Review of the manuscript "Assessment of stochastic weather forecast of precipitation near European cities, based on analogs of circulation" by M. Krouma et al.

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

This is a very interesting manuscript, owning a good potential to become a high impact paper with positive repercussions on different societal sectors. A stochastic rain generator is produced exploiting the relationship between Z500 and precipitation in different European cities. The work is worth publication, but it needs a substantial revision about three distinct points:
1 - an improved description of the methodology is needed, in order to better understand the workflow and some of the choices that have been employed.
⇒ We will take care of this and rewrite the methodology in order to be more self contained.

2 - the use of the prolonged ERA5 dataset (since 1950) is urged, in order to understand whether the differences in skill with NCEP are actually due to the length of the analog database, or to the database itself.
⇒ The first reviewer also pleaded for this. We will extend the analogue search in ERA5 to 1950. It should be acknowledged that a few colleagues at ECMWF advised against such an extension due to an undocumented (yet) potential discontinuity before 1979.

3 - a thorough and comprehensive revision of the English language is needed. Many subject-predicate inconsistencies, missing s', wrong sentence structures make some parts of the manuscript very hard to read.
⇒ OK. We will be more careful with the English language in the revision.

Specific comments are stored in the attached files. In bold font, those that pertain to the above-mentioned major observations.
Please also note the supplement to this comment: https://gmd.copernicus.org/preprints/gmd-2021-36/gmd-2021-36-RC2-supplement.pdf

Lines 79-80 and figure 1. Why exactly that region? Is there a process-knowledge approach behind this choice, a literature review or a regression/correlation between Z500 and rain over each station was applied? Whatever, the choice, it should be justified by means of references or, in the last case, with a graph/map that certifies the link between Z500 and precipitation.
⇒ We justified our choice with references in the introduction. We mentioned some previous works where the relation of Z500 and precipitation was explained. We will add references to reflect this.

Line 98. Why did the author choose exactly 20 analogues? Would not be better to base the choice on a maximum Euclidean distance?
⇒ The choice of 20 analogues was based on experimental experience. First we considered 20 analogues to ensure that we have enough analogues dates for simulations,

second, we do not find changes on the Euclidean distance when we exceed some number of analogues. A theoretical study of Platzer et al. (2020, https://arxiv.org/pdf/2101.10640) shows that for complex systems the use of a large number of analogues (exceeding 30 analogues) doesn't make a big change for a forecast with analogs. Basing the results on a variable number of analogs (with a threshold distance value) would require complicated tests to choose an appropriate threshold, which might be season dependent.

Line 100. The 4-day time embedding is not clear to me. Why is it necessary? Why does it preserve the temporal derivative of the atmospheric field? Please explain.
⇒ **A 4 day embedding enhances a better simulated persistence and yields better skill scores for the forecast. This was explained by Yiou et al. (Ensemble reconstruction of the atmospheric column from surface pressure using analogues. Clim. Dyn., 41, 1419-1437, 2013).**

Lines 113-118. Despite the mechanism is quite clear to me, the sentences "In order to go […] precipitation between t0 and t0 + T" are not well formulated. Since this is a crucial part for the understanding of the method, I would rephrase and expand this part.
⇒ **This will be better explained in the revision.**

Line 120. "of the properties" is redundant.
⇒ **OK.**

Line 123. More than the average value, the persistence consists in the anomaly between t0 - T and t0. Also, the climatological forecast takes.
Lines 128-134. Again, it is not clear upon which basis those domains (in Fig. 1b) have been chosen and the final domain selected among the four attempts. Also, the entire paragraph needs a language revision.
Line 150. A verb is missing (meet?).
⇒ **We will take care of those three comments. This will be rephrased and we will correct the mistakes.**

Figure 2. A visual legend is needed. Also, what's in the y-axis? Are the 5th and 95th percentiles calculated over the 1948-2019 time series?
⇒ **We will add a visual legend to make the figure clearer. The y-axis represents the mean of the precipitation computed over each lead time. And percentiles are computed over 1948 -2019.**

Line 184: six??
⇒ **We mean four (a mistake).**

Lines 201-205. It seems to me that part of the methodology is described here, where a description of results is expected.
⇒ **We will move this part to a new subsection on "parameter optimization".**

Line 206 and following. Very little description is given for figure 3. First of all, I think it is useful to illustrate all the stations in the main text, instead of showing only Orly while relegating the others to the Appendix. This is one of the main results of the study and deserves a better stage (instead, I would recommend to place Fig. 4 the Appendix, since the WRs are not a result of this study). Besides, fig. 3 and fig. A1 show a very interesting characteristic that should be discussed: most of the times, in fact, the summer SWG forecast vs. persistence improves with lead times, which is somewhat unexpected. Any thoughts about it?

⇒ **We thought that it is better to illustrate one station in the main text in order to not repeat the same information. Regarding the persistence, we find the same thing for other variables, and we think that could be related to the seasonality.**

Line 217. De Bilt?
Line 225. They help describe.
Line 232. Missing year for the Cassou reference.
⇒ **we will correct this and the missing reference.**

Lines 246-255. Fig. 5 (B1) should be described with more care. What do "Good forecasts (low quantiles of CRPS)" and "The low quality forecasts (high quantiles of CRPS)" mean? The caption for figure 5 is totally unclear and does not describe the plot. On the top of the blocking bar (panel b) a group of dots appear.

⇒ **We will redo the plot, what we meant here is that we attributed high and low quantiles of CRPS with respectively good and bad quality of the precipitation forecast as by definition the values of CRPS close to zero indicates a good forecast. The captions will be clarified.**

Lines 257-276. This paragraph is hard to read, there are many inconsistencies between subjects and predicates and other grammar errors. Also, the first paragraph is not clear: what is it meant to demonstrate? Maybe the lower skill of the ECMWF forecast? The latter is calculated over the entire European domain, how can it be compared with a forecast over single stations? A table with both ECMWF and SWG CRPSS would be more informative than a few words, if the authors find the way to make a fair comparison between the two.

⇒ **We will take care of the grammar errors and clarify this part by adding a table with CRPSS. We extracted the ECMWF forecasts at single points that have the same coordinates as the studied stations and we made the comparison.**

Line 279. The input of our model was analogs of geopotential heights at 500 hPa (Z500). This sentence should be rephrased.
⇒ **OK. we will correct this.**

Line 283. I cannot accept this conclusion. The only way to test it is to compute the analogs with the 70-year ERA5 dataset, now available since 1950. This is a very important test that should be included in this study, because it clarifies the role of the different reanalyses as

well as the role of database length.

⇒ As we said before, we will compute analogs from the ERA5 dataset including data from 1950 to 1978. At a first glance, the results with ERA5 (1950-2019) and NCEP (1948-2019) are very similar.