Hydrol. Earth Syst. Sci. Discuss., 12, C6841–C6854, 2016
www.hydrol-earth-syst-sci-discuss.net/12/C6841/2016/

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Simultaneous calibration of hydrological models in geographical space" *by* A. Bárdossy et al.

A. Bárdossy et al.

yingchun.huang@iws.uni-stuttgart.de

Received and published: 21 February 2016

We would like to sincerely thank Referee 1 for his review of the paper "Simultaneous calibration of hydrological models in geographical space". We have considered the reviewer's comments and revised our manuscript. The detailed answers to the general and specific comments are presented as below.

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

Main issues

1- The authors do not do a sufficient job putting their work in context. The literature review is outdated and not very useful in setting the paper in the current context.

Thanks for the comments. We have partly rewritten the literature review of this manuscript. The revised version contains an updated introduction, referring to the ongoing progress of the study for prediction in ungauged basins and the regional calibration.

2- The discussion lacks in depth. Results should be compared to other studies and discussed. As now, the discussion is mainly a recap of the results.

We have extended the discussion of the results in the revised version of the manuscript. We have now described in more details about the regionalization of the parameter η and its application in ungauged basins. We have also compared and discussed our study results to previous work on catchment classification and regionalization.

Also, in many places, English proofreading should be performed as some sentences are difficult to understand and interpret.

We have asked some English native speakers to help correct the grammar and improve the clarity of the sentences.

HESSD

12, C6841–C6854, 2016

Specific questions / issue

How does the loss in performance compare to other regionalization methods? Is the robustness gained worth it if many catchments offer suboptimal performance compared to a multi-donor regionalization approach?.

The performance of simultaneously calibrated model parameters is slightly worse than the individual calibration, but the transferred-simultaneous calibration is better than most of the parameter transferred from neighboring catchments. Research shows that simultaneously calibrated model parameters are more reliable than transferred model parameters from similar single catchment. As described in numerical experiment 1, the model parameters are sometimes more specific for the calibration time period and their relation to catchment properties seems to be unclear. This makes parameter transfers or parameter regionalization based on individual calibration difficult.

How does catchment similarity impact performance in calibration/validation? The paper states that the climate data dominates over catchment characteristics, but can the authors quantify the correlation or relationship to catchment descriptors?.

We applied simultaneous calibration in two different sets of catchments. Comparing the results of simultaneous calibration using 96 catchments to that obtained using 15 catchments in a relatively small region, we found that the increase of catchments considered for the simultaneous calibration led to a decrease of the model performance both in calibration and validation. For a specific ungauged basin, simultaneous calibration using a more careful selected donor catchments likely leads to good parameter transfers.

HESSD

12, C6841–C6854, 2016

Table 1: I do not feel that relative humidity is an acceptable physical catchment descriptor. Perhaps change to "physioclimatic" or something of the sorts to indicate that there is also climate data taken into account. Also, using base flow index as a descriptor while working with ungauged basins seems like it is cheating. Perhaps clearly indicate that catchment descriptors are not used for the parameter transfer. In this manner there will be no conflict..

Thanks for the suggestion, we have removed relative humidity in the revised version. We list the base flow index in the table because it is a very important hydrological property and it varies among these catchments. We have indicated that the catchment descriptors are not used for the parameter transfer in the revised version.

Introduction:
References are dating, lots of research has been done in the past few years regarding this subject.

Yes, we have partially rewritten the introductory part and updated the literature review in the revised version.

It would be nice to see a range (histogram perhaps) of the 10000 calibrated parameter sets. For example, in figures 5 and 6, the large spread of values would lead to believe that the NSE values are very heterogeneous. In figure 12, we see that NS skill ranges from 0.2 to 0.8. What would the difference be if the best (0.8 NS) parameter set was selected?.

The model calibration procedure was carried out using the ROPE algorithm (Bárdossy and Singh, 2008). This parameter optimization method could obtain a pre-determined number of parameter sets that perform very similar for the model, though these parameters are very heterogeneous. Figure 1 shows an example of the NS model perfor-

HESSD
12, C6841–C6854, 2016

mance for 10000 optimal parameter sets for both calibration and validation period using the HBV model. As almost all these parameter sets perform well during calibration, we used all of them for model validation and parameter transfer to other catchments. The mean of the model performances of these 10000 parameter sets represents the simulation results. For the results in Figure 12, each catchment have 10000 simulated NS performance. These 10000 parameter sets result in a small range of NS values for a specific catchment, but the average NS value for different catchments are very different, which ranges from 0.2 to 0.8. We will list the feasible range of the model performances for all calibrated parameters as supplement.

11226 Lines 22-23 : Missing "is" 11226 Lines 22-23 : Missing "is"
11227 Line 9 : Make a single sentence out of the two.
Landscapes are formed during long time through climate, and are thus in a kind of quasi equilibrium.

Thank you for these detailed suggestions and corrections. They have be incorporated in the revised version.

How about 2 very different catchments? Do you expect water dynamics to be similar for a steep catchment vs a flat catchment? Must there not be a preprocessing of similarity index for the catchments? While at it, why not go one step further and do physical similarity regionalization?

We are agree with you that for very different catchments the dynamic responses might be different and that the simultaneous calibration of these dissimilar catchments may lead to poor model performances. As suggested in the conclusion, the selection of the donor catchments for simultaneous calibration is important for the application in ungauged basins and actually we estimated the similarity between different catchments

HESSD
12, C6841–C6854, 2016

at the very beginning of this study, and we then attempted to validate the similarity measurements using hydrological models. However, the transferred results indicate the asymmetry of the parameter transfer matrices that is mainly due to the different climate conditions.

11229

I do not understand this part of the sentence (totally > 1010-year discharge calculations)

It represents the huge number of calibrations we have tested in our study. We have removed this sentence in the revised paper.

11233

Line 14: "...This is necessary as it is thought to establish correct water balances". But what do you make of equifinality? Surely this equation will produce different n values depending on the calibration parameter set.

The η value is estimated through the simulation procedure. For different dynamic parameters, the simulated discharge may be different and it leads to different η values.

11235

How do you compute the long-term water balance if the catchment is ungauged? The way I see it, there are two options. Either the n parameter is adjusted based on the actual gauged data (biasing the results since the parameter set received will need to conform to the n parameter) or there is another way to estimate the value of n at an ungauged site, namely using other regionalization techniques. It is imperative that this be discussed beforehand.

HESSD

12, C6841–C6854, 2016

This is a part of the conclusion in the discussion section. The η is a water balance related parameter but still independent of the dynamic parameters and thus in the numerical experiments the long term discharge volumes were treated as known variables for both gauged and ungauged catchments. And instead of regionalizing η for ungauged basins, we suggested for the regionalization of discharge coefficients which relate discharge volumes to precipitations. At the end, we found that the discharge coefficients show a smooth spatial behavior in the study area and the regionalization of this parameter does not seem to be difficult. Afterwards, the long term discharge volumes may be calculated and the parameter η could be estimated for each common parameter set.

11236

Can you explain the differences observed? What happens when the "good basin" parameters are transferred to the "bad basin" and that the modelling fails? What do you observe in the hydrograph? Why is this not seen in the reverse order?

In model calibration procedure, we always adjust the hydrological model according to the observations and the climate conditions during the calibration period have high influence on the model parameters. For a catchment without enough information about the flood events or extremely dry conditions during calibration period, the model is still possible to achieve high performance values. But model validation for very different climate or transfer to some other catchments is problematic. Here we took the bad receiver which is catchment 12 and calibrated it by HBV model using NS model performance as an example. From the observation data, we found that catchment 12 is under relatively dry climate conditions during the calibration and validation time periods. Figure 2 shows a small part of the runoff hydrographs obtained using individual calibration and transfer parameter set from catchment 1. From the simulated hydrographs, we can see clearly that the parameter set calibrated on catchment

HESSD

12, C6841–C6854, 2016

1 could not capture the dynamic behavior of catchment 12 as the low flows were underestimated for most of the time and the peak flows were obviously overestimated.

11241
I do not understand the sentence: "This is as expected that there is less common behavior of a large set of catchments as for a few"

By comparing the model performance of simultaneous calibration of 15 catchments and 96 catchments, we found that the common parameter sets calibrated by 15 catchments in a reasonable geographic proximity perform better than the parameter sets calibrated by the 96 catchments.

11242
"But for the Rottweil catchment, model performance is worse than for the Fils catchment. It indicates that there is some skill in the transferred parameters, but the differences are substantial. Figures 15 and 16 show part of the observed and the modeled hydrographs using the NS performance measure. We can see the transfer is reasonable and the dynamics of the discharge are similar to the US case. This experiment demonstrated that even very distant and different catchments may behave similarly.
// **Not sure that this is what is implied from the text. The second sentence says that the differences are substantial, whereas the last sentence says that the catchments may behave similarly. Also note the strong underestimation of peak flows.**

Thanks for the comments. We have received multiple comments and suggestions about this section. We are also the idea that only two supporting examples seem not sufficient to discuss the parameter transfer to other continents and thus we think

HESSD
12, C6841–C6854, 2016

the numerical experiment 4 and the associated result should be removed in the revised version of our manuscript.

11246 - Conclusion

Lines 7-11 : I do not agree with this assessment. Are the authors implying that very different catchments (mountainous vs flat, forest vs grasslands, difference in lithography and geology, etc.) react the same to similar rainfall? Could it simply be that by selecting the lowest common "acceptable" parameter set, the method neglects key differences, thus skewing the results towards this conclusion? More details are needed to justify this point.

We have rewritten the conclusion: In this study, three lumped hydrological models with three different performance measures were tested on the daily time scale. The results show that many catchments behave similar as the same dynamical parameter sets could perform reasonable for all of them. This means that hydrological behavior on the daily scale is dominated by precipitation characteristics and actual evapotranspiration. In our study area, it also indicates that the differences in catchment properties cannot be captured well by simple lumped model parameters.
We concluded that many catchments share common parameters which describe their dynamical behavior does not mean that they have the same dynamical behavior. The model output still highly depends on the water balance parameter η. The common parameters performs well for catchments with different characteristics in daily reactions. We are also interested in the hourly reactions of the common parameters, but increased errors will make the task more difficult.

Discussion: The discussion must be improved significantly and expanded:

Discussion of the results has been extended in the revised version. We also compared our study results with the usual classification and regionalization methods.

HESSD
12, C6841–C6854, 2016

9.1 -> Does this "deepest parameter set" have stronger ties to physical catchment descriptors than other parameter sets?

The deepest parameter set gives a better and more structure combination of parameters.. But no clear relationship is shown between the depth of parameter set and the catchment descriptors.

9.2 -> It is critical that the authors discuss the estimation of n at ungauged sites. How is the parameter estimated if there is no streamflow? Does it use observed streamflow to estimate properly and then only the dynamics parameters are fitted? If so, how does conditioning the dynamic parameters to the n parameter impact the result? What if we use a "bad" n?
9.3 -> Ok, place 9.3 before 9.2 or talk about this point much earlier. It is absolutely critical for understanding the paper. Also, if the n parameter is easily regionalized through space based on proximity, why not use the spatial proximity regionalization method for the other parameters? One can also combine spatial proximity and physical similarity with multiple donors to improve performance, such as described in Oudin et al. 2008 and applied in Zhang and Chiew 2009; Arsenault and Brissette 2014; Zelelew and Alfredsen 2014, etc.

Thanks very much for the suggestions. We have rewritten these sections to make the manuscript more understandable. This study only need to regionalize parameter η, it simplifies the application for prediction in ungauged basins. Oudin et al. (2010) compared two different versions of similarity: the apparent similarity defined on the basis of observable catchment properties, and behavioral similarity judged through the use of hydrological models. Their result shows that the overlap between the two kinds of similarity is significant for only 60% of the catchments. As shown in numerical experiment 2, the model parameters are often overestimated due to climate condition during calibration period. Therefore, the transferability of dynamic parameters cali-

HESSD
12, C6841–C6854, 2016

brated on a number of catchments is more robust than the parameter sets calibrated on individual catchment.

Speaking of which, the authors should point out explicitly how regional calibration instead of direct regionalization, based on past results (Parajka et al. 2007; Ricard et al 2013, Gaborit et al. 2015) which discuss regional calibration and its strengths/weaknesses.

Thanks for the suggestions, we have added the discussion in the revised paper. More present results about regional calibration have been addressed in the revised paper.

How would traditional regionalization methods fare if allowed the advantage of forcing the "n" parameter as in this case?

Compared with traditional regionalization method, we found out that simultaneous calibration of catchments only need the regionalization one single parameter that control the long-term water balance, and that the results are relatively reliable.

In my opinion, the discussion needs to be improved substantially and should have references to the current state-of-the-art to better relate the results in this paper to the literature.

Discussion of the results have been extended in the revised version. The study results were compared to the usual classification and regionalization methods.

HESSD
12, C6841–C6854, 2016

References

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrol. Earth Syst. Sci., 12, 1273–1283, doi:10.5194/hess-12-1273-2008, 2008.

Oudin, L., Kay, A., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, Water Resour. Res., 46, W11558, doi:10.1029/2009WR008887, 2010.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 11223, 2015.

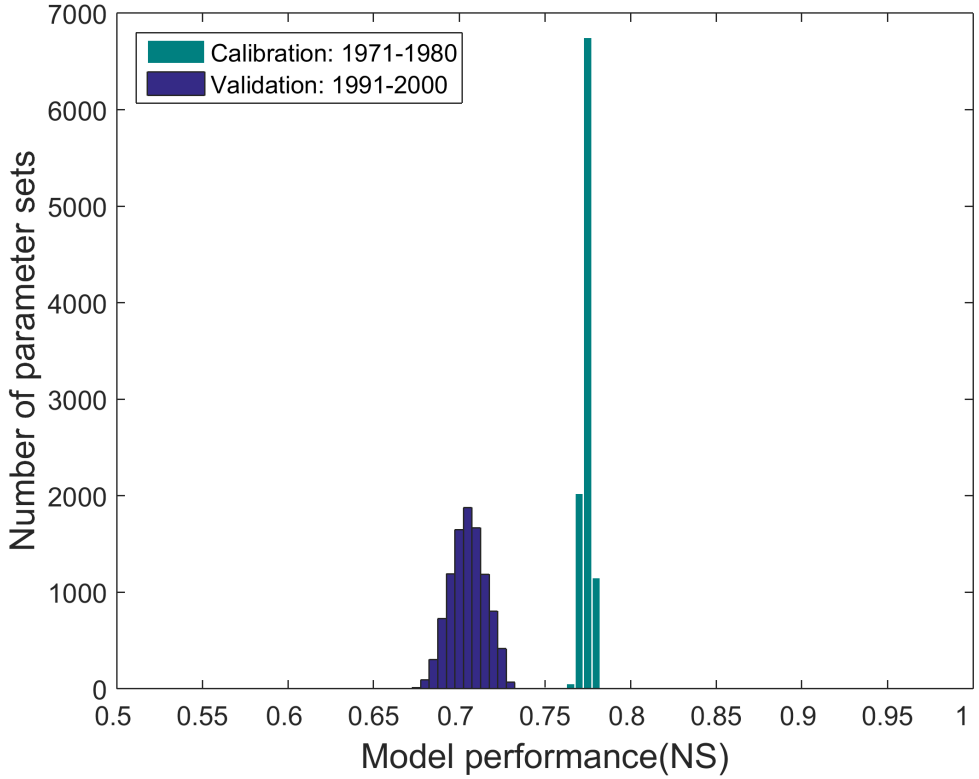

Fig. 1. NS model performance for catchment 01548500 using HBV model.

HESSD
12, C6841–C6854, 2016

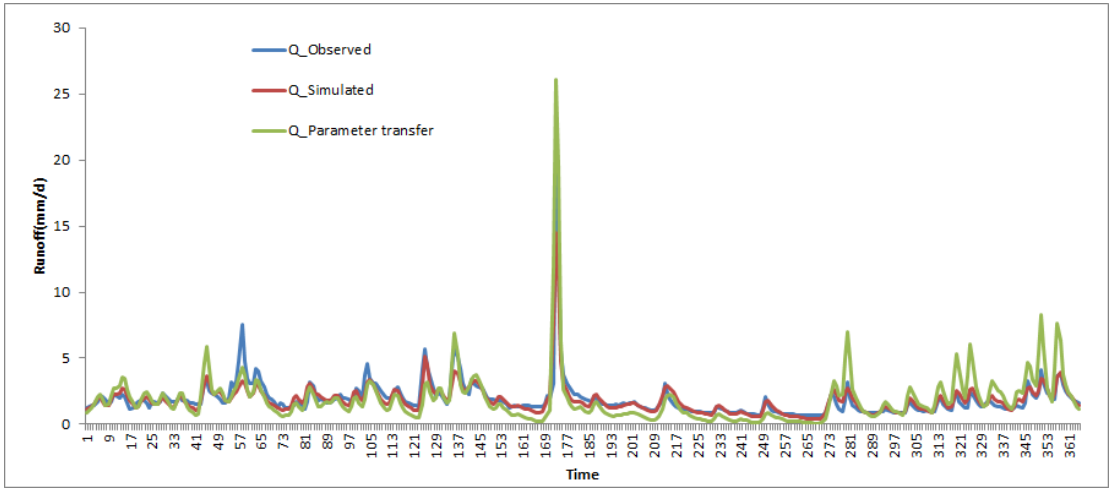

Fig. 2. Runoff hydrographs for catchment 12.

HESSD
12, C6841–C6854, 2016