Response to Reviewer Comment #3 (and Erratum to Response to Comment #2)

Thank you for your comments and constructive suggestions. In your comments you point out three major aspects: first the readability, second the depiction of the algorithm and third the model application.

Your first point has been stated by all reviewers so far and we truly accept this critique. As already mentioned in our previous statements a linguistic revision is in progress. Results of this revision will be uploaded separately, after incorporation of all other changes.

The second point intersects with the readability. You suggest a more technical, mathematical description of the algorithm. In the former submission of this manuscript we focused on a technical description of the algorithm and its functionality (but based on the reviews we decided to withdraw the manuscript and focus more on modelling application). A major point of the revisions for the previous manuscript was a too technical depiction of the algorithm that did not match the requirements of a research paper. Hence, we omitted the detailed description and offered a "verbose" qualitative description. We added the appendix to show the functionality of the involved tools. In its actual form, the manuscript is intended to offer a trade-off between technical detail and facile qualitative description.

Nevertheless, the sequence of the algorithm is shown and described in the manuscript. For readers with more interest in the algorithm and its implementation we made the code available as supplementary to this paper.

Besides the structure of the paper, you pointed out that the background of the threshold Ω and the non-linearity coefficient e is not made clear.

As described in Sect. 3.2.1 the threshold Ω is used to distinguish between distance classes comprising "high" and "low" variance. Since we do not know the exact amount of tolerable variance we had to come up with a concept to value our target. There are several ways to do so. One way might be to just take an arbitrary percentile of present variance, a valuation by cluster analysis or by setting it to a fixed number.

But the threshold, and hence the number of ascertained sub-basins, is dependent on the purpose of the performed partition. For some applications a detailed subdivision, bringing the heterogeneity to a minimum (will result in more computational effort), might be desirable while for other purposes a facile partition might be sufficient. We intended to implement a flexible function that unites the requirements for both purposes (and in between).

The weighting factor ω offers this kind of flexibility and e is its parameter. If e tends to zero, the threshold Ω tends to the average of present standard deviation. The factor e set to this value should be used in applications were only regions with significantly higher heterogeneity are to be captured. If e increases the threshold lowers and hence the number of subdivisions is likely to increase. With an increasing factor e more heterogeneity is captured by the algorithm.

Though it is interesting which value for e is suitable for what kind of application we omitted deliberately a detailed analysis. In the present state of the manuscript e is a user-dependent specification and its introduction aims to lower the range of possible arbitrary threshold values. In the following figure Fig. RC1 we compared the value of Ω to percentiles of present standard deviation values in the Mulde catchment. As it can be seen we approximately bisect the range of possible values. All figures in the manuscript involving Ω use three different values of e, namely 0., 0.5 and 1. Our intention was to show a possible range of Ω in our application, though we constantly set $e = 0.5$. All this is indeed missing in the manuscript because we ascribed only minor importance on this factor.

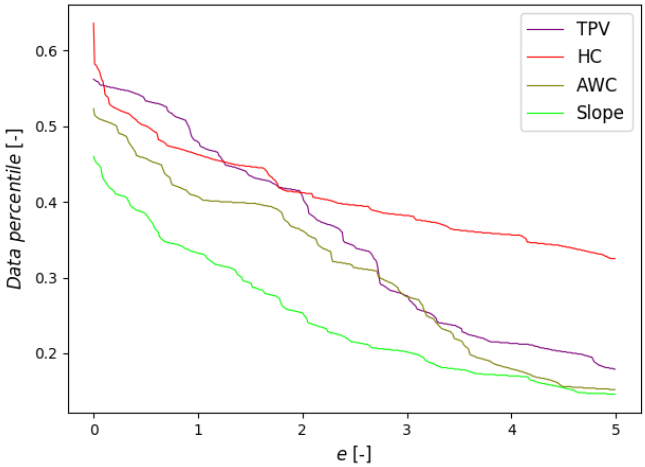

Figure RC1: Percentile of *Ω* against non-linearity factor *e*

To make the concept of the non-linearity factor *e*, its chosen value in the application and its constant depiction in the distance-factor functions understandable we added the following lines on page 7, line 8:

> "Dependent on the purpose it is recommended to ~~apply the proposed algorithm multiply in the process of subdivision for a stepwise reduction of Ω and hence the remaining variance.~~ vary the value of the non-linearity factor e in a range of [0; ∞], though values between 0 and 1 approved applicability in our case studies. If e is set to zero, Ω is equal to the average of σ(C) in the basin. As e is increased Ω lowers. The choice of e is dependent on the intention of the purpose of partition. If a detailed subdivision is required to capture the majority of heterogeneity, e should be set to a value greater or equal to 1. Otherwise, if only regions with a significantly higher heterogeneity are to be captured, e is recommended to be set to 0. However, in the presented case studies e has been set to 0.5. To show the range of other possibilities Ω is shown in all distance-factor functions for values of e ~ 0., 0.5 and 1.0.
>
> ~~Note that each application will lead to a higher number of subdivisions.~~"

To your last sub-point of your critique on the depiction of the algorithm we would like to give a detailed response. The revision of mathematical equations has been pursued from the response to Comment #2 as follows:

We introduced a set *B* of stream-flow distances x_S to the description of the distance-factor function (Page 5, Line 25):

$$B = \left[i \cdot \Delta s; (i+1) \cdot \Delta s \right] \tag{1}$$

The variable *i* is an integer ranging from 1 to N_S, the number of distance classes in the basin. Please note that this equation differs from its first appearance in the Response to Comment 2. The Equation had not been compiled properly since a "Δ" in the upper boundary had been missing.

This set is used afterwards in the equations for *E(C)* and *σ(C)*:

$$E(C)_i = \frac{1}{w(i \cdot \Delta s)} \sum_{(j | x_{s;j} \in B)} C_j \tag{2}$$

$$\sigma(C)_i = \sqrt{\frac{1}{w(i \cdot \Delta s)} \sum_{(j | x_{s;j} \in B)} \left(C_j - E(C)_i \right)^2} \tag{3}$$

Equation 3 (now 4), the beforehand discussed threshold Ω, had also not been compiled correctly. As you pointed out the factor *e* has to appear in the nominator and denominator:

$$\Omega = \frac{\sum_{i=1}^{N_S} \omega_i^{\,e} \cdot \sigma(C)_i}{\sum_{i=1}^{N_S} \omega_i^{\,e}} \tag{4}$$

Furthermore, the weighting factor Eq. 4 (now 5) on page 6, line 25, has been changed to make the range of Min/Max Function clear:

$$\omega_i = \frac{\sigma(C)_i - \max\limits_{1\le i\le N_S}\left(\sigma(C)_i\right)}{\min\limits_{1\le i\le N_S}\left(\sigma(C)_i\right) - \max\limits_{1\le i\le N_S}\left(\sigma(C)_i\right)} \tag{5}$$

The objective function Eq. 5 (now 6) has been updated analoguesly:

$$Z = \left|\left[i \mid 0 \le i \le N_S ; \sigma(C)_{\mathrm{i}} > \Omega\right]\right| \rightarrow \min \tag{6}$$

On page 9, line 9, Eq.6 (now 7) the evaluation scheme has been updated as well. We added a new index *p* to denote parallel sub-basins (Note that this is also a change to the response to comment #2. We changed the index character to prevent a mistake with the set B (Eq. 1)). Index *p* is an integer ranging from 1 to P_i, the number of parallel sub-basins within the original basin *i* ($i \in \left[0; N_S\right]$). Remaining indices have been matched with the general notations:

$$\sigma_S(C)_i = \frac{1}{\sum_{p=1}^{P_i} w(i\cdot\Delta s)_p} \cdot \sum_{p=1}^{P_i} \sigma(C)_{i;p} \cdot w(i\cdot\Delta s)_p \tag{7}$$

Your last comment addresses the model application. You suggest to substitute the NSE-based calibration study with a parameter uncertainty analysis or to omit the model application.

Likewise to your critique of the manuscript structure and algorithm depiction we shared your opinion. Moreover the previous submission of the paper changed our attitude. While the detailed technical description was heavily criticised in the previous manuscript, the missing model application "where a model with the new subdivision outperforms another model" was strongly recommended. So (again) this chapter has been incorporated to meet such expectations. A reduction to the variance topic would be contrary to the aims of this manuscript, defined in the introduction (and title), that clearly sets for modelling purposes. Since we expected that many readers would ask for that missing modelling aspect we would like to keep this section in the manuscript.

Nevertheless, our main intention is to present the method, its ability to capture catchment heterogeneity and the insights on catchment organisation. This is why we only performed a short and simple modelling study with 3 (now 5, see Response to reviewer comment #1) calibrated model setups.

The problem with NSE-based comparisons is indeed their dependence on the applied calibration procedure. Therefore, we tried to apply the same calibration strategy to both spatial setups, titled "coupled" calibration for ACS and the benchmark setup. In both applications the same 6 parameters are coupled and the identical calibration tool and the same number of iterations has been used. The only thing changed is the coupling-information, the land cover in the benchmark setup and TPV-values in the ACS setup. This procedure is described on page 16 from line 9 to 26.

Additionally, we allowed the benchmark setup to exploit its full number of parameters in an unconstraint calibration run, involving the same calibration tool and number of iterations (as described from line 27 to 29). As reviewer #1 pointed out we had a parameter mismatch in this calibration study and we therefore introduced a second parameter coupling applied to all zonal-parameters. The results and changes in manuscript can be found in our response to the first reviewer.

Your suggestion on parameter uncertainty seems quite appealing, though. Therefore, we performed a Monte-Carlo Simulation with 10'000 random parameterisations and saved the resulting NSE and parameters for several sub-basin in the Mulde catchment. The Monte-Carlo simulation has been applied to the all-parameter coupling scheme. We evaluated our results scantly with a GLUE approach. We used the NSE as generalised likelihood function and set a threshold of 1% for behavioural parameter sets. Parameter uncertainty has been calculated with the following equation:

$$U(Par) = \sqrt{\frac{\sum_{x \in Nx} \left(Par_x - \bar{P} \right)^2 \cdot NSE_x}{\sum_{x \in Nx} NSE_x}}$$

where *Par* is the analysed parameter, *Nx* is the behavioural parameter set, NSE_x is the generalised likelihood of the parameter set and $\bar{P}$ the average parameter in set *Nx*:

$$\bar{P} = \frac{1}{|Nx|} \cdot \sum_{x \in Nx} Par_x$$

To compare the uncertainty of all parameters their actual values have been normalised by their parameter constraints. In Tab. RC2 all parameters are given with their lower and upper boundaries, grouped by their usage in the model. "Single" parameters are used in the ground water levels and post-processing of the model, "Coupled" parameters are used in each individual zone. Since we applied the all-parameter coupling scheme, these parameters are estimated once per sub-basin and by a linear coupling coefficient transferred to all zones (See response to Comment #1). The coupling parameters are not shown in Tab. RC2 but their values are uniformly in the interval [0, 2].

Table RC2: Parameter constraints HBV$_{96}$ model

Parameter	Lower Boundary	Upper Boundary
Single		
ALPHA	0	2
K4	0.001	0.5
KHQ	0.01	0.5
PERCMAX	0	5
MAXBAZ	0	1
Coupled		
BETA	1	4
CFLUX	0	2
CFMAX	1	6
CFR	0	2
DTTM	-2	2
ECORR	0.95	1.05
EPF	0	0.2
ERED	0	0.5
ETF	0	0.2
FC	100	500
ICMAX	1	20
LP	0.75	1
RFCF	0.95	1.05
SFCF	0.95	1.05
TT	-2	2
TTINT	1	5

WHC 0 2

Following figures RC3-6 we plot $\overline{P}$ in the normalised parameter space for ACS setup and the benchmark setup for a single sub-basin, where the shown error bars represent calculated *U(Par)* values.

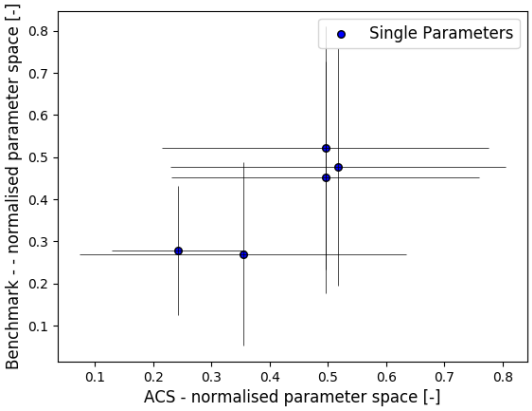

Figure RC3: Single parameters

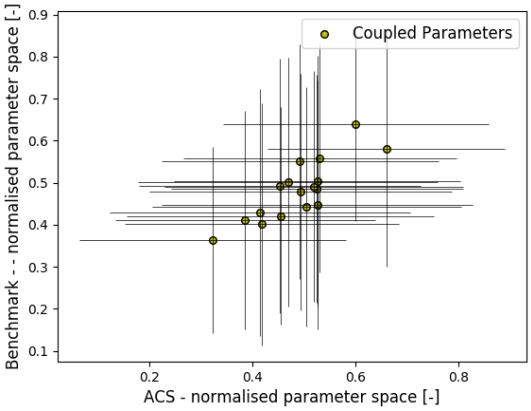

Figure RC4: Coupled parameters

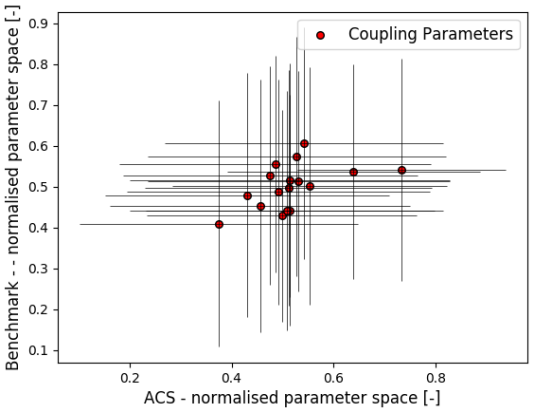

Figure RC5: Coupling parameters

As it can be seen in the figures above the uncertainty indicators are symmetrical for the most parameters, meaning that the uncertainty is equal for both models. Uncertainty values for the benchmark and ACS setup plotted against each other are given in the following Fig. RC6:

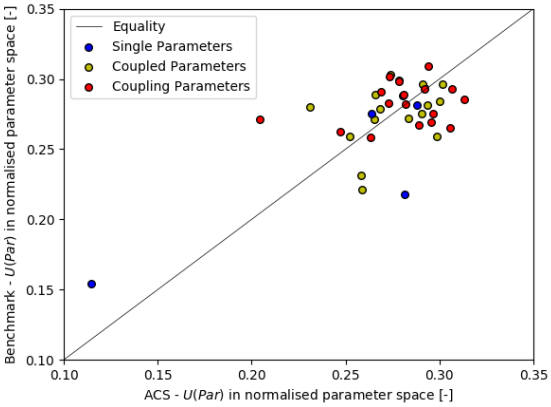

Figure RC6: Parameter Uncertainties

In Fig. RC6 several points are located above the “Equality” line, meaning that parameter uncertainty for an individual parameter is higher for the benchmark model than for ACS-model. But nearly the same number is below the line.

These results indicate that the parameter uncertainty could not be lowered (significantly). But this is only a fast and simple evaluation and its result might change with an elaborate technique.

Considering the length of the manuscript we would prefer not to include these results (or similar) in the paper, because its explanation (method, model, parameters, etc.) would effort a lot of additional information that are irrelevant for the ACS algorithm.