Authors' response to Referee1:

The paper addresses a relevant scientific question and is well presented. Within the paper, the results of different approaches for the automatic identification of dominant runoff processes are compared. My general comments refer to three subjects (suggestion for minor revisions):

We would like to thank the anonymous reviewer for the positive and constructive review. We agree in most points with the reviewer's comments and believe that the suggestions will help us to significantly improve our paper. In the following, we will respond to each point of the review and indicate how we will consider the reviewer's contribution in the revised manuscript.

General comments

1) In a first step similarity measures are used to compare the reference map, which had been generated by intensive field work (Scherrer & Naef 2003 – SN03-map), and the automatically derived maps. The used mapcurve approach was slightly changed as the SN03-maps were always used as reference maps although Hargrove et al. (2006) recommend to change compared and reference map and to use the higher MC scores.
In the discussion section you mention "reliability problems of mapcurves". Would the results change (and possibly show less reliability problems), if you would have followed the recommendations of Hargrove et al. (2006)?

Reply: We want to thank the anonymous referee for putting attention on this topic, which will be fully addressed in the revised manuscript. We choose the Mapcurves (MC) score to allow a comparison of maps with a different number of classes. This score depends on the coarseness of the compared maps, where the coarseness, in this context, "depends on the average size and number of the patches in each category" according to Hargrove et al. (2006). They postulate that the direction of comparison must be the one that gives the higher MC score, without demonstrating why this should be so. The risk in following these recommendation is to endorse the coarser maps, whereas, in our opinion, the refinement of a map instead of its coarseness should be rewarded. Given this context, we decided for consistency to keep the direction of comparison fixed.

Differently, the reliability problems we found refer to the sensitivity of the MC scores. As a result of our study we found out that the degree of overlap of the classes belonging to the automatically derived maps reflects the one on the left side of Fig. 1 of Hargrove et al. (2006). In this case, significant increases of the degree of overlap entail only small increases of the MC score. There is therefore a need for a Goodness-of-Fit score capable of (1) comparing maps with different number of classes and (2) detecting improvements even if the degree of spatial overlap between maps being compared is moderate. We will point out this need in a clearer way in the revised manuscript.

2) The usage of PREVAH via differing model parameterizations representing different DRP-maps is an interesting strategy. However, I am not sure if the synthetic runoff simulations really reveal what they are expected to. You argue, that strong model assumptions had to be made, e.g. the assumption of completely saturated catchments, and that thus a calibration against measured runoff would be meaningless. As the parameter sets were chosen within realistic bandwidths, you expect the model results to be meaningful. However, to my opinion a validation of the model results with the aid of measured runoff values (e.g. for events with high antecedent soil moisture contents) would still be important in order to ensure that the chosen model parametrizations work well. I would thus recommend adding a model validation procedure. If no measured runoff values are available, other validation strategies should be used (at least such simple approaches like envelope curves, comparison runoff coefficients..).

Reply: We would like to thank the reviewer for the comment on the modelling strategy used for this study. Here we think that there is an issue concerning terminology on the model parameterizations needing clarification. In our study, each model parameterization (referred to in the manuscript as “parameter set”) is a plausible a priori definition valid for the different runoff type (RT) and not directly for a different DRP-map. The fact that each DRP-map consists of different extents and distributions of the same RTs leads (only indirectly) to the statement on the first line of this general comment. What the PREVAH model results show are thus not properly “differences between different parametrizations”, but rather differences arising due to the use of different DRP-maps.

Certainly, due to compensation effects between parameters and RTs, the simulation results could look different if other parameter values for each RT would have used. However, this would not change what is inferable from the synthetic simulations, that is as follow: Since DRP-maps can be seen as a tool for regionalisation on ungauged catchments, one should care about the extent and distribution of the RTs, because they can have a significant effect on the simulation results.

At the moment, the PREVAH model results just show differences between different parametrizations, the proof is not yielded yet, that the model works right for the right reason.

The well-known expression “*the model works right for the right reason”* appears twice in our manuscript, ones in the introduction and ones in the discussion section. However, it was never referred to the model results of the adapted version of PREVAH used in this study. In fact, given the model configuration we used and the strong assumptions we undertook, we are (perfectly) aware of the fact that the model results are not ready yet to be compared against measured runoff values. On the contrary, we believe that a validation procedure would distract the reader from the real focus, which is, as already pointed out in the manuscript, exclusively on how the different DRP-maps influence the simulated runoff. Hence, we will try to explain in a better way in the revised manuscript what the synthetic runoff simulations are expected to yield, and for what they are not meant for. “Synthetic” means, that runoff simulations are here a benchmark we adopt to learn something about the DRP-maps, and not to learn something about model structure, parameter uncertainty or efficiency against any observed time series. These last points are the next goals we are pursuing.

3) Not surprisingly, among all three automatically derived maps, the map SF07 shows the smallest differences to the reference map SN03 as the approach of Schmocker-Fackel et al. (2007) shows – among the three automatic approaches - the strongest resemblances to the approach of Scherrer & Naef (2003). The identification of the DRP after Scherrer & Naef (2003) strongly depends on detailed field investigation of soil profiles. The approach of Schmocker-Fackel et al. (2007) also strongly relies on – naturally less detailed – soil information (soil map of Zurich 1:5.000 with information on soil-water regime as described in the method's section), which is scarcely available in the same quality outside of the canton Zurich. Missing information may be calculated by a method of Margreth et al. (2010, literature source not easily accessible). Thus applications of the approach of Schmocker-Fackel et al. (2007) outside the canton Zurich might show different results. I would recommend referring to this data restriction in the discussion section.

Reply: We agree with the reviewer here and will therefore address this topic in the revised manuscript.

Specific comments

p. 13259, line 21 f.: To my understanding, the topographic wetness index allows to identify areas prone to saturation overland flow (although a lot of publications show problems of the accuracy of this method). Areas with low topographic wetness index values must not necessarily be areas prone to Hortonian Overland Flow.

Reply: Agreed. We will adjust this paragraph accordingly.

p. 13264, line 2: Can you please explain why the reference maps do not take the rainfall characteristics into account although the Scherrer and Naef-method contains different decision trees depending on rainfall characteristics (Scherrer 2006)?

Reply: In Scherrer (2006), decision trees are defined with respect to rainfall characteristics, discerning between long-lasting events (I < ca. 20 mm/h) and intensive events with short duration (I > ca. 20 mm/h). The most relevant differences between these two decision trees concern areas where the soil shows gleying characteristics or where infiltration hindrances can be found. If these areas are often negligible in terms of runoff contribution during steady rain, they can become relevant at high precipitation intensities. However, these areas are not easy to recognise and their extension can vary from time to time, depending for instance on tracks of agricultural machines, cattle etc. (Scherrer, 2006; Hümann and Müller, 2013). Since the effort of producing a second DRP-map for a catchment with the Scherrer and Naef-method is rather high, in practice only the decision tree valid for long-lasting events is used. The statement we wrote on p. 13264, line 2 is therefore imprecise and will be corrected. However, the rainfall data we used for the synthetic runoff simulations never exceeded the threshold of 20 mm/h (cp. Figure 12 of the manuscript). The decision tree used is therefore consistent with the simulated events.

p. 13264, line 11: Why did you use 1.2 m? How sensitive are the results to the choice of this threshold value? For the approach of Gharari et al. (2011) a sensitivity study was carried out for the choice of HAND thresholds. Would this also be useful for the threshold used with the Müller et al. (2009)-method?

Reply: In their paper, Müller et al. (2009) declare that “along the stream network on both sides of the stream a D_{SOF1} area is assigned, which represents the riparian zone.” Unfortunately, they do not quantify the extension this area should have.

In our study, the riparian zone in the MU09-maps is taken into account by assigning SOF1 (and thus RT1 according to table 2 of our manuscript) to areas with a HAND lower than 1.2 m. This value was estimated by visual inspection on Fig. 4a of Müller et al. (2009). The use of a HAND-threshold led to a better result than (compared with) the definition of a constant buffer around the river network.

A sensitivity study for the choice of this HAND-threshold could furnish more insights into its effect on the results. However, it would not be easy to define against which RT should the riparian zone on MU09-maps be compared, given that this zone is mapped differently on the reference maps (mostly as RT2). On one hand, the optimisation of RT1 of MU09-maps against the RT2 of the reference maps would not be fully justifiable, given that SOF1 belongs exclusively to RT1 (Naef et al., 2000; and Table 2 of our manuscript). On the other hand, the choice of optimising the extension of RT1 of MU09-maps against RT1 of the reference maps would have resulted in meaningless results.

These considerations brought us to the conclusion, already stated in the manuscript, that a standardised definition of DRP is needed, given that the perception of some DRP-classes varies among different authors.

p. 13266, line 8: Is the "corresponding degree of similarity" the same as the "expected similarity E"? In this case the usage of the same wording would support understanding.

Reply: The fuzziness of category is incorporated in the Fuzzy Kappa method proposed by Hagen-Zanker (2009) by means of a similarity matrix, in which a degree of similarity is specified for each pair of classes. The degree of similarity corresponds therefore to a number between 0 (totally distinct classes, e.g. RT1 and RT5) and 1 (completely identical classes).

Contrarily, the expected similarity E (called expected agreement in Hagen-Zanker, 2009) must be seen as a weighting factor, which takes into account the spatial autocorrelation in both compared maps and avoids the fuzzy kappa to assume negative values. For a detailed description of the expected agreement, we refer to Hagen-Zanker (2009).

In the revised manuscript we will use the same terminology as Hagen-Zanker (2009) to support understanding. We will therefore rename E as "expected agreement".

p. 13267, line 21: I would omit the word "fully" and just name it a "distributed model". Comparing the size of the catchments and the PREVAH-Gridsize of 500 m as well as regarding the way of implementation of the Runoff Type-information, I would not regard the model to be fully distributed.

Reply: We agree with the reviewer here. Contrarily to the model version described in Viviroli et al. (2009), the model used for this study is a gridded version of PREVAH. This spatial discretisation allows, on one hand, allows the consideration of the variability of the meteorological input. On the other hand, it enables the separate calculation of both runoff concentration and routing, which is fundamental for the application of the model on ungauged catchments. We will therefore omit the word "fully" in the revised manuscript.

Figure 12: Which model parametrization (4.1?) led to the simulated runoff of SN03, SF07 and MU09?

Reply: Yes, the model parameterisation is the 4.1 of table A2. We will state it more explicitly in the revised manuscript.

p. 13274, last sentence of the discussion: This suggestion is without doubt appropriate. But is it really suggested by your findings (with regard to the general remarks 2)?

Reply: Accordingly to our reply to the 2nd general comment, we will rephrase the last sentence of the discussion.

Table A3: Different writing of theta leads to confusion. Please explain the subscripted numbers (presumably they are referring to the runoff types).

Reply: Agreed. We will add an explanation of the subscripted numbers and adapt the writing of theta in the caption.

Technical corrections

Fig. 1 and 2: if possible, using the same scale would be helpful.

Reply: Agreed.

Additional technical correction: Citation of Dobmann, J.: Hochwasserabschätzung in kleinen Einzugsgebieten der Schweiz, Interpretations- und Praxishilfe, Südwestdeutscher Verlag für Hochschulschriften, Saarbrücken, 2015.

The correct year is 2010. https://www.svh-verlag.de/catalog/details/store/pt/book/978-3-8381-1420-0/hochwasserabschaetzung-in-kleinen-einzugsgebieten-derschweiz?search=dobmann

Reply: We really appreciate the attention the referee put in reviewing our manuscript.

References

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, Hydrol. Earth Syst. Sci., 15, 3275-3291, doi: 10.5194/hess-15-3275-2011, 2011.

Hagen-Zanker, A.: An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation, Int. J. Geogr. Inf. Sci., 23, 61-73, doi: 10.1080/13658810802570317, 2009.

Hargrove, W. W., Hoffman, F. M., and Hessburg, P. F.: Mapcurves: a quantitative method for comparing categorical maps, Journal of Geographical Systems, 8, 187-208, doi: 10.1007/s10109-006-0025-x, 2006.

Hümann, M., and Müller, C.: Improving the GIS-DRP Approach by Means of DelineatingRunoff Characteristics with New Discharge Relevant Parameters, ISPRS International Journal of Geo-Information, 2, 27, doi, 2013.

Müller, C., Hellebrand, H., Seeger, M., and Schobel, S.: Identification and regionalization of dominant runoff processes – a GIS-based and a statistical approach, Hydrol. Earth Syst. Sci., 13, 779-792, doi: 10.5194/hess-13-779-2009, 2009.

Naef, F., Scherrer, S., Thoma, C., Weiler, W., and Fackel, P.: Die Beurteilung von Einzugsgebieten und ihren Teilflächen nach der Abflussbereitschaft unter Berücksichtigung der landwirtschaftlichen Nutzung - aufgezeigt an drei Einzugsgebieten in Rheinland-Pfalz. Untersuchung im Auftrag des Landesamts für Wasserwirtschaft, Rheinland Pfalz. Report 003, 2000.

Scherrer, S.: Bestimmungsschlüssel zur Identifikation von hochwasserrelevanten Flächen, Report 18/2006, Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht Rheinland-Pfalz LUWG, Mainz, 2006.