

EGU21-4825

<https://doi.org/10.5194/egusphere-egu21-4825>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



How reliable are crowd-sourced data in hydrology? Lessons learned from a citizen science experiment

Julian Klaus¹, David Hannah², and Kwok Pan Chun³

¹Luxembourg Institute of Science and Technology, Environmental Research and Innovation Department, Esch-sur-Alzette, Luxembourg (julian.klaus@list.lu)

²University of Birmingham, School of Geography, Earth and Environmental Sciences, Birmingham, United Kingdom (D.M.HANNAH@bham.ac.uk)

³Department of Geography, Hong Kong Baptist University, Hong Kong, China (kpchun@hkbu.edu.hk)

Crowd-sourcing of hydrological data with volunteering citizen scientists has the potential to overcome severe data limitations in space and time. However, several aspects on the reliability, quality, and value of crowd-sourced data are under debate. In this contribution, we present results of a citizen science experiments involving 300 high school students in Luxembourg. The students relied on self-build rainfall collectors to sample precipitation over selected 24-hour periods covering Luxembourg at national scale (~2500 km²) and subsequently measured the amount. Following data collection and archiving, we evaluated the quality of the data by benchmarking the crowd-sourced values to data collected with a dense network of ~50 tipping buckets across the country. This was done by kriging both data sets. We found that the aerial precipitation at national scale derived from both data sort was consistent, however with a rather systematic bias between the two data sources. The bias was in the same range as the bias between tipping bucket data and average amounts from several measurements with a self-build sampler at the same location. The students' data showed a clearly higher variance compared to the national data but was still able to resolve finer scale variations compared to the national network. We observed the highest differences between both data sets in urban settings. Here, it is not clear if the student's data was less robust when acquired in an urban setting or if the difference arose from urban rainfall processes that were not observed by the national network, where stations are placed at open sites. With our proposed experiment and the statistical data analysis, we were able to quality control crowd-sourced precipitation data and showed that they are reliable. This increases confidence for many studies relying on similar samplers. Yet, some samples of individuals showed a rather high deviation from the kriged national network, showing that sampling with only a few citizen observers could lead to higher uncertainty in the data. While some limitations exist, we showed that data from citizens are of high quality and provide valuable information for hydrological studies.