**Hydrology and
Earth System
Sciences
Discussions**

Interactive comment on "Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria" *by* A. Viglione et al.

A. Viglione et al.

viglione@hydro.tuwien.ac.at

Received and published: 7 April 2013

We would like to thank the reviewer for her/his constructive comments on the manuscript. In the following Referee #1's comments are in *italic* and our responses in plain text.

The manuscript "Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria" by Viglione et al. compares how a process-based and a statistical modelling approach can predict various runoff signatures, characteristic for different time scales in a cross-validation approach for more than 200 catchments, spanning a significant gradient of hydroclimatic and geologic/pedologic conditions. It is a very interesting topic which is quite timely and illustrates the importance of not only getting one aspect of the runoff right in models, but ideally many more simulta-

neously to make sure that the entire characteristics of the system are well captured. The manuscript is very well written and structured, however, although the some of the methods used were common practice for several decades now, I am not sure if some details of these methods can, in the light of recent advances in this area, still be called "sound" in the strictest sense. Furthermore, although the manuscript starts off very promising it does in the end not quite meet the expectations, as unfortunately a wider and deeper discussion of the results is not provided, i.e. what are the wider implications of the results? How do the results help to improve modelling strategies? Do the results contribute anything to how we can better predict? This makes the paper a nice standalone comparison of the predictive power of two modelling approaches in a certain environmental setting but the authors unfortunately do not try to generalize the results in some way or another. Having said that, I think the manuscript should definitely be published eventually, as it will make a nice contribution to literature if the authors gave some considerations to the detailed comments below.

Referee #1 comments are very valuable. We wrote a "too enthusiastic" introduction which will be revised in order to better explain the objectives of this paper and match the results accordingly. For instance, while the other two companion papers concentrate on individual signatures across many geographic regions, the motivation for this paper is to assess prediction in ungauged basins for several signatures in the same geographic region. Analogously, the results will be discussed more deeply in a discussion section. For instance, we will discuss more deeply the implications, such as that extremes are harder to estimate with process-based methods than with statistical ones. This is reflected on the fact that all studies considered in Salinas et al. (2013) for regionalisation of extremes use statistical methods.

Regarding the regionalisation methods used, the focus of the paper is not on the choice of models (and their parametrisation) but on the assessment of existing methods. This is analogous to what we did in the other two companion papers but with the difference that the two methods are applied here on the same region (these are two methods

which were evaluated as the best ones in Austria in previous studies such as Merz and Blöschl, 2004, and Parajka et al., 2005). We will clarify this strongly in the introduction and we will add an extended description on the methods and their parameterisation in section 3, along with the relevant literature.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, Journal of Hydrology 287, 95–123, doi:10.1016/j.jhydrol.2003.09.028, 2004.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrol. Earth Syst. Sci., 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.

Salinas, J. L., Parajka, J., Viglione, A., Rogger, M., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrol. Earth Syst. Sci. Discuss., 10, 411–447, doi:10.5194/hessd-10-411-2013, 2013.

(1) p.451, l.5: please provide a references for the theory behind cross-validation and possibly one example of successful application

Cross-validation is a classic statistical technique extensively used in hydrology. We added a citation to Efron, B., Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. The American Statistician, 37(1), 36-48.

(2) p..454, l.15-16: should maybe better read:"were the central part of the flow duration curves are particularly flat"

Yes. Corrected.

(3) p.456, l.5ff: this part raises some questions and it is in fact my biggest concern. Unfortunately, the reader is not given any detailed information on the calibration strategy and on how the parameterizations used in the end were chosen. Were the models calibrated on the hydrograph or on all signatures? Which objective function(s) were used – the same as for the rest of the analysis? If only one objective function was

used (e.g. Nash-Sutcliffe Efficiency), how can it be expected that the predicted catchments perform well with respect to the other signatures? If we want to ensure space-transferability, shouldn't we first make sure that our parametrerizations are transferable in time, i.e. to make sure that the chosen parameterizations give adequate results in an independent test (validation) period as stressed by Klemes (1986), Andreassian et al. (2009) and in an innovative approach recently addressed by Gharari et al. (2013)? If such tests are not carried out, we run the risk of choosing the "optimal" parameterization as a result of a mere mathematical fitting exercise ("mathematical marionettes" as termed by Kirchner (2006)), rather than on basis of an adequate process representation, thus leading to limited predictive power. Similarly, Beven (2006) and Andreassian et al. (2012) noted that frequently the most suitable parameterization for a model in a given catchment is "sub-optimal"! Of course, it has a long tradition just to calibrate models and declare the parameterization with the highest performance the most suitable one – BUT: we could do SO MUCH BETTER! This is in my opinion especially true in an analyses as the presented one, were prediction is at its core. Maybe this has all be done by the authors, but then I think it should be prominently commented on as these are crucial details. Further, although there is, for practical reasons, in principle no strong argument against one-fits-all modeling approaches in studies like this one, I could, however, imagine that predictions could be improved by at least introducing 6-7 different model classes depending on the dominant runoff pattern (similar to what was recently shown by Ye et al., 2012). Having said that, I do not necessarily want the authors to redo their entire analysis, although this would a fantastic effort to increase the relevance of the results, but I would at least like to encourage them to give the above mentioned concerns some consideration and discuss the limitations of their methodology accordingly.

As discussed above, the focus of the paper is not on the choice of models (and their parametrisation) but on the assessment of existing methods. However the remarks of Referee #1 will be discussed since they are very relevant and we completely agree with them. We will add an extended description on the methods and their parameterisation

in section 3, which were performed before and independently from our analysis, and which are published in Parajka et al. (2005-2007), Merz et al. (2011) and will be published in Parajka et al. (2013, in prep). For the process-based method, temporal validation has been performed and discussed recently in Merz et al. (2011). We will add some of the details in Section 3. For the regionalisation exercise in our manuscript, the cross validation procedure has been applied using parameters in the calibration period for the donor catchments (see also Merz et al., 2004, and Parajka et al., 2005).

Although we assess the performances on the signatures, the models have not been calibrated to them. In the revised papers we will discuss this point and add references to the relevant literature (e.g., papers of Thorsten Wagner's group). We think that the assessment on signatures is not the result of mathematical fitting, since the calibration was not performed on signatures, but captures strengths and weaknesses of the methods in capturing hydrological variability in Austria.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, Journal of Hydrology 287, 95–123, doi:10.1016/j.jhydrol.2003.09.028, 2004.

Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, Water Resour. Res., 47, W02531, doi:10.1029/2010WR009505, 2011.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrol. Earth Syst. Sci., 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrol. Process., 21, 435–446, doi:10.1002/hyp.6253, 2007.

Parajka et al. (2013, in prep.) Optimal station density for runoff regionalisation by Topkriging, in preparation, 2013.

(4) p.457, l.17 l.23: please adjust table numbering in the correct order

Yes, Table 1 and 2 have been switched.

(5) p.457, l.23, Table 1: maybe more instructive to show as boxplots rather than in a table

We prefer the table because boxplots would not add information and would be confusing because of different scales for each variable

(6) p.460, l.15: why 0.368? please explain where this value comes from

0.368 is simply 1/e rounded at the 3rd digit

(7) p.464, l.10ff: comparing the results to catchment characteristics is a very important and instructive part of the paper. However, I was surprised that the analysis was not done with some more depth, for example also including some measures of catchment organization (e.g. drainage density, or the the inverse of that, the average flow path lengths; average flow path gradients; some indicator of soil types; or height above nearest drainage – HAND (Renno et al., 2008) as proxy for the hydraulic head). This could have given some insight in which types of catchments predictions work better or worse.

For consistency, we used the same climatic/catchments indices used in the first 2 papers (from the literature review). In the revised paper we will add other measures, e.g., the drainage density, and see if they can be related with the ability of the methods in getting the signatures right.

(8) p.467, l.5: maybe better "the bars contain the interquartile range of the values..."

We changed the sentence to: "the bars contain the interquartile range (i.e., 50

(9) p.468, l.16-17: may be add Euser et al. (2012) as example of how this could be done

Citation added.

(10) Please comment in detail on the wider implications of the results: How do the results help to improve modelling strategies? Do the results contribute anything to how we can better predict in general?

As discussed above, the purposes of the paper will be clearly stated in the introduction. These questions arose because of the lack of clarity about the objectives of the paper in the earlier version. By the way, one general implication of the paper (along with the two companion ones) is that our process-based rainfall-runoff models are not good enough to estimate extremes to the same accuracy as the statistical methods can. This does not mean that process-based methods should not be used for extreme estimation, because they are suitable to address problems where causality can be explicitly accounted for (e.g., change). We will discuss this issue in the new discussion section.

(11) Figure 2: ellipse and arrows not quite clear. Please explain the connections. For example, why winter in (d) is connected with snow in (b) but not with snow in (f)?

The way is reasoned is that you have winter/summer low flows in Tyrol, which affects q95 and the shape of FDC. Of course q95 is related to FDC but not in a cause-effect sense.

References

Andreassian, V., Perrin, C., Berthet, L.,LeMoine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M., Valery, A., Crash tests for a standardized evaluation of hydrological models, Hydrol. Earth Syst. Sci. 13, 1757-1764, 2009

Andreassian, V., LeMoine, N., Perrin, C., Ramos, M., Oudin, L., Mathevet, T., Lerat, J., Berthet, L., All that glitters is not gold: the case of calibrating hydrological models, Hydrol. Process. 26, 2206-2210, 2012.

Beven, K., A manifesto for the equifinality thesis, J. Hydrol. 320, 18-36, 2006

Euser, T., Winsemius, H.C., Hrachowitz, M., Fenicia, F, Uhlenbrook, S., Savenije, H.H.G., A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci. Discuss. 9, 12989-13036, 2012

Gharari, S., Hrachowitz, M., Fenicia, F., Savenije, H.H.G., An approach to identify time consistent model parametersa: sub-period calibration, Hydrol. Earth Syst. Sci. 17, 149-161, 2013.

Klemes, V., Operational tresting of hydrological simulation models, Hydrol. Sci. J. 31, 13-24, 1986

Renno, C.D., Nobre, A.D.,Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J, Waterloo, M.J., HAND, a new terrain descriptor using SRTM-DEM: Mapping terra firme rain forest environments in Amazonia, Remote Sens. Environ. 112, 3469-3481, 2008

Ye, S., Yaeger, M.A., Coopersmith, E., Cheng, L., Sivapalan, M., Exploring the physical controls of regional patterns of flow duration curves – part 2: role of seasonality and associated process controls, Hydrol. Earth Syst. Sci. 16, 4447-4465, 2012