AUTHORS' REPLIES TO RC3 SUPPLEMENT

General comments

RC3	*The paper describes experimental data sets of urban flooding experiments. The experimental set up covers a 500 m^2 area which provides data at almost a 1:1 scale. Such data sets are not very common and can be of high interest to test urban flood models in conditions close to field conditions. The authors should emphasize this point in their description of the data set. I agree with the authors that such a data set is very valuable to assess the validity of urban flood models.* *If we refer to the scope of ESSD, it is stated that "Any interpretation of the data is outside the scope of a regular paper….. Any comparison to other methods is beyond the scope of regular data." Therefore, it is not expected to see comparisons of the data with numerical models in the paper, nor any complete description or interpretation of the data. However, ESSD expects good quality data sets and the description should be clear and complete enough to provide the relevant information useful for potential users. I believe the authors did not really succeed to reach this last goal for the following reasons:*
Authors:	The authors value the helpful suggestions provided by the reviewer. Thanks to the general and detailed feedback, the manuscript has been modified to provide a clearer understanding of the urban flooding data we are sharing.

RC3.a	*the experimental set up description provided in the paper is not clear and complete enough to really understand the experimental set up and the various components of the experiment. The photo provided in Figure 1 does not provide the location of the sensors and does not allow understanding how the experiment was conducted and how water was flowing. Figure 2 provides a schematic transverse view of the experiment but it gives the impression that it is only a connection of pipes and valves, and does not inform on the lateral extension of the experiment. My point of view is that the authors should provide one or more plan view of their experiment showing the interconnections between the various elements, the 2D extension of their experiment and of water flow and also the locations of the various sensors (and possibly the view angles of the cameras). The authors make reference to the floodX documentation data set that provides plans of the experimental set up, but I believe this information should be provided in the main paper, in a schematic manner, but providing information on the main elements of the experiments (reservoirs, pipes, valves, manholes, etc..), of the location of the sensors and of the main directions of the flow.*
Authors:	*We agree with the reviewer that Figure 1 did not serve its purpose of giving a good overview of the flood facility. A computer rendering of the facility was created to replace Figure 1. With such a rendering, the complete floodable area is visible.*
Changes:	Figure 1 has been replaced with a 3D computer rendering of the flood facility supplemented with labels indicating the most important hydraulic components and sensors. The same

symbols as for Figure 2 have been used, making the two figures easier to interpret together.

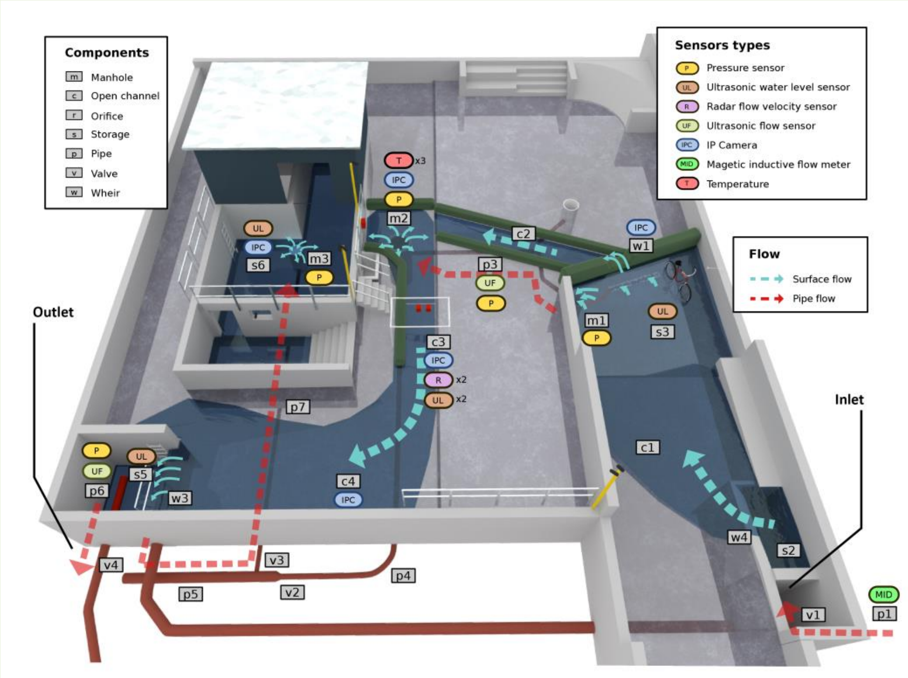

Figure 1: Overview of the floodable area of the floodX flood facility, illustrating main hydraulic flows on the surface (blue arrows), and in pipes (red arrows). The labels indicate major hydraulic components and the placement of sensors, with the IPC labels indicating what component is in the centre of each camera view.

RC3.b *in the abstract the authors mention 37 experiments, but the paper does not provide an overview of the content of those experiments. The reader should refer to the data set documentation for that. I believe that a summary table, providing the main features of the experiments (duration, input discharge, specific configurations, etc.) should be provided in the paper, together with information on the data reliability of each data set (see next point). The potential user of the data should have guidelines to determine which experiment is reliable and which one is relevant for his/her specific needs.*

Authors: We agree with the reviewer that such a table is lacking from the manuscript, so such a table was included in the revised version of the manuscript. Based on the feedback from reviewers 2 and 3, we have defined two high-quality standards for experiments, one being suitable for flood monitoring research and the other being suitable for flood model calibration research as well as the flood monitoring; this information was also included in the table. Additionally, since only 21 of the experiments pass the newly defined quality standard criteria, the abstract has been adapted to reflect these changes.

Changes: A table has been created and added to the section in which the experiments are described.

Table 3: Selection of high-quality experiments conducted, including the duration of the flooding, the total volume of water introduced in the system, and the experiment quality. The experiments are sorted by their total flood volume.

#	Duration	Volume [m³]	Description	Lighting	Experiment quality
29	00:09:50	3.5	small event	night + infrared	calibration
14	00:07:00	3.8	small event	overcast	monitoring
35	00:15:00	3.8	small, empty sewers, manhole cover on m2	overcast	monitoring
16	00:05:00	4.0	small event, full sewers at start	direct sun	monitoring
30	00:05:15	4.2	small event	night + flood light	calibration
26	00:07:00	4.8	small event, full sewers at start	evening sun	calibration
17	00:06:00	5.0	two small events, full sewers at start	direct sun	monitoring
25	00:06:00	5.3	small event	evening sun	calibration
27	00:07:00	6.6	two small events	evening sun	calibration
15	00:07:00	7.0	two small events	overcast; direct sun	monitoring
34	00:20:00	12.8	multiple small, slow events	overcast	calibration
31	00:16:00	24.2	simple dam overflow with pre-impulse	night + flood light	calibration
18	00:15:00	25.3	simple dam overflow	direct sun	monitoring
21	00:17:00	26.5	simple dam overflow with pre-impulse	direct sun	monitoring
28	00:23:00	27.6	simple dam overflow	evening sun	calibration
23	00:19:00	31.4	slow event	overcast; direct sun	monitoring
33	00:20:40	32.2	large event followed by a small one	overcast	calibration
20	00:16:00	33.3	multiple large events	overcast; direct sun	monitoring
22	00:20:40	34.5	large event followed by a small one	direct sun	monitoring
19	00:15:00	44.9	multiple large events	direct sun	monitoring
32	00:22:00	64.1	multiple large events	night + infrared	calibration

RC3.c *the authors also provide the codes used to pre-process the data. They provide the raw and pre-processed data which I find a good point, as the potential users have the possibility to take the raw data and make their own pre-processing if they do not agree with the authors' one. However, the pre-processing data presentation is somehow puzzling for the reader. As the authors are honest and do not hide the problems encountered with their data, the reader is left with the idea that none of the data set is of enough quality. I think the authors should provide clearer information on the data set quality, and if some experiments are not reliable enough, they should consider removing them from the data set.*

Authors: We agree with the reviewers (reviewer 2 had a similar comment).

Changes: First, the table added on the basis of comment RC3.b includes information about the quality of the data and its recommended use, be it for flood model calibration or flood monitoring. Experiments of insufficient quality have been removed from the dataset. Second, we have reviewed the manuscript and removed parts that give unnecessary weight to standard data issues such as temporal shift that we were able to resolve. Third, we have changed the dataset structure to include two distinct data packages, each containing a cleaned selection of data of different quality standards, the first being appropriate for flood monitoring research and the second being appropriate for model calibration research. Finally, we have reduced the number of experiments we "advertise" in the abstract to reflect the number of high-quality experiments only.

RC3.d *the authors explain with quite details, the problems they encountered with their experiments. This is a good point for the future users. However, the presentation is not balanced enough and the reader ends its reading with serious doubts about the interest of the data set. I believe the authors should also spend some times explaining what makes their data set valuable for other users, and what the strengths of their data sets are. Figure 3 is a nice summary of the collected data, but should be commented more in details to explain what can be seen in the figure. In particular the use of the temperature sensors is presented as new, but this would gain being illustrated with some examples. I also believe the OCR treatment of the data loggers is an original application that contributes to the originality of the data set.*

Authors: This comment links to the previous one (RC3.c) with regards to the unbalanced presentation of the weaknesses and potential of the data. We acknowledge the comment and have added a section before the conclusion highlighting possible uses of the data, supported by recent literature.

Changes: The following section has been added to the manuscript:

7. POTENTIAL APPLICATIONS OF DATASETS

Monitoring data for historic urban pluvial floods is typically limited to the (underground) drainage network because most sensors are designed specifically for that setting. Overland flow and accumulation is of much more interest than the drainage network when modelling urban pluvial floods, but the lack of suitable sensors means that flood hydrologists must often calibrate and validate their models with very limited or partial information. The data collected in the floodX project will be used to develop and investigate both image-based flood monitoring methods and flood model calibration schemes that can assimilate non-standard overland flow data. The tools developed in these two lines of research are necessary for using social media images and surveillance videos from real flood events to create more reliable flood models.

A trove of overland flow information lies in existing surveillance infrastructure and social media in the form of images and videos. With appropriate processing methods, flow and water depth could be automatically estimated from these data, and the floodX datasets[1] provide an opportunity to research such methods. For example, the measurement of shallow overland flows with Large Scale Particle Image Velocimetry could be investigated in channel c3 with the two cameras and two radar systems (Figure 1). For the moment, LSPIV has been investigated in urban settings only for large flows and without direct validation data (Guillén et al., 2017; Perks et al., 2016), or for seeded flows (Branisavljević and Prodanović, 2006). Another example is the use of deep learning to estimate flood water levels through semantic scene interpretation, e.g. by interpreting the immersion level of objects of known dimensions in snapshots and videos (Figure 2).

[1] floodX Preprocessed Monitoring Data, floodX Flooding Videos, floodX Flooding Images

Figure 1: View from camera CAM2 in which channel c3 is visible, as well as the scaffolding holding two radar-based flow measurement systems. The same channel is also visible from camera CAM3.

Figure 2: View from camera CAM1 in which a bicycle is visible in the flood water behind the dam. Deep learning could make it possible to automatically estimate the water level from such an image. An ultrasonic sensor above the water and a pressure sensor in manhole m1 provide water level data.

The floodX data can also be used for urban pluvial flood model calibration research since the flood monitoring setup was contained within a hydraulic system comparable to an urban catchment. Thanks to this unique setup and the subset of calibration-quality data[2], it will be possible to test model calibration concepts capable of assimilating the overland flooding data delivered by the novel monitoring methods. The specific questions one may need to address are, for example, the choice of appropriate objective functions and of an appropriate weighting strategy.

In summary, the research made possible by the floodX data will both contribute to urban flood monitoring innovation and improve the reliability of urban flood modelling, thereby increasing the effectiveness of urban flood management services, such as flood forecasting, response, and risk management.

RC3.e *the authors should try to get a final version of the paper only focused on the data set description. So they may provide additional information (such as file naming, sensors drift, ..) in a supplementary material.*

Authors: This point was also brought up several times by reviewer 2 and we have made changes so that the paper focuses more on the data and less on technicalities such as experimental challenges and data processing.

[2] floodX Preprocessed Calibration Data

Changes: Many technical details have been moved out of the manuscript and into the supplementary material. Concerned passages include data logging with OCR (section 4.4.1), preprocessing of time series (section 4.4.2), preprocessing code (section 4.4.3). The discussion of temporal shift is more to-the-point (section 5.1), and we have removed anecdotal information about sensors that is not relevant for the data (section 5.2.2).

RC3.f *at the very end of the paper, the authors mention water balance mass errors of about 20%. This is a big issue that may compromise the usefulness of the data set, and could require further analysis (in particular when putting doubts on the validity of the p6 discharge measurements). Shouldn't it be possible to make additional verifications to see if the proposed explanation may hold?*

Authors: The comment is valid and we have analyzed the data and the facility plans again to scrutinize the proposed explanations. We concluded that the main reason for discharge differences between pipes p1 and p6 had to be coming from the discharge measurement errors in pipe p6. The possibility of leaks is omitted because the volumes involved are too large. The discharge measurement in pipe p1 is judged accurate because it is confirmed by the measurement in pipe p3 and because the measurement conditions in pipes p1 and p3 are much better than in pipe p6. The measurement in pipe p6 was highly constrained because of the facility layout, and despite having installed additional piping, the stabilization distance before and after the measurement was not ideal. Reinforcing this conclusion is the discovery of blatant measurement errors in the discharge data of pipe p6.

Changes: The discharge data measured in pipe p6 has been removed from the preprocessed datasets but was left in the raw data. Because of this decision, the description of the issue has been shortened and mentioned in the context of section 5.3 "Data omitted from the datasets":

> Over all experiments, the discharge measured by the ultrasonic discharge sensor p6_q_us_nivus at the facility outlet deviates substantially from the discharge measured at the inlet of the flood facility in pipe p1. Different hypotheses to explain the volume differences were brought forward, such as residual water in the system and pipe leaks, but investigation of the data and the facility plans led to the conclusion that such factors could not fully explain the discrepancy. The only remaining explanation is that the constrained measurement conditions, especially the short stabilization distance before and after the sensor, the presence of a valve at the end of the pipe, and the frequent changes between full and partially filled pipe, caused the measurement to be erroneous. This conclusion is corroborated by the discovery of artefacts in the discharge data. Since the volume differences are larger than the expected measurement error for the technology (DWA, 2011), the data was judged to be of insufficient quality and removed from the datasets.

RC3.g	*In conclusion, I believe that, provided the authors better document the quality of the various data sets and better describe the experiments, their experiments provide valuable data sets at a 1:1 scale that are of great interest for the science community, in particular for evaluating urban flood models. The data set has also some potential to develop LS-PIV technique. However, in its present state, the data set presentation suffers too many weaknesses for the paper to be published. I recommend major revision of the paper, following the suggestions provided above and below, before possible publication in ESSD.*
Authors:	We would like to thank the reviewers for the time and effort invested in the review of the manuscript. The quality and value of the manuscript and the data has been greatly improved based on the reviewer's comments.

Specific comments

RC3.1	*Section 1.1. The authors may also consider the following references (Bazin et al., 2014; Mignot et al., 2013)*
Authors:	The publications are very interesting but unfortunately do not fit in line with the introduction in its current form.
Changes:	No changes were made.

RC3.2	*P. 2, line 12 "that covers the majority of the components of the hydraulic system": be more specific, which components?*
Authors:	The statement was imprecise and has been changed. Figures 1 and 2 have been updated to give an idea of the sensors' distribution in the hydraulic network. (See next comment RC3.3)
Changes:	The sentence has been changed: "Second, the experiment has a relatively high density of sensors providing information on the major storage nodes and flow channels (Figure 1) so a comprehensive picture of the flooding dynamics can be gained."

RC3.3	*Figure 1, p.3 line 13: the figure is not clear enough and we do not see the elements mentioned by the authors. As mentioned before, one or several plan view of the experiment is necessary for the reader to understand how the experiment was working and the water flow paths. This is all the more necessary that Figure 2 gives the impression that the experiment is only a connection of reservoirs and pipes.*
Authors:	Figure 1 was indeed unhelpful and has been replaced with a 3D computer rendering of the facility and now includes the same labels as Figure 2.
Changes:	See reply to general comment RC3.a.

RC3.4 *P.5 put the section on files naming in a supplementary material, so that it is accessible to the reader without needing to download the whole data set.*

Authors: This information was already included in the "floodX Documentation" package, but can also be provided as supplementary material to the paper.

Changes: This nomenclature information has been removed from the manuscript since it was also questioned by Reviewer 2. A document containing this information will be uploaded as supplementary material with the paper.

RC3.5 *P. 6: section 4.3: I suggest that the authors provide a summary table with the various experiments that were conducted, their main characteristics and information about data quality.*

Authors: See reply to general comment RC3.b.

RC3.6 *Section 5.1: you could illustrate what the images look like by providing some examples.*

Authors: The bulk of the OCR description has been moved to the supplementary material based on a comment from Reviewer 2, which falls in line with general comment RC3.e (that the paper should focus on the data and not experimental issues/data processing). An example image taken from one of the webcams is now provided in the supplementary material.

RC3.7 *Figure 3 is a nice example of the collected data but it is hardly described and analyzed in the paper. The figure should be enlarged so that the authors can highlight the part of the curves that are providing useful information. Such a description would strengthen the paper, by providing an insight on how the data can be interpreted and used to the readers. In addition, the authors mention the novelty of the temperature sensors data, but they do not demonstrate it. An explanation/discussion of the corresponding figure would provide elements to really appreciate the interest of such data.*

Authors: Figure 3 has been removed based on a comment from reviewer 2. However, as proposed by Reviewer 3, a section has been added discussing the utility and potential of the data.

Changes: see reply to general comment RC3.d.

RC3.8 *P.9 line 19 " … was unknowingly moved.." . What does unknowingly means?*

Authors: "Unknowingly" means it was moved without our knowing it at the time, but by looking at the videos and data we could determine a posteriori when and how it was moved. Data from this sensor are not included in the preprocessed datasets newly defined after the reviews, and so this sensor is not discussed anymore in the paper.

Changes: Mentions of this sensor have been removed.

RC3.9 *Figures 4 and 5: the points the authors want to highlight are not very clear*

Authors: This is a valid point also raised by reviewer 2.

Changes: The Figures have been made clearer and the legends made more descriptive.

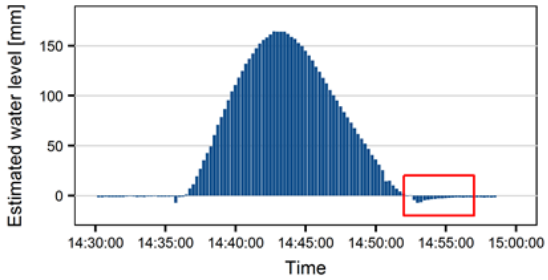

Figure 4: Measured water level in basement s6 by ultrasonic sensor s3_h_us_maxbotix presents abnormally low values after flooding (negative water levels of up to -7 mm can be seen in the red rectangle).

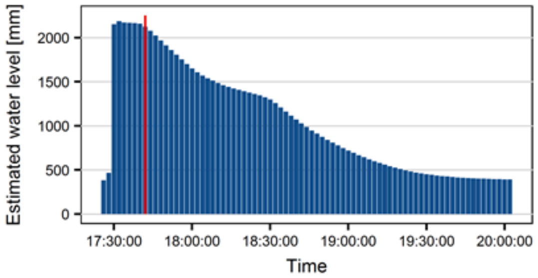

Figure 5: Signs of leakage visible in the water level at manhole m2 (sensor m2_h_p_endress_minilog). The pipe network should not be able to drain but over the course of two hours, starting at the red line, the water level in manhole m2 falls by around 1.7 meters.

RC3.10 *Section 6.4: the authors should provide information on the experiments that were affected by this problem (in the summary table that would be added to the manuscript)*

Authors: Please refer to our answer to general comment RC3.f. The problem is due to unreliable discharge data from pipe p6 and this data has been removed from the preprocessed data.

RC3.11 *Section 6.6: among the difficulties mentioned by the authors, the water balance closure problem is certainly the most critical one. The authors propose several explaining factors, but none of them is fully convincing. In particular, problems with discharge measurements at the outlet are one plausible explanation. But could the authors make additional analysis to see if this explanation really holds?*

Authors: Please refer to our answer to general comment RC3.f.

RC3.12 *Figure 7: I don't understand what the "trend" line in the figure is.*

Authors: The trend is a logarithmic model fitted to the data. This figure has been removed from the paper since the discharge data from p6 is now discarded from the published datasets.

RC3.13 *P.13, line 1: I don't understand what the authors mean with this sentence.*

Authors: Thank you for pointing out the confusing sentence. We meant that for experiment setup and execution tasks that are novel in some way (in that experience is lacking with the method), three times more time should be planned for executing them.

Changes: This point is probably not very helpful since it is very subjective and for this reason has been removed.

References

Branisavljević, N. and Prodanović, D.: Large Scale Particle Image Velocimetry – Measuring Urban Discharge, Vodoprivreda, 38(4–6), 233–238 [online] Available from: http://www.vodoprivreda.net/large-scale-particle-image-velocimetry-merenje-urbanog-oticaja/ (Accessed 18 March 2016), 2006.

DWA: Merkblatt DWA-M 181 Messung von Wasserstand und Durchfluss in Entwässerungssystemen, DWA Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V., 2011.

Guillén, N. F., Patalano, A., García, C. M. and Bertoni, J. C.: Use of LSPIV in assessing urban flash flood vulnerability, Nat. Hazards, doi:10.1007/s11069-017-2768-8, 2017.

Perks, M. T., Russell, A. J. and Large, A. R. G.: Technical Note: Advances in flash flood monitoring using unmanned aerial vehicles (UAVs), Hydrol. Earth Syst. Sci., 20(10), 4005–4015, doi:10.5194/hess-20-4005-2016, 2016.