for EXP-RS and EXP-ATOVS*2.

Finally, verification scores are also obtained for the same variables but at different levels (see Fig. 6 and Table 2). It is found that the impact of data targeting, although positive, decreases at low levels. For instance, at 850 hPa the number of improvements is clearly greater than the number of deteriorations, but the forecast error reduction is lower than that for the 500 hPa counterparts. Overall, taking into account all

Fig. 5. Top: for each case, RMSE of 500 hPa geopotential height forecasts (in metres) for EXP, EXP-RS and EXP-ATOVS*2 experiments **(a)**, and relative RMSE (in %) of 500 hPa geopotential height forecasts for data targeting experiments EXP-RS and EXP-ATOVS*2 **(b)**. Bottom: relative RMSE (in %) of 925, 850, 700, 500 and 300 hPa geopotential height forecasts for data targeting experiments EXP-RS and EXP-ATOVS*2 for 2008 cases **(c)** and 2009 cases **(d)**.

the variables and all vertical levels, the average improvement for the selective satellite targeting is −6.5 % , approximately doubling that one obtained when only targeted radiosondes are assimilated (i.e. −3.6 %). This benefit is larger (lower) when 2008 (2009) cases are separately considered. While for MEDEX-DTS cases the average improvement is slight for both EXP-RS and EXP-ATOVS (−1.4 % and −1.7 % respectively) for PREVIEW-DTS cases the average improvement is significant for EXP-RS (−4.8 %), but especially for EXP-ATOVS2 (−8.9 %).

In summary, targeted observations are able to significantly reduce the forecasts error of upper-air variables. The improvement is larger when not only targeted radiosondes are assimilated, but also enhanced density satellite observations are assimilated in sensitive regions. Improvements are more noticeable for the 2008 PREVIEW cases. The benefits are large for geopotential height and temperature, and to a lesser extent, for wind speed and relative humidity. Finally, for all variables, the larger improvements are found in the medium to upper troposphere.

4.2 Objective verification of forecasts using SYNOP surface observations
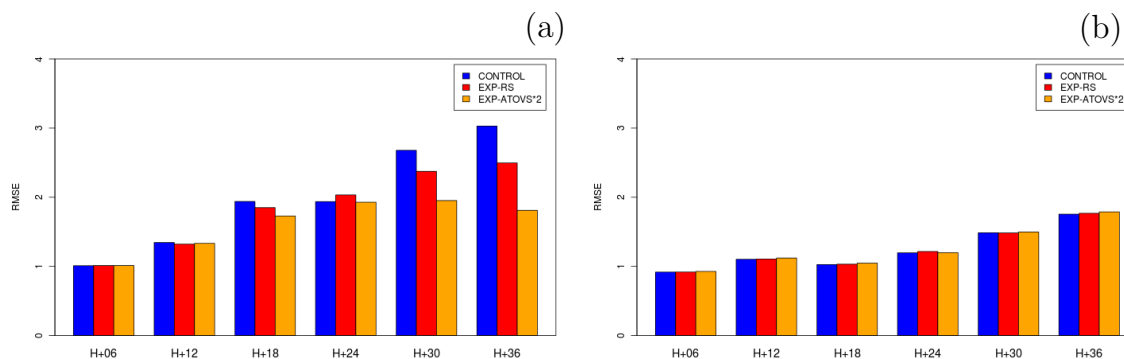
The forecasts performance during the high-impact weather events has also been assessed using surface observations reported by the meteorological stations included in the mobile VA at all the VTs (see Sect. 2.2.2). The objective verification scores for some relevant surface parameters like mean sea level pressure, surface wind and precipitation are presented here. They have been obtained for the two periods in 2008 (PREVIEW) and 2009 (MEDEX) separately.

4.2.1 Mean sea level pressure

For the 2008 PREVIEW cases, the assimilation of targeted observations is able to reduce RMSE: to improve forecasts (see Fig. 7a). The improvement is significant, and more evident when double-density satellite data in sensitive areas are assimilated. Moreover, improvements are larger for longer forecast lengths. However, for the 2009 MEDEX cases, (see Fig. 7b) neither an increase of targeted radiosondes in sensitive areas nor a higher density of ATOVS data assimilated in target regions produce a significant impact in surface pressure forecast skill.

Fig. 6. Relative RMSE (%) of 925, 850, 700, 500 and 300 hPa geopotential height **(a)**, temperature **(b)**, wind speed **(c)** and relative humidity **(d)** forecasts for data targeting experiments EXP-RS and EXP-ATOVS*2 for 2008 and 2009 cases.

Fig. 7. RMSE for MSLP forecasts (in hPa) against forecast length for EXP, EXP-RS and EXP-ATOVS for **(a)** PREVIEW-DTS cases and **(b)** MEDEX-DTS cases.

4.2.2 Surface wind

ETS has been calculated for different wind speed thresholds. For the 2008 PREVIEW cases, targeted observations seem to improve (degrade) the surface wind forecast for weak (moderate and strong) winds. On the other hand, for the 2009 MEDEX cases almost no impact is found due to any targeting strategy tried in EXP-RS and EXP-ATOVS*2 (not shown).

4.2.3 Precipitation

ETS has been calculated at different thresholds for 12 h accumulated precipitation (from 6 h before to 6 h after the verification time). Again, a different behaviour for PREVIEW and MEDEX cases is observed. In 2008, targeted observations clearly increase the forecast skill for all precipitation amounts (Fig. 8a). In the 2009 MEDEX cases, both EXP-RS and EXP-ATOVS*2 experiments seem to improve the control experiment EXP for light precipitation, but for moderate

and intense rainfall both experiments behave clearly worse than the control experiment (Fig. 8b).

4.3 SAL verification of QPF using high-resolution rain gauge data

A feature-based insight on QPF performance assessment follows here, trying to provide quantitative and detailed comparisons on key aspects of the precipitation field, to assess the impact of the targeting on QPF performance.

The SAL method has been applied to the set of forecast-observation field pairs. The information is summarized in the so-called SAL plot. It can help understand and assess the huge amount of information about precipitation patterns that SAL method provides. In a SAL plot, x and y axes depict S and A components, respectively, while L is the colour of the points in the plot (see Figs. 9–11). Any point in the plot describes SAL results for one comparison (in this study, a $500\,\text{km} \times 500\,\text{km}$ sub-domain of the VA). The dashed lines show the medians of S and A distributions, while the shadowed rectangle their Inter-Quartile-Range (IQR). The IQR for L is placed on the location colour legend inset, as three vertical black lines that depict 25th percentile, median, and 75th percentile. The contingency table in the down-right inset shows the joint distribution of forecasts and observations for rain – no rain in the whole domain, and reflects the ability of the model to capture the daily rainy – dry character.
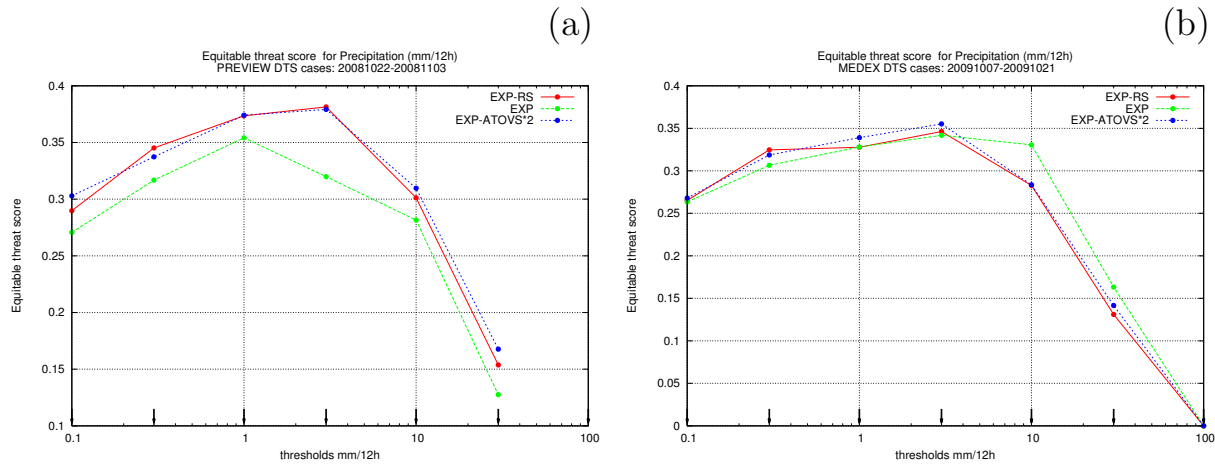
A first overall comparison between experiments is done gathering together in the same SAL plot all the cases (2008 and 2009) for each one of the experiments. Then, 2008 and 2009 cases for each experiment are separated in different SAL plots to further compare PREVIEW-DTS (2008) and MEDEX-DTS (2009) behaviour. Finally, some findings are commented for the different weather cases.

The first overall experiment comparison (gathering 2008 and 2009) is shown in Fig. 9a–c which correspond respectively to control (EXP), EXP-RS and EXP-ATOVS*2. For the three experiments, the overall location of the cloud of points is a first indication: they cover mainly quadrants I, III and IV leaving almost empty quadrant II (quadrants are defined here in the usual way, quadrant I with S and A positive, and then a counter-clockwise numeration). No one seems to be significantly more centred than the others. Looking at the shadowed rectangle that indicates the IQR for S and A, both are biased in all the experiments: S to the right (overestimation) and A downwards (underestimation), respectively indicating forecasts with larger and/or flatter structures than observed, and a negative bias in the amount of precipitation in the object space (most operational models show on average a positive bias in the amount of precipitation, see e.g. Navascués et al., 2013, but this study deals with a selection of extreme precipitation cases, thus no consistency with the average behaviour is expected on this aspect). Slight differences can be observed both in the median positions and in the rectangle sizes: while EXP-RS S median (dashed vertical line) compares to that one for EXP, EXP-ATOVS*2 one is a bit larger (0.3 with respect to 0.2), indicating a slightly worse representation of structure. EXP-ATOVS*2 rectangle size is smaller than that of EXP, and EXP-RS even smaller, thus showing a more consistent performance. Location components are also similar within eyeball significance, EXP-ATOVS*2 and EXP-RS improving EXP slightly (see solid black vertical lines on location colour legend). A clear difference between experiments comes from the number, position and colour of the outliers, those individual cases (points) laying far apart from the main cloud and/or with clear higher values of L component: while EXP shows 5 green points (location greater than 0.2) and a higher population in quadrant III, EXP-RS and EXP-ATOVS*2 presents only 3 green points and few points on quadrant III. Therefore, EXP-RS and EXP-ATOVS*2 provide a clear better location in the QPF objects predicted than that of EXP. The contingency table inset on the bottom-right corner shows in all the cases a perfect discrimination of dry-rainy days.

The second inter-comparison looks separately at PREVIEW-DTS (2008) and MEDEX-DTS (2009) campaigns in different SAL plots, for the purpose of splitting 2008 and 2009 points. Figure 10a–c show 2008 SAL plots for control (EXP), EXP-RS and EXP-ATOVS*2, respectively, while Fig. 11a–c show the same for 2009. Both 2008 and 2009 show overestimation of S and underestimation of A. For 2008 and according to the median S, EXP seems to better represent structures, though EXP-RS shows less variability in S performance (smaller IQR rectangle). For A, no apparent differences can be highlighted, except for less variability in EXP-RS. Concerning location, both EXP-RS and EXP-ATOVS*2 outperform EXP, with medians, 25th and 75th percentiles below EXP ones (solid black lines on location legends). The difference in number and behaviour of outliers described for 2008–2009 is similar here. For 2009, while there are no differences in the median S, the median A for EXP-RS is slightly better. In the location component, EXP performs slightly better than the others. The overall differences between 2008 and 2009 campaigns (regardless differences in experiment performance), are related to a different predictability in 2008 and 2009 cases as a whole, accordingly to their different flow regimes. The shapes of the clouds of points differ clearly, showing in the three experiments for 2008 a cloud that extends an arm down to quadrants III and IV (mainly A outliers, few S ones), while 2009 experiments show much more centred clouds (the absence of S and A outliers in this case corresponds possibly to a higher predictability).

Given that a very clear improvement for S and A components can not be shown from SAL plots, further diagrams have been plotted showing a case-by-case relative difference of these components (Fig. 12). Only those cases with a S or A difference greater than 0.1 (0.05 for L) have been plotted, avoiding as well those with very low S, A or L (very low denominators). For each graph, the abscissa

Fig. 8. Equitable threat score (ETS) for precipitation (in mm $12\,h^{-1}$) for EXP, EXP-RS and EXP-ATOVS for **(a)** PREVIEW-DTS cases and **(b)** and MEDEX-DTS cases.

corresponds to each case and the ordinate is the relative improvement/degradation (negative/positive). Rows in Fig. 12 correspond to S, A and L (a, b and c respectively) components while red dots depict EXP-RS and black crosses depict EXP-ATOVS*2 with respect to EXP relative differences. This is a way to reveal slight differences that can be important. Figure 12 shows that EXP-ATOVS*2 experiment produces more frequently an impact on precipitation. There are more improvements in S for both experiments, A is clearly improving in EXP-RS for most of cases and in EXP-ATOVS*2 improvements prevail. Location was rather similar in control and DT experiments for most of the cases studied; however, L is always improved when the experiments differ, in agreement with SAL plots. In general, it can be observed that the highest positive impact corresponds to some 2008 cases included in the event c described in Sect. 3.1. Those showing a larger impact (and mostly positive) on S and A parameters in 2009 took place at the end of the period studied, but not during the event f also presented in Sect. 3.1. However, it has to be taken into account that the available rain gauge data from Eastern Mediterranean countries were not enough to verify some of the high-impact cases corresponding to this event.
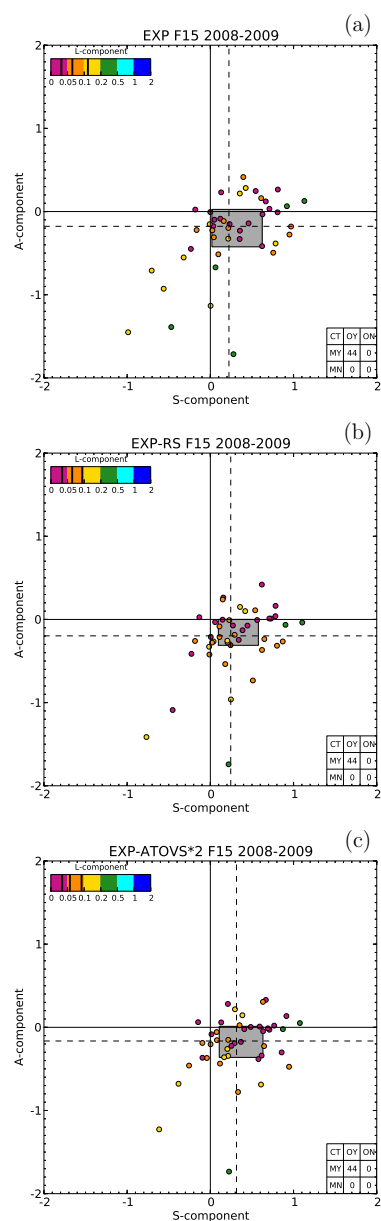
Finally, the sensitivity of SAL results to the selection of precipitation threshold has also been studied. In the SAL method this threshold is field-dependent; the precipitation objects are those areas of the precipitation field that exceed the threshold $R* = (1/f)\,R_{95}$, where R_{95} is the 95th percentile of the precipitation field. Values in the range 5–20 are often checked (Wernli et al., 2008) for this purpose. Here, similar ranges have been tested (not shown), yielding clear differences for values below 10 (camel effect present, see Wernli et al., 2008) but no significant sensitivity of SAL results with respect to values of f in the range 13–15. Therefore it was eventually selected $f = 15$ for the results shown.

5 Conclusions and discussion

In the present study, the impact of targeted observations on the short-range forecasts for a large set of case studies of high-impact weather events in the Mediterranean has been assessed. We have conducted a set of observing system experiments based on the HIRLAM NWP system. As targeted observations we have used (i) additional radiosondes derived from two-week periods of PREVIEW-DTS and MEDEX-DTS field campaigns, and (ii) enhanced satellite data in SV-based sensitive regions. Forecasts from the different experiments have been compared against both verifying analyses and surface observations. Furthermore, quantitative precipitation forecasts have been verified by means the novel feature oriented quality measure SAL. The main findings are summarized here as follows:

- Results show an impact of the extra observations assimilated. In general, targeted observations improve the forecast skill of high-impact weather events in the Mediterranean.

- The sign and magnitude of the impact highly depend on the case.

- The forecast skill is increased when not only targeted radiosondes but also enhanced satellite radiances selected with an adaptive observations screening are assimilated ($-3.6\,\%$ and $-6.5\,\%$, respectively).

- The improvement for geopotential height and temperature fields is higher than for wind speed and humidity. Moreover, at upper-levels, all the fields gain more benefit from targeted observations than at lower ones.

- Improvements are found to be larger for the PREVIEW-DTS (2008) cases than for the MEDEX-DTS (2009) cases (average improvements of $-1.7\,\%$
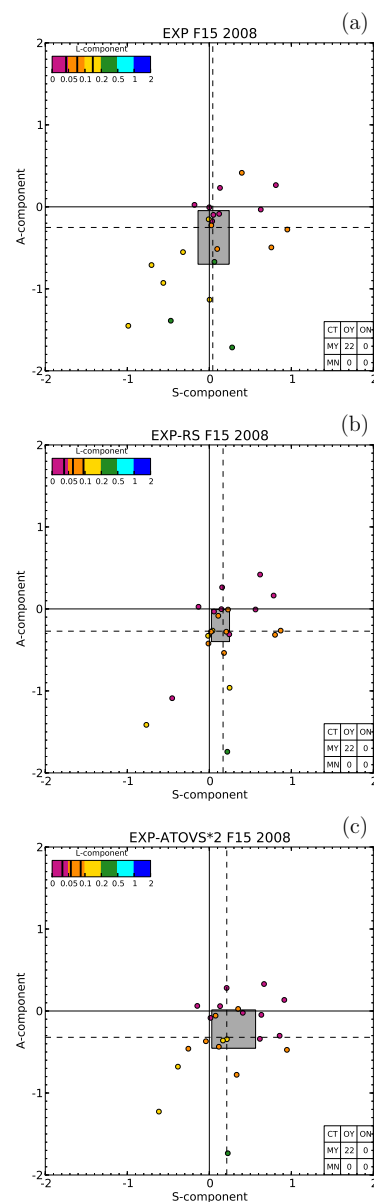
Fig. 9. SAL plots for the three experiments gathering together 2008 and 2009 cases, **(a)** control (EXP), **(b)** EXP-RS and **(c)** EXP-ATOVS∗2. Structure S and amplitude A components in x and y axes, respectively, location L in colour scale (legend). Dashed lines depict S and A medians, while the shadowed rectangle comprises the IQRs. Median, percentile 25 and 75 for L are vertical solid black lines in the legend (see text for details).

Fig. 10. Similarly to Fig. 9, SAL plots for the three experiments but now selecting only 2008 cases, **(a)** control (EXP), **(b)** EXP-RS and **(c)** EXP-ATOVS∗2.

and −8.9 %, respectively, for the experiment with enhanced satellite data).

– It is found that a large set of additional radiosondes does not imply a better forecast skill. The added value per targeted radiosonde is different for each case; they are more valuable when there are extra observations

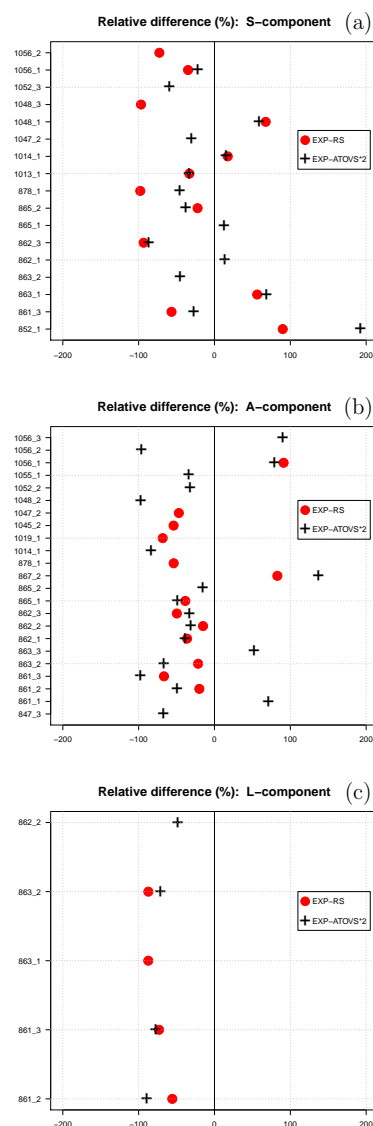located in sparse data areas and during periods of low predictability.

– The conclusions about experiment performance obtained with the different verification metrics applied both to upper-air variables and to surface meteorological parameters are in agreement. For all parameters studied, apart from surface wind, forecasts improve when targeted observations are assimilated.

– When spatial verification methods are applied to QPF (SAL was applied as a feature oriented method), the

Fig. 11. Similarly to Figs. 9 and 10, SAL plots for the three experiments now selecting only 2009 cases, **(a)** control (EXP), **(b)** EXP-RS and **(c)** EXP-ATOVS*2.

Fig. 12. Case-by-case relative difference of the three components S, A and L. X axes depict cases and y axes relative differences. Rows correspond to S, A and L (**a–c**, respectively). Relative differences (with respect to EXP) for EXP-RS (red dots) and EXP-ATOVS*2 (black crosses).

benefit of targeted observations is confirmed: both the amplitude and the location components of the precipitation patterns improve clearly (the improvement for S is also present), as well as the overall performance.

A positive impact due to targeted radiosonde observations is therefore observed in all variables and the addition of satellite targeted observations is able to double it. However, its magnitude is larger in mass variables and smaller in wind and humidity. The DA algorithms and error statistics are essential to determine the atmospheric initial state. Extra wind observations only come into the analysis from a few additional radiosondes, being the underlying error constraints in the DA (the adjoint of tangent linear of the forecast model and the background error statistics) also responsible for the wind analysis increments. Humidity analysis is less multivariate than that for mass and wind, and the length scales of its error statistics are much shorter. The number of humidity observations is considerably larger in the satellite targeted observation experiment, EXP-ATOVS*2 and a larger impact could be expected. However, the forecast error growth in humidity is not only due to the initial state but also to advection and other model errors.

The results obtained in this study are consistent with those of many other targeting observation field campaigns developed over different extratropical regions like the Fronts and Atlantic Storm Track Experiment (FASTEX; Montani et al., 1999), the North Pacific Experiment (NORPEX; Langland et al., 1999), the Winter Strom Reconnaissance Project (WSRP; Szunyogh et al., 2002) and the Atlantic-THORPEX Regional Campaign (A-TReC; Petersen and Thorpe, 2007). The present study is, to our knowledge, the first one to assess the impact of targeted observations on the forecasts over the Mediterranean region in an operational environment.

The improvement of forecast skill is appreciably larger for PREVIEW-DTS cases than for MEDEX-DTS ones, especially if not only additional radiosondes but also enhanced satellite data are assimilated. Two reasons can be argued to explain this difference. First, a different location of the sensitive regions which entail a distinct relationship between targeted observations and the rest of assimilated observations; and second, 2008 cases present lower predictability than 2009 ones (see for instance RMSE of 500 hPa geopotential in Fig. 5a).

As it would be expected, the existing composite observing system seems to be crucial for the impact produced by the targeted observations. In PREVIEW-DTS cases, targeted radiosondes are located in sparse areas, where few radiosondes (over remote islands or over the ocean as ASAPs) may greatly improve the forecast skill. However, in MEDEX-DTS, most of the targeted radiosondes are located in well-observed regions and this could be the reason for the small improvements found. Moreover, in this last case, sometimes targeted radiosondes can be redundant either by operational radiosondes or AMDAR ascents/descents. This is in agreement with some studies conducted by EUCOS to define a new upper-level network design. In these studies, OSEs were run for a number of scenarios of the composite observing system with different NWP models. Results show that removing radiosonde sites collocated close to 3-hourly visited E-AMDAR airports show almost no degradation in forecast skill (Kleinet and Klink, 2010).

However, when the target areas are oceanic, targeted radiosondes can not fully sample them. The assimilation of enhanced satellite observations (usually not used due to thinning) can overcome the under-sampling of these sensitive regions, and in principle, to diminish the initial condition uncertainty. Nevertheless, data assimilation is a complex process where the inclusion of new data heavily depends on the other assimilated data (as conventional radiosondes or AMDAR data). In PREVIEW-DTS, most of the sensitive maritime regions were located in subtropical latitudes, and hence in sparse data regions. Instead, in 2009 cases, sensitive regions spread over Central Europe and higher latitude oceanic areas which are better sampled, especially by AMDAR data. This might explain why improvements for EXP-ATOVS*2 are much larger for 2008 cases than for 2009 ones.

These findings concerning the impact of satellite data are in agreement with other observing system studies performed. Some intercomparison experiments to find the impact on short-range forecasts skill of the diverse observation types in different global forecast systems have been carried out recently (see Gelaro et al., 2010). In these experiments, the contribution of each observational datum to the forecast quality is found using an adjoint based method. The results obtained were consistent across the different NWP participating systems: on a global sense, the largest forecast error reductions are due to the assimilation of satellite radiances, and most of the improvement resulted from a large number of observations that had a small impact per observation.

Nonetheless, North Africa still remains as a large region with practically no upper-level data coverage. This fact may suppose a drawback in cases of high-impact weather in the Mediterranean, where sensitive regions are located in such regions. This is the case when a large upper-level trough disrupts in western North Africa and south-western Mediterranean, usually evolving to a cut-off (Homar et al., 2007), producing heavy precipitation in coastal regions.

As a consequence of a different large-scale flow in the two periods, predictability regimes were also rather different. During the MEDEX-DTS field campaign, the weather was rather predictable. This results in small uncertainties in the numerical weather forecasts and therefore there was little room for forecast improvement. On the other hand, during PREVIEW-DTS field campaign the atmosphere was less predictable and thus a larger improvement can be expected from targeted observations. Buizza et al. (2007) found that the observations taken in SV-based sensitive regions over the Atlantic Ocean are on average more valuable on short-range forecasts over Europe than observations taken in randomly selected areas. If the baseline observing system is data void (data rich) over the Atlantic, then the average value of observations taken in SV-based target areas is very high (rather small) over Europe. But that investigation does not exclude the possibility that during certain weather regimes a larger value of targeting in SV-based areas may occur. Cardinali et al. (2007), in a successive paper of 3-part series, showed that in certain weather regimes, particularly during periods of tropical to extratropical transitions, the removal of observations in SV-based sensitive regions degraded the skill of 2 day forecasts over Europe significantly more than randomly selected observations. Thus, it seems that to identify a priori cases in which large impacts are expected should result in a more effective use of observations. Studies to diagnose and understand predictability, and the characteristics of forecast error propagation and growth in the context of evolving observing systems and data assimilation seem to be necessary, especially in low predictability regimes (Majumdar et al., 2011).

The procedure of selecting targeted observations is complex and imperfect. An important issue is predicting the optimal time and location for targeting, prior to the deployment

of observations (Majumdar et al., 2011). As it is mentioned in Sect. 3, SVs and ETKF techniques were used during the PREVIEW campaign, and only the SV-based one were employed for MEDEX. The influence of potential differences due to the application of different sensitive area calculation/prediction methods was not addressed in this study. A simultaneous work by Garcies and Homar (2013) presents a theoretical assessment of the targeting guidance provided by different sensitivity methods based on single sounding assimilation experiments over a particular MEDEX-DTS case study.

At last, it should be mentioned that the impact due to targeted observations may strongly depend on the assimilation algorithms and the usage of observations. A first round of experiments over the MEDEX period was carried out with the HIRLAM 3-D variational data assimilation (Gustafsson et al., 2001; Lindskog et al., 2001), and the impact found was much smaller than the results presented in this paper (not shown). This fact points to the importance of flow dependency to fully exploit the potential of targeted observations. The sensitivity to the thinning strategy of raw satellite data demonstrates how the observations usage within the data assimilation also plays a crucial role on it.

Edited by: G. Panegrossi
Reviewed by: two anonymous referees

References

Bauer, P., Buizza, R., Cardinali, C., and Thépaut, J.-N.: Impact of singular-vector-based satellite data thinning on NWP, Q. J. Roy. Meteor. Soc., 137, 286–302, doi:10.1002/qj.733, 2011.

Bech, J., Pineda, N., Rigo, T., Aran, M., Amaro, J., Gayà, M., Arús, J., Montanyà, J., and van der Velde, O.: A Mediterranean nocturnal heavy rainfall and tornadic event. Part I: Overview, damage survey and radar analysis, Atmos. Res., 100, 621–637, doi:10.1016/j.atmosres.2010.12.024, 2011.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, Mon. Weather Rev., 129, 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2, 2001.

Buizza, R. and Montani, A.: Targeting observations using singular vectors, J. Atmos. Sci., 56, 2965–2985, doi:10.1175/1520-0469(1999)056<2965:TOUSV>2.0.CO;2, 1999.

Buizza, R., Cardinali, C., Kelly, G., and Thépaut, J.-N.: The value of observations. II: The value of observations located in singular-vector-based target areas, Q. J. Roy. Meteor. Soc., 133, 1817–1832, doi:10.1002/qj.149, 2007.

Cardinali, C., Buizza, R., Kelly, G., Shapiro, M., and Thépaut, J.-N.: The value of observations. III: Influence of weather regimes on targeting, Q. J. Roy. Meteor. Soc., 133, 1833–1842, doi:10.1002/qj.148, 2007.

Cherubini, T., Ghelli, A., and Lalaurette, F.: Verification of Precipitation Forecasts over the Alpine Region Using a High-Density Observing Network, Weather Forecast., 17, 238–249, doi:10.1175/1520-0434(2002)017<0238:VOPFOT>2.0.CO;2, 2002.

Dando, M. L., Thorpe, A. J., and Eyre, J. R.: The impact of targeted satellite observations on numerical weather prediction, Q. J. Roy. Meteor. Soc., 133, 1945–1957, doi:10.1002/qj.176, 2007.

Davis, C., Brown, B., and Bullock, R.: Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas, Mon. Weather Rev., 134, 1772–1784, doi:10.1175/MWR3145.1, 2006.

Fresnay, S., Hally, A., Garnaud, C., Richard, E., and Lambert, D.: Heavy precipitation events in the Mediterranean: sensitivity to cloud physics parameterisation uncertainties, Nat. Hazards Earth Syst. Sci., 12, 2671–2688, doi:10.5194/nhess-12-2671-2012, 2012.

García-Moya, J. A., Callado, A., Escribà, P., Santos, C., Santos-Muñoz, D., and Simarro, J.: Predictability of short-range forecasting: a multimodel approach, Tellus A, 63, 550–563, doi:10.1111/j.1600-0870.2010.00506.x, 2011.

Garcies, L. and Homar, V.: Are current sensitivity products sufficiently informative in targeting campaigns? A DTS-MEDEX-2009 case study. Q. J. Roy. Meteor. Soc., doi:10.1002/qj.2148, 2013.

Gelaro R., Langland, R. H., Pellerin, S., and Todling, R.: The THORPEX Observation Impact Intercomparison Experiment, Mon. Weather Rev. 138, 4009–4025, doi:10.1175/2010MWR3393.1, 2010.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E.: Intercomparison of Spatial Forecast Verification Methods, Weather Forecast., 24, 1416–1430, doi:10.1175/2009WAF2222269.1, 2009.

Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M., Navascués, B., Mogensen, K.-S., and Thorsteinsson, S.: Three-dimensional variational data assimilation for a limited area model. Part I: General formulation and the background constraint, Tellus A, 53, 425–446, doi:10.1111/j.1600-0870.2001.00425.x, 2001.

Gustafsson, N., Huang, X. Y., Yang, X., Mogensen, K., Lindskog, M., Vignes, O., Wilhelmsson, T., and Thorsteinsson, S.: Four-dimensional variational data assimilation for a limited area model, Tellus A, 64, 14985, doi:10.3402/tellusa.v64i0.14985, 2012.

Homar, V. and Stensrud, D. J.: Sensitivities of an intense Mediterranean cyclone: Analysis and validation, Q. J. Roy. Meteor. Soc., 130, 2519–2540, doi:10.1256/qj.03.85, 2004.

Homar, V., Ramis, C., and Alonso, S.: A deep cyclone of African origin over the Western Mediterranean: diagnosis and numerical simulation, Ann. Geophys., 20, 93–106, doi:10.5194/angeo-20-93-2002, 2002.

Homar, V., Jansà, A., Campins, J., Genovés, A., and Ramis, C.: Towards a systematic climatology of sensitivities of Mediterranean high impact weather: a contribution based on intense cyclones, Nat. Hazards Earth Syst. Sci., 7, 445–454, doi:10.5194/nhess-7-445-2007, 2007.

Jansà, A., Arbogast, P., Doerenbecher, A., Garcies, L., Genovés, A., Homar, V., Klink, S., Richardson, D., and Sahin, C.: A new approach to sensitivity climatologies: the DTS-MEDEX-2009 campaign, Nat. Hazards Earth Syst. Sci., 11, 2381–2390, doi:10.5194/nhess-11-2381-2011, 2011.

Jolliffe, I. T. and Stephenson, D. B. (Eds.): Forecast Verification: A Practitioner's Guide in Atmospheric Science, Wiley, 274 pp., 2011.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, J. Appl. Meteor., 43, 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Kleinet, T. and Klink, S.: Upper-air network redesign. EUCOS/TS/2010-002,18/09/2010 (available at: http://www.eucos.net/cln_033/nn_779844/EN/Home/restricted/Downloads/studies/Upper__Air__redesign/Final__doc__UANR,templateId=raw,property=publicationFile.pdf/Final_doc_UANR.pdf), 2010.

Langland, R. H.: Issues in targeted observing, Q. J. Roy. Meteor. Soc., 131, 3409–3425, doi:10.1256/qj.05.130, 2005.

Langland, R. H., Toth, Z., Gelaro, R., Szunyogh, I., Shapiro, M. A., Majumdar, S. J., Morss, R. E., Rohaly, G. D., Velden, C., Bond, N., and Bishop, C. H.: The North Pacific Experiment (NORPEX-98): Targeted Observations for Improved North American Weather Forecasts, B. Amer. Meteorol. Soc., 80, 1363–1384, doi:10.1175/1520-0477(1999)080<1363:TNPENT>2.0.CO;2, 1999.

Lindskog, M., Gustafsson, N., Navascués, B., Mogensen, K.-S., Huang X.-Y., Yang, X., Andrae, U., Berre, L., Thorsteinsson, S., and Rantakokko, J.: Three-dimensional variational data assimilation for a limited area model. Part II: Observation handling and assimilation experiments, Tellus A, 53, 447–468, doi:10.1111/j.1600-0870.2001.00447.x, 2001.

Majumdar, S. J., Aberson, S. D., Bishop, C. H., Cardinali, C., Caughey, J., Doerenbecher, A., Gauthier, P., Gelaro, R., Hamill, T. M., Langland, R. H., Lorenc, A. C., Nakazawa, T., Rabier, F., Reynolds, C. A., Saunders, R., Song, Y., Toth, Z., Velden, C., Weissmann, M., and Wu, C.-C.: Targeted observations for improving numerical weather prediction: an overview, World Weather Research Programme/THORPEX Publication No. 15, 37 pp., 2011.

Montani, A., Thorpe, A. J., Buizza, R., and Undén, P.: Forecast skill of the ECMWF model using targeted observations during FASTEX, Q. J. Roy. Meteor. Soc., 125, 3219–3240, doi:10.1002/qj.49712556106, 1999.

Navascués, B., Calvo, J., Morales, G., Santos, C., Callado, A., Cansado, A., Cuxart, J., Díez, M., del Río, P., Escribà, P., García-Colombo, O., García-Moya, J. A., Geijo, C., Gutiérrez, E., Hortal, M., Martínez, I., Orfila, B., Parodi. J. A., Rodríguez, E., Sánchez-Arriola, J., Santos-Atienza, I., and Simarro, J.: Long-term verification of HIRLAM and ECMWF forecasts over Southern Europe: History and perspectives of Numerical Weather Prediction at AEMET, Atmos. Res., 125–126, 20–33, doi:10.1016/j.atmosres.2013.01.010, 2013.

Nuissier, O., Ducrocq, V., Ricard, D., Lebeaupin, C., and Anquetin, S.: A numerical study of three catastrophic precipitating events over southern France. I: Numerical framework and synoptic ingredients, Q. J. Roy. Meteor. Soc., 134, 111–130, doi:10.1002/qj.200, 2008.

Petersen, G. N. and Thorpe, A. J.: The impact on weather forecasts of targeted observations during A-TreC, Q. J. Roy. Meteor. Soc., 133, 417–431, doi:10.1002/qj.31, 2007.

PaiMazumder, D. and Mölders, N.: Theoretical Assessment of Uncertainty in Regional Averages due to Network Density and Design, J. Appl. Meteorol. Clim., 48, 1643–1666, doi:10.1175/2009JAMC2022.1, 2009.

Prates, C., Richardson, D., and Sahin, C.: Final report on the PREVIEW observation Data targeting System (DTS), ECMWF Tech. Memo. 581, 2009.

Ramis, C., Romero, R., Homar, V., Alonso, S. , and Alarcón, M.: Diagnosis and numerical simulation of a torrential precipitation event in Catalonia (Spain), Meteorol. Atmos. Phys., 69, 1–21, doi:10.1007/BF01025180, 1998.

Reiter, E. R.: Handbook for Forecasters in the Mediterranean; Weather Phenomena of the Mediterranean Basin, Navy Weather Research Facility, Tech. Pap. 5/75, 1975.

Rudari, R., Entekhabi, D., and Roth, G.: Terrain and multiple-scale interactions as factors in generating extreme precipitation events, J. Hydrometeorol., 5, 390–404, doi:10.1175/1525-7541(2004)005<0390:TAMIAF>2.0.CO;2, 2004.

Schyberg, H., Landelius, T., Thorsteinsson, S., Tveter, F. T., Vignes, O., Amstrup, B., Gustafsson, N., Harvinen, H., and Lindskog, M.: Assimilation of ATOVS Data in the HIRLAM 3-D-Var System. HIRLAM Tech. Rep. Nr 60, available at: http://www.hirlam.org/index.php?option=com_docman&task=doc_download&gid=260&Itemid=71, 2003.

Szunyogh, I., Toth, Z., Zimin, A., Majumdar, S. J., and Persson, A.: Propagation of the Effect of Targeted Observations: The 2000 Winter Storm Reconnaissance Program, Mon. Weather Rev., 130, 1144–1165, doi:10.1175/1520-0493(2002)130<1144:POTEOT>2.0.CO;2, 2002.

Undén, P., Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., García-Moya, J.-A., Jones, C., Lenderlink, G., McDonald, A., McGrath, R., Navascués, B., Nielsen, N.-W., Odegaard, V., Rodríguez, E., Rummukainen, M., Rõõm, R., Sattler, K., Sass, B.-H., Savijärvi, H., Schreur, B.-W., Sigg, R., The, H., and Tijm, A.: HIRLAM-5 Scientific Documentation, available from SMHI, S-60176 Norrköping, Sweden, 2002.

Vié, B., Nuissier, O., and Ducrocq, V.: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on initial conditions and lateral boundary conditions, Mon. Weather Rev., 139, 403–423, doi:10.1175/2010MWR3487.1, 2011.

Wernli, H., Paulat, M., Hagen, M., and Frei, C.: SAL – A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts, Mon. Weather Rev., 136, 4470–4487, doi:10.1175/2008MWR2415.1, 2008.

Wernli, H., Hofmann, C., and Zimmer, M.: Spatial Forecast Verification Methods Intercomparison Project: Applica-

tion of the SAL Technique, Weather Forecast., 24, 1472–1484, doi:10.1175/2009WAF2222271.1, 2009.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Volume 100, Third Edition (International Geophysics), Academic Press, 2011.

Zhang, M., Lin, W., Bretherton, C. S., Hack, J. J., and Rasch, P. J.: A modified formulation of fractional stratiform condensation rate in the NCAR Community Atmospheric Model (CAM2), J. Geophys. Res, 108, 4035, doi:10.1029/2002JD002523, 2003.