Reviewer 2

This study comprehends an interesting analysis of the impact of assimilating 4 different observation systems in the AROME-WMED reanalysis for the Autumn 2012, period of the first HyMeX SOP.
The main results of this publication include the good performance of the GNSS-ZTD Data Assimilation (3DVar), the improvement gained over the Iberian Peninsula due to the assimilation of the Spanish radar network and the weak impact of assimilating the wind profiles and Lidar measurements. Also noteworthy, is the weak but positive impact of GNSS Data Assimilation on wind correction given this measurement system provides information on the integrated atmospheric moisture column.
Even though I consider this manuscript has potential for publication, it must undergo major revisions to have a sufficient quality.
There are two overall problems. First, the composition and structure need to be substantially improved as the main guideline of the paper is not clearly shown. Second, an overarching conclusion encompassing the findings for each of the observations systems is needed.

We thank Reviewer 2 for his/her comments which helped to improve, we hope, the quality of the manuscript. Please find below our response to your comments. Reviewer 2's comments are in bold font, our answers are written with normal font.

Regarding the composition of the paper, the introduction/justification for this particular manuscript is not highlighted. It is clear that it has been produced in the framework of the AROME-WMED second reanalyses. But the reason why these observations impact studies were carried out is not said nor in the abstract, or in the introduction. The abstract would be more attractive if at least it was mentioned why these four observation types were selected. To this end it would be useful using the sentence in the first paragraphs of the conclusions "Previous studies such as Duffourg and Ducrocq (2011) Ricard et al. (2012) or Bresson et al. (2012), have shown the interest of an accurate description of the low-level moist flow feeding mesoscale convective systems. In this study the impact of various data set related to humidity and wind on the forecast quality from this comprehensive reanalysis is investigated over the 2-month period".
The authors agree with the reviewer that the rationale for this study and the types of observations selected were not sufficiently explained in the text.
The first paragraph of the abstract has been reformulated to give the explanation of the selection of these 4 observation types: A reanalysis with a convective model AROME-WMED was performed which assimilated most of all available data for a 2 month period corresponding to the first Special Observation Period of the field campaign (Fourrié et al., 2019). Among them, observations related to the low level humidity flow which are important for the description of the feeding of the convective mesoscale systems with humidity (Duffourg and Ducrocq, 2011, Bresson et al., 2012 and Ricard et al.,2012), were assimilated. Among them there were a dense reprocessed network of high quality Global Navigation Satellite System (GNSS) Zenithal Total Delay (ZTD) observations, reprocessed data from wind profilers, lidar-derived vertical profiles of humidity (ground and airborne) and Spanish radar data.
It has also been rearranged following the suggestion of the reviewer given in the specific comments.
Sentences of Lines 51-55 have been changed to justify the choice of the four studied observation types:
As previously mentioned, an accurate description of the low-level humidity flow is required to well simulate the evolution of the mesoscale system. The aim of the study presented here is to quantify the impact of four observation systems on the quality of precipitation simulation. These observation data sets were assimilated in the AROME-WMED reanalysis of SOP1 and provided information on this low-level flow. The observations are the reprocessed ZTD from the ground based GNSS (Bock et al., 2016), the humidity profiles from ground based and airborne lidars (Chazette et al., 2016 and Di Girolamo et al., 2016), reprocessed wind profiler data (Saïd et al., 2016) and the Doppler winds and reflectivities from the Spanish radars. To achieve this, a number of denial data

assimilation experiments, consisting in removing one observation type, were carried out during the 2-month period of SOP1.

Second, the methodology of the experiments is chaotically explained. Throughout the manuscript different denominations are used for the same simulation. Hence, I would advise sticking to one nomenclature for the simulations, First Guesses (FG), Analysis (AN) and the different data denial experiments. In section 2, the logic order of presenting the experiments would be, first a brief description of the experiments, then the model, then observations and finally the main concept of the data denial experiments and its nomenclature.

As suggested by the reviewer, section 2 was modified. Here is the new structure of the section:

1. Observing System Experiment Methodology
2. AROME-WMED configuration
3. Description of the studied observing systems
   - 3.1. GNSS Zenithal Total Delays
   - 3.2. Wind profilers
   - 3.3. Lidars
   - 3.4. Spanish radars
4. Description of the experiments

Moreover, the text was checked to have more consistency between the different denomination concerning the model and the experiments. The last subsection describes in more details the different experiments and the associated nomenclature (Please see specific comments).

Regarding the conclusions, <u>an overarching statement as to what would be the best(s) observation system(s) to use in future Data Assimilation experiments or operations is needed.</u> This is crucial as only rarely there is such a high availability of different observations. In most of the cases efforts and resources must be concentrated and this paper would be helpful in providing some guidance for decision making. For instance, in case of having to choose between one of the four observation systems, which one would bring more added value? In the assimilation procedure should some of these observations be given more weight than others? Is the temporal frequency of any of these playing a special role?

Thank you for point out the interest of our study in providing guidance for decision making. In case of having to choose between one of the four observation systems, we have shown the clear interest of the assimilation of GNSS ZTD data because it represents a frequent data set well spread over the domain. The comparison of the impact between NOWPROF, NORADSPAIN and NOLIDAR suggest that the frequent availability of the data could play an important role to get a significant impact on the forecast. In summary, both temporal and horizontal availability are needed to influence the analysis and then the forecast. A paragraph on the main conclusions of our study had been added in the conclusion.

"With the examination of the impact of the assimilation of 4 different data sets over a two-month period in the meso-scale AROME-WMED, our study shows that it is required to have well spatially distributed and frequent data sets such as the GNSS ZTD data set to get, with its assimilation, an overall impact in terms of analysis and forecast skills. This result agrees with the findings of Mahfouf et al. (2015) who show that the assimilation of GNSS systematically improves the atmospheric humidity short-range forecasts despite the small fraction of GNSS observations assimilated in AROME. A high temporal availability and a regular horizontal distribution are both needed to get a significant impact on the forecast scores. When the data set is available frequently but not well spread over the model domain such as the Doppler winds and reflectivities from the Spanish radars or winds from profiler radars, its assimilation may lead to a positive impact on the precipitation forecast but it remains local. Finally, marginal impact from local and sporadic data sets such as humidity profiles from water vapour Lidars can be obtained but it is not visible on "global scores". To get a material impact on the forecast in a mesoscale model from a set of observation through its data assimilation, our study suggests to select data sets which are frequently available at each analysis time and also well spread over the domain."

Concerning the assimilation procedure, the good results with the GNSS and the Spanish radar data suggests to follow efforts to improve their assimilation. Since 2019, the GNSS are bias corrected with an adaptive bias correction updated through the minimisation, which has improved their assimilation and their impact. Since Summer 2020, Spanish radar data (and data from other countries) are assimilated through the OPERA, the European radar programme of EUMETNET. This OPERA processing allows to get higher quality data in the assimilation and to enhance the impact of the foreign radars.
The impact of the above mentioned data could be further improved. For example, the impact of GNSS in AROME-France has been recently improved with the use of variational bias correction in replacement of the static bias correction used in this study (P Moll, personal Communication). In addition radar data from foreign countries are now assimilated in AROME since July 2020. The distribution of these data by the OPERA (the EUMETNET Radar programme) allows to get data of high quality in the data assimilation and thus to increase their impact in the AROME model (Martet et al, 2019).

Specific comments

Title - I would strongly advice including the word assimilation in the title. "Data Assimilation Impact studies with AROME-WMED reanalysis of the HyMeX SOP1"

The title of the paper has been modified according the suggestion of Reviewer.

Abstract

(L01-05) – In addition to the missing appropriate justification of this paper (see general comments), the introduction is too specific on the terminology of HyMeX. Using acronyms such as OSEs, AROME and HYMEX might be familiar to readers in the community but not necessarily to a broader audience. I would suggest starting the abstract mentioning the global topics of the paper: Data assimilation, the four observation systems and the validation of first guesses/analyses and forecast range.

The first paragraph of the abstract was rephrased as followed: This study was performed in the frame of HyMeX (Hydrological cycle in the Mediterranean Experiment) which aimed to study the heavy precipitation that regularly affects the Méditerranean area. A reanalysis with a convective model AROME-WMED was performed which assimilated most of all available data for a 2 month period corresponding to the first Special Observation Period of the field campaign (Fourrié et al., 2019). Among them, observations related to the low level humidity flow which are important for the description of the feeding of the convective mesoscale systems with humidity (Duffourg and Ducrocq, 2011, Bresson et al., 2012 and Ricard et al.,2012), were assimilated. Among them there were a dense reprocessed network of high quality Global Navigation Satellite System (GNSS) Zenithal Total Delay (ZTD) observations, reprocessed data from wind-Profilers, lidar-derived vertical profiles of humidity (ground and airborne) and Spanish radar data. The aim of the paper is to assess the impact of the assimilation of these four observation types on the analyses and the forecasts from the 3h forecast range (first guess) up to the 48-h forecast range. In order to assess this impact, several OSEs or also-called denial experiments, were carried out by removing one single data set from the observation data set assimilated in the reanalysis.

L51-52 - What is discussed in the paper is really the impact of 4 observation systems (GNSS, LIDAR, Wind Profiler, Spanish radars) not "the many observation data sets which were assimilated" in the second reanalysis. Please, rephrase to constrain appropriately the scope of the paper. Moreover, "Quantify the contribution" sounds ambiguous. The aim could be rephrased to "quantify the impact of four observation systems on the quality of precipitation simulation". Finally, it would be appropriate writing that the impact of the data denial experiments is studied for the quality in the simulations and not on processes. Which might be understood from the title of the paper.

Thanks for this remark the sentence was changed as suggested: The aim of the study presented here is to quantify the impact of four observation systems on the quality of precipitation simulation.

L85-91 - As explained in the general comments the general data denial procedure and the nomenclature have to be well introduced and explained in this section and kept during the remainder of the manuscript. More clarity is needed in the nomenclature.

As suggested by Reviewer Section 2 was written with a different outline. The first paragraph introduces the concept of denial experiments.

To study the contribution of the observations on the analysis and forecast quality of the heavy precipitating events of the SOP1, denial experiments have been devised. These experiments consist of removing one observation data set and to compare the forecast quality with the one originating from assimilating all the observations. Here, denial experiments were conducted on the following four observation types: the ground-based GNSS ZTD, the wind profilers, the water vapour lidars and Spanish radars. They were performed with the AROME-WMED model.

Now the description of the data denial experiments is as follows. :

"Table 2 summarizes the names of the denial experiments and the observations considered. Five experiments were conducted over the 2-month period of SOP1 (from 5 September 2012 to 5 November 2012). They all used the same configuration of AROME-WMED, the differences lying in the observations assimilated. For each experiment, it differs only one observation type from the reanalysis (REANA) used as the reference. This allows to evaluate the impact of this observation type on the analysis and the forecast. Among the five experiments, two experiments deal with the impact of GNSS ZTD. The first one, NOGNSS is obtained by removing the GNSS ZTD from the assimilation. The second, called OPERGNSS, aimed to evaluate the impact of the reprocessed data set provided by (Bock et al., 2016) compared to the operational data set provided by E-GVAP. The E-GVAP data set was thus assimilated in replacement of the Bock et al., (2016)'s one in OPERGNSS. The NOLIDAR experiment is the run with no airborne nor ground-based Lidar data in the data assimilation. The NOWPROF experiment is obtained by removing the wind profiler data and the NORADSPAIN experiment was run without any data from the five Spanish radars."

2.2 Observing System Experiment Description – Some relevant information is missing regarding the description of the observational data sets and its assimilation procedure. For example, for GNSS a short description of what is ZTD and its relationship to humidity would be desirable. For Wind profiles, some notions on what is the measurement technique is needed. The same applies for the Spanish radar. Overall more details on how these variables are assimilated is advised. For example, no information is given about the forward operators and specific prognostic variables for each of the 4 data observation systems.

GNSS ZTD provides useful information on precipitable water and pressure in all weather conditions at a high temporal frequency. Information about their assimilation was also added in the text as shown below:

GNSS ZTD provides useful information on precipitable water and pressure at a high temporal frequency and in all weather conditions. In REANA2, we considered here reprocessed data (REPRO-GNSS in the following) with a homogeneous reprocessing using a single software and more precise satellite orbits position and clocks (Bock et al., 2016), which were available for the whole SOP1. Additional data were also considered compared to the operational and data set available in near real-time. This data set, called hereafter OPER-GNSS, is provided by E-GVAP(EUMETNET EIG GNSS (Global Navigation Satellite System) water vapour programme and ZTD data for one reception station may be available for more than 10 processing centres. These ZTD data are assimilated according the methodology described in Mahfouf et al. (2015). The model equivalent is computed with the following equation Mahfouf et al. (2015):

$$ZTD = 10^{-6}\int_{0}^{z_{top}} (k_1\frac{p}{T} + k_3\frac{e}{T^2})dx$$

where p is the pressure, T the temperature, e the water vapour pressure, k1 = 0.776 Pa^{-1} K, and k3=3730 Pa^{-1} K^2 , x is the height above the ground and z_{top} is the height of the model top. After a monitoring of the difference between observations and model equivalent, observations with good statistics are selected in a 'white list'. ZTD data are also bias corrected and an updated bias correction for each GNSS station was also computed in the REANA2 version. They are finally assimilated if they pass the first guess quality control which rejects data too far from the model background equivalent. Only one observation per 3-h assimilation and per surface station is assimilated for each analysis. Please refer to Mahfouf et al. (2015) for more information on the data assimilation of GNSS ZTD in AROME.

Concerning wind from profilers
The data of 8 wind profiler radars were considered in the reanalysis. These profilers provided vertical profiles of wind vector that were assimilated after a quality control reprocessing performed by Saïd et al (2016) to remove spurious data. The paragraph on wind profilers was extended as follows:
Data from eight wind profiler radars (sounding in VHF or UHF bands) were assimilated in AROME-WMED. These profilers provided vertical profiles of wind vector, turbulence, precipitation and the height of the atmospheric boundary layer and tropopause (Saïd et al., 2016). The measure principle is described in Annex 1 of Saïd et al. (2016)'s paper. Profilers measure the Doppler radial spectra of the returned signal backscattered by various types of targets. In order to derive the three components of the wind, most of the HyMeX profilers use five beams. These data were available for the whole SOP1 in real-time and have been 125 reprocessed after the SOP1 by Saïd et al., (2016) with an improved quality control to remove spurious data. Here, observations from 8 wind radars (UHF and VHF) mainly located in the South of France, in Corsica and Menorca (Figure 1) were considered. These observations are assimilated as vertical profiles of horizontal wind.

Spanish radars:
As for the previous observation types, more information on data and the way they are assimilated is given as shown below:
"Doppler radial winds and reflectivities from five Spanish radars, located in Barcelona, Valencia, Almeria, Murcia, Palma de Mallorca and provided by AEMET were assimilated in REANA. After a strict quality control check to exclude data with gross errors, only the three lowest elevations have been considered for the assimilation. Doppler wind are assimilated in the 3D-Var of AROME according the method described by Montmerle and Faccani (2009) and reflectivity data are assimilated as pseudo-observations of relative humidity profiles as proposed in Caumont et al. (2010) and implemented in Wattrelot et al. (2014).
Several procedures are applied to raw data in order to avoid as much as possible erroneous measurements entering the minimization. An observation operator allows to simulate radial Doppler winds measurements from the model horizontal wind based on Caumont and Ducrocq (2008). Only measurements performed within 150 km to the radar are considered due to the broadening of the beam with increasing distance and the lack of reliability. An observation error variance proportional to their distance from the radar is applied in the minimization. Reflectivities are not directly assimilated but they are used to retrieve pseudo-observations of relative humidity from surrounding simulated reflectivity profiles through a unidimensional Bayesian inversion. A horizontal thinning on the data (Doppler winds and retrieved profiles of relative humidity) is performed to avoid horizontal correlation of observation errors: only one profile, having the most important number of elevations that passed the quality control, is selected in each 15 × 15 km 2 box."

Figure 1 – Add a legend of the observations shown in the figure. Change caption to show that also GNSS is used in the study even if the coverage is not shown in Figure 1.
The legend was added in the figure and the mention of GNSS is made in the caption: "Location of observations considered in this study, with the exception of GNSS Zenithal Total Delays."

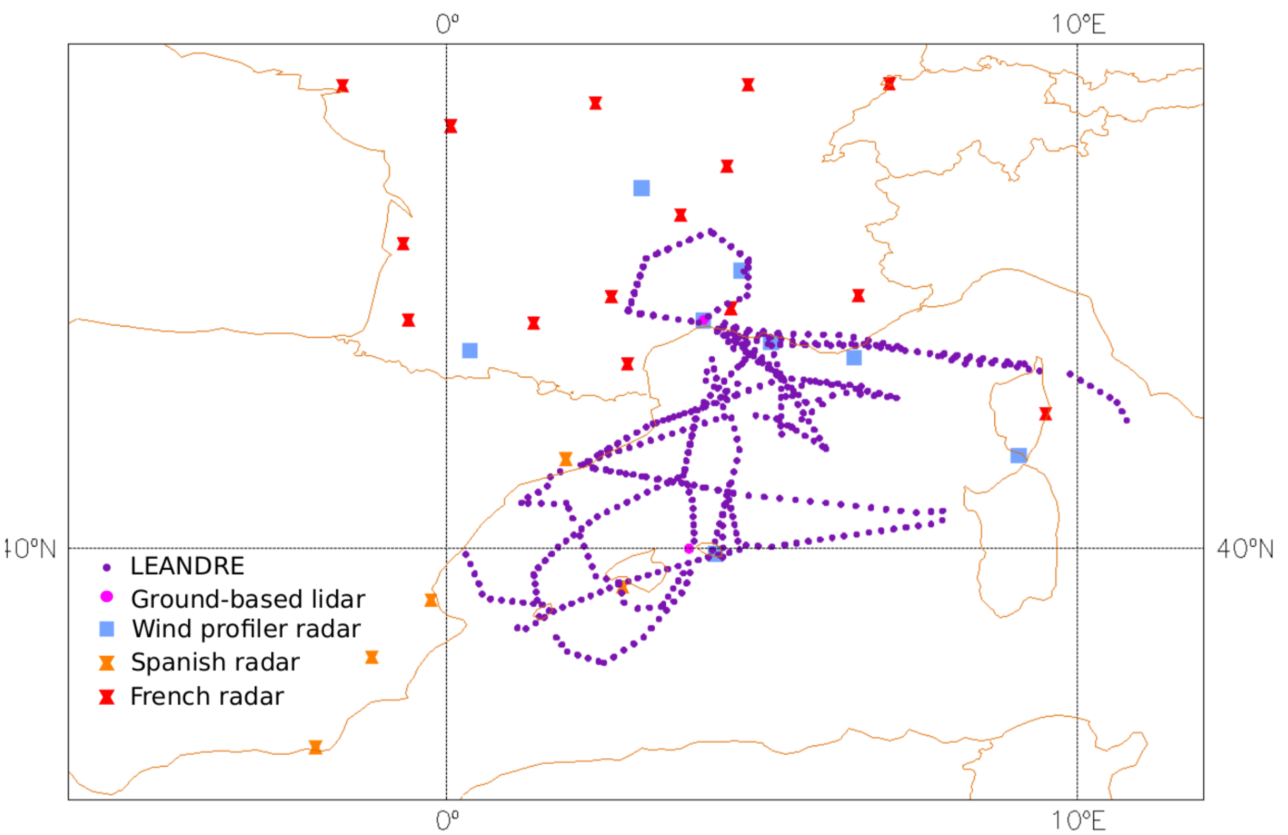


Figure 2 – Adds no relevant information.
The figure has been removed.

2.3 ***Validation protocol – The terminology is somewhat confusing. First of, the term validation should be used for observations vs. simulations comparisons, not for a comparison against the reference run REANA. Second, the authors should differentiate between a validation against dependent and independent observations. My suggestion would be starting the description of the validation protocol talking about the comparison against observations (dependent and independent) and afterwards about the evaluation of the impact against REANA and not otherwise.***

The subsection was rearranged as suggested by the Referee:

"As a first step, the performance of the data assimilation system is validated by comparing the various Analysis (AN) and First-Guess (FG) values against available observations which can be independent from REANA (i.e. not assimilated) or on the contrary assimilated in REANA. One of the key tool used to evaluate the performance of the assimilation system is to examine the FG departure (O-FG) and the AN departure (O-A) in terms of mean and root-mean square (RMS) values, O standing for Observation with the other assimilated observations. Statistics of departures are computed at the observation location.

Those statistics were also computed using few available independent data. The first source comes from the vessel Marfret-Niolon, which was an instrumented commercial ship of opportunity, cruising regularly between the southern France harbour of Marseille and two Algerian harbours (Algiers and Mostagadem). Please refer to Figure 14 of Fourrié et al. (2019) for the trajectories of the vessel during SOP1. Two autonomous systems were installed in order to provide atmospheric and oceanic measurements, in the context of the HyMeX Long Observation Period (LOP). A GNSS antenna was installed at the front on the vessel Marfret Niolon for the duration of the HyMeX campaign. An example of the operational measurements which started on January 2012 are provided in Figure 2 with figures ranging from 2.2 m to 2.6 m. The data were post-processed in kinematic Precise Point Positioning with the software provided by Natural Resources Canada

(Kouba and Héroux, 2001) and using high-resolution products provided by the International GNSS Service.
The second source of independent data comes from wind data obtained from an airborne Doppler cloud-profiler radar named RASTA (Radar Airborne System Tool for Atmosphere (Bouniol et al., 2008; Protat et al., 2009; Delanoë et al., 2013)) that flew 45 days during SOP1. This airborne radar was on board the Falcon 20 research aircraft. It allows the documentation of the microphysical properties and the horizontal components of the wind field in terms of vertical profiles.
The operational data assimilation monitoring procedure also provides FG and AN departure statistics for assimilated observations in the experiments, which are are described in a companion paper (Fourrié et al., 2019).
In a second step, the forecast (range between +3 to +54 hours) quality is assessed in terms of surface parameters and precipitation scores. The surface parameters (temperature and relative humidity at 2 m and wind at 10 m) come from the HyMeX database which provides surface synoptic observations available over the AROME-WMED domain, together with additional hourly observations from Météo-France, AEMET and MeteoCat mesoscale networks. Some of these observations were assimilated to produce surface analyses. For the evaluation of the precipitation quality, the dense surface data set rain gauge network available in the HyMeX data base (V4 version) has been used. Scores of 3 hourly accumulated precipitation from all analysis times on a given day are compared to the corresponding observed 24-h accumulated precipitation.
The evaluation of the various denial experiments is compared with the reference REANA run. This allows to get the impact of each considered observation type on the analysis and the forecast.

L149-150 - The sentence "This data set represents the largest one in terms of total amount, even though it represents a small fraction of assimilated data (1.85%)" needs further explaining. Considering the information conveyed in Table 1, where the satellite are the most numerous assimilated observation, why is it here said that the GNSS data set is the largest in terms of amount of observations? And finally, why do they "represent only a small fraction of assimilated data"? Did these data not pass the quality control of the assimilation system? Are they rejected due to blacklisted reports?
Please elaborate.
The Referee is right, it is incorrect to says that GNSS represents the largest data set in terms of total amount of observations. The correct sentence is : "This data set represent the largest one in terms of the number of studied observation types, event though in the end it represents only a small fraction of assimilated data (1.85%) in the analyses (Table 1)." because satellite data are the most numerous assimilated data in AROME. There are other observation types which are denser like the surface stations or which provide information one the vertical at different levels.
This data set represents the largest one in terms of the number of studied observation types, even though in the end it represents only a small fraction of assimilated data (1.85%) in the analyses (Table 1). As seen in Table 1, satellite data are the most numerous, followed by surface stations data, radar data from the French network, aircraft data and radiosondes ones. Even if surface data provide information only for one level, the network is very dense over France and was reinforced in other countries like Spain or Italy. The other observation types provide information at different levels all along the vertical

L157-159 - Why is there a minimum in the correlation at 15 h? Were the measurements less accurate at that time of the day and therefore dismissed? Were there less observations assimilated? Is it due to any physical process with a diurnal cycle? The explanation should be mentioned in the manuscript even if just briefly.
The correlation (respectively the standard deviation) is lower (resp. stronger) in the afternoon at 15 UTC. This time corresponds to the early stage of the maximum of the convection. A weak diurnal cycle of the scores is noticed with a maximum correlation (resp. minimum standard deviation) around 09 UTC and a minimum (resp. maximum) around 15 UTC, the latest one occuring during the maximum activity of the diurnal convection. The text was modified as below: This minimum of correlation and the maximum of standard deviation correspond to the time of the early stage of the convection.

L163-165 - ***This sentence should be reformulated. Why no impact can be seen in the analysis of RMS differences, but a small impact is present for FG RMS differences?***

No impact can be seen on the analysis RMS differences. This absence of impact can be explained by the fact that radiosondes are reference observations for assimilation and all the analyses are very constrained by these observations. However, a small positive impact is present on the FG RMS difference 3-h later. Lowest differences are obtained with REANA simulation, the largest ones with NOGNSS. The OPERGNSS differences are close to REANA one but slightly larger, showing on the one hand that the assimilation of GNSS data is beneficial (OPER-GNSS data set or REPROC-GNSS one) and on the other hand that the reprocessing of the data bring a small improvement in the comparison of FG with humidity of radiosondes. This shows that the modifications in the analysis brought by the GNSS at other places than radiosondes ones are beneficial and kept in the 3-h forecast.

The explanation above was provided in the paper.

L165-167 - This information is not understandable. If the largest benefit of assimilating GNSS data occurs in the layer, between 600 hPa and 850 hPa how is it possible that "the slight benefit of assimilating reprocessed GNSS data appears between 700-850 hPa."? The 700 hPa – 850 hPa layer is contained within the 600 hPa – 850 hPa.

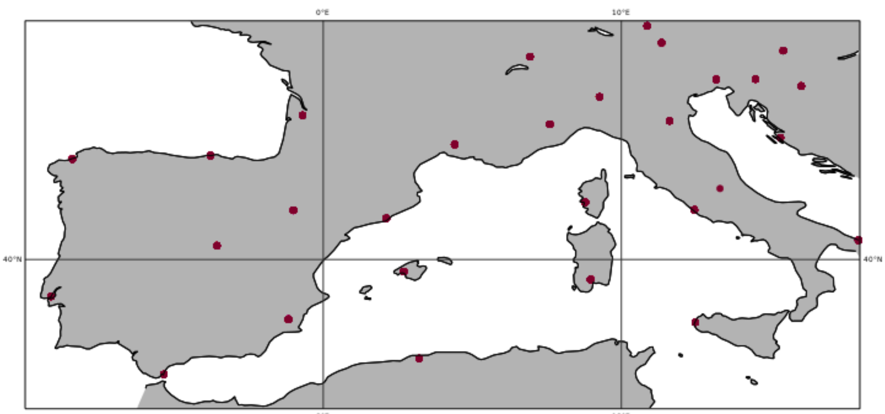
The largest improvement of the assimilation of GNSS data (OPER-GNSS data set and REPROC-GNSS one) is found between 600 and 850 hPa. In addition, a slight benefit of assimilating REPROC-GNSS data versus OPER-GNSS ones appears between 700 and 850 hPa.

Overall, a clearer description of when the reprocessed data is used or not should be needed. Again, it would be advisable choosing one nomenclature for each data set and stick to it.

OPER-GNSS will now refer the operational E-GVAP dataset and REPRO-GNSS to the reprocessed one.

Figure 4 – No appreciable differences exist in the profiles of the AN runs for REANA, NOGNSS, OPERGNSS. Furthermore only small differences (0.05 g/kg) can be found for the FG runs. If there are no differences in the profile, why is the standard deviation of the integrated moisture so different (up to 1 mm) in Fig. 3? Also more information about the radiosondes used is needed. Where are they located? What is the temporal resolution?

Even though larger differences in analysis are seen in Figure 3, there are no appreciable differences exist in the profiles of the AN runs for the 3 experiments. It is because we are not looking at the same place. In Figure 3, we study the impact of the assimilation at the REPRO-GNSS location and there much less radiosondes (29 in the AROME-WMED domain, please see below and Figure 1 in Fourrié et al (2015) than GNSS ground stations (698 were considered here). Radiosondes are mainly launched at 00 and 12 UTC. Some additional radiosondes were launched on request at 6 and 18 UTC. In addition analysis is strongly driven by the radiosondes, that is why the statistics for the 3 analysis are very close.



We propose the following modifications in the text: “No impact can be seen on the analysis RMS differences. It can be explained by the fact that radiosondes are reference observation for assimilation and all the analyses are very close to these observations. However, a small positive impact is present on the FG RMS difference 3-h later.”

Figure 5 – What is the explanation for the large bias in mean ZTD between the independent station Marfret-Niolon and the set of simulations? This is especially evident after 12 UTC. This would imply a wet bias in the REANA, NOGNSS and OPERGNSS simulations compared to the independent observations.

We recognize that there is a positive bias (2,46 m) which implies that the 3 REANA, NOGNSS and OPERGNSS have a wet bias compared to independent observations. We have no clear explanations for that. Marfret Niolon observations are located over the sea where few observation are available and the model is less corrected with the data assimilation and thus has with stronger errors. An explanation attempt was provided in the text:

This could be explained by the fact that there few assimilated observations over the sea which results in a more biased model.

Figure 6 – Just as Fig. 5 and Fig.11 the validation against the Marfret-Niolon is too noisy to convey any conclusive information. Either quantify or change the graph to show real evidence of the good(bad) performance of the OPERGNSS and NOGNSS assimilations.

Figure 6 was replaced and Figure 5 was supplemented with a new table, Table 3, see below.

Experiments	AN ZTD Correlation	Mean ZTD (m)	FG ZTD Correlation
REANA	0.967	2.4617	0.961
NOGNSS	0.961	2.4642	0.958
OPERGNSS	0.962	2.4654	0.958
OBS	1	2.4606	1

Table 3. Correlation of the differences between zenithal total delays (ZTD) between REANA, NOGNSS and OPERGNSS analyses and corresponding Marfret-Niolon observations computed over the 8 analysis slots (first column), mean ZTD for REANA, NOGNSS and OPERGNSS analyses and Marfret Niolon observations and correlation between ZTD forecasted by AROME-WMED at the 3-hour forecast range and observations from Marfret Niolon.

The following comment was added in the paper:

Table 3 shows the mean correlation of REANA, NOGNSS and OPERGNSS AN and FG with Marfret-Niolon observations. The higher correlation is obtained with REANA for both AN and FG. When comparing the mean value of ZTD at the Marfret-Niolon places, the closest value to the observed one is obtained with REANA, even if a small moist bias is observed (0.9 mm). This bias is larger for NOGNSS (3.6 mm) and OPERGNSS (4.8 mm). This could be explained by the fact that there few assimilated observations over the sea which results in a more biased model.

Table 3 Shows Interesting information as GNSS observations have no wind information. How can the improvement in the wind description be explained? Is it a direct impact of the 3DVar assimilation? Is there any physical process explaining the improvement? If so, please mention it in the manuscript. Also reference to other publications if this effect had been addressed before.

The effect of mass field information assimilation on the wind field is essentially created during model integration because there is little coupling between these fields during the analysis (e.g., Borderies et al. 2019). This indirect effect has already been demonstrated by Wattrelot et al. (2014), for example, who noted a positive impact on the wind field when assimilating pseudo-observations of relative humidity. Lindskog et al. (2017) also reported—but did not show—a positive impact on wind forecasts when assimilating ZTD data.

As suggested this explanation was added in the text : « As GNSS observations do not provide any wind information, the improvement observed in wind field can be explained by the effect of mass field information assimilation on the wind field, essentially created during model integration. There is indeed a little coupling between these fields during the analysis (Borderies et al. 2019b). This indirect effect was already demonstrated by Wattrelot et al. (2014), for example, who noted a positive impact on the wind field when assimilating pseudo-observations of relative humidity. Lindskog et al. (2017) also reported—but did not show—a positive impact on wind forecasts when assimilating ZTD data. »

Figure 8 - This result is very interesting. Especially striking is the loss of skill for OPERGNSS with larger daily precipitation. How can this be explained? Is this a result of the lower number of occurrences for heavier precipitation events? Please, add an explanation in the manuscript.
The loss of skill for OPERGNSS simulation for larger daily precipitation from the 8 3-h forecast is surely due to the lower number of occurrences for heavier precipitations. It was written in the text : However this threshold represents only few cases. We replace it with : However, this is the result of the lower number of occurrences for heavier precipitation events.

L205-207 - Indeed the comparison is noisy, with this graph is not possible to see which simulation has the largest (smallest) correlation. Please, add a table with the average values or present the differences against REANALYSIS, otherwise the statement "the correlation for the NOGNSS is lower and the standard deviations are in general higher for the NOGNSS" is unsupported. The same applies to Figures 5 and 6.
Figure 10 was replaced with the following Table 5:

Parameter	REANA	NOGNSS	OPERGNSS
Correlation (1-24h)	0,962	0,957	0,957
Correlation (25-48h)	0,922	0,917	0,919
Correlation (49-54h)	0,906	0,902	0,905
Standard deviation (forecast -observation, 1-24h)	0,0152	0,0164	0,0160
Standard deviation (forecast -observation, 25-48h)	0,0221	0,0226	0,0223
Standard deviation (forecast -observation, 49-54h)	0,0244	0,0249	0,0244

Table 5. Correlation and standard deviation of integrated water vapour content between AROME-WMED forecasts and reprocessed GNSS observations averaged over forecast ranges.

The text was modified : "Compared to the observed ZTD from the Marfret-Niolon ship, the signal is more noisy because of a smaller dataset but when comparing to values average over the forecast ranges (Table 5, the correlation for the NOGNSS is lower than REANA and OPERGNSS, which provides it-self lower correlation than REANA. The standard deviations are higher for the NOGNSS forecasts. In addition, a decrease of the correlation (respectively an increase of the standard deviation) is seen for forecast range over 24-h."

L231-236 – There is no evidence for the results conveyed of the NOLIDAR experiment. Please add the corresponding evidence.
The following table was added in the paper:

	REANA	NOLIDAR
Correlation for the forecast 0h	0.968	0.960
Correlation for the forecast 1-24 h	0.962	0.961
Correlation for the forecast 25-48 h	0.923	0.924
Correlation for the forecast 49-54 h	0.906	0.907
Standard deviation for the forecast 0h	0.0144	0.0167
Standard deviation for the forecast 1-24 h	0.0152	0.0154
Standard deviation for the forecast 25-48 h	0.0221	0.0220
Standard deviation for the forecast 49-54 h	0.0243	0.0243

Table 6. Correlation and standard deviation of integrated water vapour content between AROME-WMED forecasts and reprocessed GNSS observations averaged over forecast ranges (0 h, 1-24 h, 25-48 h and 49-54 h).

L238-245 – This result is revealing. Why the improvement of the Spanish radars can only be observed over Spain? why is the impact so Local? does it depend on the assimilation system (3DVar)? How is the localization of the data assimilation working in your system? More information is needed on how the 3DVar system is implemented for these cases.

No clear impact on the global scores can be seen but the REANA and NORADSPAIN forecasts are different. I would like to remind that the assimilation of Spanish radar data in AROME-WMED was made on a research mode as only French radars were assimilated at the time of the HyMeX campaign and the reanalyses. The Spanish radar data were directly provided by AEMET on a ftp site. These data represent only 0.6% of the assimilated data. This is three times less than REPROC-GNSS data. In addition only three elevations are used for these radars. To get a positive impact, a strict quality control from these data has to be applied, that induces a decrease of the number of potential assimilated radar data. The impact result would have been certainly different if we had consider all the radar data including the French radar network. Currently Spanish radar data are provided by the European Radar programme OPERA which proposed a unified processing of the data which provides a better impact of the radars. Additional information on the 3D-Var was added in the AROME-WMED description section 2.2.

A comment on this has been added to the text. Even if we do not obtain significant impact at the HyMeX domain scale but a significant one over the Iberian Peninsula, it is interesting to remind that the assimilation of Spanish radar data in AROME-WMED was made on a research mode as only French radars were assimilated at the time of the HyMeX campaign and the reanalyses. These data represent only 0.6% of the assimilated data. This is three times less than REPROC-GNSS data.

Technical corrections

Abstract (L13-14) –For clarity, add information of lead time (number of hours) considered as a "very short term forecast" and a "short term forecast".

The very short term correspond to the 3-h lead time and the short term to the 30-h one. This was clarified in the text.

L16 - "Copyright Statement. TEXT". Removed

L21 - Correct the typo ": : :6 November 2012) in the north western : : :". Done

L31 - Bad double parentheses style. It should read "(Application of Research to Operations at Mesoscale; Seity et al., 2011)". https://guides.library.nymc.edu/c.php?g=567729&p=4609898.
Corrected

L35 - Acronym of Innovative Observing and Data Assimilation Systems is missing.
IODA-MED has been added in the text.

L36 - Substitute "With a view of" by "with the aim of". Modified

L36-39- Sentence is too long. Split it and rephrase ": : :due to a system upgrade in the middle of the SOP1). The second one included in addition a maximum of observations: : :"
Done

L39 - Substitute "This latter" for "The latter". Done

L56 - The description of the paper's sections ("This paper is arranged as follows: : :") should start in a new paragraph. Done

L108-109. Sentence ": : :data is 75 m but for assimilation, data were thinned at 75 m below 2000 m up to 450 m above 5000 m." should read ": : :data is 75 m but for data assimilation above 5000 m the resolution was thinned to 450 m."

Changed with The raw vertical resolution of the data is 75 m but for assimilation above 2000m, the resolution was thinned starting from 75 m to 450 m above 5000 m.

Table2 - Should be placed at the beginning of Section 2.2, outside the bullet point 2.2.4 as it shows the overview of all observation systems, not only on the Spanish radar.
Done

L171 - English style is not correct because of double enclosure (see above).
Corrected.

L171-175 - Show the plots with the SEVIRI results in the supplementary material.
Please find below the table of RMS (K) for FG and AN departures for SEVIRI channels 2 and 3

	REANA	NOGNSS	OPERGNSS
FG channel 2	1,27	1,26	1,27
AN channel 2	0,44	0,44	0,44
FG channel 3	1,24	1,24	1,27
AN channel 3	0,48	0,47	0,47

As you can see, no impact of GNSS can be seen. I add the Table in the supplementary material.

L176 - For the "various AROME-WMED ZTD analyses" use the designated nomenclature AN.
Done

L177 - Please rewrite "is slightly and consistently higher".
The correlation between the various AROME-WMED ZTD AN and corresponding independent (not assimilated) Marfret-Niolon observations is higher for REANA than for NOGNSS and even for OPERGNSS (Figure 5).

Figure 4 - Instead of top and lower panel should show left and right. Corrected

L185 - It should read "This airborne radar was on board the Falcon 20 aircraft". Done.

Figure 5 caption - Write explicitly that the Marfret-Niolon observations are independent. The caption is now : Correlation of the differences between zenithal total delays (ZTD) between REANA (Black), NOGNSS (red), OPERGNSS (blue) analyses and corresponding Marfret-Niolon independent observations as a function of analysis time in the left panel; mean value in the right panel, the grey line corresponding to observations.

Line 195 - It is not clear to which data set the authors are referring to with the terms "reprocessed data" and "real-time ones", is the NOGNSS or OPERGNSS ?. Be consistent in the nomenclature of the data sets.
The text was changed into : When comparing the assimilation of REANA to OPERGNSS , the ETS for precipitation is slightly better with the reprocessed data set but the differences are not significant except for the 40 mm/day threshold.

L201 - Instead of "IWC" it should be "IWV". Corrected

L.209 - Instead of wind it should be "humidity" or "relative humidity"
Not changed :The only impact on the surface parameters is on relative humidity at 2 m. No impact was found on temperature at 2 m or on wind at 10 m.

Figure 11 - Instead of top (bottom) panel it should be left (right). Corrected.

L230-236 - Indicate which Figure is being analysed when saying "The denial NOLIDAR experiment results are close to the reanalysis ones as these data represent very few additional data and are located over ocean where few observations are available for the comparison".
This does not refer to a figure but now to Table 6.

L242 - It should read "This impact does not remain at longer : : :." Corrected

L250 - Should read "Catalonia". Corrected

L251 - Rephrase sentence "on, in the evening the Italian Ligurian coast.".
"on, and then hitting the Italian Ligurian coast in the evening"

Figure 20 - has a bad quality, letters cannot be read and are blurry.
Figure 20's quality was improved as shown below and in the pdf file.

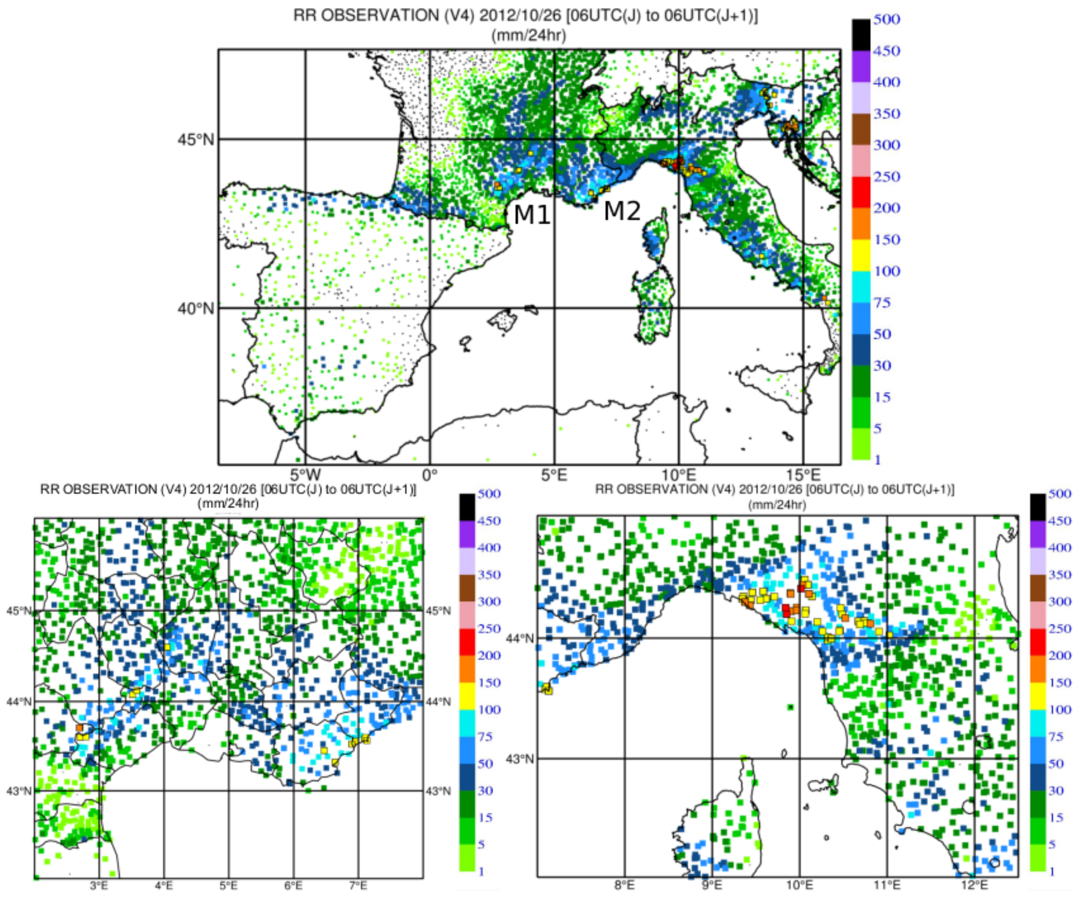


Figure 20. 24 h accumulated precipitation (mm) between 26 October 06 UTC and 27 October 2012 at 06 UTC over the AROME-WMED domain (upper plot) and zoom over the Cevennes region (left lower plot) and over North of Italy (right lower plot).

L304 – Substitute "Iberic Peninsula" for "Iberian Peninsula" Modified