Review ESSD-2019-187

Very interesting data set, unusual perhaps for ESSD but of potential high value none-the-less.

Data download easily from Zenodo. Good data organisation and formats, easy to open and use both the .dat file and the Python scripts. Good metadata headers in discharge and VWC files but missing (and much needed - you do not want users like me guessing at the data columns) in Arduino and rainfall files. In downloading MacOSX versions I found that changing suffixes from .dat to .csv made files much easier to use in text editors, spreadsheets, GIS software, etc. Consider .csv rather than .dat? Or include a hint for users about changing .dat to .csv?

Overall, with small improvements in metadata, a good useful data product. Presentation however remains weak. Improvements in descriptions and narrative will help many users.

Overall comments:

- The manuscript highlights large areal coverage (e.g. 1 ha) of BGW but in fact the data only cover 3k m^2 (e.g. figure 2 and line 252). The area under measurement here still exceeds prior studies by at least a factor of 5, but advertised 1 ha (10k m^2) while having data for 'only' 3k m^2 seems misleading or perhaps even dishonest?

- The time period of this data set (2018 February to May) misses the usual period of heaviest rainfall for Paris: intense afternoon late-afternoon thunderstorms in mid- to late-summer? Impressive that these authors achieved such high data collection rates (e.g. section 3.2 on times series performance) but do they contend that these measurements cover the full range of precipitation events? If not, they should inform readers about context of these particular months. What would happen (has happened) in heavy (rain rates greater than 20 mm / h) summer rainfall events? Soil / substrate erosion? Aerial flooding? Storage unit 1 fills and overflows to storage unit 2? Ultrasonic proximity/distance sensors in pipe or in storage unit 1 get immersed? Why did the measurements end in May 2018? Particularly curious about this statement at line 381: "this operation is done during a dry period".

- The authors rightly give high attention to retention / detention issues: water storage and run-off delays due to BGW. But, unfortunately, nowhere does a user find hints that these data might actually allow one to calculate retention or detention. Data providers know area, substrate, depth to impervious layer, soil moisture content, rainfall inputs, etc. But they leave it to users to try to calculate e.g. retention? Or they leave the impression that, despite qualify of measurements, one can not actually derive retention / detention? E.g at the time resolution used in figure 4, discharge looks simultaneous / instantaneous with rainfall. The system provides no detention? Or, the data do not allow user to calculate detention. Having raised the issue often and prominently in the introduction and justification, the authors seem remiss to not address whether their data prove relevant to those questions? Give us an example or address what one would need differently or additionally to actually calculate the BGW impacts on retention / detention? We see reference to these values (as outputs from the Python scripts) at lines 357 to 360 but the authors should give us a graphic example with specified uncertainties? Does the system actually produce useful numbers?

Specific comments:

The manuscript needs many small changes / improvements in language. Proofreading will catch many but not all of these errors. I record several specific concerns below. No doubt I missed others.

Line 43 “reaching the network”. I believe the authors refer here to the stormwater management network but - unfortunately - the manuscript displays too many possible terms and explanations: sometimes ‘network’, sometimes ‘rainfall network’, sometimes ‘stormwater network’, rarely ‘stormwater management network’. Settle on and define a standard language, then use it throughout.

Line 140, figure 1: In the upper right the figure lists 32 soil moisture sensors but - at left center of the figure and in text lines 200 to 210 - the authors show and explain use of only 16 sensors. Make 16 sensor the default configuration with parenthetical note or footnote about why 32 sensors seemed to exceed logger bandwidth? Fix sensor number in figure 1?

Line 221: “a nominal range of 250 mm”. Clever to use ultrasonic distance / proximity sensors to measure water height but most ultrasonic sensors have dead zone or null zone close to the sensor face. Data sheets for ultrasonic sensors often specify “little or no dead zone” but more careful analysis suggests working dead zone of 2 cm. This represents nearly 10% of the working range of the UM18. Can manufacturers or authors certify linear response outside of that dead zone out to the maximum range? Have authors in this case relied entirely on manufacturer data sheets? If so, tell the user? Do the ultrasonic sensors, particularly in the pipe or in storage unit 1, get wet or get immersed? What happens then? Why do some file names include the term ‘Arduino’ (which I know well)? Arduino MPU to control the UM18, sending serial data to Campbell data logger? Or, does Arduino refer to the “Unused data coming from a non operational sensor.” Evidently the term ‘Arduino’ applies to storage data but not pipe data? Sensor operated differently or data recorded differently in the two situations?

Line 247 and 252, figure 2: total contributive area of 3511 m^2. See comment above about measured area vs total area.

Line 272 figure 3: If x axis legend of this figure is correct (e.g. Q2 in liter / second) as I think it is, then figure legend (“downstream discharge Q1”) seems wrong? Should read ‘downstream discharge Q2’?

Line 375, 376: “heterogeneousness of the substrate, due to its granular composition and its wavy-form”. Perhaps, but also including sensor-level uncertainties / imprecisions in measuring soil moisture? To the extent “granular composition” and “wavy-form” have an influence, do those represent features of the original BG roof or features that have evolved during time of existence?