Response to interactive review comment by F. Nardi (Referee)

This research presents an investigation concerning the information content of stream level classes, potentially observed by citizens and/or using video-cameras, for improving hydrologic modelling performances in ungauged basins. The presented methodology and results show the potential value/capacity of informal hydrologic crowd-sourced observations - as respect to the case where/when high resolution flow monitoring or other standard hydrologic data are available - for monitoring and modelling river channel flows, especially in low contributing area river basins that are nowadays still lacking of adequate monitoring networks, also in developed regions.

The manuscript is well structured, presented and written and the subject/goals of the research, considering the actual importance of the active citizenship topic in hydrology (and not only), is of absolute interest for HESS. Nevertheless, there are some general, yet minor, issues and further few specific comments that I'm inserting hereafter that I strongly suggest authors to consider to improve the readability and clarity of the submitted work .

> We thank the reviewer for the positive assessment of our manuscript and appreciate the valuable comments on how to improve and clarify the text. Below we respond to the three general comments. Of course we will also consider all minor comments in the pdf when revising our manuscript.

General Comments

GC.1) I fully agree with the first reviewer that the description of the calibration methodology is not clear. The performance parameter (Spearman rank), the modeling parameters used while performing the simulation used in the calibration process among others (see specific comments in the attached pdf) should be explained in more detail.

The methodology description relies heavily on referenced works while the reader should be guided in independently following the manuscript without accessing other papers to understand data, methods and results.

> We will add more details on the use of the Spearman rank coefficient as an objective function and rewrite the methods section to better explain the methodology.

GC.2) The characterization of the conversion of stream flow data into classes and the relationship of this crucial step with the stream flow level classification should be also better explained.

> We will explain the conversion of the measured streamflow to the water level class data better. For the first simulations, with two classes, we converted all streamflow values above the median to water level class 2 and all streamflow values below the median to water level class 1. For all situations with more than two classes, we also assigned the classes so that there were an equal number of measurements for each class. See Figure 1 below for three examples of the time series of the measured streamflow and the assigned streamflow classes for the case that there were two, three, or five water level classes.
>
> Later (for Figures 4 and 5 of the manuscript), we changed the class boundaries for the case of two and three water level classes so that there were a different number of data points in each class.

The modeling results are presented only in the form of performance measures and this doesn't allow the reader in understanding the real "information content" of citizen-observed hydrologic monitoring data. Together with comments already introduced by first reviewer and already partially addressed by authors regarding the temporal sampling of flow data in both the monitoring and modeling process, I'd like to add a further major concern I have that is related to the quality/accuracy of the source (informal crowd-sourced) data itself within the proposed research framework. To be more clear: results show that from 4-5 classes and up the modeling performances of the citizen-derived data are or may be "good", but in minor upstream rivers 5 classes of flow levels should be hard to be observed. While I approve the general concept and idea of the presented work I'd like to invite authors to express their view on the practical applications and related issues of the proposed method with specific regard to the issues of citizens gathering 5+ classes of flow level observations in upstream, often inaccessible, vegetation-dense creeks and very minor channels. In this regard a sample picture from a real case study with a visual cross sections representing the potential analysis of the classes or a synthetic figure eventually associated with a flow chart to better depict the authors' view could constitute a solid improvement for this work.

> We agree that in some streams it might be difficult to distinguish five or more water level classes. However, our results show that already two or three classes can be informative and useful for model calibration and also that this is the case regardless of where the exact class boundaries are. While we talk about the water level being above or below a rock, in reality there are often multiple rocks that could be used to determine the water level class (see Figure 2 below). We describe in the manuscript (section 4.1) that it is good news for citizen science projects that two to three classes are already informative for model calibration because citizens are likely able to distinguish between two to three classes but not 20 classes. We will try to stress this point further in the revised manuscript.
>
> With regard to people's ability to observe stream level classes, we want to refer to the 'CrowdWater-Game' as a quick test and demonstration: https://docs.google.com/forms/d/e/1FAIpQLScJ_xYFeYRvBMZEEMoUI3BYddjhpSRRpnW0styvFBJvqg8GTQ/viewform?c=0&w=1. This game includes photos of potential observation sites in Switzerland. The practical issues related to how to observe stream level classes are a central part of our CrowdWater project (see http://www.crowdwater.ch/).
>
> Also, please note that the smallest catchment that is included in our database is 1.2 km^2 and that these are thus well defined channels and not tiny headwater streams for which the water level may only rise a few cm.

GC.3) I understand authors are proposing a novel framework and testing the performances of flow level classes as calibration parameter for hydrologic models gathered from citizen science/data. And I assume the presented synthetic case study doesn't allow to dig into data, but I'd be glad to insert in the manuscript a river flow data/level plot comparing the different curves of hydrologic modeling results built upon the different monitoring datasets (highly detailed/resolution flow data vs citizen data ect). This would also help in addressing GC.1 for better describing the temporal/spatial sampling of parameters and results.

> We want to emphasize that we used real data from 100 catchments, the only synthetic aspect was that we generated stream level class data out of the real high-resolution time series (see Figure 1 below for an example). Given the number of catchments, it is difficult to

show time series (actually this would be difficult already for one catchment). Therefore, we argue that the summarizing assessment using model performance measures is more informative. However, we could include several plots showing observed and modeled responses for selected catchments in the supplementary material.

Specific/Minor comments

See attached PDF

We thank the reviewer for these comments and will address them in the revised version of the manuscript.

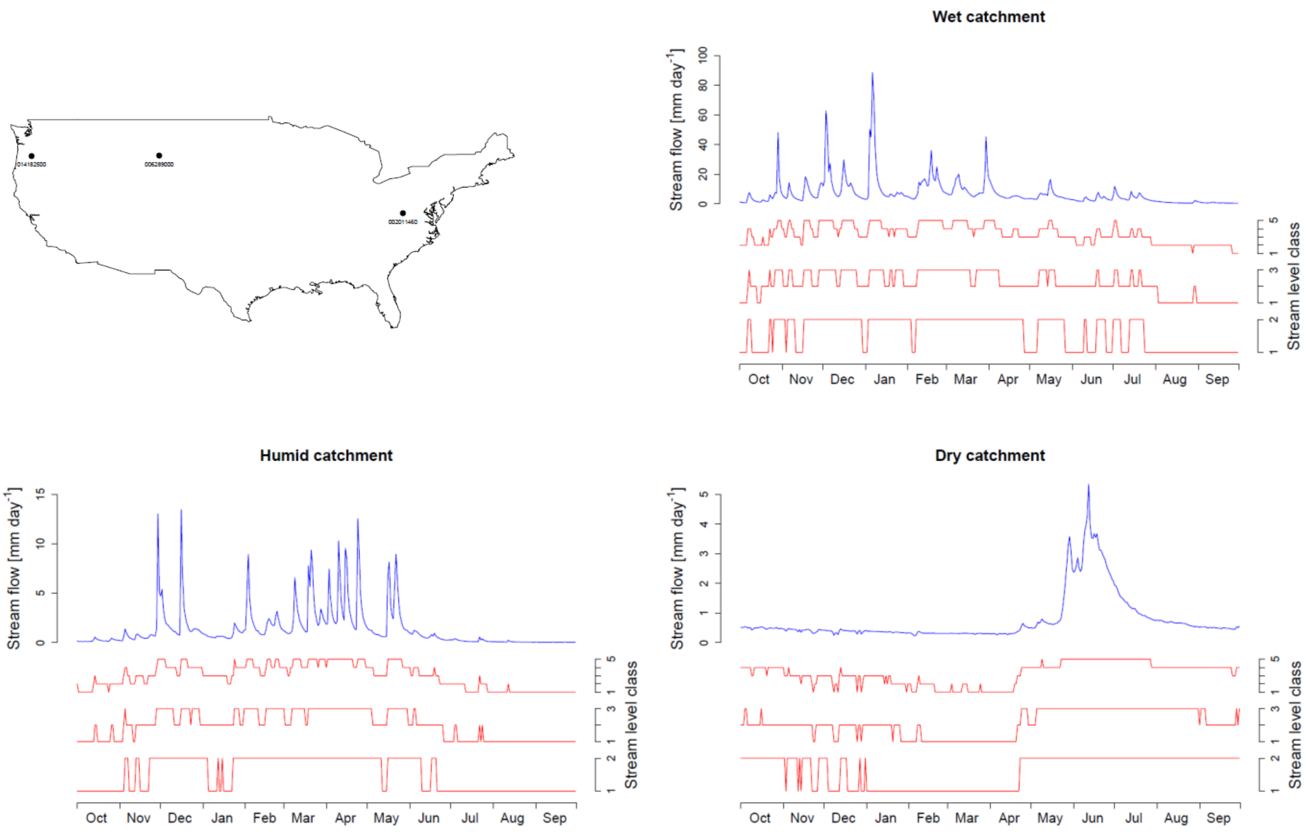

Figure 1. Time series of the observed streamflow (blue) for the first year of simulation (October 1982 – September 1983) for three median sized catchments (wet: 292 km^2; humid: 235 km^2; dry: 472 km^2) and the derived time series of stream level class for the case of two, three and five level classes (red), where the stream level is in each water level class for, respectively, 50%, 33% and 20% of the time. The inset shows the location of the three selected catchments. The aridity index of the catchments is 2.94 (humid), 1.33 (wet) and 0.71 (dry). Note that the scale for the observed streamflow is different for the three catchments.

Figure 2. Pictures of streams showing that multiple features can be used to define two to five stream level classes. For more pictures we refer to the CrowdWater game (https://docs.google.com/forms/d/e/1FAIpQLScJ_xYFeYRvBMZEEMoUI3BYddjhpSRRpnW0styvFBJvqg8GTQ/viewform?c=0&w=1).