Review of HESS Manuscript

'Catchment classification by runoff behaviour with self-organizing maps (SOM)'

by R. Ley, M. C. Casper, H. Hellebrand and R. Merz

Dear Authors, dear Editor,
I have reviewed the aforementioned work. My conclusions and comments are as follows:

1. Scope

The work is well within the scope of HESS.

2. Summary

The proposed paper deals with the classification of catchments a) using information of catchment response (expressed by 12 indices derived from event discharge coefficients and flow duration curves) and b) using hydroclimatic (mean annual rainfall, potential evapotranspiration) and physiogeographic (mean slope, field capacity and groundwater recharge) catchment properties.
The analysis and classification is in both cases carried out with self-organizing maps (SOM) followed by hierarchical clustering to subdivide the SOM into clusters (catchment classes). Altogether 5 clusters are identified and interpreted/discussed with respect to meaningfulness.
The results of the two classification approaches are then compared by counting the number of catchments assigned to the same clusters; the overlap is about 67%.
The method is presented at the example of 15 years of hourly data from 53 medium-sized, partly nested catchments (9-1469 km²) in Rhineland-Palatinate.

3. Overall ranking

The work is ranked **'Major revision'**. This is due to some aspects as explained below.

4. General evaluation

Scientific significance

The work contributes to the task of catchment classification, a topic of ongoing and intense research activity. The interesting and novel aspects of the work are that a) the classification is done via two separate approaches (one based on response behavior, one based on hydroclimatic and physiogeographic catchment characteristics) and b) the use of SOMs and hierarchical clustering. The first allows a direct comparison of the resulting classifications, which allows evaluating the main hypothesis of the work, namely that catchment functioning and hydroclimatic / physiogeographic catchment characteristics are related. The use of the latter for catchment classification is relatively new and shows promising potential.
Thus the work contains significant and novel aspects.

Scientific quality

The scientific approach and the carrying out of the work is mainly clearly explained, self-contained and valid. However, I want to raise some issues here that apply to catchment classification in general and also to proposed work:

- **What shall be classified and for what purpose?** The catchment as a physiogeograpic unit described by e.g. slope, soil, vegetation, land use or the combination of the catchment and its hydroclimatic conditions? Of course the two are linked as the catchment evolved under the

hydroclimatic conditions, but it makes a strong difference on the applicability of the classification: If we classify solely based on catchment properties, we may group catchments that show different response behavior as they may be exposed to different hydroclimate. If we classify based on the combination, the two aspects become inseparable. Which of the two approaches is appropriate depends on the intended use of the classification system. The authors apply the second approach in the paper, which makes sense as the intention is to compare it to catchment response behavior. On the other hand, it would be the wrong approach if the purpose of the classification is model parameter transfer (e.g. soil hydraulic parameters), as they are (should be) mainly related to the catchments physiogeographic characteristics.

  - **In summary:** The issue of what shall be classified and for what purpose should be clearly stated in the paper. All further decisions (choice of the metrics etc.) in the work should be discussed in the light of this purpose.

- **Catchment size matters:** In my eyes, any catchment classification should consider size as one of the first and most important criterion of separation/grouping. This is for the following reasons:
  - First, the influence of channel (and floodplain) flow on the response behaviour (i.e. the discharge) changes with catchment size. It is imaginable that two catchments of different size appear similar with respect to flow duration curves (FDC) even though this similarity is not caused by the same underlying catchment functioning (coined 'process equifinality' by the authors).
  - Given the same hydroclimate and physiogeographic properties, two catchments of different size may produce very dissimilar response behaviour if the typical areal extend of rainfall is much smaller than the size of the larger catchment: Then, mean areal rainfall does not sufficiently explain rainfall and the location and areal extend of rainfall within the catchment plays a role, but only in the larger catchment. In this case, the two catchments could be dissimilar with respect to FDC only because of the differing size. I recommend to discuss this matter in the text (maybe rainfall extend here is typically in the range of most of the catchment sizes and therefore it does not matter). Maybe it is worthwile to add a metric of rainfall heterogeneity to the classification process.
  - Also, the larger a catchment the less representative the mean catchment slope may be: With the same mean slope, a concave, convex or uniformly inclined catchment may produce differing responses. Maybe it is worthwile to include some metric of slope heterogeneity in the classification process.
  - **In summary:** It would be helpful to either proof that within the range of catchment sizes in this study, size effects are not critical or to include size as a classification criterion.
- Normalization of FDC's
  - As indicated by the authors on page 3061/14-25, the choice of the normalization of the FDC has a large impact on the classification result. This deserves and requires a more detailed discussion in the text. Again, the question 'What shall be classified and for what purpose' comes into play: The authors state for example that normalization by catchment area shows (reveals?) a strong influence of mean annual precipitation. The question 'Is this an important criterion of classification or not' can again only be answered if the purpose of classification is clear.
  - **In summary**: Please discuss and justify the choice of the normalization in more detail in the light of the intended use.

Minor issues are:

- 3056/18: The definition of events contains a fixed parameter of duration. This makes sense, but could be a problem if the investigated catchments cover a large size/travel time range (see above). Please discuss.

- 3058/1-15: Please explain the choice of the indices in more detail: how was the selection and weighting process carried out, what were the underlying goals?

Presentation quality

The work is structured in a logical and comprehensive manner and good to read. It cites relevant literature and gives a good overview on the state of the art. However, there are some points that deserve further consideration:

- 3053/25: Please add information about the number of raingauges used
- 3050/20: Current investigations on the application of SOMs to catchment classification are e.g. done by
  - Di Prinzio, M., Castellarin, A., and Toth, E.: Data-driven catchment classification: application to the PUB problem, Hydrol. Earth Syst. Sci. Discuss., 8, 391-427, doi:10.5194/hessd-8-391-2011, 2011
  - Toth, E. (2009): Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting. Hydrology and Earth System Sciences 13 (9), 1555-1566, 2009
- 3055/9: Definition of events: Can you show some result that supports the statement of agreement of the automated and manual event identification?
- 3063/7-13: The results are very interesting and at this point the reader expects a discussion/interpretation of WHY the classification is as it is. The authors discuss this in section 4.4, therefore a reference to this section would be helpful.
- 3064/3: Which catchments are borderline catchments and why?
- 3064/13: Please explain the choice of the indices in more detail: how was the selection and weighting process carried out, what were the underlying goals?
- 3064/15: How was long-term ET calculated?
- 3065/5: It is hard for the reader to become familiar with the catchments and the distribution of indices in the test area: Please add at least a map with the DEM and the mean annual precipitation.
- 3065/21: Please show some results to allow the reader a direct comparison with the behavioral classification. E.g. a figure like Fig 4c) and a dendrogram/cluster like Fig. 5.
- 3067/25-27: it was not clear to me how this was done → please explain
- 3068/10-13: This is exactly the point I wanted to make in the above discussion. Maybe you can discuss this at the example of two catchments in this study that are similar with respect to behavior, but not properties.

5. <u>Minor comments</u>

- Literature
  - 3051/28: van Dijk (2010) → reference missing
  - 3057/22: Please add the Kohonen reference
  - 3070/20: Gottschalk and Weingartner (1998) → not used
  - 3071/24: Merz and Blöschl (2004) → not used
  - 3071/24: Merz and Blöschl (2005) → not used
  - 3072/16: SCS (1972) → not used
  - 3072/23: Uhlenbrook et al (2004) → not used
- Text comprehension / spelling (leading number indicates page/line)
  - 3049/13: hydrologic → hydrologically

- 3058/9: The measure of topologic error is described on the next page. For better understanding, please relate to this.
- 3074/Table 1
  i) QSH: Is this really the 1.0 minus the 0.2 quantile? I would have expected 1.0 minus 0.8 for the high range
  ii) MWL: Why is the range 0.7 to 0.9 (not 1.0)? Is this due to outliers you want to avoid?
- 3080/Fig 3: Please add in the description that this is for the entire year (not winter or summer)
- 3080/Fig 4:
  i) explanation of 4b) and 4c) need to be switched
  ii) please add a legend description
- 3080/Fig 4 b) and c), 3083/Fig 7: I assume the order of the neurons is the same as in Fig 4 a) (1, 7, 13, …, 30), so neuron 1 is occupied with catchments 14, 36, 45, neuron 7 is not occupied). Is this correct? As it was not entirely transparent for me, you should maybe explain it in the text.
- 3082/Fig 6: limp → limb, potted → plotted
- 3067/19: I suggest: 'The comparison of clusters based on i) response behavior and ii) physical …'

Concluding remark

(Please see this as suggestion, not something that needs to be considered in the revised version).

So far, rainfall dynamics, i.e. the temporal distribution of rainfall during an event is not considered in the classification. Picking up the above discussion about 'What shall be classified and for what purpose', it may be worthwile to include some metric of rainfall temporal heterogeneity in the classification. Not considering this implies the assumption of homogeneous rainfall characteristics throughout all catchments, which may or may not be valid. At least the mean annual precipitation shows a clear gradient in the test area, and this might be the case with event-scale rainfall characteristics as well. If we want to compare catchment physiogeographic properties and catchment behaviour, then only cases of equal meterorological forcing should be compared for classification.

Yours sincerely,

Uwe Ehret