Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-67-RC2, 2019


Hydrology and Earth System Sciences
Discussions

# *Interactive comment on* "Calibration event selection for green urban drainage modelling" *by* Ico Broekhuizen et al.

**Anonymous Referee #2**

Received and published: 18 April 2019

General comments

This manuscript presents an analysis of the impact of selecting different sets of calibration data for the SWMM urban hydrological model. Selection is based on a variety of hydro-meteorological characteristics of the available storm events. In addition, the calibration is performed either adjusting all calibration parameters simultaneously, or at two stages where parameters related to pervious and impervious areas are calibrated separately. Finally, the results are analyzed against a backdrop of other sources of uncertainty besides the calibration dataset.

The idea of calibrating impervious area parameters separately using such data where the role of previous areas is presumably insignificant is promising, and in my opinion the results related to this represent the most valuable contribution of the present



manuscript. On the other hand, I struggle to find a novel scientific contribution in the analysis of the calibration event selection in combination with other causes of uncertainty. As argued in the specific comments below the results are inconclusive and it is hard to find any other take-home message than the fact that selection of calibration data has an impact on model parameter values and model performance. This has been established already in existing hydrological literature, as acknowledged also by the authors themselves.

The readability and the quality of the English language are at a very good level.

Specific comments

Study site and data

It would be useful to show somewhere a brief summary of the storm events (e.g. duration, cumulative rainfall depth, cumulative runoff, peak runoff, runoff percentage). The runoff percentage in particular would be interesting as it is used in selecting events for the two-stage calibration. Also, it would be interesting to see to which extend the permeable areas are activated during more intensive events (i.e. runoff-% > 12%).

Event selection

To me the most promising aspect of this manuscript lies in the idea of calibrating parameters related to pervious and impervious areas separately. It is obvious that with a greater runoff percentage than 12% other than just directly connected areas need to contribute. For events with less than 12% runoff it is not equally evident that ONLY directly connected areas contribute. Still, this is a feasible assumption and probably holds to a sufficient extent. There is ample evidence that in urban setting for small events (directly connected) impervious areas predominantly contribute to stormwater flow and for major events also permeable areas are activated.

A couple of issues require further clarification. Did you check whether in the model any runoff was generated from permeable areas when the runoff-% was below 12%? If it is argued that no runoff is produced outside of the (directly connected) impervious areas for low runoff-% events it should be checked that the model result is consistent with this assumption. Second, the large range of rainfall multipliers (0.48 – 2.92) can make determining the runoff-% somewhat ambiguous. Presumably, the 12% runoff threshold was based on the measured values of precipitation and discharge before applying the rainfall multipliers. Did it happen that a smaller than the unity rainfall multiplier changed the initially below 12% runoff event to exceed the 12% threshold after rainfall multiplier calibration? If yes, should such an event be included in the first stage calibration?

Other sources of uncertainty

The reasoning in including some of the uncertainty sources while leaving others out is not quite clear to me. Also, the take-home what readers should learn from this exercise should be clarified.

Rainfall input

The authors report that reducing flow measurements by 40% leads to 37% reduction in the mean value of rainfall multipliers, and increasing flow measurements by 40% results in a 33% increase in the rainfall multiplier mean value. This seems like rather a trivial result. A more justified description about the purpose of scaling the discharge by a constant multiplier, which causes a corresponding change in the rainfall depth scaling parameter, is needed.

Calibration data measurement uncertainties

See comment above.

Conceptualization / model discretization

While I agree that SWMM is a well established model for urban drainage I do not think that its applicability to areas clearly dominated by pervious areas is equally evident. Presumably in the SWMM runs of the current manuscript the groundwater module has been turned off and infiltration is based on the Green-Ampt equation with infiltration



continuing with a rate appraoching assymptotically the hydraulic conductivity value. It can be questioned whether this is realistic for longer storm events when the soil becomes more saturated. Transpiration is also not accounted for but evaporation only occurs from the depression storage. I am not suggesting that it would feasible to take into account all aspects related to modelling uncertainty. But in my mind the authors' statement "...it is safe to assume that the SWMM conceptualization is appropriate for urban drainage modelling and there was no need to consider this issue further" is in the context of such a low density urban area questionable and does not constitute a valid argument for making a choice about which uncertainty sources are included/excluded in/from the analysis.

Calibration algorithm

The authors state that SCE-UA "...has been widely applied in hydrological applications with great success, so there was no need to subject it to scrutiny in this paper." While I agree that SCE-UA is a powerful tool with an extensive pool of hydrological modelling applications, it is not a sound, objective argument for leaving it out of study. The authors themselves admit that calibration against RMSE can yield a higher NSE than calibration against NSE itself, indicating that the algorithm does not always converge to the optimum value.

Validation performance

Validation performance should be the main argument for improved calibration strategy. If a calibration strategy leads to improved parameter identifiability this should be visible in better results against independent validation data. The authors state that "the two calibration strategies that performed best in the validation period were two-stage strategies" and "...calibrating impermeable and green area parameters in two separate steps may improve the model performance in the validation period...". I think that currently the results about the validation performance for one-stage and two-stage calibrations are inconclusive. The authors use the sum of ranks from several performance criteria as a proxy for overall performance. Are the results shown in any Table? If yes, I missed them. Also, I would prefer a more quantitative statistic than a sum of class variables (ranks). As NSE is used as the objective criterion for the baseline calibrations it would be a logical choice also for comparing the validation performance. The authors state in Section 3.5. about the validation performance "In terms of NSE, the single-stage calibrations performed better...". On the other hand the 'NSE joint' criterion, typically used for validation (performance over the entire validation data set), seems to be higher for two-stage strategies in Table 6. It is hard for the reader to find guidance here what would be the preferred calibration strategy.

Recommendation

In its current form the manuscript is not in my mind publishable in HESS. The following major changes would be required:

A more informative description of the hydrometeorological data to allow the readers to understand differences between different calibrations

A better justified reasoning for inclusion/exclusion of different error sources

Most importantly, a clear statement about the scientific novelty value of the manuscript where it becomes obvious what are the new findings over just showing that different calibration data lead to different model parameter values and validation performance

Technical comments

Mostly technical comments. The comment for Figure 4 also relates to the content of the manuscript.

Figure 1. Remove the text below the figure (1 map catchment.png). Increase the font size/figure resolution. The legend is hard to read.

Figure 2. Remove the text below the figure (2 example hydrographs run130.pdf).

Figure 3. Remove the text below the figure (3 VE PFR histograms.pdf). In the figure



caption it is stated peak flow ratios to be on the left whereas in the figure the left panel shows the volume error. Please correct.

Figure 4. Remove the text below the figure.

It is hard to interpret with the given information what is causing the negative NSE for the right panel. Is there a timing difference invisible to the eye? Why does the modelled flow stay at zero for the beginning of the event? Clearly there is rain (left panel), so is the diminished rainfall multiplier and/or increased depression storage value causing all rain falling on the directly connected impervious area to be captured in the depression storage?

Figure 5, 6, 7, 8. Remove the text below the figure.

Table 11. Mistake in the NSE single-stage value for D_prec (0.41)? The corresponding value in Table 6 is 0.43?

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-67, 2019.