Review of the revised manuscript HESS-2019-223 "Regional ensemble forecast for early warning system over small Apennine catchments on Central Italy" by Ferretti R, Lombardi A, Tomassetti B, Sangelantoni T, Colaiuda V, Mazzarella V, Maiello I, Verdecchia M, Redaelli G

28 February 2020

Specific comments

1. *The statistical analysis of the results has been slightly reinforced by introducing the analysis of ROC. However, I am worried about some results shown in Figure 7. Please, excuse me, if I am missing some details and I misunderstand the result, but I think there is some mistake in data processing. Some basic remarks follow.*

   (a) ROC curves should be monotonically non-decreasing, shouldn't they? The probability of detection and the false alarm rate for a threshold (e.g., 10 mm) are always smaller than those for lower thresholds (e.g., 5 mm), aren't they? This is not the case for the blue curve in Figures 7a and 7b and for the red curve in Figure 7b.

The ROC curves have been recalculated due to a problem with the data files. The new results fix the evolution of ROC curves for ensemble forecast, except for the high resolution. This still shows some inconsistency between 10 mm/3h and 5 mm/3h thresholds because of high spatial variability of precipitation over a complex orography terrain. We point out that the aim of this paper is to present a meteo-hydro ensemble and eventually to validate the ensemble without performing a comparison between the HR and the ensemble. Therefore, we decide not to show the ROC curves for the HR simulations, which are not necessary for an evaluation of the ensemble but to evaluate the ensemble and the CNTR simulation. We rewrite the related manuscript section (pag. 9, lines 3-8)

   (b) How can be computed the area under the blue curves, which have multiple values on the y axis for a given interval on the x axis?

The new curves do not show this problem. Therefore, the Area Under the Curve (AUC) can be computed exactly. A table with the values of the AUC, calculated for two-time steps, has been added to the paper (pag. 9, table 1)

   (c) Why are the data of the threshold 15 mm not shown for the red and pink curves?

The 3 hourly accumulated rainfall does not reach the threshold value, that is why we did not show the 15mm related curve

2. *I am sorry, but I did not put enough attention to equation (1) in the previous review. This equation is evidently wrong. The following remarks should also be considered.*

   (a) Notice that freq = (t_{mem+1} + 1)/2, because $\sum_{j=1}^{n} j = n \times n(n+1)/2$

   (b) Why is the sum extended to t_{mem+1} if the number of members of the ensemble is t_{mem}? Is t_{mem+1} the control member?

a) We are deeply sorry, but we made a mistake in writing the equation and we did not notice it in reviewing the paper. Thanks for pointing it out! We correct equation 1 in the paper (from pag 6, line 30 to pag.7 line 5).

b) Yes, the summation is plus 1 because the control member is also added. We added a statement in the paper

3. *Page 7, lines 28 to 30. The sentence "all the time series show a variability among ensemble members much smaller than the difference between ensemble mean and observations at the maximum of the rainfall" implies that the variability among the ensemble members can be used to estimate the uncertainty among the predictions, but it does not provide an accurate estimate of the spread between the observed and the predicted quantities. Unfortunately, I think that this is a very weak point of the manuscript.*

We performed the computation of the rank histogram (Fig. 1) to verify the variability of the ensemble respect to the observed values and we found that there is too little spread as you suggested, but no bias. We would expect to have a uniform distribution for a properly calibrated ensemble even if this is not a sufficient criterion for thoroughly assessing the ensemble reliability (Hamill, 2001). Moreover, if a single step rank histogram is considered a closer distribution to the uniform is found, except for member 21 (Fig. 2).

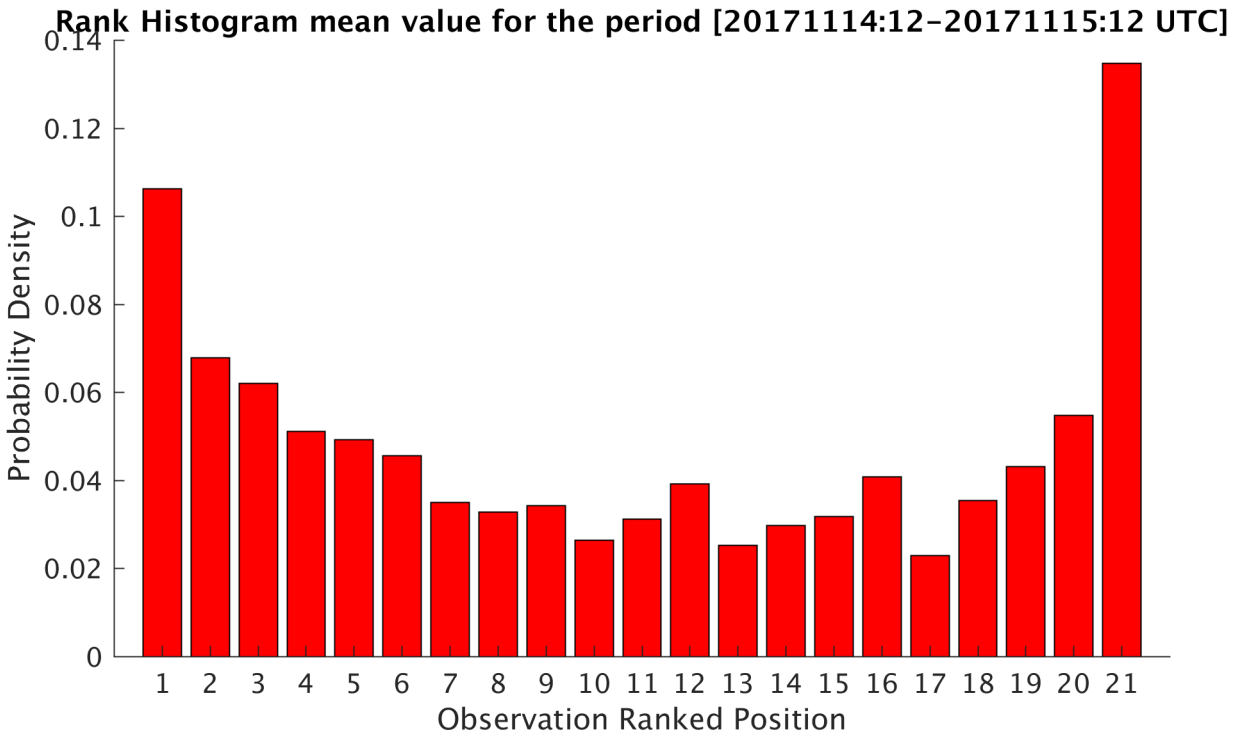

Figure 1

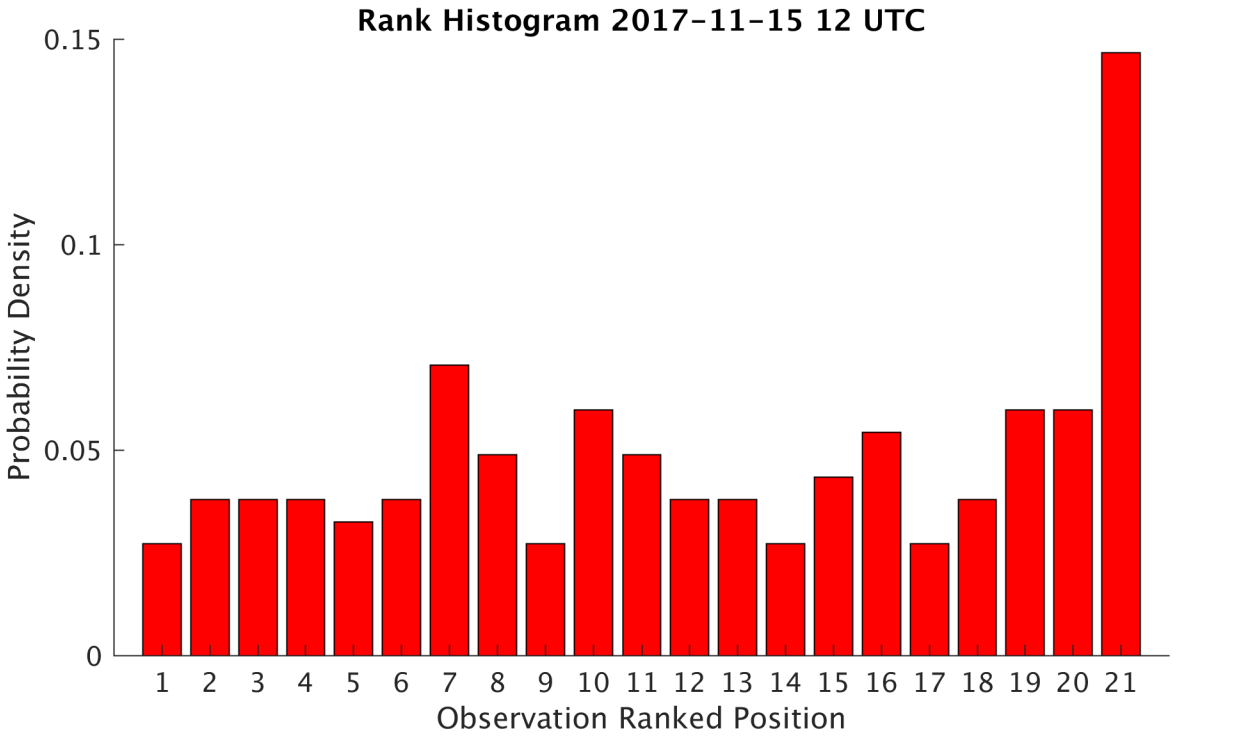

Figure 2

In this regard a comment has been added on pag. 8 lines 16-21.

We want to stress out that this paper aims to present a new meteo-hydro ensemble forecast, where the regional meteorological ensemble represents the driving field of the hydrological ensemble, which on the contrary shows a larger variability. In any case, being the meteorological ensemble an intermediate step to build the hydrological one, which is the final aim of this work, a conclusive assessment on the ensemble reliability can be expressed according to the benefits on the hydrological forecast, if any.

4. *In their answer, the Authors claim that "there are not yet operational system based on coupled regional meteorological ensemble and hydro ensemble in Italy", but this is not enough to support the innovation content of a scientific paper in an international high-quality journal.*

Probably, we did not state clearly the following point: actually, this paper is not on the regional meteorological ensemble, but on the development of a high-resolution hydrological ensemble on a complex topography, forced by a regional meteorological ensemble. The newly developed hydrological ensemble applied to the small Apennine catchments and the probability index (BDD_{prob}) specifically developed for assessing the uncertainty of the flood risk, are part of the innovation content of the paper. We added a few statements to make it clearer (pag 1 lines 4-8; pag. 4, lines 1-2; pag.15, lines 31-33; pag. 16, lines 1-5)

Moreover, to the aim of preventing misunderstandings in the purpose of the paper, we also changed the title to: "Regional meteo-hydro ensemble forecast for early warning system over small Apennine catchments on Central Italy"

3 Technical comments

1. *Page 5, lines 1 to 3. Acknowledge in the text the data source for SRT and rain gauges. Should not SRT be substituted with STR (Surface total rainfall)?*

On the DEWETRA platform is defined as SRT and we prefer to keep as it is

2. *Page 5, line 13. In which sense the following configuration provides the best results?*

The best results in term of precipitation forecast. We clarified this in the paper.

3. *Page 8, line 11. The ROC method was proposed much earlier than the book by Jolliffe and Stephenson (2012), which, therefore, could not be the best reference.*

Thanks for the suggestion, we changed the references in the paper for the use of ROC in atmospheric science (pag 21, lines 23-24).

4. *Page 8, line 15. I expect that radar data are spatially distributed over the area and do not need interpolation. Which interpolation method is used for the rain gauges data?*

The data are interpolated to the model grid, this has been done also for the radar data. Inverse Distance Weighting (IDW) Conservative method (Jones, PW. 1999) has been used for this purpose.

References:

Hamill, 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. Mon. Wea. Rev., https://doi.org/10.1175/1520-0493(2001)129%3C0550:IORHFV%3E2.0.CO;2

Jones, PW. 1999: First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates Mon. Wea. Rev. , 127, 2204-2210