In their paper, Pedersen et al. present an approach for assessing the performance of urban drainage models at the local level (i.e. at different sites) based on a variety of criteria. The manuscript is well-written and structured, and it is easy to follow. There are many other papers that deal with model validation at local scales in rural and urban areas, so I found the novelty of the paper to be a bit weak. My rating for the paper's suitability for HESS is "medium", as I am not sure its content will be of interest to most of the journal's readers. Apart from that, I have no major concerns with the paper's content, as both the methods and the analysis are sound. Below you will find some specific comments.

Figure 2. This is a cross-section, I assume. Include labels to indicate ground level, pipe borders, circle dashed lines, etc.

Table 1. Units should be added to the variables.

Figure 4. Figure 4 is not very informative. Could you please elaborate on the different symbols in the figure caption? This figure should also be moved to the supplementary
material - it does not add any value to the article.

Figure 5. The 1:1 line should be in a different color or presented as a solid line.

Figure 6. Can also be transferred to the supplementary material.

Sites names. I suggest simplifying the names of the sites in the text (e.g. in line 336) and in the figures (e.g. Figure 7). For example, it would be easier to read "Site A" instead of "F64F46Y".

Lines 467-472 and Figure 12. From these statistics, what can we learn? Does it tell us anything about the model's capabilities? I suggest removing this part from the manuscript if not.

Section 4.4 and Figure 13. I don't understand this section. The figure shows what exactly? I cannot follow the rationale for plotting based on physical properties, signatures, and slopes here.

Conclusion section. Currently, it's a summary, mirroring the abstract a bit. I suggest shortening it to one paragraph, summarizing the main findings.