Reply to referee 1

Comments of referee 1 are in black.

Replies of the authors (AR) are in blue.

Comments of referee 1

https://hess.copernicus.org/preprints/hess-2024-172#RC1

Review of Technical Note: An illustrative introduction to the domain dependence of spatial Principal Component patterns by Lehr and Hohenbrink.

Major comments

This manuscript attempts to extend the study of how analyzing data on various shaped spatial domains affects the principal component loading patterns. The extension is both in content, as new material is added to the existing literature and the authors hope to gain the audience of hydrologists who, by and large, have not been exposed to such a concept. The importance of the work lies in several areas (expanded on below) but the key one is that if the PC loading patterns match those that are expected to arise from the shape of the domain, rather than the covariance fields, the recommendation should be a full stop on continuing. Therefore, understanding domain dependence is a necessary, but not sufficient condition, for physical interpretation of PC loadings.

Let me add that I like this paper and believe it can be a useful addition to the literature, helping analysts to interpret their eigenanalyses. Therefore, I hope the authors view my extensive comments with that in mind. If I come across as opinionated it is because of my lengthy work in this area and if it seems direct, that is my nature. Regardless, I like this manuscript and hope it gets published after further revisions.

AR: Thank you a lot for your comments and the appreciation of our work. Your comments are really helpful and the literature you suggested as well. Thank you also for the kind and respectful personal comment directly above, putting the degree of detail and accuracy in your comments into context.

We like to take the opportunity to put our replies to your comments into context as well. You spend a lot of effort to examine our work in detail and you offer us a lot of additional information in a high degree of detail. We highly appreciate that. If we refrain from some of your suggestions it is sometimes simply because it goes beyond the scope of what we want to offer here in the journal HESS. Our work is meant as an illustrative introduction for PCA users in the field of hydrology who probably never heard of domain dependence and its effects on the explained variance distribution, contrasts of the spatial PC patterns, degenerated multiplets, etc. - not to mention different scalings of PCA eigenvectors or scores, the congruence coefficient, etc.

To our feeling, the manuscript is already quite full with all the aspects we included. In fact, during the writing process we were even discussing to omit the whole multiplet aspect. Thus, we prefer not to extend it further with new aspects. Furthermore, we want to find a balance with referee 2 who advocated mainly for restructuring the current manuscript instead of extending it further.

Now for the general comments. The paper builds upon the pioneering work of C. Eugene Buell. Those papers are cited. Buell (1979) left the reader with this final thought on the subject of domain dependence in the last line of his conclusions, stating that unless domain dependence was accounted for, on interpreting EOFs, "*Otherwise, such interpretations may well be on a scientific level with the observations of children who see castles in the clouds*". That is a pretty direct and strong statement. Digging deeper into why that can occur, the manner in which individual EOFs were being analyzed in the 1970s,...,2020s is by inferring physics by visual inspection of the magnitudes and gradients of the EOFs when plotted on maps. There was no external or internal validation of the patterns, only conjecture. With over 50 years of this practice, little attention was paid to whether this was a wise idea and thousands of such EOF studies emerged, with claims of the importance of the magnitudes and shapes of the patterns, many of which looked suspiciously like those patterns Buell generate. However, we should be wiser today and the authors are telling the investigator that if the covariance fields vary across a given domain shape but the same basic Buell patterns emerge, perhaps it is castles in the clouds rather than physics. However, there may be something more than a chimera, a mixture of signal and domain dependence. We come to learn later in the manuscript that a third confounding factor, namely the degeneracy of PC loading patterns with closely spaced eigenvalues, playing a role. It is good to see these factors considered.

Next, let's discuss PCA as a technique. According to those who understand the method, there is general agreement that PCA is useful for data reduction. In other words, in the type of analysis in the manuscript, the time series at n gridpoints or locations can have their covariances explained in k PCs where k<

1. Given the above prologue, the authors on lines 408-409 discuss "heavy constraints" of PCA that inhibit physical interpretation. To that good list, I'll add that it has been shown the leading PC, by virtue of the constraint of maximal variance can pull multiple unrelated sources of variation onto that leading PC, confounding physical interpretation. This should be added. The Karl and Koscielny citation (in your reference list already) shows this in their Appendix. Further details are given in the annotated manuscript (attached).

AR: Thank you very much. We will add it.

2. There is a general lack of agreement on terminology for eigenmodels, that leads to massive confusion among users of these techniques. At first when reading this manuscript, I thought the authors were applying EAOFs, only to change my opinion later in the manuscript that they were applying the PCA model. The original paper where EOFs were named EOFs, is generally attributed to Lorenz (1956). However, in that report, Lorenz refers to the displays as EOFs of space, and EOFs of time, to define what have now mutated somewhat into what are called "EOFs", and "Principal Components", respectively. Assuming a spatial analysis, those EOFs of space are unit length (sum of the squares of each EOF's coefficients = 1), whereas the EOFs of time are orthogonal vectors, each with a mean of zero and variance equal to the associated eigenvalue. In contrast, the PCA model, generally attributed to both Pearson (1901) and more fully to Hoteling (1933). weights (postmultiplies) the unit length eigenvectors (EOFs) by the square root of the corresponding eigenvalue to give "PC loadings". That seemingly minor change in the spatial patterns (keeping with the definition of space and time given for EOFs)

results in the time series calculation and properties being different. Those time series in the PCA model are called "PC scores" and have mean 0 and variance 1. They are also orthogonal. Flip the space and time definitions of these displays if the analysis is temporal. Because the two models result in different space and time patterns, they cannot be compared directly and the precise equations used are necessary to attempt to reproduce the findings of others. **I urge the authors to state clearly what model they are using immediately after the introduction and show the equation**. The situation becomes more complicated as users of these techniques tend to grab EOF/PCA code off of various statistical packages or Python code libraries, that often mislabel the results, never checking the specifics, thereby perpetuating the confusion. For the current paper, one must know if the analyses are applied to EOFs (unit length eigenvectors) or PC loadings (unit length eigenvectors postmultiplied by the square root of the corresponding eigenvalues). Further, it would be helpful to know if any of the results for domain dependence change as a function of the specific model invoked. **There is considerable confusion about this topic when reading this paper. It is important the model being used herein is stated unambiguously at the outset of this paper and the equation added in the methods section to avoid such confusion. Further adopt the correct terminology for that model and don't list any alternative terminology that might confuse the reader.**

AR: We agree, there is a lack of agreement in the literature regarding the terminology. For example, the distinction in PCA and EOF model you are suggesting, is only one option that can be also critizised (Jolliffe, 2002). Most often, we found the information that PCA and EOF have different roots but are interchangeable and that the related terms are used interchangeably (e.g. Hannachi et al., 2007; Wilks, 2006). It also appears in a paper that you are highlighting (Huth and Beranova, 2021). Thus, to our understanding it is not the main point here to decide between the models you are suggesting, but to clearly define the applied terms and to use them consistently throughout the manuscript. Here, we certainly agree and thank you for pointing out inconsistencies.

The PCA performed with function "prcomp" in R gives unit length eigenvectors (what you call EOFs). These are termed "loadings" in the documentation of the function in R. To our knowledge, this is also the way it is commonly used among PCA users in hydrology. These are the coefficients used to calculate the PCs. The PC scores have mean zero and the variance equals the associated eigenvalue (what you call PCs). Thus, prcomp applies what you call the EOF model.

As postprocessing step, we multiply the unit length eigenvectors with the square root of the corresponding eigenvalues (what you call PC loadings). Thus, the sum of the squared correlation loadings of a PC equals its eigenvalue. They are equivalent to the Pearson correlation of the PCs and the analysed variables, since we apply correlation matrix PCA. To emphasize this, we call them "correlation loadings".

We are aware that the term "correlation loadings" is is not commonly used. However, given the lack of agreement regarding the terminology, we prefer that the reader might stumble upon "correlation loadings" and is forced to check our definition rather than using the term "loadings" where the reader might think of either unit length eigenvectors or the scaled version.

We use correlation loadings here for several reasons. They provide the Pearson correlation range from –1 to 1 which is for most users easy to grasp. The common range also enables to directly compare the contrasts of the spatial PC patters from different PCs or PCAs (L167). Finally, it is prerequisite for the calculation of the stochastic DD reference patterns (L208).

In our set up here, the correlation loadings define the spatial PC patterns, the eigenvectors the unscaled spatial PC patterns and the PCs the temporal PC patterns. Note however, that the focus in our work is on the spatial PC patterns. The analysed time series are all z-scaled white noise series, resulting in the temporal PC patterns being white noise as well. Furthermore, the mean correlation loadings from the stochastic method, cannot be used to calculate scores, like with the classical loadings (L211).

For clarification, we will add the equations for the PCA in section 2.1 and for the correlation loadings in section 2.1.2 and rephrase the second paragraph of section 2.1.2 to:

"In correlation matrix based PCA, normalizing the unit length loadings a_j of a PC j by multiplying it with the square root of its eigenvalue λ_j is equivalent to the Pearson correlation between the scores of that PC and the analysed variables. Thus, the loadings are normalized to the commonly well-known Pearson correlation range from -1 to 1 which simplifies reading and interpretation of the PCA results. The sum of the squared correlation loadings of a PC equals its eigenvalue. Note that these normalized loadings are different from the "classical loadings", used in the linear combination to calculate the PC scores, which are not normalized to a common range. To prevent confusion, we use the term "correlation loadings" for the normalized loadings. In the following, the spatial PC patterns are described with correlation loadings c only.

$$c_j = a_j * \sqrt{\lambda_j}$$

For S-mode PCA, the normalization enables direct comparison of the contrasts of spatial patterns from different PCs or PCAs. Here, we define the contrast of a spatial PC pattern as the range between the minimum and maximum of the correlation loading values of that PC. Thus, the maximum contrast possible would be 2."

Hannachi, A., Jolliffe, I. T. and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, International Journal of Climatology, 27, 1119-1152, https://doi.org/10.1002/joc.1499, 2007.

Jolliffe, I. T.: Principal Component Analysis, 2nd ed., New York, Springer, 2002.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, 2nd ed., Elsevier, 2006.

3. The treatment of eigenvalue degeneracy is generally well addressed with one exception that potentially plagues nearly every applied eigenanalysis: eigenvalue degeneracy at the truncation point (k). If those PCs associated with closely spaced eigenvalues between k and k+1 have information that is intermixed, problems arise and data is intermixed with noise on the kth retained PC loading vector. Your paper presents 10 PCs, therefore, the spacing between the 10th and 11th eigenvalues should exceed the North et al. criterion. Does it? Let the reader know.

Further, this needs to be mentioned because it can cause the loss of a domain dependence pattern simply because the way eigenvalues are ordered in descending order makes them more likely to be closely spaced as the smallest eigenvalues head toward the tail (presumably noise) where the analyst would normally truncate the analysis to discard the k+1,...,nth eigenvalues, perhaps using some other criterion (e.g., based on percent variance extracted, eigenvalue magnitude).

Related to this, I wonder why eigenvalue degeneracy is not addressed earlier in the paper as it seems to affect domain dependence. If that is the case, then consider moving it earlier in the paper as those PC loadings arising from degenerate multiplets should not be expected to exhibit the domain dependent patterns but the multiplet may be dominated by the domain dependent patterns and those are intermixed into new patterns that don;t seem to be domain dependent patterns.

AR: Thank you. We will include the aspect of eigenvalue degeneracy at the truncation point. Therefore, we will expand the warning to split multiplets in L425 with an explicit statement about the truncation point:

"In particular, special care has to be taken that the truncation point of a PCA does not split a multiplet."

Here, we show the first ten PC patterns merely for illustration. We found it to be a good balance between showing the DD pattern sequences and some degree of detail, but not too much detail that it is still visually easy to grasp. Also, to our experience most S-mode PCA applications in hydrology use substantially less than ten PCs, our casual guess would be around four. To clarify this, we will add an explanation to the truncation point used in our study in L260:

"Note, that here and in the following we show the results for the first ten leading PCs. The decision was taken merely for the illustrative purpose. We found it to be a good balance between showing the DD pattern sequences and some degree of detail, but not too much detail that it is still visually easy to grasp. There was no other specific truncation criterion, e.g. based on eigenvalue magnitude or percent variance extracted, applied."

We did not check for degenerated multiplets formed by PC 10 + PC x, because we did not analyze the PC 10 patterns further and used them only as examples for illustration. We will state this explicitly in L473, including your hint on intermixing of signal and noise:

"Note also, that intermixing might be easier overlooked for the smaller eigenvalues that are more closely spaced. If the analysist selects PCs to separate noise from signal, this could possibly result in truncation within a multiplet and consequently intermixing of noise and signal in the last considered PCs. Here, we selected the first ten PC merely for the illustrative purpose (section 4.1). If the goal would be to further analyze PC 10, it would be necessary to check its patterns for intermixing - also with the subsequent PCs, in particular PC 11. Indications for intermixing in the PC 10 pattern can be seen in the stability plots of Figures 4a, 10c, S1a+c. In case of Figure 10c, PC 9 does not show sign of intermixing, thus, in this case the intermixing is probably with PC 11."

Thank you also for your hint that intermixing might mask (expected) DD patterns. We will add a short paragraph on that in L453:

"Note, however that degeneracy might cause domain dependent patterns that don´t seem to be DD patterns because they are intermixed into new patterns. For example, in Figure 13 the patterns of the multiplet pairs of simulations 1, 4 and 5 exhibit different patterns than those of simulations 2 and 3."

Regarding the order of our sections, we like to point out again that the motivation of our work is to provide an illustrative introduction for PCA users in hydrology. We come back to this because it relates a lot to what we are presenting in which order. We start with simple examples to introduce the general phenomenon. Then, we get more specific and complex. Step by step, we focus on different aspects of DD and link it to PCA features we assume to be of interest for the PCA practitioners in hydrology and regularly used.

In section 4.1, we use the Buell patterns to introduce the general phenomenon, the concept of stability of the PC patterns and the use of the scripts. In section 4.2, we continue with the domain shape aspects, including irregular distribution of the locations. In section 4.3, the ratio of domain size versus spatial correlation length and its effects on the explained variances and the contrasts of the DD patterns comes into play. Thus, in each subsection of section 4, we introduce new aspects, building on the earlier ones.

We assume that the effectively degenerated multiplets will be the most abstract and difficult to grasp part for most of our readers, and probably also the furthest away from what they are used to. That is why the degenerated multiplets come in their own section as the last of the phenomena we want to introduce.

4. Comparison of PC loading patterns is accomplished with correlations. S-mode PC loading (and that of EOFs) interpretation depends on the magnitude of the PC loadings plotted on a map (and in general, the magnitude of the PC loadings/EOFs is important in any mode). Therefore, correlations subtract each PC loading/EOF vector mean (pattern mean), so two patterns with different means can have their large correlations, yet their magnitude patterns will be much different and the grid boxes (I think what you refer to as cells) with the maximum PC loadings will be in different geographical (or topological) locations in your domains. If that is the case, the the correlation is suboptimal for such comparisons. Find a better metric that includes magnitude in terms of comparison. I suggest the congruence coefficient, though others exist that preserve the vector magnitudes.

AR: One of the major benefits of Pearson correlation is that it is well known and the results in terms of r or R^2 can easily be contextualized by the reader. We assume that most, if not all, readers of the manuscript will know it from their own work. This includes the well-known sensitivity to outliers (see your comment on L175). To our experience Pearson correlation is regularly used in hydrology for the comparison of spatial patterns from PCs and potential explanatory variables or other patterns of interest.

Note also that all PC patterns in our study here are (1) from correlation matrix based PCA, thus, all analyzed variables are z-scaled, and (2) correlation loadings, thus, scaled to the common range [-1, 1]. Furthermore, all PC patterns that are compared with Pearson correlation are based on data sets simulated with either (a) identical spatial correlation properties and identical domain (step 2 of the stochastic approach) or (b) identical spatial correlation properties (the correlation exercise with the all cells variant and the spatially homogeneous and heterogeneous subsampling variants in section 4.2.).

We assume that for our examples here, neither differences in magnitude of the patterns nor the effect of the pattern mean subtraction by the correlation analysis versus deviations from zero by the congruence coefficient (see your comments on L192 and 336) are much of an issue. Thus, for our analysis and the introductory purpose here, we think simple Pearson correlation is sufficient.

Therefore, we prefer to keep it simple and stay with Pearson correlation for the presented analysis and results and include a discussion on the limitations of using Pearson correlation and the benefits of the congruence coefficient in section 5.1.2..

5. It seems odd that after the paper establishes the details and importance of domain dependence, it has no results on how rotating those PCs affects such dependence. There is only a scant mention of the possibility of this near the end of the paper, mostly in the context of rotating degenerate multiplets. However, rotation can be applied to PC loadings associated with non-degenerate eigenvalues and it will affect domain dependence patterns. Please consider adding a section on rotation and show those patterns to comment about how domain dependence is addressed by post processing the PC loadings with a rotation.

AR: In our study here, there are no physical structures, hydrological signals, modes or processes integrated in the simulated data sets. Thus, there are no signals to detect. The idea of our work was to demonstrate to PCA users who are not aware of DD and the related discussion that even without any physical structures, processes or modes, suggestive patterns can emerge. The numerical experiments are designed for that. Adding new numerical experiments to evaluate the performance of unrotated versus rotated PCs in identifying (hydrological) signals is beyond our introductory scope here.

This is a basic difference to the extensive study of Compagnucci and Richman (2008) who simulated data sets with a series of typical atmospheric flow patterns (plasmodes) to test the performance of unrotated versus rotated PCs, both from S- and T-mode PCA, in recovering the plasmodes. We agree that it would be very interesting and valuable to perform a similar study with typical hydrological signals instead of the atmospheric plasmodes. However, we think that this is material for another standalone study.

Here, we will extend the section 5.2.2 on rotation and include your literature recommendations and some of your thoughts and hints. See also our reply to your specific comments to L341, 557, 558, 573, 574.

6. The manuscript discusses accounting for domain dependence prior to attempting **physical interpretation**. Both the abstract and the introductions discuss how ignorance about domain dependence can easily lead to the wrong interpretations of PCA results (e.g., "Ignorance about DD can easily lead to the wrong interpretations of PCA results. DD patterns are distinct, with strong gradients and contrasts, and therefore highly suggestive to indicate physically meaningful drivers or properties of the analyses system". I agree with this statement and, assuming it is valid, the reader will want to know abut the right interpretations of PCA results. The manuscript further states (correctly) that the analyses proceed from data that are formed into a correlation (or covariance) matrix, either explicitly and implicitly and that matrix (or the standardized data in the case of SVD) are decomposed into eigenvectors that should be capable of summarizing the correlations/covariances of the data (after ensuring they do not represent domain dependence patterns). Therefore, some additional discussion of how to interpret those eigenvector (in the case of the present manuscript, PC loadings and PC scores), after passing a domain dependence assessment, must be added. It seems the majority of patterns shown in the paper suffer from domain dependence or from the effects of eigenvalue degeneracy combined with domain dependence. Would that be the null hypothesis for other investigators?

The main recommendation to assess such a hypothesis of domain dependent patterns (according to the manuscript) seems to be to visually assess the similarity but it leaves the reader asking, "then what do I do?". Presently, there is a suggestion to visually assess the analyzed patterns and compare to the domain dependent patterns for a similarly shaped domain. Two issues with visual assessment are (a) the reliability of the same pattern under the eyes of different analysts may well have one analyst believing there is a strong resemblance, and the pattern should not be further interpreted, yet a second analyst may think it has some resemblance but not that much to reject it as domain dependent. Further, (b) the nature of a qualitative visual assessment means any one analyst can see some resemblance to domain dependent patterns in their visual assessment and then discount it based on personal bias. A more quantitative approach to avoid (a) and (b) would be a direct numerical comparison using a matching coefficient (e.g., congruence coefficient). In that case, a recommendation could be made, such as, if the congruence coefficient exceeds some value (e.g., > 0.8), the analysis is dominated by domain dependence and the unrotated PC loadings/EOFs should not be analyzed physically. The assessment of the physical interpretation gets even trickier at this point. If the PC loading pattern based on either visual assessment or congruence coefficient value is thought not to be sufficiently contaminated be domain dependence, it does not mean it is physically interpretable as a meaningful mode without further investigation. Recall what the PCA does. It summarizes the correlation/covariance structure into a set of k PC loadings and k PC scores. Do we know if any of those structures relate well to the correlation/covariance matrix from which they were drawn? Without such a step, physical interpretation would seem unwise (we're back to the castles in the clouds but now from the "heavy constraints"). Because the manuscript is motivated by finding physically important modes, a revised manuscript should address or provide some suggestions on how to confirm if a mode is physically realistic or related to the correlations/covariances (or not). There is some literature on this topic, ranging from never physically analyze any PC structures (in that case domain dependence is moot because domain don't affect the ability of PCA to extract most of the variance from a dense correlation/covariance matrix) to, in many cases, the PC structures can be analyzed after confirming similarity to the correlations/covariances . I suggested examining the Compagnucci and Richman (2008) and Huth and Beranova (2021) papers for starters. The latter asks the specific question about what is a "true mode" whereas the former addresses the question about if certain analysis modes can retrieve the modal patterns. Of course, there are other alternatives, such as using a technique not rooted in eigenvectors. However, if the paper offers a path to identifying domain dependence that undercuts physical interpretation, some remedy should be offered.

AR: Yes, the suggestion is to use DD reference patterns as null hypothesis (see L175, 520-527, 582-589 and Appendix A). In the revised manuscript, we will include a discussion on the limitations of the visual assessment and the use of Pearson correlation and the congruence coefficient for numerical comparison in section 5.1.2.. See also our reply to your major comment 4.

We agree that DD is just one of the aspects that should be checked prior physical interpretation. It is not enough to check whether the patterns are sufficiently free from DD. Or as you stated in the beginning of your major comment section, it is “a necessary, but not sufficient condition for physical interpretation" (see also your comment on L575). Thank you for the references and the different hints in your comments to the physical interpretation of the PC patterns in L23, 72, 81, 408, 428, 574, 575 and your major comment 1. We will include them, expanding the discussion on physical interpretation of the PC patterns in Lines 406-409 and move the whole discussion to a new paragraph after L73 in the introduction.

Regarding the identification of physically realistic modes, we will include references to the work of Compagnucci and Richman (2008) and Huth and Beranova (2021) in the extension of section 5.2.2. (see our reply to your major comment 5).

In hydrology, spatial PC patterns have been also used to describe the spatial variability of distinct hydrological signals, processes or physical properties (L78-80). Building on the idea with the plasmodes (see our reply to your major comment 5), we think it would be very interesting to conduct more numerical experiments with hydrological simulation models to test whether any of the implemented hydrological features of the model can be uncovered with the patterns of the PCs. The test data could be, for example, spatially distributed groundwater level series simulated with a groundwater model. Again, this could include the comparison of the performance of unrotated versus rotated and / or S- versus T-mode PCA. In the revised manuscript we like to include these ideas as an outlook to future research at the end of the conclusion (see also our reply to your comment on L573).

Specific comments

Numerous specific comments are listed in the annotated manuscript (attached).
Citation: https://doi.org/10.5194/hess-2024-172-RC1

L23:
What is the proper "interpretation" for PCA? At one end of the spectrum, some might claim PCA is simply a data compression technique with little or no possibility of physical interpretation. At the other end, some may claim that the individual PCs can be interpreted as physically meaningful entities. Some discussion of this might be in order prior to the discussion of domain dependence, particularly as Buell (1979), in the final sentence of his conclusions, states, "Otherwise, such interpretations may well be on a scientific level with the observations of children who see castles in the clouds."

AR: We will elaborate on that in a new paragraph after L73 in the introduction. See also our reply to major comment 6.

L36:
This may be the situation; however, there is a general lack of agreement on these models, that leads to massive confusion among users of these techniques.

The original paper where EOFs were named "EOFs", is generally attributed to Lorenz (1956). However, in that report, Lorenz refers to the displays as EOFs of space, and EOFs of time, to define what have now mutated somewhat into what are called "EOFs", and "Principal Components", respectively. Assuming a spatial analysis, those EOFs of space are unit length (sum of the squares of each EOF's coefficients = 1), whereas the EOFs of time are orthogonal vectors, each with a mean of zero and variance equal to the associated eigenvalue.

In contrast, the PCA model, generally attributed in idea to Pearson (1901) and more fully to Hoteling (1933). The PC loadings in that PC model weights (postmultiplies) the unit length eigenvectors (EOFs) by the square root of the corresponding eigenvalue to give "PC loadings". That seemingly minor change in the spatial patterns (keeping with the definition of space and time given for EOFs) results in the time series calculation and properties being different to close the PC model. Those time series in the PCA model are called "PC scores" and have mean 0 and variance 1. They are also orthogonal.

Flip the space and time definitions of these displays if the analysis is temporal.

Because the two models result in different space and time patterns, they cannot be compared directly and the precise equations used are necessary to attempt to reproduce the findings of others.

The situation becomes more complicated as users of these techniques tend to pull EOF/PCA code off of various statistical packages or Python code libraries, that often mislabel the results, never checking the specifics, thereby perpetuating the confusion.

For the current paper, one must know if the analyses are applied to EOFs (unit length eigenvectors) or PC loadings (unit length eigenvectors postmultiplied by the square root of the corresponding eigenvalues). Further, it would be helpful to know if any of the results for domain dependence change as a function of the specific model invoked. **There is considerable confusion about this topic when reading this paper. It is important the model being used herein is stated unambiguously at the outset of this paper and the equation added in the methods section to avoid such confusion. Further adopt the correct terminology for that model and don't list any alternative terminology that might confuse the reader.**

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Project Rep. 1, MIT Department of Meteorology, 49 pp.

Pearson K. On lines and planes of closest fit to systems of points in space. Philosophy Magazine. 1901;2(6):559-72.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441. http://dx.doi.org/10.1037/h0071325

AR: We addressed this comment in our reply to major comment 3.

L72:
Although I agree with this sentiment, showing that it is reasonable to expect any physical conclusions to be drawn from PCA should be discussed first. Assuming the basic covariance structure carries important information about the physical processes, do the PC loading patterns (or EOFs) relate well to the underlying covariance structures? The investigator must confirm this prior to expectation of the patterns being related to physical processes. Add a few sentences on that.

AR: We will do so in a new paragraph after L73. See also our reply to major comment 1 and 6.

L81:
True, but domain dependence is one of several factors that hampers physical interpretation of the system. Others include:

(1) Data are related either explicitly or implicitly (by expressing them in anomaly or standardized anomaly form) via covariances or correlations. Such matrices express only the linear relations in the data. Further, subtracting a mean in either covariances or correlations assumes stationarity of the mean and variance, often violated by processes such as climate change and hydrology affected by climate change.

(2) The eigenanalysis technique is limited to only linear relationships between the covariances or correlations and the EOFs/PC loadings.

(3) The first eigenvector extracts maximal variance and often pulls in different sources of variability onto the leading vector, mixing the different sources. This is shown clearly in Karl and Koscielny (1982) in their Appendix Fig. 14A (top panel), where their data vectors X1, X2 and X3 are highly related as group 1 and where their vectors X4, X5 and X6 are also highly related as group 2, but groups 1 and 2 are nearly orthogonal. However, as they show, the first PC lies directly between groups 1 and 2, thereby describing neither accurately and introducing distortion by merging them so that their projections (PC loadings) would all be positive of nearly equal magnitude on PC loading 1, suggesting to the investigator that there is only one grouping.

(4) All eigenvectors, beyond the first, are orthogonal to all the previous eigenvectors, and the hydrological processes are rarely if ever orthogonal.

AR: We will include these aspects in a new paragraph after L73. See also our reply to major comment 1 and 6.

L93:

This would include the degree of linear association and the scale of the spatial correlation or covariance field with respect the the domain size. One might hypothesize that processes that are either weakly linear or mostly nonlinear and the linear part is relatively small, that domain dependence might dominate. Similarly, if the size of the spatial covariance/correlation data function is nearly the same size as the spatial domain used, the domain dependence might be different than for small scale processes, of perhaps 1/3rd of the domain size.

AR: We demonstrated this and analyzed the effect of domain size and spatial correlation length in section 4.3. The analysis of non-linearity is beyond our introductory scope here.

L96:

I'm not sure other papers have examined this. If that is true, add it as a unique aspect of this research.

AR: We do not examine this here. However, the blurring effect of measurement errors can be simulated with the stochastic method, using relatively short time series with rather unstable spatial PC patterns. We will include a statement on this in L284.

L99:

Define "low ranked PCs". I'm assuming these are PCs associated with smaller eigenvalues. If so, state that. If not, define it explicitly. If it is for the eigenvectors associated with small eigenvalues, the work of North et al. (1982) [cited extensively in your manuscript] and Quadrelli et al. (1989) on degenerate multiplets claim intermixing of the variance structures when he eigenvalues between adjacent eigenvectors are close in magnitude. That may be what you are seeing here.

Quadrelli, Roberta, Christopher S. Bretherton, and John M. Wallace. "On Sampling Errors in Empirical Orthogonal Functions." Journal of Climate 18, no. 17 (September 1, 2005): 3704–10. http://dx.doi.org/10.1175/jcli3500.1.

AR: Yes, your assumption is right. We will define it there and include a few sentences to the intermixing and rotation together with the references you provided:

"This gets less clear for those of the PCs with smaller eigenvalues (low ranked PCs). They are more finely detailed and less robust against deviations from Buell´s settings. Furthermore, there might be intermixing of the variance structures when the eigenvalues from successive eigenvectors are of very similar size (North et al., 1982; Quadrelli et al., 2005). These PCs which are not well separated with the PCA are called effectively degenerated multiplets (North et al., 1982). For their separation, additional post-processing is required, e.g. rotation of eigenvectors (Richman, 1986; Jolliffe, 1989)."

L99:
becomes

AR: We will change according to your suggestion.

L101:
Yes but that could be the aforementioned intermixing of the variances. In fact, it is possible that even Buell patterns for large sample sizes may intermix if the eigenvalues between successive eigenvectors are very close in magnitude.

AR: We agree. This is addressed in our response to your first comment on L99 above.

L116:
Here you mean "PCs which are not separable without additional post-processing (e.g., rotation of the PC loadings to separate the sources of variability for all eignvectors or for those eigenvectors with closely spaced eigenvalues)." Richman (1986) and Jolliffe (1989). You cite both these papers presently but the logical conclusion is missing.

AR: Yes. We will include your specification on the post-processing / rotation in the newly added lines where we introduce the effectively degenerate multiplets (see our reply to your comment on L99) and reduce the sentence here to "... and c) effectively degenerate multiplets."

L135:
See my earlier comment. It is more serious than terminology. The original models of EOF versus PCA have specific terminology and give different results. Those model display names have mutated over time and the terminology has been intermixed. Unless one state the mathematics of the model invoked, the reader has little idea what model is invoked. That also means that any conclusions for the EOF model need to be verified for the PCA model as the EOFs vs. PC loadings have different magnitudes and properties and the PCs vs. PC scores have different magnitudes and properties.

AR: Please see our reply to major comment 2.

L138:
It appears this is the EOF model. If so, despite Jolliffe's terminology, the majority of the literature envoking EOFs will call the displays EOFs and PCs (for space and time displays, respectively). Wilks textbook has a reasonable section on the varied terminology of EOF/PCA, although even that excellent book it is not exhaustive in this regard. You might try to simplify the sentence here where eigenvalue, scores, eigenvectors are mentioned. Alternatively (and perhaps the superior solution) would be to show the compact equation(s) in this manuscript for the model being invoked and that would clarify any confusion by the readers.

Addendum... After reading further, it seems you may have used actual PC loadings and PC scores. If so, ignore my comments about not using PC scores but, in that case, drop the discussion of EOF beyond the beginning of the introduction, as it serves only to confuse the readers. Further, clearly state what model you are using in section 2.1 with the equation for that model.

AR: Please see our reply to major comment 2.

L140:
See earlier comment. Please refrain from using the term "PC scores" if you are using an EOF model. PC scores are defined only for the PC model that weights the unit length eigenvectors by the square root of the corresponding eigenvalues. For the present manuscript, if you are using EOFs, then you can call them PCs and drop the "i.e., the PC scores". However, if you are using the PCA model, then use the terminology PC loadings and PC scores but drop other terms that will serve to confuse.

Addendum...See comment above. There is a need to clearly define the model used, define the appropriate terminology of the model displays and use on those terms throughout.

AR: Please see our reply to major comment 2.

L143:
This sentence seems to be awkward or a fragment at a minimum. Please clarify.

AR: "The eigenvectors of all PCs define the orthogonal basis of the new ordination system into which the analysed data is projected (orthogonality constraint)".

L143:
"mapped to" what? This sentence seems incomplete.

AR: Please see our reply to your comment on L143 above.

L147:
Yes, see earlier comment. Often the process of maximal variance extraction runs counter to interpretation of the sources of variability in physical systems (unless there is a single physical mode that encompasses the full extent of the domain where PC 1 can explain it -- rarely the situation). I mentioned this earlier and pointed to the Appendix of Karl and Koscielny (that you cite in this manuscript).

AR: Please see our reply to your comment on L81.

L154:
This is vague as "PC series" is undefined.

In the **EOF model:**

(1) the EOFs are not uncorrelated because their mean is not zero. However, they are orthogonal (if fact, orthonormal) by either column or by row as V'V and VV' = I.

(2) The PCs are uncorrelated as their means are zero. Additionally, their variance is the eigenvalue. They are uncorrelated by column and therefore orthogonal by column.

==========

In the PCA model (with eigenvectors scaled by the square root of the eigenvalue):

(1) the PC loadings are not uncorrelated because their mean is not zero. However, they are orthogonal by column only (their diagonal is the eigenvalue), V'V = D.

(2) the PC scores have mean = 0 and variance = 1. They are uncorrelated by column. Because they're uncorrelated, they're orthogonal by column with the diagonal equal to the degrees-of-freedom (normally, that would be n-1 if the correlation/covariance matrix is not singular).

Given these differences, hopefully you can appreciate the importance of stating unequivocally the specific model invoked. Specifically, where you say: All PC series are linearly uncorrelated with each other" is incorrect for both the EOF and for the PCA model. for the EOFS or for the PC loadings, as neither has zero mean column vectors.

Therefore rephrase "PC series" in this sentence and be precise to specify uncorrelated by row or by column.

AR: The term "PC series" means here the temporal PC patterns. It evolved somehow informal among colleagues as a short form for the PC scores which in the S-mode PCA case are time series of the same length as the analysed time series. Thank you for pointing out that it causes confusion without this background. We will replace it the two times it appears in L154:

"All temporal PC patterns are linearly uncorrelated with each other, each temporal PC pattern is associated with a spatial pattern and all spatial PC patterns are orthogonal to each other."

Please see also our reply to your major comment 2 about the PCA terminology.

L155:
Two comments:

1. The discussion prior to this section seemed ambiguous as to which eigenmodel was being invoked. Please fix that.

2. Does this imply that other modes of PCA do not suffer from domain dependence? For example, Q-mode is a field x station data matrix, giving a station x station covariance matrix.

AR:

To 1.: Please see our reply to your major comment 2.

To 2.: We did not investigate that and it is beyond the scope of our work here. However, we would hypothesize that in case of a homogeneous correlation structure, it can be an issue there as well.

L163:
Earlier, it seems that you were using EOF, here is suggests you are using PCA. State clearly from the outset which model is being used and stick with that terminology. If it is the PC model, them PC loadings and PC scores.

AR: Please see our reply to your major comment 2.

L167:
Two comments:

1. The desire to have PC loadings within the same range would fit the idea of EOFs better as those are all unit length eigenvectors. It may explain why Buell used EOFs, rather than PCs, to describe domain dependence. That said, there is nothing preventing any arbitrary scaling of the eigenvectors, as long as point (2) is noted.

2. The loadings normalized to anything other than the square root of the corresponding eigenvalue will no longer close the PC model using the standard formulation. I think that is what you are attempting to say in the last sentence, but it could be clarified.

AR: Please see our reply to your major comment 2.

L175:
Two comments:

1. Pearson correlation is leveraged by outliers. If two maps are being compared with relatively few common points on both maps extreme in the same direction, but the remainder of the gridpoints not in agreement, the correlation may be large.and exceed some t-test at alpha = 0.05. Because of that, field significance should be examined for difference fields of the maps, the pairwise comparison of spatial patterns of the combinations of PCs, or use a resistant statistic. For the field significance, here is an excellent test:

On “Field Significance” and the False Discovery Rate: By D.S. Wilks, Journal of Applied Meteorology and Climatology, Volume 45, Issue 9, 2006, pages 1181–1189.

2. t-tests assume Gaussian distributions. Correlation distributions are not Gaussian, particularly in the tails (where most of the matches would occur) because its range is limited to -1 to +1. You could Fischer z-transform the correlations first to partly mitigate this or, better yet, apply a permutation test to the maps, as the permutation test is distribution free.

AR: We think for our purpose here, simple Pearson correlation and the t-test are sufficient. This combination was also used in the study by Huth and Beranova (2021) you recommended.

For a general comment on why we use Pearson correlation, please see our reply to your major comment 4.

L192:
Because the mean of the PC patterns (i.e., the mean of each vector of PC loadings) is not zero, and the interpretation of the PCs is a function of the magnitude of the PC loadings, the correlation of PC loading vectors by subtracting out the mean, is an inferior metric for PC loading comparison. Lorenzo-Seva and ten Berge (2006) make a good case for the congruence coefficient, which does not remove the mean prior to the comparison. This metric has been used in the geosciences literature for such comparison.

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

AR: Thank you for the literature recommendation. For our reply, please see our reply to major comment 4.

L193:
Explain why the ranks are substituted for the values.

AR: We are not sure what you mean here. Assuming that you are asking why the analysis is performed separately for each PC rank, this is why we want to identify the stability of the spatial patterns and of their ranking prior calculating the mean spatial PC patterns.

L194:
Given the earlier comment about congruence coefficient and the use of correlations here, using the word "congruence" here is not good form.

AR: We will replace it with "similarity".

L199:
Similar comment about the distributions of the variability. Do you test for symmetry to determine the variance in each tail is approximately similar (and hence a single standard deviation holds)? One way to test that is to calculate the skewness of each of these patterns to decide if the skewness magnitude exceeds 0.5 and therefore would not be sufficiently symmetric to assume symmetry.

AR: No, we did not test that. We assume that in our case here, in which the compared PC patterns stem from data sets simulated with an identical parametrisation, it is negligible, especially for the mean spatial PC patterns which are calculated based on a large number of simulated data sets (100) with rather long time series length (10 000).

L225:
Great. That was my earlier suggestion. Extend the commentary to the instability resulting from degeneracy arising from closely spaced eigenvalues.

AR: Thank you. Assuming you are referring to your comments in L99-116, please see our reply there. To improve the structure of the method section, we will move L222-228 "Confidence limits ... both eigenvalues." in to a new section 2.3 "North´s rule of thumb".

L242:
Does "prcomp" give unit length eigenvalues or PC loadings? I suggest checking this manually as some of the R codes I have investigated say the output is one thing but really supply something else or simply wrong.

AR: "prcomp" gives unit length eigenvectors. See also our reply to your major comment 2.

L250:
Define "cell". It appears in two previous Figure captions too. I assume cell means grid box but a formal definition is required.

AR: Yes, you are right. We will add the following definition in L182:

"The grid cells (cells) of the random field represent the locations of a data set."

L255:

Two comments:

(1) The x-axis on Figure 4 presents some challenges because the behavior normally asymptotes after about 2000 observations for the first few PCs. However, by showing 10,000, the reader cannot pick out the number of observations required to find sufficient stability in those leading few PCs.

(2) By the time an analyst extracts PCs beyond the first few, one wonders how many degenerate multiplets emerge. The problem with not knowing that is if the adjacent eigenvalues are separated sufficiently to exceed the North et al. criterion, then fewer observations than Figure 3 suggest are required to provide stability.

Conversely, if the leading eigenvalues in PCs 1-3 are separated by less than the North et al. criterion, the results shown in Figure 3 may be too optimistic compared to such cases with degenerate multiplets in the first few PCs.

Some comment on the eigenvalues and their separation in these examples is critical to interpret Figure 3 and the unravel the effects of degeneracy arising from multiplets from that of domain dependence. Ideally this would precede the discussion of domain dependence as that effect seems to affect domain dependent patterns.

AR: Figure 4 is meant as an overview figure. Section 4.1. is meant as starting point from which we introduce step by step new aspects in the following sections. Thus, at this point in the manuscript we do not want to get more detailed. We also do not want to open the topic with the degenerate multiplets at this early stage. We explain the logic of the order of the sections and why the degenerate multiplets come last in our reply to your major comment 3. Please see there.

L260:

Useful information but for applied research rarely are there 10,000 observations, much less 10,000 independent observations. Is there advice for the analyst using 100 to 1000 observations (more in the typical range)?

AR: We would advise the analyst to fit a spatial covariance function to the data and calculate stochastic DD reference patterns with that function for the spatial domain of interest. If the time series of the observed data are shorter than the time series length that is required to reach stable DD patterns, the PCA results would "have to be interpreted with the reservation that the DD might be stronger than the comparison with the reference suggests." (L514–518). We provide this information in section 5.1.2.. Here in section 4.1, we do not want to open this topic already. We explain the logic of the order of the sections in our reply to your major comment 3. Please see there and our reply to your comment on L255.

L265:

If I'm interpreting the variance percentages correctly, it seems the patterns for PC2 and PC3 may have very closely spaced eigenvalues, not separated sufficiently according to the North et al. criterion. There are a few others too that seems degenerate.

AR: You are right. But we do not want to open this topic in this section already. Please, see our reply to your comment on L255 and your major comment 3.

L273:
That may be part of it but the degeneracy of the adjacent eigenvalues may be a major contributor and therefore it needs to be factored into the explanation. Perhaps that would make it more intuitive?

AR: Please, see our reply to your comment on L255 and your major comment 3.

L283:
Good but without doing that, the conclusions drawn in this section are subject to unnecessary uncertainty. Perhaps remove the conclusions from the paragraphs above and save them until after the degeneracy analysis. Alternately, run the degeneracy analysis first to clarify the instability in these results.

AR: Please, see our reply to your comment on L255 and your major comment 3.

L285:
production of

AR: We will change that.

L302:
Are these "PC patterns" the PC loadings? Please define the terminology, then use that same terminology throughout the paper, including figures and tables.

AR: These are the stochastic DD reference patterns calculated as mean correlation loadings. Please see our reply to your major comment 2.

L318:
delete "and kind of intuitive"

AR: We will change that.

L319:
predictable, a priori,

AR: We will change that.

L319:
more

AR: We will change that.

L321:
deleted space in front of %

AR: In accordance with the HESS guidelines, we will keep the space in front of the %.

L323:
Two comments:

(1) You use heterogenous here (check for a typo, as I think you really want to use heterogeneous) and inhomogeneous in other locations below. If those are the same thing, stick with the exact same word throughout.

AR: Thanks for the typo hint. Yes, we meant that and will change the phrasing consistently to heterogeneous.

(2) Motivate **why** homogeneous versus heterogeneous are both used. The former would apply to those analyses using **gridded datasets** with a regularly spaced grid, whereas the latter might apply to those analysts using actual irregularly spaced **station data**.

AR: Thanks for the hint. We will add a sentence for motivation in L319:

"...visual recognition is limited. This holds in particular for spatially irregular distributed locations which is the common case in hydrology."

L328:
predictable.

AR: We will change that.

L331:
as

AR: We will change that.

L332:
whereas

AR: We will change that.

L334:
I follow the thought but these sentences read awkwardly. Please reword more succinctly.

AR: We will rephrase it to:

"The similarity of patterns formed by congruent selections of cells from the different variants is of particular interest. It addresses the question whether the spatial PC patterns calculated from two different domains result in different relations between the values at locations with coincident coordinates. This is visually only poorly assessible. Therefore, we correlated the patterns of the subsampling variants with the patterns formed by ..."

L336:
Important. As stated earlier, PC loading (and that of EOFs) interpretation depends on the magnitude of the PC loadings. Correlations subtract each vector mean (pattern mean), so two patterns with different means can have their large correlations, yet their magnitude patterns will be much different and the grid boxes (I think what you refer to as cells) with the maximum PC loadings will be in different geographical (or topological) locations in your domains. If that is the case, the the correlation is suboptimal for such comparisons. Find a better metric. I suggested the congruence coefficient, though others exist that preserve the vector magnitudes.

AR: Please see our reply to your major comment 4.

L341:
Three comments:

(1) Perhaps, but if those "lower ranked" PCs are associated with closely spaced eigenvalues, we have no idea if the lower correlations arise from less DD or more degenerate patterns.

(2) If the results arise from PC loadings associated with closely spaced eigenvalues, Richman (1986) and Jolliffe (1989) have shown that those can be rotated to remove the degeneracy, and then the DD measured. This is mentioned near the end of the paper but it should be mentioned here too.

(3) In fact all the PC loadings could be rotated and then the DDs assessed as rotation has been shown to remove DD (at least beyond how the domain shape affects the correlation structure itself). Most people who interpret the PC loadings first rotated their PCs, so the lack of inclusion of rotated PCs is a shortcoming of the present manuscript.

AR: To (1): We agree.

To (2): At this point in the manuscript, we prefer to stay with unrotated PCs and discuss rotation later in its own subsection as part of the section how to consider DD.

To (3): We understand that you like to highlight the possibility of rotation. However, our focus here is to provide an introduction to DD and its side effects for PCA users in hydrology. For this, we believe that it makes sense to focus on unrotated PCs. Because, to our knowledge, many of the EOF / PCA techniques, and their characteristics and issues, are much less common in the hydrological literature than in the atmospheric sciences literature. We argue that your claim "most people who interpret the PC loadings first rotated their PCs" holds maybe in the atmospheric sciences literature, but not in the hydrological literature. For example, all the hydrological case studies we listed in our introduction were performed with unrotated PCs. We ourselves don´t have noteworthy practical experience with performing and interpreting rotational PCA. In fact, we debated whether to mention PC rotation at all during manuscript preparation, because we felt that the manuscript developed further away from the background and practical experience of the audience we like to address. However, we are aware that the atmospheric sciences literature is rich in this regard and that there is a lot to discover. Therefore, we came up with the compromise to restrict the examples in our introduction to unrotated PCs, but mention rotation as one possibility to continue from there on (section 5.2.2). We will gladly expand the section on rotation with your literature recommendations and some of your thoughts and hints.

We addressed this also in our reply to your major comments 3 and 5.

L348:
Will this be a footnote?

AR: A footnote would be linked to Figure 10. We are not sure whether this would work well. So, we like to leave it as it is.

L364:
Replace correlations with some metric that incorporates magnitude (e.g., congruence coefficient).

AR: Please see our reply to your major comment 4.

L383:
delete "it is simple"

AR: We will change that.

L402:
This makes sense because you are using unrotated PCs. Unless the autocorrelation is constructed to coincide with one of those PCs, there is little hope of isolating noise on any one PC. The situation may be less problematic if the PCs are rotated. That is one solution to the problem highlighted in this work.

AR: We will address that in section 5.2.2.

L403:
be salient against

AR: We will change that.

L408:
Two comments:

(1) This question of retrieval of the correct features of interest has been addressed in at least two published articles: Compagnucci and Richman (2008) and Huth and Beranova (2021).

Compagnucci, R. H., and M. B. Richman, 2008: Can principal component analysis provide atmospheric circulation or teleconnection patterns? Int. J. Climatol., 28, 703–726, https://doi.org/10.1002/joc.1574.

Huth, R. & Beranová, R. (2021). How to recognize a true mode of atmospheric circulation variability. Earth and Space Science, 8, e2020EA001275. https://doi.org/10.1029/2020EA001275

(2) You can add "maximal variance" to this list. It is clear that the first PC (associated with the largest eigenvalue) often merges several unique sources of variability when the spatial scale of the variability is smaller than the domain scale. This was mentioned in a previous comment in the work of Karl and Koscielny (in their Appendix), a citation in your list.

AR: Thank you for your comment and the literature. We will include both. Please see our reply to your major comment 1 and 6 and L81.

L419:
This section might be more logical to move this section to closer to the beginning of the paper because the intermixing of the unrotated PC loading signals needs to be addressed/accounted for before one can assess domain dependence.

Personally, I think both are issues but if the paper might apportion the percent of distortion associated with each of degeneracy and domain dependence (in percentage of distortion for example), the utility of this work would be enhanced.

AR: Regarding the order of the sections, please see our reply to your comments on L255-283 and to your major comment 3.

Regarding the apportion of the percent of distortion associated with degeneracy and domain dependence, we agree that this would be interesting for future work. Here, it is beyond the scope of our manuscript.

L420:
If it is not well separated by the "consecutive eigenvalues" in PCA.

AR: We are not sure what you mean here. Do you want some rephrasing?

L428:
It is surprising you say this given in the last paragraph you said, "For example, for physical processes or modes of geosystems, the S-mode PC properties orthogonality of spatial patterns and linear uncorrelatedness of temporal patterns are heavy constraints (Buell, 1979; Jolliffe, 2002; von Storch and Zwiers, 2003; Hannachi et al., 2007; Monahan et al., 2009)."

That said, it is possible but requires that the PC pattern modes must be assessed for their veracity (validity) by determining if those modes are similar to the patterns embedded in the correlation/covariance matrix from which the PCs were drawn.

AR: We will move the paragraph you mention to a new paragraph after L73 in the introduction where we will include your other comments regarding the physical interpretation of the PCs. Your second comment here will be addressed in the extension of section 5.2.2..

See also our reply to your major comment 6.

L435:
In situations where (1) domain dependence exceeds the physical signal and (2) the leading modes have closely spaced eigenvalues, the situation becomes intermixed. It would be ideal to unmix those sources in this work. At the least, issue a caveat.

AR: We are not working with physical signals here. The simulated data sets are designed to produce DD patterns and multiplets for demonstration. Adding new numerical experiments to perform hydrological signal identification is beyond our introductory scope. In the revised manuscript, we will issue a caveat in the extended discussion on rotation and physical signal identification. Please see also our reply to your major comment 5 and 6.

L439:
small variations of what?

AR: "small variations in the analysed data"

L440:
delete "d" in "degenerate~~d~~ pair"

AR: We will keep "degenerated pair" instead of the suggested "degenerate pair".

L449:

Please note here and below my previous criticisms of using correlations to compare PC loading vectors. Often correlating two PC loading vectors results in high correlation for magnitude configurations that don't match well. This occurs because correlations measure only the gradients and ignore the magnitudes, yet the interpretation of the PC loading vectors depends heavily on the locations of the maximum magnitude loadings.

AR: Please see our reply to your major comment 4.

L456:

How many patterns represent signal in applied research? There is a point in the eigenvalue spectrum where the associated eigenvectors represent either noise or signal with less variance than noise variance, and such eigenvectors would never be analyzed. This is why analysts nearly always truncate their n eigenvectors at k<<n.

AR: We agree. We show the 10 PC patterns for the introductory purpose of our work here, because we aim to demonstrate the DD phenomenon and its side effects. Please see also our reply to your major comment 3.

L465:

Weigh this finding against the average length of data analyzed in applied geophysical research. What is the advice to that analyst with say 100 or 500 independent observations?

AR: We would advise the analyst to consider subsampling a less symmetric domain (L459–462). We will add this suggestion in section 5.2.1. and rewrite L543–548:

"Analysing a subsampled data set with enlarged minimal distance between the locations can be used to diminish the DD of the PCA results. Reducing the symmetry of the domain can remove effective multiplets. Both can help to carve out features other than DD. On the other hand, informative local details might be filtered out together with the excluded locations. If there is still DD, the new DD patterns of the subsampled data set might be harder to recognize visually because of the smaller number of locations per area. The selected minimal distance, respectively the selection of locations, is critical for the analysis. Depending on the choice, different features in the results might stick out, get diminished or even disappear."

L465:

If the eigenvalues are exceedingly close on two adjacent eigenvectors, no sample size is sufficient for those eigenvectors to resolve the degeneracy - something that should be mentioned (e.g., see Richman, 1986, his Table II, where 10,000 observations were insufficient to resolve the true patterns once intermixed by degeneracy).

AR: We will include a sentence on that aspect at the end of L465 and add a paragraph break after the new sentence:

"However, for very symmetric domains no sample size might be sufficient to resolve the degeneracy (see Richman, 1986 and PC 2+3 and 5–10 of the square domain in Figure 4a)."

L482:
This seems to be part of the conclusions. If so, fold it into a larger conclusions and suggestions section.

AR: It is the suggestion section. Based on a suggestion of referee 2, it will be renamed to "Approaches to consider DD".

L484:
See previous comment on Q-mode, where the PC loadings are mapped spatially.

AR: We did not investigate that and will restrict us here to S-mode. See also our reply to your comment on L155.

L490:
This is a strange section. Yes, T-mode PCA is possible and even used by some (e.g., see previous citation to Huth and Beranova, 2021) but the idea of applying T-mode is made and never examined in the paper to determine if there is domain dependence in T-mode or how that would manifest. If you say this, then provide evidence that domain dependence may or may not be an issue in T-mode (or other modes). Assuming this is not added to a revision, at best much of this section can be reduced and placed into the conclusions under a paragraph on future research ideas.

AR: We see your point. In the revised version we plan to shorten it substantially.

L496:
rarely applied

AR: We will change that.

L500:
To

AR: We will keep our phrasing.

L500:
deleted "is to"

AR: We will keep our phrasing.

L501:
Such a comparison

AR: We will change that.

L502:
deleted "to perform"

AR: We will change that.

L554:
; Huth and Beranova, 2021

AR: We will add this.

L556:

This is more complicated than stated here. In the EOF model, the eigenvectors (EOFs) are orthogonal by column and by row. Under the PC model, the PC loadings are only orthogonal by column (not by row). Once rotated, both EOFs and PC loadings are no longer orthogonal by column. However, in the EOF model, the PCs are uncorrelated. Under the PC model, the PC scores are uncorrelated. Under orthogonal rotation, the PC scores are uncorrelated (and hence orthogonal by column). Under oblique rotation, the PC scores are correlated by column (and hence not orthogonal by column). This is all interesting mathematically, but neither the atmosphere or the hydrologic system follow anything remotely close to orthogonality. Therefore, this is a **validity** issue. If the PC patterns are not valid to represent the physical processes, all the mathematical niceties are meaningless if the PC loadings are to be analyzed individually. PCA is simply the incorrect model to represent the physics on each vector. If you are willing to forego physical analysis of each PC loading vector, then the PCA is an efficient linear representation of the total space but, in that case, domain dependence is not important. Once an analyst wants to add physical interpretation of each PC loading vector, all those maximal variance, and orthogonality features become useless in most cases, but now accounting for patterns with domain dependence becomes important. Again this topics of extracting the known sources of variability and of true modes are addressed in Compagnucci and Richman (2008) and Huth and Beranova (2021), among others.

AR: Thank you for your elaboration. For our scope here, we like to keep the simple more general statement regarding the side effects of rotation and add Wilks (2006) as second source. But we will extend the discussion on rotation in section 5.2.2. and physical interpretation, including the literature you suggested. See also our reply to major comment 5 and 6.

L556:

are relaxed

AR: We will keep our phrasing.

L557:

Perhaps that "redistributed variance" is the variance of the true modes of variability? Huth and Beranova would support such an interpretation. If that is the case, it is a more important aspect of the physical system than the eigenvalues (or the percent variance associated with each eigenvalue). One way to assess this is (assuming the correlations capture the physically meaning variations in the data) to determine if the PC loadings from a solution (unrotated, rotated) represent the underlying correlation functions. Once that is assessed, and if a PC loading pattern is associated with a correlation/covariance patternr, the statistics associated with these patterns are what describes the physical system.

AR: We will include these aspects in the extension of section 5.2.2.. See also our reply to major comment 5 and 6.

L558:

If you examine the results in this paper, for unrotated PCA, selection of k PCs to avoid truncation of degenerate multiplets is also critical, so the criticism of truncating PCs holds in general (unrotated PCA, rotated PCA) for all cases where the n PCs are not retained. Normally, k<<n PC are retained in unrotated solutions too, so selecting k is still an issue for unrotated PCA. Your results suggest that unless k is selected at a location in the ordered eigenvalues at a

location where the eigenvalue spacing exceeds the North criterion, too little eigenvalue spacing confounds the assessment of domain dependence.

If one rotates their PC loadings, previous research suggests that all the domain dependence seems to disappear. If that is the case, you could rotate and check your analyses for domain dependence and report on its reduction under rotation in this manuscript. You could also check the amount of the correlation functions applied that emerge with both the unrotated and rotated PC loadings and report those values and on the differences found.

AR: We agree that defining the truncation point is always an issue, be it for unrotated or rotated PCA. We will include the aspect of eigenvalue degeneracy at the truncation point. Therefore, we will expand the warning to split multiplets in L425 with an explicit statement about the truncation point. Please see our reply to your major comment 3.

Regarding rotation, we will expand section 5.2.2.. Please see our reply to major comment 5.

L564:
If you claim this, then can you point to a table of results that partitions the variance into correlation structure variance versus domain dependence variance)?

AR: We cannot point to such partition table. We refer to what we showed here, that is sequences of pure DD patterns in which the leading PCs were associated with substantial amounts of variance.

L569:
Should you add the following corollary? *"Without knowledge about the effects of degenerate multiplets, DD can be misinterpreted"*

AR: We assume, you are referring to the intermixing that can mask the expected DD patterns. Thus, we would rephrase your suggestion and add it to the list:

"Without knowledge about the effects of degenerate multiplets, DD can be overlooked because the degeneracy can mask the expected DD patterns."

L573:
Regardless of the assessment of effective multiplets (including at the truncation point) and DD, not analyzing how well the PCs resemble the underlying covariance or correlation structure will often lead to the wrong hyrological interpretations.

AR: To specific comments on L573–575.

We agree and understand that this is an important point for you. We think so as well. We will add a new paragraph at the very end of the conclusion. It seems to us, that the terminology "underlying covariance or correlation structure" points to atmospheric mode identification. In hydrology, mode identification is not as common as in the atmospheric sciences. Therefore, we will phrase our statement differently.

"However, it has to be noted that passing the check for DD and accounting for effective multiplets in the selection of the PCs are necessary but not sufficient conditions to assure physical meaningfulness. When single PCs, or combination of PCs, are assigned to distinct hydrological features, it should be carefully checked whether the S-mode PCA constraints orthogonality of spatial patterns and linear uncorrelatedness of temporal patterns support such interpretation. Building on this study, a next research task could be a numerical experiment to evaluate which PCA variants (unrotated vs. rotated, S-Mode versus T-Mode) and which matching coefficients to compare the spatial PC patterns (Pearson correlation vs. congruence coefficient) work best for hydrological feature identification."

L574:

PCs are a fine method for data reduction or compact orthogonal description of data onto k PCs. Once the analyst jumps from such a well-accepted interpretation to analyzing or interpreting each individual PC, some assessment of how well each PC represents the data covariability must be performed. Even in cases with no degenerate multiplets and small DD, that does not guarantee (even hint at) an accurate portrayal of a physical process on an individual PC. Such a determination must be made after the analysis. This needs to be added to the conclusions to inform the reader that physically analyzing individual unrotated PCs is a suggested path for enlightenment about the physical system. Recall, the cautionary statment in the conclusion of Buell (1979): "Otherwise, such interpretations may well be on a scientific level with the observations of children who see castles in the clouds."

Sadly, Buell's comment holds for unrotated PCs in general, because of all those "heavy constraints", eigenvalue degeneracy and domain dependence. Again, I urge you to examine Compagnucci and Richman (2008) and Huth and Beranova (2021).

AR: Thank you for your elaboration and the references. We will include this. Please see our reply to your comments to L573 and your major comments 5 and 6.

L575:

Perhaps necessary but certain not sufficient to show physical meaningfulness.

AR: We will add this. Please see our reply to your comments to L573 and 574 and your major comments 5 and 6.

L598:

What does "clean structure" mean?

AR: We will rephrase it to "smooth pattern".