Supplement of Hydrol. Earth Syst. Sci. Discuss., 12, 13301–13358, 2015
http://www.hydrol-earth-syst-sci-discuss.net/12/13301/2015/
doi:10.5194/hessd-12-13301-2015-supplement

Hydrology and Earth System Sciences
Open Access
Discussions

Supplement of

The importance of topography controlled sub-grid process heterogeneity in distributed hydrological models

R. C. Nijzink et al.

Correspondence to: R. C. Nijzink (r.c.nijzink@tudelft.nl)

Ranked Probability Score

The Ranked Probability Score (S_{RP}; Wilks, 2005) was adapted as a measure for the magnitude of the expected model improvement or deterioration. Originally, S_{RP} was designed to estimate the "distance" between an observation and an empirical cumulative distribution function, based on the area enclosed between the two (Figure 6). The Ranked Probability Score is given by:

$$S_{RP} = \frac{1}{M-1}\sum_{m=1}^{M}\left[\left(\sum_{k=1}^{m} p_k\right) - \left(\sum_{k=1}^{m} o_k\right)\right]^2, \quad (8)$$

where M is the length of the distribution of performances of a certain signature, p_k the probability of a certain signature performance to occur and o_k the probability of the observation to occur. In our case o_k is a step function, which is either 1 or 0. For example, E_{NS} has it's optimal value at 1. Thus, as there is only one time series, it has a step distribution function at 1. The model runs will have a cumulative distribution function, as multiple sets of parameters are considered as feasible. This distribution function will be close to 1 in case of model with a relatively good performance. The difference of the S_{RP} between two models was used here as a measure to identify and quantify improvement.

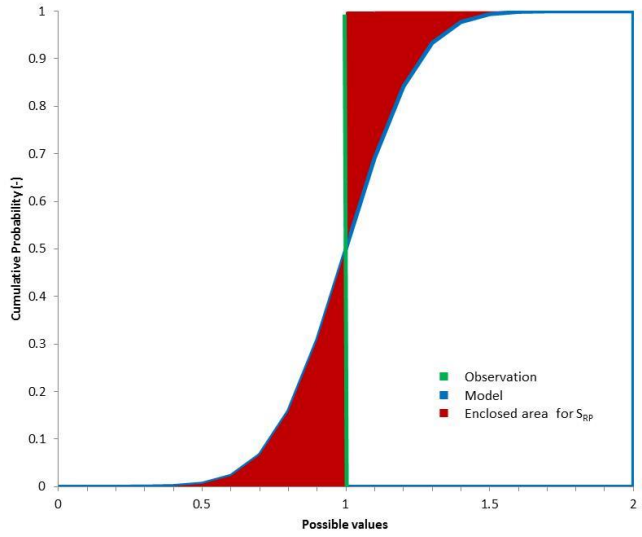


Figure S1. Graphical illustration of the ranked probability score S_{RP}. The enclosed area (red) between model (blue) and observation (green) determines the score.

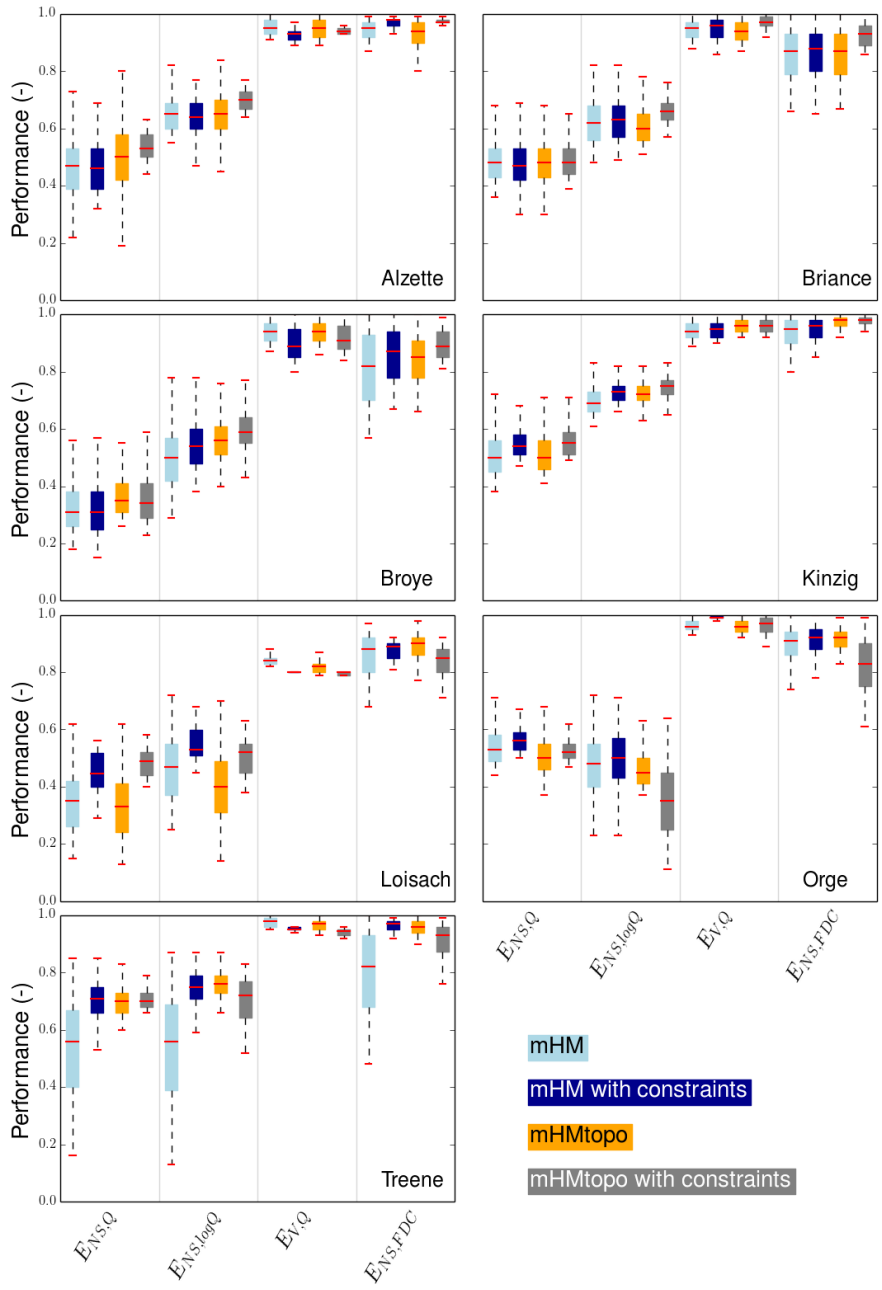
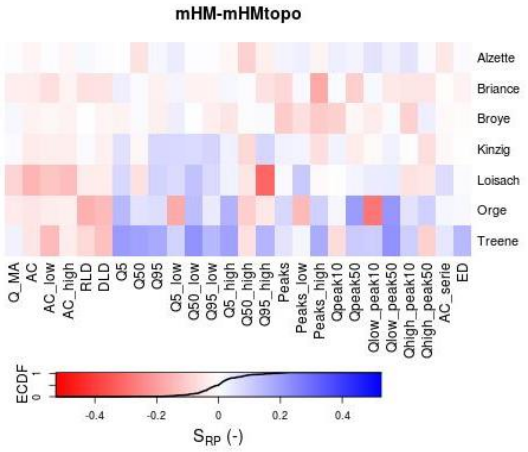
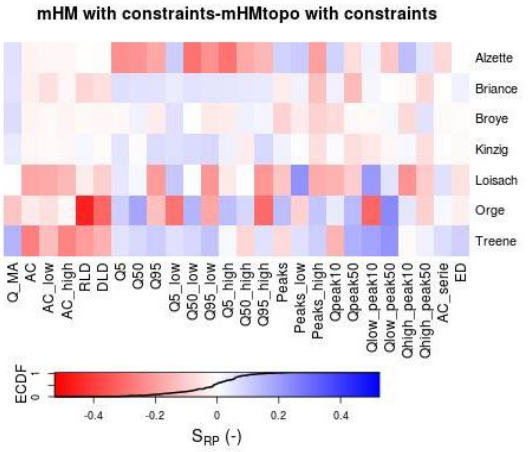


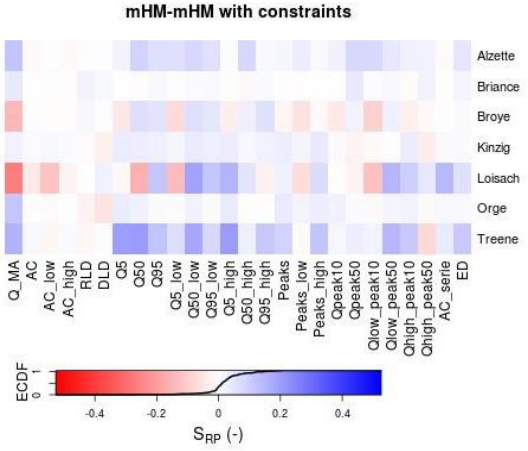
Figure S2. Nash-Sutcliffe efficiency ($E_{NS,Q}$), log Nash-Sutcliffe efficiency ($E_{NS,logQ}$), volume error ($E_{V,Q}$) and log Nash-Sutcliffe efficiency of the flow duration curve ($E_{NS,FDC}$) for the seven catchments in the calibration periods. The optimal value for all four criteria is 1, whereas 0 is regarded to have a low performance. The boxplots are formed by the Pareto space spanned by the four objective functions.



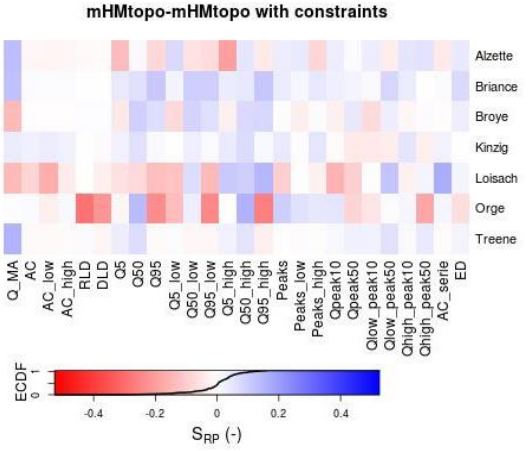
1 a)



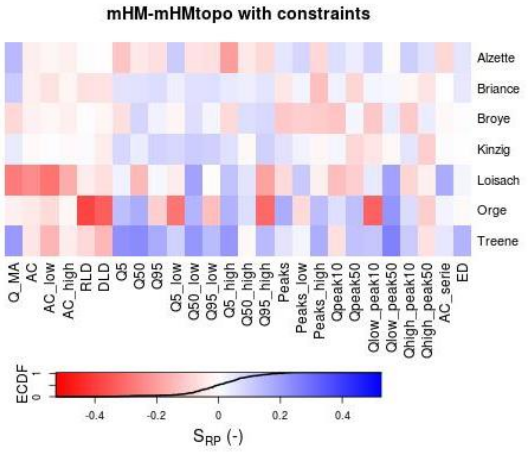
b)



2 c)



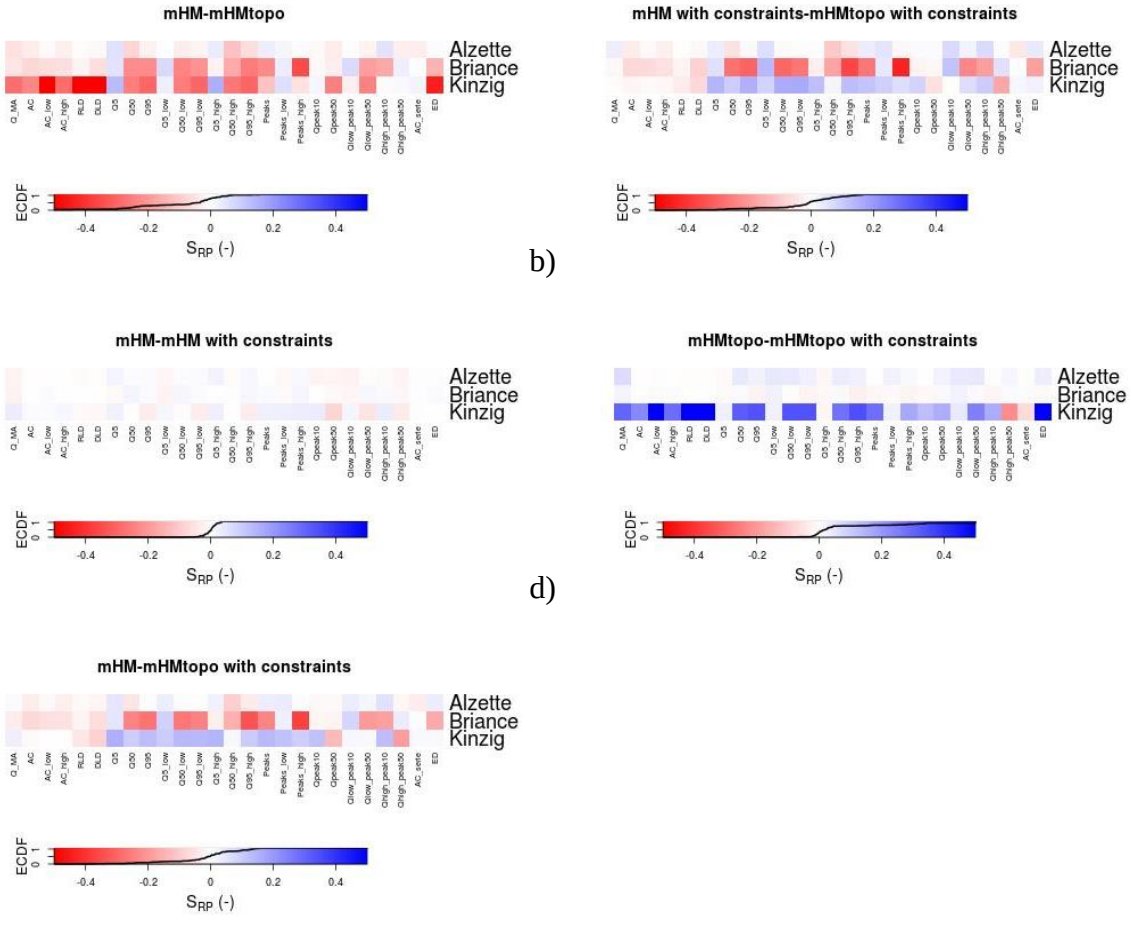
d)



3 e)

4 Figure S3. Difference in Ranked Probability Scores between (a) mHM and mHMtopo without
5 constraints and (b) with constraints, (c) mHM with and without constraints, (d) mHMtopo
6 with and without constraints (e) the base case mHM with the constrained mHMtopo case. The
7 colours are linearly related to scores between the most negative values (darkred), 0 (white)
8 and the most positive values (darkblue), where positive values indicate an improvement. An

empirical cumulative distribution function based on all values has been added to assess the distribution of occurring score differences.



a) b)

c) d)

e)

Figure S4. Difference in Ranked Probability Scores between (a) mHM and mHMtopo without constraints and (b) with constraints , (c) mHM with and without constraints, (d) mHMtopo with and without constraints and (e) the base case mHM with the constrained mHMtopo after the transfer of global parameters. The colours are linearly related to scores between the most negative values (darkred), 0 (white) and the most positive values (darkblue), where positive values indicate an improvement. An empirical cumulative distribution function based on all values has been added to assess the distribution of occurring score differences.