Thank you very much Dr Kratzert for your comments.

Reviewers' comments are shown in black. Authors' responses are shown in green

> During the review process, we identified an error in the computation of one of the hydrological indicators ($VCN30_5$). We corrected the error. The change rate values of the VCN30 are now more consistent with those of the QMNA (both low-flow indicators) than before. We will update the figure 7 and the values of changes rate of the $VCN30_5$ in the results section of the manuscript. We will also update the appendix C.

The manuscript, by Laurent Strohmenger et al., presents results of a survey, in which 42 hydrologists were asked to annotate anomalies in streamflow time series. In my eyes, the main results of this manuscript:

1. A dataset of anomalies in streamflow time series annotated by human experts.
2. Another proof of the subjectivity and inconsistency of human experts, when tasked to rate/compare/annotate hydrographs.
3. Recommendations for future studies of this kind.

Overall, the manuscript is very well written, easy to follow. Given the already published comments by Martin Gauch, I have very little (see below) to add and recommend the publication of this manuscript after considering these minor comments.

Data sharing.

The authors state in L 354ff. The following:

An automatic detection of anomalies could avoid these issues of subjectivity and weariness. Using the bias between model simulations and measured time series could be a starting point for identifying potential anomalies. Unfortunately, to our knowledge, these techniques still require improvements. Such an algorithm should be flexible regarding the types of anomaly to identify, and might be trained for each study to avoid the risk of removing data of interest (e.g., using a visual inspection such as the one reported in this study). Ideally, hydrologists should share a common library of anomaly types such as suggested by Wilby et al. (2017). "

I 100% agree with this statement and e.g. the automatic quality control functions in the GSIM paper yield very suboptimal results. I therefore value that the authors opted for publishing the annotations from their study, which could be beneficial when testing future approaches for automatic detection algorithms. In this regard though, I agree with the review comment by Martin Gauch, that it would be very helpful if you would also include the streamflow time series in the published data, and maybe the GR5J simulations. As Martin Gauch mentioned, the linked homepage is in French, which e.g. I do not speak. I tried to use a translation tool but after 10 minutes of trying to figure out how to download data, I gave up. It also seems like there is no API access for downloading the data, which makes the effort to get the ~600 station time series quite cumbersome. If this data already exists in the hands of the authors and there are no constraints from the data provider that prohibits the publication of their data, then why not include the streamflow time series as well. Otherwise I see limited use in the published annotations, which would be a shame.

Thanks you very much for your kind words about the paper.

As suggested to Martin Gauch's commentary (RC1):

Unfortunately, we are not allowed to provide for download data that we did not produce and that we do not own.

Regarding hydrological data, we would like to mention the existence of the hub'eau API to collect streamflow time series over France from the HydroPortail. It is a French service that aim at simplifying access to water data:

https://hubeau.eaufrance.fr/page/api-hydrometrie

A tutorial is available here (unfortunately also in French, but translation tools seem to be efficient in providing an English version):

https://hubeau.eaufrance.fr/page/api-qualite-cours-deau-tuto

We will mention the hub'eau API in the data availability section.