Responses to J. Seibert

First of all, we thanks for your thoughtful comments. The responses to the comments are listed point by point as following:

Major comments

1. How are physical characteristics estimated at the grid scale?

As I understand one grid cell is about 30 x 30 km2. How do you get values for Ksat and other soil characteristics at this scale? Several previous studies have shown that measured point values do not agree with effective values at larger scales. Therefore I would argue that the model parameter values you estimate at best might be indicative, but not really physically based. Furthermore, the average subgrid slope seems, at this scale, a poor proxy for the hydraulic gradient (note also that the subgrid slope depends largely on the resolution of the DEM being used!). Please describe in more detail which data sources were used and how as well as discuss the limitations.

The description of data sources would be added to section 4.1 study area and dataset. The soil physical characteristics were estimated by the relationships between soil characteristics (Ksat, etc.) values and each soil type.

1) The soil data was extracted from the FAO two-layer 5-minute 16-category global soil texture maps. In this datasets, the soil was classified into 16 categories, the first 12 kinds of which were used in this study. The resolution of the soil data was 5-minute, but it was 30-second in USA.

2) The relationships between soil characteristics (Ksat, etc.) values and each soil type were referenced to Rawls et al. (1998). Page 7025, line 16. The resolution of the grid cell in this study was $0.25^{o}\times0.25^{o}$. Thereby, the Ksat value in each grid cell could be averaged by the Ksat value with the resolution of 5-minute.

3) The land cover data was obtained from the University of Maryland's 1km Global Land Cover data. There were 14 kinds of land cover types. The resolution was 30 seconds.

4) The DEM data was obtained from SRTM 90m Digital Elevation Data. Based on the comments 2, the slope data would not be used.

Because the soil and DEM data were global, this framework for baseflow estimation could be used widely. But the relationships between soil characteristics (Ksat, etc.) value and each soil type might not be adequate accurate. This could be revised by the improvement of agrology.

2. Equation 8 is not correct

While the units might look ok, I do not think the equation is correct. In the Darcy law you need to multiply by the area through which the flow occurs. This should here be the width of a grid cell multiplied by the depth of the saturated zone (i.e. area in the xzdirection). The reason Dm has the unit length per time, as the right side of the equation has, is that the water flow is seen over the area of the grid cell in xy directions.

We agreed with you. D_m was the daily maximum subsurface flow, which occurred when the third soil layer moisture was saturated. By the Darcy's Law, when the third soil layer was saturated, the base flow, Q_b, could be calculated by:

$$Q_b = -K_s A \frac{\partial \psi(z)}{\partial z} \quad (1)$$

where,

$$A = w_x \times w_y \,, \quad \frac{\partial \psi(z)}{\partial z} = \frac{\psi_2 - \psi_1}{h} = -1 \quad \text{(Fig. R1)} \quad (2)$$

where, A was the area through which the flow occurred; w_x and w_y (m) were the width of the grid cell in the x and y direction, respectively. Therefore, D_m could be calculated as:

$$D_m = \frac{Q_b}{A} = K_s \quad (3)$$

The Eq. 3 would be used in the revised study.

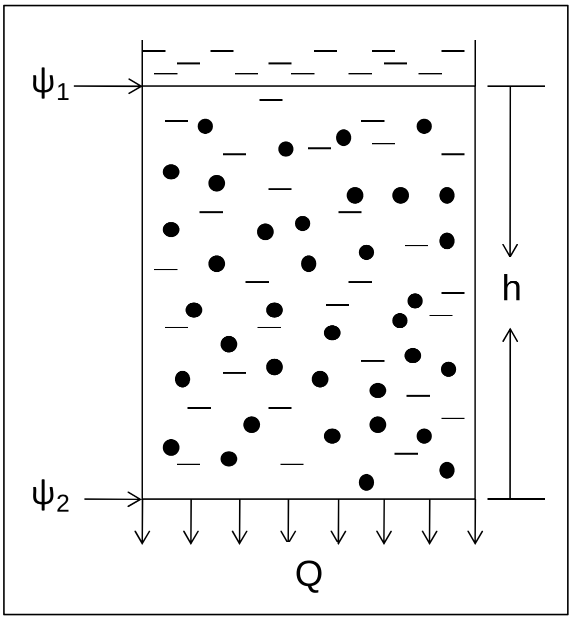

Fig. R1. The maximum baseflow generation in the VIC model.

3. Test of model performance

Given that there is a considerable equifinality it is not surprising that model performances are about the same when the three less sensitive parameters were not calibrated but fixed before the calibration. This is because you can find good fits all over the parameter space. It is also expectable that parameter uncertainties decrease once you have fixed some parameters. The interesting question is more whether fixing these parameter values in the way presented here is any better than fixing the values to other values. This could be addressed by using some average values for all 24 catchments (=1 test) and/or using parameter values derived for another (i.e., wrong) catchment (=23 tests). It would also be interesting to see whether the parameter sets which partly have been derived based on other data than runoff would perform better or worse during conditions outside the calibration conditions (see Seibert, 2003).

The framework presented in this study was for predictions in ungauged basins. During calibration in gauged basins, the highest value of *Nsc* could be got no matter nearly whatever values of the three baseflow parameters were set (Fig. 5, page 7045-7047). That was because of the equifinality. Therefore, it would be inaccurate for transferring baseflow parameters from gauged basins to ungauged basins. The test of model performance would be illustrated in another paper. With the framework in this paper, the baseflow parameters would directly estimated in each catchment without calibration. Regarding one catchment as "ungauged catchment", the regression relationships between other three calibrated parameters and the basic catchment characteristics would be constructed in other 23 catchments. The other three parameters in this "ungauged catchment" could be estimated by the regression relationships and the basic catchment characteristics. Then the model performance could be tested by comparing to the results under calibrated conditions.

Minor comments

1. I find the structure of the manuscript rather confusing. Some of the methods are mentioned first in the result section and a discussion section is missing. In the end I would prefer conclusions rather than a summary. The manuscript would also benefit from improving the language.

The methods would be introduced in section 2 and 3. The discussion part would be added. The last part would be revised as a conclusion

2. The Mnc criteria (Eq 6) is a suitable approach to ensure a better fit in terms of the water balance than using just Nash-Sutcliffe. However, I wonder whether the volume error does not become to dominating. It certainly does if Re is used as defined in Eq 5 (multiplied by 100%), but even otherwise. Lindström et al. found a weighing factor of 0.1 to provide best results.

There was not a widely used criteria considering both *Nsc* and *Re*. Despite in Eq 5 *Re* was multiplied by 100%, but in Eq 6, it was still used as decimal fraction and would not become to dominating. For example, if in one result, the *Nsc*=0.9, *Re*=15%, in another result, *Nsc*=0.89, *Re*=6%, it might regarded as the second result was better than the first result. The *Mnc* would be 0.875 and 0.915 in the first and second result, respectively; the *RV* (Lindström et al., 1997) would be 0.885 and 0.884 in the first and second result, respectively. Therefore, in this section, the *Mnc* might be better than *RV*.

3. Several times the term sub-catchments is used. As (most or even all?) of the catchments are not nested I would recommend to just use catchment Please provide the units of all variables and parameters used in the manuscript. This is not the case at the moment.

"sub-catchment" would be replaced by "catchment". The units of the variables and parameters would be added.

4. P 7023, 19. Please provide rather a real reference than a website link.

The VIC model was detailed described in Liang et al., 1994, 1996; Liang and Xie, 2001 (P 7018 line 20). The web link in P 7023, line 19 was the official website of the VIC model.

5. What do you mean by 'bulk water'? (p7025, 19)

"bulk water" (P7025, 19) meant gravitational water, i.e., free water.

6. In Eq 9 W is used with different subscripts and the different W have different units, which is confusing.

W_f and W_m in Eq 9 would be replaced by other letters.

7. P 7029, 21/22. Do not provide too many digits.

The digit in P 7029 line 21/22 would be reduced.

8. Table 1: for three parameters for the unit N/A is written, should be [-]

N/A for three parameters in Table 1 would be replaced as [-]

9. Figure 11: What is the unit on the y-axis

In fig. 11, the y-axis indicated the *H* value calculated by Eq 7 (P 7024 line 24). Therefore the unit would be [-].