Responses to Reviewer #2's Comments:

Please see general responses to your helpful comments below in blue (original comments in black). Once all the edits (per the details below) are finalized, a marked-up version of the manuscript showing all changes, along with specific responses to reviewers' comments will be provided.

The paper presents results on three simple and easy to use discharge estimation methods appropriate for citizen science (SC) that the authors applied in the Kathmandu Valley, Nepal. They assessed the agreement between the methods and compared estimated discharge to selected measurements using a doppler radar device. The text is short but mainly well written, the graphical presentation is clear and appealing. I recommend to state explicit research questions at the end of the introduction (currently missing). I have also some major concerns about parts of the analysis and the interpretation of the results:

1. While the authors do well in terms of reporting statistical significance of their results, the use of the Pearson Correlation Coefficients seems not appropriate for the properties of the dataset. I therefore recommend the non-parametric Spearman Rank Correlation and the associate non-parametric statistical test.

RESPONSE: Following the suggestion of Reviewer #1, we will remove the correlation analyses and will present the information in a tabular summary with average absolute error. For the newly collected data (see response below point 4 below), we will present the information as box plots showing the distribution of error for each site.

2. I question whether it is meaningful or informative to correlate the slope of the salt dilution calibration k to latitude or longitude and elevation and would suggest to skip (or better explain this analysis).

RESPONSE: This analysis has been removed.

3. Instead I recommend to also show the comparison of discharge estimated by salt dilution and by the Bernoulli method.

RESPONSE: The new box plots show these comparisons between float, salt dilution, and Bernoulli in a more comprehensive way.

4. I would ask the authors to quantitatively prove that they can compare discharge estimates taken during the CS-campaign with doppler radar observed discharge taken +/- one month (!!) before/after the campaign or skip that part. In the discussion they state themselves that the flow might have decreased during that time.

RESPONSE: Yes, the lack of timely reference flow measurements was a significant challenge with the initial data set. It was timely to receive your comments when we did because we were just starting our post-monsoon Citizen Science (CS) Flow campaign in Kathmandu. Based on your comments, and

Reviewer 1's feedback, we were able to design and implement additional data collection which was performed from 18 to 20 of September 2018. We feel that once these data have been incorporated along with the other suggested edits the revised manuscript will be strengthened.

From 18 to 20 of September 2018, we facilitated measurements at 15 sites in two different watersheds in the Kathmandu Valley. Ten CS Flow groups, each comprised of three students, performed all three methods (i.e. float, salt dilution, and Bernoulli) at each site. At the same time, an "expert" group (authors) performed the same three methods at the same sites, along with a FlowTracker ADV reference flow measurement. After the field measurements, all the CS Flow participants completed a survey about their experiences with (and perceptions of) each simple measurement method.

As the remaining analysis is probably too short for a full publication, I suggest the authors to check whether their dataset would allow additional analysis e.g., on the difference of the quality of the measurements taken by experts and citizens (see my suggestions in the pdf). The current paper is interesting, the dataset promising but the current state of analysis is not enough for a full publication. I therefore encourage the authors for major revisions and additional data analysis.

RESPONSE: As stated above, we have been able to collect additional data that we believe overcomes the challenges associated with the initial dataset.

In the following I summarize my suggestions for the individual sections and ask the authors to also check my detailed comments and suggestions that I have included in the pdf (uploaded as supplement):

1. Introduction: The introduction is on the short side and starts a bit philosophical. I would focus more on streamflow and introduce to the problem that large parts of the words still have limited number of gauging stations (especially remote and developing countries) and that measuring devices - while still decreasing in costs - have their limitations. Some other citizen science studies are briefly mentioned but the findings of other studies could be described a bit more in detail. The same applies to the existing methods for low-cost streamflow assessments. Their pros and cons could be compared using a table. I also do not agree that the are no tools on the market, that allow direct measurements of discharge with smartphones and added one link to an example. The research questions should be clearly formulated at the end of the introduction. Please also see more specific comments directly marked in the pdf (uploaded as supplement).

RESPONSE: We have removed the first paragraph. We also describe in more detail the problem of limited gauges, high costs, and other limitations. We have strengthened our description of the findings of other studies. The pros and cons of low-cost streamflow methods has been summarized in a table in the Introduction. A reference to the other project using smartphones for discharge measurements has been included in this summary table. The research questions are more clearly stated at the end of the introduction. Your helpful and detailed comments in the supplemental material have also been addressed in the revised manuscript.

2. Methods: The method section needs a better description of the experimental setup; study area, test with students, repetitions y/n etc. (parts are mentioned at the end of the method section but should be stated at the beginning)! The catchments and streams used for testing need a better description (see my suggestions in the pdf). The same applies to the training of the students. The explanation of the different methods is long but can be useful for some non-hydrological readers. I suggest to consider to present all this information in the introduction section. I would however include a list of objective criteria why these three and not other methods have been selected. Please also see more specific comments directly marked in the pdf (uploaded as supplement). The method section should be clearer about the two datasets collected a) dataset with n=20 samples (I assume collected by the authors themselves = exports) and the CS-campaign with n=145 samples collected by citizens. One issue seems critical to me: Authors compare observed discharge using the doppler radar with CS-discharge measurements done +/- 1 moth earlier/later. The authors should prove statistically that the mean daily flow in the month before and after the CS-discharge measurements is not significantly different. In fact authors state in the discussion section that flows decreased during that period.

RESPONSE: Based on your comments, we have updated the methods section. We have improved the description of the catchments and the training of the citizen scientists (CS). The explanation of the methods has been moved to the introduction. The objective criteria has been included in a flow measurement method summary table in the introduction. An additional table has been added to clarify the three phases (and datasets) of the study: (1) initial evaluation (authors), (2) citizen scientist evaluation (authors and CS Flow groups), and (3) citizen scientist application (CS Flow groups).

#	Phase	Description	Performed By	Period	Season
1	Initial Evaluation	Initial evaluation of three simple flow measurement methods (i.e. float, salt dilution, and Bernoulli) along with FlowTracker ADV reference flow measurements at 20 sites within the Kathmandu Valley. Reference flows ranged from 6.4 to 240 L s^{-1}.	Authors	March/ April 2017	Pre-monsoon
2	Citizen Scientist Evaluation	Citizen Scientist evaluation of three simple flow measurement methods (i.e. float, salt dilution, and Bernoulli) along "expert" and FlowTracker ADV reference flow measurements at 15 sites within the Kathmandu Valley. Reference flows ranged from 4.2 to 896 L s^{-1}.	Authors for "expert" and reference flows PLUS 10 Citizen Science Flow groups for simple methods	September 2018	Post-monsoon
3	Citizen Scientist Application	Salt dilution measurements at roughly 150 sites in the 10 perennial watersheds of the Kathmandu Valley. Float measurements with a small number of sub-sections (e.g. 3 to 5) performed at each site to determine salt quantities.	17 Citizen Science Flow groups (7 from April and 10 from September)	April and September 2018	Pre and Post Monsoon

The comparison of measurement methods has been edited, so that only measurements taken within +/- one day of each other are used, and then only if precipitation didn't occur.

3. Results: Graphical presentation of the results is good and I appreciate that the authors report about statistical significance of their results. For some of the dataset I suggest to used the Spearman Rank Correlation Coefficient as the assumptions for using the Pearson Correlation seem to be not fulfilled. I also suggest to mention that, while statistically significant, some of the relations show relations difficult to interpret (definitely not linear or exponential but complex or clustered). I ask the authors to explain why they think Figure 2 is informative to the reader expect for presenting the measurements. I question whether it is meaningful or informative to correlate the slope of the salt dilution calibration k to latitude or longitude and elevation and would suggest to skip this analysis. Instead I recommend to also show the comparison of discharge estimated by salt dilution and by the Bernoulli method. As mentioned in the method section I would ask the authors to quantitatively prove that they can compare discharge estimates taken during the CS-campaign with doppler radar observed discharge taken +/- one month before/after the campaign or skip that part. Please also see more specific comments directly marked in the pdf (uploaded as supplement).

RESPONSE: Based on comments from Reviewer 1, the scatter plots have been removed. We have also removed Figure 2. The K correlation analysis (Figure 5) has been removed. Yes, the time difference between CS Flow group and verification measurements was a significant challenge in the initial dataset. The results from the newly collected dataset allow for a more confident comparison between the results from the simple flow measurement methods and the reference (or true) flow. The remainder of the detailed and helpful comments in the supplemental material have been addressed.

4. Discussion: The discussion is short and not very into depth. Parts of it would better fit into the result section. While the background concentration is certainly affected by the geology I have strong doubts that correlating k with latitude and longitude or elevation is meaningful. At least better explain why the authors think these predictors are meaningful and the correlations not spurious. I suggest the authors to check whether their dataset would allow additional analysis e.g., on the difference of the quality of the measurements taken by experts and citizens (see my suggestions in the pdf).

RESPONSE: The discussion section has been revised, and parts have been moved to the results section. Additional details have been included in the discussion. The K correlation analysis has been removed. The new dataset has been used to evaluate variability in CS Flow group measurements with box plots, as compared to both “expert” measurements using the same method, and reference flows from the FlowTracker ADV.

5. Summary and future work Is well written in general. However, addressing the outcome of this work in the light of research questions (that I suggest to include in the introduction) would improve this section.

RESPONSE: The summary now returns to the original research questions stated at the end of the introduction.

I hope these suggestions are useful to work on a more advanced version of the manuscript!

RESPONSE: Indeed, these comments, along with your detailed supplementary comments have been extremely helpful towards improving the quality of this manuscript! We deeply appreciate your sincere efforts and investment of time.

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-425/hess-2018-425-RC2-supplement.pdf

RESPONSE: As mentioned earlier, a marked-up version of the manuscript showing all changes, along with specific responses/edits based on your comments provided in the supplement will be provided once all the edits are completed and the newly collected data is incorporated.