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 will consider 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" etc.) without having a known solution to achieve (e.g. Weiler and McDonnell, 2004; Lanni et al., 2012; Frei et al., 2012). In our study, the focus is not on calibrating a model to a measured discharge. Contrarily, we want to investigate 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 will state 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 will therefore reformulate 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 will rephrase the last sentence of the discussion as also declared in the reply to the other reviewer.

“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, not shown in the manuscript) and what we found is that the same values found by Gharari et al. (2014) were applicable on our study areas.

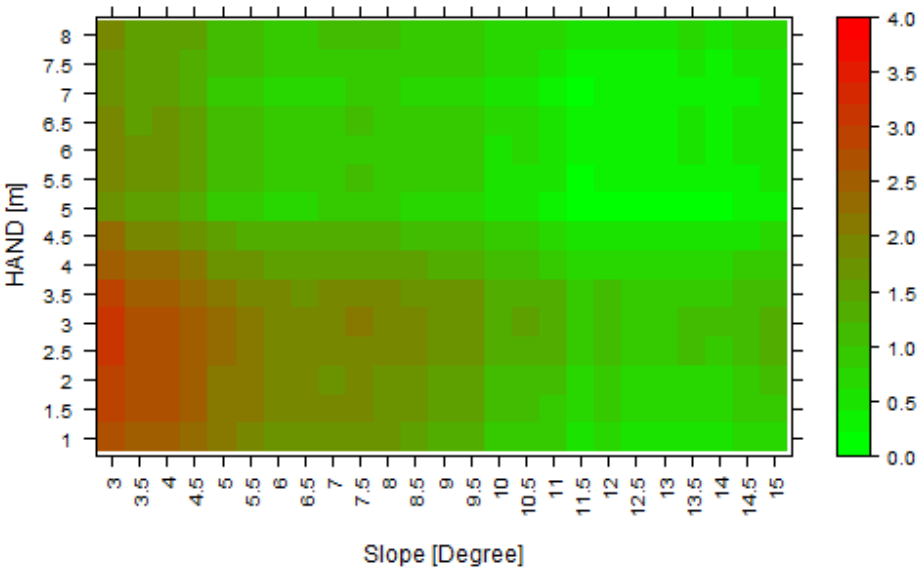

Figure 1. Sensitivity analysis of the threshold values for the HAND-based landscape classification on the Reppisch catchment. The level plot shows the percentage of deviation from the maximal MC-score (0.2023) obtained by comparing GH11-maps with the reference maps.

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.

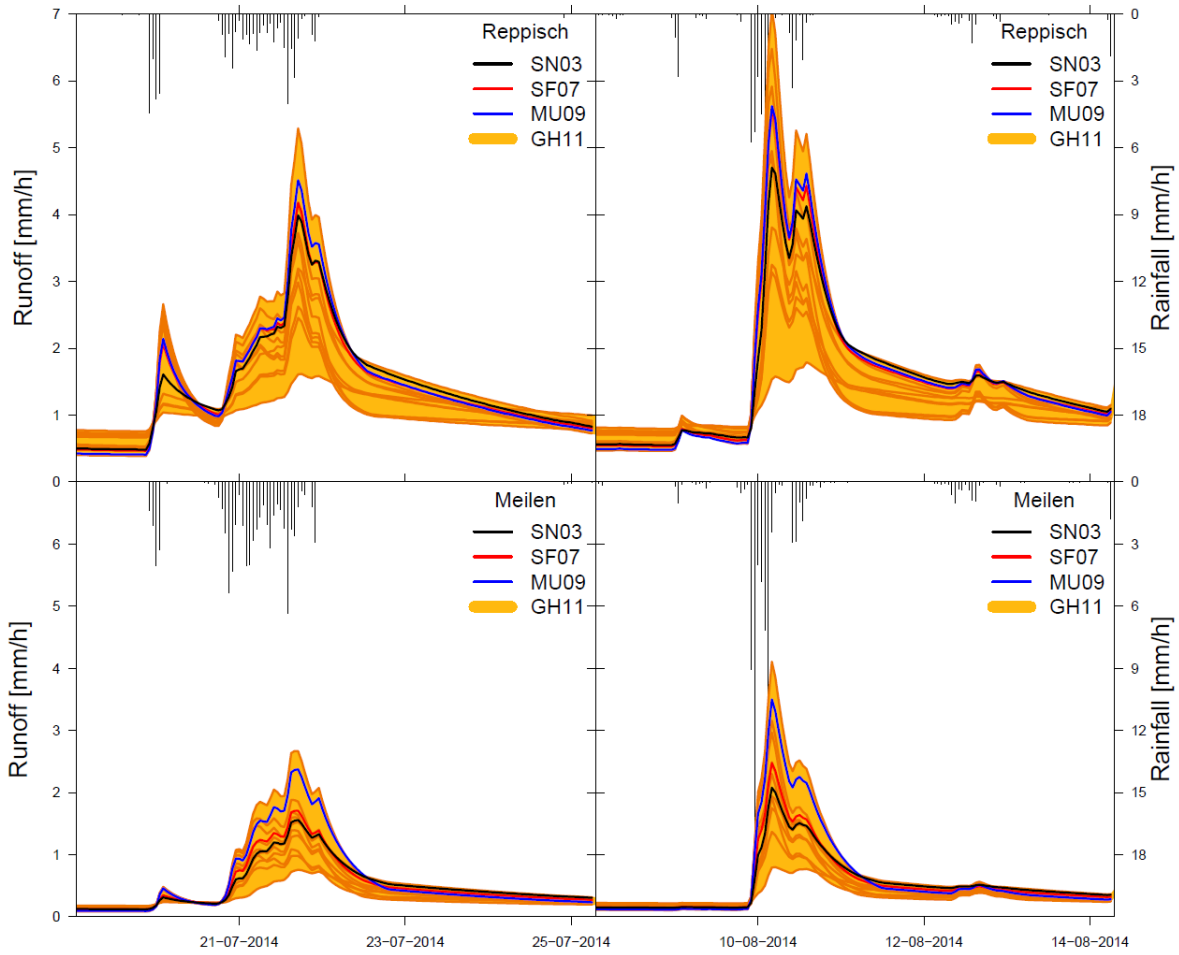

Figure 2. Simulated runoff during the two heaviest rainfall events of the simulation period, obtained from the different DRP-maps for the two study sites by varying the parameter values for each RT.

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.

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.

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.