Note to the Editor:

The reference list of both the revised manuscript and the marked-up version is not updated, with the exception of five new references (Bahremand, 2016; Beran, 1990; Fatichi et al., 2016; Mosley, 1981; Scherrer, 2006) and the modified one (Dobmann, 2010).

This is because the typesetting process took place on the pdf files.

For the updated reference list we therefore refer to the typeset manuscript (“hess-2015-507-discussions-typeset_manuscript-version3.pdf”).

Thank you for your understanding,

Manuel Antonetti, on behalf on the coauthors.

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 helped us to significantly improve our paper. In the following, we will respond to each point of the review and indicate how we considered 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 was better 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 which 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 association of the maps we compared 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, on one hand, comparing maps with different number of classes and, on the other hand, detecting improvements even if the degree of spatial overlap between maps being compared is moderate. We pointed 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 types (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 different parameter values for each RT would have been used. However, this would not change what is inferable from the synthetic simulations, that is as follows. 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 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. In the revised manuscript, we tried to explain in a better way what the synthetic runoff simulations are expected to yield, and for what they are not meant for. In a recent paper, Fatichi et al. (2016) also deal with this topic and clearly state that synthetic simulations (called "virtual experiments" in their paper) "are different from studies aiming at comparing models among themselves or validating model results". "Synthetic" means therefore, 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 addressed 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 adjusted 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 was therefore imprecise and was 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 used the same terminology as Hagen-Zanker (2009) and renamed 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, 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 therefore omitted 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 stated it more explicitly.

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 rephrased 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 added an explanation of the subscripted numbers and adapted the writing of theta both in the caption and in Eq. 3.

Technical corrections

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

Reply: With the revised manuscript, we furnished a resized version of figures 1 and 3, where the same scale as for figures 2 and 4 is used. According to layout requirements, the old versions of figures 1 and 3 could be preferred to the new ones by the typesetter.

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

References

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, Journal of Hydrology, 537, 45-60, doi: 10.1016/j.jhydrol.2016.03.026, 2016.

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 Delineating Runoff Characteristics with New Discharge Relevant Parameters, ISPRS International Journal of Geo-Information, 2, 27, doi: 10.3390/ijgi2010027, 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.

Authors' Response to Shervan Gharari:

The papers tries to evaluate different existing models for dominant runoff process (DRP) and compare it with an expert driven map which have been already existed. This DRPs then are used for comparison in a synthetic case study. I believe the study is important and relevant and deserved to be published.

Reply: We would like to thank Shervan Gharari for his valuable comments. His only amendments concern the part of the manuscript related to the synthetic runoff simulations, concerning Figures 11, 12 and related methods and tables. We are pleased that the referee had no objections on the first part of the study (map comparison with similarity measures). In the following, we will respond to each point of the review and indicate how we considered the reviewer's contribution in the revised manuscript.

Let me start from the part which I don't really understand. The synthetic case study itself. To my point of view the design of synthetic case study in many cases, which this case is one of them, is not that easy. The synthetic case study means that you have a known solution for which you seek to achieve via a model or strategy and it is a isolated test which cannot say much about the real system behavior. The question is that in this specific case how the authors are making sure that their synthetic data is truly reproducible by the three different models (this is the fairness of the comparison) and not biased to one of the existing models. I guess this is almost impossible to show, because we don't know what the true mapping is.

Reply: We do not fully agree with the definition of "synthetic case study" given by the referee. In literature, several studies have made use of similar approaches (sometimes called as "virtual experiments", "numerical experiments", "synthetic experiments", "virtual laboratories", etc.) without having a known solution to achieve (e.g. Weiler and McDonnell, 2004; Lanni et al., 2012; Frei et al., 2012, and the recent work of Fatichi et al., 2016). In our study, the focus is not on calibrating a model to a measured discharge. Contrarily, we want to test the influence of different hypotheses concerning the extension and distribution of DRPs on the model output using runoff simulations as a benchmark.

Of course we do not know what the true mapping is, but we can assert with a high degree of acceptance, that DRP-maps derived by the decision schemes of Scherrer and Naef (2003) correspond to the product of highest quality(closest to reality) a modeller can get from an experimentalist (Seibert and McDonnell, 2002). How a modeller should best deal with such maps will be object of further studies.

Another point, the perception of DRP are also time scale dependent, meaning that if you generate a longer synthetic case study bringing in the evaporation, transpiration and other processes which happen in a longer period of time you may end up having different conclusion.

Reply: The focus of our study is on floods and we stated it more clearly in the revised manuscript. We agree that another time scale would have led to different results, but we believe that the conclusions of the manuscript could be applied also to other temporal scales.

In other words, using expert knowledge should play a crucial role in each phase of the whole modelling process at each temporal scale.

At the event scale, more than by evaporation and transpiration, a crucial role is played by soil moisture and, to a lesser extent, by interception. As shown for instance in Scherrer et al. (2007), the antecedent wetness can strongly affect the DRP on a site. A biased mapping of DRPs can therefore lead to further uncertainty in case of unsaturated conditions. This issue will be investigated in a further study.

For me what is more interesting than the synthetic case study is to see how our available expert knowledge is transferable to different DRPs established by different automated mapping models and how it affects the final outcome, in this case discharge, given the real data. I elaborate my point; imagine that using GH11 the authors can make the model which contains three DRPs and as they did they find a way to introduce the expert knowledge to the model by setting some constraints, the same can be done for the other models using similar strategy but of course different set of constraints as they are more detailed (based on land use and geology). These models should provide ensembles which then can be used to compare the different models.

Reply: The referee will agree that expert knowledge can be used at every step of the modelling chain: landscape classification, model building and model constraining. In our study we tried to investigate how different degrees of expert knowledge applied on landscape classification affect the final outcome of hydrological simulations, while not variating the level of expert knowledge for the other two steps (model building and constraining). The point suggested by the referee will be investigated in further studies. The focus here is exclusively on DRP mapping approaches characterized by different levels of expert knowledge.

Anyway we cannot find which mapping model is more realistic based on expert knowledge but what we can do is to see how those mapping together with our expert knowledge in the model can reproduce the output which then can be used as a proxy of how close we are to the output and how sufficient is the mapping complexity to hold all of our knowledge about the processes. This way the different models can be compared. Having said that I do not agree with statement as such: “We therefore recommend not only using expert knowledge for model building and constraining but also trying to obtain spatially distributed landscape classifications that are as realistic as possible.” in my point of view we can never have the confidence to say which one is more realistic that the other one as long as our understanding is biased. However we can say which mapping is sufficient for the specific purpose. That is the entire point I wanted to hint at.

Reply: Here we do not agree with the referee’s statement: “*we cannot find which mapping model is more realistic based on expert knowledge”.* This is not what we did. What we did, in the first part of the paper, is following:

We started from the assumption that the manually derived map is the most realistic representation of the distribution of DRPs in a catchment. This is irrefutable, since manually derived map are based on extensive field work and sprinkling experiments. Several studies used the same approach for validating simplified mapping approaches (Hümann, 2012; Schmocker-Fackel et al., 2007; Müller et al., 2009). Then, we derived similarity measures, class comparisons, and deviation maps to find out which automatic mapping approach is the most similar to the reference, manually derived, map. Hence, we did not base our comparison on expert knowledge.

Expert knowledge is a concept which does not belong exclusively to the modellers. On the contrary, it is at the base of the dialog between experimentalists and modellers (Seibert and McDonnell, 2002). Traditionally, expert knowledge is used for model building (one should think, for instance, to the "perceptual model" described in Beven, 2012). Recently, expert knowledge was used in the phase of model constraining, and the referee must be acknowledged for that (e.g. Gharari et al., 2014). But the dialog must go on, and more interfaces must be created between dry and wet hydrologists. One possibility concerns the landscape classification. We do believe, that the use of expert knowledge in the phase of reading and trying to understand our catchment would finally lead to an improvement of our simulations. To investigate the influence of the landscape classification on rainfall-runoff simulations, we designed a small synthetic case study. We demonstrated that the DRP-maps have a strong influence on the results, meaning that maybe it is worth to increase the realism also towards the landscape classification.

As we already stated in the discussion, landscape classification and model complexity are strongly linked together. We agree that with models tailored on every single DRP-map the result could look differently. However, this was not the scope of this study, which was focussed exclusively on the mapping of DRPs.

"Once a model structure and its parameters have been identified for each landscape in a gauged catchment, they can be transferred to an ungauged catchment where the landscapes have similar hydrological behaviour." not that accurate statement in my point of view. Still some caution is needed.

Reply: With this sentence we wanted to furnish a vision on how the DRP-maps could be used for regionalisation purposes. The sentence reflects the definition of regionalization given by different authors (e.g. Beran, 1990; Mosley, 1981; Viviroli et al., 2009). However, we acknowledge that the form of the sentence can lead to misunderstanding, and we therefore reformulated it in the revised manuscript.

"As the results of this study suggest, the use of expert knowledge should not be limited only to the phases of landscape classification or model building and constraining, but should play a crucial role in each phase of the whole modeling process." please clarify this. What do the authors mean by this sentence? Can you make such a conclusion from your study?

Reply: Since both the extent and distribution of the RTs can have a significant effect on the results of hydrological simulations, one should strive for the most realistic landscape classification as possible. This is what one can conclude from our study. In support of this, we also identified different controversial points, where automatic mapping approaches usually fail in the assessment of the runoff intensity. We rephrased the last sentences of the discussion.

"However, the adaptation of these classifications to the characteristics of our study sites was beyond the scope of this study" but the authors already did? Right?

Reply: What we meant in this paragraph with "adaptation of classifications" is to intervene in the classification design, e.g. by adding or removing input data and modifying accordingly the classification rules (like for instance what Gao et al. (2014) did for the Upper Heihe, China). In this respect, we did not adapt the classifications and to date we are not working on this issue. On the contrary, we performed a sensitivity analysis for the two parameters controlling the GH11-maps (Fig. 1) and what we found is that the same values found by Gharari et al. (2014) were applicable on our study areas.

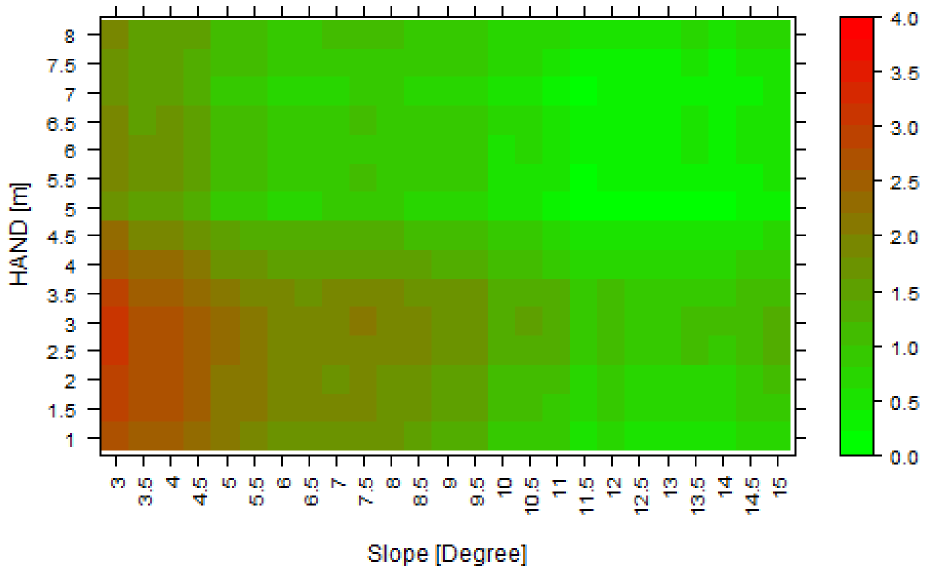


Figure 1. Sensitivity analysis of the threshold values for the HAND-based landscape classification on the Reppisch catchment as a whole. The level plot shows the percentage of deviation from the maximal MC-score (0.2023) obtained by comparing GH11-maps with the reference maps.

In the revised manuscript, we added Fig 1. in the Appendix and we stated more clearly what is meant with the term "adaptation".

I would like to see figure 12 with its distribution for GH11 simulations. Are most of the simulations closer to upper or lower limit?

Reply: Fig. 2 is an updated version of figure 12 of the manuscript and includes now the 10 lines corresponding to the simulations driven by GH11-maps. As already mentioned in the manuscript, due to the consistency assumption that no interflow is expected on wetlands and plateaus, too much water remained in the storage and runoff peaks were mostly underestimated. We updated this figure in the revised manuscript.

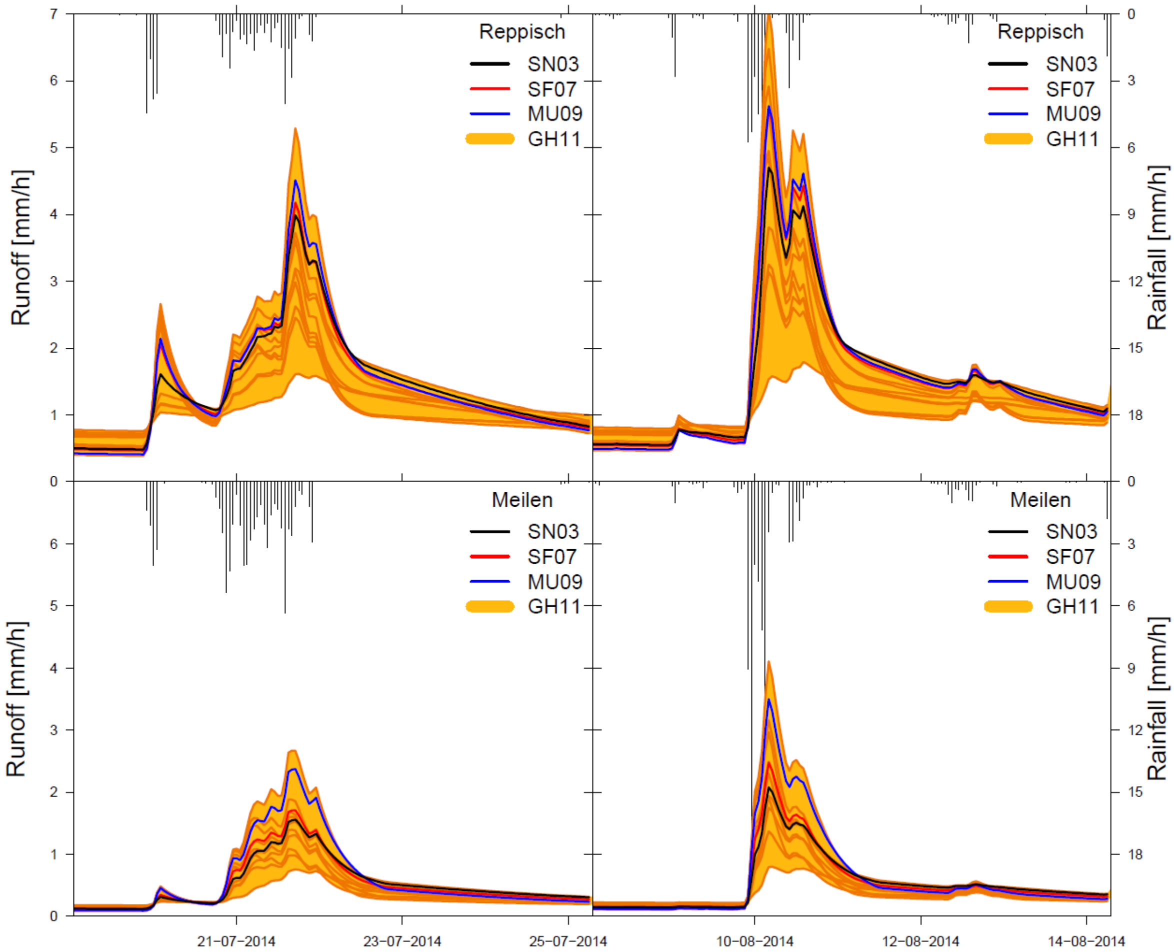


Figure 2. Simulated runoff during the two heaviest rainfall events of the simulation period, obtained from the different DRP-maps for the two study sites by varying the parameter values for each RT.

Overall the paper is interesting and I believe it can be published after major revision specially for the set up of the synthetic case study and its related conclusions.

With kind regards

Shervan Gharari

PS. I should thank Dr. M. Shafiei who helped in reviewing this paper.

Reply: We would like to thank again Shervan Gharari for his review of our manuscript. He offered a modeller insight into our study, coming up with valuable comments and criticisms on our analysis. His suggestion of adapting model structures and model constraints to the DRP-maps is challenging and will be certainly addressed in a future study. However, as a starting point of a multi-step analysis of system uncertainties with a view to regionalisation for ungauged catchments, within this study we investigated the effects of a precise uncertainty source, i.e. the DRP-maps, on the system output, i.e. the simulated runoff, while keeping fixed the other uncertainty sources. We believe this procedure to be reasonable, and do not want to confuse the reader by shifting the focus on any other aspects besides landscape classification.

References

Beran, M. A.: New Challenges for Regional Approach, in: Regionalization in Hydrology. Proceedings of an international symposium held at Ljubljana, April 1990. Beran M.A., Becker A., Bonacci O., Eds. IASH publication 191. Wallingford, UK, 1990.

Beven, K.: Down to Basics: Runoff Processes and the Modelling Process, in: Rainfall-Runoff Modelling, John Wiley & Sons, Ltd, 1-23, 2012.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, Journal of Hydrology, 537, 45-60, doi: 10.1016/j.jhydrol.2016.03.026, 2016.

Frei, S., Knorr, K. H., Peiffer, S., and Fleckenstein, J. H.: Surface micro-topography causes hot spots of biogeochemical activity in wetland systems: A virtual modeling experiment, Journal of Geophysical Research-Biogeosciences, 117, doi: 10.1029/2012jg002012, 2012.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrol. Earth Syst. Sci., 18, 1895-1915, doi: 10.5194/hess-18-1895-2014, 2014.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18, 4839-4859, doi: 10.5194/hess-18-4839-2014, 2014.

Hümann, M.: Abflussgeschehen unter Wald – Validierung und Weiterentwicklung eines GIS-basierten Tools zur Erstellung von Abflussprozesskarten auf forstlich genutzten Standorten. Dissertation - University of Trier, ISBN: 978-3-9813264-8-2, 2012.

Lanni, C., McDonnell, J., Hopp, L., and Rigon, R.: Simulated effect of soil depth and bedrock topography on near-surface hydrologic response and slope stability, Earth Surface Processes and Landforms, 38, 146-159, doi: 10.1002/esp.3267, 2013.

Mosley, M. P.: Delimitation of New Zealand hydrologic regions, Journal of Hydrology, 49, 173-192, doi: http://dx.doi.org/10.1016/0022-1694(81)90211-0, 1981.

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.

Scherrer, S., and Naef, F.: A decision scheme to indicate dominant hydrological flow processes on temperate grassland, Hydrological Processes, 17, 391-401, doi: 10.1002/hyp.1131, 2003.

Scherrer, S., Naef, F., Faeh, A. O., and Cordery, I.: Formation of runoff at the hillslope scale during intense precipitation, Hydrol. Earth Syst. Sci., 11, 907-922, doi: 10.5194/hess-11-907-2007, 2007.

Schmocker-Fackel, P., Naef, F., and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrol. Earth Syst. Sci., 11, 891-906, doi: 10.5194/hess-11-891-2007, 2007.

Seibert, J., and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resources Research, 38, 23-21-23-14, doi: 10.1029/2001WR000978, 2002.

Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J., and Weingartner, R.: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part I: Modelling framework and calibration results, Journal of Hydrology, 377, 191-207, doi: http://dx.doi.org/10.1016/j.jhydrol.2009.08.023, 2009.

Weiler, M., and McDonnell, J.: Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology, Journal of Hydrology, 285, 3-18, doi: 10.1016/S0022-1694(03)00271-3, 2004.

Mapping dominant runoff processes: an evaluation of different approaches using similarity measures and synthetic runoff simulations

M. Antonetti[1,2], R. Buss[1], S. Scherrer[3], M. Margreth[4] and M. Zappa[1]

[1]{Swiss Federal Research Institute WSL, Birmensdorf, Switzerland}

[2]{Department of Geography, University of Zurich, Zurich, Switzerland}

[3]{Scherrer AG, Basel, Switzerland}

[4]{SoilCom GmbH, Zurich, Switzerland}

Correspondence to: M. Antonetti (manuel.antonetti@wsl.ch)

Abstract

The identification of landscapes with similar hydrological behaviour is useful for runoff and flood predictions in small ungauged catchments. An established method for landscape classification is based on the concept of dominant runoff process (DRP). The various DRP mapping approaches differ with respect to the time and data required for mapping. Manual approaches based on expert knowledge are reliable but time-consuming, whereas automatic GIS-based approaches are easier to implement but rely on simplifications which restrict their application range. To what extent these simplifications are applicable in other catchments is unclear. More information is also needed on how the different complexity of automatic DRP mapping approaches affects hydrological simulations.

In this paper, three automatic approaches were used to map two catchments on the Swiss Plateau. The resulting maps were compared to reference maps obtained with manual mapping. Measures of agreement and association, a class comparison and a deviation map were derived. The automatically derived DRP-maps were used in synthetic runoff simulations with an adapted version of the hydrological model PREVAH, and simulation results compared with those from simulations using the reference maps.

The DRP-maps derived with the automatic approach with highest complexity and data requirement were the most similar to the reference maps, while those derived with simplified approaches without original soil information differed significantly in terms of both extent and distribution of the DRPs. The runoff simulations derived from the simpler DRP-maps were more uncertain due to inaccuracies in the input data and their coarse resolution, but problems were also linked with the use of topography as a proxy for the storage capacity of soils.

The perception of the intensity of the DRP classes also seems to vary among the different authors, and a standardised definition of DRPs is still lacking. ~~We therefore recommend not only using expert knowledge for model building and constraining but also trying to obtain spatially distributed landscape classifications that are as realistic as possible.~~Furthermore, we argue not to use expert knowledge for only model building and constraining, but also in the phase of landscape classification.

1 Introduction

Conceptual rainfall-runoff models perform well on gauged basins but appear to be limited in reproducing the hydrological behaviour of ungauged catchments (Hrachowitz et al., 2013). Expert knowledge about the different runoff processes that can occur on a catchment can improve the hydrological simulations for such ungauged basins. For example, it can be used to design process-tailored model structures aiming to be right for the right reason (Klemeš, 1986). Furthermore, it can help to reduce the need for calibration by constraining the parameter values or modelled output to guarantee consistency with the reality (Franks et al., 1998; Seibert and McDonnell, 2002; Gharari et al., 2014; Hrachowitz et al., 2014). Hydrological classifications based on landscapes with similar hydrological behaviour ~~are~~ can be useful regionalisation tools for predictions in ungauged basins. In this case, o~~O~~nce a model structure and its parameters have been identified for each landscape in a gauged catchment, they ~~can be~~are transferred to an ungauged catchment where the landscapes have similar hydrological behaviour (e.g. Beran, 1990; Mosley, 1981; Viviroli et al., 2009).

In recent decades, several methods have been developed to quantify the spatial extent and to identify the distribution of areas where a specific runoff process occurs. The topographic wetness index (Beven and Kirkby, 1979), as an example of index-based methods, allows areas prone to ~~infiltration excess (Hortonian) overland flow (HOF) to be distinguished from areas prone to~~ saturation overland flow (SOF) to be identified using only topographical information. Similarly, Woods et al. (1997) developed a topographic index for areas where subsurface flow (SSF) occurs. Another well-established methodology involves the explicit definition of hydrological response units (HRUs), which can be identified according to geological, ecological, pedological and/or topographical criteria (e.g. Ross et al., 1979; Flügel, 1995). For example, (Markart, 2011) developed a method for assessing surface runoff coefficients and surface roughness in case of extreme precipitation events. Similarly, Dobmann (2009) introduced a way to map runoff disposition, defined as “the tendency of water to become displaced downstream due to gravity in such a way as to cause damage” (Kienholz, 1998).

Although these methods represent an important basis for the determination of runoff peaks and return periods of flood events, they cannot reproduce the full range of runoff responses that can be observed on a site. To improve the HRU approach, several hydrological classifications have been developed based on the concept of ~~Dominant~~ dominant ~~Runoff~~ runoff ~~Process~~ process (DRP), i.e. the runoff generation mechanism that contributes most to runoff (Blöschl, 2001).

DRP classifications may be manual or automatic (Table 1). Manual approaches are based on extensive field investigations, and the interpretation and the upscaling of the results on expert knowledge (e.g. Scherrer and Naef, 2003). In contrast, automatic methods generally rely on GIS and on algorithms based on simplifications of expert knowledge (e.g. Peschke et al., 1999).

Automatic approaches differ in which data they require. Some rely on topographical information only (e.g. Gharari et al., 2011), while others use all the available information for an area (e.g. Schmocker-Fackel et al., 2007). The data requirement is closely linked to the time it takes to map the DRPs, ranging from a few hours with simple data input to months if the data are derived from extensive field investigations (e.g. Tezlaff et al., 2007).

The output classes of the classifications also differ. All methods distinguish at least between infiltration excess (Hortonian) overland flow (HOF) and SOF, and between SSF and deep percolation (DP) (e.g. Gharari et al., 2011; Gao et al., 2014). Several approaches also provide information on the intensity of the SOF and SSF processes, where the numbers from 1 to 3 represent the delay in their reaction to rainfall, with 1 representing an almost immediate reaction, 2 a slightly delayed one and 3 a strong delayed one (e.g. Scherrer and Naef, 2003; Schmocker-Fackel et al., 2007; Müller et al., 2009; Hümann and Müller, 2013). Boorman et al. (1995), however, classified expected hydrological behaviour according to 29 classes in the Hydrology Of Soil Types classification of Great Britain.

Several algorithms have been developed exclusively for specific catchments, and are therefore not suitable for regionalisation purposes. For instance, Tilch et al.'s (2002) classification is based on the genesis of the hillslope and its covering material. Similarly, Waldenmeyer (2003) determined DRPs from a forestry site map, and Gao et al. (2014) linked the presence of forest to the hillslope exposition in the barely inhabited Upper Heihe catchment in China. These simplifications limit the applicability of the methods to other catchments.

All these methods aim to map the spatial distribution of DRPs in a realistic way, but only few have investigated the transferability of the algorithms to other catchments. Furthermore, it remains unclear how the different time and data requirements of the mapping approaches affect hydrological simulations. The objective of this paper is therefore to (i) test the suitability of different automatic DRP-mapping approaches for mapping ungauged catchments, and (ii) quantify the uncertainty of hydrological simulations due to different spatial representations of DRPs.

DRP-maps were produced for two catchments on the Swiss Plateau using the automatic approaches of Schmocker-Fackel et al. (2007), Müller et al. (2009) and Gharari et al. (2011). These were then compared with reference maps produced using manual mapping according to Scherrer and Naef (2003). To assess how similar the automatically derived DRP-maps are to the reference maps, a measurement of agreement, Fuzzy Kappa (Hagen-Zanker, 2009), a measurement of association, Mapcurves (Hargrove et al., 2006), and a class comparison were carried out. Furthermore, the effects of the differences between the DRP-maps on synthetic runoff simulations were investigated with an adapted version of the well-established PREVAH model (Viviroli et al., 2009b).

2 Study Sites

Our analyses are performed on two small catchments on the Swiss Plateau. The Dorfbach Meilen is a creek which drains a 4.6 km^2 catchment and flows into Lake Zurich (Fig. 1). The elevation of the catchment ranges from 409 to 850 m a.s.l.. It is mainly covered by grassland (49.4%) and forest (39%) and, to a lesser extent, arable land (3.6%) and settlements (8%). The basin is characterised by Upper Freshwater Molasse with conglomerate in the shallow subsurface (Hantke et al., 1967). A large part of the catchment is covered by brown earth soils with normal permeability and storage capability. Soils with less permeable soils and wetlands are less widespread but play an important role in runoff generation.

The Reppisch catchment up to Birmensdorf is situated in the southwest of Canton Zurich, Switzerland (Fig. 2). It has an area of 22 km^2, of which 48 % is covered by forest, 42 % by grassland, and 7 % by settlements. The elevation of the catchment ranges from 467 to 894 m a.s.l.. The geological substructure of the catchment forms the Upper Freshwater Molasse, composed of sandstone and marl, and is covered in most cases by glacial sediments (Hantke et al., 1967; Pavoni et al., 1992; Bolliger et al., 1999). Gravel deposits can be found along the Reppisch river, while a number of smaller alluvial fans were accumulated by its many tributaries. Brown earth soils with normal permeability and storage capability cover most of the catchment, while soils with low permeability are less widespread.

3 Data and Methods

3.1 DRP-mapping approaches

Manually derived DRP-maps based on the decision scheme of Scherrer and Naef (2003), referred to here as SN03-maps, are available as shape-files for both study sites and were used as reference maps (Fig. 3a and 4a). These DRP-maps are developed in different steps as follows: 1) Information about the land-use, vegetation, soil, geology, hydrogeology and topography of the catchment are collected. 2) Based on these data, the DRPs are initially estimated using expert knowledge, and locations where estimations are not straightforward are identified. 3) On these sites, soil profiles are investigated and the DRP at the plot-sites identified according to the decision schemes for long-lasting events,~~-~~ i.e. with precipitation intensity less than ca. 20 mm/h, of Scherrer ~~and Naef~~ (2006~~3~~). 4) After the analysis of the field investigations, the DRPs can be determined for the hillslopes and finally for the whole catchment. 5) The DRPs are reclassified into five different runoff types (RTs) with respect to the runoff intensity (Table 2).

Schmocker-Fackel et al. (2006) developed a strategy to simplify the decision schemes of Scherrer and Naef (2003) and determine the DRPs automatically within a GIS environment. Basically, the method relies on a soil map with high resolution (1:5000) of Canton Zurich and information about the soil water regime, soil depth, and the soil's physical and chemical properties. Where information on soil is lacking, an expert-based soil prediction model was used to derive DRPs from information about forest communities, the slope and shape of hillslopes, the surface water network and the geology (Margreth et al., 2010). This step is relatively time-consuming, since the soil prediction model has to be adapted to each catchment according to the information available. Therefore, several days of fieldwork are necessary. The DRP-maps derived with this approach for this study are available as shape-files, referred to hereafter as SF07-maps (Fig. 3b and 4b).

Müller et al. (2009) proposed a further simplification of the Schmocker-Fackel et al.'s (2007) approach based on GIS, and valid for prolonged rainfall events. The method combines information on the permeability of the geological substratum, land-use and slope, but excludes soil information. It results in the same DRP classes as those proposed by Scherrer and Naef (2003), and involves: first, using a DTM analysis to identify classes of slopes; then, classifying the geological substrata of the catchments as either permeable or impermeable; and finally, combining the pre-processed digital data to obtain the DRP (Table 3). Hümann and Müller (2013) extended the approach proposed by Müller et al. (2009) to forested areas and to different~~. They modified the classification criteria according to the~~ event types. ~~and thus made the DRP-maps dependent on the event characteristics. The reference maps do not, however, take the rainfall characteristics into account, which is why Müller et al.'s (2009) approach was used for this study.~~ Since the reference maps refer to long-lasting events, the Müller et al.'s (2009) approach was used in this study.

DRP-maps based on Müller et al. (2009), referred to here as MU09 (Fig. 3c and 4c), were derived for the two study sites with a spatial resolution of 25 m based on following assumptions: (i) Riparian zones, i.e. the spots around the river network, were classified as SOF1. The extension of these areas were defined by taking into consideration the cells with a Height Above the Nearest Drainage (HAND), i.e. the height of a DTM-cell less the elevation of the river network where the cell drains (Rennó et al., 2008), that is lower than 1.2 m. (ii) Settlement areas were not considered in the current study as the resolution of the land-use map used (100 m) was not high enough to obtain a realistic representation of their spatial distribution.

As a further simplification, topography-based classifications were developed with the assumption that the topography can be seen as a proxy for the geology, soil, land-use, climate and, consequently, DRPs (Savenije, 2010). In addition to traditional topographical descriptors (e.g. elevation, slope and exposition), these methods are based on the HAND value, which represents, in turn, a rearrangement of the “elevation-above-stream” proposed by Seibert and McGlynn (2006). HAND-based classifications have been used to define classes of soil water environments, where a single runoff generation mechanism dominates (Nobre et al., 2011; Gao et al., 2014). Gharari et al. (2011) found that the combination between HAND and slope provided the most suitable descriptors for a topography-based classification of DRPs. The mapping approach distinguishes between three landscape classes. Areas below a certain HAND threshold value are called “wetland” (subject to SOF). The remaining regions are further divided into two classes: “hillslope”, subject to SSF, and “plateau”, subject to DP, depending on whether the slope is above or below a certain threshold value. Since these threshold values are not unconditionally transferable to other catchments, a sensitivity analysis was carried out on both study sites. Different combinations of threshold values were tested, and the resulting maps were compared with SN03 at a spatial resolution of 25 m. We selected the maps with the best Mapcurve-score (cf. 3.2) for this study, and refer to them as GH11 (Fig. 3d and 4d). The threshold values obtained are in agreement with those of Gharari et al. (2011) in a central European catchment (Fig. A1~~not shown~~).

3.2 Map comparison

To test the suitability of different approaches for automatically mapping the DRPs on ungauged catchments, a class comparison between automatically derived DRP-maps and the reference maps was carried out for the two study sites. The percentage of total catchment area assigned to each RT, and the percentage of discrepancy between the RTs in the automatic DRP-maps and those in the reference maps were calculated. To deal with the difference in number of classes between the GH11-maps and reference maps, an expedient step was introduced. Since none of the three classes of GH11-maps (wetland, hillslope and plateau) is necessarily comparable to a specific class of the reference maps, the 5 RTs of the SN03-maps were reclassified into 3 classes covering every possible combination (Table A1), resulting in 6 new reference maps. These were compared one by one with the GH11-maps. In addition, the discrepancies between the MU09-maps and the reference maps were highlighted in a deviation map to identify the spots where the difference in the RTs is greater than 2 and to help identify the possible causes of incorrect mapping.

To account for fuzziness in the definition of the RTs, a measure of agreement, fuzzy kappa (K_{fuzzy}), was used. The method was proposed by Hagen-Zanker (2009) to extend the well-established Cohen's Kappa (Cohen, 1960), and to take into account the fuzziness of categories, allowing some pairs of classes to be more similar than others, as well as the fuzziness of location, given that cells tend to be at least slightly spatially correlated. To take the fuzziness of categories into account, a similarity matrix was defined, where each pair of classes was assigned a number between 0 (totally distinct) and 1 (completely identical). The extent to which neighbouring cells influence the cell in question is defined by a distance decay function. An overall measure of similarity between two maps can be obtained by using the following equation:

$$K_{Fuzzy} = \frac{P-E}{1-E} \; [-] \qquad (1)$$

where P represents the mean agreement of the two compared maps, weighted by the expected ~~similarity~~ agreement E. K_{Fuzzy} ranges from 0 (fully distinct maps) to 1 (fully identical maps). For this study, the fuzzy kappa algorithm implemented in the software Map Comparison Kit 3 (Visser and de Nijs, 2006) was used. We assumed that contiguous RTs are similar to some extent and the corresponding degree of similarity was set to 0.25. An exponential decay function with a halving distance of one cell is adopted.

Given that the number of classes in the GH11-map is different from that in the reference maps, the goodness-of-fit (GOF) measure called Mapcurves (Hargrove et al., 2006) was used to quantify the degree of spatial concordance between the automatic DRP-maps and the reference maps. For each of the existing classes in two maps, a GOF-score [unit-less] was calculated according to the following equation:

$$GOF_X = \sum_{Y=1}^{n} \left(\frac{C}{A} \cdot \frac{C}{B} \right) \qquad (2)$$

where A is the total area [m^2] of a given class X on the map being compared, B is the total area [m^2] of a class Y on the reference map, C is the intersecting area [m^2] between X and Y when the maps are overlaid, and *n* is the total number of classes on the reference map. The sum of this product gives a GOF-value for a particular class. The overall Mapcurves (MC)-score is given by the area under the curve obtained by plotting the GOF-scores on the abscissa and the percentage of map classes with a GOF-score larger than a particular value on the ordinate. An MC-score of 1 represents a perfect fit, while an MC-score of 0 means that there is no spatial overlap between the classes of two maps. Both the shape of the Mapcurves and the MC-score differ when the compared map is used as a reference map. This is because the MC-score depends on the average size and number of the patches in each class of the maps being compared. Hargrove et al. (2006) argue that~~, in order to describe the degree of association between two maps,~~ the combination of compared map and reference map that has the highest ~~MC~~ MC-score must be chosen. However, by doing so, the coarser maps would be advantaged. ~~However~~Therefore, for this study, SN03-maps were always set as reference maps. A detailed description of the two similarity measures is reported in Hagen-Zanker (2009) and Hangrove et al. (2006), while applications in hydrology are described in Speich et al. (2015) and Jörg-Hess et al. (2015).

To identify those landscapes where automatic approaches perform better, the comparison measures were applied to the single sub-catchments, at a high spatial resolution, to take into account the added value of the finest maps. For this reason, the shapefiles were rasterised and the coarser maps were resampled to a grid resolution of 2 meters.

3.3 Synthetic runoff simulations

To assess how the differences between the automatic DRP-maps affect a hydrograph, synthetic runoff simulations were carried out. This approach was inspired by Weiler and McDonnell (2004), who suggested using numerical experiments to isolate hypotheses and investigate their influence on the model output. In a recent review paper, Fatichi et al. (2016) acknowledge these studies to be different from the ones aiming at comparing performances of different models or validating model results. The word “synthetic” implies therefore that the focus is exclusively on how the different DRP-maps influence the simulated runoff, and not on how well the model reproduces a measured discharge.

The model used for this study is an adapted version of the runoff generation module of the PREVAH model (Viviroli et al., 2009a). It is ~~fully~~ distributed (500 m grid resolution) to take into account the spatial variability of the input data, which consists of a combination of radar and traditionally measured rainfall data (Sideris et al., 2014). For each cell, the percentage of each RT is taken into account to avoid losing information because of the grid resolution.

The model does not take interception, evapotranspiration and soil moisture into consideration (Fig. 5). The rainfall directly recharges the upper zone (unsaturated) runoff storage (SUZ), where the storage times for the surface runoff (K0H) and subsurface runoff (K1H) regulate the generation of the runoff. The threshold for quick runoff formation (SGRLUZ) determines the separation between surface runoff (R0) and subsurface runoff (R1). A maximum percolation rate (CPERC) controls the percolation to the groundwater storage, which is divided into a quick-leaking storage (SLZ1) and two slow-leaking storages (SLZ2 and SLZ3; Schwarze et al., 1999). The storage capacity of SLZ1 is limited by a maximal storage charge (SLZ1MAX), while its contribution to the slow runoff (R2) is regulated by the storage time for quick baseflow (CG1H). SLZ2, which only receives the fraction of percolation not absorbed by SLZ1, is controlled by the storage time for slow baseflow (K2H). With this model configuration, it is possible to detect the effects of differences between the different maps in terms of both extent and distribution of RTs. The difference in extent of RTs gives more weight to one or other of the parameter sets. If the RT extent is the same, the location of the RTs on the catchment plays a role since the rainfall input can vary from cell to cell.

We assume that the properties of the different RTs can be represented by varying the parameter values of the model employed. For example, the tendency for RT1 and RT2 to generate overland flow was represented by assigning low values of SGRLUZ and CPERC. Furthermore, the K0H values assigned to RT1 and RT2 were set as low since the fast contributing areas were assumed to be close to the river network. On areas where either HOF or DP dominates, the subsurface flow was neglected and K1H was set to higher values (e.g. 1000 h). As the baseflow generation does not necessarily depend on the RTs, the parameters of the SLZ1, SLZ2 and SLZ3 were defined a priori as averaged values for both catchments and kept constant for the simulations. The values selected were based on the results of Viviroli et al. (2009a), who identified a range of suitable values for each parameter of PREVAH for flood estimation in ungauged mesoscale catchments in Switzerland.

To investigate the sensitivity of the model output with respect to the definition of parameter values based on the RTs, the parameters were defined in a stepwise process, resulting in 16 different parameter combinations (Table A2). First, the 5 RTs were assigned the same set of parameter values and no information about the RTs was thus included. In the second step, the value of each parameter controlling the SUZ was defined with respect to the RT one at the time, and the value of the other parameters was left unchanged. The same procedure was then repeated by defining the values based on the RTs of two, three and finally all the parameters at the same time. As in the class comparison (see section 3.2), an expedient step was introduced to take into account the fact that there were fewer classes of GH11-maps. Every possible combination of the five predefined values for each parameter was covered, provided that the parameters fulfilled the following condition:

$$\vartheta\theta_{WETLAND} \leq \vartheta\theta_{HILLSLOPE} \leq \vartheta\theta_{PLATEAU} \qquad \vartheta\theta = SGRLUZ, K0H, K1H, CPERC$$

(3)

This resulted in 10 different runs for each parameter combination (Table A3), with one exception: the storage time for the subsurface flow K1H. This was set at 1000 h for wetland (SOF) and plateau (DP), since no subsurface flow was expected there.

Synthetic simulations were carried out on the two study sites over the time period which ranges from 16/06/2014 to 15/08/2014. A modified version of the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), in which the observed runoff is replaced by the runoff simulated with the reference maps, was therefore used as objective function (Eq. 4).

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{SN03,i} - Q_{DRP,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{SN03,i} - \overline{Q_{SN03}}\right)^2} \; [-] \qquad DRP = SF07, MU09, GH11 \tag{4}$$

4 Results

According to the reference (SN03) maps, the two study sites differ slightly in their RT distributions (Fig. 6). In the Reppisch catchment, areas with a delayed runoff contribution (RT3) prevail (45% of the catchment area), while, in the Meilen catchment, areas with strongly delayed runoff contribution (RT4) cover 55.3% of the catchment. SF07-maps reproduce the RT distribution fairly, although they slightly overestimate the fast contributing areas (RT1), and underestimate the areas with strongly delayed contribution (RT4) in the Meilen catchment. The RT distribution of the MU09-maps deviate from the one of the reference maps. They considerably overestimate the delayed contributing areas (RT3) and, to a lesser extent, the fast ones (RT1), at the expense of the remaining RTs. The runoff contribution is consistently overestimated especially in the Meilen catchment, whereas in 64% of the whole catchment the RT is faster compared with the SN03-map (Fig. 7).

The distribution of landscape classes of GH11-maps in the Meilen catchment (Fig. 6b) agrees well with the reference map, if the landscape class “hillslope” is assumed to correspond to RT3, “wetland” to the union of RT1 and RT2, and “plateau” to both RT4 and RT5. However, this consideration no longer holds true in the Reppisch catchment, where the percentage of the total catchment mapped as “hillslope” (68%) markedly exceeds the one mapped as RT3 in the reference map (45%). Considering each possible reclassification into 3 classes of the 5 RTs of the SN03-maps (Table A1), the GH11-maps, on average, estimate the runoff contribution as lower than the SN03-maps estimate (Fig. 7).

Figure 8 shows a map of the Reppisch catchment highlighting areas where the discrepancy between the RTs in the MU09-map and the SN03-map is higher than 2 (Table 4). The RT assigned to area 1 is too fast as the glacial sediments were assumed to be always impermeable. Similarly, area 3 was mapped as a non-contributing area as the alluvium was assumed to be always permeable. However, previous investigations showed the local permeability of the glacial sediments was high and the one of the alluvium was low due to clayish sediments (Scherrer AG, 2006). Area 2 is located on a steep hillslope and is therefore mapped as contributing with a slight delay. In contrast, area 4 is on a flat plateau, so that its contribution to the runoff was assumed to be strongly delayed. However, field investigations found the soil was very thick indicating a high storage capacity in area 2. In contrast, the mixture of brown-earth, stagnosol and gleysol resulted in a low storage capacity in area 4 (Scherrer AG, 2006). In area 5, the river network derived with the DTM analysis differs considerably from the actual river path. The runoff contribution there was therefore overestimated by MU09. Similarly, the runoff contribution of area 6 was overestimated because the depiction of the lake was wrong due to the coarse resolution of the land-use map.

The measures of association and agreement obtained by comparing the automatically derived DRP-maps with the reference maps for the sub-catchments of the two study areas differ (Figure 9). The scores of the SF07-maps are higher than those obtained by the comparison of MU09-maps and GH11-maps with the reference maps. The highest scores in the Reppisch catchment were in sub-catchment 1 due to the presence of a lake, which is mapped as RT1 in every mapping approach. As the values of the MC-score obtained with MU09-maps and GH11-maps are nearly equal, these two mapping approaches seem to be interchangeable for both of the two study areas.

Comparing the MC-scores for each RT reveals which RTs can be clearly identified by the automatic mapping approaches (Fig. 10). The higher MC-scores for classifications with the same number of classes should ideally be located along the main diagonal of the output matrices, meaning that each RT of an automatically derived DRP-map is spatially best associated with its equivalent in the reference map. This is mainly the case for the SF07-maps, with the exception of the fast RT1 and RT2. These are identified as more similar to the next slower RTs of the reference maps. The MU09-maps's overestimation of the general runoff intensity of the whole catchment can be attributed to RT2 and RT4 in the Reppisch catchment and RT1 and RT3 in the Meilen catchment. These were spatially associated with the next slower RTs of the reference map. On both study sites, the landscape classes "wetland", "hillslope" and "plateau" of the GH11-maps fit best with RT2, RT3 and RT4 of the reference maps, respectively.

Since the extent and distribution of areas with the same RT differ, using automatically derived DRP-maps in runoff simulations affects the results of the simulations themselves (Fig. 11). Simulations driven by the SF07-maps showed the smallest deviation in comparison with simulations driven by the SN03-maps. The tendency of the MU09-maps to overestimate the runoff contribution (Fig. 7) led to higher peaks in the Meilen catchment since overland flow was activated on areas with delayed runoff contribution during the two heavy rainfall events on 21 July 2014 and 10 August 2014 (Fig. 12). This did not happen in the Reppisch catchment as the precipitation intensity in the catchment was lower. The GH11-maps were very sensitive to the storage time for subsurface flow K1H due to the consistency assumption, i.e. no interflow is expected on wetland and plateau areas, which are prone to SOF and DP, respectively. As a result, too much water remained in the storage and runoff peaks were mostly underestimated.

5 Discussion

One of the main purposes of this study was to test how well automatic approaches can map small catchments. The most complex automatic DRP-maps, i.e. the one derived according to Schmocker-Fackel et al. (2007), proved to be most similar to the reference maps derived manually with Scherrer and Naef (2003), according to both the class comparison and the similarities measures. This result is not surprising, considering that the method of Schmocker-Fackel et al. (2007) was developed for the canton of Zurich, where the two catchments of the present study are located. However, the method was successfully tested also outside the canton of Zurich (e.g. on the Swiss Prealps, Scherrer et al., 2013).

The DRP-maps derived with simplified mapping approaches, that included no soil information, differed significantly in terms of both extent and distribution of the DRPs from the reference maps.

These differences are clearly linked to the quality of the input data. Geological maps are often not fine enough to depict geological formations and possible variations in permeability within the same formation. Furthermore, if the resolutions of the DTM and the land-use map are too coarse, significant biases may result. However, using input data with high resolution would not necessarily improve the results, if the classification concept itself is too coarse and generic. Since topography does not seem to be a good proxy for the storage and infiltration capacity of the soils on the study sites, the approaches developed by Müller et al. (2009) and Gharari et al. (2011) often overestimated the runoff intensity on steep sites and underestimated it on flat sites. These approaches were developed on basins, located in Rhineland-Palatinate (Germany) and in the Grand Duchy of Luxembourg, with different soil properties and event characteristics than those investigated for this study. However, the adaptation of these classifications to the characteristics of our study sites (e.g. by adding or removing input data and modifying the classification criteria accordingly) was beyond the scope of this study.

The high MC-scores obtained by certain pairs of different RTs (Fig. 9), as well as the visual inspection of the DRP-maps, suggest that the perception of the intensity of DRPs varies among different authors. For example, the riparian zones on the reference maps were mostly mapped as RT2, but, where they were completely saturated and at least slightly sloped, they were mapped as RT1. In contrast, on MU09-maps and on SF07-maps the riparian zone was mostly mapped as RT1. Similarly, areas prone to DP on GH11-maps fitted best with RT4 areas of the reference maps, which represent areas where strongly delayed SOF or SSF, but not DP, occur. Since a straightforward, standardised definition of DRPs is missing, not only do the classification criteria vary, but also the classes. This can be misleading, especially if different classes have the same DRP names.

The MC-score ranking of the automatic mapping approaches is similar to the fuzzy kappa ranking, but the differences between the MC-scores were not as significant as those between the fuzzy kappa values (~~Table 6~~Fig. 9). This is because the degree of association of the maps we compared is moderate. In this case, significant increases of the degree of overlap entail only small increases of the MC-score (see Fig. 1 in Hargrove et al., 2006). This ~~confirms the reliability~~ problem~~s~~ was encountered also by ~~of Mapcurves, which~~ Speich et al. (2015). ~~also encountered~~. There is therefore a need for a Goodness-of-Fit score capable of comparing maps with different number of classes, while detecting improvements, as well, even if the degree of spatial overlap between maps being compared is moderate.

To keep the rainfall-runoff model as simple as possible strong assumptions had to be made. These included no interception, no evapotranspiration and completely saturated catchments. A calibration against measured runoff would have thus been meaningless. However, recent studies suggest that using expert knowledge in selecting parameter values and introducing constraints can increase the performance of conceptual models even without traditional calibration (Bahremand, 2016; Gharari et al., 2014; Hrachowitz et al., 2014). Therefore, the choice of realistic parameter values according to Viviroli et al. (2009a) and the introduction of parameter constraints allow the simulation results obtained to be plausible.

The complexity of the model structure is necessarily usually linked to the complexity of the DRP-mapping approaches. Two research directions have recently received attention, one using expert knowledge mainly in the phase of DRP identification and the other using this knowledge in the modelling phase. Hellebrand et al. (2011) used expert knowledge to determine the spatial distribution of DRPs as realistically as possible, as they assumed that with a more realistic DRP classification the modules representing each DRP in the model could be simplified. Gharari et al. (2014), in contrast, adopted a relatively complex combination of modules and fluxes to compensate for the rather simple classification they used. They, then, used expert knowledge to constrain both the model fluxes and parameters, to force the model to work well for the right reason by neglecting the actual spatial localisation of the DRPs.

As the results of this study suggest, the use of expert knowledge should not be limited only to the phases of landscape classification or model building and constraining, but should play a crucial role in each phase of the whole modelling process.

In this study, the same model structure and model constraints were applied to different DRP-mapping approaches. By doing so, it was possible to investigate the effects of a specific uncertainty source (i.e. the DRP-maps) on the system output (i.e. the simulated runoff), while keeping the other uncertainty sources fixed.

As the results indicate, the simplified classification approaches mostly fail in representing the spatial localisation of the DRPs and have a large impact on the simulated runoff. This finding suggests that investing more efforts in the landscape classification could enhance runoff predictions on ungauged catchments by improving the model realism. This topic will be further investigated during future research, by addressing the uncertainties linked to different input data, model structures, model parameters, and model constraints, as well as their interaction.

6 Conclusions

Mapping DRPs manually produces robust results but is time-consuming. Several ways of mapping DRPs automatically have been developed. They differ in terms of how much input data they require for mapping, their classification criteria, and the number of output classes.

In this study, three approaches to mapping DRPs automatically were compared in two catchments on the Swiss Plateau to determine which produces the most realistic results. The DRP-maps derived automatically with the most complex and most data demanding approach (Schmocker-Fackel et al. 2007) were most similar to the reference maps derived according to the manual approach based on Scherrer and Naef (2003), and resulted in the lowest deviations from them when used as input data for synthetic runoff simulations. The DRP-maps produced using Müller et al.'s (2009) simplified mapping approach, which requires no soil information, and those produced using Gharari et al.'s (2011) topography-based approach differed considerably and similarly from the reference maps in terms of DRPs' extent and distribution. The differences arose from the inaccuracy and the coarse resolution of the input data. The simplifying assumptions these two approaches require also limit their usefulness in automatically mapping small catchments.

The runoff simulations performed with these simplified DRP-maps significantly differed from those performed with the reference maps. It can be speculated that, i~~I~~t would be ~~therefore~~ worthwhile investing efforts and using expert knowledge to obtain hydrological landscape classifications that are as realistic as possible. A standardised definition of DRPs, moreover, would be helpful to avoid mapping bias due to researchers different perception of DRP intensity.

Author contribution

M. A. and M. Z. designed the comparisons and simulations, while R. B. and M. A. performed them. S. Sch. produced the reference maps, M. M. the SF07-maps, and M. A. and R. B. the

MU09-maps and GH11-maps. M. A. prepared the manuscript with contributions from all co-authors.

Acknowledgements

This work was funded through a PhD grant from the Swiss Federal Office for the Environment (FOEN). The meteorological data used as input was provided by the Swiss Federal Office of Meteorology and Climatology MeteoSwiss. Prof. Dr. Peter Molnar supervised the contribution of Rahel Buss. For the calculation of Mapcurve scores, the algorithm available online developed by Dr. Emiel van Loon (University of Amsterdam) was used. We are also grateful to Daniel Farinotti and Anil Yildiz for their helpful comments and feedback, and to Silvia Dingwall for the English revision of the manuscript. We thank Shervan Gharari and another anonymous reviewer for their comments and suggestions on a previous version of the manuscript.

References

Bahremand, A.: HESS Opinions: Advocating process modeling and de-emphasizing parameter estimation, Hydrol. Earth Syst. Sci., 20, 1433-1445, doi: 10.5194/hess-20-1433-2016, 2016

Beran, M. A.: New Challenges for Regional Approach, in: Regionalization in Hydrology. Proceedings of an international symposium held at Ljubljana, April 1990. Beran M.A., Becker A., Bonacci O., Eds. IASH publication 191. Wallingford, UK, 1990.

Beven, K. J., and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43-69, doi: 10.1080/02626667909491834, 1979.

Blöschl, G.: Scaling in hydrology, Hydrological Processes, 15, 709-711, doi: 10.1002/hyp.432, 2001.

Bolliger, T.: Geologie des Kantons Zürich. Stiftung Geologische Karte des Kantons Zürich, Ott Verlag Thun, 1999.

Boorman, D. B., Hollis, J. M., and Lilly, A.: Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom., Institute of Hydrology, Wallingford, 146, 1995.

Cohen, J.: A Coefficient of Agreement for Nominal Scales, Educational and Psychological Measurement, 20, 37-46, doi: 10.1177/001316446002000104, 1960.

Dobmann, J.: Hochwasserabschätzung in kleinen Einzugsgebieten der Schweiz. Interpretations- und Praxishilfe., Südwestdeutscher Verlag für Hochschulschriften~~ AG Co. KG.~~, Saarbrücken, 2010~~09~~.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, Journal of Hydrology, 537, 45-60, doi: 10.1016/j.jhydrol.2016.03.026, 2016.

Flügel, W.-A.: Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the

River Bröl, Germany, Hydrological Processes, 9, 423-436, doi: 10.1002/hyp.3360090313, 1995.

Franks, S. W., Gineste, P., Beven, K. J., and Merot, P.: On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resources Research, 34, 787-797, doi: 10.1029/97WR03041, 1998.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrol. Earth Syst. Sci., 18, 1895-1915, doi: 10.5194/hess-18-1895-2014, 2014.

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.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18, 4839-4859, doi: 10.5194/hess-18-4839-2014, 2014.

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.

Hantke, R. e. a.: Geologische Karte des Kantons Zürich und seine Nachbargebiete in 2 Blättern 1:50'000., Kommissionsverlag Lehmann, 1967.

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.

Hellebrand, H., Müller, C., Matgen, P., Fenicia, F., and Savenije, H.: A process proof test for model concepts: Modelling the meso-scale, Physics and Chemistry of the Earth, 36, 42-53, doi: 10.1016/j.pce.2010.07.019, 2011.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resources Research, 50, 7445-7469, doi: 10.1002/2014WR015484, 2014.

Hrachowitz, M., Savenije, H. H. G., Bloschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)-a review, Hydrolog Sci J, 58, 1198-1255, doi: 10.1080/02626667.2013.803183, 2013.

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.

Jörg-Hess, S., Griessinger, N., and Zappa, M.: Probabilistic Forecasts of Snow Water Equivalent and Runoff in Mountainous Areas, Journal of Hydrometeorology, 16, 2169-2186, doi: 10.1175/JHM-D-14-0193.1, 2015.

Kienholz, H., Keller H., Ammann W., Weingartner R., Germann P., Hegg Ch., Mani P., Rickenmann D.: Zur Sensitivität von Wildbachsystemen. Schlussbericht NFP 31, VDF, Hochschulverlag an der EZH Zürich, 1998.

Klemeš, V.: Dilettantism in hydrology: Transition or destiny?, Water Resources Research, 22, 177S-188S, doi: 10.1029/WR022i09Sp0177S, 1986.

Margreth, M., Naef, F., and Scherrer, S.: Weiterentwicklung der Abflussprozesskarte Zürich in den Waldgebieten. Amt für Abfall, Energie, Wasser und Luft des Kantons Zürich (AWEL), 2010.

Markart, G., B. Kohl, B. Sotier, K Klebinder, T. Schauer, G. Bunza, H. Pirkl, R. Stern: A simple Code of Practice for the Assessment of Surface Runoff Coefficients for ALpine Soil-/Vegetation Units in Torrential Rain (Version 2.0), Department of Natural Hazards, Federal Research and Training Centre for Forest, Natural Hazards and Landscaper (BFW), 127, 2011.

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.

Mosley, M. P.: Delimitation of New Zealand hydrologic regions, Journal of Hydrology, 49, 173-192, doi: http://dx.doi.org/10.1016/0022-1694(81)90211-0, 1981.

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.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282-290, doi: 10.1016/0022-1694(70)90255-6, 1970.

Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M., and Saleska, S.: Height Above the Nearest Drainage – a hydrologically relevant new terrain model, Journal of Hydrology, 404, 13-29, doi: 10.1016/j.jhydrol.2011.03.051, 2011.

Peschke, G., Etzenberg, C., Töpfer, J., Zimmermann, S., and Müller, G.: Runoff generation regionalization: analysis and a possible approach to a solution, IAHS Publ. 254 (Regionalization in Hydrology), 1999.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, Remote Sensing of Environment, 112, 3469-3481, doi: 10.1016/j.rse.2008.03.018, 2008.

Ross, B. B., Contractor, D. N., and Shanholtz, V. O.: A finite-element model of overland and channel flow for assessing the hydrologic impact of land-use change, Journal of Hydrology, 41, 11-30, doi: 10.1016/0022-1694(79)90101-X, 1979.

Savenije, H. H. G.: HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)", Hydrol. Earth Syst. Sci., 14, 2681-2692, doi: 10.5194/hess-14-2681-2010, 2010.

Scherrer AG: Ermittlung massgebender Hochwasserabflüsse der Reppisch. Technical Report commissioned by AWEL, Ct. Zurich 2006.

Scherrer, S.: Bestimmungsschlüssel zur Identifikation von hochwasserrelevanten Flächen, Report 18/2006 commissioned by LUWG, Mainz, 2006.

Scherrer, S., and Naef, F.: A decision scheme to indicate dominant hydrological flow processes on temperate grassland, Hydrological Processes, 17, 391-401, doi: 10.1002/hyp.1131, 2003.

Schmocker-Fackel, P., Naef, F., and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrol. Earth Syst. Sci., 11, 891-906, doi: 10.5194/hess-11-891-2007, 2007.

Seibert, J., and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resources Research, 38, 23-21-23-14, doi: 10.1029/2001WR000978, 2002.

Seibert, J., and McGlynn, B.: Landscape Element Contributions to Storm Runoff, in: Encyclopedia of Hydrological Sciences, John Wiley & Sons, Ltd, 2006.

Sideris, I. V., Gabella, M., Erdin, R., and Germann, U.: Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland, Quarterly Journal of the Royal Meteorological Society, 140, 1097-1111, doi: 10.1002/qj.2188, 2014.

Speich, M. J. R., Bernhard, L., Teuling, A. J., and Zappa, M.: Application of bivariate mapping for hydrological classification and analysis of temporal change and scale effects in Switzerland, Journal of Hydrology, 523, 804-821, doi: 10.1016/j.jhydrol.2015.01.086, 2015.

Tetzlaff, D., Soulsby, C., Waldron, S., Malcolm, I. A., Bacon, P. J., Dunn, S. M., Lilly, A., and Youngson, A. F.: Conceptualization of runoff processes using a geographical information system and tracers in a nested mesoscale catchment, Hydrological Processes, 21, 1289-1307, doi: 10.1002/hyp.6309, 2007.

Tilch, N., Uhlenbrook, S., and Leibundgut, C.: Regionalisierungsverfahren zur Ausweisung von Hydrotopen in von periglazialem Hangschutt geprägten Gebieten, Grundwasser, 7, 206-216, doi: 10.1007/s007670200032, 2002.

Visser, H., and de Nijs, T.: The Map Comparison Kit, Environmental Modelling & Software, 21, 346-358, doi: 10.1016/j.envsoft.2004.11.013, 2006.

Viviroli, D., Mittelbach, H., Gurtz, J., and Weingartner, R.: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results, Journal of Hydrology, 377, 208-225, doi: 10.1016/j.jhydrol.2009.08.022, 2009a.

Viviroli, D., Zappa, M., Gurtz, J., and Weingartner, R.: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, Environ. Model. Softw., 24, 1209-1222, doi: 10.1016/j.envsoft.2009.04.001, 2009b.

Waldenmeyer, G.: Abflussbildung und Regionalisierung in einem forstlich genutzten Einzugsgebiet (Dürreychtal, Nordschwarzwald), Karlsruher Schriften zur Geographie und Geoökologie, IFGG, 2003.

Weiler, M., and McDonnell, J.: Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology, Journal of Hydrology, 285, 3-18, doi: 10.1016/S0022-1694(03)00271-3, 2004.

Woods, R. A., Sivapalan, M., and Robinson, J. S.: Modeling the spatial variability of subsurface runoff using a topographic index, Water Resources Research, 33, 1061-1073, doi: 10.1029/97WR00232, 1997.

Table 1. List of hydrological classifications based on DRPs, the data they require and the number of output classes (A = Automatic; M = Manual).

	Approach	Topography	Land-use	Geology	Soil maps	Drainage maps	Forest-vegetation maps	Extensive field investigations	Number of output classes
Boorman et al. (1995)	A				x				29
Peschke et al. (1999)	A	x	x	x	x				7
Tilch et al. (2002)	M	x		x					6
Waldenmeyer (2003)	A	x					x		7
Scherrer and Naef (2003)	M	x	x	x	x	x	x	x	9
Schmocker-Fackel et al. (2007)	A	x	x	x	x	x	x	x	12
Tetzlaff et al. (2007)	A	x	x	x				x	5
Müller et al. (2009)	A	x	x	x					9
Gharari et al. (2011)	A	x							3
Hümann and Müller (2013)	A	x	x	x					10
Gao et al. (2014)	A	x	x						4

Table 2. Reclassification of DRPs according to runoff types (HOF = Hortonian Overland Flow; SOF = Saturation Overland Flow; SSF = Subsurface Flow; DP = Deep percolation; 1 represents an almost immediate reaction, 2 a slightly delayed one and 3 a strong delayed one). Adapted from Naef et al. (2000).

Runoff type (RT)	DRP	Runoff intensity
1	HOF1/2, SOF1	Fast
2	SOF2, SSF1	Slightly delayed
3	SSF2	Delayed
4	SOF3, SSF3	Strongly delayed
5	DP	Not contributing

Table 3. Dependency of the DRP on the slope and permeability of the substratum for grassland, arable land and forest, according to Müller et al. (2009).

Slope	Impermeable substratum		Permeable substratum
[%]	Grass- and arable land	Forest	Grass-, arable land and forest
0 – 3	SOF3	SOF3	DP
3 –5	SOF2	SSF3	DP
5 – 20	SSF2	SSF2	DP
20 – 40	SSF1	SSF2	DP
> 40	SSF1	SSF1	DP

Table 4. List of areas identified in Fig. 8 with the automatically and manually derived DRPs (RTs), and a possible explanation for their deviation.

Area	DRP (RT) on MU09-map	DRP (RT) on SN03-map	Explanation
1	SSF2 (RT3)	DP (RT5)	Moraine not necessarily impermeable
2	SSF1 (RT2)	SSF3 (RT4)	Although high slope, high storage capacity of soil
3	DP (RT5)	SSF2 (RT3)	Alluvium not necessarily permeable
4	SOF3 (RT4)	SOF2 (RT2)	Although low slope, low storage capacity of soil
5	SOF1 (RT1)	SSF2 (RT3)	Coarse resolution of DTM
6	SOF1 (RT1)	SSF2 (RT3)	Coarse resolution of land-use map

Appendix A

Table A1. Reclassification of the reference maps for the class comparison with the GH11-maps.

Combination	1	2	3	4	5	6
Wetland	RT1	RT1	RT 1	RTs 1, 2	RTs 1, 2	RTs 1, 2, 3
Hillslope	RT 2	RTs 2, 3	RTs 2, 3, 4	RT 3	RTs 3, 4	RT 4
Plateau	RTs 3, 4, 5	RTs 4, 5	RT 5	RTs 4, 5	RT 5	RT 5

Table A2. Parameter values used for the 16 runs of the synthetic runoff simulations. The simulation names are of the form “i.j”, where i refers to the number of parameters defined based on the RTs and j refers to the different combinations.

Simulation name		0.1	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.5	2.6	3.1	3.2	3.3	3.4	4.1
SGRLUZ1	[mm]	30	5	30	30	30	5	5	5	30	30	30	5	5	5	30	5
SGRLUZ2	[mm]	30	15	30	30	30	15	15	15	30	30	30	15	15	15	30	15
SGRLUZ3	[mm]									30							
SGRLUZ4	[mm]	30	100	30	30	30	100	100	100	30	30	30	100	100	100	30	100
SGRLUZ5	[mm]	30	200	30	30	30	200	200	200	30	30	30	200	200	200	30	200
K0H1	[h]	20	20	5	20	20	5	20	20	5	5	20	5	5	20	5	5
K0H2	[h]	20	20	10	20	20	10	20	20	10	10	20	10	10	20	10	10
K0H3	[h]									20							
K0H4	[h]									20							
K0H5	[h]									20							
K1H1	[h]	100	100	100	10^3	100	100	10^3	100	10^3	100	10^3	10^3	100	10^3	10^3	10^3
K1H2	[h]	100	100	100	50	100	100	50	100	50	100	50	50	100	50	50	50
K1H3	[h]									100							
K1H4	[h]	100	100	100	150	100	100	150	100	150	100	150	150	100	150	150	150
K1H5	[h]	100	100	100	10^3	100	100	10^3	100	10^3	100	10^3	10^3	100	10^3	10^3	10^3
CPERC1	[mm/h]	0.12	0.12	0.12	0.12	0.04	0.12	0.12	0.04	0.12	0.04	0.04	0.12	0.04	0.04	0.04	0.04
CPERC2	[mm/h]	0.12	0.12	0.12	0.12	0.08	0.12	0.12	0.08	0.12	0.08	0.08	0.12	0.08	0.08	0.08	0.08
CPERC3	[mm/h]									0.12							
CPERC4	[mm/h]	0.12	0.12	0.12	0.12	0.16	0.12	0.12	0.16	0.12	0.16	0.16	0.12	0.16	0.16	0.16	0.16
CPERC5	[mm/h]	0.12	0.12	0.12	0.12	0.2	0.12	0.12	0.2	0.12	0.2	0.2	0.12	0.2	0.2	0.2	0.2
CG1H	[h]									600							
SLZ1MAX	[mm]									150							
K2H	[h]									2500							

Table A3. Parameter combinations for the simulations driven by the GH11-maps. ~~θ~~ϑ = SGRLUZ, K0H, K1H, CPERC. Subscripted numbers refer to the RTs.

Combination	A	B	C	D	E	F	G	H	I	J
$\vartheta_{WETLAND}$	ϑ_1	ϑ_1	ϑ_1	ϑ_1	ϑ_1	ϑ_1	ϑ_2	ϑ_2	ϑ_2	ϑ_3
$\vartheta_{HILLSLOPE}$	ϑ_2	ϑ_2	ϑ_2	ϑ_3	ϑ_3	ϑ_4	ϑ_3	ϑ_3	ϑ_4	ϑ_4
$\vartheta_{PLATEAU}$	ϑ_3	ϑ_4	ϑ_5	ϑ_4	ϑ_5	ϑ_5	ϑ_4	ϑ_5	ϑ_5	ϑ_5

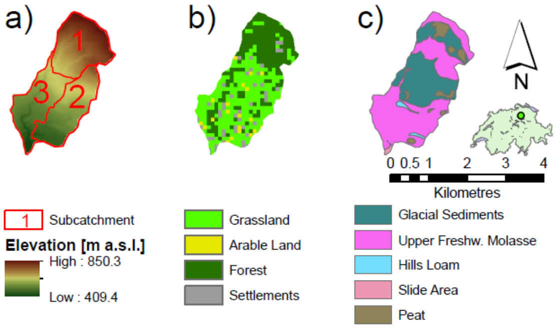


Figure 1. Overview of the Meilen catchment, Switzerland. a) Digital Terrain Model (25 m resolution) subdivided into 3 sub-catchments; b) Land-use map (100 m resolution); c) Geology map (data: BFS GEOSTAT/Federal Office of Topography swisstopo)

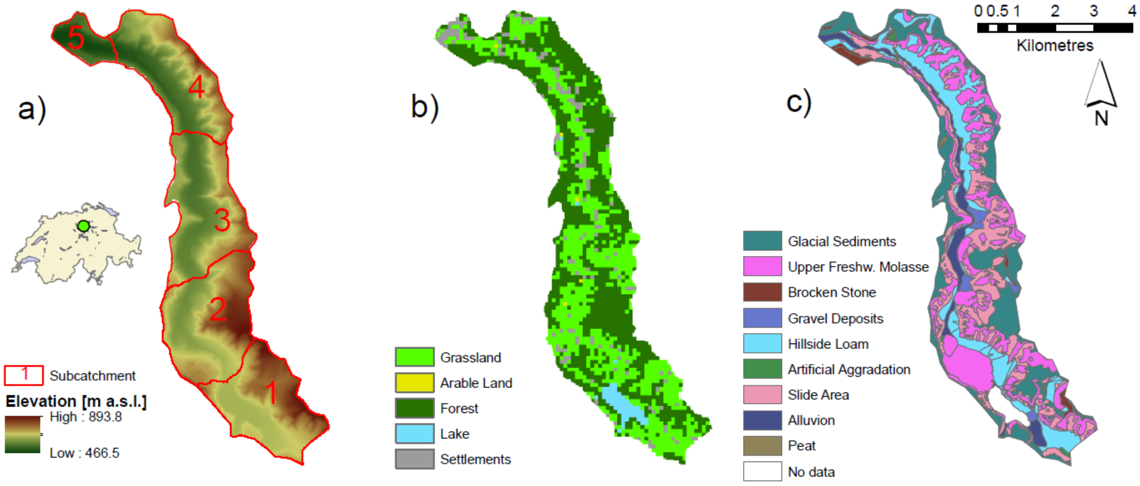


Figure 2. Overview of the Reppisch catchment, Switzerland. a) Digital Terrain Model (25 m resolution) subdivided into 5 sub-catchments; b) Land-use map (100 m resolution); c) Geology map (data: BFS GEOSTAT/ Federal Office of Topography swisstopo)

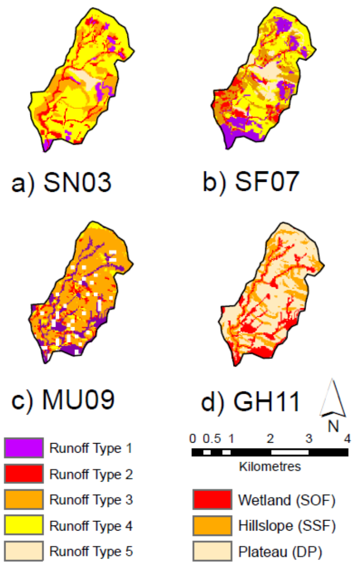


Figure 3. DRP-maps for the Meilen catchment: (a) Reference map according to Scherrer and Naef (2003) and automatically derived map according to (b) Schmocker-Fackel et al. (2007); (c) Müller et al. (2009); (d) Gharari et al. (2011).

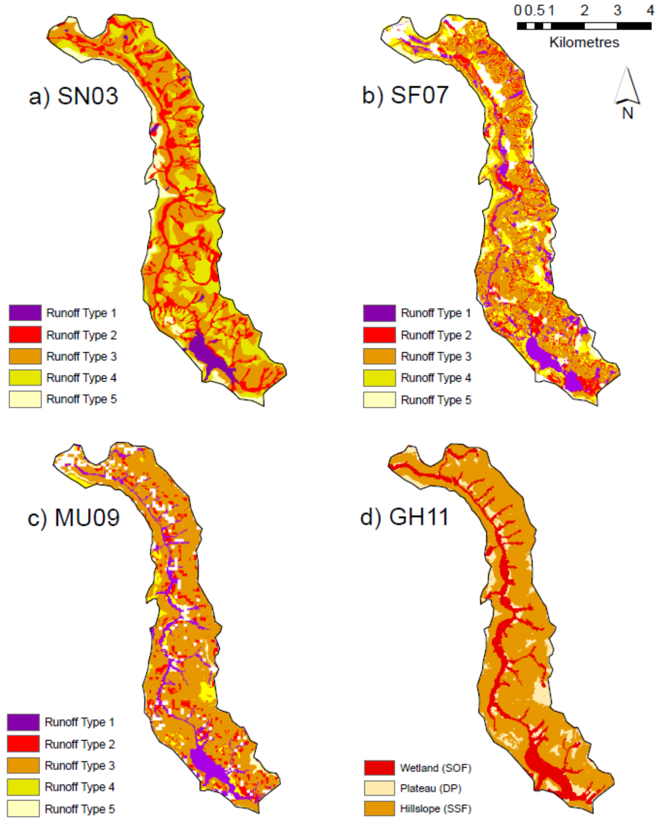
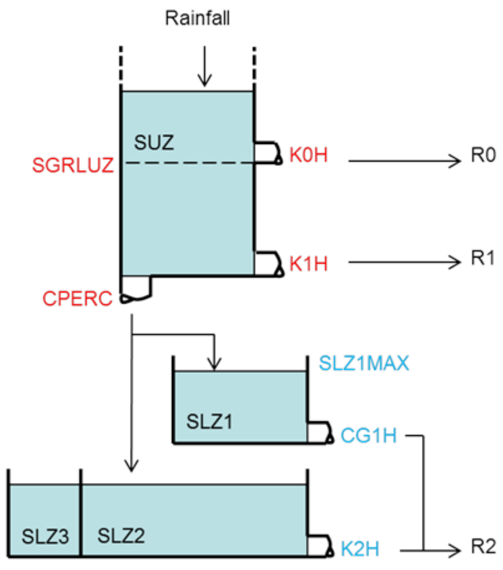


Figure 4. DRP-maps for the Reppisch catchment: (a) Reference map according to Scherrer and Naef (2003) and automatically derived map according to (b) Schmocker-Fackel et al. (2007); (c) Müller et al. (2009); (d) Gharari et al. (2011).



	PARAMETERS
SGRLUZ	Threshold for quick runoff formation [mm]
K0H	Storage time for surface runoff [h]
K1H	Storage time for subsurface runoff [h]
CPERC	Percolation rate [mm/h]
SLZ1MAX	Max. content of quick baseflow storage [mm]
CG1H	Storage time for quick baseflow [h]
K2H	Storage time for slow baseflow [h]
	STORAGE MODULES
SUZ	Upper zone (unsaturated) runoff storage [mm]
SLZ1-3	Lower zone (saturated) runoff storages [mm]
	FLUXES
R0	Quick runoff (surface runoff) [mm/h]
R1	Delayed runoff (subsurface runoff) [mm/h]
R2	Slow runoff (baseflow) [mm/h]

Figure 5. Runoff generation module of PREVAH, adapted from Viviroli et al. (2009). Parameters in blue are averaged for the whole catchment, while parameters in red are adapted stepwise to the RTs.

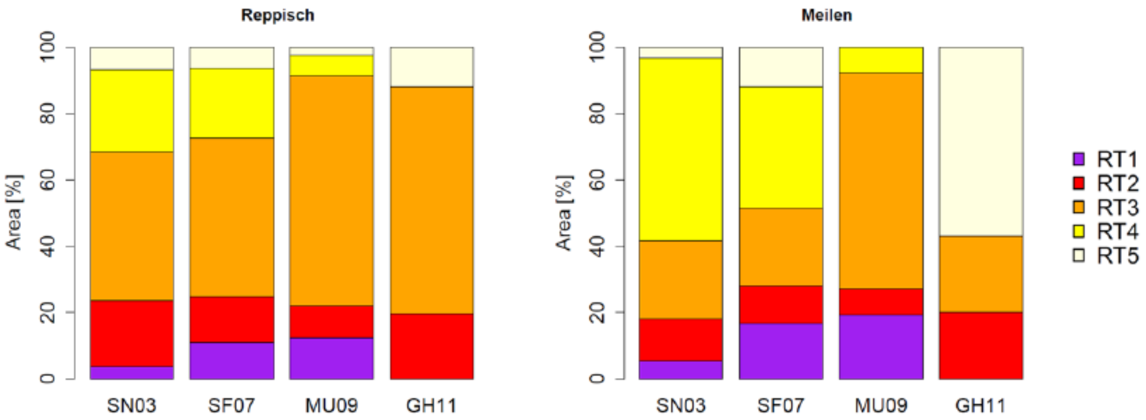


Figure 6. Percentage of total catchment area assigned to each runoff type in the (a) Reppisch (b) Meilen catchments with the four different mapping approaches.

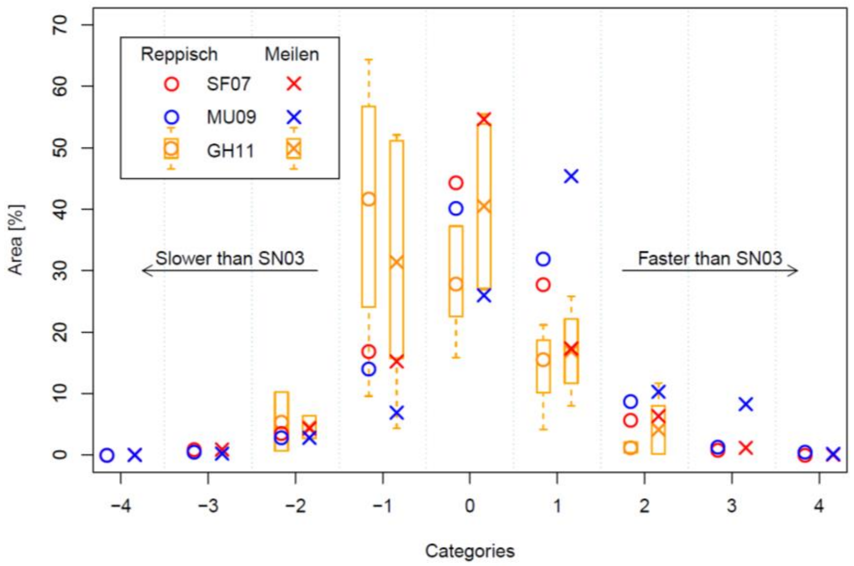


Figure 7. Distribution of the class deviations of the different automatic mapping approaches from the reference maps (circles refer to the Reppisch catchment and crosses to the Meilen catchment). The boxplots show median and interquartile ranges from the comparison between GH11-maps and the reclassified reference maps.

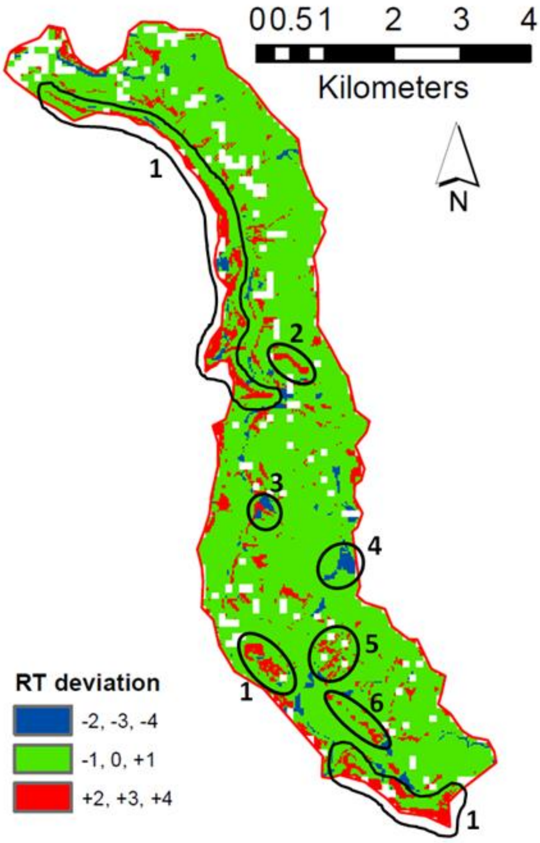


Figure 8. Deviation map between the MU09-map and the reference map. In the numbered areas the runoff contribution was either over- (red) or underestimated (blue).

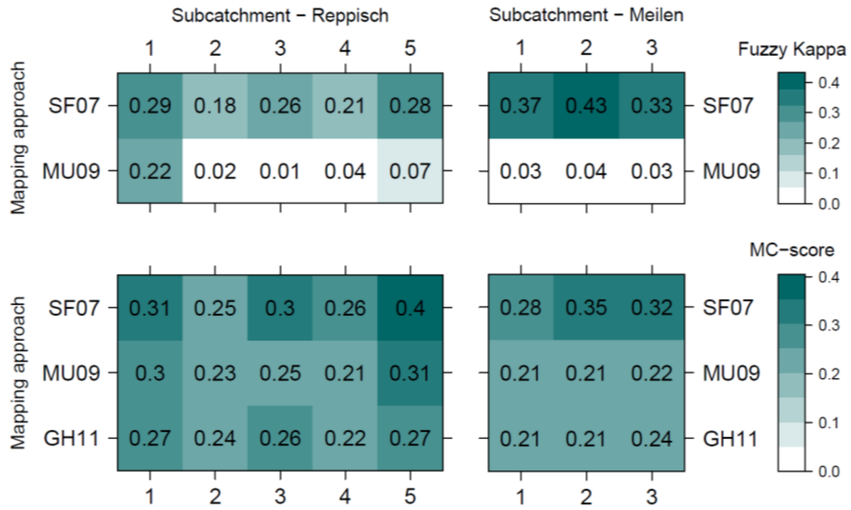


Figure 9. Agreement scores K_{Fuzzy} and MC-scores obtained by comparing the maps derived with the automatic mapping approaches SN07, MU09 and GH11 with the reference maps (SN03) for the sub-catchments of the two study areas.

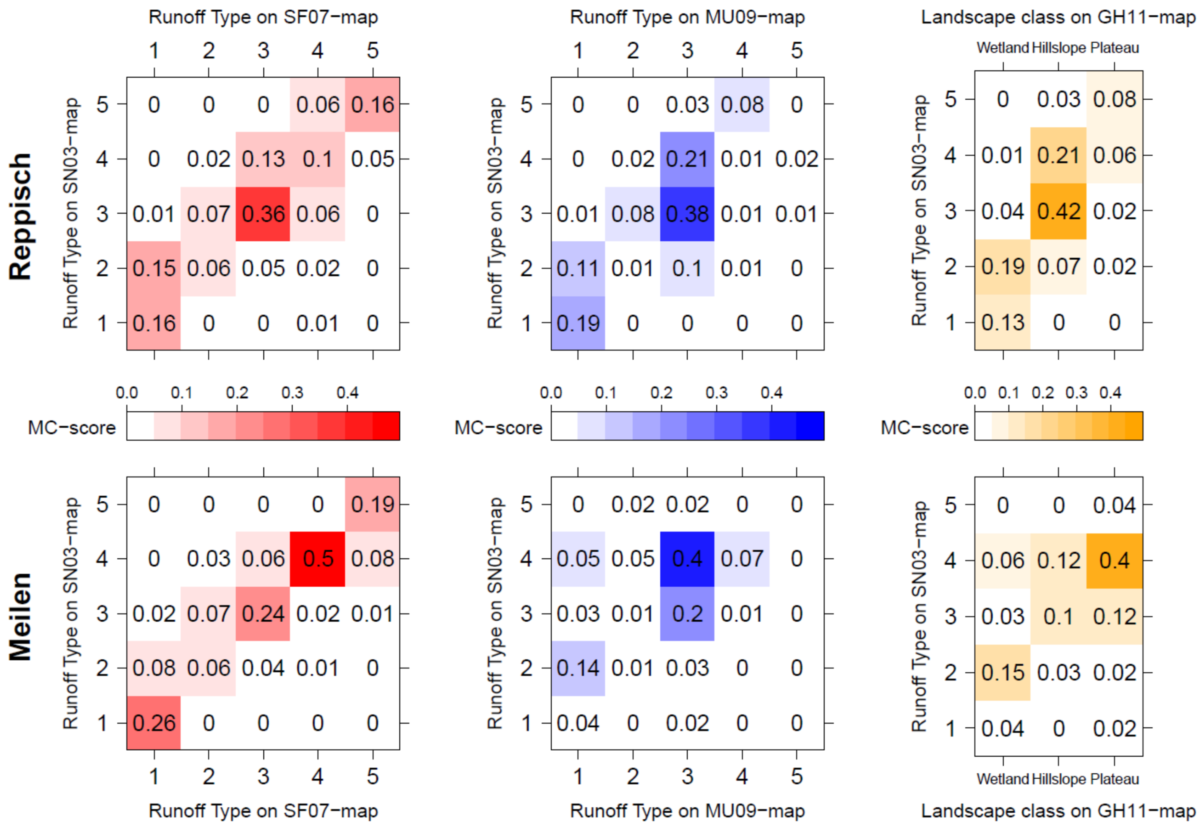


Figure 10. MC-scores related to each RT obtained by comparing the maps derived with the automatic mapping approaches SN07, MU09 and GH11 with the reference maps (SN03) for the two study sites.

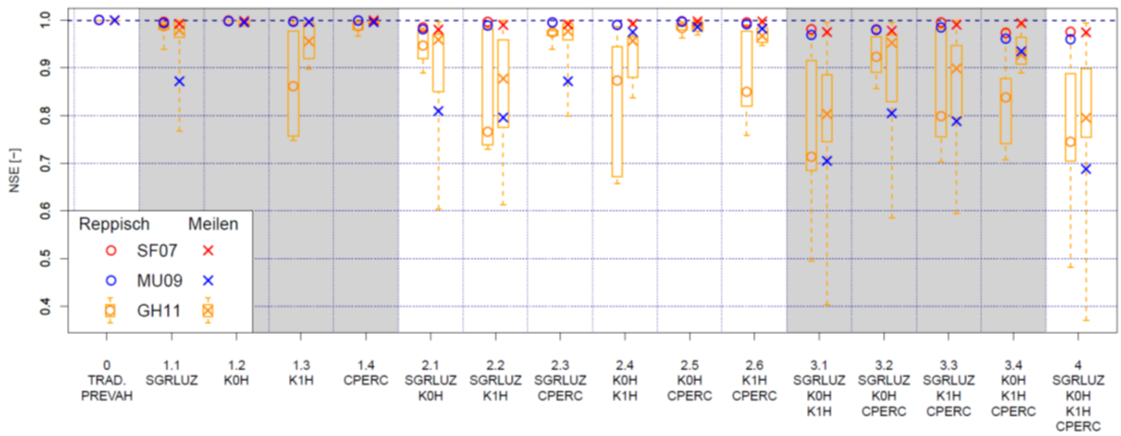


Figure 11. Modified NSE obtained by comparing the runoff simulated with the automatic DRP-maps with that simulated with the reference maps, in the two study sites (simulation period 16/06/2014 - 15/08/2014). The boxplots show the medians and the interquartile ranges of the simulations driven by GH11-maps, while the labels on the abscissa show the model parameters, whose values were defined based on the RTs.

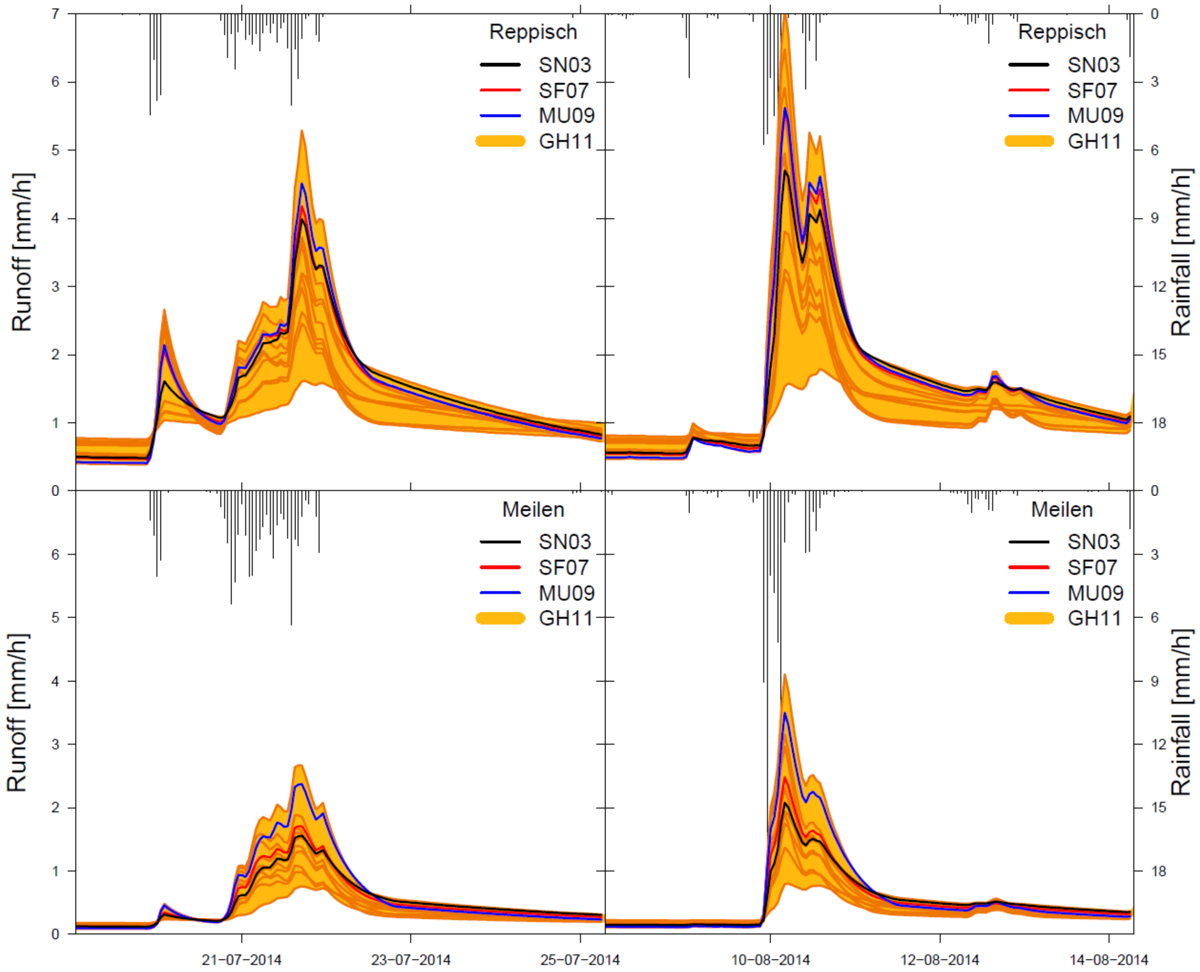


Figure 12. Simulated runoff during the two heaviest rainfall events of the simulation period, obtained from the different DRP-maps for the two study sites by varying the parameter values for each RT (simulation 4.1 of Table A2). The bands represent the minimum and maximum runoff values obtained with the different parameter combinations for the simulations driven by GH11-maps.

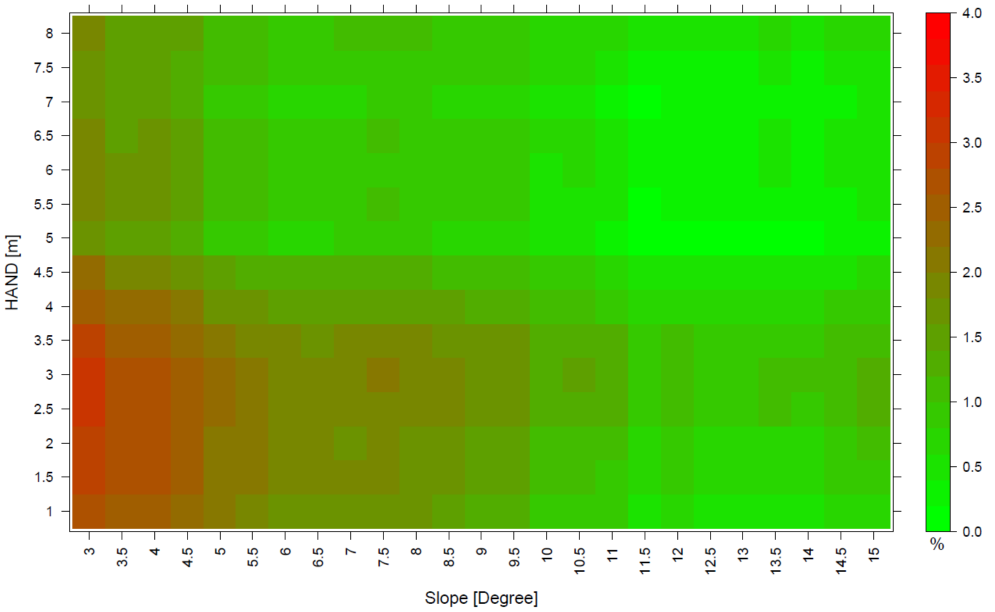


Figure A1. Sensitivity analysis of the threshold values for the HAND-based landscape classification on the whole Reppisch catchment. The level plot shows the percentage of deviation from the maximal MC-score (0.2023) obtained by comparing GH11-maps with the reference maps.