Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-508-RC2, 2016

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Interactive comment on "Towards simplification of hydrologic modeling: identification of dominant processes" *by* S. L. Markstrom et al.

B. Guse (Referee)

bguse@hydrology.uni-kiel.de

Received and published: 23 February 2016

This article applies a parameter sensitivity analysis using the FAST algorithm on different model outputs of a distributed-parameter hydrology model (DPHM). It focuses on presenting spatial patterns of parameter sensitivities for the different model outputs and extracts the dominant processes for United States of America.

I really like the core idea of this study and think this manuscript is worth to be published. However, in my opinion, the article might benefit from a couple of improvements to emphasize the major outcomes more precisely. I also think that the interpretation and discussion of the results could be more clearly to make the study interesting for a broader audience. At some points, I presented some ideas which might worth discussing about it. Thus, I encourage the authors to consider the following remarks.

Major comments:

I encourage the authors to improve the readibility of the abstract to present the idea of this study in a clearer way.

Please think about the use of the notation "objective function" for mean, CV,... . In my understanding, these are statistical values describing different model outputs without giving information of the model performance. The use of the term "objective function" indicates an evaluation of the model performance according its common use in hydrological modelling. I propose to use "fundamental daily streamflow statistics (FDSS)" as mentioned in the text instead of "objective function".

A table with the model parameters and their corresponding processes is missing. I see that you refer to another article. However, this manuscript would be more readable, if the reader has an idea of the parameter used for this study. When stating that a certain number of parameters is required "to account for 90% of the parameter sensitivity" is necessary to know how many parameters for this process are included in the model structure. For example, assuming that there are only two snow parameters, then it is not surprising when the number of required parameters is two. However, let's say that are eight parameters for the snow process then it is interesting to know that only two parameters are required.

Furthermore, in chapter 4.2, you should mention whether the parameters (accounting for 90%) are identical for a certain process or vary (P. 10, L.5-6).

It is really interesting to see a systematic in the number of parameters as stated on P. 10, L.20-23. Could you explain it? At best in relation to the model structure? Are you expect a different result for different models (structures)?. While this result is reasonable for snowmelt, it is really surprising that you only need a small number of parameters to explain the soil moisture behaviour.

I think that the article would benefit if you could relate the results (e.g. P.10, L.24-30) to the process heterogenity in the different parts of the CONUS. There are certainly regions with very complex process patterns and other with a clear dominance of a single process. Are there other studies looking at process dominance or process heterogeneity in the CONUS? Maybe you can make a comparison with these studies?

It is certainly required to discuss the relationship of model parameters and the corresponding processes. The stronger this relationship is, the more sensitive a parameter might be for this process. Could you mention how the parameter-process relationship affect your results?

By summing up the first-order partial variance and using this value as indicator to estimate the dominant process, you do not consider the parameter interactions (second and higher order sensitivities). However, the parameter interaction depends (among others) on the parameter selection. Could you explain how this aspect affect you results?

The interpretation of table 1 needs to be reworked. I do not agree at least with the sentence on P. 11, L.16-18 that a count of dominant parameters shows how important a parameter is. Assuming that a parameter is strongly related to a certain process, e.g. snowmelt, and is thus relevant for the three objective functions related to snowmelt, but not to the other processes (maybe except of runoff), it is still an important parameter for this specific process. This interpretation and also of the fig. 5 aggregates the results in my opinion in a strong way. It might be more interesting to look at the relationship of model parameters to the processes. To how many processes you can related a parameter? Are these results reasonable when looking at the model structure? An idea of how to relate model parameters and corresponding processes is given in the figures and tables in Pfannerstill et al. (2015).

Concerning the discussion of the spatial heterogeneity in parameter sensitivity (subchapter 5.1), it might worth looking at the expert knowledge on dominant processes in the CONUS. It is not surprising when a HRU with a complex hydrological situation with

relevant contributions from different runoff components provides a different results as a HRU with a strong dominance of one hydrological component. Here, I think that a general discussion of process dominance is missing and a discussion in the context of former studies on dominant processes in the CONUS (if existing).

Maybe you can think about presenting the results in Tab. 1 and Figs. 4 and 5 in a different way, so that the most important outputs are more emphasized. It is rather difficult to extract information of the relationship of parameter and processes from Tab. 1 and a counting how often a parameter occurs is also time-consuming. But in my opinion this information is required to make Fig. 5 more informative.

Fig. 4: Is it maybe relevant thinking about the variability, e.g. in the snowmelt subplot? It is stated that on average 2.25 parameters are required to explain 90%. The map (subplot 4M) shows that in most of the HRUs only 2 or 3 parameters are required. However in the snow-dominated northern parts up to 10 parameters are required. It might be worth thinking about extracting additional information from this idea. One way would be to add an additional line in the subplots 4A-4H which is only related to HRUs which have certain relevance of this process (kind of threshold exceedance approach or something similar).

Fig. 6: Could you explain why infiltration is the inferior process in many HRUs. I cannot imagine a hydrological situation in which the infiltration process is less relevant than total runoff, all runoff components, ETP, soil moisture.

It might be interesting to think about the following results of the Fig 4-5: According to Fig. 4 only 4.15 parameters are required to explain soil moisture, which is a relative low value keeping in mind that the soil moisture interacts with almost all other processes. Furthermore, there are 7.05 parameters needed for infiltration. Then, it is stated in Fig. 5 that soil_moist_max is overall the most important parameter. Do this mean that the relationship between soil_moist_max and soil moisture is extremely high so that only a few additional parameters (about 3) are needed to reproduce the soil moisture conditions?

Minor comments:

Abstract:

Page 2, Line 2: The first sentence of the abstract could be written more clearly. Why not only writing: "The Precipitation-Runoff Modeling System as a distributed-parameter hydrologic model has been applied to the conterminous United States.

P. 2, L. 4-5: Whilst it is certainly clear that the number of parameters is an aspect of model complexity, this is not fully clear for the "interpretation of the model output". Is this really an aspect of complexity? Do you assume that the model which provides a higher number of model outputs is more complex?

P. 2, L. 5-8: To make the abstract more readable, I would suggest to subdivide this sentence into two separate ones. There are too many aspects in this sentence (parameter sensitivity for simplification, parameter identification and its relationship to dominant processes, spatial patterns)

P. 2, L. 9-10: I do not think that this sentence is understandable when reading the abstract at first before knowning the whole article. What do you mean with "processes correspond to variables"? Which type of variables?

P. 2, L. 11: The notation "categories" is not clearly described in the abstract.

P. 2, L. 12-13: How do you estimate the "model performance" by visualizing categories? This part needs to be improved.

P. 2, L. 16: The benefit of a reduction of the dimensionality of output variables or objective functions is not clear.

P. 2, L. 22: I would encourage the authors to add a final sentence to emphasise the general advantage of this study.

Introduction:

P. 2, L. 28: The article would be benefit from a clear definition of "input parameters". Is an input parameter related to a driver of the hydrologic cycle such as precipitation or solar radiation or more to a real model parameter? In all cases, it is better to avoid potential misunderstandings.

P. 3, L. 1: References are missing such as for constraining parameter in models, e.g. Hrachowitz et al. (2014) and for stating that different parameter good have a comparable impacts on the model results.

P. 3, L. 6: The three references are related to studies which investigate performance measures more precisely. It might be good to also have a reference to studies which are directly investigating the model output.

P. 3, L. 11-12: Please also add the study from Reusser et al. (2009).

P. 3, L. 14: Please indicate that you consider uncertainty in this study only on input parameter uncertainty and not on structural uncertainty in the model.

P. 3, L. 18-28: It might be good to mention here that it is at least at this scale impossible to support the results with adequate measurements in addition to the total discharge.

P. 4, L. 1: References are here missing, e.g. Wagener et al. (2003), Reusser et al. (2011), Guse et al. (2014).

P. 4, L. 11: Reference of Reusser et al. (2011) is missing.

P. 4, L. 20-22: As mentioned before, it is not clear why you aimed "to reduce the number of inputs and outputs". I think the overall aim should be a clearer characterization of the model parameters and to focus on the dominant processes.

Methods:

P. 4, L.29- P. 6, L.7: Please check carefully if you could reduce the subchapter 2.1 in length. Do you really need this information for this article?

P. 6, L.8-25: The selection of the eight output variables is reasonable and seems to be representative for hydrological studies with distributed models. Maybe you can emphasize this to give the article a more general character.

P. 7., L. 18: Please also add the reference of Guse et al., 2014, since it is the initial study for Pfannerstill et al. 2015.

Results:

P. 8, L. 17: Please think about a more precise title for the subchapter 4.1.

P. 8, L. 20-23: This sentence is not understandable. It is understandable that you have calculated the sum of the first-order partial variance. However, it is not clear how you can estimate an average value (average of what?).

P. 8, L. 23: The total sensitivity is one, is it? Why do you need to scale the sum of the sensitivities to the total sensitivity?

P. 8, L. 23: "category of modeled process“ instead of "category of process“.

P. 8, L.28-30: I recommend to be more precisely here: You have calculated the sum of all partial sensitivities for a certain HRU for each process. Then, the process with the highest sum of the first-order sensitivity is indicated as "dominant process“. To make this clear, you should add that the dominant process is the process with the largest sum of all first-order partial variances (sensitivities). This is required since the sensitivity of a single parameter is not shown here.

P. 9, L.17-18: Can you extract a systematic pattern in these results?

P. 10, L.24-25: Please add that this statement is not valid (or only to a low extent) to fig 4J and 4N.

P. 11, L. 6-9: Do you see a general systematic why the spatial patterns of parameter

sensitivity are different for the different objective functions. It might be interesting to give further statements on this.

P.11, L. 28-32: When stating that the parameter "soil_moist_max“ is the most important and a model calibration should be focused on it, then it is required to know for which process this parameter is relevant. Assuming that a typical calibration uses discharge as target variable, a focus on "soil_moist_max“ helpful in the case of a dominance of "soil_moist_max“ on runoff. However, to include this information in a calibration in the case of a dominance on other process but not on runoff?

P. 12, L.2-8: The part on the least sensitive parameter can be removed since the reader does not receive any details about the parameters. Or could you extract some further information from the fact that these parameters have a low sensitivity?

P. 12, L. 9-14: I think that the authors should add here some more details. It is really helpful if a parameter can be precisely characterized by saying that it is only dominant in a very specific case (e.g. for one process). But this information cannot currently not be extracted from article.

P. 13, L.8-12: I like this part. Maybe you can in addition relate it to the concept of vertical water redistribution (Yilmaz et al., 2008, Pfannerstill et al., 2015).

P. 14, L. 22-23, Step 1: Summed in time?

P. 14, L. 24-25, Step 2: How to you obtain a score for each process? Do you assign each parameter to a certain process? If yes, then you have to mention somewhere which parameter is related to which process.

P. 16, L. 31: Spelling error: Mishra (2009)

Figures:

Fig. 1: Could be removed. I do not see an advantage of it. Maybe you can transfer it to the supplementary material.

Fig. 2: Does the last row and column present the average values along the row/column? Do you maybe have to change "process average“ and "objective function average“?

I recommend to show the figure 3 before the figure 2, since fig. 3 provide a general map of the USA whilst, fig. 2 already show the distributed results.

Figure 4 would benefit from knowning which parameters are within the 90% and how variable the parameters belonging to this 90% are?

Fig. 4: The legend needs to be graphically improved.

I do not really see a real benefit of fig. 5. Maybe you can extract the results in a better way. One point might be that the model parameters are not explained and even the related processes are not highlighted in Fig. 5. In particular, it is not clear which information you can derive from the last place occurrence.

It is not fully clear which information you can derived from investigating the most inferior process. It seems to be that this is either clear such as snowmelt parameter for California or related to the model structure.

Reference list:

Guse, B., Reusser, D. E., and Fohrer, N.: How to improve the representation of hydrological processes in SWAT for a lowland catchment – Temporal analysis of parameter sensitivity and model performance, Hydrol. Process., 28, 2651–2670, doi:10.1002/hyp.9777, 2014.

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

Pfannerstill, M., Guse, B., Reusser, D., and Fohrer, N.: Process verification of a hy-

drological model using a temporal parameter sensitivity analysis. Hydrology and Earth System Sciences 19: 4365–4376, 2015.

Reusser, D. E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, Hydrol. Earth Syst. Sci., 13, 999–1018, doi:10.5194/hess-13-999-2009, 2009.

Reusser, D.E., and Zehe, E.: Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. Water Resources Research 47(7): W07550. DOI:10.1029/2010WR009946, 2011.

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V.: Towards reduced uncertainty in conceptual rainfall–runoff modelling: dynamic identifiability analysis. Hydrological Processes 17: 455–476, 2003.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007WR006716, 2008.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2015-508, 2016.