Dear referees,

We would like to thank you for your careful reviews and constructive suggestions with regard to our manuscript. Your thoughtful remarks and suggestions enabled us to enrich the paper and bring complementary information and details that make it more straightforward. We therefore took into account all your comments, queries and suggestions.
We also thank you for the clarity of your reports that made the correction work easier, as well as our point-by-point responses that follow. Most of the time, our replies are presented under the form of modified texts ready to be inserted to obtain a revised version, which will also take into account the editor's comments and suggestions. Referees' comments appear in italics and black colour, our reply in normal font and blue colour, and corresponding modifications proposed for the text in the manuscript are in underline font and red colour.

Responses to Referee 1
Received and published: 8 January 2021

Referee comments 1 :
In this article, the authors aim to investigate the uncertainty of hydrological responses in various NBS scenarios resulting from the spatial variability of rainfall and the heterogeneous distribution of NBS at the urban catchment scale. I find the manuscript to be quite suitable for HESS and presents a straight forward method which is intended to provide means to modelling works for NBS at the urban scale.
Overall, the manuscript is well presented. I would like to suggest a Minor revision for this paper, but there are a few points that should be addressed to improve its quality.
1. You use 25-m resolution DEM but the model was implemented with a 10 m spatial resolution, could you please make a comment or discussion about this? Moreover, 25-m resolution DEM seems relatively quite rough to use for an urban area. Do you think this is the limitation of your study? If you use a more detailed DEM, will it impact your result?

Author's answer:
Thanks to the referee's suggestions, the authors added the following underlined sentences (at lines 193-194 of the previous version) to enhance the comprehension and readability of the manuscript:

"Besides the land use, the elevation is also assigned to each pixel of the model. For this purpose, the interpolation was used to downscale the raw DEM data from 25 m to 10 m (DEM25-10) to incorporate it with the model resolution. More precisely, each pixel was first subdivided into 25 equal sub-pixels as a proxy of the 5 m resolution, then the elevation data were up-scaled 4 by 4 pixels to produce the 10 m interpolation of the original elevation.
While the 25 m resolution DEM may seem too coarse to use for an urban area, it did not limit the study in any way because the catchment is relatively flat. To test this, we up-scaled the raw 5 m DEM data to adapt them to the model resolution (DEM5-10). Table. 1A presents the results of the statistical analysis of DEM25-10 and DEM5-10, which are so similar that the difference could not impact the results. For instance, the Root Mean Square Error is about 0.26, and the correlation coefficient is around 0.99. Besides, the ensemble of the data actually available for the Guyancourt watershed would need to be more detailed to make it worth going to a higher resolution of the model.

Statistic metrics	DEM25-10	DEM5-10
Median	143.3	143.4
Mean	160.1	160.1
Maximum	175.4	175.9
Minimum	143.0	143.3
Standard deviation	80.2	80.2
Root Mean Square Error	0.26	
Correlation coefficient	0.99	
Maximum difference	5.3	
Mean difference	0.01	

Table 1A . The statistical comparison of DEM5-10 and DEM25-10. This table plans to be inserted in the revised manuscript after the Table 2.

<u>Referee comments 2:</u>

2. Figure 5, In the legend it should include an abbreviation of each NBS measures e.g., Porous pavement (PP), Rain garden (RG),. . .

<u>Author's answer:</u>

Thanks to the referee's suggestion, we have revised the Figure 5 including the abbreviation of each NBS measures.

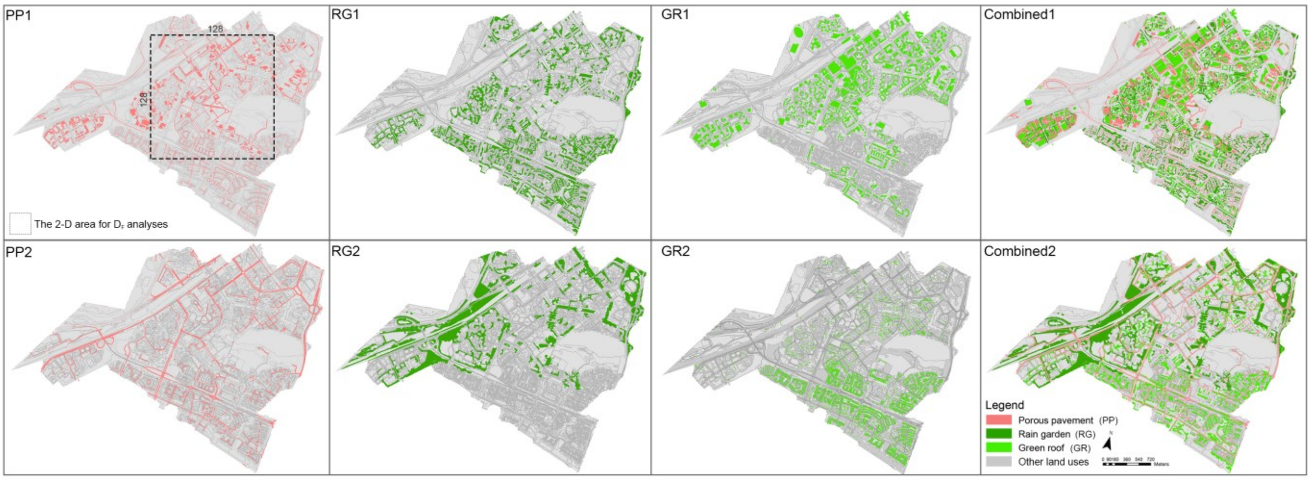

Figure 5. NBS scenarios including PP1, PP2, RG1, RG2, GR1, GR2, Combined1, and Combined2. The rectangular area presented in the PP1 scenario is the example area for applying fractal analysis.

Referee comments 3:

3. Please consider the results of validation from Line 328 to the Results section as this is results of the validation of your baseline scenario.

Author's answer:

Thanks to your suggestion. We have moved the results of validation (from line 328 to line 340) to the Results section and added some discussions in the revised version of the manuscript.

4 Results and discussion

4.1 Validation of baseline scenario

Regarding the water levels observed and those simulated in the baseline scenario, the model indeed performs well for the studied area. The *NSE* coefficients and the *PE* indicators validated Multi-Hydro's performance (see Table 4). Indeed, for the three distributed rainfall events (Figure 9), the *NSE* are larger than 0.9, and *PE* are lower than 5 %. For the uniform rainfall event of EV2, the model represents the water levels with *NSE* equal to 0.95, and *PE* equal to 1.96 %: only a slight overestimation of the observed water levels is observed between hours 4 to 7. For the uniform rainfall of EV1 and EV3, the temporal evolutions of simulated water levels slightly underestimate the observed ones, with *NSE* around 0.8, as well as *PE* around 7 %. Regarding the temporal evolutions of simulated water levels under the distributed rainfall of EV1 and EV3, they are more consistent with the observed ones. The reason is that the rainfall intensities of the distributed rainfall are generally higher than those of the uniform rainfall at the storage basin location. Namely, in uniform rainfall events, the accumulated water levels in the storage basin are less than that of in distributed rainfall events. Overall, the distributed rainfall gives slightly better results, and the simulated water levels using uniform rainfall also match sufficiently well the observed ones to validate the Multi-Hydro implementation in the Guyancourt catchment.

Regarding the validation results, the scalability of Multi-Hydro allowed us to define the optimal resolution to finely reproduce the spatial heterogeneity of the watershed. Remember that this resolution is the ratio between the external scale of the watershed and the scale of the grid. The heterogeneity mentioned above propagates from the smallest scale to the largest, impacting the simulation results in any through the hierarchy of spatial scales of the watershed. It should be understood that the selected 10 m grid scale is not the smallest scale possible, but the optimal one to ensure a good balance between, for example, sufficient heterogeneity and the required quantity of the data required, again in precision valuable and involved computing time involved. As discussed in Section 3.2, the spatial heterogeneity for each of the NBS scenarios evolves with the fractal dimension on two scale ranges: the asset implementation scales (10m - 80m) and the larger basin scales. Such an evolution remains fully compatible with the intrinsic scalability of Multi-Hydro, which makes it particularly suitable and sufficiently reliable to study the impacts of the spatial variability of hydrological responses in different NBS scenarios.

Referee comments 4:

4. For Figure 12 and 16, can you discuss more on why percentage errors of peak flow are much higher than total runoff volume.

Author's answer:

Thanks to the referee's suggestion. The authors added the discussions for Figures 12 and 16 to enhance the comprehension and readability of the manuscript (in our revised manuscript, the verification metrics of percentage error was changed to percentage difference as the metrics throughout the manuscript. The corresponding abbreviation for percentage difference on peak flow was modified to 'PD_{Qp}', and the percentage difference on total runoff volume was modified to 'PD_V').

As shown in Figure 12 and Figure 16, the percentage difference on peak flow is much higher than that of the total runoff for each scenario. These results can be explained by the fact that the spatial variability of rainfall intensity at the largest rainfall peak is strong in all three rainfall events, while the total rainfall volume for the gridded and uniform rainfall inputs is the same. Being an integrative variable of the rainfall over the watershed, the total volume of runoff should not differ much for gridded and uniform rainfall inputs. This small PD_V is influenced by the differences on the grid scale (storage capacity, infiltration, etc.), which are differently modelled when the input is uniform or non-uniform. For these three rainfall events, the spatial variability of total rainfall depth is less pronounced with respect to the spatial variability of the rainfall intensity at the largest rainfall peak, and also there is no highly localized storm cell. Figure 3 (top) displays the rainfall intensity at the largest rainfall peak (per radar pixel) over the Guyancourt catchment area for the three studied rainfall events. It is noticed that the highest rainfall peak of the gridded rainfall is very variable in space, which enlarged the discrepancy with the corresponding uniform rainfall data, resulting in a significant impact on the peak flow of each NBS scenario that simulated with two different rainfall data. However, the cumulative rainfall of gridded rainfall data is not very variable in space (see Figure 3 middle). For instance, the standard deviation (*SD*) of the cumulative rainfall of the three rainfall events is around 1 mm, which indicates that the spatial variability of the distributed rainfall is not very pronounced at most of the time steps. Thus, the difference between gridded rainfall data and uniform rainfall data is relatively small during the whole rainfall period. Finally, the simulated flow of NBS scenarios under two different rainfall inputs is similar in most time steps, resulting in the percentage difference on the total runoff volume of NBS scenarios (simulated by gridded rainfall input and uniform rainfall input) is not significant.

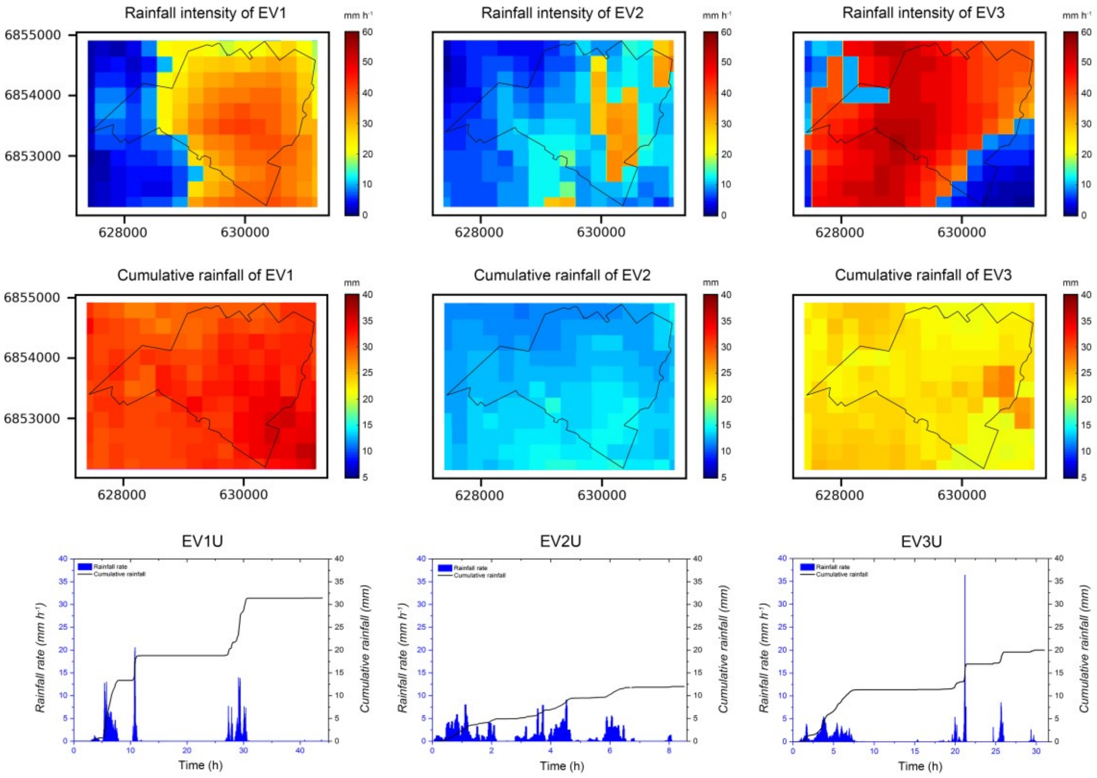

Figure 3: Top: The rainfall intensity at the largest rainfall peak (per radar pixel) over the Guyancourt catchment area for the three studied rainfall events. Middle: Cumulative rainfall depths (per radar pixel) over the Guyancourt catchment area for the three studied rainfall events. Bottom: Time evolution of rainfall rate (mm h^{-1}) and cumulative rainfall (mm) of the three uniform rainfall events over the whole catchment.

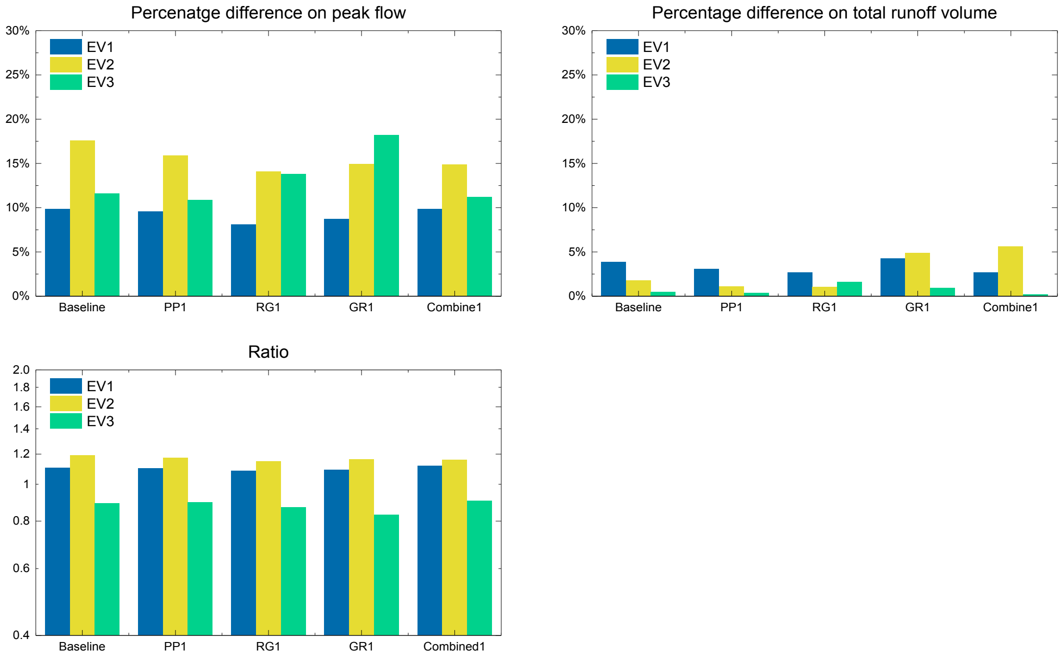

Figure 12. (a) Percentage difference on peak flow of the baseline scenario and the first set of NBS scenarios under the three distributed rainfall events and the three uniform rainfall events. (b) Percentage difference on total runoff volume of the baseline scenario and the first set of NBS scenarios under the three distributed rainfall events and the three uniform rainfall events. (c) The ratio of peak flow between the scenarios under the distributed rainfall and the scenarios under the uniform rainfall.

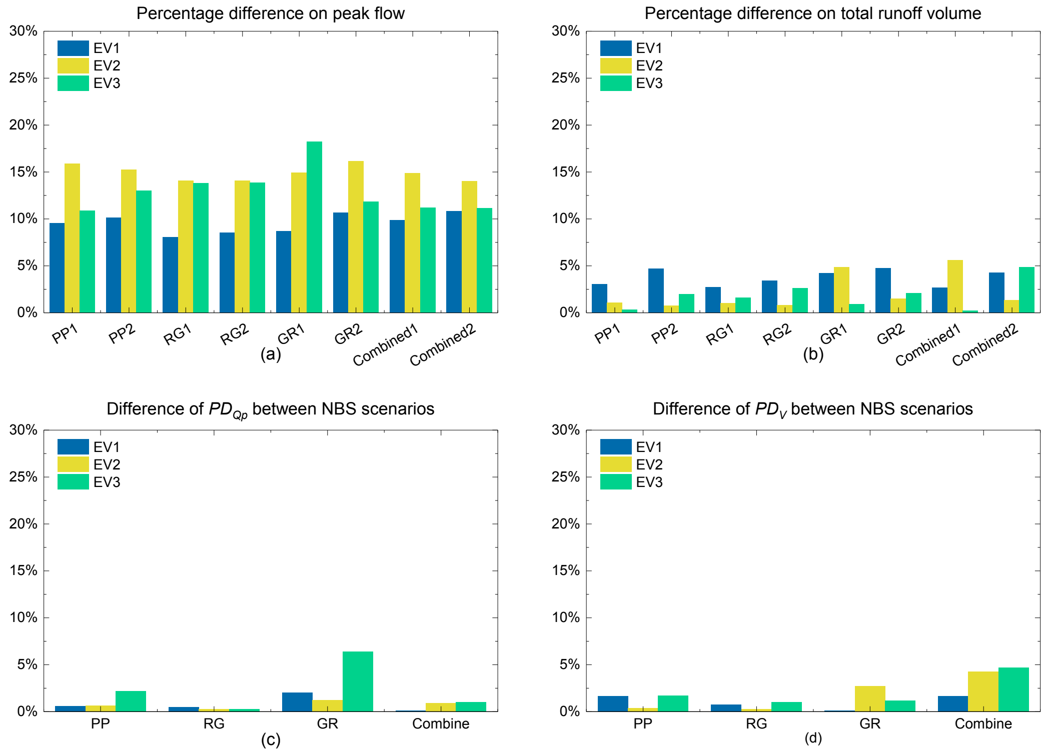

Figure 16. (a) Percentage difference on peak flow of all NBS scenarios under the three distributed rainfall events and the three uniform rainfall events. (b) Percentage difference on total runoff volume of all NBS scenarios under the three distributed rainfall events and the three uniform rainfall events. (c) Difference of PD_{Qp} between the same types of NBS scenario. (d) Difference of PD_V between the same types of NBS scenario.