Review of « Assessment of stochastic weather forecast of precipitation near European cities, based on analogs of circulation » by Krouma et al.

⇒ We thank the reviewer for the positive and constructive comments.

General comments :
The article assesses the skill of a stochastic weather generator to forecast precipitation in 4 cities of Western Europe. The SWG is based on random sampling of analogs of geopotential height. It was developed in another article by Yiou and Déandréis 2019, where it was applied to temperature. This study complements the latter for precipitation. As a refinement, a time embedding of 4 days is considered in the distance for the search of analogs (however the considered distance is not a mathematical distance anymore - see below). Skill scores are evaluated for lead times of 5 to 20 days. Results show positive skills up to 10 days. A comparison to ECMWF forecasts is provided but I have some concerns about this part (see below). The study is interesting, clear and well written. Precipitation forecasting is an important subject of research and I support the idea basing SWG on analogs. However I have several main concerns :
- The results are mainly shown for NCEP which used to cover a longer period than ERA5. However ERA5 is now available since 1950. Given its much better resolution, I recommend considering ERA5 for all the results.
⇒ We will extend the search of analogs to 1950 using the ERA5 database. Our idea was just to verify the robustness of our methodology to compute analog from different sources. We mention that the new database of ERA5 from 1950 to 1978 is still preliminary.

- I'm surprised that all the applied tests (Table 1 and p 14) have pvalues equal to 2.2^{-16} (I guess you mean 2.2 10^{-16}?). Isn't that strange ? More importantly, I doubt that Kolmogorov Smirnov gives such low pvalues given the differences in the CDF of Figure 6.
⇒ The pvalues of table 1 are obtained from the correlation test where the null hypothesis is that the correlation is equal to 0. Given the correlations observed, it is not unreasonable that the p-values of the tests are this low. However, we will instead use a confidence interval in the revision, since they provide the uncertainty around the estimate which is more appropriate in this case. We will use a confidence interval. For the Kolmogorov Smirnov (K-S) test, the p-values were given by the K-S test. In fact the K-S test with low p-values meant the rejection of the null hypothesis that the two distributions are equal. Hence, the Figure 6 and the conclusion of the KS test are consistent. However we can add the D value between CDF as a better indicator and the Student test between median of distributions.

- I'm concerned about the comparison with ECMWF forecasts since ECMWF are gridded data, whereas SWG is based on point data (ECAD). Have you considered comparing ECMWF forecasts with SWG based on E-OBS, since both have an horizontal resolution of 0.25°x0.25° ?

⇒ No, we have not used E-obs in this study. We did an extraction of the ECMWF forecast at the station level, we computed the CRPS and we did the comparison. We mention that we compared the values of CRPS (we found for ECMWF) to the CRPS provided by ECMWF for each month. It is true that E-obs have the same horizontal resolution as ECMWF forecast, but a resolution of 0.25°x 0.25° does not consider the information at the station level. As the E-OBS data are interpolations of ECA&D, both datasets are close near the four stations we consider (see table of correlation below). Moreover, we used E-obs to forecast precipitation near Madrid as an example, and the values of CRPSS/ correlation do not change that much. We will mention that in our results.

Station	Correlation between E-OBS and ECA&D (pearson)	Correlation between E-OBS and ECA&D (Spearman)
Orly	0.77	0.82
Berlin	0.69	0.77
Madrid	0.87	0.85
Toulouse	0.74	0.80

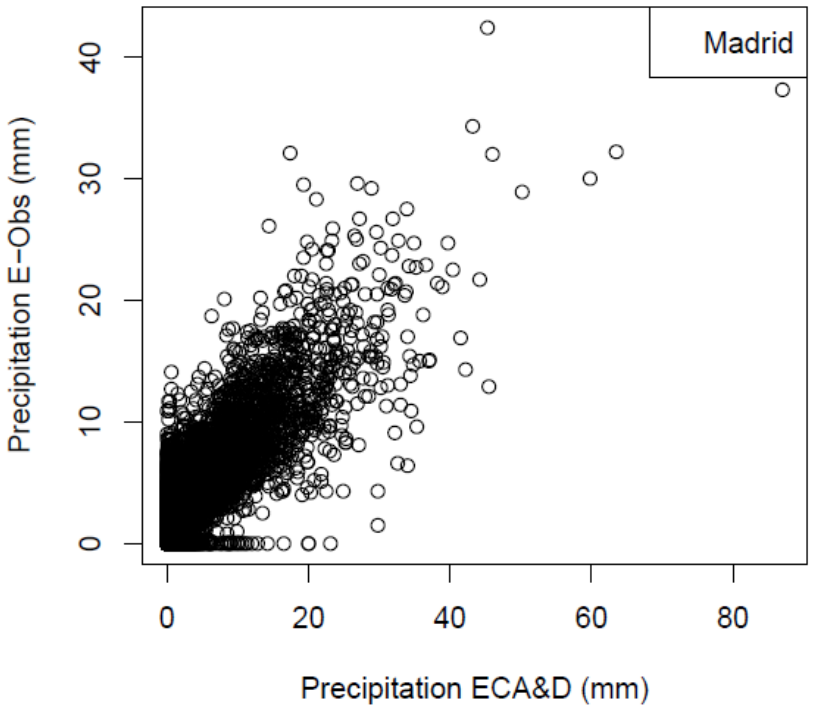

Fig.Scatter plot of precipitation of Madrid from 1950 to 2020.

- The evaluation of CRPSS conditional on weather regimes is interesting but I wonder whereas considering the weather regime of the last day of the sequence (t0+T) is representative for the weather regime of the whole sequence.
⇒ Thank you for this interesting suggestion, as explained below we will consider this and we will change our methodology: we will consider the "dominating" weather regimes for each simulation (rather than at the last day) and re-verify this relation between CRPS and weather regimes.

More minor, some references are missing (see below). There are issues in the units of CRPS. There are several equation issues.

Detailed comments :
- l 65 : Reference to Klein Tank is missing
- l 67 : please specify that ECAD provides point (station) data
- l 72-81 : actually ERA5 is now available since 1950.
- l 84 : Reference to Herschbach is missing. By the way, is it the right reference ?
- l 87 : so ECMWF forecasts have the same resolution as EOBS.
⇒ We will take care of the references and those five comments

- l 104 eq (1) : This is a good idea to account for several days in the distance, however D in (1) is not anymore a mathematical distance. Of course this is not mandatory for analog search, however why not using [sum_x sum_i {Z500(x,t+i)-Z500(x,t'+i)²}], which is a mathematical distance? By the way, could you pelase provide a comparison of the results with the Euclidean distance (based on 1 day) vs. the D distance (based on 4 days)? And why 4 days ?
⇒ We beg to disagree with the referee. The formula we use (square root of a sum of squares) is a distance in the mathematical sense, as it corresponds to a Euclidean distance in an augmented space. The formula given by the referee is NOT a distance because it does not satisfy the triangular inequality. An elementary counter example is to consider three distinct aligned points A, B, C, and B is between A and C. It should be clear to the reviewer that d(A,C) > d(A,B) + d(B,C) when using the above formula (with no square root), because there is an equality with sqrt(d). This contradicts the triangular inequality that is a property of mathematical distances.
We will show how the skill scores are improved with d=4, with respect to d=1. Yiou et al. (Clim. Dyn. 2013) argued for such delays for analog computations to simulate atmospheric fields. It is true that the distance based on 1 day is lower. However, 4 days help to better catch the persistence. That helps to obtain better skill scores for the forecast.
We prefer to use d<5 days, as we make forecasts for lead times T of 5 and 10 days.

- l 111-118 : explanations are quite confusing. I had to read Yiou and Déandréis to understand Please consider rewritting the method.
⇒ We will rewrite this part.

- Figure 1 : a) please consider placing the red rectangle somewhere else within the 30 days for clarity since its date is not necessarily the same as the target day. b) the largest window doesn't match the coordinates given l 80. I would be happy to see some results on the other windows of analogy. Otherwise I think it's not worth showing them. Also there are several syntax issues in the caption

⇒ We will modify the figure to make it clearer. For the results on the other windows of analogy, the small blue box is not different from the red rectangle. That is why we considered it. However there is a difference for the big blue rectangle, and we have put this information in table 2.

- l 122 : please specify that persistence is computed over year k (unlike the climatology which is computed over all years)

⇒ ok

- l 124 : « control forecast » I dont't understand

⇒ Indeed. It is not the appropriate word. We meant climatology and persistence forecasts. This will be changed.

- l 137 : I guess that averaging the 100 trajectories smooths out the predictions. So at the end, is there a real gain (in terms of CRPSS) compared to considering only one analog ? (maybe that's already studied in another article, I haven't checked)

⇒ The CRPS/CRPSS are based on the 100 trajectories, not by considering only the mean. The "averaged" trajectory is just for illustration purposes, and does not influence the CRPS score computations. The correlations are computed over the averaged trajectories.

- l 150 : P(x) should be P(x,t) for day t. Please also rephrase the sentence

⇒ OK.

- l 153 eq (2) : the equation is confusing. Should be CRPS(P,t) and t should be in the right side as well. The inferiori limit is 0 for precipitation.

⇒ we will verify and correct the equation to make it consistent.

- l 159 : seasonality → climatology

⇒ OK.

- l 162 eq (3) : Equation issues. there should a sum over the days (or mean) in the numerator and denominator

⇒ We will correct all this.

- Table 1 : is it Pearson correlation ? I'm surprised that all pvalues equal 2.2^{-16} (I guess you mean 2.2 10^{-16}?).

⇒ It is a Spearman (rank) correlation. We will use confidence intervals, which are more informative.

- l 198-201 : this paragraph should go after l 204. Please refer to Fig 3.
⇒ **We will correct this.**

- l 212 : syntax issue
⇒ **we will correct this.**
- Table 2 : I guess this is for NCEP ?
⇒ **yes, indeed.**
- l 228 : remove the brackets
⇒ **OK.**

- l 230 : so you obtain 100 classifications. How do you deal with that ?
⇒ **We will clarify this paragraph. Indeed, we followed the procedure of Yiou et al. (NPG, 2008): we obtain 100 classifications. Then we classify the centroids of those 100 classifications and determine the most probable classification. This Monte-Carlo procedure helps "stabilizing" the classification into weather regimes.**

- l 241 : is the weather regime at time t0+T representative of the sequence from t0 to t0+T. Why don't you consider the most frequent WR within t0 to t0+T ?
⇒ **This is a very interesting suggestion. Actually we took the weather regime at t0+T and t0+T/2, which is not useful, as you point out. We decided to consider the most frequent weather regime in each simulation and determine how that will change the results.**

- Fig 3 : please consider plotting both reanalyses on the same plot for ease of comparison. Actually for persistence ERA5 seems to give larger CRPSS. Caption : persistence in lowercase letter. Please use either « reeference » or « baseline » along the article. For the boxplots, why don't you consider correlation with the mean instead of the median (as the predictions of SWG) ? (but it should not change much)
⇒ **We will add another plot containing both reanalyses, and we will take care of using reference in the whole paper.**

- l 243 : don't you mean below the 25th quantile ? Where is this used ?
⇒ **We noticed that all the values of CRPS below the 75th quantile are very low, so it doesn't matter to consider this quantile. This will be changed with the redefinition of the "most frequent weather regimes".**

- Figure 4 caption : blank space after (BLO)
⇒ **OK.**

- l 251 and followings : I see there are differences depending on the WR but it seems to depend on the city. Can we have explanations why, e.g. CRPS for BLO is better in Orly ?

⇒ We related this to the difference in the local weather. However as we will make changes in the way of attributing CRPS to weather regimes we will see how that will change.

- Figure 5 : Units of CRPS are not mm. What is the lead time here ? Given l 243-244, I would have expected here to see boxplots for the two classes of predictability.
⇒ lead time of 5 days, that's what we are showing . This will be clarified. The units of CRPS are mm/day.

- l 257 : what is the reference for CRPSS ? (I guess climatology)
⇒ yes climatology. This will be clarified.

- l 261 : do I understand correctly that in ECMWF forecasts, CRPSS is given for the whole of Europe whereas CRPS are avalaible at every grid point ? As said above, I find difficult comparing the skills of ECMWF vs SWG given that the horizontal resulution is different (0.25°x0.25° vs point data). Comparison of ECMWF with EOBS at the same resolution may be easier .
⇒ Yes ECMWF forecasts are given for the whole of Europe, we extracted forecasts in single points which have the same coordinates then the studied stations, then we did the comparison. ECAD data are provided at station level (This is what interests us in this study). Moreover, the E-OBS are made from the ECAD data. By comparing the E-OBS and ECAD there is a strong correlation between data.

- Table 3 : please specify the reference. You may want to add here the CRPSS of Europe with ECMWF.
⇒ climatology. This will be clarified.

- l 264 : CPRSS are actually hard to compare since they are not based on the same data (different resolution)
⇒ We cannot find literature that explains how CRPS/CRPSS values should depend on data spatial resolution. Those scores are connected to the temporal resolution of variables and the size of ensembles. We will do the simulations with E-OBS data (that yields the same horizontal resolution as the ECMWF forecast) for comparison purposes. The main difference stems from the ensemble size.

- l 266 « We found... » I don't understand the sentence (syntax issues). Anyway according to the CDF of CRPS in Fig 6, ECMWF seems significantly better (a much larger proportion of low values)
⇒ We will correct the syntax issues to clarify this point.

- l 270 again 2.2^{-16} ? Anyway, I think something's wrong here because the CDFs in Fig 6 do seem different. A difference of 0.2 between CDFs is large actually.
⇒ small p-values of Kolmogorov Smirnov indicate that the null hypothesis is rejected, and it is in agreement with the D (difference between CDFs)

- l 276 and Fig 7 : I think something's wrong because ECMWF shows a much larger proportion of small CRPS for Toulouse and Madrid (see Fig 6). The difference in CRPSS for Orly between 5 and 10 days is very surprising.
⇒ **We will compare what we find for precipitation with temperature in order to better understand this relation; for Toulouse and Madrid, we will verify this.**

- l 300 : designed
⇒ **ok**

- Some references are missing. There is no year for Cassou.
⇒ **We will take care of the references.**