Summary:

This paper leverages a highly resolved database of urban land cover to evaluate how effectively different SWMM model set-ups capture the hydrologic processes of urban watersheds with GI at a small (I'll call 'hillslope') scale. Once the ideal model set up (ideal in that balances complexity with *perceived* accuracy) is identified, this model set up is used to simulate flows at the watershed scale. The watershed scale simulations are done to evaluate parameter sensitivity and compare source area contributions to hydrographs generated by GI vs. non-GI scenarios.

General comments:

The research questions posed are certainly worth answering. The data used, along with the author's demonstrated modeling skill, is suitable for answering the questions. Specifically, the author's attempts to improve representation of "reality" (i.e., an HRE) of urban surface connectivity to GI within the SWMM framework (i.e., a subwatershed) is valuable. However, the paper is often bogged down by an extensive discussion of the methods used, relative to an exploration of the results generated. Also, conclusions are often made without a reference to a specific metric of evaluation.

Study Limitations

There are some considerable limitations to the methods used in this manuscript. Specifically:

1. The 6 model set ups are evaluated at the hillslope scale using a design storm approach. Model "accuracy" (as the authors refer to it) seems to be based solely the author's expectations of physical processes in the watersheds. While I'm sure the authors have an excellent understanding of these processes in a historically well-studied watershed - using observed flows to validate the model set-ups would be a far better approach.

   If there is some other way that model accuracy is assessed (i.e., beyond comparing the output to expected behavior) in this portion of the manuscript (Sections 3.2) it needs to be stated explicitly. Are results compared relative to model set-up "Option 1" because this modeling framework is most complicated? Are they compared to the mean? Also - the “MR5subs" metric is not defined. Nor is “Variation from 5 Subs” on the y-axis in Figure 10. Is this how accuracy is assessed? Whatever assumptions are made to compare, they must be stated explicitly (and justified)

   To that end: I would not suggest comparing model output to other model output. I would suggest comparing the 6 model set-up options at the **watershed scale** - and see which set-up is best able to represent the **observed flows** in the July 2009 period with and without calibration. Then, the can truly be assessed, and there is potential for a discussion of the complexities associated with the various model set ups and calibrations.

   Alternatively, the authors could re-frame this entire section of the manuscript as “critical evaluation of model physics” or something – an avoid using the term accuracy. However, the assumptions that go into the “critical evaluation” need to be explicit and justified with literature and/or specific observations at the site.

2. With regard to the hydrograph separation method: separating the sources of runoff is definitely valuable, and the way the authors do this is clever. However, is SWMM not able to output these variables explicitly? It is a physical model. Why do the authors need to back-calculate these fluxes by adding/subtracting different model scenario output? If either SWMM cannot output these variables directly or if it is too complicated to aggregate the output from all the subwatersheds at the watershed scale: a justification for the "back-calculation" approach should be provided in the introduction.

3. I do not see the value in "calibrating" – or even restricting - BPA width. The actual BPA will change with storm size (as the authors point out). Why not just leave the BPA as the total pervious area – which is a better representation of the potential to infiltrate run-on. Then, just change the connectivity to the pervious area in the modelling scenarios and allow the model's infiltration routine calculate how much water is infiltrated vs. how much runs off the surface? This is more realistic. Some specific comments on this can be found in the next section of the review. Additionally, the specific metrics by which different BPA widths were compared is not explicitly mentioned in the text.

Manuscript Organization

I also have a series of suggestion regarding the organization of the manuscript. The manuscript, as it is, is very, very long. It includes 14 figures. I'm aware of HESS's page limitation, but it seems like this manuscript could be streamlined. Some suggestion I have for removing/reorganizing content are:

1. Great length is spent regarding the watershed scale model parameterization and set up (i.e. Fig 6 and its paragraphs of discussion). It seems to me like much of the detailed methods discussed are standard practices in SWMM. I would suggest the authors determine which of these steps is novel vs. which are just typical urban hydrologic modeling practices, and focus on just writing up the novel components. Direct the reader to the SWMM user manual, or assume they have some basic understanding of hydrologic modelling. I have some specific suggestions in the "Specific Comments" section of this review.

2. The SWMM model and its components (i.e., parameters, use of subwatersheds, etc. – Section 2.5) is introduced AFTER these components have been discussed extensively during the description of the 6 model set-ups (Section 2.4). Perhaps moving this ahead of the set-ups would be a better organizational set up, and allow for the removal of duplicate text from the current Section 2.4.

3. The ratio of methods to results in this paper is extremely large. I understand that the steps presented in the manuscript are meant to be used as a model to future SWMM users modelling GI, but the introduction frames the manuscript are purely evaluating SWMM model set up for evaluating GI. Perhaps some text in the introduction that describes how this manuscript is really a suggested method for modelling GI in SWMM - with some critical evaluation - would be more suitable for the actual contents of the manuscript.

4. Reference the Lee, 2017 report when describing the watershed's land cover instead of documenting it in this manuscript

5. It is often unclear which spatial scale (i.e., hillslope vs. watershed) specific parts of the manuscript are describing. Clarifications should be added in almost all of the manuscript's sub-sections.

Specific comments:

A note before beginning – continuous line numbers would make completing a future review of this manuscript much easier

Abstract

- Pg 1, Ln 17-21: Add clarification on spatial scale: design storms = hillslope; continuous = watersheds

Introduction

- Pg 2, Ln 9-10: What about infiltration as an objective of GI?
- Pg 2, Ln 12: Each installation is relatively less expensive, but also has relative less impact. Should add some detail as to what costs your describing while making this claim
- Pg 3, Ln 12: "modeld" >> "modeled"
- Pg 3, Ln 28: "After the HRE delineation is performed in SWMM, each one undergoes" >> Clarify sentence structure so "one" is clear
- Pg 3, Ln 34: Clarify "conventional objectives"

Methods and Material

- Pg 5, Ln 28: What is a stormwater detention area vs. a dry/wet pond?
- Pg 6, Figure 2: Possibly add some of the shapefile IDs to the map? Generally, though, I'm not sure this figure is needed. The text adequately describes the spatial database.
- Pg 7, Ln 5-10: BPA is variable based on storm size (as the authors point out) – so why try limit yourself to one BPA width? Why not just allow for the entire pervious area to potentially infiltrate? As would be the case in real life. If the run-on is greater than the infiltration capacity at any time, the infiltration model will handle this appropriate and generate surface runoff.
- Ph 7, Ln 11: Where do these widths come from?
- Pg 7, Figure 3: While it is certainly attractive, this figure is not that valuable (especially considering there are so many in the manuscript already). I think the readers can conceptualize what a 0.3m buffer looks like without the visual aid.
- Pg 7, Figure 4: Consider adding the inlets to this map
- Pg 8, Figure 5: For the subcatchment boxes: do the sub-subcatchment areas (i.e., "TPA" and "TIA" in Option-4) run in series or in parallel? There is discussion of this (somewhat) in the text when discussing the SWMM model connectivity scenarios – but including a set of arrows and/or arranging the sub-subcatchment areas vertically when connected in parallel and horizontally when in series would be of great value to this figure

- Pg 9, Ln 33: This could be my own lack of understanding – but I'm unclear on what the "cumulative" precipitation values represent. Either way - I'm not sure it adds considerable value to include this in the text and in Table 1. Percentile is likely sufficient.
- Pg 10: As mentioned in the "General Comments" – the SWMM model and its components/parameters should probably be introduced prior to the discussion of the 6 model set-ups
- Pg 10: This subwatershed set up was done for the entire watershed – may want to clarify this at some point. Same with the discussion of parameterization on Page 11
- Pg 10, Ln 30: Why bother introducing the 16 land cover types at all? Just introduce the 10 the first time, and remove this later clarification.
- Pg 11, Figure 6: This figure seems too fundamental to the process of modelling to warrant inclusion. As mentioned: unless the paper is reframed as a "new method" for running SWMM with GI – I would remove this. I would actually advocate for this framing because it aligns better with the content (namely, all the nitty-gritty set up and parameterization details included). However, I'm not sure HESS would be the place for such a manuscript
- Pg 11, Ln 15-32 and Pg 12, Ln 1-24: There is far too much SWMM set up 101 in this section. Is it possible to reduce this to a table of parameter values? And just refer the user to the SWMM manual when describing what each parameter is and how it was determined?
- Pg 12, Ln 31: Hillslope or watershed scale?
- Pg 13, Ln 7: Here is evidence that one BPA width is probably not the best way to model this. Again – I'd advocate allowing all of the BPA to infiltrate water, and let the model physics determine how much of that water makes it to the subsurface – this seems more representative of reality
- Pg 13, Ln 9-15: Refer to SWMM manual
- Pg 13, Ln 27: PCSWMM should be mentioned earlier in the manuscript if it is a critical step to the model set up
- Pg 13, Figure 7: Is it possible to combine Figure 4 and 7?
- Pg 14, Ln 9: BPA is not in Figure 6
- Ph 15, Ln 1: What is the spatial scale these simulations were done at? And what model set-up was used (out of the 6 tested)?
- Pg 16, Ln1-10: See "General Comment" on including justification for back-calculating these values vs. taking model output directly. A good place to do with would be in the last paragraph of the Introduction on page 4

Results and Discussion

- Pg 16, Ln 19ff: This not a result, it is a description of the watershed. Aren't all of these details in an EPA report somewhere (e.g., Lee, 2017)?
- Pg 16, Ln 22: What scale was the BPA calibration performed at? If watershed scale, then I would assume the July 2009 time period was used. What size storm events occurred in this period? If they're all small – I would expect an underestimation of the BPA width. If it occurred at the hillslope scale – what design storms were used?
- Page 16, Ln 22: These results do not mean "that the runoff from ICIA is discharged

to the adjacent 0.61 m of pervious area" – soften the language. E.g.: "a 0.61m buffer width best mimicked hydrologic behavior, evaluated by metric XYZ"
    - Note: what was used to compare these runoff widths? Total runoff? Peak flow? NSE? Include this, it seems important.
- Pg 17, Ln 1: What spatial scale did this occur at?
- Pg 17-18: See General Comment #1 on the use of the word "accuracy" in this section 3.2
- Pg 18, Ln 17: Define MR5subs – it is important
- Pg 19, Ln 3-4: "Fig 10" >> "Fig 11"
- Pg 19, Ln 1-10: 3% change in total runoff doesn't seem like a lot of sensitivity. This should be noted. How much did total runoff change during calibration?
- Pg 19, Ln 13-24: Move calibration discussion of calibration ahead of discussion of sensitivity analysis. Alternatively, combine and condense these paragraphs.
- Pg 20, Ln 12: "Fig. 12b and 13" >> "Fig 13b and 14"

Conclusions

- Pg 21, Ln 3: "accuracy" is not the best word