Reviewers' comments:  
  
  
**Reviewer #2:** This manuscript examined the effectiveness of green roofs to attenuate stormwater runoff across a large metropolitan area. The topic is worth investigating. However, there are some suggestions to improve the manuscript:   
1.There is a lack of in-depth analysis of the mechanism of green roofs delaying and retaining stormwater during the hydrologic simulation in the "Introduction" section.

RESPONSE:

2.It is suggested that more detailed information about the model and its structure should be supplemented to improve the "Watershed Model" section. For example, what are the characteristics of the model? How to explain the mechanism of the model? How the reliability of the model？

RESPONSE:

3.The data used in present work includes national elevation dataset, urban streams, land use/land cover, temperature and precipitation, etc. However, the spatial resolution of these data is different, and how to deal with the scale problem among them? Meanwhile, I noticed that DEM data with high spatial resolution was used in the study. In addition to the structural or percentage factors of the green roofs, can the influence of other factors such as slope be considered in the study?

RESPONSE:

4.The observed streamflow data on line 214 is significance in the hydrological simulation calibration. However, this data lacks relevant introductions and is not shown in Table 2.

RESPONSE:

5.The process of the model calibration is unclear. The authors seem to be using the calibration results of one watershed (Taylor Creek) for other study areas. And how to eliminate or evaluate the errors produced by this process?

RESPONSE:

6.It is necessary to further analysis research results rather than only display the statistical results. It is suggested that a separate discussion section for further analysis of this study.  
RESPONSE:

7.The thumbnail in Figure 1 is unclear. It is recommended to adjust the size of the thumbnail or add text description to enhance the readability of the thematic map.

RESPONSE:

8.The final calibrated value of the petparam 1 (Conifer) (0.99) exceeds the calibrated range (0.27-0.92)?

RESPONSE:

**Reviewer #4:** Thank you for the opportunity to review your paper "Modeling the hydrologic effects of watershed-scale green roof implementation".  The manuscript is well written and organized in a way that is easy to follow.    
RESPONSE: Thank you for your supportive comments.

Introduction:  
You mention in multiple places (including your abstract) that modeling studies of green roofs at the catchment scale is lacking and you cite very few papers.  A quick literature search "green roof catchment scale" turned up multiple studies (First page of results reproduced below):  
  
\*Schmitter, P., Goedbloed, A., Galelli, S., & Babovic, V. (2016). Effect of Catchment-Scale Green Roof Deployment on Stormwater Generation and Reuse in a Tropical City. Journal of Water Resources Planning and Management, 142(7), 5016002-. <https://doi.org/10.1061/(ASCE)WR.1943-5452.0000643>  
\*Masseroni, D., & Cislaghi, A. (2016). Green roof benefits for reducing flood risk at the catchment scale. Environmental Earth Sciences, 75(7), 1-11. <https://doi.org/10.1007/s12665-016-5377-z>  
\*Ercolani, G., Chiaradia, E., Gandolfi, C., Castelli, F., & Masseroni, D. (2018). Evaluating performances of green roofs for stormwater runoff mitigation in a high flood risk urban catchment. Journal of Hydrology, 566, 830-845. <https://doi.org/10.1016/j.jhydrol.2018.09.050>  
\*Versini, P., Gires, A., Tchinguirinskaia, I., & Schertzer, D. (2016). Toward an operational tool to simulate green roof hydrological impact at the basin scale: a new version of the distributed rainfall-runoff model Multi-Hydro. Water Science and Technology, 74(8), 1845-1854. <https://doi.org/10.2166/wst.2016.310>  
\*Versini, P., Ramier, D., Berthier, E., & de Gouvello, B. (2015). Assessment of the hydrological impacts of green roof: From building scale to basin scale. Journal of Hydrology, 524(524), 562-575. <https://doi.org/10.1016/j.jhydrol.2015.03.020>

(Others that look at optimizing location of GI or combinations of GI that appeared on the first page of the search)  
\*Giacomoni, M., & Joseph, J. (2017). Multi-Objective Evolutionary Optimization and Monte Carlo Simulation for Placement of Low Impact Development in the Catchment Scale. Journal of Water Resources Planning and Management, 143(9), 4017053-. <https://doi.org/10.1061/(ASCE)WR.1943-5452.0000812>  
\*Palla, A., & Gnecco, I. (2015). Hydrologic modeling of Low Impact Development systems at the urban catchment scale. Journal of Hydrology, 528, 361-368. <https://doi.org/10.1016/j.jhydrol.2015.06.050>  
  
This is not intended imply these specific studies are all applicable they were simply the first to come up; however, there is relevant literature and therefore the statements regarding the lack of literature are not entirely valid - there is less research than in other areas but relevant literature does exist and should be cited.  You are urged to conduct your own literature search and add to this portion of the introduction to adequately explain the novel contribution of this work.

RESPONSE: Thank you for your comment and for providing examples of studies related to evaluation of green roofs and LID at catchment scales. We’ve

General questions / comments on the model construction:  
Did you model the existing GI (ponds, etc.) in VELMA when adding the GR coverage?  
RESPONSE: No. The only land uses included in the model were buildings, ground-level impervious surfaces (e.g., roads, sidewalks, parking lots), grass, and trees.

We added a sentence in the manuscript to clarify this point.

“Four land use categories were included: grass, trees, buildings, and other impervious surfaces (e.g., roads, sidewalks, parking lots). Therefore, existing GI (e.g., established green roofs, ponds, impervious pavement) were not included in the baseline model simulations.”

The 1m land-use land cover data was resampled based on "majority rule".  Please provide a short description of what the majority rule does.

RESPONSE:

VELMA is not as commonly used for urban stormwater modeling as SWMM (for example).  It would be helpful for the reader if you provide a little more background on the underlying VELMA model, particularly as it pertains to the relevant rainfall-runoff modeling for this application.

RESPONSE:

More description in methods section.

How did you determine where the green roofs were placed?  Was it a percentage of every roof or were full roofs randomly chosen until the required percent coverage was met?  The second approach would produce a range of results based on the location of the green roofs but is arguably more realistic.

RESPONSE: The first approach you describe is what was implemented in our simulations. VELMA simulates 10-m cells throughout the watershed, some of which are designated as building cells. Of these building cells, some percentage (e.g., 25%, 50%, 75%, 100%) were chosen randomly to be converted with green roofs. This means that some green roof implementations included only a portion of a roof while others included full roof implementations. This is less realistic than the second approach you describe in which only full buildings are chosen together, but it is much simpler to implement in our modeling context.

We’ve added the following portion to further clarify our choice.

“Note that the method of random placement applied to 10-m cells, and therefore some green roof implementations only covered a portion of the total roof area of a given building.”

You have a petparam for conifer and grass - you state this accounts for ET.  Did the model account for ET from the green roofs?

RESPONSE:

Line 217: "McKane et. al in preparation" is problematic in a methodology section.  It seems these are being published out of order.  If this is to be published first, a better description of the semi-automatic calibration process is needed.  
RESPONSE: The reference “Mckane et al. in preparation” is a model manual rather than a peer-reviewed publication. We’ve

Replace description with MOEA citations as well as Deb NSGA-II.

Results:   
The results in general match much of the literature including the inevitable saturation under heavy rainfall.  However, you have a set of anomalous results for Pipers creek.  Given this is the first time the VELMA model is used to model GRs, anomalous result must be fully explored and understood.  The lack of sewers does not explain the decrease in runoff retention when increasing the green roof coverage.  In addition, the application of parameters which were not calibrated to the specific watershed is not a strong argument given it only occurs in Pipers creek for the extensive GR. There is also no explanation for why the model produces a data point that is above the baseline case, how did adding 100% green roof coverage make things worse?   
  
Ultimately, this needs to be explored and explained given some of the choices that were made while creating the models:  
1.Allowing lateral flow into and out of the green roof parcels   
2.Ignoring storm sewers / pipes   
3.Calibration of only one model and then transferring the parameters to others, all models should be calibrated and validated.   
4. No sensitivity analysis is included which is fairly common in modeling studies  
Combining all four of these concerns with the lack of explanation of the anomalous results, the results and their applicability to reality in an urbanized setting are suspect.  A more thorough exploration of the model behavior is warranted.

RESPONSE:

**Reviewer #5:** Dear Authors,  
I have reviewed your manuscript, and found it very interesting, even if modelling is not my top expertise, I could understand perfectly form your description, how you applied the model to these two different green roof tops. I think your work is useful to any urban planning policies, beyond«d the cases you analysed empirically.

My only comment is that, exactly for that reason, I missed a more comprehensive framing of the GI - green roofs within the planning processes, challenges, and policies' opportunities, by exploring more its ecosystem services (that you list) and a broader view of how these simple techniques can effectively contribute to buffer climate change impacts, energy savings, and solar energy for example, among others. I think that in some parts of the introduction, and the conclusions those urban scale and policy level links/mentions would enrich you manuscript.    
  
I congratulate you all, and wish the best to your publication.  
The reviewer  
RESPONSE: Thank you for your supportive comments. ADD URBAN SCALE AND POLICY LEVEL LINKAGES.

**Reviewer #6:** The manuscript "Modeling the hydrologic effects of watershed-scale green roof implementation" by Barnhart et al. investigates the effectiveness of green roofs on storm water attenuation in urban watersheds. Using modeling approach, the paper shows quite significant reduction of mean annual runoff (15-25%), indicating that green roof implementation can be an effective tool for stormwater management in the area. The results provide some new insights about the hydrologic impacts of green roofs at catchment scale. However, the methods and results need to be presented in much greater details to make this work more robust. This paper need substantial revision before accepting for publication on JEMA.  
  
Comments:  
1. The authors used a spatially distributed eco-hydrologic model (VELMA) to simulate daily surface runoff but did not incorporate underground drainage/sewer systems in the model. Is there any underground sewer/drainage system in the study area? Many urban areas has sewer systems that significantly affect overland flow and runoff. This point should be mentioned clearly in the method section.

RESPONSE: Yes, the Greater Seattle Metropolitan area

2. Information shown in Figures 3-6 is not helpful at all. Probably, information about land cover, weather stations, etc can be shown in another Figure for Taylor Creek only (Other watersheds can be presented separately in the Supplementary). Also, did you randomly select buildings for green roof for 25%, 50%, and 75% scenarios?

RESPONSE: We have moved Figures 4-6 to a supplementary section. Also, we included additional language that explains the choice of green roof coverages for the different scenarios. We randomly allocated green roofs to cells (not entire buildings) in the 25%, 50%, 75%, and 100% scenarios. See our response to Reviewer #4 regarding the random placement of green roofs in the scenarios.

3. The authors should present a couple of figures to discuss the variability of temporal data input across scales. Variability of hydrologic response from green roofs should also be discussed in details as well.

RESPONSE:

4. The results for 75% and 100% extensive scenarios in the Pipers Creek indicates that the model used is not robust and should be fixed. Otherwise, the results for other watersheds are not reliable as well.

RESPONSE:

5. Figure 8: Are the trends in each subplots statistically significant? It would be better to add more information about the linear regression lines (slope, R2) in the subplots. The current figure is not informative.

RESPONSE:

6. Line 175-176: What do you mean by "require complex processing to ensure reliable outputs" for lidar. Theoretically, higher-resolution topographic data is better. The reason you choose 10-m resolution is simply a balance between accuracy and simulation time.

RESPONSE: The original intent of this section was to explain that state agencies and municipalities oftentimes do not have the expertise or resources to process lidar data. However, you make a good point, and we have revised this section to only specify the compromise between accuracy and simulation time.

“A 10-m digital elevation model (DEM) was acquired from the USGS (Table 2). This product was chosen over lidar-based digital terrain models, which provide higher spatial resolution, because higher resolutions require more voxels to be simulated within VELMA, which in turn increase the total simulation time. Therefore, we chose a 10-m DEM, and more generally 10-m voxels throughout the watershed, as a compromise between accuracy and simulation time.”