Reviewer 2:
We thank Reviewer #2 for his/her constructive comments.
Our responses are given below in red.

Page 2, line 16: "...at each kilometer levels..." Please explain better what this means?

We replaced "at each kilometer levels" with "a high vertical resolution".

Page 2, line 35: What is different compared to the operational version? Just a brief, short explanation would be good.

The AROME-WMED model was specifically designed for the HyMeX-SOP1 field campaign to support the instrument deployment. The main differences between AROME and AROME-Wmed are:

- The domain: the AROME-Wmed domain is centred on the area of interest (northwestern Mediterranean area)
- The Background error covariance matrix, which has been calculated during an autumnal period in October 2010 characterised by heavy rainfall events.
- The number of assimilated observations: AROME-Wmed assimilates more observations in the southern part of the domain.

We added "specifically designed for the HyMeX-SOP1" in the introduction section (L6, P3). We also added the above informations in section 3.1

Page 3, line 30: 7.8 m/s is a rather low value for the unambiguous velocity. In fact this is one of the main challenges in Doppler wind assimilation. A de-aliasing, or unfolding algorithm can work fine for unfolding once but what if the wind speed is high enough to fold twice? Then it will be much more uncertain. Are the authors confident with the algorithm used and/or that there are no wind speed above this limit? Another complicating factor is that the aircraft is moving and this also needs to be taken into account in the unfolding. Perhaps it is outside the scope of the paper to discuss this in detail but a brief discussion about this is necessary since it is crucial when using the data.

In situ wind measurements at flight altitude are used to check the number of foldings for the first valid gate and then by applying a gate to gate correction for the next ones. The authors are confident with the algorithm because most of the time RASTA was collecting data in cloudy areas. In addition to that the combination of the three non colinear beams is used to verify potential unfolding issues as the retrieval would be locally inconsistent.

The exact speed of the aircraft and the pointing angles allow one to rigorously determine the component related to the aircraft's movement. The algorithm is further described by Bousquet et al. (2016). We added some more information in section 2.1.

Bousquet, O. , Delanoë, J. and Bielli, S. (2016), Evaluation of 3D wind observations inferred from the analysis of airborne and ground-based radars during HyMeX SOP-1. Q.J.R. Meteorol. Soc., 142: 86-94. doi:10.1002/qj.2710

Page 5, line 21: "every 3 time steps" What does this mean?

First, super-observations are calculated using the median filter (median value of all data available along the aircraft track within a box of 2.5 km length between the two half model levels surrounding each model level). After this median filter, a thinning is applied to these super-observations. It has been decided to select one super-observation out of three.

We modified the text in section 4.1 by:
"After this pre-processing, to satisfy assumptions about observation error covariances, which are supposed to be 0~m²/s², a thinning is applied to RASTA wind "super-observations". One super-observation out of three is then assimilated, which is equivalent to approximately one observation every 5~km to 9~km depending on the aircraft speed."

Page 5, lines 25-28: It could also be so that “important” data is collected late (or early) in a longer assimilation window if the aircraft e.g. flies into a convective cell. In studies like this it would also be very beneficial to run with FGAT (First Guess at Appropriate Time) or even better using a 4D-Var assimilation scheme.

Reviewer 2 is right, FGAT is a way to improve the handling of the time dimension in a 3D-Var scheme as it allows to compute the innovations (i.e. the observation-guess differences) at the time of the observations for different times during the assimilation window. For the AROME model, the FGAT option has been evaluated by Brousseau (2012) for moving platforms, but without any positive improvement in the subsequent forecasts (Brousseau et al. 2016, section 2). For observations from static platforms, the 3DVar without FGAT only uses the observations performed at the middle of the assimilation window. The FGAT option allows to estimate innovations for sub-hourly data from the same instrument at the same location. More observations are assimilated, but the 3D-Var minimisation, without time dimension, uses these several innovations at the middle of the assimilation window. This leads to an averaging and a smoothing effect on these observations and a loss of information on the temporal details, which is not desirable in a convective DA system. Therefore, in this study we decided to use conventional 3DVar to assimilate all the different kinds of observations in the same way.

The 4DVar assimilation scheme is numerically too costly for the AROME model (Brousseau et al. 2016).

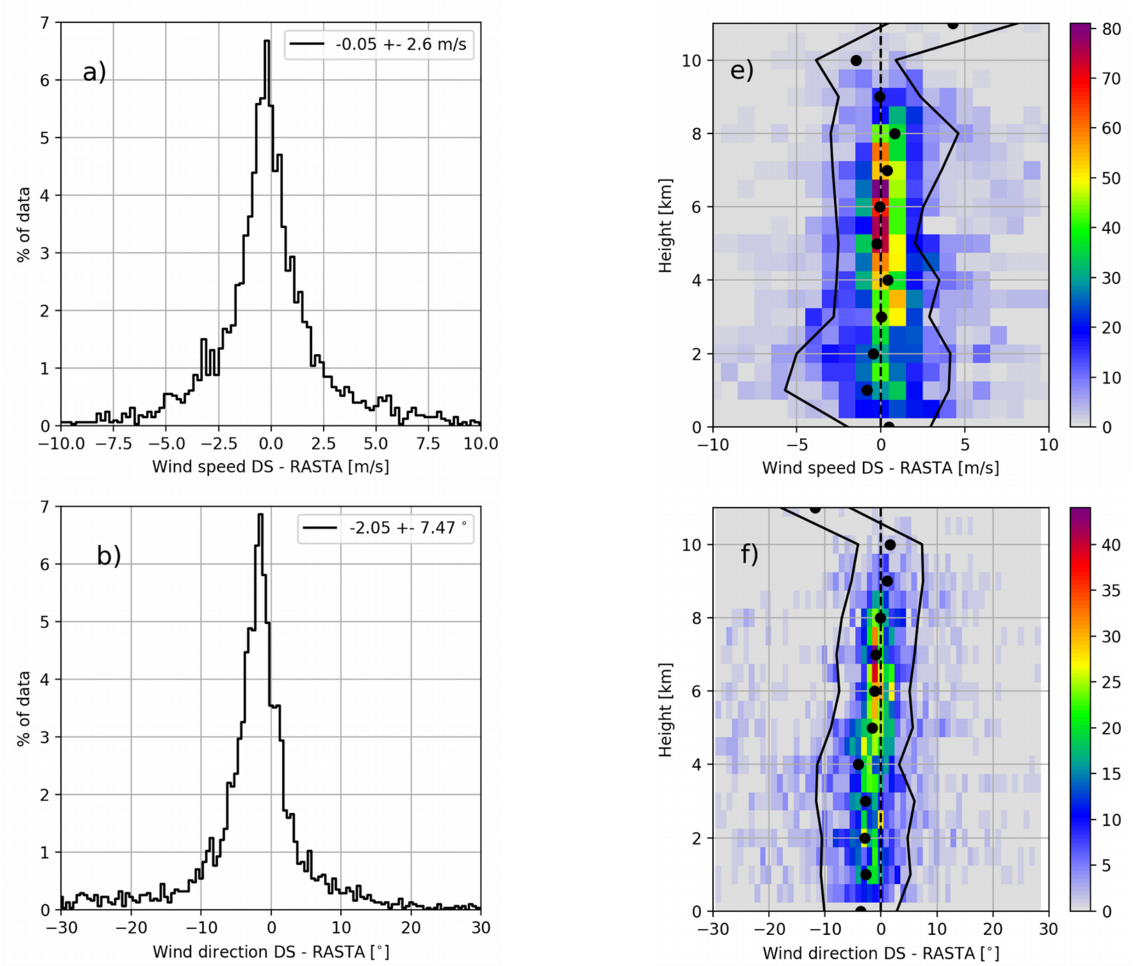
Brousseau, 2012: Propagation of observed information into the AROME data assimilation and atmospheric model, PhD thesis, Université de Toulouse III – Paul Sabatier

Brousseau, P. , Seity, Y. , Ricard, D. and Léger, J. (2016), Improvement of the forecast of convective activity from the AROME-France system. Q.J.R. Meteorol. Soc., 142: 2231-2243. doi:10.1002/qj.2822

Page 6, line 10: The same observation error as radiosondes. Isn’t this a bit optimistic?

RASTA wind data during the HyMeX-SOP1 field campaign have been compared against ground-based Doppler radars by Bousquet et al. (2016). Results of their study show that “The low values of the bias error suggest that errors are close to Multiple-Doppler wind synthesis and should remain comprised between 1 and 1.5m/s” (see section 3.2, page 93). These values are smaller than the radiosonde ones (between 1.8 and 2.52m/s). We added these values in section 4.2. *“Bousquet et al. (2016) demonstrated that the bias error of RASTA wind data is comprised between 1 and 1.5 ms −1 . In this study, it has been decided to use the same observation error as the one used for radiosondes, which increases with the altitude (from ≈ 1.8 ms −1 at 900 hPa to ≈ 2.52 ms −1 at 200 hPa).”*

RASTA wind data have also been evaluated during the NAWDEX field campaign which occurred in Iceland (http://www.pa.op.dlr.de/nawdex/). In the following figures, RASTA wind retrievals were compared against radiosonde measurements. These Figures demonstrate that the observational error for RASTA wind data is of the same order of magnitude as that of radiosondes.



Bousquet, O. , Delanoë, J. and Bielli, S. (2016), Evaluation of 3D wind observations inferred from the analysis of airborne and ground-based radars during HyMeX SOP-1. Q.J.R. Meteorol. Soc., 142: 86-94. doi:10.1002/qj.2710

Page 6 lines 11-13: Is there any other quality control applied to the observations? If yes, what and how. If no, why not?

There is no other quality control applied to the observations. We did not apply any other quality control because after visual inspection, we did not see any remaining spurious observations.

Page 6, lines 20-24: The last two sentences in the paragraph is really hard to follow. Please re-write to make it more clear.

The text has been modified. We hope it is now clearer:

"A larger assimilation window results in assimilating data more frequently, but the time lag between the observation time and the analysis time is greater than one hour. On the other hand, a smaller assimilation window constrains the number of analyses to those for which the observations are valid near the analysis time. Therefore, the percentage of analyses in which RASTA wind data were assimilated decreases with the length of the assimilation window from 9.5% in the RASTA_3h experiment to 7.2% in the RASTA_1h experiment. Finally, the last column of Table 1 represents the percentage of RASTA wind data which were assimilated among the total number of assimilated data (conventional, GNSS, radar, satellite, RASTA, etc.) over the entire AROME-WMed domain (represented in Figure 1).

This percentage is quite small because of the already dense observing network used in AROME-WMed."

Page 7, table 1: In the column of assimilated data it says Conventional and Conventional + RASTA. Are only conventional observations assimilated apart from the RASTA observations? In section 3.2 there are many more observations mentioned that are not consider to be conventional, e.g GNSS and satellite data.

All the observations which are mentioned in section 3.2 are assimilated.
Table 1 has been rectified.

Page 7, lines 10-14: The data collection starts at 06:10 and the analysis time chosen to study is 06:00. This means that the 3 hour window only will be a 1.5 hour window. Is there no better example where one can find a data collection more centered around the analysis time. Why not show an example from 09:00? Then the data collection will also be skewed but there will at least be data available on both sides of the analysis time.

For the case study, we looked for a situation in which the Falcon 20 aircraft was collecting data around the location where the increments are advected as the forecast range increases. Such a configuration only happens twice during the entire HYMEX-SOP1 period: the IOP7a case study at the analysis time of 06 UTC, and the 2012-10-11 case study at the analysis time of 18:00 UTC, in which the observation time start at 17:50 UTC.

There are other examples in which RASTA observations are more centred around the analysis time, for example at the same date at 09:00UTC. However, there are no observations available after 9:30 UTC to validate the subsequent forecasts. Therefore, we decided to only show the 06:00 UTC analysis results because RASTA observations are available at 07,08 and 09 UTC to validate the forecasts.

Page 7, line 20: "...expected to improve the forecast..." Is this really the case? It depends on how he data is introduced, observation errors and how the model performed without the data.

The authors agree with reviewer 2. We wrote this sentence because the increments are advected at approximately the same place of which the rainfall event took place. Therefore, the impact of the assimilation of RASTA data is expected to have a noticeable impact on the rainfall event. This is the main reason why this case study has been selected. However, at this stage, it is too early to talk about any possible improvement.

We replaced "expected to improve the forecasts" with "expected to have an impact on the forecasts".

Section 5.2: It would be interesting to see the same case in a cycled period. If the cycled run builds up its own "climate" could the results be even better?

The results of this case study in the cycled run (starting from September 24) are indeed better for the CTRL and for the different RASTA experiment cycled runs. However, in this section we want to demonstrate the impact of the assimilation of RASTA wind data at a specific analysis time. Therefore, to disentangle the benefits brought by the cycling effect from the impact of the assimilation of RASTA wind data, results are shown for the non-cycled experiment runs for the case study.

In the statistical study, results are shown for the cycled experiment runs.

Page 8, lines 19-25: Please explain figure 5 better.

The explanations have been modified. Fig 5a (now 6a) is first described is section 5.2: *"Figure 6A represents the wind speed increments at approximately 4 km of altitude (model level 30) between the RASTA_3h and the CTRL analysis. Wind directions are also indicated by the green (resp. black) arrows for the CTRL (resp. RASTA_3h) analysis. The data points assimilated in the RASTA_3h experiment until 07:30 UTC are also represented by the black data points.".*

Then, Fig.6 B-D are explained at the beginning of section 5.2: *"Figure 6 (panels B to D) represents the wind speed differences of the RASTA_3 h 1-, 2- and 3-h forecasts and the CTRL ones. At each forecast term, the black data points indicate the different RASTA locations which are available during a 1-h time window centred on the forecast time (forecast term ± 30 minutes)."*

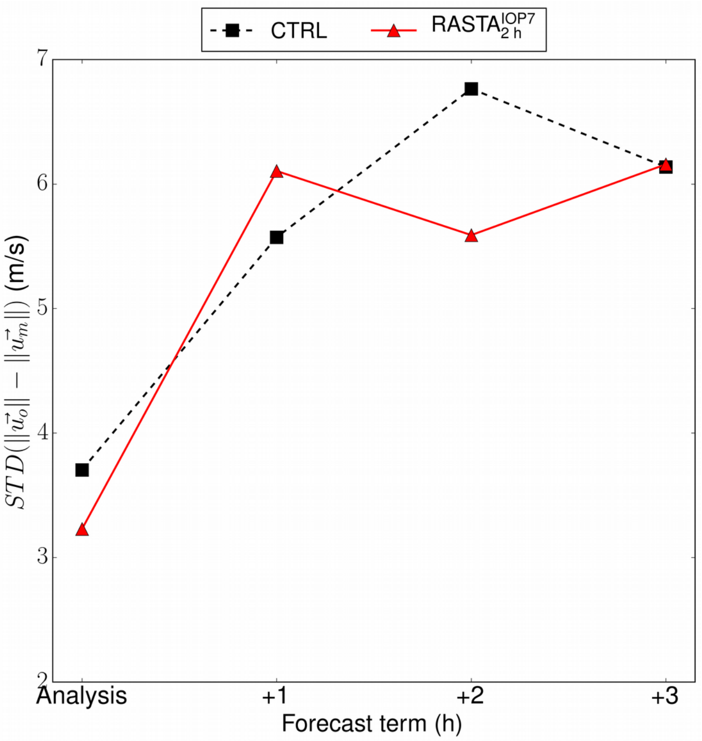
We hope it is now clearer.

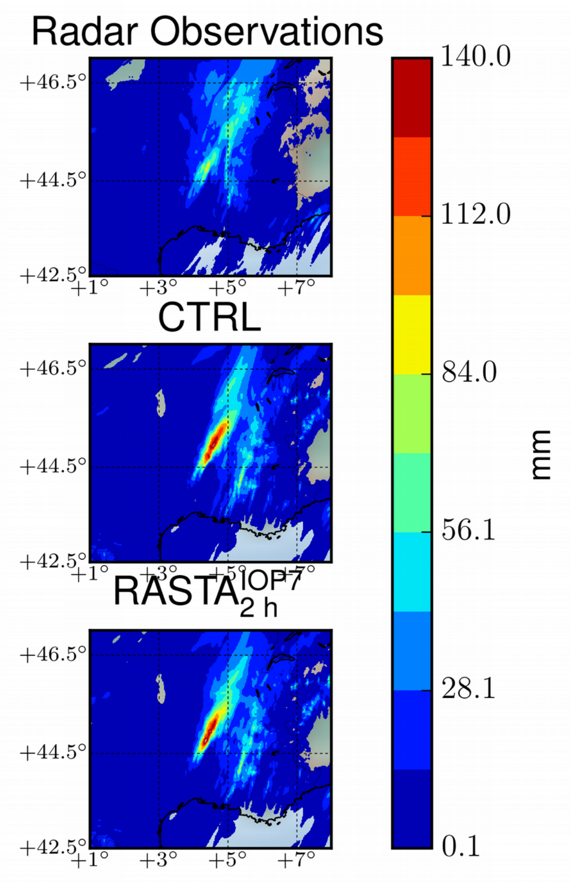
Page 9, line 9: The observations assimilated are not from ± 30 minutes from 06 UTC. They are from +30 minutes. Right?

Yes they are from +30 minutes. We rectified the text.

Page 9 and figure 6: This is a typical behavior when observations are assimilated with a too small observation error. The analysis is adjusted to fit the RASTA observations too much but as soon the model starts running it adjust itself to its own more comfortable state. The analysis will look very good, especially compared to RASTA observations, but there will be a spinup to the model state as seen in the figure. Why not run the same experiment with different observation errors too see if that can reduce the spinup and improve the forecasts, not only the analysis?

We ran the same experiment with a larger observation error of 6 m/s for the IOP7 case study (with an assimilation window of 2h). Results are displayed in the following figures. The comparison against RASTA wind data (Figure on the left) shows that the spinup is not reduced and the 3-h forecast is not improved. Besides, the comparison against radar observations (Figure on the right) indicate that the assimilation of RASTA data with a larger observation error doesn't improve at all the 12-h accumulated rainfall.





Page 9, line 28: What forecast lengths are used for the 12 hour accumulation? Interesting to know in view of the above discussion about spinup.

> We used the 12 hour rainfall forecasts for the comparisons in Figure 7 (between 06:00 UTC and 18:00 UTC). However, we agree that there is a spinup problem; and hence the first hour of rainfall accumulation should not be taken into consideration for the calculation. Therefore, we now compare the 11-h accumulated rainfall forecasts between 07:00 UTC and 18:00 UTC in Figure 7 (now Figure 8). The results are similar. We rectified the text in section 5.4.
>
> In the statistical study, we also removed the 1st hour to calculate the scores.

Page 10, line 18: Calculations are only performed over the 35 runs (I assume that this means analysis times) with RASTA observations. Why?

> As pointed out by Reviewer 1 in his/her major comment #1, the impact of the assimilation of RASTA wind data is limited in space and in time. To maximise our chances to see an impact of the assimilation of RASTA data on the scores, we constrained the number of assimilation times to only those for which RASTA wind data were assimilated. This is the reason why the scores are only calculated over the 35 analysis times in which it was possible to assimilate RASTA data. For the same reason, the scores are now calculated over a RASTA-limited area. This area contains the aircraft flight path +/- 0.5° both in longitude and latitude.

Page 13, line 11: Again quality control is mentioned and that it is important. What quality control was applied to the data assimilated here (see also comment above)?

> The quality control applied to the observations depends on the first guess departure (|Obs – Guess|). This quality control depends on the observation error and on the background error. Since the observation error increases with the altitude, the quality control also depends on the altitude.