Response to Referee 1

We would like to thank Reviewer 1 for his/her useful comments and suggestions, which helped to improved the quality of the manuscript. Reviewer comments are reproduced in italic text. Answers are in plain text.

"Generally the paper is well written but can be improved by being more precise at some places (see detailed comments). Since lots of acronyms are used within the paper and not all of them have been defined before their first occurrence a list of acronyms would be useful.
We acknowledge the used of many acronyms, a full list of them will be appended at the end of the article and defined when first used.

Appendix B: List of Acronyms
AEMET: Agencia Estatal de METeorologìa
ALADIN: Aire Limitée Adaption Dynamique et développement InterNational
AMV: Atmospheric Motion Vector
AROME-France: Application of Research to Operations at MEsoscale, France
AROME-WMED: Application of Research to Operations at MEsoscale, WestMEDiterranean sea
ARPEGE: Action de Recherche Petite Echelle Grande Echelle
BLPB: Boundary Layer Pressurized Balloon
BLLAST: Boundary-Layer Late Afternoon and Sunset Turbulence
BSS: Brier Skill Score
CNES: Centre National d'Etudes Spatiales
CNRM-GAME: Centre National de Recherches Météorologiques-Groupe d'études de l'Atmosphère MEtéorologique
COPS: Convective and Orographically-induced Precipitation Study
DTS: Data Targeting System
ECMWF: European Centre for Medium-range Weather Forecasts
E-GVAP: EUMETNET EIG Global navigation Satellite System water VApour Programme
EUMETNET EIG: EUMETNET Economic Interest Group
ETS: Equitable Threat Score
EUCOS: former EUMETNET EIG Observation Programme
GPS-ZTD: GPS Zenith Total Delay
GTOPO30: Global 30 Arc-Second Elevation Data Set
GTS: Global Telecommunication System
HyMeX: HYdrological cycle in the Mediterranean EXperiment
IASI: Infrared Atmospheric Sounding Interferometer
IOP: Intense Observation Period
MAP-D-PHASE: Mesoscale Alpine Programme-Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Event
SBL: Surface Boundary Layer
SEVIRI: Spinning Enhanced Visible and InfraRed Imager
SOP(1/2): Special Observation Period (1: Autumn 2012 / 2: Winter 2013)
SURFEX: Externalized Surface (surface scheme)
SYNOP: surface synoptic observations
WRF: Weather Research and Forecasting

Since no error analysis of the observational data in discussed within this paper it should be mentioned somewhere in the text that the observations are assumed to represent the truth."

The reviewer is right, though observations are also subject to errors but we do not have any other source of comparison for our evaluation. We will mention that fact in the introduction of paragraph 3. “Models were evaluated against observation data, which are subject to errors and biases but in this study they are used as a reference and assumed to represent the truth.” In addition we have chosen the word evaluation instead of validation since observations are not perfect as every one knows.

Detailed comments:

1. *Within this work the standard deviation (σ) is used as a measure for the forecast errors. However, it is more common to use the root mean square error (RMSE) which is the same as the standard deviation in the case there is no bias; usually RMSE is larger than σ. Why was σ used instead of RMSE?*

We decided to separate the origin of errors (bias and variability), however we also computed the root mean square error (RMSE). RMSE is very close to the standard deviation (σ), excepting in the case where the bias is large, between 9 and 15 UTC. We can replace (σ) with RMSE. In addition, we found that relative humidity data used in figures (Fig 7 and 15) did not correspond to those used for temperature and wind, we replaced panels for humidity by the right ones. Here are the corresponding figures.

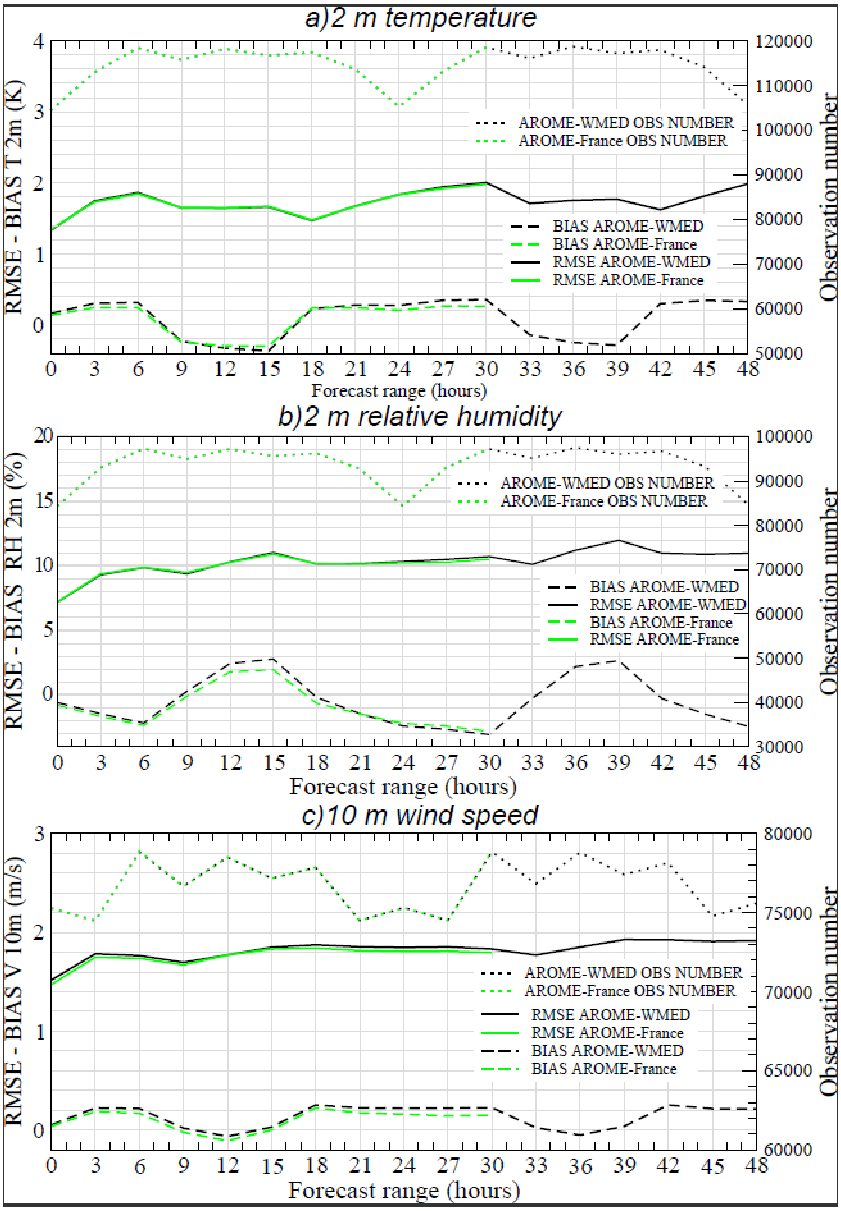

Fig7: Bias (dashed lines) and Root Mean Square Error (RMSE, solid lines) computed with 2 m temperature (a), 2 m relative humidity (b) and 10 m wind speed (c) with respect to the forecast range for the AROME-WMED model (black) and the AROME-France model (green) for SOP1 from 5 September to 6 November 2012. Dotted lines denote the number of observations used for the comparison (right y axis).

The paragraph will be modified as follow:

Temperature biases and root mean square errors are affected by the diurnal cycle. The bias in both models is positive during night-time with a maximum at 06:00UTC and negative during day-time with a minimum at 15:00UTC (Fig. 7a). The absolute bias values are slightly larger for AROME-WMED than for AROME-France, between 0.02 and 0.03 °C on average for AROME-WMED. The RMSE are similar for AROME-WMED and AROME-France. The biases for the 24–48 h ranges follow the same pattern as those of the first 24 h, but RMSE increases for the 24–48 h range (about 0.2 °C). A minimum in relative humidity bias is found at 6 h and maximum at 15 h (Fig. 7b). In that case, the error difference between both models nearly reaches 1 %. As for 2 m temperature, the RMSE are similar between both models and increase for the 24–48 h range. Concerning the wind at 10 m (Fig. 7c), the AROME-WMED mean bias is lower at 12:00UTC and larger during night-time between 18:00 and 06:00UTC

with an 0.2ms^{-1} overestimation. The RMSE, varying between 1.5 and 1.9ms^{-1}, is also a little larger than the one in AROME-FRANCE.

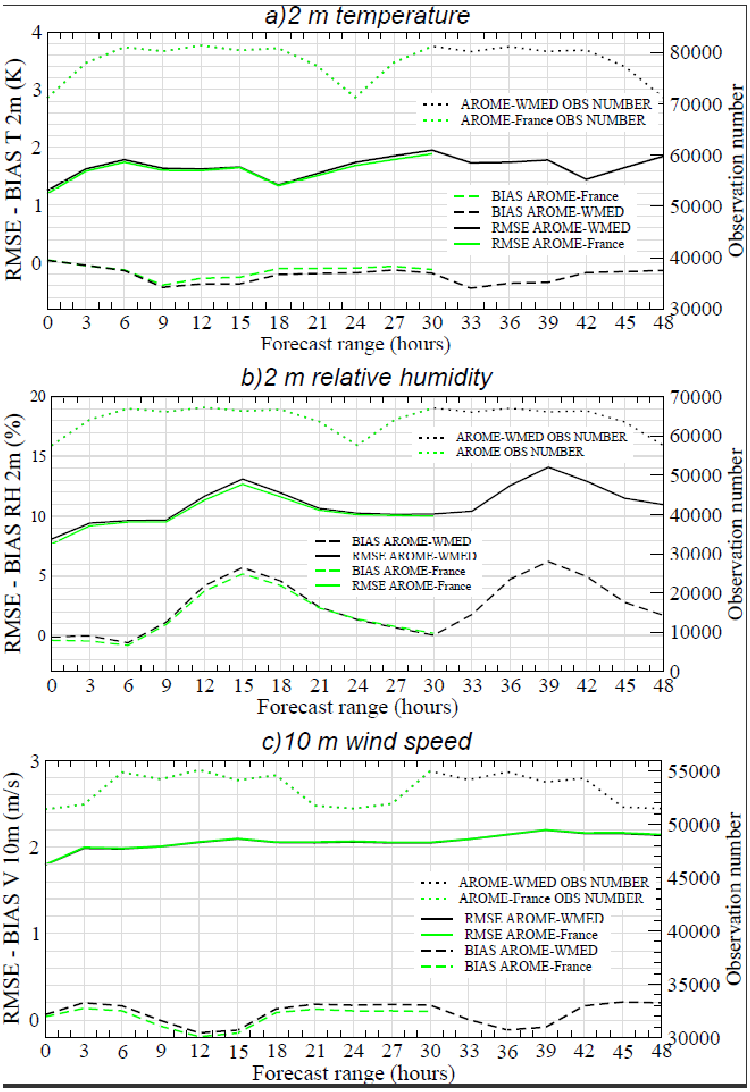

Fig 15: Bias (dashed lines) and Root Mean Square Errors (RMSE, solid lines) computed with 2 m temperature observations (a), 2 m relative humidity (b) and 10 m wind (c) with respect to the forecast range for the AROME-WMED model (black) and the AROME-France model (green) for SOP2 from 1 February to 15 March 2013. Dotted lines denote the number of observations used for the comparison (right y axis).

The paragraph will be corrected. “Though the 2 m temperature RMSE are similar (around 2°C) between AROME-WMED and AROME-France, they are slightly larger (5.6% on average) for AROME-WMED beyond the 24 h forecast range. The negative 2 m temperature error bias value becomes larger for AROME-WMED beyond the 12 h range. The difference in temperature biases of both models is around 0.1 °C for forecast ranges over 12 h (Fig. 16a). The pattern of the negative bias follows a diurnal cycle, which is less pronounced than during SOP1 (Fig. 7a). Its values are however identical between the 0–24 h and 24–48 h forecast ranges. Concerning the relative humidity, the bias cycle with respect to time is stronger than the temperature bias cycle (Fig. 16b). The minimum is obtained at 6 h and the maximum at 15 h. Moreover the RMSE in relative humidity is larger for ranges from 24 to 48 h than for the day-1 range. The RMSE maximum is reached at 15:00UTC (15 and 39 h ranges). AROME-France and AROME-WMED have a quite similar behaviour, with a better fit for AROME-

France, as shown by a smaller RMSE. As for other parameters, the wind error RMSE in AROME-WMED is larger, ranging from 1.8 to 2.2ms^{-1} during SOP2 (Fig. 16 c). The differences in error bias are more pronounced."

2. *In line 16 of the abstract it is stated that "The overall performance or AROME-WMED is good….". What does "good" mean? Same for "…similar to…" (line 16) and "… less accurate …" (line 18). It would be useful to state some hard numbers here.*

We propose to modify these sentences by "The overall performance of AROME-WMED is good for SOP1 (i.e. mean 2m temperature root mean square error (RMSE) of 1.7 °C and mean 2m relative humidity RMSE of 10% for the 0-30-h forecast ranges) and similar to those of AROME-France for the 0 to 30 h common forecast range (maximal absolute difference of 2m temperature RMSE of 0.2 °C and 0.21 for the 2m relative humidity). For the 24 to 48 h forecast range is of course less accurate (relative loss between 10% and 12% in 2m temperature and relative humidity RMSE, and ETS for 24-h accumulated rainfall) but it remains useful for scheduling observation deployment."

3. *P 1803, line 29: "A specific…". This sentence is unclear and should be rephrased.*

We propose to modify this sentence with this short paragraph: "To be able to make forecast during MAP-D PHASE and COPS experiments, an AROME domain was created over the Alps. This model was initialised using ALADIN-France, which was at the time the operational regional Météo-France model, taking its lateral boundary conditions from ARPEGE and its initial state from a three-dimensional variational data assimilation (3D-Var) scheme (Fischer et al. 2005). This AROME model was run during 6-months (June-November 2007)."

4. *P 1804, line 22: This sentence (…In Sect. 3, the…) is unclear and should be reformulated.*

We propose the following clarification: "In Sect. 3, the performances of AROME-WMED and AROME-France models are evaluated during the SOP1 over a common area. The comparison is based on Météo-France operational scores and on scores computed with additional surface observations from the HyMeX database."

5. *P 1805, line 2: Since different domains and grid points are used it might be useful to mention that both model have a 2.5 km grid. This is only mentioned for AROME-France.*

Lines 1 to 3 have been replaced by "AROME-WMED is based on AROME-France, which is a limited area model that rests upon non-hydrostatic equations (Bénard et al 2010). Both models have a 2.5 km grid and 60 vertical levels ranging from 10 m above ground to 1 hPa. They use a 1-moment microphysical…."

6. *P 1807, line 24: "…over a long period." Could this be more precise?*

We have replaced "over a long period" with "2 week period". To introduce the flow dependency into the background error covariances proves to be too costly, hence, a climatological background error representation (i.e. spatially and temporally averaged statistics over a 2 week period) is used in AROME-France and in AROME-WMED instead.

7. *P 1809, last paragraph: "EUCOS", "BLBPs" have not been defined before. Same for "IASI" on p1810, line 24. All acronyms should be checked for explanation before first occurrence and, as mentioned above, a list of acronyms should be included.*

A list of acronyms will be included in annex 2 as mentioned in the answer of the specific comment. EUCOS is now defined p 1804 (EUMETNET Composite Observing System) but

BLPB were already defined in page 1804 lines 10-11 (Boundary Layer Pressurized Balloons). We have carefully verified that every acronyms was explained at their first appearance.

8. *P 1811, line 25ff: It is stated that a code change has been performed during SOP1. Did this affect the results in a noticeable way? Was some verification done to show that this change in code does not affect the results? One or two sentences for clarification would be useful (either in this paragraph or in the Concluding Remarks on p 1823 where this issue is also addressed in line 18).*

For technical reasons, it was not possible to have a single version of the AROME-WMED during SOP1. The main changes in the AROME model reside in the revision of the cloud scheme with a realistic increase of intermediate cloudiness in addition to changes concerning observation use. As it was not possible to run simultaneously both cycles of the AROME-WMED, we could not quantify the impact of the code change on the AROME-WMED forecast during SOP1. However, the evaluation of the code change in the operational AROME-France suite has shown that low-level cloud fields were altered and the precipitation were slightly improved. These clarification sentences will be added in paragraph 2.3

9.P 1817, line 10: However, missing data do not occur…". This sentence is not clear. Does this mean that there is no day with missing data at all stations? Maybe this can be formulated more clearly.

Readers should be aware that only surface stations with daily precipitation for all the 62 days of the SOP1 are plotted (as in fig.6 for the temperature, some raingauge data were missing the date depending of the station). We chose to be very strict and to discard all stations with missing data to make a fair comparison between observation and model.

10. P 1819, line 10: Looking at Figure 16 the maximum of FBIAS is 1.8 (1.3 is stated). How is the rapid increase in FBIAS at higher thresholds in Fig 16 explained when compared with the decrease of FBIAS at high thresholds shown in Fig 10?

For the lowest x-axis thresholds (i. e. <20mm/24hr) corresponding to comparison samples exceeding 1000, the frequency bias is fairly similar for both SOP1 and SOP2, that is around at 1.2 - 1.3. For higher thresholds, AROME-WMED always exhibits a higher frequency bias reaching 1.8 for the 60 mm/24h. However, for threshold above 20 mm/24h, the smaller the sample (between 100 and 50 verification data), the larger the margin error, hence the larger observed differences are not significant. On the contrary, during SOP1, there were 700 verification data for the 60 mm/24h threshold, making the comparison more robust. We propose to add this comment in the text.

11. P 1823, line 2ff: One sentence explaining the reason why AROME-France gives better results in terms of temperature, humidity, wind and precipitation for SOP2 should be included (in this paragraph or in paragraph 4, "Forecast evaluation during the second Special Observation Period").

In our opinion, AROME-France benefits in this case from a more adequate B-Matrix than AROME-WMED during SOP2. Indeed, the B-Matrix of AROME-WMED has been computed over an autumn period whereas the AROME-France one was made over many different meteorological situations. The following sentences will be added: " During winter period, AROME-France model benefited from a B-matrix, computed over different

meteorological situations (including anticyclonic and stable situations), more representative of the meteorological conditions encountered during SOP2 (see paragraph 2.2.1)."

Textual comments:

1. *P 1803, line 20: "…of Mediterranean Sea…" should be ""…of the Mediterranean Sea…"*

The modification is accepted.

2. P 1808, line 20: Is "estimation of the estimation of the error…" correct?
"Of the estimation" has been removed

3. P 1814, line 10: "Rainaud et al. (2014)" should be "… (2015)" as in the list of references.
The modification has been made.

4. P 1822, line 22: "Once the field campaign over" should be Once the field campaign was over"
It has been corrected

5. P 1824, line 18: Better "Frequency BIAS (FBIAS) and Equitable Threat Score (ETS) …"
The clarification has been made.

6. P 1829, line 18: "Murphy, A. H.: A new vector partition of the probability score, J. Appl. Meteorol., 12, 595–600,1973." listed in the reference list is not mentioned within the paper. Remove or include reference somewhere in the text.
This reference has been removed.

Figures & Figure captions:

1. *P 1834, Figure 3: Use "(lower panel)" instead of "(left panel)".*

The modification has been made.

2. *P 1842, Figure 11: There are no stars in this figure. The figure caption should mention this in a way (e.g. by adding "No differences are statistical significant at the 90% level in this case").*

The figure caption has been changed into:
"Brier skill score computed in a neighbouring distance of 54 km for AROME-France (green) and AROME-WMED (black) computed with rain gauge data from France as a function of 6–30 h rain rate threshold during the SOP1 period from 5 September to 5 November 2012. In this case, there is no significant differences at the 90 % level."

3. *Some of the figures are very small and details are hard to see (especially Figures 6, 7, and 8) while others are quite large but only have little detail (especially Figures 11, 13, 14, 17 and 18). It should be checked if these figures could be resized.*

Figure 6 has been enlarged:

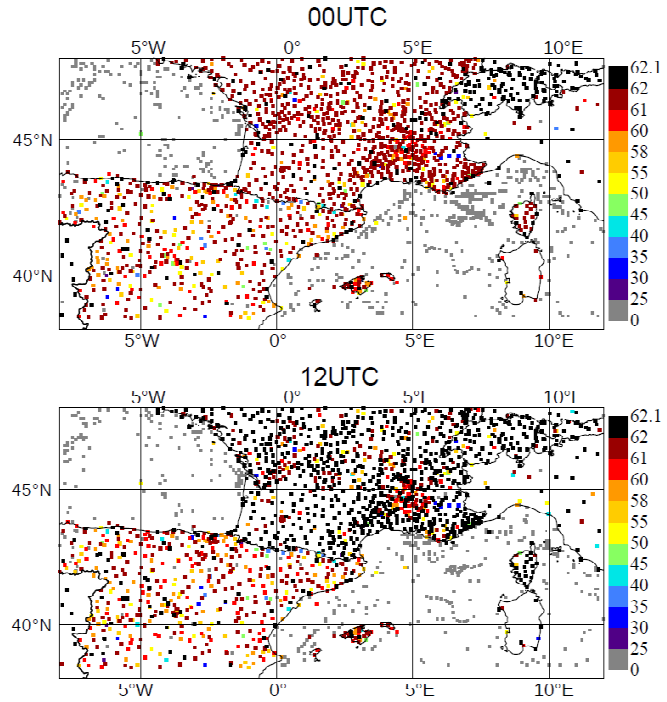

Fig6:

Figures 7 and 15 are adapted to portrait vision, but, if necessary, can be split into 3 panels. They have also been enlarged (please see first detailed comment).

Figure 8 could be split into two parts as shown below:

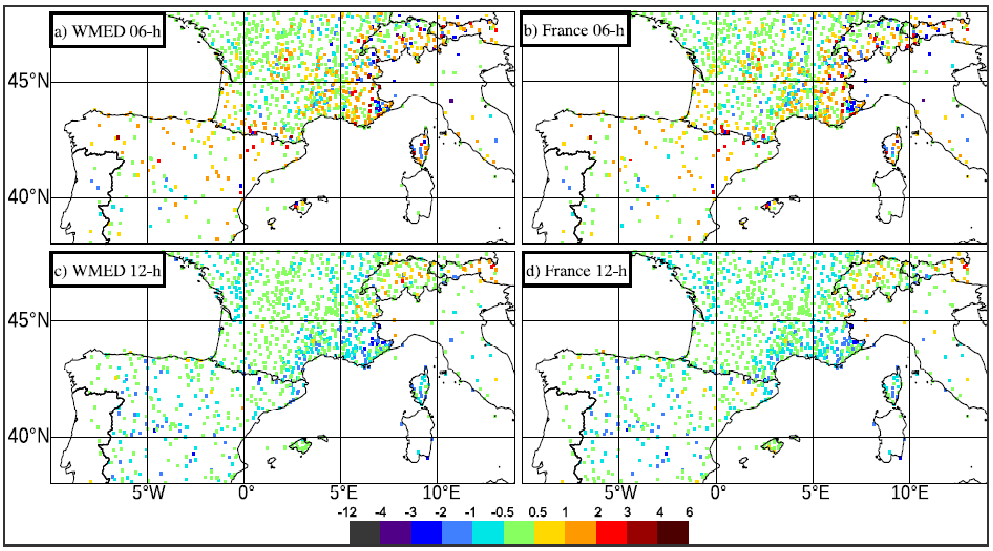

Fig8a:

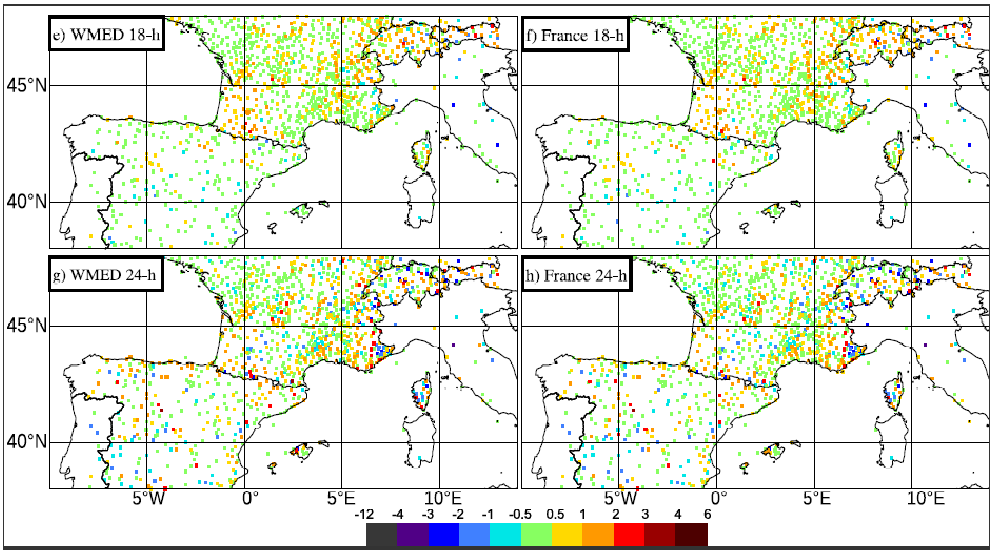

Fig8b:

We propose to interact with the Editor to propose the adequate size of the figures.