Review 1:

Similar to clustering studies that meticulously compare cluster patterns, the paper is at times hard to follow for the reader, due to the tedious nature of going through detailed dissemination of specific patterns found in data. While this reflects the subjective impression of me as a reader, from the standpoint of the reviewer I can attest that the paper has a clear structure and exhibits straight and deep thinking about governing processes that is valuable for publication.

Its scientific significance is good: There are numerous studies tracing streamflow and groundwater time series patterns to overall governing processes – but few compare streamflow and groundwater side-by-side.

The scientific quality is fair, but can be good once the main general comment below addressing selective reporting of results is addressed.

Presentation quality is good and could be excellent if language and reasoning would be more concise and on point, and more aware of clearly communicating complicated mental concepts and connections instead of assuming the reader being "in the know" – at least in some parts of the results section. However, I would not draw necessity for improvement. It is good enough.

- Thanks for the thorough review of the paper and providing many helpful comments. We did our best to take up the criticisms to improve the paper.

General comments

There is a fundamental problem of selective use of analysis methods in this paper: for example, the author does not report correlations of PC2-6 with raw time series means, but only the correlation of PC1 raw time series means is reported (line 173). As another example goes the analysis of autocorrelation (only PC2 & PC4, Figure 3 & Figure 6) or correlation with trends (only PC2, Figure 4) or the reporting of distribution of loadings w.r.t. aquifer type (only PC4, figure 6) or overall the selective reporting of correlations to individual catchment attributes throughout the paper, and several other examples. This is to say that in this paper, the (shown) application of methods is selective, and while selective reporting makes for a good storytelling, the conclusions draw in the study (PC1 captures "mean behavior", PC2 the "dampening" etc.) can only be drawn from comparative analysis, e.g. when correlation with time series mean is much better for PC1 than for PC2-6, and analogously for the other examples. It may or may not be that comparative analysis has been done in the background and only the significant results are shown (line 230 indicates that – stating that autocorrelation "not shown" for PC3). But to make sure that the results are not selective constructions, uniform method application and reporting across all PCs and is required. For the most part, at best in the appendix, to not blow up the paper, although some referencing in the main text of the manuscript will be needed. To be clear, I am not suggesting scientific misconduct. The displayed results seem to be reasonable, however for reasons of scientific rigor and for the reader to be able to reconstruct the conclusions properly, the paper simply needs verification via negative elimination through comparative analysis for reasons of completeness.

- For each of the depicted principal components we provided only those pieces of information that we deemed helpful to test our hypotheses for plausibility, e.g., correlation with other variables, the effect of single principal components on autocorrelation or trends of the respective time series. These effects have been checked for all of the depicted principal components in the same way, but for the sake of brevity only the interesting results are presented in the manuscript. We will provide that information in the supplement.

Minor comments

Introduction section

Line 33: WFD citation missing; maybe add half a sentence of explanation what WFD is for intercontinental audience; explain that for WFD the definition of water bodies includes GW

- The reference will be included in the manuscript as well as an additional explanation: "The European Water Framework Directive (WFD) provides a legally binding framework for all member states of the European Union. It aims at achieving and maintaining a good water quality and good water quantity status in freshwater systems, coastal waters and groundwater bodies. Among others, the WFD demands a regular inspection of the "good quantitative" status of water bodies. To that end …"

Line 33-45: In addition to citations to studies diagnosing the pitfalls of heterogeneities in groundwater, the author is advised to include citations to papers that also try to grasp it, e.g. this Gothenburg-based research group – but happily also others:

- Barthel, R., Haaf, E., Giese, M., Nygren, M., Heudorfer, B. and Stahl, K., 2021. Similarity based approaches in hydrogeology: proposal of a new concept for data-scarce groundwater resource characterization and prediction. Hydrogeology Journal, 29(5), pp.1693-1709. https://doi.org/10.1007/s10040-021-02358-4

- Giese, M., Haaf, E., Heudorfer, B. and Barthel, R., 2020. Comparative hydrogeology–reference analysis of groundwater dynamics from neighbouring observation wells. Hydrological Sciences Journal, 65(10), pp.1685-1706. https://doi.org/10.1080/02626667.2020.1762888

- Haaf, E., Giese, M., Heudorfer, B., Stahl, K., & Barthel, R. (2020). Physiographic and climatic controls on regional groundwater dynamics. Water Resources Research, 56, e2019WR026545. https://doi.org/10.1029/2019WR026545

- We are sorry to have overlooked these papers (see more detailed comment to the Discussion section below) and will incorporate them in the manuscript.

Line 79-80: for signature-based analyses of hydrologic similarity, essential (and more suitable) citations that need to be included to connect the reader to the large body of literature of earlier decades on hydrologic similarity, are the following; they are also good evidence for the bold claim that signatures are "hardly ever checked for relevance".

- Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications, 19(2), 101–121. https://doi.org/10.1002/rra.700

- Olden, J.D., Kennard, M.J. and Pusey, B.J., 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. Ecohydrology, 5(4), pp.503-518. https://doi.org/10.1002/eco.251

- We are aware that there are some examples for a check for relevance but consider these examples rather scarce. We admit that we overlooked these papers (see more detailed

comment to the Discussion section below) and will incorporate them in the manuscript as counter examples.

Line 68-71: This attempt can brilliantly be motivated based on literature as well, e.g. by this older call to interdisciplinary studies on hydro(-geo)logical similarity:

- Barthel, R., 2014. HESS Opinions" Integration of groundwater and surface water research: an interdisciplinary problem?". Hydrology and Earth System Sciences, 18(7), pp.2615-2628. https://doi.org/10.5194/hess-18-2615-2014

- We are sorry to have overlooked this paper (see more detailed comment to the Discussion section below) and will incorporate it in the manuscript.

Line 78-83: Maybe reiterate the method of choice (PCA) here.

- Much of the relevant information including references to the literature is provided in the Method section. For the sake of brevity, we will add only two phrases here: “In numerous studies principal component analysis has proven its great potential to extract the prevailing features in large sets of interrelated variables or time series. This data-driven approach allows to differentiate between generic and site-specific features without any pre-defined assumptions.”

Data section

Line 101-114: Would be informative to state the total number of samples and total number or share of interpolated samples that you had in there in the end.

- The number of dates per time series is given in line 109, the number of sites in line 113 – 114. Daily mean values of discharge have been calculated based on readings at 15 or 60 minutes intervals. In 44 out of 3,248,865 cases (15,695 days times 207 sites) one or more readings per day were missing. Only in one case did the gap cover a whole day, requiring interpolation between the preceding and the following day. Note that only 1/7 of these discharge data have been used eventually, that is, the values from every Monday. Groundwater data mostly exhibited daily resolution, except for seven wells with weekly intervals. However, data gaps were more abundant. In total, 75% of the groundwater measurement days matched exactly the final time axis of weekly data. All other had to be interpolated, usually over a time span of a few days. As described in l. 105- 108 the maximum length of the gap to be filled by interpolation was defined by the autocorrelation of the respective time series.

Line 108: which analogous approach? Standard? Or citation? Unknown to me.

- Standard autocorrelation analysis requires gapless time series with regular intervals and thus could not be applied to many of our time series. Thus we followed an approach analogously to that used in semivariogram analysis in geostatistics where irregular (spatial) spacing between data points is the rule rather than an exception: Pairwise correlation between time series is determined only for those lag widths where data were available, and then the results were summarized for different lag width classes.

Method section

Line 130: “it has rarely been used in hydrology” -> that is simply untrue, as per the very large body of literature that opens up to the reader once they trace the literature starting from the Olden 2003/2012 paper provided above.

- Principal Component Analysis is in fact widely used in hydrology to reduce the dimensionality of large sets of hydrograph or catchment indices. Olden et al. (2012) is a nice review in that regard. But that’s not what this phrase is referring to. Rather, it refers to PCA application directly to a set of time series without the necessity of prior definition of any metrics. That approach is widely used, e.g., in climatology where it is coined as Empirical Orthogonal Function. But it is still not very common in hydrology. Otherwise we wouldn’t encounter so many problems publishing respective papers.

Line 140-143: Belongs to data section.

- We consider data normalization an integral part of the parameterization of the PCA. PCA could be applied to non-normalized data as well, but that would change the interpretability of the results.

Line 145: “unlike in some other studies”: cannot be stated like this without citations. Citation or rephrasing necessary.

- In fact weighting loadings by eigenvalues is the default setting of the R routine that we used. We wanted to make clear that our approach differed from that without going further into details which would go beyond the scope of the paper. We suggest to delete that clause.

Results section

See general comment about selective vs. comparative analysis

- See our reply to these comments.

Discussion section

Reading through this discussion, I can only double my suggestion above to read the works of the referenced Gothenburg group, especially since they use data from the adjacent region. Haaf 2020 can be brilliantly connected to the overall discussion points, and the Giese 2021 paper to the subsurface section 5.3 especially, which is lacking discussion with literature overall. Also the Barthel 2014 paper belongs next to the Berkowitz and Zehe 2020 in line 444. Crazy how people who write about the same can be completely unaware of each other (Berkowitz/Zehe and Barthel, that is).

- Although we do our best to keep an eye both on hydrological and hydrogeological research we have to admit that we overlooked these papers. Within the PUB framework catchment similarity has been studied and discussed extensively. But to the best of our knowledge it has been much less of a burning topic in hydrogeology so far. Thus, we focused our literature search on the hydrology community. Thanks for drawing our attention and that of future readers of the manuscript to this work. In spite of good will on either side bridging the gap

between these two sub-disciplines obviously is not without frictions when it comes to practice.