

Interactive comment on “Calibration event selection for green urban drainage modelling” by Ico Broekhuizen et al.

Ico Broekhuizen et al.

ico.broekhuizen@ltu.se

Received and published: 1 June 2019

We sincerely thank the referee for their extensive comments on the manuscript, which we reply to point-by-point below. The referee's comments have been numbered (intentionally starting at 10) for easy reference.

General comments

(10) Referee's comment: 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

C1

impervious areas are calibrated separately. Finally, the results are analyzed against a backdrop of other sources of uncertainty besides the calibration dataset.

Authors' response: this summarizes well the contents of the manuscript.

Changes in manuscript: -

(11) Referee's comment: 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.

Authors' response: this is addressed in our response to points 17-21, 24, 28 and a newly added emphasis on innovative aspects of our manuscript.

Changes in manuscript: see below (points 17-21, 24, 28)

(12) Referee's comment: The readability and the quality of the English language are at a very good level.

Authors' response: we thank the referee for their supportive comment.

Changes in manuscript: -

Specific comments

C2

Study site and data

(13) Referee's comment: 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 extent the permeable areas are activated during more intensive events (i.e. runoff-% > 12%).

Authors' response: we agree that a table summarizing rainfall-runoff events could be useful to the reader. The table can also contain a rough estimate of how many mm runoff was generated by the green areas. The extent to which green areas are activated can be estimated in a limited way from the data directly (see last column in table C1 in the supplement).

Changes in manuscript: such a table will be added in the methods section, see Table C1 in the supplement.

Event selection

(14) Referee's comment: 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.

Authors' response: it can indeed be the case that green areas contribute some runoff even when the percentage runoff is less than 12%. Even impervious surfaces will not generate 100% runoff, so if runoff is exactly 12% it is reasonable to expect that at least

C3

a small part of runoff has come from green areas instead. We agree with the referee that the amount of runoff from green areas is small enough that assuming it zero is a feasible assumption. In any case, it would be difficult to determine by how much the 12% threshold should be lowered to ensure that no green area runoff is included, since this would also depend on the antecedent conditions in the catchment. Given a lack of other measurements (e.g. soil moisture, standing water in swales) in the catchment it is not possible to tell the initial wetness of the catchment from measurements. Estimating initial conditions using the model itself would lead to the undesirable situation where the value of the threshold (and therefore potentially the set of events to use) would be different for each model run. A fixed percentage is therefore much more workable and probably of more practical use.

Changes in manuscript: add the following sentence in section 2.3 on line 11: "(It is conceivable that there is some contribution of green areas when the percentage runoff is less than 12%, and in that case the threshold should be set at a lower value, but since the amount of green area runoff and the appropriate value of the threshold would be highly dependent on antecedent conditions this was not included here.)"

(15) Referee's comment: 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.

Authors' response: there are several items to check here:

First, during the first stage calibration (i.e. with default values for green area parameters) there was no runoff from green areas for any of the calibration events in any of the calibration scenarios, and so the first stage calibration attributed all runoff to impervious areas.

C4

Second, using the calibrated parameter values for both impervious and green areas, there were some first-stage events where some runoff was predicted from green areas:

1. When runoff was disabled from both directly and indirectly connected impervious areas /by setting their depression storage to 1000 mm) there were three calibrated models runs (2 for T32S_D_prec, 1 for T32s_Q_60m) that actually generated some runoff from green areas (i.e. the runoff did not originate on impervious areas draining to green areas), but since this was $\leq 2\%$ of the total simulated runoff volume this was considered negligible.

2. When runoff was disabled only for directly connected impervious areas, a total of 12 calibrated model runs showed non-zero runoff from green areas. This was $<5\%$ of total simulated runoff volume for 4 runs, $<10\%$ for an additional 3 runs, and 11.6%, 11.7%, 21.7%, 22.9% and 25.7% respectively for the remaining 5 runs. However, almost all of this was runoff that was generated on impervious areas draining onto green areas (see point 1 above).

Regarding the last mentioned 5 runs, it should be noted that these concerned 3 different events with a percentage runoff between 11% and 12%. Such events may be expected to include some green area runoff and it could be considered to exclude these from the first stage calibration as discussed in comment 14. In addition, all three events were also included in other first-stage calibrations that did not result in any significant simulated green area runoff (0, 0 and 3.4% of total simulated runoff respectively). Removing these events from the first stage of calibration based on initial calibration results would therefore result in the same event being included in different stages for different calibration scenarios, which we considered undesirable.

Overall we believe that, although the assumption that all runoff is from directly connected impervious areas when $QV_{ppP} < 12\%$ is violated in some cases, the assumption that these events are suitable for calibrating impervious area parameters does hold to a sufficient degree, as also evidenced by the good first-stage calibration per-

C5

formance (mentioned on p 10, l. 2-3). In addition, checking for green area runoff as done here is only possible after calibration, and taking it into account when selecting events would thus create a more complex, iterative calibration procedure which limits the practical applicability of the approach. We considered this to be beyond the paper's original scope of examining different strategies for calibration event selection. It could however be considered as a potential avenue for further research on multi-stage calibration procedures.

Changes in manuscript: add a (shorter) version of our response above on page 10, l. 3.

(16) Referee's comment: 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?

Authors' response: the 12% runoff threshold was indeed applied directly to the measured values of precipitation and discharge.

There were two events where the rainfall multiplier was less than 1 and reduced rainfall so that the new percentage runoff exceeded 12%. This can also be displayed in an extended version of Table 4 from the manuscript, see Table C2 in the supplement. It is of course possible to exclude such events from their respective stages in the calibration and replace them with another event. Being consistent about considering the percentage runoff as calculated using the calibrated rainfall multipliers would also require the following three adjustments as well:

C6

1. It would have to be applied 'in both directions', i.e. second-stage calibration events where the calibrated multiplier was large enough that runoff % was reduced below 12% would have to be excluded from the second stage. (This was the case for two events. For these events they would first have to be considered as replacement for a first-stage event, and the first stage calibration re-run, before redoing the second stage of the calibration. (Depending on the results from this the whole procedure might have to be repeated as well.)

2. All event characteristics related to rainfall (i.e. P_sum, PI_mean, PI_30m, QV_ppP) would have to be re-calculated and the related CSs determined and run again if the event set changed.

3. Out of the 32 events that were available for use in calibration scenarios, only 22 were actually selected by one or more CSs, so calibrated multipliers are not available for the other 10 events. It would be necessary to somehow obtain a calibrated multiplier value for them too so that they may be reconsidered for use in the calibration.

Although this might improve the overall results of the proposed calibration procedure, it would also increase the complexity and raise several new issues, such as how to obtain a calibrated rainfall multiplier for the 10 events that have not yet been used. We considered this to be beyond the paper's original scope of examining different strategies for calibration event selection. It could however be considered as a potential avenue for further research on multi-stage calibration procedures.

Changes in manuscript: clarify that event selections were fixed beforehand and not adjusted based on initial calibration results. Add a short version of the explanation above in section 2.3, page 6, end of line 18.

Other sources of uncertainty

(17) Referee's comment: 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

C7

should learn from this exercise should be clarified.

Authors' response: some of the issues described have been investigated before for urban drainage models (e.g. data uncertainties by Kleidorfer et al. (2009) and Dotto et al. (2014), model resolution by e.g. Krebs et al. (2014), Petrucci and Bonhomme (2014), Sun et al. (2014) and Tscheikner-Gratl et al. (2016)). The idea behind including other sources of uncertainty was (primarily) to see if different calibration event sets showed different sensitivity to these issues and (secondarily) to see if the findings also applied to a different data set and catchment (more dominated by green areas).

Although we considered it an interesting experiment at the time, the impact of what objective function is used in calibration of urban drainage models has not been investigated extensively before (Barco et al. (2008) made some short remarks), so we would remove this from the manuscript. (A thorough investigation of this would be an interesting topic for a different study.) However the different objective functions used for validation phase (e.g. volume error, peak flow) would still be included since they provide additional insight into the simulation results.

Removing the parts on objective function would also allow to describe in more detail the effect of the model resolution, since it is an interesting finding that some of the benefits of the two-stage calibration (better flow volume and peak flow in validation phase) are stronger for the low-resolution model.

The take-home messages from this are:

1. The impact of perturbed calibration data impact appears small (confirming findings by Dotto et al. (2014)), but we do see interaction between calibration data selection and the model discretization.

2. The two-stage calibration gives better results in terms of flow volume and peak flow in the validation phase, and this effect is much stronger for the low-resolution models.

Changes in manuscript:

C8

1. Section 2.4 (other sources of uncertainty):

1a. Describe aim of including other sources: i.e. check if earlier findings are sensitive to different calibration data sets and if they also apply for a different data set and a greener catchment.

1b. Add references to previous studies on rainfall input uncertainty effect on urban drainage modelling in lines 10-12 (Dotto et al., 2014; Kleidorfer et al., 2009).

1c. Lines 20-24: describe a bit more the Dotto and Kleidorfer papers that are referred to, including that they used more pipe-based drainage systems and a fixed set of events.

1d. Lines 28 and further: add references to articles dealing with model resolution (Krebs et al., 2014; Petrucci and Bonhomme, 2014; Sun et al., 2014; Tscheikner-Gratl et al., 2016)

2. Remove parts that deal with the calibration using RMSE as alternative objective function:

2a. page 7 lines 25-27

2b. page 9 lines 2-5

2c. Section 3.1.2, including table 3.

2d. Table 5 column 3: "RMSE as obj. func." + update column "total"

2e. Section 3.3.3, including table 7 and figure 8.

2f. Conclusion page 24 lines 10-11

2g. mention in abstract

Rainfall input

(18) Referee's comment: The authors report that reducing flow measurements by 40%

C9

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.

Authors' response: since flow data is obviously an important part in the calibration process, we wanted to see if earlier results from Dotto (2014) and Kleidorfer (2009) would be sensitive to different sets of calibration events. Our findings mainly confirm their work. For urban catchments these issues have only been investigated to a limited extent (i.e. with a single set of events and for rather impervious catchments) so additional support of earlier findings is useful. Other disturbances of / errors in the calibration are conceivable, but were deemed beyond the scope of this study. In addition, the correlation between the adjustment in rainfall and the adjustment in rainfall multipliers also supports the idea that the rainfall multipliers are compensating (even in the baseline run) for a mismatch between observed and best-fitting rainfall (as discussed in section 3.2.2), and therefore that they are a suitable way of accounting for this mismatch. A better description of why this aspect is considered is also addressed in our response to comment 17 above.

Changes in manuscript: see above.

Calibration data measurement uncertainties

(19) Referee's comment: See comment above.

Authors' response: See response above.

Changes in manuscript: See response above.

Conceptualization / model discretization

(20) Referee's comment: 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

C10

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 approaching asymptotically 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 be 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.

Authors' response: we appreciate the distinction between the application of SWMM to pervious and impervious areas. It is correct that the groundwater module in SWMM was not utilized in this study, and that therefore only the Green-Ampt equation + drying is used to account for infiltration. As pointed out by the referee, recovery of the infiltration capacity is not based on evapotranspiration, but is instead based on the soil's saturated hydraulic conductivity (Rossman and Huber, 2016).

Our original formulation was perhaps too optimistic, but we still believe that it is reasonable not to treat model structure as an uncertainty source in this article for the following reasons:

1. Unlike input and calibration data and model resolution, model structure uncertainty has not been addressed extensively in the urban drainage modelling literature.
2. There is a lack of methods for considering model structure uncertainty other than using different models, which is outside the scope of this study. The catchment and the high-resolution model also require certain features (e.g. routing runoff from one subcatchment to another subcatchment, support for automated runs) that are present

C11

in SWMM and not in other models. Model runtime is also a limiting factor.

3. The Green-Ampt method itself has been in use for many years. Other infiltration models are available (e.g. Horton or SCS curve number in SWMM) but going into these would be outside the scope of this study. Ideally a study on infiltration models in urban drainage modelling would also make use of infiltration and/or soil moisture measurements which are not available here.

Changes in manuscript: Replace p8, lines 5-7 with: Although model structure is also a recognized source of uncertainty (Deletic et al., 2012), it was not considered here since (a) there is a lack of previous research on this topic for urban drainage modelling that could be referred to and (b) there is a lack of methods to address this other than using different models in parallel, which was considered outside the scope of this study, and would in any case be difficult since the catchment model requires some SWMM features (e.g. routing runoff from one subcatchment to another, good support for automated runs) which are not always present in other models.

Calibration algorithm

(21) Referee's comment: 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.

Authors' response: in relation to the improved description of why different sources of uncertainty are included it's good to mention that (like for objective functions) there is a lack of studies examining the effect of calibration algorithms on urban drainage modelling (Deletic et al., 2012). (And even to some extent in general hydrology (Houska et al., 2015)). A thorough examination of the effect of the calibration algorithm would

C12

require implementing many different algorithms. Since there is a lack of earlier studies here to refer to it is acceptable to leave the calibration algorithm out altogether.

Changes in manuscript: the statement on the exclusion of the calibration algorithm as source of uncertainty (page 8, lines 4-7) will be rephrased: the calibration algorithm used in this study (SCE-UA) has been widely applied in hydrological applications. Although the calibration algorithm is recognized as a source of uncertainty (Deletic et al., 2012; Houska et al., 2015), there is a lack of previous research on this topic for urban drainage modelling to refer to, and therefore it is not considered here.

Validation performance

(22) Referee's comment: 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.

Authors' response: the overall ranking is shown for the calibration phase in Table 3, and shown for both calibration and validation phase in Table 9. Presentation of validation results could be presented better by having all results relating to the baseline calibration (HR model) in one table, i.e. combining tables 6, 9 and 11. A similar table could also be made for all results with the low-resolution model (i.e. replacing table 8) to better illustrate the benefits that the two-stage calibration offers there in terms of flow volume and peak flow performance in the validation phase, since these are more pronounced for the low-resolution model. The results from the low-resolution model

C13

were not described extensively in the manuscript but we believe they can strengthen a take-home message for the readers and should therefore be included in more detail. The results are currently best illustrated by Table 4.2 in the corresponding author's licentiate thesis (Broekhuizen, 2019), so the re-organizing of the tables with validation results should also include the data shown there. The table is included below (Table C3 in the supplement) for easy reference, but the data will be organized differently (i.e. one table for the HR model and one for the LR model) in the paper.

Changes in manuscript: re-organize tables as described above, and describe in more detail in the text the effects of the single- and two-stage calibrations for the low-resolution model.

(23) Referee's comment: 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.

Authors' response: the validation NSE is already presented in Table 9 to allow for comparison of the different CSs. The problem with any single validation characteristic is that it would either ignore some aspects of model performance or it would have to combine different statistics (i.e. NSE, volume error, peak flow error) in some arbitrary way. E.g. using only the NSE for validation performance would ignore that two-stage strategies perform better in terms of total flow volume and peak flow. We think it is interesting that different statistics give a different view of which calibration strategies perform better and this should be reflected in the manuscript.

Changes in manuscript: The discussion will be focused more on discussing the individual performance statistics to highlight that different criteria give a different picture of the effects of calibration data selection.

C14

(24) Referee's comment: 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.

Authors' response: although the single-stage performs better in terms of mean NSE (i.e. NSE calculated for each event, then averaged), it performs worse in terms of joint NSE (i.e. all events collated into a single time series for which NSE is then calculated), joint volume error and mean peak flow ratio. As discussed in section 3.3.2 the downside of the joint NSE is that it can give good scores even when several events are poorly predicted. Therefore joint NSE may be considered too optimistic which is why we did not use it extensively in this paper. In terms of a take-home message it is important to point out that the two-stage calibration is much faster since it reduces the dimensionality of the calibration problem compared to the single-stage calibration. In addition to this it has sometimes slightly poorer validation performance in terms of NSE but typically better performance according to other characteristics. The take-home message can also be strengthened by highlighting more the differences between HR and LR models (or rather that the benefits of the two-stage calibration are stronger for the LR model). This is currently best illustrated by Table 4.2 in the corresponding author's licentiate thesis (Broekhuizen, 2019), so the re-organizing of the tables with validation results (see comment 22) should also include the data shown there. The table is included in the supplement as table C3 for easy reference, but the data will be organized differently (i.e. one table for the HR model and one for the LR model) in the paper:

Changes in manuscript: changes according to the previous two paragraphs.

Recommendation

(25) Referee's comment: In its current form the manuscript is not in my mind publish-

C15

able in HESS. The following major changes would be required:

Authors' response: we believe that the major changes requested by the referee can be implemented in a new version of the manuscript, as detailed for the individual comments.

Changes in manuscript: see individual points below.

(26) Referee's comment: A more informative description of the hydrometeorological data to allow the readers to understand differences between different calibrations

Authors' response: see our response to comment 13 above.

Changes in manuscript: Include table C1 (supplement) in the manuscript's methods section.

(27) Referee's comment: A better justified reasoning for inclusion/exclusion of different error sources

Authors' response: this is addressed in our response to points 17-21, 24, 28.

Changes in manuscript:

(28) Referee's comment: 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

Authors' response: aspects to highlight in the conclusion and abstract:

- Two-stage calibration is faster, and can provide some performance benefits: e.g.

C16

better match of flow volume and peak flow in validation phase.

- Benefits of two-stage calibration are stronger for the LR model.

- Confirmation of earlier findings regarding input and calibration data from Dotto et al. (2014) and Kleidorfer et al. (2009) for a different data set and site (more green area). Findings are independent of the calibration event selection which provides support for their general applicability.

Changes in manuscript: changes according to the response above.

Technical corrections

(29) Referee's comment: Mostly technical comments. The comment for Figure 4 also relates to the content of the manuscript.

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

Authors' response: The text below the figure is added automatically by the Copernicus Latex template used for the submission and would not appear in the final published version of the article. This applies to the other figures as well.

The font size in the legend can be increased.

Changes in manuscript: increase font size in legend.

(31) Referee's comment: Remove the text below the figure (2 example hydrographs run130.pdf).

Authors' response: see above.

Changes in manuscript: -

C17

(32) Referee's comment: 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.

Authors' response: this will be corrected.

Changes in manuscript: this will be corrected.

(33) Referee's comment: 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?

Authors' response: The main reason why NSE is low is that the low flow rates in the event mean that the variance of observations is low, see also section 2.5, lines 13-14. For the baseline run (left panel), the variance of the observation is $1.2 \text{ L}^2 \text{ s}^{-2}$ while for the right panel it is just $0.33 \text{ L}^2 \text{ s}^{-2}$. The variance of the errors meanwhile is $0.25 \text{ L}^2 \text{ s}^{-2}$ (left) resp. $0.38 \text{ L}^2 \text{ s}^{-2}$ (right). NSE is calculated as $\text{NSE} = 1 - (\text{var err.} / \text{var obs.})$ so the variance of the observations is used as a scaling factor and it is mainly the difference in this factor that causes the degradation in NSE in this example.

Changes in manuscript: Add the variance of observations in figure and refer to section 2.5 in the caption for explanation of NSE and discuss this in the text of section 3.1.4, page 12, line 7.

(34) Referee's comment: Figure 5, 6, 7, 8. Remove the text below the figure.

C18

Authors' response: see above.

Changes in manuscript: -

(35) Referee's comment: Table 11. Mistake in the NSE single-stage value for D_prec (0.41)? The corresponding value in Table 6 is 0.43?

Authors' response: the correct value is 0.43.

Changes in manuscript: correct this to 0.43.

References

Barco, J., Wong, K.M., Stenstrom, M.K., 2008. Automatic Calibration of the U.S. EPA SWMM Model for a Large Urban Catchment. *Journal of Hydraulic Engineering* 134, 466–474. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2008\)134:4\(466\)](https://doi.org/10.1061/(ASCE)0733-9429(2008)134:4(466))

Broekhuizen, I., 2019. Uncertainties in rainfall-runoff modelling of green urban drainage systems: Measurements, data selection and model structure (Licentiate thesis). Luleå University of Technology, Luleå.

Deletic, A., Dotto, C.B.S., McCarthy, D.T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T.D., Rauch, W., Bertrand-Krajewski, J.L., Tait, S., 2012. Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth, Parts A/B/C* 42–44, 3–10. <https://doi.org/10.1016/j.pce.2011.04.007>

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., 2014. Impacts of measured data uncertainty on urban stormwater models. *Journal of Hydrology* 508, 28–42. <https://doi.org/10.1016/j.jhydrol.2013.10.025>

Houska, T., Kraft, P., Chamorro-Chavez, A., Breuer, L., 2015. SPOTting Model Parameters Using a Ready-Made Python Package. *PLOS ONE* 10, e0145180.

C19

<https://doi.org/10.1371/journal.pone.0145180>

Kleidorfer, M., Deletic, A., Fletcher, T.D., Rauch, W., 2009. Impact of input data uncertainties on urban stormwater model parameters. *Water Science and Technology* 60, 1545–1554. <https://doi.org/10.2166/wst.2009.493>

Krebs, G., Kokkonen, T., Valtanen, M., Setälä, H., Koivusalo, H., 2014. Spatial resolution considerations for urban hydrological modelling. *Journal of Hydrology* 512, 482–497. <https://doi.org/10.1016/j.jhydrol.2014.03.013>

Petrucci, G., Bonhomme, C., 2014. The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data. *Journal of Hydrology* 517, 997–1007. <https://doi.org/10.1016/j.jhydrol.2014.06.019>

Rossman, L.A., Huber, W.C., 2016. Storm Water Management Model Reference Manual. Volume I: hydrology (Revised). U.S. Environmental Protection Agency, Cincinnati.

Sun, N., Hall, M., Hong, B., Zhang, L., 2014. Impact of SWMM Catchment Discretization: Case Study in Syracuse, New York. *Journal of Hydrologic Engineering* 19, 223–234. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000777](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000777)

Tscheikner-Gratl, F., Zeisl, P., Kinzel, C., Leimgruber, J., Ertl, T., Rauch, W., Kleidorfer, M., 2016. Lost in calibration: why people still do not calibrate their models, and why they still should – a case study from urban drainage modelling. *Water Science and Technology* 74, 2337–2348. <https://doi.org/10.2166/wst.2016.395>

Please also note the supplement to this comment:

<https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-67/hess-2019-67-AC2-supplement.pdf>

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

C20