Hydrol. Earth Syst. Sci. Discuss., 9, C6063–C6068, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C6063/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Flood-initiating catchment conditions: a spatio-temporal analysis of large-scale soil moisture patterns in the Elbe river basin" *by* M. Nied et al.

Anonymous Referee #1

Received and published: 27 December 2012

Nied, M., Y Hundecha, and B. Merz, 2012, "Flood-initiating catchment conditions: a spatio-temporal analysis of large-scale soil moisture patterns in the Elbe River," Hydrology and Earth System Sciences.

General comments

The primary objective of this paper is to identify classifications of soil moisture patterns or conditions that contribute to flooding in the Elbe River. The objective of this paper is a good one and certainly warrants attention. If soil moisture conditions that present a high risk for flood occurrence could be identified, communities could make

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

suitable preparations for floods even before rainfall occurs. The authors aim to use a combination of principal component analysis (PCA) and cluster analysis to achieve their objectives.

I have three main concerns about that this paper that I believe should be addressed before final publication in HESS. 1. All the results are based on simulated soil moisture patterns. While the authors clearly state this assumption, they provide no support that the simulated patterns are realistic. They calibrate the discharge that is produced by the model to the observed discharge, but I don't think they have any observations of soil moisture against which to compare. How do we know that the simulated patterns have any resemblance to the real patterns? This is a critical limitation on the analysis because everything the authors conclude relies on the validity of those patterns.

2. The methods in this paper are not adequately described. Numerous important details missing, but more importantly, the overall rationale for the approach is not given. I am very familiar with PCA, but I could not determine exactly what the authors have done from their description. There are many ways that PCA can be performed with a given dataset (for example, see the details of the papers that the authors cite), and these various approaches ultimately lead to different principal components (PCs) and different interpretation of those PCs. In particular, the authors say they performed PCA on the "spatial linear Pearson correlation matrix." If I understand them correctly, this is an unusual way to do the analysis. Why was this matrix selected? What is the rationale behind this unusual choice? What are the implications for the PCs and eigenvectors (i.e. how should the reader interpret those)? The authors also choose to include multiple simulated soil moisture patterns in the PCA rather than performing different PCAs for the different simulated patterns. Why was that approach selected, and what are the implications? I am less familiar with cluster analysis and their description of this topic is much shorter. I could not understand what was done.

3. The implications of the results seem rather limited. The authors mainly focus on the direct results of the analysis such as the variance that is explained by the different PCs

HESSD

9, C6063–C6068, 2012

and so forth. The deeper implications of those results are not well explored. In the end, the main conclusion of the paper is that "these results underline the importance of catchment state for flood initiation and severity." This conclusion is relatively unsatisfying when compared to the worthy objectives of the paper. Can the authors provide more precise and impactful conclusions?

Specific comments Figure captions need to be more descriptive.

Section 2: Study area Page 10057, lines 15-17: Climate is described both as temperate and as transitioning from maritime to continental. Are these descriptions consistent or contradictory?

Section 3: Hydrological modeling Page 10060, lines 24-25: "accumulated soil transpiration" over what duration? Page 10060, line 25: Should "potential soil transpiration is reduced" be "actual soil transpiration. . ."? Also, wording is incorrect with ". . .in dependence of the number of days. . ." Page 10061, line 9: cite chapter of reference, not entire reference Page 10061: The model used is calibrated for discharge, not soil moisture patterns and as far as I can tell there no actual soil moisture values are used in the paper for validation of the modeled moisture patterns. The lack of soil moisture measurements to verify to any degree the modeled patterns means that the results presented in the manuscript remain speculative. The authors did note the possibility of verification of the patterns by remotely-sensed products. Page 10061, line 15: clarify what "behavioral model performances" means Page 10061, line 16: How were the nine sensitive parameters identified? Page 10062, line 11: Page 10060, lines 8-9 indicate that multiple soil layers are used in the model, so need to explain what constitutes a daily soil moisture value. Is it an arithmetic average? Or a layer-depth weighted average value? Page 10063, lines 13-15: This is a valuable sentence explaining the relationship between eigenvectors and PCs. That relationship could be further clarified by explicitly stating that PCs are time-series and the eigenvectors are spatial patterns. Page 10064, line 13: climPCk and PCk are time-series, why is the correlation being done with a median value? What is a "decomposed" PC? Is it the same as the decom-

HESSD

9, C6063–C6068, 2012

position on page 10065, line 12? If so, then explain the decomposition at this point and how the median is calculated. Page 10064, line 14: The formula for the Spearman's rank correlation coefficient is unintelligible. Page 10064, 3.3 Cluster analysis: Explain more thoroughly, but briefly, cluster analysis in general. Many readers will not be familiar with that technique. Page 10064, line 18: Why up to 15 PCs used? Page 10064, line 19: What is the weight that is being assigned to a PC? Page 10064, line 22-23: Why is a distance metric needed and what are you measuring the distance between? What is the hierarchical Ward algorithm and provide a reference for it. Page 10065, lines 11-17: How are the "leading PCs" determined? How many are there? Also, what are the results of the cluster analysis? It seems that each PC, a time-series, is divided into clusters consisting of days that are somehow similar to each other. In the second cluster analysis, what is a cluster centroid? An SMI value? One should not have to assume that the numbers of clusters in two analyses are the same but should be able to control that number. Page 10065, lines 18-26: Is cluster enumeration somehow ordered? Otherwise, cluster 1 in one analysis may be equivalent to cluster 8, for example, in another analysis. But in order to assign a single cluster number to a day, the enumeration should have some meaning or order to it. Move section 3.4 Flood event identification to the beginning of section 3 so the ordering agrees with section 4. Page 10066, lines 12-14: Provide a brief justification why the flood identification method used was chosen from the five methods listed.

Section 4: Results Page 10067, lines 20-22: Are the validation gauges a completely separate set from the calibration gauges? The validation gauges need to be identified on Figure 1, along with specifying what 'identification' gauges are. This is an example of a paragraph that is too short and needs to be expanded or combined with the subsequent paragraph. Page 10068, lines 12-19: Selection, justification and use of a consistent number of PCs throughout the manuscript will aid in reader comprehension. For example, currently forty-three PCs are significant according to this paragraph but twenty are shown in Figure 3, five are shown in Figure 4, and up to fifteen are used in the cluster analysis. Page 10068, line 20 – Page 10069, line 12: Ensure correct refer-

HESSD

9, C6063–C6068, 2012

Interactive Comment

ence to eigenvectors and PCs throughout this range of text. For example, the sentence beginning with “The second PC. . .” indicates that the PC shows a north-south partition; however, the PC in this analysis is a time-series and the eigenvector is the spatial as shown in Figure 4. Accurately referencing the different products of the PCA is required to avoid confusion about the manner in which the PCA is performed. Page 10070: Could you note the divergent results of the different cluster validation methods but pick one to use in the analysis? Page 10069-10072: I am unclear as to exactly how the cluster analysis is performed and what comprises a cluster. Is a cluster a range of PC anomaly values? I need a better description of cluster centroids and how they lead to SMI patterns. In general, the description of the cluster methodology needs a good deal of work to clarify the techniques, products and results. I am lost in varying numbers of clusters, probabilities and patterns. Without a better explanation of the technique and the analyses performed and their purpose, I am unable to assess the validity of the results.

Figure 1 Legend needs better formatting Identify validation gauges

Figure 3 The number of PCs shown on the figure seems arbitrary. Twenty is already less than the total number of significant PCs but the variance explained could be considered negligible after only a few PCs. Figure 4 This is a good figure for illuminating the form of eigenvectors (spatial patterns) and PCs (time-series). Center “Principal Components” title text over figures

Figure 5 A table containing this data would be easier to understand.

Figure 6 These figures are too complicated. The number of PCs shown is arbitrary and not consistent with other portions of the manuscript. The numbers of clusters shown seem arbitrary and random and there are too many different numbers. I have difficulty determining the message that the figures are attempting to convey.

Figure 7 I am unclear as to what a cluster centroid is and how it leads to a spatial moisture pattern.

HESSD

9, C6063–C6068, 2012

Interactive Comment

Figure 8 I believe this figure shows the distribution of the values in PC1 and PC2 among the different clusters (i.e., each box and whisker indicates the distribution of anomaly values from PC1 or PC2 within each numbered cluster). If so, the caption should be amended to make that more clear. Also, describe what the features of the boxes and whiskers represent (e.g., median or mean, interquartile range or 1 standard deviation, etc.)

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 10053, 2012.