

## Reply to Anonymous Referee #2

Thank you very much for your comments and suggestions for this paper.

- 1. The authors should discuss the implications of their results on the quality of the original data. A definition of “faulty input data and observation” should be provided.**

We agree and will add a few lines of definition:

We omitted raw data which were obviously faulty i.e. when the measured velocity or level was zero. This happens occasionally when objects such as toilet paper etc. clogs/attaches to the flow meter. Of course there might be cases when the gauge is only semi-clogged and hence unreliable measurements are sampled and included for the analysis but such data can be very hard to separate from good data. The implication of using all the original data would thus have meant a lower containment ratio and assumingly larger prediction bounds.

- 2. The rationale for substituting the distributed model with the lumped one should be more clearly outlined, and its limitations explained.**

The primary scope with the paper is to test the usability of the GLUE methodology as a tool for uncertainty analysis and estimation in complex urban drainage modelling, that is, by evaluating how well the GLUE methodology performs when the behavioural parameter sets deduced from a calibration period are applied to generate prediction bounds in a validation period. For this test we decided to keep the model simple and compare observed and modeled flow at just one place downstream the considered catchment. Hence we inferred that a state-of-the-art physically distributed model that calculates flow and levels in every pipe of the sewer system would be overly complex for the purpose considering both the computational requirements and the risk of over-parameterisation. Instead a simple modeling approach was chosen yet complex enough to describe the major flow components (diurnal wastewater variations, fast rainfall-runoff from paved areas and slow Infiltration-inflow from unknown sources).

Moreover in Section 2.2 we refer to a study by Thorndahl et al. (2008) that showed for a similar GLUE investigation of an urban drainage system using a distributed model that the hydraulic parameters (Manning number and minor losses) played an insensitive role when extracting the behavioural parameters of the model, while the surface runoff part of the model (particularly the hydrological reduction factor and time of concentration) were very sensitive. This suggests that simple lumped modelling as applied here is adequate. The limitation of using such a simplistic modelling approach is that the model may be too simplistic, e.g. in cases when system components such as weirs, gates, pumping stations and storage tanks play a significant impact on the observed flow or in cases with heavy backwater effects. We will add a few lines about this in the Section 2.2.

- 3. The choice of likelihood measures introduced in (1) (linear and exponential) should be motivated and contrasted with other possible choices.**

We will elaborate a bit more on our choice of likelihood measures but we do not find it necessary to contrast with other possible choices as this is not the scope with this paper. We have here applied some commonly used likelihood measures for the test of consistency in the

GLUE derived uncertainty bounds between calibration and validation periods. There are other papers that have dealt thoroughly with contrasting the different likelihood measures.

The linear likelihood was chosen for the dry weather case because of the desire to fit the dry weather diurnal flow pattern well whereas an exponential likelihood was chosen to fit the peaks of the hydrographs better. The exponential likelihood accentuates the peaks, and weights them higher compared to local minima. The flow peaks are normally an important output in sewer flow modeling to assess surcharge and flooding. We will add a comment on this.

**4. The same holds for the choice of a multiplicative form of the overall likelihood (2)**

The multiplicative form of the overall likelihood was chosen because we wanted to give equal weight to performance in dry and wet weather periods thereby dealing with the biased representation of dry- and wet weather days of the calibration period. A few lines about this will be added.

**5. Please check for consistency or typos the following sentences: - Title of Tables 5-6 is the same; - p. 8593, line 26, check “although”.**

Thank you for spotting these typos. Title in Table 5 should be changed to: “Minimum and maximum of posterior wet weather parameter ranges for different numbers of retained parameter sets”. Title in Table 6 is correct.