Department of Integrated Water Systems & Governance

Thaine Herman Assumpção
t.hermanassumpcao@un-ihe.org

IHE Delft Institute for Water Education
Westvest 7
2611 AX Delft
The Netherlands

Author's response – HESS-2017-456

Date:
04 January 2018

Dear Editor,

Thank you for the evaluation report and for giving us the opportunity to revise the manuscript. We would like to thank the referees for taking time to review the manuscript. Their comments delivered insightful and enriching recommendations on how to improve the manuscript's content, scientific quality and readability. We also appreciated the comments made by a reader who was not a referee. A revised manuscript was prepared based on the comments we have received.

In this response letter we present the response, as follows: Section 1 contains the explanation and responses to all points raised by reviewers; Section 2 contains a list of all relevant changes made in the manuscript; Section 3 contains the revised manuscript with track changes; and Section 4 contains the revised manuscript text without track changes, for easy reading.

We hope the comments and requirements for publication are met in the revised manuscript. In case there are more concerns, please let us know for further corrections and improvements.

We are looking forward to your decision.

Yours sincerely,

Thaine Herman Assumpção and co-authors

1. Authors' responses to reviewer's comments

The response to all comments raised by the referees, reflected in the interactive discussion section, are presented here in detail. The page and line numbers in the authors' response refer to the revised version of the manuscript, the one that contains the marked-up changes (see Section 2 of this document).

Note: Modifications and additions to the response, as compared with the one in the interactive discussion, are highlighted in the authors' response, as underlined text.

Anonymous Referee #1 – RC1

We thank the reviewer for taking the time to review this paper and for providing useful feedback. Your input is valuable in improving the scientific quality of the paper and its readability. Please find below our answer addressing your comments.

Comment #1: *Dear editor, I went through the paper entitled "Citizen observations contributing to flood modelling: opportunities and challenges" by Assumpcao et al. Bringing people's idea and their involvement in science (citizen science) is becoming significant globally. This paper is exactly what lies behind the role of citizen science in combating the flooding by modelling. However, I find the paper is quite difficult to follow in its current form. This also has no such in-depth assessment of the role of science in mitigating climate-induced flood events/hazards.*

Authors' response: We acknowledge that the assessment of the role of citizen science in mitigating climate-induced flood events/hazards is not addressed in the present article and that is because the focus of the paper is different, particularly it is to review the existing scientific literature regarding the actual and potential crowdsourced data for flood modelling. From that perspective, climate-induced flood events/hazards do not bring different challenges for citizens' data collection compared to "regular" flood events. Of course citizen science is much broader than only crowdsourcing of data, but such broad perspective is outside of the scope of this article. Regarding mitigation, in the article's Introduction (page 2), we are mentioning the review of Horita et al. (2013); and the studies of Dashti et al. (2014) and Oxendine et al. (2014). They are addressing disaster management and damage data collection, including the role of citizen science for mitigation of floods in general. In the present article the analysis is made for model improvement, but the model may have multiple purposes (e.g. flood risk or ecosystem conservation). The paper determines what are the benchmarking difficulties and benefits of collecting flood-related data by citizens and of integrating them into models, for the purposes of model set up, calibration, validation, simulation and forecasting.

The improved explanation emphasizing the aim of the review is added in the Introduction section on page 3, lines 1-5.

Comment #2: *The synthesis/review would have been much useful and interesting if this were focused on one or two key objectives. For example, how citizen science would link to model building process based on crowdsourced data and how citizens themselves would be benefitted provided the feedbacks for the model improvement.*

Authors' response: The approach taken was to group and analyse the studies in which crowdsourced data was integrated into each part of the flood modelling process. We could not take a different approach because unfortunately the literature on specific parts is scarce (e.g. in Table 5, Page 19, we

found just 6 studies on the model building process). Hence, the review could not be limited to few particular aspects. Similarly, although a paper on citizen's benefits from model improvement would be useful and interesting, this is a recent topic that has not been explored enough and there are not enough publications to date so that a review is required or can be made.

Some specific comments:

Page 2 Line 10-15:

Comment #3: *what are the valuable contributions? elaborate*

Authors' response: As suggested we further elaborated, on page 2, lines 15-19, of the revised manuscript. For example, the CITI-SENSE project managed to simultaneously collect perception data and acoustic measurements in an approach that can be used to develop citizen empowerment initiatives in case of noise management (Aspuru et al., 2016).

Page 2 Line 22-26:

Comment #4: *what are three projects? provide the summary*

Authors' response: The manuscript was changed to include such a summary, on page 2, lines 31-32.

Page 4 Line 19:

Comment #5: *please define 'CAPTCHA plug in framework', not all readers would necessarily know about it*

Authors' response: A footnote was added to the manuscript in order to clarify the concept of a CAPTCHA plugin (page 6, footnote 2):

"CAPTCHA stands for 'Completely Automated Public Turing test to tell Computers and Humans Apart'. It is a test evaluating if the subject is human, which is used in websites to provide security. After the test is done the user can be asked to perform extra tasks, for example, tag images."

Page 10 Line 12-17

Comment #6: *what level of citizens will get involved to generate data globally as many citizens are devoid of IT technology?*

Authors' response: Iwao et al. (2006) did not provide any information on the profile of citizens, nor on engagement strategies, although the lack of data in certain regions was shortly addressed. However, as stated in the Citizen Science section of the manuscript (page 3), the review did not discuss the mechanisms of citizen engagement and participation, as this is a research topic on its own and we focus on data integration. To address this issue, also raised by a comment of a reader in the HESSD interactive discussion, explanations were added on page 4, lines 20-22.

Page 15, Fig. 6:

Comment #7: *perhaps Fig. 6 holds the core concept of the paper, where the citizen science link to modelling and its application*

Authors' response: Though the figure is a core concept of the paper, the paper structure is such that first the wider scope of the paper is defined, laying all the literature that has the potential to contribute to flood modelling in terms of flood-related data. This literature is characterized and analysed for advantages and disadvantages. Then, it is presented an in-depth analysis of the scientific contributions to each part of the modelling cycle. The existing literature is evaluated in terms of its information

content and analysed to check how much it matches model requirements. Finally, opportunities and challenges are identified. Following this structure, Figure 6 is presented in a later section.

Page 18 Line 23:

Comment #8: *please provide what consequences of uncertainty in data mining and how this is improved?*

Authors' response: The consequence of uncertainties, including the ones of data mining, is low model performance. We consider that the higher the uncertainty, the lesser the quality of the data. Hence, because data obtained through data mining has, in general, more sources of uncertainty (from value, geotagging and timestamping), they can potentially be of lesser quality and result in models with low performance. As suggested by a reader, this was further extended in the new version of the manuscript on page 21, lines 7-8; and on page 22 lines 15-16.

To date, in modelling studies, there are only few studies that quantify the uncertainty from crowdsourced data, the impact on model performance or that consider methods for its reduction. To remain neutral, we did not include in the manuscript anything beyond what is in the literature, thus we do not include a discussion on how to improve the situation in modelling.

Anonymous Referee #2 – RC2

We would like to thank the reviewer for the revision. We appreciate the comments provided, that deliver insightful and enriching recommendations on how to improve the content of the paper. We have addressed your comments individually in the text below.

Comment #1: *This paper present an interesting and fairly complete review on the use of crowdsourcing for flood modelling purposes. The effort to try and characterise the reliability and uncertainty associated to different types of data and different methods of involving citizens in collected them is worth highlighting. I would limit my review to three general comments: (1) There is no mention in the paper of the diversity of models that are used for flood modelling, and whether they are more or less suited for integrating the different types of citizen observations. Arguably, one of the challenges for hydrologists could be to design models specifically for that purpose. At least, it would have been interesting to have some information of the kind of models used in the studies analysed in the paper.*

Authors' response: The manuscript was modified to include an explanation on types of flood models (fluvial, pluvial, coastal and drainage) on page 17, lines 8-14. The matter of suitability is not addressed, mainly because the considered papers are not addressing the suitability. However, we found this comment very valuable and we added more information on the kind of models used in the reviewed studies (page 19, table 5).

Comment #2: *(2) the question of time is only very briefly discussed, while in flood modelling, and particularly for real time flood forecasting, this is an critical issue: models not only require the highest water level or the maximal flooded area extension (which are, I guess, when most of the observations are done), but high resolution data during the rising part of the hydrograph. What have been done to collect this information, and/or what type of participatory approach should be organised to do so?*

Authors' response: In the section 'Crowdsourced data information content' on pages 21-22, we discuss the question of time within each part of the flood modelling cycle. Flood forecasting is not included because citizens cannot provided forecasting data. We acknowledge that we do not consider

calibration and validation for specific purposes and thus do not consider them done specifically for obtaining an operational model for flood forecasting. <u>With that in mind and in view of the reviewer's comment, we changed on page 21, Table 6, in the column 'Calibration Validation', the temporal coverage to 'Discrete/Continuous' and the spatial coverage to 'Discrete/Distributed'. A remark was added to the table mentioning that the data properties for calibration and validation depend on the purpose of the model.</u>

<u>Moreover, the discussion was extended to accommodate such view and answer the question on what has been done to collect this time sensitive information (page 21, lines 23-26; page 22, lines 1-3).</u> Organisation of participatory approaches are not discussed as they are outside the scope of the proposed article.

Comment #3: *(3) In the same line, rainfall is almost absent in the discussion. As far as I know, crowdsourcing have also been used to obtain spatially distributed rainfall, and many extreme storm events are characterised by a high spatial variability of rainfall, so I suspect that this type of citizen observation could be useful.*

Authors' response: <u>We agree that contextualization of the rainfall component is lacking and this was added to the manuscript (page 7, lines 7-11). We mentioned its importance for certain types of flooding and provided pointers to articles on crowdsourced data for rainfall. We acknowledge that citizen contributions could be useful for observation of this variable, however, we did not include rainfall in the flood-related crowdsourced data section because it was partially covered by the review of Buytaert et al. (2014) and totally covered by the review of Muller et al. (2015).</u> Rainfall is a variable of greater importance for hydrological models, whilst the review focusses on a hydrodynamic representation of floods.

Anonymous Referee #3 – RC3

We thank the reviewer for providing feedback on the quality of the paper. The review is valuable for making the paper clearer and more structured and the comments are highly appreciated. Please find below our response to the provided comments.

Comment #1: *This paper addresses a very timely and interesting topic: citizen science and its use in flood modelling. It will provide some guidance to researchers struggling with the lack of traditional data and at the same time resistant to adhere to alternative data sources. Overall, the text is rather fluid and well written, but in topic 3, "crowded source data in flooding modeling", the explanation of some uses of citizen data in modeling is confusedly described and could benefit from a restructuring of description of uses.*

Authors' response: <u>Following the reviewer's suggestion, the description of uses in Section 3 was restructured (page 19, lines 6-7; page 20, lines 1-32; page 21, lines 1-4).</u>

Comment #2: *Also, despite the relatively large number of papers gathered, the revision process and papers selection is not fully described. Thus, for a synthesis paper, it will be worth proving a perspective on how exhaustive were the efforts undertaken in the collection and selection of relevant studies, and the data sources consulted.*

Authors' response: <u>The manuscript was extended to inform that the papers' collection was done through multiple platforms (e.g. Scopus and Google Scholar), exemplifying used keywords (page 3, lines 7-14). Additionally, explanation on the selection criterion for consideration was given, which is</u>

the generation/use of flood-related crowdsourced data, as well as explanation on why certain articles were not selected.

A few minor points include:

Comment #3: *In Figure 1, only level one is termed crowdsourcing, not level 2, as stated in the text (page 3, lines 30-31).*

Authors' response: The sentence was rephrased (page 4, lines 22-24).

Comment #4: *It is not clear how the CAPTCHA plug in works as a volunteered contribution; please provide a better explanation.*

Authors' response: Clarification regarding the CAPTCHA plug-in was done by means of a footnote, as also requested by another reviewer (page 6, footnote 2).

Comment #5: *Figure 2 does not seem relevant, I suggest excluding it; while Figure 6, in its present form, does not seem very informative.*

Authors' response: Figure 2 was included as an introductory example of framework for analysing crowdsourced data. We acknowledge that it does not attend other purposes in the previous version of the manuscript. As per suggestion of a reader that commented on HESSD interactive discussion, we included a modified version of Figure 2 further in the text, changed to include the reviewed literature (page 14, figure 3). The motivation behind increasing the relevance of such a figure is two-fold: exposition to the interested reader of classification systems of citizen science approaches; connect at a superficial level with social studies that evaluate these classifications, to increase the integration among disciplines.

Figure 6 presents visually two types of information: the components of the flood modelling process and the data necessary for each component; citizen contributions within the process. We consider that the first type of information is essential for scientists in the field of citizen science that do not have a background in modelling (but that can, for example, research data collection methods to address modelling needs). The second type of information is an essential component of the manuscript and, although described via text, making it explicit visually fulfils the objective of highlighting it in the paper. We are open to suggestions on how this image could be enhanced.

Comment #6: *I suggest merging Section 1.2 - Article outline with the end of the Introduction (page 2, line 30).*

Authors' response: In HESS interactive discussion we said we would consider this suggestion. The outline has been merged with the end of the Introduction (page 3, lines 16-22).

Comment #7: *There are some unnecessary wording throughout the paper, for example: "We have seen in the previous section that" and "In this section we intend to" (page 14, lines 4-5).*

Authors' response: Thank you for the suggestion, the paper was thoroughly scanned for unnecessary wording and changed accordingly (page 17, lines 4-7, page 18, lines 7-8).

Anonymous Referee #3 – RC4

Thank you for the follow-up feedback. Please find below the answers to your comments.

Comment #1: *The issue concerning the CAPTCHA plug in is not really about its definition, but about HOW it will be used as a "volunteered contribution". How, for example, random images of a deforested area chosen for security reasons will contribute to the monitoring of land use?*

Authors' response: We clarified the manuscript's text by saying that the process of tagging images for land use is uncorrelated to the CAPTCHA, to the test of distinguishing computers from humans. Tagging is a task performed after the test, on the same platform (page 6, lines 2-3).

Comment #2: *Additionally, despite the content of the outline, it makes more sense that it comes at the end of the introduction, providing readers with an initial and general idea of what will follow. The current topic 1.1 could come together with the description of data sources used and papers selection (to be included), as part of a methodology section.*

Authors' response: In HESS online discussion, we proposed to keep section 1.1 and to create a section "1.2 Review approach". Upon revision of the manuscript, we realize that the review approach could be summarized in a paragraph and that there was no need for a separate section. Thus, as mentioned in the response to the previous comment, we have added to the end of the introduction the review approach and the article outline (page 3, lines 7-22). We would like to reiterate here, and strengthened this in the text of the manuscript (page 3, lines 1-3), that the intention is not that citizen science is the focus of the manuscript, but the data obtained from it, thus maintaining its discussion within the introduction section.

Interactive comment – SC1

Thank you for finding the paper timely and for the appreciation of the review paper. Authors would like to thank M. Moy de Vitry for taking time to review the paper and add to the ongoing discussion. The comments and suggestions received are of high value, and based on them we made improvements to the manuscript. Please see below the answers to the comments.

Comment #1: *The review of how citizen observations have been used in flood modelling research is useful and very timely. The main value of the review is in mapping out the different case studies, identifying trends, and pointing out research gaps. Minor revisions are recommended:*

Page 1 line 27:

Comment #2: *Do the authors refer to the general need for data in modelling, or specifically to monitoring data used for calibrating the models?*

Authors' response: Authors are referring to general data needs for modelling floods, no special distinction for calibration is made. Thank you for pointing out the confusion. In order to clarify this issue to the reader an additional statement was added to the manuscript (page 1, lines 28-29).

Comment #3: *The example in the second sentence "This is especially true.." requires some explanation.*

Authors' response: More explanation was added, by rephrasing the sentence (page 1, line 29; page 2, line 1).

Page 3 line 26:

Comment #4: *Effort is made to present two classification systems. However, these classifications are not used in sections 2 and 3.*

Authors' response: These classifications are not introduced for the purpose of further classifying other papers, but for opening the discussion and debate on the existing reviewed literature. The first classification system (i.e. level of engagement), aims to explicitly say that discussion on advantages/disadvantages of collection/analysis methods, as well as their purposes, is strictly addressing contributions in terms of quantitative data (i.e. contributions towards flood modelling); and it does not address the advantages/disadvantages of contributions from other types of involvement. For example, it is out of the scope of the article to discuss tacit knowledge or social media mining having the (possible) disadvantage of not fostering awareness. For further clarification, the new version of the manuscript was amended (page 4, lines 20-22).

The second classification system was made to provide a reflection of such components (implicitly/explicitly geographic and implicitly/explicitly volunteered) when data is obtained from citizens. Based on this and a follow-up comment, we added a Figure where we place on the framework the studies cited in this paper; and we provided an analysis of such result (page 13, lines 12-16; page 14, figure 3).

Page 4 line 10

Comment #5: *It is unclear why geo-tagged information is not explicitly geographic.*

Authors' response: For clarity, in the beginning of such paragraph an explanation was added (page 5, lines 5-6).

Page 4 lines 15-20:

Comment #6: *It does not seem appropriate that SCENT is given a prominent position in this review paper, which should review published literature and not ongoing projects.*

Authors' response: As mentioned in the acknowledgements, this review and research related to it are supported by the H2020 project, SCENT. Therefore, it is natural that the ideas generated within the project, which aims at covering scientific gaps, are properly acknowledged in the paper text as well. The inclusion of SCENT has as objective to illustrate the classification system, taking advantage of the fact that in the project the four classes are being covered. For clarity, it was not chosen to include published literature in this part of the article without analysing it first. As per suggestion of the reviewer, we presented the same scheme later on, where such literature was included (page 14, figure 3).

Page 5 Figure 2:

Comment #7: *Fig 2 illustrates nicely how specific examples are classified within Craglia et al.'s definition, and therefore more examples would be beneficial. It would be even better if the examples were taken from literature.*

Authors' response: Thank you for this suggestion, we took it into account and expanded in the second version of the manuscript (page 14, figure 3).

Comment #8: *SCENT should be removed from the figure.*

Authors' response: The justification of SCENT's inclusion in the figure has been provided in a previous comment. This figure sets the scene for the second one that was added based on the reviewer's suggestion.

Comment #9: *it is unclear why the CAPTCHAs are neither implicit nor explicit.*

Authors' response: In the image CAPTCHA plugin is both implicit and explicit. The text was modified for clarification (page 6, lines 1-2).

Page 6, lines 1-2:

Comment #10: *Have studies such as Merkuryeva et al. (2015) been included in the review? please specify.*

Authors' response: No, they have not been included. The text was modified for clarification (page 3, lines 11-14).

Comment #11: *The citation is not necessary.*

Authors' response: We acknowledge that the citations do not serve a purpose other than being examples. However, as a review paper, we consider that different aspects of the literature should at least be exemplified, in a way that the interested reader may wish to explore topics not covered in the review.

Page 6, line 18-20:

Comment #12: *It is unclear why the text example is provided in the same paragraph as the images/videos and not in the previous paragraph.*

Authors' response: Thank you for bringing up this misunderstanding. The text examples are related to non-quantitative text that is converted to quantitative measures. As the section's first paragraph is about quantitative crowdsourced data and the second is about qualitative ones, this information fits better in the second paragraph. For clarity, the second paragraph was modified (page 7, line 26-31; page 8, line 1).

Page 7, Table 1:

Comment #13: *It would be good to split the column 'case study' into two columns 'location' and 'flooding type'*

Authors' response: The columns were split in all tables into 'Flood Type' and 'Location' and studies with flood-related crowdsourced but without floods will be classified as 'No flooding'.

Comment #14: *What ordering is used in the table? publication year might make sense.*

Authors' response: The ordering used in the table was done by grouping papers with similar measurement/analysis methods, followed by the order monitoring, mapping and modelling. This is similar to the way the analysis is done.

Page 12, Figure 3:

Comment #15: *The review extends to April 2017 - has the publication count for the year 2017 been normalized?*

Authors' response: No, it has not been normalized. We understand your reasoning, but our focus is on the content and interpretation, not on a precise, numerical analysis of the contributions. Thus, because of the small numbers of contributions per type of publications, for simplicity, we decide not to normalize.

Page 13, line 7:

Comment #16: *Flickr and Picasa are products, it is better to refer to photo sharing services.*

Authors' response: <u>We changed to the proposed terminology.</u>

Comment #17: *what is exactly meant with 'mining', and how does that entail low-quality data?*

Authors' response: Mining refers to the extraction of specific data from a dataset. For example, tweets can be mined from Twitter for a certain period of time and for tweets that contain the word 'flood'. <u>We expanded the first appearance of this term to include such qualification and make it clearer (page 8, lines 3-4).</u> Crowdsourced mined information has the possibility of not having a precise time-stamp or geotag. Thus, there is uncertainty related to it. We consider that the higher the uncertainty, the lesser the quality of the data.

Page 18, lines 19-25:

Comment #18: *The discussion on reliability and volume of data is interesting and necessary, but the statements do not seem to make good of the review that was conducted. Do none of the papers attempt to quantify uncertainty?*

Authors' response: Yes, some of the papers do. <u>We expanded this discussion to include more information (page 22, lines 15-22).</u>

Comment #19: *Is the volume of data per type (water level, velocity, etc.) available comparable to the number of case studies?*

Authors' response: We have not computed the volume of data for each data type. At the moment we estimate that they are directly proportional to the number of case studies. Unfortunately, it is not possible to get the exact number as in some cases more than one variable is collected and no distinction in the overall count is provided.

Page 19, lines 20-26:

Comment #20: *The language used is imprecise*.

Authors' response: <u>The language was rephrased</u>. See below.

Comment #21: *"interactions between citizen science and water resources"*

Authors' response: <u>It was rephrased (page 23, lines 20-21).</u>

Comment #22: "Deal with uncertainty"

Authors' response: <u>It was rephrased (page 23, line 26)</u>.

2. List of relevant changes made in the manuscript

The relevant changes made in the manuscript are described per article section.

1. Introduction

In this section we emphasized the article's aim, included an explanatory paragraph on the review approach and removed section 1.2 by putting the article's outline at the end of the introductory text. Changes were also made to improve the clarity of some concepts.

2. Flood-related crowdsourced data

The beginning of this section was changed to discuss precipitation. The tables on the sub-sections were changed: the column 'Case Study' was split into 'Flood Type' and 'Location'. In the last sub-section, on Summary Analysis, a figure similar to Figure 2 was added, displaying the discussed studies in the framework presented in Figure 2 and analyzing the results.

3. Crowdsourced data in flood modelling

In this section, explanation on types of flood models was added, as well as information on the types of flood models used in the discussed papers. The text description of uses of crowdsourced data in the reviewed studies was restructured for clarity. Lastly, in the sub-section on crowdsourced information content, temporal dimension considerations in calibration and validation were introduced and the discussion on uncertainty and volume of crowdsourced data was expanded.

4. Opportunities and challenges

No relevant changes were made.

5. Conclusions and recommendations

No changes were made.

3. Marked-up version of the manuscript

This section provides the marked-up version of the manuscript. The following notation was used:

- Text that was inserted appears in red;
- Text that was deleted appear in strikethrough red;
- Black vertical track lines in the left margin indicate a change on the adjacent line.

Citizen observations contributing to flood modelling: opportunities and challenges

Thaine Herman Assumpção[1], Ioana Popescu[1], Andreja Jonoski[1], Dimitri P. Solomatine[1, 2]

[1]Integrated Water Systems and Governance, IHE Delft, Delft, The Netherlands
5 [2]Water Resources Section, Delft University of Technology, Delft, The Netherlands

Correspondence to: Thaine Herman Assumpção (t.hermanassumpcao@un-ihe.org)

Abstract. Citizen contributions to science have been successfully implemented in many fields – and water resources is one of them. Through citizens, it is possible to collect data and obtain a more integrated decision-making process. Specifically, data scarcity has always been an issue in flood modelling, which has been addressed in the last decades by remote sensing
10 and is already being discussed in a citizen science scenario. In this context, this article aims to review the literature on the topic and analyse the opportunities and challenges that lie ahead. The literature on monitoring, mapping and modelling, was evaluated according to the flood-related variable citizens contributed to. Pros and cons of the collection/analysis methods were summarised. Then, pertinent publications were mapped into the flood modelling cycle, considering how citizen data properties (spatial and temporal coverage, uncertainty and volume) are related to its integration into modelling. It was clear
15 that the number of studies in the area is rising. There are positive experiences reported in collection and analysis methods, for instance with velocity and land cover, and also when modelling is concerned, for example by using social media mining. However, matching the data properties necessary for each part of the modelling cycle with citizen generated data is still challenging. Nevertheless, the concept that citizen contributions can be used for simulation and forecasting is proved and further work lies in continuing developing and improving not only methods for collection and analysis but certainly for
20 integration into models as well. Finally, in view of recent automated sensors and satellite technologies, it is through studies as the ones analysed in this article that the value of citizen contributions is demonstrated.

1 Introduction

The necessity to understand and predict the behaviour of floods has been present in societies around the world. This comes from the fact that floods impact their surroundings - in negative or in positive ways. The most common way used nowadays
25 to better understand and often predict flood behaviour is through modelling and, depending on the system at hand, a variety of models can be used (Teng et al., 2017).

In order to have adequate representation of floods, most models require large amounts of data, both for model building and model usage. This is especially true for pluvial flood modelling, where flooding may not occur in gauged rivers and hence,

1

flow gauging stations outside of flooded zones may be of little use~~flow gauging stations may end up being of little use~~. Remote sensing technologies are a part of the solution, as they offer spatially distributed information. However, their availability may be limited, also in terms of space and time, and their uncertainties often are not quantifiable (Di Baldassarre et al., 2011; Grimaldi et al., 2016; Jiang et al., 2014; Li et al., 2017). Thus, acquiring the necessary data for simulations and predictions can still be expensive, particularly for rapidly changing systems that require frequent model updates.

In this context, sources of data coming in abundance and at low-costs are needed, together with modified modelling approaches that can use these data and can adapt to changes as fast as they occur. Citizen Observatory (CO) is an emerging concept in which citizens monitor the environment around them. It is often considered under the umbrella of Citizen Science (including citizen participation up to the scientist level) and it is also related to the concept of crowdsourcing (distributing a task among many agents). With technology at hand, it is possible to empower citizens to not only participate in the acquisition of data but also in the process of scientific analysis and even in the consequent decision-making process (Evers et al., 2016). Citizen Observatories have been researched in several EU-funded projects. Finished projects (CITI-SENSE, Citclops, COBWEB, OMNISCIENTIS and WeSenseIt) already resulted in valuable contributions to the field (Alfonso et al., 2015; Aspuru et al., 2016; Friedrichs et al., 2014; Higgins et al., 2016; Uhrner et al., 2013). For example, the CITI-SENSE project managed to simultaneously collect perception data and acoustic measurements in an approach that can be used to develop citizen empowerment initiatives in case of noise management (Aspuru et al. 2016); while in COBWEB project processes of quality assurance, data conflation and data fusion were studied and recommendations were made (Friedrichs et al., 2014). The currently running CO projects (Ground Truth 2.0, LANDSENSE, SCENT and GROW Observatory) propose to investigate this concept further.

Citizen science concepts have been researched and applied in various fields such as ecology and galaxy inspection (Lintott et al., 2008; Miller-Rushing et al., 2012). Volunteer Geographic Information (VGI), as one of the most active citizen science areas, has developed over the past decade and several researchers reviewed the state of the art of citizen science in the field of geosciences (Heipke, 2010; Klonner et al., 2016). There is also a part of the scientific community dedicated to investigating damage data crowdsourced after flood emergencies (Dashti et al., 2014; Oxendine et al., 2014) and evaluating the cycle of disaster management (Horita et al., 2013). In the context of water resources, Buytaert et al. (2014) reviewed and discussed the contribution of citizen science to hydrology and water resources, addressing the level of engagement, the type of data collected (e.g. precipitation, water level) and case studies where more participatory approaches are being implemented. Le Coz et al. (2016) provided examples and reflections from three projects related to flood hydrology and crowdsourcing, which involve the derivation of hydraulic information from pictures and videos in Argentina, France and New Zealand.

The present review aims to look at studies that had citizen science connected to floods. Specifically, it focusses on the data collected by citizens that are relevant in a flood modelling context, benchmarking difficulties and benefits of their collection and integration into models. Integration is considered for the purposes of model set up, calibration, validation, simulation and forecasting. ~~and analyse in detail how the contributions were made so far in a modelling context. Moreover, we aim to~~

5 ~~detect the opportunities and challenges related to exploring citizen science for modelling the hydrodynamics of floods.~~

The review process involved defining web platforms, keywords and criteria for searching and selecting publications. The main platforms used were Scopus and Google Scholar. The keywords are a combination of words related to citizen science (e.g. "citizen science" and crowdsourcing) and to flood-related variables (e.g. "water level" and "flood extent"). The

10 obtained articles were scanned for their content. Articles were selected mainly if crowdsourced data was obtained for quantitative use in monitoring, mapping or modelling. There were studies that were not selected because they just mention the use of crowdsourced data and do not provide more relevant information on collection, analysis, use and quantity of data, such as Merkuryeva et al. (2015). The same is the case of studies that evaluate variables qualitatively, in ways that cannot be directly associated with modelling (Kim et al., 2011). This review included articles published up to April 2017.

15

Further in this section, we introduce the citizen science concept and related classification systems. In Sect. 2 of the article, we overview studies on citizen contributions for flood modelling, classifying them according to the flood-related variable the contributions were made, followed by a summary of the pros and cons of measurement and analysis methods. Section 3 aggregates the studies that involve flood modelling and analyses the contributions considering the component of the

20 modelling process where they were used, also including a discussion of the factors that affect flood modelling. Section 4 describes the challenges and opportunities of using data contributed by citizens in flood modelling, and finally, Sect. 5 presents the conclusions and recommendations.

1.1 Citizen Science

25 Buytaert et al. (2014) defined citizen science as "the participation of the general public (i.e. non-scientists) in the generation of new knowledge". In the same manner that the involvement of citizens can be diverse, such is the way their participation is found in the scientific literature:

- Citizen Science (Buytaert et al., 2014)
30 - Citizen Observatory (Degrossi et al., 2014)
- Citizen Sensing (Foody et al., 2013)
- Trained volunteers (Gallart et al., 2016)

- Participatory data collection methods (Michelsen et al., 2016)

- Crowdsourcing (Leibovici et al., 2015)

- Participatory sensing (Kotovirta et al., 2014)

- Community-based monitoring (Conrad and Hilchey, 2011)

5
- Volunteered Geographic Information (Klonner et al., 2016)

- Eye witnesses (Poser and Dransch, 2010)

- Non-authoritative sources (Schnebele et al., 2014)

- Human Sensor Network (Aulov et al., 2014)

- Crowdsourced Geographic Information (See et al., 2016)

10

Some of the terms used by the above-mentioned articles have specific definitions that are used to delineate debates on the social mechanisms of citizen participation. Others are just the best form the researcher found to characterise the contribution or the citizen (e.g. eye witnesses). Citizen Science and adjacent areas have become fields of research in themselves that, for instance, focus on understanding the motivation of citizens or its interaction with public institutions (Gharesifard and Wehn,
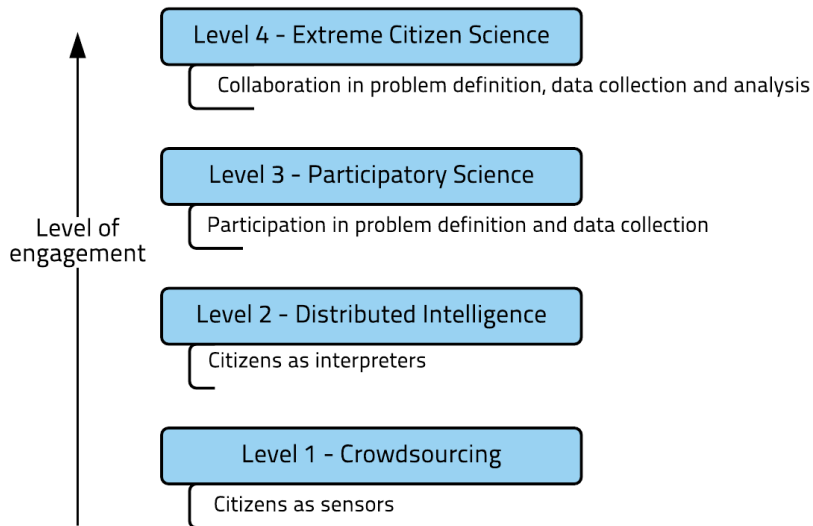
15   2016).

In this field, one of the classifications of citizen science is by level of engagement. Haklay (2013) built a model that has four levels (Fig. 1), in which the first one refers to the participation of citizens only as data collectors, passing through a second level in which citizens are asked to act as interpreters of data, going towards the participation in definition of the problem in

20   the third level and finally, being fully involved in the scientific enterprise at hand. The aim of the review presented in this current article is focused on the contribution towards flood modelling only, coming most prominently from the two lowest levels of engagement. We do not discuss topics related to engagement for the generation of (quantitative) data. Further in this article, for readability, only the term crowdsourced data is used to refer to data from these two levels of engagement will be termed as crowdsourced data.

25

4

Figure 1: Levels of participation and engagement in citizen science projects. Adapted from Haklay (2013).

Another way to classify citizen science initiatives (within the context of VGI) is by setting them as implicitly/explicitly volunteered and implicitly/explicitly geographic (Craglia et al., 2012). In this classification system, geographic refers to the main information conveyed through the contributed data, therefore, geo-tagged data is not necessarily geographic. For example, in the Degree Confluence Project (Iwao et al., 2006), citizens were oriented to go to certain locations, take pictures, make notes and deliberately make available their material on the project's website. In this case, the information is explicitly volunteered and explicitly geographic. Most land use/cover projects related to citizen science are explicitly geographic. Differently, in the study conducted by Lowry and Fienen (2013) citizens would also willingly send text messages to the researchers, in this case providing water level readings from installed water level gauges. Although explicitly volunteered, the message was non-geographic (just geo-tagged). Another type of implicitly geographic information was derived from Twitter by Smith et al. (2015) to obtain water level, velocity and flood extent estimates. As the citizens did not make the information public with the specific purpose to provide estimates, it is implicitly volunteered.
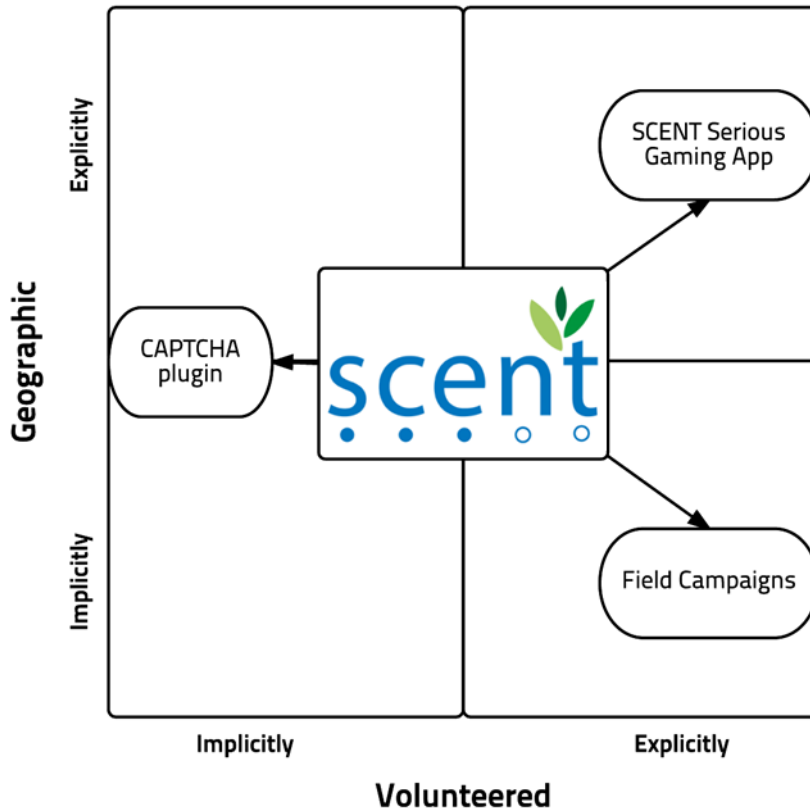
The concepts defined by Craglia et al. (2012) can be graphically represented as in Fig. 2. The SCENT project[1] (Smart Toolbox for Engaging Citizens in a People-Centric Observation Web) is one of the four Horizon 2020-funded projects focussing on citizen observatories. It lies in the middle of this quadrant as it encourages citizens to participate in gaming to collect land cover/use data, in field campaigns to collect other implicitly geographic information (e.g. water level), and also

---

[1] https://scent-project.eu/

aims to obtain implicitly volunteered contributions through a CAPTCHA[2] plugin, in which citizens tag images, ~~related~~e.g. of ~~to~~ land cover/use or water level, in order to access online content. Tagging images is uncorrelated to the CAPTCHA, it is a task performed after the test, on the same platform.

Figure 2: SCENT project represented in the typology of VGI (Volunteered Geographic Information)

~~1.2 Article outline~~

~~After this introduction, in Sect. 2 of the article, we overview studies on citizen contributions for flood modelling, classifying them according to the flood related variable the contributions were made, followed by a summary of the pros and cons of measurement and analysis methods. Section 3 aggregates the studies that involve flood modelling and analyses the contributions considering the component of the modelling process where they were used, also including a discussion on the~~

---

[2] CAPTCHA stands for 'Completely Automated Public Turing test to tell Computers and Humans Apart'. It is a test evaluating if the subject is human, which is used in websites to provide security. After the test is done the user can be asked to perform extra tasks, for example, tag images.

~~factors that affect flood modelling. Section 4 describes the challenges and opportunities of using data contributed by citizens in flood modelling, and finally, Sect. 5 presents the conclusions and recommendations.~~

2 Flood-related crowdsourced data

There are many types of data which relate to floods that can be collected by citizens. Likewise, there are many ways to
5 collect, analyse and use them (for monitoring, mapping and modelling). In the next sub-sections we address how these aspects were explored in the scientific literature. Each sub-section discusses a data type corresponding to a flood modelling variable: water level, velocity, flood extent, land cover and topography. Depending on the type of flooding, other variables are relevant, such as precipitation. The scientific literature already shows that citizens' contributions could be useful for observation this variable (Muller et al., 2015; De Vos et al., 2017). However, rainfall is not included in this section because it
10 was already covered by the review of Muller et al. (2015). Moreover, in general it is a variable of greater importance for hydrological models, whilst the present review is focussed on a hydrodynamic representation of floods. ~~It needs to be noted that there are studies that just mention the use of crowdsourced data and do not provide more relevant information on collection, analysis and quantity of data, such as Merkuryeva et al. (2015). Some of the studies evaluate variables qualitatively, in ways that cannot be directly associated with modelling, therefore such studies are not included (Kim et al.,~~
15 ~~2011). Finally, t~~There are articles mentioned and reviewed in more than one section because they evaluated more than one variable, as it is, for example, the case of Smith et al. (2015). ~~It is worth mentioning that this review includes articles published up to April 2017.~~

2.1 Water level

Table 1 gives an overview of the articles about collection of water level data. The studies presented started to involve
20 citizens in the collection of water level data with the explicit goal of improving flood management. This is due to the ease of collecting such data, which mostly consists of comparing the water level with a clearly defined reference. In some cases, the reference is a water level gauge, the comparison is made by the citizen, and readings are being submitted to the researchers (Alfonso et al., 2010; Degrossi et al., 2014; Fava et al., 2014; Lowry and Fienen, 2013; Walker et al., 2016). Such kind of reading practically do not require further analysis, although they entail the installation of water level gauges.

25

In other cases, the citizen provides qualitative data that will be compared to references by researchers. ~~m~~Mostly during flooding situations, citizens provide pictures (Fohringer et al., 2015; Kutija et al., 2014; Li et al., 2017; McDougall, 2011; McDougall and Temple-Watts, 2012; Smith et al., 2015; Starkey et al., 2017) or videos (Le Boursicaud et al., 2016; Le Coz et al., 2016; Michelsen et al., 2016). In the case of pictures/images, the water level is compared with objects in the images
30 that have known or approximately known dimensions. For videos, although water level was estimated, the main goal was to obtain discharge values, via estimates of flow velocity. In two cases, texts from citizens were used (e.g. water over the knee),

to calculate~~provide directly quantitative~~ water level values or to assum~~ing~~ a certain value when no value was provided (Li et al., 2017; Smith et al., 2015). This sort of data (text, pictures and videos) was mostly collected through social media and public image repositories~~,~~. Gathering data from such sources requir~~es~~ing mining of the relevant material (i.e. extraction of specific data from a dataset) and dealing with uncertainties in the spatio-temporal characterization of the data of interest.

5

One aspect that varies across the studies is the level of detail in the comparison method used for determining the water level measurement. For example, McDougall (2011) and McDougall and Temple-Watts (2012) explicitly state that field visits to the selected photo locations are required in order to properly analyse the image and extract water level values. On the other hand, Fohringer et al. (2015), Smith et al. (2015) and Starkey et al. (2017) do not mention any method.

10 In most cases, crowdsourcing has been used to monitor water level, followed by the use of such data for modelling and lastly for mapping. In the case of Starkey et al. (2017), although hydrological modelling was done and water levels were converted into discharge to allow for comparisons, only qualitative comparisons were made.

Table 1: Scientific literature on citizen contributions to measurement and analysis of water level

Study	Measurement/analysis methods	Type	Purpose	Flood type	Location~~Case Study~~
Alfonso et al. (2010)	Citizen's reading of water level gauges sent by text message	1D	Monitoring	No flooding	~~Polders in~~ The Netherlands
Lowry and Fienen (2013)	Citizen's reading of water level gauges sent by text message	1D	Monitoring	No flooding	~~Watersheds in the~~ USA
Deg~~G~~rossi et al. (2014)	Citizen's reading of water level gauge sent through app/webpage	1D	Monitoring	No flooding	~~Flood Citizen Observatory in~~ Brazil
Walker et al. (2016)	Citizen's reading of water level gauge collected and provided by the community	1D	Monitoring	No flooding	~~Dangila *woreda* region in~~ Ethiopia
Fava et al. (2014)	Citizen's reading of water level gauge sent through app/webpage	1D	Modelling	Flood forecasting	~~Flood forecasting in~~ Brazil
Le Boursicaud et al. (2016)	LSPIV analysis of video collected from social media (YouTube)	1D	Monitoring	Flash flood	~~Flash flood in~~ France
Le Coz et al. (2016)	LISPIV analysis of video sent through webpage	2D	Modelling	Fluvial flood	~~Flash flood in~~ Argentina
Michelsen et al. (2016)	Analysis of images extracted from videos collected from social media (YouTube) and own photographs	Neither	Monitoring	No flooding	~~Cave in~~ Saudi Arabia
Li et al. (2017)	Analysis of texts and pictures collected from social media (Twitter)	2D	Monitoring	Flood map	~~Flood map in the~~ USA
Starkey et al. (2017)	Citizen's reading of water level gauge and analysis of pictures and videos collected from social media (Twitter) and	2D	Monitoring	Flood	~~Flood in the~~ UK

8

	crowdsourced (email, webpage and mobile app)				
McDougall (2011), McDougall and Temple-Watts (2012)	Analysis of texts and pictures collected from social media (Twitter, Facebook) and crowdsourced (email, text message, Ushahidi, Flickr and Picasa)	2D	Mapping	Flood map	~~Flood map in~~ Australia
Kutija et al. (2014)	Analysis of pictures collected by the University and City Council	2D	Modelling	Pluvial and drainage flood	~~Pluvial flood in the~~ UK
Aulov et al. (2014)	Visual analysis of texts and pictures collected from social media (Twitter and Instagram)	2D	Modelling	Coastal flood	~~Storm surge forecasting in the~~ USA
Fohringer et al. (2015)	Visual analysis of pictures collected from social media (Twitter) and crowdsourced (Flickr)	2D	Mapping	Flood	~~Flood in~~ Germany
Smith et al. (2015)	Analysis of texts and pictures collected from social media (Twitter)	2D	Modelling	Pluvial and drainage flood	~~Pluvial flood in the~~ UK

2.2 Velocity

As velocities and discharges traditionally require more complex measuring methods, the collection of this type of data by citizens has not been explored on a scientific basis. However, it is common to include direct measurements of velocity in protocols to monitor the environment and water quality, as it is the case of Hoosier Riverwatch (IDEM, 2015). In these cases, the citizens perform measurements that involve more processing (e.g. definition of transects to measure flow, use of formulas).

To the best of the authors' knowledge, only three studies were found that make use of velocity data collected by citizens, all for the study of floods, as presented in Table 2. Le Boursicaud et al. (2016) evaluated the surface velocity field in a channel from a YouTube video, using the LSPIV methodology (Large Scale Particle Image Velocimetry), an established method to obtain velocity from a sequence of images. For enabling this analysis, information about the camera (model and lens type) is needed, visible, fixed elements are needed to be used as reference points and it is also required that both river banks are visible. Although the method calculates the velocity in two dimensions, in Table 2 we referred to it as 1D because it was carried out in a channel, which in a context of flood modelling is considered as a 1D domain. A complementary project was discussed by Le Coz et al. (2016), in which the same technique is applied to a video crowdsourced by a citizen, this time using the result to estimate discharge and the latter to calibrate a 1D hydraulic model. For this, a visit to the location was needed to extract cross-sectional data. In this context, Yang and Kang (2017) developed a method for crowd-based velocimetry of surface flows, based on Particle Image Velocimetry, in which citizens mark features in the picture. The method has not been tested with citizen collected data yet.

The third study, conducted by Smith et al. (2015), selected Twitter messages that include terms of semantic value related to the citizen location, water depth (e.g. knee-deep) and velocity. The terms were then associated with quantitative values/ranges. The authors did not go into detail on discussing the reliability and uncertainty in such data, even though the issue is recognised.

Table 2: Scientific literature on citizen contributions to measurement and analysis of velocity

Study	Measurement/analysis methods	Type	Purpose	Flood type	Location~~Case Study~~
Le Boursicaud et al. (2016)	LSPIV analysis of video collected from social media (YouTube)	1D	Monitoring	Flash flood	~~Flash flood in~~ France
Le Coz et al. (2016)	L~~I~~SPIV analysis of video sent through webpage	2D	Modelling	Fluvial flood	~~Flash flood in~~ Argentina
Smith et al. (2015)	Analysis of texts and pictures collected from social media (Twitter)	2D	Modelling	Pluvial and drainage flood	~~Pluvial flood in the~~ UK

2.3 Flood extent

Flood extent, similarly to water level, is a variable that is simple to measure as it consists of binary values: flooded or non-flooded area. As a 2D variable, it needs a lot of spatial information and it is the main reason related studies gather flood extent estimates in data rich environments, through social media/photo sharing services~~Flickr/Picasa~~ mining, as shown in Table 3. In some cases, the citizens act only as sensors, providing pictures to be analysed by the research team, while in other cases they also act as interpreters by providing the flooded/non-flooded information. As can be expected, all studies found were carried out in urban areas.

In some of the studies the text and images are indicating the location of their origin as being flooded (georeferenced or inferred) (Aulov et al., 2014; Smith et al., 2015; Yu et al., 2016), whilst in others (Cervone et al., 2016; Li et al., 2017; Rosser et al., 2017; Schnebele et al., 2014; Schnebele and Cervone, 2013) there is processing of the information to infer the surrounding inundated areas. Additionally, the last group of studies mentioned fused flood extent data from citizens with satellite data or with gauge data.

Table 3: Scientific literature on citizen contributions to measurement and analysis of flood extent

Study	Measurement/analysis methods	Purpose	Flood type	Location~~Case~~

				Study
Cervone et al. (2016), Schnebele et al. (2014); Schnebele and Cervone (2013)	Analysis of pictures and videos collected from social media (Facebook and YouTube) and crowdsourced (Flickr)	Mapping	Flood map	Flood maps in USA and Canada
Li et al. (2017)	Analysis of texts and pictures collected from social media (Twitter)	Mapping	Flood map	Flood map in the USA
Rosser et al. (2017)	Analysis of crowdsourced pictures (Flickr)	Mapping*	Flood map	Flood map in the UK
Aulov et al. (2014)	Visual analysis of texts and pictures collected from social media (Twitter and Instagram)	Modelling	Coastal flood	Storm surge forecasting in the USA
Smith et al. (2015)	Analysis of texts and pictures collected from social media (Twitter)	Modelling	Pluvial and drainage flood	Pluvial flood in the UK
Yu et al. (2016)	Citizen's visual identification of flooded/non flooded location collected by governmental Chinese website	Modelling	Pluvial and drainage flood	Flood in China
Padawangi et al. (2016)	Citizen information	Monitoring	Flood	Flood in Indonesia

* A statistical model is created, but in this study we consider only physical models in the modelling category

2.4 Land cover/Land use

Land cover is not a variable in flood-related models but we include it in this review for its importance in inferring roughness. Other valuable aspects of land use data are the information on roads and structures that can be obstacles to floods, which can
5 be incorporated in the model structure; and the information on vulnerability (e.g. hospitals, dense residential areas, industrial zones), which can be used to obtain flood risk maps. According to Klonner et al. (2016), when reviewing the literature on VGI for natural hazard analysis, there are few studies for vulnerability analysis. The aspects of land use related to vulnerability and risk are complex and study topics on themselves, so these aspects are not discussed further in this article. Table 4 presents the articles considered for this review. Compared to previously discussed variables, the contribution of
10 citizens to land cover maps generation has been already proved as a concept (Albrecht et al., 2014; Fritz et al., 2012), nowadays being researched further for quality of data (Salk et al., 2016) and fusion of maps (Lesiv et al., 2016).

One of the first publications on the subject was from Iwao et al. (2006), in which they describe the Degree Confluence Project. The objective was to generate a global land cover map, which implies obtaining ground truth data from around the globe. For obvious reasons, it was unfeasible to make field campaign or analyse low-resolution images with sufficient resolution. Thus, they launched a webpage that invited citizens to visit integer coordinates (e.g. 25° W, 25°) locations, take

5   photos from the four cardinal directions and provide comments on the region. They discovered that citizen-generated data was having quality similar to that provided by specialists.

Another significant project in the area is GeoWiki. It started in 2009 as a platform for people to validate global land cover maps, by comparing their classification to high-resolution images (Fritz et al., 2009). The project has grown since and has

10   recently achieved its main goal: to generate a hybrid global land cover map by fusing existing maps and performing calibration and validation using the analyses made by citizens (See et al., 2015). Current initiatives in the GeoWiki project include gamification and analysis of pictures uploaded onto the platform (See et al., 2015). Many studies stemmed from the data collected, generally focused on specific land cover types. A similar approach is taken by Dong et al. (2012), that analyses pictures uploaded by citizens using a different web application. The research conducted by Dorn et al. (2014) goes

15   one step further, as it attributes roughness values to multiple land cover maps, including Open Street Maps (a website where citizens can modify the current street and land cover map).

Table 4: Scientific literature on citizen contributions to measurement and analysis of land cover/land use

Study	Measurement/analysis methods	Purpose	Flood type	LocationCase Study
Iwao et al. (2006)	Visual interpretation of crowdsourced tagged pictures sent through app/webpage (Degree Confluence Project website)	Mapping	No flooding	Global land cover map
See et al. (2015b)*	Visual interpretation of Google Earth and pictures sent through app/webpage (GeoWiki)	Mapping	No flooding	Global land cover map
Dong et al. (2012)	Analysis of tagged pictures from Global Geo-Referenced Field Photo Library (DCP citizen pictures + field trip pictures)	Mapping	No flooding	Forest cover map in Asia
Dorn et al. (2014)	Use of Open Street Maps	Modelling	Fluvial flood	Flood in Austria

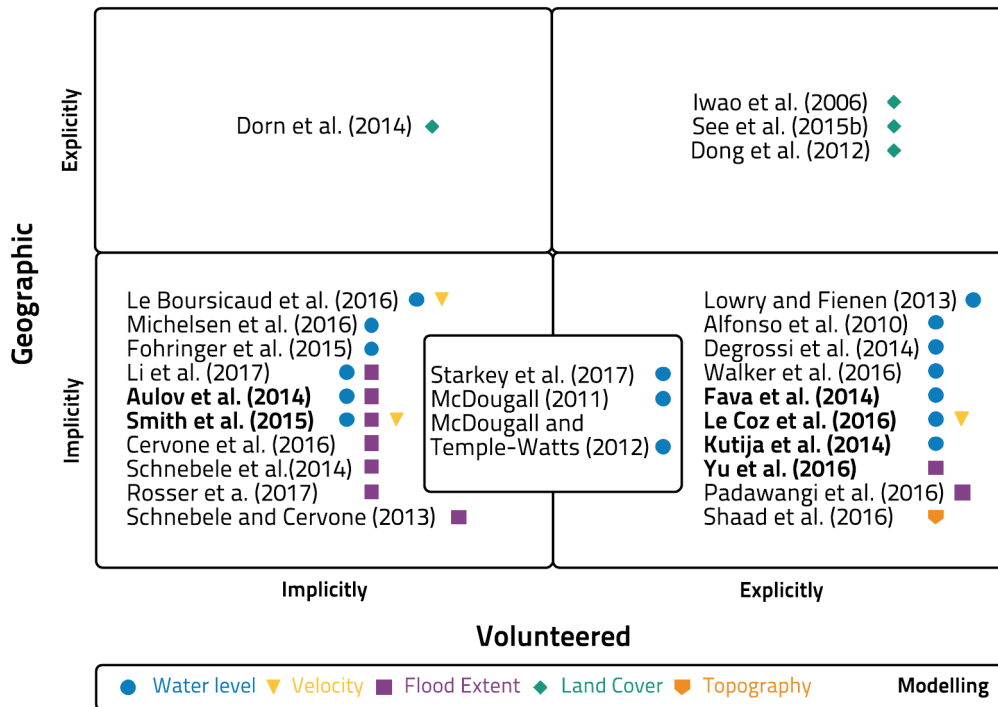* Many other articles related to crowdsourcing through GeoWiki

2.5 Topography

The Digital Elevation Model (DEM) is one of the most important components in flood modelling, as it generally heavily influences flood propagation. It is particularly important in urban settings, where spatial variability in refined scales has a considerable effect on the direction of water flows. Unfortunately, this is a complex variable to measure that so far relies either on fully trained professionals to go to the field, or on expensive airborne technologies. Recently, Shaad et al. (2016) studied a terrain capturing low-cost alternative to LiDAR remote sensing images and other expensive methods. The low-cost technique is the ground-based close-range photogrammetry (CRP) that consists of collecting images/videos from the ground, post-processing them and obtaining terrain information. Volunteers made the videos in a designated location, where even Unmanned Aerial Vehicles (UAVs) would not be able to collect data. After comparing the results to other methods, they concluded that the result has an acceptable quality.
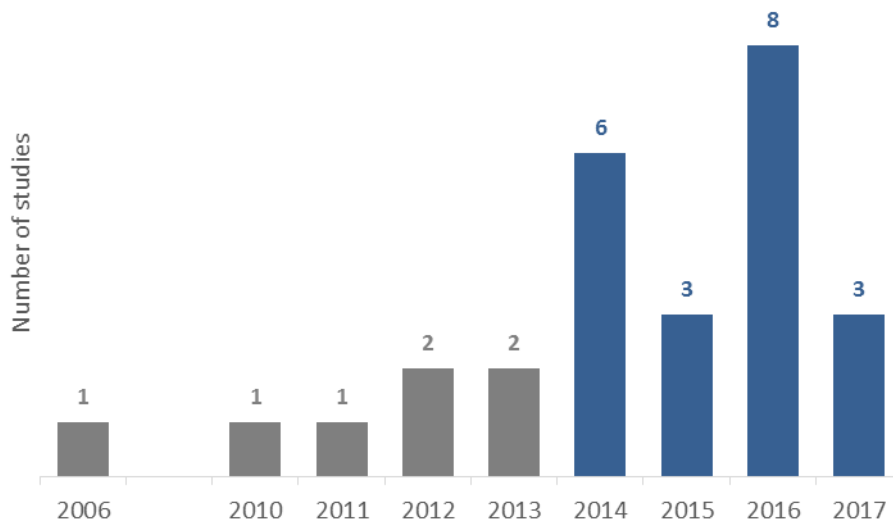
2.6 Summary analysis

By classifying the discussed studies according to Craglia et al. (2012), there is an overall similarity in the number of studies that crowdsource data implicitly and explicitly (Fig. 3). It is visible though that this aspect does not translate into homogeneous distribution per flood-related variables, with most implicitly volunteered contributions being related to flood extent and most explicit being related to water level. There is a slightly higher concentration of modelling studies that are explicitly volunteered, but not enough to be able to draw any conclusions.
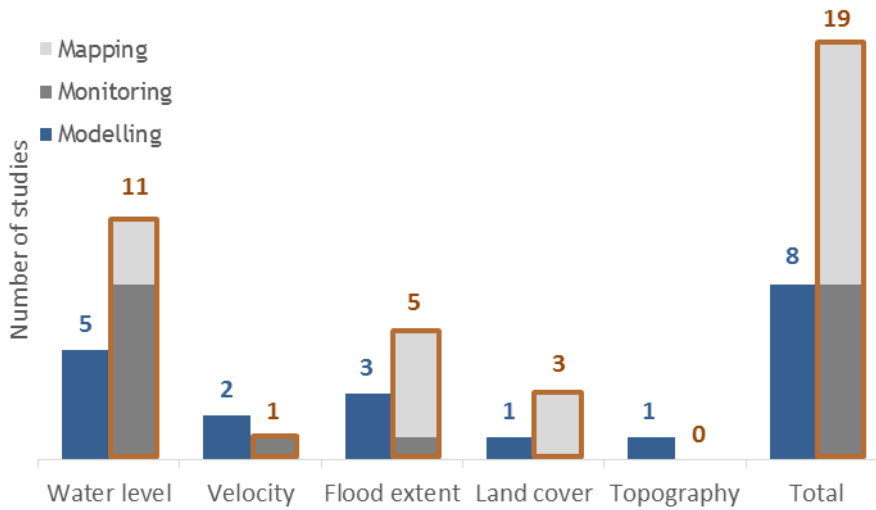
Figure 3: Selected studies represented in the typology of VGI (Volunteered Geographic Information)

Considering the temporal distribution of studies evaluated in this review, it is evident that there is a trend: the rise in ~~a~~ number of studies from 2014 onwards (Fig. 4~~3~~). This relates to the initial barrier in acknowledging citizen data as having quality that is high enough for scientific studies (Buytaert et al., 2014). This resistance is reducing over time as such data is being proved~~has been proven~~ useful, protocols are being designed and the data uncertainty is being better understood and quantified.

Figure 4: Number of studies analysed per year

If the analysed studies are aggregated into categories (Fig. 54), it can be seen that modelling studies amount to approximately the same quantity as monitoring ones, but they are only about a third of all studies reviewed. This is expected because to use data in models it is necessary to monitor them first. Also, monitoring and mapping applications attend to more general end uses. Specifically for land cover, there is an unexplored field in modelling (there are more mapping studies than the ones in the graph, see Sect. 2.4). The reason behind may be that modellers do not tend to validate their own land cover maps and thus will not do it with citizen science data. What can be noted though, is the lack of exploration of velocity and topography variables, which, as mentioned, can be due to the complexity in analysing and setting up the experiment.

Figure 5: Number of studies analysed per flood-related variable per category: mapping, monitoring and modelling

Related to that, previous sub-sections discussed in detail the methods for collection and analysis of flood-related data obtained through crowdsourcing. For example, water level data obtained from reading a water level gauge is easy to collect and easy to analyse. On the other hand, it requires the installation of gauges (Fig. 65). In summary, whenever data is collected from the Internet, there is the disadvantage of needing social media/photo sharing servicesFlickr/Picasa mining, entailing computational efforts and dealing with a high percentage of data that is not georeferenced or time stamped. Further, in the case of water level and velocity, some studies suggest that also field visits are necessary and the methods to analyse data are complex. Considering crowdsourced data on flood extent, land cover and topography, it is straightforward to measure and analyse them, although their delivery to the interested parties may requires a smartphone app or a web-site to be set up and maintained (with the exception of Open Street Maps).

	Collection	Analysis
Reading water level gauge	Measurement (+) Instalation of gauges (-)	Analysis (+)
Collection of data from social media/Flickr/Picasa	Social media/Flickr/Picasa mining (-) Location and time (-)	Flood extent analysis (+) Water level and velocity analysis (-) Field visit for water level (-)
Collection of data from Youtube	Youtube mining (-) Location and time (-)	Established method (+) Field visit (-) Camera and other set ups (-)
Crowdsourced data	Measurement (+) App/website (-)	Analysis (+)
Visual analysis of tagged images		Analysis (+) App/website (-)
Open Street Maps editing	Measurement (+) OSM land cover types (-)	Analysis (+)
Crowdsourced video	Measurement (+) App/website or youtube mining (-)	Established method (+)

(Left-side labels connected by arrows: Water level, Velocity, Flood extent, Land Cover, Topography)

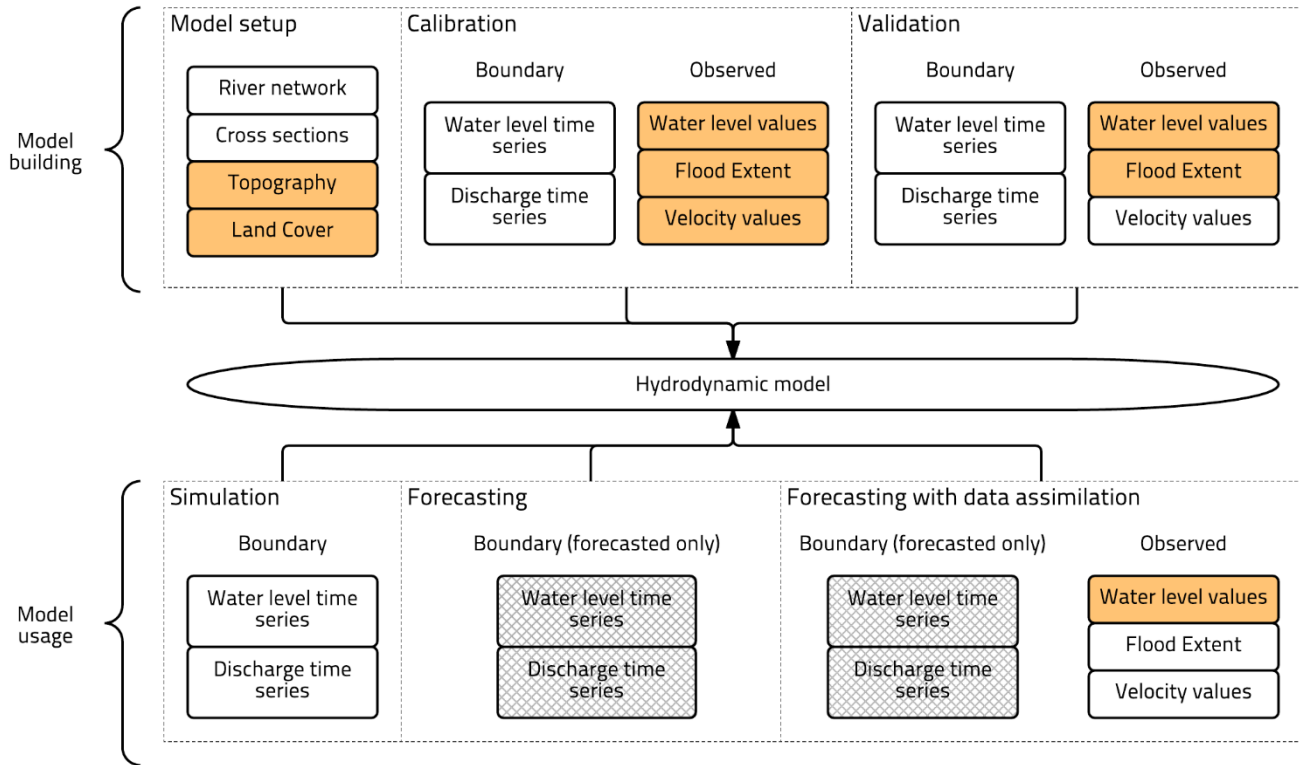**Figure 6: Pros and cons of collection and analysis methods used to collect flood-related data by citizens**

3 Crowdsourced data in flood modelling

We have seen in the previous section that data related to flood modelling can be collected for many reasons, mainly monitoring, mapping and modelling. In this section we intend to explore in detail how the data was integrated into models. By concentrating on the studies in which modelling was performed, we explore in detail how crowdsourced data was integrated into each component of flood models.

There is a variety of flood models developed to deal with different types of flood, including: fluvial, pluvial, coastal and drainage floods. The main driver of fluvial floods is upstream river discharge, of pluvial floods it is precipitation and of coastal floods it is storm surges. In urban drainage floods, the flows inside, through and outside of drainage systems are pivotal for flood representation. Moreover, there are complex cases where more than one flood process needs to be represented. Although in physically-based flood models water flow is computed by the same principles, different sets of data are needed for different types of flood models. We focus on a general hydrodynamic model definition and its common inputs but present what was the flood type evaluated in the scientific literature (Table 5).

The flood modelling process typically, typically involving hydrodynamic models, has two parts: model building, and model usage. (Fig. 67). Model building starts by defining the model setup (boundary conditions, parameters, schematization, input data), followed by calibration and validation of the water level and velocity fields (dependent variables) with observed values. Calibration and validation can be performed for both simulation and forecasting models. Once the model is ready, simulations can be run by using different boundary conditions or introducing designed measures for better flood management; or forecasts can be made by using forecasted water levels or discharges as boundaries. In a simulation setting,

model parameters are assumed to be constant in time, while in a forecasting setting the parameters, inputs or states (water levels) can be updated while the model is in use, using data assimilation.

5   **Figure 7: Flood models data requirements. Orange coloured tiles correspond to data that citizens have contributed to in a flood modelling context and gridded tiles correspond to data citizens cannot contribute to (forecasted water levels and discharges).**

~~In view of this process, we analyse how the studies that were carried out in a modelling context included crowdsourced data into the model (Table 5).~~ From the studies analysed (Table 5), three consider 1D channels and the others worked in a 2D setting. Most of them analyse only one variable, except Smith et al. (2015) that evaluate water level and velocity. Moreover,

10   most of them model urban floods, some in a pluvial and others in a fluvial context.

15

Table 5: Scientific literature on crowdsourced data used in flood modelling

Use in modelling	Study	Measurement method	Type	Variable	Flood type	Location~~Case study~~
Model setup	Dorn et al. (2014)	Use of Open Street Maps	2D	Land cover	Fluvial flood	~~Flood in~~ Austria
	Shaad et al. (2016)	Analysis of pictures captured by volunteers at selected location	2D	Topography	Fluvial flood	~~Flood in~~ Indonesia
Calibration	Smith et al. (2015)*	Analysis of pictures and tweets collected from social media (Twitter)	2D	Water level and velocity	Pluvial and drainage flood	~~Pluvial flood in the~~ UK
	Le Coz et al. (2016)	L~~I~~SPIV analysis of videos sent through webpage	1D	Velocity	Fluvial flood	~~Flash flood in~~ Argentina
	Yu et al. (2016)	Citizen's visual identification of flooded~~/non flooded~~ location provided through Chinese website	2D	Flood extent	Pluvial and drainage flood	~~Flood in~~ China
Validation	Kutija et al. (2014)	Analysis of pictures collected from the University and City Council	2D	Water level	Pluvial and drainage flood	~~Pluvial flood in the~~ UK
	Yu et al. (2016)	Citizen's visual identification of flooded~~/non flooded~~ location provided through Chinese website	2D	Flood extent	Pluvial and drainage flood	~~Flood in~~ China
Data assimilation	Aulov et al. (2014)	Visual analysis of texts and pictures collected from social media (Twitter and Instagram)	2D	Water level and flood extent	Coastal flood	~~Storm surge forecasting in the~~ USA
	Mazzoleni et al. (2015, 2017)	Simulated citizen reading of water level gauge sent through app	1D	Water level	Flood forecasting without flood model	~~Flood forecasting in~~ Italy and USA
	Fava et al. (2014)	Citizen's reading of a water level gauge sent through app or webpage	1D	Water level	Flood forecasting without flood model	~~Flood forecasting in~~ Brazil

* It is classified as calibration because, in the classical sense, it improves the model according to observations. However, what actually is done is the fine-tuning selection of the precipitation field that fits the observations better.

5

Considering model building, specifically the model setup, citizens contributed to improving/updating land cover (and consequently roughness) and topography information ~~the datasets that are used in the model, both for Land Cover (Dorn et~~

al., 2014) and the Digital Elevation Model (Shaad et al., 2016). Dorn et al. (2014) used the land cover information contained in ~~The first case uses~~ Open Street Maps³, for modelling a fluvial flood.~~an online platform that provides maps, including land cover, which can be changed by citizens at any time.~~ They~~Dorn et al. (2014)~~ do not analyse how much contribution was made by the citizens and data processing is restricted to attributing land cover classes to the features displayed in the maps.

5 In the study of Shaad et al. (2016), which addresses topography, there is only one citizen contribution (low-cost alternative) in one selected location that is merged with an existing DEM and then used in the model. In both cases, ~~T~~the objective was to compare the performance of this low-cost alternative against the performance of consolidated technologies~~' performance~~ when used for hydrodynamic simulations.

10 Crowdsourced data has also been used to calibrate and validate flood models in four studies. ~~We have found four studies that gather and use this information,~~ ~~o~~One study gathered such data through social media and public image repositories mining and the others through data uploaded by citizens on specific platforms. Smith et al. (2015) ~~aimed to do real time urban modelling to identify possible flooded areas due to rainfall. Storm events were~~ identified storm events through social media, triggering shock-capturing hydrodynamic model runs with various rainfall intensities. The results were compared with social

15 media data on water level/velocity. The comparison consisted of defining~~They defined~~ a buffer zone around the crowdsourced observation location, built a histogram of simulated cell values within it and evaluat~~ing~~ed the overlap of ~~the~~ crowdsourced value/range and the histogram 70-95th percentile range. As most citizen contributions did not have a water level.velocity value, they received a minimum water level value. Because of that, ~~T~~the selected simulation was the one ~~with less rainfall,~~ with more 'overlaps' and that would not perform better than a simulation with rainfall slightly higher. ~~This was~~

20 ~~done because most contributions were considered as a minimum water level criterion.~~

Yu et al. (2016) collected flooded~~/non-flooded~~ data through a Chinese website and divided it into calibration and validation data sets for a pluvial flood model verification. There is no mentioning on how this data is provided (e.g. text or image). Le Coz et al. (2016) obtained a discharge value for calibration of a hydraulic model based on the surface velocity data obtained

25 by a video uploaded to a specific website. Kutija et al. (2014) collected pictures uploaded by citizens and extract from them water levels by comparison with reference objects, such as cars (no further detailing on the method of extraction is made). Water level data is then used to validate a pluvial flood model.

~~T~~All ~~t~~he described approaches so far consider citizen data for model building and its possible extension for recalibration and

30 revalidation. Four ~~The~~ studies ~~of Mazzoleni et al. (2015, 2017)~~ went one step further, integrating crowdsourced data in model usage. Mazzoleni et al. (2015, 2017) used synthetically generated data to represent citizen observations, which were~~as they incorporate crowdsourced data while the model is being used.~~ incorporated in the model ~~This is done~~ through data

³ Open Street Maps (OSM) is an online platform that provides street maps and other information. The maps provided can be edited by the users at any time

20

assimilation algorithms, adapted to deal with the intermittent nature of crowdsourced data. Aulov et al. (2014) and Fava et al. (2014) also used the data for simulation/data assimilation instead of setup, but the methods used are not detailed in the studies. However, the studies of Mazzoleni et al. (2015, 2017) and Fava et al. (2014) were made for flood forecasting through hydrological models and not using hydrodynamic models.

3.1 Crowdsourced data information content

If we aim at integrating data into model, data accuracy, volume and temporal and spatial coverage should be at a certain level. When these data properties are inadequate, data integration would not provide useful results (i.e. the model performance can be low). Although most modelling variables vary in time and space, the data does not need to cover all dimensions in all parts of the modelling process. For instance, in model setup, topographic data is not needed every 15 minutes, hourly or daily; it can be provided in a discrete time coverage, from months to years. We analyse four data properties: temporal coverage, spatial coverage, volume and uncertainty (Table 6). Although same for all parts, the last two properties vary significantly when analysing the information content of crowdsourced data and that is why these properties are included (Table 6).

Table 6: Data properties currently required in the modelling process

	Setup	**Calibration & Validation**[1]	**Simulation**	**Data assimilation**	**Data assimilation**
	Topography Land Cover	*Water Level Velocity Flood Extent*	*Water Level Velocity*	*Water Level Velocity*	*Flood Extent*
Temporal coverage	Discrete	Discrete/Continuous	Continuous	Variable	Variable
Spatial coverage	Distributed	Discrete/Distributed	Discrete	Discrete	Unknown
Uncertainty	The lower the better				
Volume	The higher the better				

[1] Dependent on purpose of the model

Analysing crowdsourcing studies by their information content, it is possible to draw the following conclusions:

- Model setup: for integration of topographic and land cover data, it is necessary to have spatially distributed data. While this has been achieved within land cover studies, there is only one study involving topography and the data obtained so far have discrete spatial coverage.

- Calibration and validation: through mining and crowdsourcing of water level and flood extent estimates from social media and open image repositories, spatially distributed crowdsourced data have already been obtained for calibration/validation of simulation models acquired and integrated and that is why there are more studies related to this modelling stage.. The accuracy of the time stamp was considered vital (Kutija et al., 2015) and results in time

21

have a preliminary good level of agreement with citizen observations (Yu et al., 2016). However, even though these studies compare the results with citizen observations in time, this is done qualitatively and there is no focus on reporting and evaluating the temporal coverage.

- Simulation: traditional modelling efforts require time series of data at specific frequencies, which has only been achieved through crowdsourcing in the realm of community-based approaches, in which water levels are measured at 6 a.m. and 6 p.m. in agreement with the community (Walker et al., 2016). However, this type of data has been only monitored and not used in a modelling context so far.
- Data assimilation: it generally assimilates data provided with a fixed time frequency, but there are a few studies that consider intermittent data to be assimilated (Mazzoleni et al., 2015, 2017). However, similarly to simulation, the temporal coverage of crowdsourced data is insufficient for data assimilation efforts.

Considering uncertainty, this is highly dependent on the collection/analysis method. For example, obtaining water level values from pictures of in flooded areas (2D) is uncertain, as it mostly involves the selection of what constitutes a good reference point to be made by the citizen. Flood extent, on the other hand, tends to be less uncertain to measure, due to its binary nature. The collection through data mining (and sometimes crowdsourcing) has, in general, more sources of uncertainty: from geotagging, timestamping and the observed value. To deal with the first two, Aulov et al. (2014) used only data that contained proper geotag and time stamp. Kutija et al. (2014) classified non-timestamped data as during or after the event, based on picture visual inspection, defining an observation time range. Smith et al. (2015) dealt with uncertainty in location by generating a histogram of simulated values around the observed point. Yu et al. (2016) acknowledged these sources of uncertainty. Regarding uncertainty in value, existent in all sources of crowdsourced data, most studies used the (processed) observations as were, without indication of uncertainty. Smith et al. (2015) defined ranges, although these are not discussed. Mazzoleni et al. (2015, 2017), used uncertain synthetic crowdsourced data with variable uncertainty.

Regarding volume of data collected, this is an issue for all modelling processes, although data mining has again been able to provide a better coverage. Besides the challenge of uncertainty, The challenge of data mining has also the challenge , however, lies in providing less uncertain data, in terms of value, geo-referencing and time stamp, and also in providing data in conditions that are not extreme, as most of the contributions are done in floods situations and . Data miningit is also limited to certain variables (water level, flood extent and velocity). Some of the studies were proof of concepts and integrated up to 3 crowdsourced observations each (Le Coz et al., 2016; Fava et al.; 2014; Shaad et al., 2016). Others ranged from 12 to 298 observations (Kutija et al., 2014; Smith et al., 2015; Yu et al., 2016) and in some cases it was not possible to define the exact number (Aulov et al., 2014; Dorn et al., 2014).

22

4 Opportunities and challenges

In the last years, the interest in citizen science and the number of citizen science studies in the water resources context has risen considerably. The main factors affecting its use in flood modelling are the degree of how difficult it is to acquire and evaluate these data and their integration into the models. Our analysis of the existing literature allows for pointing out a
5   number of positive experiences from which we can derive opportunities to:

- ▪ Explore and improve the existing methods to obtain water velocity and topography from videos
- ▪ Explore calibration and validation employing data collected through social media in urban environments
- ▪ Explore the possibilities of setting up the models with the use of land cover maps validated with citizen science
10 ▪ Make use of apps/websites already developed for citizen science

The first one is based on small scale but successful studies related to using well-developed techniques in a citizen science scenario. The relevant experience in data gathering and analysis can be updated to fit the needs of flood modelling. Also, social media and public image repositories mining has proved to be successful in calibration and validation in modelling
15 studies, proving the concept and opening the opportunity to investigate how large this contribution is. As mentioned previously, in the field of land cover map generation citizen data has been used to validate maps and this successful example could be used to obtain new roughness maps in a modelling context. Lastly, technological development of apps, websites and techniques could be shared and put to public use, to be tested further and to avoid duplicated work.

20 There are aspects of <u>the integration of crowdsourced data into flood modelling</u>~~interactions between citizen science and water resources~~ that are still challenging. These are:

- ▪ Explore the use of citizens as data interpreters
- ▪ Improve methods to estimate water level from pictures
25 ▪ Harmonise the time frequency and spatial distribution of models with the ones of crowdsourced data
- ▪ <u>Quantification of</u>~~Deal with the~~ uncertainty
- ▪ Increase the volume of data gathered, mainly in non-urban environments

Most of the analysed studies regard the citizen as a sensor, with the exception of studies about land cover related data, in
30 which the citizen also acts as an interpreter. For other variables, some studies have already started evaluating the ability of citizens to provide interpreted information (Degrossi et al., 2014), but these are few. Regarding water levels, readings from rulers and extraction from pictures are described differently in the literature, with varying degrees of thoroughness, indicating a need for development and testing of water level measurement methodologies in the context of citizens'

23

contributions. The third point brings up a challenge that concerns not only citizen science but also modelling: what is the necessary temporal and spatial distribution? Is the traditional modelling approach definitive in terms of data requirements and citizen science approaches should adapt to it, or, the modelling process can be adapted to receive citizen science data? The fourth challenge relates to the quality of data and, again, in the area of global land cover maps some articles have already discussed the subject (Foody et al., 2013), but still, when modelling is concerned, the crowdsourced data are treated as traditional data and the issue of quality is hardly addressed (albeit recognized as an issue). To which extent does this assumption hold? What is the uncertainty in citizen science data? Lastly, there is a challenge mentioned by many studies but not really addressed in itself and it is the volume of data. Although the volume of data necessary depends on the objective of the modelling effort, the volume of crowdsourced data tends to be low, lacking temporal/spatial coverage for integration into models. This leads to the question: How to increase the volume of data? Considering this limitation, it is also natural to move towards the question: How much data is needed to improve the model significantly?

Application of citizen science in modelling brings an extra challenge of interdisciplinary. Among similar technical fields (e.g. geosciences and hydrodynamic modelling) there is an issue of technology transfer to be addressed, and there are discussions on underlying assumptions and uncertainties that need to be considered. Additionally, hard and soft sciences are also very linked, as the quality and value of the citizens' observations and their temporal/spatial coverage are intrinsically related to social drivers such as why citizens engage, for how long, with which frequency and what is the role of various stakeholders.

5 Conclusions and recommendations

Citizen science has successfully made its way in many scientific domains and it is only fair that the contribution of citizens to modelling floods is also investigated, due to the related intensive data needs. Analysis of literature clearly shows an increasing number of scientific studies in this area. Successful examples of using existing measurement and analysis methods (e.g. velocity and land cover) and of modelling floods with citizen science data (e.g. social media mining) have been published and are seen as a good basis for further exploration. There is a clear need to standardise and consolidate methodologies and there are challenges involving temporal and spatial distribution of data, uncertainty and volume.

It can be observed that the role of citizen contributions is not only in providing information about the current state of the environment, in monitoring and mapping studies, but also in providing data that can be used in its modelling and forecasting. Studies reviewed in this article showed that crowdsourced data can be integrated: in model building, to improve their overall performance; and directly into models (by data assimilation), to improve immediate forecasts. These are promising studies, however still too few, and they highlight the need for further work in this direction. The integration of crowdsourced data

into flood models is a viable way to help solve issues of data scarcity in both ungauged catchments and systems subject to change.

One of the challenges worth mentioning is the integration of citizen data with other more traditional data sources like gauging and remote sensing. It is also necessary to analyse cases in which citizens are involved at higher levels of engagement (e.g. participating in the problem definition, analysis of results and even in the decision-making process) and to evaluate the trade-off between model data needs and levels of engagement.

Finally, there is the challenge to make citizen contributions valuable in a time where automation in gaining increasing space. One may say that citizens are not needed because of automated sensors. At the same time, there are situations where crowdsourced data are very valuable. One of the non-technical challenges that we see here is to demonstrate such situations and increase acceptance of crowdsourced data by water managers.

References

Albrecht, F., Zussner, M., Perger, C., Dürauer, M., See, L., McCallum, I., Fritz, S. and Wolfgang, K.: Using Student Volunteers to Crowdsource Land Cover Information, in GI_Forum 2014. Geospatial Innovation for Society, pp. 314–317, Austrian Academy of Sciences Press., 2014.

Alfonso, L., Lobbrecht, A. and Price, R.: Using Mobile Phones To Validate Models of Extreme Events, in 9th International Conference on Hydroinformatics, pp. 1447–1454, Tianjin, China., 2010.

Alfonso, L., Chacón, J. C. and Peña-Castellanos, G.: Allowing citizens to effortlessly become rainfall sensors, in E-proceedings of the 36th IAHR World Congress, The Hague, The Netherlands., 2015.

Aspuru, I., García, I., Herranz, K. and Santander, A.: CITI-SENSE: methods and tools for empowering citizens to observe acoustic comfort in outdoor public spaces, Noise Mapp., 3(1), 37–48, doi:10.1515/noise-2016-0003, 2016.

Aulov, O., Price, A. and Halem, M.: AsonMaps: A Platform for Aggregation Visualization and Analysis of Disaster Related Human Sensor Network Observations, in ISCRAM 2014 Proceedings - 11th International Conference on Information Systems for Crisis Response and Management, Pennsylvania, USA., 2014.

Di Baldassarre, G., Schumann, G., Brandimarte, L. and Bates, P.: Timely Low Resolution SAR Imagery To Support Floodplain Modelling: A Case Study Review, Surv. Geophys., 32(3), 255–269, doi:10.1007/s10712-011-9111-9, 2011.

Le Boursicaud, R., Pénard, L., Hauet, A., Thollet, F. and Le Coz, J.: Gauging extreme floods on YouTube: application of LSPIV to home movies for the post-event determination of stream discharges, Hydrol. Process., 30(1), 90–105, doi:10.1002/hyp.10532, 2016.

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., De Bïvre, B., Bhusal, J., Clark, J.,
5 Dewulf, A., Foggin, M., Hannah, D. M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., Tilahun, S., Van Hecken, G. and Zhumanova, M.: Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development, Front. Earth Sci., 2, 26, doi:10.3389/feart.2014.00026, 2014.

Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J. and Waters, N.: Using Twitter for tasking remote-sensing data
10 collection and damage assessment: 2013 Boulder flood case study, Int. J. Remote Sens., 37(1), 100–124, doi:10.1080/01431161.2015.1117684, 2016.

Conrad, C. C. and Hilchey, K. G.: A review of citizen science and community-based environmental monitoring: issues and opportunities, Environ. Monit. Assess., 176(1–4), 273–291, doi:10.1007/s10661-010-1582-5, 2011.

Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R.,
15 Dramais, G. and Braud, I.: Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand, J. Hydrol., 541, 766–777, doi:10.1016/j.jhydrol.2016.07.036, 2016.

Craglia, M., Ostermann, F. and Spinsanti, L.: Digital Earth from vision to practice: making sense of citizen-generated content, Int. J. Digit. Earth, 5(5), 398–416, doi:10.1080/17538947.2012.712273, 2012.

Dashti, S., Palen, L., Heris, M. P., Anderson, K. M., Anderson, T. J. and Anderson, S.: Supporting Disaster Reconnaissance
20 with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods, in Proceedings of the 11th International ISCRAM Conference, Pennsylvania, USA., 2014.

Degrossi, L. C., Porto De Albuquerque, J., Fava, M. C. and Mendiondo, E. M.: Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil, in SEKE, pp. 570–575., 2014.

Dong, J., Xiao, X., Sheldon, S., Biradar, C., Duong, N. D. and Hazarika, M.: A comparison of forest cover maps in Mainland
25 Southeast Asia from multiple sources: PALSAR, MERIS, MODIS and FRA, Remote Sens. Environ., 127, 60–73, doi:10.1016/j.rse.2012.08.022, 2012.

Dorn, H., Vetter, M. and Höfle, B.: GIS-Based Roughness Derivation for Flood Simulations: A Comparison of Orthophotos, LiDAR and Crowdsourced Geodata, Remote Sens., 6(2), 1739–1759, doi:10.3390/rs6021739, 2014.

Evers, M., Jonoski, A., Almoradie, A. and Lange, L.: Environmental Science & Policy Collaborative decision making in
30 sustainable flood risk management: A socio-technical approach and tools for participatory governance, Environ. Sci. Policy, 55, 335–344, doi:10.1016/j.envsci.2015.09.009, 2016.

Fava, M. C., Santana, G., Bressiani, D. A., Rosa, A., Horita, F. . E. A. and Mendiondo, E. M.: Integration of Information Technology Systems for Flood Forecasting With Hybrid Data Sources, in 6th International Conference on Flood Management, pp. 1–11, São Paulo, Brazil., 2014.

Fohringer, J., Dransch, D., Kreibich, H. and Schröter, K.: Social media as an information source for rapid flood inundation mapping, Nat. Hazards Earth Syst. Sci., 15(12), 2725–2738, doi:10.5194/nhess-15-2725-2015, 2015.

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C. and Boyd, D. S.: Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project, Trans. GIS, 17(6), 847–860, doi:10.1111/tgis.12033, 2013.

Friedrichs, A., Busch, J. A., Henkel, R., Heuermann, R., John, C. and Zielinski, O.: Measuring fluorescence by means of smartphones with the new Citclops-Application, in Ocean Optics XXII, Portland, USA., 2014.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F. and Obersteiner, M.: Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover, Remote Sens., 1(3), 345–354, doi:10.3390/rs1030345, 2009.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F. and Obersteiner, M.: Geo-Wiki: An online platform for improving global land cover, Environ. Model. Softw., 31, 110–123, doi:10.1016/j.envsoft.2011.11.015, 2012.

Gallart, F., Llorens, P., Latron, J., Cid, N., Rieradevall, M. and Prat, N.: Validating alternative methodologies to estimate the regime of temporary rivers when flow data are unavailable, Sci. Total Environ., 565, 1001–1010, doi:10.1016/j.scitotenv.2016.05.116, 2016.

Gharesifard, M. and Wehn, U.: To share or not to share: Drivers and barriers for sharing data via online amateur weather networks, J. Hydrol., 535, 181–190, doi:10.1016/j.jhydrol.2016.01.036, 2016.

Grimaldi, S., Li, Y., Pauwels, V. R. N. and Walker, J. P.: Remote Sensing-Derived Water Extent and Level to Constrain Hydraulic Flood Forecasting Models: Opportunities and Challenges, Surv. Geophys., 37(5), 977–1034, doi:10.1007/s10712-016-9378-y, 2016.

Haklay, M.: Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation, in Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice, edited by D. Sui, S. Elwood, and M. Goodchild, pp. 105–122, Springer Netherlands., 2013.

Heipke, C.: Crowdsourcing geospatial data, ISPRS J. Photogramm. Remote Sens., 65(6), 550–557, doi:10.1016/j.isprsjprs.2010.06.005, 2010.

Higgins, C. I., Williams, J., Leibovici, D. G., Simonis, I., Davis, M. J., Muldoon, C., Van Genuchten, P. and O 'hare, G.: Citizen OBservatory WEB (COBWEB): A generic infrastructure platform to facilitate the collection of citizen science data for environmental monitoring, Int. J. Spat. Data Infrastructures Res., 11, 20–48, doi:10.2902/1725-0463.2016.11.art3, 2016.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. F. G., Zipf, A. and Porto De Albuquerque, J.: The use of Volunteered Geographic Information and Crowdsourcing in Disaster Management: a Systematic Literature Review, in Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, USA., 2013.

IDEM: Volunteer Stream Monitoring Training Manual., 2015.

Iwao, K., Nishida, K., Kinoshita, T. and Yamagata, Y.: Validating land cover maps with Degree Confluence Project information, Geophys. Res. Lett., 33(23), L23404, doi:10.1029/2006GL027768, 2006.

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J. and Xiao, T.: An automated method for extracting rivers and lakes from Landsat imagery, Remote Sens., 6(6), 5067–5089, doi:10.3390/rs6065067, 2014.

Kim, S., Robson, C., Zimmerman, T., Pierce, J. and Haber, E. M.: Creek watch: Pairing Usefulness and Usability for Successful Citizen Science, in Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11, p. 2125, ACM Press, New York, New York, USA., 2011.

Klonner, C., Marx, S., Usón, T., Porto de Albuquerque, J. and Höfle, B.: Volunteered Geographic Information in Natural Hazard Analysis: A Systematic Literature Review of Current Approaches with a Focus on Preparedness and Mitigation, ISPRS Int. J. Geo-Information, 5(7), 103, doi:10.3390/ijgi5070103, 2016.

Kotovirta, V., Toivanen, T., Järvinen, M., Lindholm, M. and Kallio, K.: Participatory surface algal bloom monitoring in Finland in 2011–2013, Environ. Syst. Res., 3(1), 24, doi:10.1186/s40068-014-0024-8, 2014.

Kutija, V., Bertsch, R., Glenis, V., Alderson, D., Parkin, G., Walsh, C. L., Robinson, J. and Kilsby, C.: Model Validation Using Crowd-Sourced Data From a Large Pluvial Flood, in 11th International Conference on Hydroinformatics, New York, USA., 2014.

Leibovici, D. G., Evans, B., Hodges, C., Wiemann, S., Meek, S., Rosser, J. and Jackson, M.: On Data Quality Assurance and Conflation Entanglement in Crowdsourcing for Environmental Studies, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., II-3/W5, 195–202, doi:10.5194/isprsannals-II-3-W5-195-2015, 2015.

Lesiv, M., Moltchanova, E., Schepaschenko, D., See, L., Shvidenko, A., Comber, A. and Fritz, S.: Comparison of Data Fusion Methods Using Crowdsourced Data in Creating a Hybrid Forest Cover Map, Remote Sens., 8(3), 261, doi:10.3390/rs8030261, 2016.

Li, Z., Wang, C., Emrich, C. T. and Guo, D.: A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods, Cartogr. Geogr. Inf. Sci., 0(0), 1–14, doi:10.1080/15230406.2016.1271356, 2017.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P. and Vandenberg, J.: Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, Mon. Not. R. Astron. Soc., 389(3), 1179–1189, doi:10.1111/j.1365-2966.2008.13689.x, 2008.

Lowry, C. S. and Fienen, M. N.: CrowdHydrology: Crowdsourcing Hydrologic Data and Engaging Citizen Scientists, Ground Water, 51(1), 151–156, doi:10.1111/j.1745-6584.2012.00956.x, 2013.

Mazzoleni, M., Alfonso, L., Chacon-Hurtado, J. and Solomatine, D.: Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models, Adv. Water Resour., 83, 323–339, doi:10.1016/j.advwatres.2015.07.004, 2015.

Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M. and Solomatine, D. P.: Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?, Hydrol. Earth Syst. Sci., 21(2), 839–861, doi:10.5194/hess-21-839-2017, 2017.

McDougall, K.: Using volunteered information to map the Queensland floods, in Surveying & Spatial Science Biennial Conference, Wellington, New Zealand., 2011.

McDougall, K. and Temple-Watts, P.: The use of LidAR and Volunteered Geographic Information to map flood extents and inundation, in ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 251–256, Melbourne, Australia., 2012.

Merkuryeva, G., Merkuryev, Y., Sokolov, B. V., Potryasaev, S., Zelentsov, V. A. and Lektauers, A.: Advanced river flood monitoring, modelling and forecasting, J. Comput. Sci., 10, 77–85, doi:10.1016/j.jocs.2014.10.004, 2015.

Michelsen, N., Dirks, H., Schulz, S., Kempe, S., Al-Saud, M. and Schüth, C.: YouTube as a crowd-generated water level archive, Sci. Total Environ., 568, 189–195, doi:10.1016/j.scitotenv.2016.05.211, 2016.

Miller-Rushing, A., Primack, R. and Bonney, R.: The history of public participation in ecological research, Front. Ecol. Environ., 10(6), 285–290, doi:10.1890/110278, 2012.

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. R.: Crowdsourcing for climate and atmospheric sciences: Current status and future potential, Int. J. Climatol., 35(11), 3185–3203, doi:10.1002/joc.4210, 2015.

Oxendine, C. E. ., Schnebele, E. ., Cervone, G. . and Waters, N. .: Fusing non-authoritative data to improve situational awareness in emergencies, in Proceedings of the 11th International ISCRAM Conference, pp. 762–766, Pennsylvania, USA., 2014.

Padawangi, R., Turpin, E., Herlily, Prescott, M. F., Lee, I. and Shepherd, A.: Mapping an alternative community river: The case of the Ciliwung, Sustain. Cities Soc., 20, 147–157, doi:10.1016/j.scs.2015.09.001, 2016.

Poser, K. and Dransch, D.: Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation, Geomatica, 64(1), 89–98, 2010.

Rosser, J. F., Leibovici, D. G. and Jackson, M. J.: Rapid flood inundation mapping using social media, remote sensing and topographic data, Nat. Hazards, 87(1), 103–120, doi:10.1007/s11069-017-2755-0, 2017.

Salk, C. F., Sturn, T., See, L., Fritz, S. and Perger, C.: Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game, Int. J. Digit. Earth, 9(4), 410–426, doi:10.1080/17538947.2015.1039609, 2016.

Schnebele, E. and Cervone, G.: Improving remote sensing flood assessment using volunteered geographical data, Nat. Hazards Earth Syst. Sci., 13(3), 669–677, doi:10.5194/nhess-13-669-2013, 2013.

Schnebele, E., Cervone, G., Kumar, S. and Waters, N.: Real Time Estimation of the Calgary Floods Using Limited Remote Sensing Data, Water, 6(2), 381–398, doi:10.3390/w6020381, 2014.

See, L., Schepaschenko, D., Lesiv, M., McCallum, I., Fritz, S., Comber, A., Perger, C., Schill, C., Zhao, Y., Maus, V., Siraj, M. A., Albrecht, F., Cipriani, A., Vakolyuk, M., Garcia, A., Rabia, A. H., Singha, K., Marcarini, A. A., Kattenborn, T., Hazarika, R., Schepaschenko, M., van der Velde, M., Kraxner, F. and Obersteiner, M.: Building a hybrid land cover map with crowdsourcing and geographically weighted regression, ISPRS J. Photogramm. Remote Sens., 103, 48–56, doi:10.1016/j.isprsjprs.2014.06.016, 2015a.

See, L., Fritz, S., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., Duerauer, M., Sturn, T., Karner, M., Kraxner, F. and Obersteiner, M.: Harnessing the power of volunteers, the internet and Google Earth to collect and validate global spatial information using Geo-Wiki, Technol. Forecast. Soc. Change, 98, 324–335, doi:10.1016/j.techfore.2015.03.002, 2015b.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H. and Milˇ, G.: Crowdsourcing , Citizen Science or Volunteered Geographic Information ? The Current State of Crowdsourced Geographic Information, , doi:10.3390/ijgi5050055, 2016.

Shaad, K., Ninsalam, Y., Padawangi, R. and Burlando, P.: Towards high resolution and cost-effective terrain mapping for urban hydrodynamic modelling in densely settled river-corridors, Sustain. Cities Soc., 20, 168–179, doi:10.1016/j.scs.2015.09.005, 2016.

Smith, L., Liang, Q., James, P. and Lin, W.: Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework, J. Flood Risk Manag., doi:10.1111/jfr3.12154, 2015.

Starkey, E., Parkin, G., Birkinshaw, S., Large, A., Quinn, P. and Gibson, C.: Demonstrating the value of community-based ("citizen science") observations for catchment modelling and characterisation, J. Hydrol., 548, 801–817, doi:10.1016/j.jhydrol.2017.03.019, 2017.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D. and Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environ. Model. Softw., 90, 201–216, doi:10.1016/j.envsoft.2017.01.006, 2017.

Uhrner, U., Grosso, G., Romain, A. C., Hutsemekers, V., Delva, J., Kunz, W., De Groof, A., Arnaud, Y., Valoggia, P., Johannsen, L., Stevenot, B. and Ledent, P.: Development of an environmental information system for odour using citizen and technology innovative sensors and advanced modeling, in CEUR Workshop Proceedings, vol. 1322, Aachen, Germany., 2013.

De Vos, L., Leijnse, H., Overeem, A. and Uijlenhoet, R.: The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam, Hydrol. Earth Syst. Sci., 21(2), 765–777, doi:10.5194/hess-21-765-2017, 2017.

Walker, D., Forsythe, N., Parkin, G. and Gowing, J.: Filling the observational void: Scientific value and quantitative validation of hydrometeorological data from a community-based monitoring programme, J. Hydrol., 538, 713–725, doi:10.1016/j.jhydrol.2016.04.062, 2016.

Yang, Y. Y. and Kang, S. C.: Crowd-based velocimetry for surface flows, Adv. Eng. Informatics, 32, 275–286, doi:10.1016/j.aei.2017.03.007, 2017.

Yu, D., Yin, J. and Liu, M.: Validating city-scale surface water flood modelling using crowd-sourced data, Environ. Res. Lett., 11(12), 124011, doi:10.1088/1748-9326/11/12/124011, 2016.

4. Revised manuscript

The revised manuscript without marked-up changes in included in this section.

Citizen observations contributing to flood modelling: opportunities and challenges

Thaine Herman Assumpção[1], Ioana Popescu[1], Andreja Jonoski[1], Dimitri P. Solomatine[1, 2]

[1]Integrated Water Systems and Governance, IHE Delft, Delft, The Netherlands
5 [2]Water Resources Section, Delft University of Technology, Delft, The Netherlands

Correspondence to: Thaine Herman Assumpção (t.hermanassumpcao@un-ihe.org)

Abstract. Citizen contributions to science have been successfully implemented in many fields – and water resources is one of them. Through citizens, it is possible to collect data and obtain a more integrated decision-making process. Specifically, data scarcity has always been an issue in flood modelling, which has been addressed in the last decades by remote sensing
10 and is already being discussed in a citizen science scenario. In this context, this article aims to review the literature on the topic and analyse the opportunities and challenges that lie ahead. The literature on monitoring, mapping and modelling, was evaluated according to the flood-related variable citizens contributed to. Pros and cons of the collection/analysis methods were summarised. Then, pertinent publications were mapped into the flood modelling cycle, considering how citizen data properties (spatial and temporal coverage, uncertainty and volume) are related to its integration into modelling. It was clear
15 that the number of studies in the area is rising. There are positive experiences reported in collection and analysis methods, for instance with velocity and land cover, and also when modelling is concerned, for example by using social media mining. However, matching the data properties necessary for each part of the modelling cycle with citizen generated data is still challenging. Nevertheless, the concept that citizen contributions can be used for simulation and forecasting is proved and further work lies in continuing developing and improving not only methods for collection and analysis but certainly for
20 integration into models as well. Finally, in view of recent automated sensors and satellite technologies, it is through studies as the ones analysed in this article that the value of citizen contributions is demonstrated.

1 Introduction

The necessity to understand and predict the behaviour of floods has been present in societies around the world. This comes from the fact that floods impact their surroundings - in negative or in positive ways. The most common way used nowadays
25 to better understand and often predict flood behaviour is through modelling and, depending on the system at hand, a variety of models can be used (Teng et al., 2017).

In order to have adequate representation of floods, most models require large amounts of data, both for model building and model usage. This is especially true for pluvial flood modelling, where flooding may not occur in gauged rivers and hence,

flow gauging stations outside of flooded zones may be of little use. Remote sensing technologies are a part of the solution, as they offer spatially distributed information. However, their availability may be limited, also in terms of space and time, and their uncertainties often are not quantifiable (Di Baldassarre et al., 2011; Grimaldi et al., 2016; Jiang et al., 2014; Li et al., 2017). Thus, acquiring the necessary data for simulations and predictions can still be expensive, particularly for rapidly changing systems that require frequent model updates.

In this context, sources of data coming in abundance and at low-costs are needed, together with modified modelling approaches that can use these data and can adapt to changes as fast as they occur. Citizen Observatory (CO) is an emerging concept in which citizens monitor the environment around them. It is often considered under the umbrella of Citizen Science (including citizen participation up to the scientist level) and it is also related to the concept of crowdsourcing (distributing a task among many agents). With technology at hand, it is possible to empower citizens to not only participate in the acquisition of data but also in the process of scientific analysis and even in the consequent decision-making process (Evers et al., 2016). Citizen Observatories have been researched in several EU-funded projects. Finished projects (CITI-SENSE, Citclops, COBWEB, OMNISCIENTIS and WeSenseIt) already resulted in valuable contributions to the field (Alfonso et al., 2015; Aspuru et al., 2016; Friedrichs et al., 2014; Higgins et al., 2016; Uhrner et al., 2013). For example, the CITI-SENSE project managed to simultaneously collect perception data and acoustic measurements in an approach that can be used to develop citizen empowerment initiatives in case of noise management (Aspuru et al. 2016); while in COBWEB project processes of quality assurance, data conflation and data fusion were studied and recommendations were made (Friedrichs et al., 2014). The currently running CO projects (Ground Truth 2.0, LANDSENSE, SCENT and GROW Observatory) propose to investigate this concept further.

Citizen science concepts have been researched and applied in various fields such as ecology and galaxy inspection (Lintott et al., 2008; Miller-Rushing et al., 2012). Volunteer Geographic Information (VGI), as one of the most active citizen science areas, has developed over the past decade and several researchers reviewed the state of the art of citizen science in the field of geosciences (Heipke, 2010; Klonner et al., 2016). There is also a part of the scientific community dedicated to investigating damage data crowdsourced after flood emergencies (Dashti et al., 2014; Oxendine et al., 2014) and evaluating the cycle of disaster management (Horita et al., 2013). In the context of water resources, Buytaert et al. (2014) reviewed and discussed the contribution of citizen science to hydrology and water resources, addressing the level of engagement, the type of data collected (e.g. precipitation, water level) and case studies where more participatory approaches are being implemented. Le Coz et al. (2016) provided examples and reflections from three projects related to flood hydrology and crowdsourcing, which involve the derivation of hydraulic information from pictures and videos in Argentina, France and New Zealand.

The present review aims to look at studies that had citizen science connected to floods. Specifically, it focusses on the data collected by citizens that are relevant in a flood modelling context, benchmarking difficulties and benefits of their collection and integration into models. Integration is considered for the purposes of model set up, calibration, validation, simulation and forecasting.

5

The review process involved defining web platforms, keywords and criteria for searching and selecting publications. The main platforms used were Scopus and Google Scholar. The keywords are a combination of words related to citizen science (e.g. "citizen science" and crowdsourcing) and to flood-related variables (e.g. "water level" and "flood extent"). The obtained articles were scanned for their content. Articles were selected mainly if crowdsourced data was obtained for

10 quantitative use in monitoring, mapping or modelling. There were studies that were not selected because they just mention the use of crowdsourced data and do not provide more relevant information on collection, analysis, use and quantity of data, such as Merkuryeva et al. (2015). The same is the case of studies that evaluate variables qualitatively, in ways that cannot be directly associated with modelling (Kim et al., 2011). This review included articles published up to April 2017.

15 Further in this section, we introduce the citizen science concept and related classification systems. In Sect. 2 of the article, we overview studies on citizen contributions for flood modelling, classifying them according to the flood-related variable the contributions were made, followed by a summary of the pros and cons of measurement and analysis methods. Section 3 aggregates the studies that involve flood modelling and analyses the contributions considering the component of the modelling process where they were used, also including a discussion of the factors that affect flood modelling. Section 4

20 describes the challenges and opportunities of using data contributed by citizens in flood modelling, and finally, Sect. 5 presents the conclusions and recommendations.

1.1 Citizen Science

Buytaert et al. (2014) defined citizen science as "the participation of the general public (i.e. non-scientists) in the generation

25 of new knowledge". In the same manner that the involvement of citizens can be diverse, such is the way their participation is found in the scientific literature:

- Citizen Science (Buytaert et al., 2014)
- Citizen Observatory (Degrossi et al., 2014)
30 - Citizen Sensing (Foody et al., 2013)
- Trained volunteers (Gallart et al., 2016)
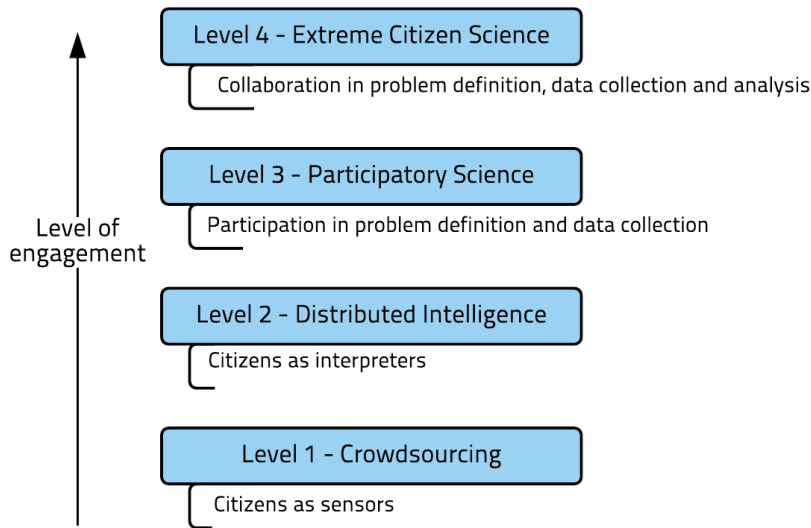- Participatory data collection methods (Michelsen et al., 2016)

- Crowdsourcing (Leibovici et al., 2015)

- Participatory sensing (Kotovirta et al., 2014)

- Community-based monitoring (Conrad and Hilchey, 2011)

- Volunteered Geographic Information (Klonner et al., 2016)

5 - Eye witnesses (Poser and Dransch, 2010)

- Non-authoritative sources (Schnebele et al., 2014)

- Human Sensor Network (Aulov et al., 2014)

- Crowdsourced Geographic Information (See et al., 2016)

10　Some of the terms used by the above-mentioned articles have specific definitions that are used to delineate debates on the social mechanisms of citizen participation. Others are just the best form the researcher found to characterise the contribution or the citizen (e.g. eye witnesses). Citizen Science and adjacent areas have become fields of research in themselves that, for instance, focus on understanding the motivation of citizens or its interaction with public institutions (Gharesifard and Wehn, 2016).

15

In this field, one of the classifications of citizen science is by level of engagement. Haklay (2013) built a model that has four levels (Fig. 1), in which the first one refers to the participation of citizens only as data collectors, passing through a second level in which citizens are asked to act as interpreters of data, going towards the participation in definition of the problem in the third level and finally, being fully involved in the scientific enterprise at hand. The review presented in this current article

20　is focused on the contribution towards flood modelling only, coming most prominently from the two lowest levels of engagement. We do not discuss topics related to engagement for the generation of (quantitative) data. Further in this article, for readability, only the term crowdsourced data is used to refer to data from these two levels of engagement.

Figure 1: Levels of participation and engagement in citizen science projects. Adapted from Haklay (2013).

Another way to classify citizen science initiatives (within the context of VGI) is by setting them as implicitly/explicitly volunteered and implicitly/explicitly geographic (Craglia et al., 2012). In this classification system, geographic refers to the main information conveyed through the contributed data, therefore, geo-tagged data is not necessarily geographic. For example, in the Degree Confluence Project (Iwao et al., 2006), citizens were oriented to go to certain locations, take pictures, make notes and deliberately make available their material on the project's website. In this case, the information is explicitly volunteered and explicitly geographic. Most land use/cover projects related to citizen science are explicitly geographic. Differently, in the study conducted by Lowry and Fienen (2013) citizens would also willingly send text messages to the researchers, in this case providing water level readings from installed water level gauges. Although explicitly volunteered, the message was non-geographic (just geo-tagged). Another type of implicitly geographic information was derived from Twitter by Smith et al. (2015) to obtain water level, velocity and flood extent estimates. As the citizens did not make the information public with the specific purpose to provide estimates, it is implicitly volunteered.
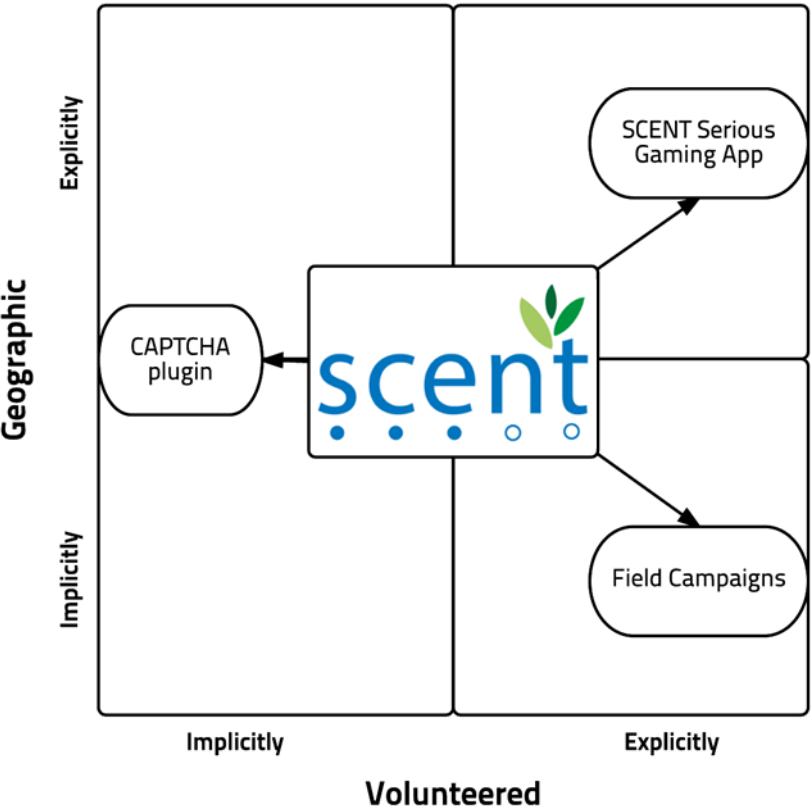
The concepts defined by Craglia et al. (2012) can be graphically represented as in Fig. 2. The SCENT project[1] (Smart Toolbox for Engaging Citizens in a People-Centric Observation Web) is one of the four Horizon 2020-funded projects focussing on citizen observatories. It lies in the middle of this quadrant as it encourages citizens to participate in gaming to collect land cover/use data, in field campaigns to collect other implicitly geographic information (e.g. water level), and also

---

[1] https://scent-project.eu/

aims to obtain implicitly volunteered contributions through a CAPTCHA[2] plugin, in which citizens tag images, e.g. of land cover/use or water level, in order to access online content. Tagging images is uncorrelated to the CAPTCHA, it is a task performed after the test, on the same platform.

Figure 2: SCENT project represented in the typology of VGI (Volunteered Geographic Information)

2 Flood-related crowdsourced data

There are many types of data which relate to floods that can be collected by citizens. Likewise, there are many ways to collect, analyse and use them (for monitoring, mapping and modelling). In the next sub-sections we address how these aspects were explored in the scientific literature. Each sub-section discusses a data type corresponding to a flood modelling

---

[2] CAPTCHA stands for 'Completely Automated Public Turing test to tell Computers and Humans Apart'. It is a test evaluating if the subject is human, which is used in websites to provide security. After the test is done the user can be asked to perform extra tasks, for example, tag images.

variable: water level, velocity, flood extent, land cover and topography. Depending on the type of flooding, other variables are relevant, such as precipitation. The scientific literature already shows that citizens' contributions could be useful for observation this variable (Muller et al., 2015; De Vos et al., 2017). However, rainfall is not included in this section because it was already covered by the review of Muller et al. (2015). Moreover, in general it is a variable of greater importance for hydrological models, whilst the present review is focussed on a hydrodynamic representation of floods. There are articles mentioned and reviewed in more than one section because they evaluated more than one variable, as it is, for example, the case of Smith et al. (2015).

2.1 Water level

Table 1 gives an overview of the articles about collection of water level data. The studies presented started to involve citizens in the collection of water level data with the explicit goal of improving flood management. This is due to the ease of collecting such data, which mostly consists of comparing the water level with a clearly defined reference. In some cases, the reference is a water level gauge, the comparison is made by the citizen, and readings are being submitted to the researchers (Alfonso et al., 2010; Degrossi et al., 2014; Fava et al., 2014; Lowry and Fienen, 2013; Walker et al., 2016). Such kind of reading practically do not require further analysis, although they entail the installation of water level gauges.

In other cases, the citizen provides qualitative data that will be compared to references by researchers. Mostly during flooding situations, citizens provide pictures (Fohringer et al., 2015; Kutija et al., 2014; Li et al., 2017; McDougall, 2011; McDougall and Temple-Watts, 2012; Smith et al., 2015; Starkey et al., 2017) or videos (Le Boursicaud et al., 2016; Le Coz et al., 2016; Michelsen et al., 2016). In the case of pictures/images, the water level is compared with objects in the images that have known or approximately known dimensions. For videos, although water level was estimated, the main goal was to obtain discharge values, via estimates of flow velocity. In two cases, texts from citizens were used (e.g. water over the knee), to calculate water level values or to assume a certain value when no value was provided (Li et al., 2017; Smith et al., 2015). This sort of data (text, pictures and videos) was mostly collected through social media and public image repositories. Gathering data from such sources requires mining of the relevant material (i.e. extraction of specific data from a dataset) and dealing with uncertainties in the spatio-temporal characterization of the data of interest.

One aspect that varies across the studies is the level of detail in the comparison method used for determining the water level measurement. For example, McDougall (2011) and McDougall and Temple-Watts (2012) explicitly state that field visits to the selected photo locations are required in order to properly analyse the image and extract water level values. On the other hand, Fohringer et al. (2015), Smith et al. (2015) and Starkey et al. (2017) do not mention any method.
In most cases, crowdsourcing has been used to monitor water level, followed by the use of such data for modelling and lastly for mapping. In the case of Starkey et al. (2017), although hydrological modelling was done and water levels were converted into discharge to allow for comparisons, only qualitative comparisons were made.

Table 1: Scientific literature on citizen contributions to measurement and analysis of water level

Study	Measurement/analysis methods	Type	Purpose	Flood type	Location
Alfonso et al. (2010)	Citizen's reading of water level gauges sent by text message	1D	Monitoring	No flooding	The Netherlands
Lowry and Fienen (2013)	Citizen's reading of water level gauges sent by text message	1D	Monitoring	No flooding	USA
Degrossi et al. (2014)	Citizen's reading of water level gauge sent through app/webpage	1D	Monitoring	No flooding	Brazil
Walker et al. (2016)	Citizen's reading of water level gauge collected and provided by the community	1D	Monitoring	No flooding	Ethiopia
Fava et al. (2014)	Citizen's reading of water level gauge sent through app/webpage	1D	Modelling	Flood forecasting	Brazil
Le Boursicaud et al. (2016)	LSPIV analysis of video collected from social media (YouTube)	1D	Monitoring	Flash flood	France
Le Coz et al. (2016)	LISPIV analysis of video sent through webpage	2D	Modelling	Fluvial flood	Argentina
Michelsen et al. (2016)	Analysis of images extracted from videos collected from social media (YouTube) and own photographs	Neither	Monitoring	No flooding	Saudi Arabia
Li et al. (2017)	Analysis of texts and pictures collected from social media (Twitter)	2D	Monitoring	Flood map	USA
Starkey et al. (2017)	Citizen's reading of water level gauge and analysis of pictures and videos collected from social media (Twitter) and crowdsourced (email, webpage and mobile app)	2D	Monitoring	Flood	UK
McDougall (2011), McDougall and Temple-Watts (2012)	Analysis of texts and pictures collected from social media (Twitter, Facebook) and crowdsourced (email, text message, Ushahidi, Flickr and Picasa)	2D	Mapping	Flood map	Australia
Kutija et al. (2014)	Analysis of pictures collected by the University and City Council	2D	Modelling	Pluvial and drainage flood	UK
Aulov et al. (2014)	Visual analysis of texts and pictures collected from social media (Twitter and Instagram)	2D	Modelling	Coastal flood	USA
Fohringer et al. (2015)	Visual analysis of pictures collected from social media (Twitter) and crowdsourced (Flickr)	2D	Mapping	Flood	Germany
Smith et al. (2015)	Analysis of texts and pictures collected from social media (Twitter)	2D	Modelling	Pluvial and drainage flood	UK

8

2.2 Velocity

As velocities and discharges traditionally require more complex measuring methods, the collection of this type of data by citizens has not been explored on a scientific basis. However, it is common to include direct measurements of velocity in protocols to monitor the environment and water quality, as it is the case of Hoosier Riverwatch (IDEM, 2015). In these cases, the citizens perform measurements that involve more processing (e.g. definition of transects to measure flow, use of formulas).

To the best of the authors' knowledge, only three studies were found that make use of velocity data collected by citizens, all for the study of floods, as presented in Table 2. Le Boursicaud et al. (2016) evaluated the surface velocity field in a channel from a YouTube video, using the LSPIV methodology (Large Scale Particle Image Velocimetry), an established method to obtain velocity from a sequence of images. For enabling this analysis, information about the camera (model and lens type) is needed, visible, fixed elements are needed to be used as reference points and it is also required that both river banks are visible. Although the method calculates the velocity in two dimensions, in Table 2 we referred to it as 1D because it was carried out in a channel, which in a context of flood modelling is considered as a 1D domain. A complementary project was discussed by Le Coz et al. (2016), in which the same technique is applied to a video crowdsourced by a citizen, this time using the result to estimate discharge and the latter to calibrate a 1D hydraulic model. For this, a visit to the location was needed to extract cross-sectional data. In this context, Yang and Kang (2017) developed a method for crowd-based velocimetry of surface flows, based on Particle Image Velocimetry, in which citizens mark features in the picture. The method has not been tested with citizen collected data yet.

The third study, conducted by Smith et al. (2015), selected Twitter messages that include terms of semantic value related to the citizen location, water depth (e.g. knee-deep) and velocity. The terms were then associated with quantitative values/ranges. The authors did not go into detail on discussing the reliability and uncertainty in such data, even though the issue is recognised.

Table 2: Scientific literature on citizen contributions to measurement and analysis of velocity

Study	Measurement/analysis methods	Type	Purpose	Flood type	Location
Le Boursicaud et al. (2016)	LSPIV analysis of video collected from social media (YouTube)	1D	Monitoring	Flash flood	France
Le Coz et al. (2016)	LSPIV analysis of video sent through webpage	2D	Modelling	Fluvial flood	Argentina
Smith et al. (2015)	Analysis of texts and pictures collected from social media (Twitter)	2D	Modelling	Pluvial and drainage flood	UK

2.3 Flood extent

Flood extent, similarly to water level, is a variable that is simple to measure as it consists of binary values: flooded or non-flooded area. As a 2D variable, it needs a lot of spatial information and it is the main reason related studies gather flood extent estimates in data rich environments, through social media/photo sharing services mining, as shown in Table 3. In some cases, the citizens act only as sensors, providing pictures to be analysed by the research team, while in other cases they also act as interpreters by providing the flooded/non-flooded information. As can be expected, all studies found were carried out in urban areas.

In some of the studies the text and images are indicating the location of their origin as being flooded (georeferenced or inferred) (Aulov et al., 2014; Smith et al., 2015; Yu et al., 2016), whilst in others (Cervone et al., 2016; Li et al., 2017; Rosser et al., 2017; Schnebele et al., 2014; Schnebele and Cervone, 2013) there is processing of the information to infer the surrounding inundated areas. Additionally, the last group of studies mentioned fused flood extent data from citizens with satellite data or with gauge data.

Table 3: Scientific literature on citizen contributions to measurement and analysis of flood extent

Study	Measurement/analysis methods	Purpose	Flood type	Location
Cervone et al. (2016), Schnebele et al. (2014), Schnebele and Cervone (2013)	Analysis of pictures and videos collected from social media (Facebook and YouTube) and crowdsourced (Flickr)	Mapping	Flood map	USA and Canada
Li et al. (2017)	Analysis of texts and pictures collected from social media (Twitter)	Mapping	Flood map	USA
Rosser et al. (2017)	Analysis of crowdsourced pictures (Flickr)	Mapping*	Flood map	UK
Aulov et al. (2014)	Visual analysis of texts and pictures collected from social media (Twitter and Instagram)	Modelling	Coastal flood	USA
Smith et al. (2015)	Analysis of texts and pictures collected from social media (Twitter)	Modelling	Pluvial and drainage flood	UK
Yu et al. (2016)	Citizen's visual identification of flooded location collected by governmental Chinese	Modelling	Pluvial and drainage	China

	website		flood		
Padawangi et al. (2016)	Citizen information		Monitoring	Flood	Indonesia

* A statistical model is created, but in this study we consider only physical models in the modelling category

2.4 Land cover/Land use

Land cover is not a variable in flood-related models but we include it in this review for its importance in inferring roughness. Other valuable aspects of land use data are the information on roads and structures that can be obstacles to floods, which can

5 be incorporated in the model structure; and the information on vulnerability (e.g. hospitals, dense residential areas, industrial zones), which can be used to obtain flood risk maps. According to Klonner et al. (2016), when reviewing the literature on VGI for natural hazard analysis, there are few studies for vulnerability analysis. The aspects of land use related to vulnerability and risk are complex and study topics on themselves, so these aspects are not discussed further in this article. Table 4 presents the articles considered for this review. Compared to previously discussed variables, the contribution of

10 citizens to land cover maps generation has been already proved as a concept (Albrecht et al., 2014; Fritz et al., 2012), nowadays being researched further for quality of data (Salk et al., 2016) and fusion of maps (Lesiv et al., 2016).

One of the first publications on the subject was from Iwao et al. (2006), in which they describe the Degree Confluence Project. The objective was to generate a global land cover map, which implies obtaining ground truth data from around the

15 globe. For obvious reasons, it was unfeasible to make field campaign or analyse low-resolution images with sufficient resolution. Thus, they launched a webpage that invited citizens to visit integer coordinates (e.g. 25° W, 25°) locations, take photos from the four cardinal directions and provide comments on the region. They discovered that citizen-generated data was having quality similar to that provided by specialists.

20 Another significant project in the area is GeoWiki. It started in 2009 as a platform for people to validate global land cover maps, by comparing their classification to high-resolution images (Fritz et al., 2009). The project has grown since and has recently achieved its main goal: to generate a hybrid global land cover map by fusing existing maps and performing calibration and validation using the analyses made by citizens (See et al., 2015). Current initiatives in the GeoWiki project include gamification and analysis of pictures uploaded onto the platform (See et al., 2015). Many studies stemmed from the

25 data collected, generally focused on specific land cover types. A similar approach is taken by Dong et al. (2012), that analyses pictures uploaded by citizens using a different web application. The research conducted by Dorn et al. (2014) goes one step further, as it attributes roughness values to multiple land cover maps, including Open Street Maps (a website where citizens can modify the current street and land cover map).

11

Table 4: Scientific literature on citizen contributions to measurement and analysis of land cover/land use

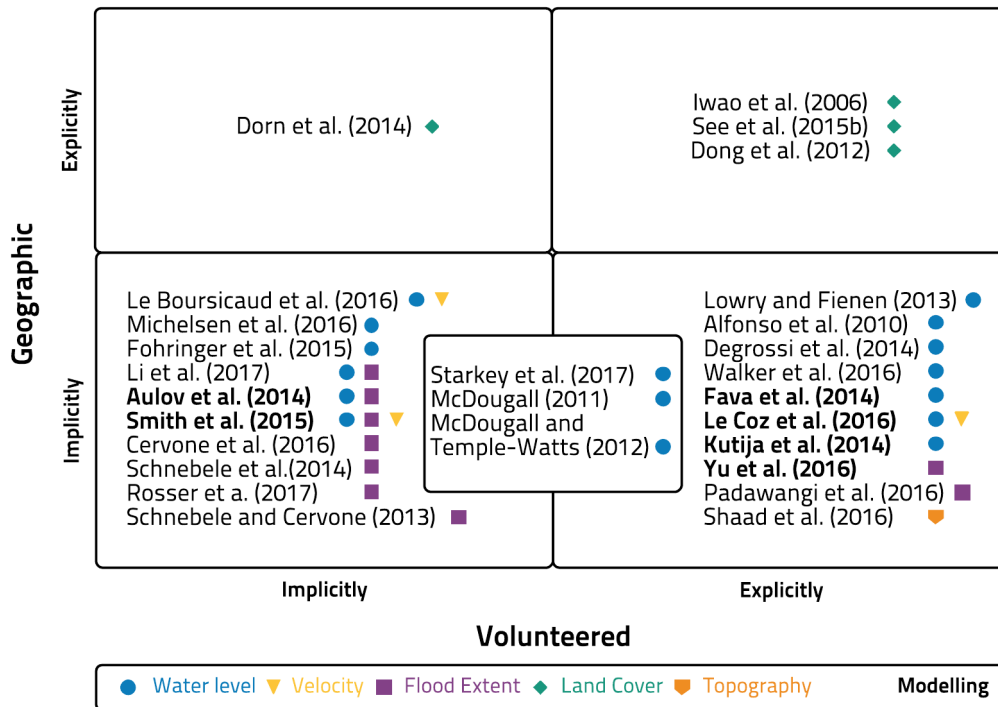| Study	Measurement/analysis methods	Purpose	Flood type	Location
Iwao et al. (2006)	Visual interpretation of crowdsourced tagged pictures sent through app/webpage (Degree Confluence Project website)	Mapping	No flooding	Global land cover map
See et al. (2015b)*	Visual interpretation of Google Earth and pictures sent through app/webpage (GeoWiki)	Mapping	No flooding	Global land cover map
Dong et al. (2012)	Analysis of tagged pictures from Global Geo-Referenced Field Photo Library (DCP citizen pictures + field trip pictures)	Mapping	No flooding	Forest cover map in Asia
Dorn et al. (2014)	Use of Open Street Maps	Modelling	Fluvial flood	Austria

* Many other articles related to crowdsourcing through GeoWiki

2.5 Topography

The Digital Elevation Model (DEM) is one of the most important components in flood modelling, as it generally heavily influences flood propagation. It is particularly important in urban settings, where spatial variability in refined scales has a considerable effect on the direction of water flows. Unfortunately, this is a complex variable to measure that so far relies either on fully trained professionals to go to the field, or on expensive airborne technologies. Recently, Shaad et al. (2016) studied a terrain capturing low-cost alternative to LiDAR remote sensing images and other expensive methods. The low-cost technique is the ground-based close-range photogrammetry (CRP) that consists of collecting images/videos from the ground, post-processing them and obtaining terrain information. Volunteers made the videos in a designated location, where even Unmanned Aerial Vehicles (UAVs) would not be able to collect data. After comparing the results to other methods, they concluded that the result has an acceptable quality.
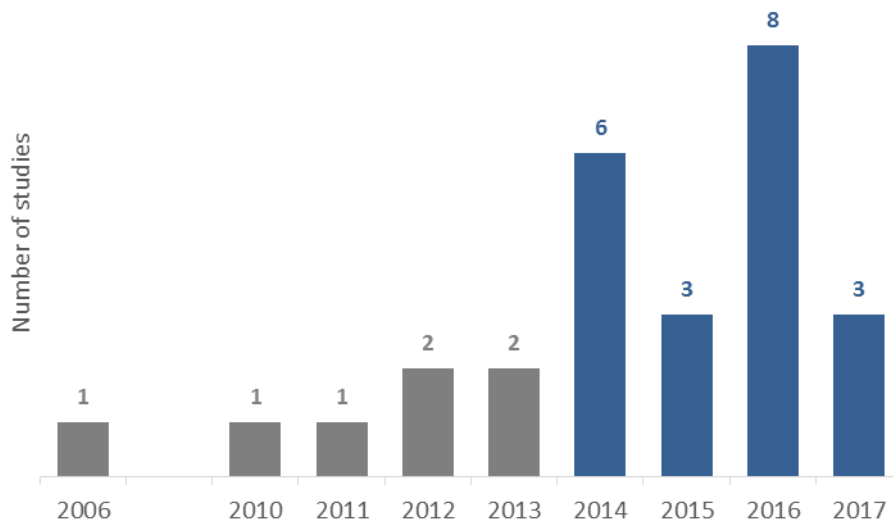
2.6 Summary analysis

By classifying the discussed studies according to Craglia et al. (2012), there is an overall similarity in the number of studies that crowdsource data implicitly and explicitly (Fig. 3). It is visible though that this aspect does not translate into homogeneous distribution per flood-related variables, with most implicitly volunteered contributions being related to flood extent and most explicit being related to water level. There is a slightly higher concentration of modelling studies that are explicitly volunteered, but not enough to be able to draw any conclusions.

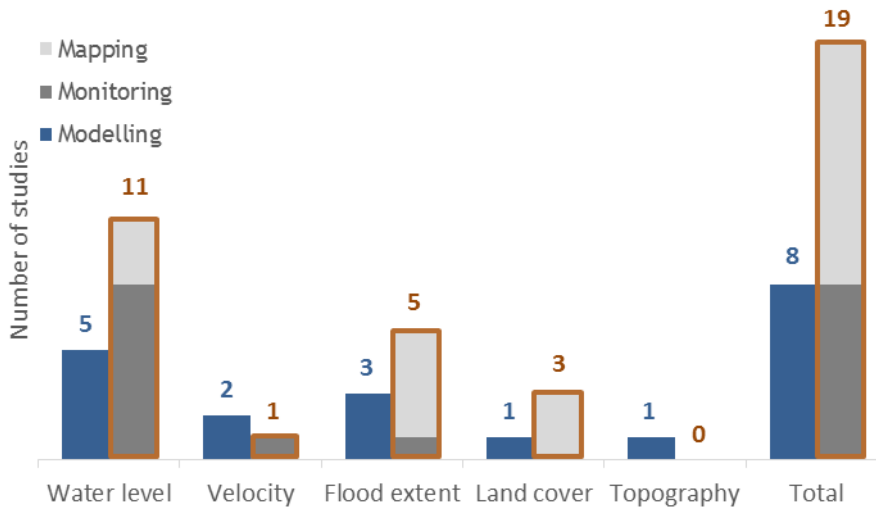**Figure 3: Selected studies represented in the typology of VGI (Volunteered Geographic Information)**

5   Considering the temporal distribution of studies evaluated in this review, it is evident that there is a trend: the rise in number of studies from 2014 onwards (Fig. 4). This relates to the initial barrier in acknowledging citizen data as having quality that is high enough for scientific studies (Buytaert et al., 2014). This resistance is reducing over time as such data is being proved useful, protocols are being designed and the data uncertainty is being better understood and quantified.

Figure 4: Number of studies analysed per year

If the analysed studies are aggregated into categories (Fig. 5), it can be seen that modelling studies amount to approximately

5   the same quantity as monitoring ones, but they are only about a third of all studies reviewed. This is expected because to use

data in models it is necessary to monitor them first. Also, monitoring and mapping applications attend to more general end

uses. Specifically for land cover, there is an unexplored field in modelling (there are more mapping studies than the ones in

the graph, see Sect. 2.4). The reason behind may be that modellers do not tend to validate their own land cover maps and

thus will not do it with citizen science data. What can be noted though, is the lack of exploration of velocity and topography

10 variables, which, as mentioned, can be due to the complexity in analysing and setting up the experiment.

Figure 5: Number of studies analysed per flood-related variable per category: mapping, monitoring and modelling

Related to that, previous sub-sections discussed in detail the methods for collection and analysis of flood-related data obtained through crowdsourcing. For example, water level data obtained from reading a water level gauge is easy to collect and easy to analyse. On the other hand, it requires the installation of gauges (Fig. 6). In summary, whenever data is collected from the Internet, there is the disadvantage of needing social media/photo sharing services mining, entailing computational efforts and dealing with a high percentage of data that is not georeferenced or time stamped. Further, in the case of water level and velocity, some studies suggest that also field visits are necessary and the methods to analyse data are complex. Considering crowdsourced data on land cover and topography, it is straightforward to measure and analyse them, although their delivery to the interested parties may require a smartphone app or a website to be set up and maintained (with the exception of Open Street Maps).

	Collection	Analysis
Reading water level gauge	Measurement (+) Instalation of gauges (-)	Analysis (+)
Collection of data from social media/Flickr/Picasa	Social media/Flickr/Picasa mining (-) Location and time (-)	Flood extent analysis (+) Water level and velocity analysis (-) Field visit for water level (-)
Collection of data from Youtube	Youtube mining (-) Location and time (-)	Established method (+) Field visit (-) Camera and other set ups (-)
Crowdsourced data	Measurement (+) App/website (-)	Analysis (+)
Visual analysis of tagged images		Analysis (+) App/website (-)
Open Street Maps editing	Measurement (+) OSM land cover types (-)	Analysis (+)
Crowdsourced video	Measurement (+) App/website or youtube mining (-)	Established method (+)

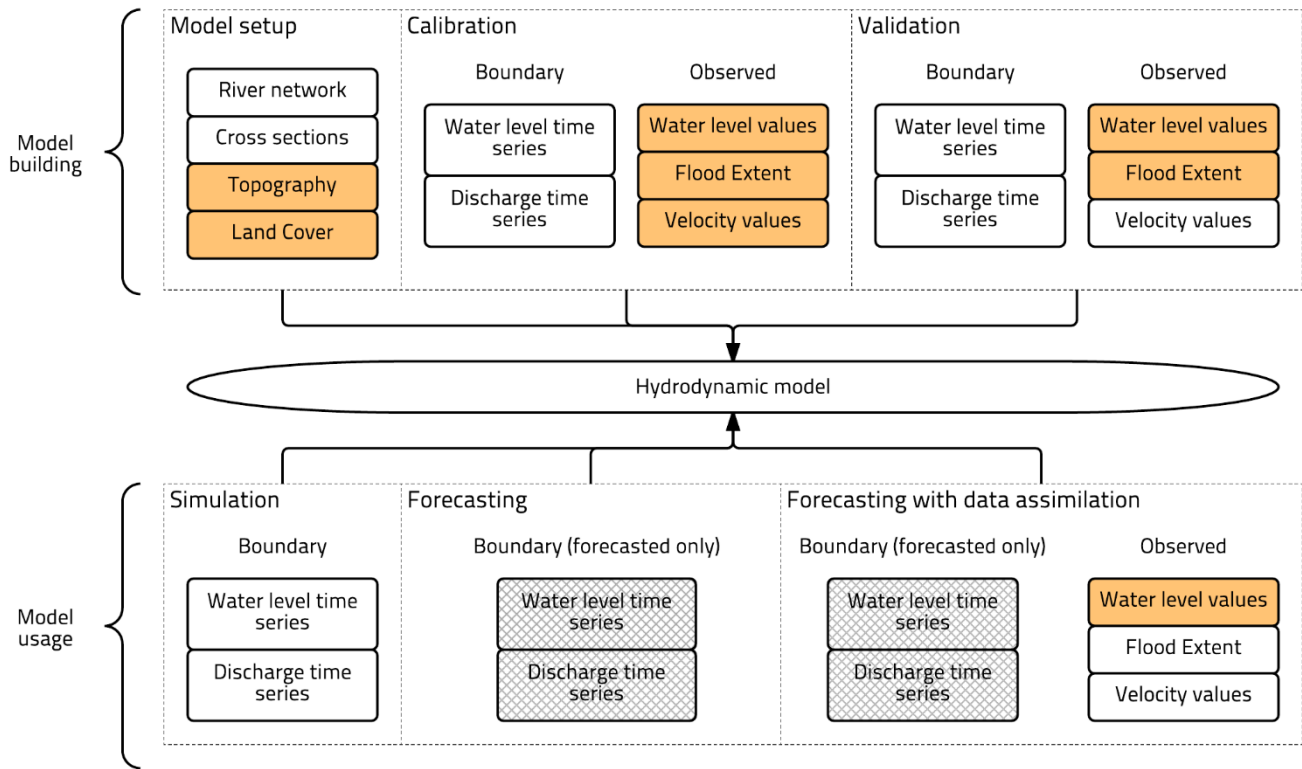(Left labels: Water level, Velocity, Flood extent, Land Cover, Topography)

Figure 6: Pros and cons of collection and analysis methods used to collect flood-related data by citizens

3 Crowdsourced data in flood modelling

By concentrating on the studies in which modelling was performed, we explore in detail how crowdsourced data was
5 integrated into each component of flood models.

There is a variety of flood models developed to deal with different types of flood, including: fluvial, pluvial, coastal and drainage floods. The main driver of fluvial floods is upstream river discharge, of pluvial floods it is precipitation and of coastal floods it is storm surges. In urban drainage floods, the flows inside, through and outside of drainage systems are pivotal for flood representation. Moreover, there are complex cases where more than one flood process needs to be
10 represented. Although in physically-based flood models water flow is computed by the same principles, different sets of data are needed for different types of flood models. We focus on a general hydrodynamic model definition and its common inputs but present what was the flood type evaluated in the scientific literature (Table 5).

The flood modelling process typically has two parts: model building, and model usage. (Fig. 7). Model building starts by defining the model setup (boundary conditions, parameters, schematization, input data), followed by calibration and
15 validation of the water level and velocity fields (dependent variables) with observed values. Calibration and validation can be performed for both simulation and forecasting models. Once the model is ready, simulations can be run by using different boundary conditions or introducing designed measures for better flood management; or forecasts can be made by using forecasted water levels or discharges as boundaries. In a simulation setting, model parameters are assumed to be constant in time, while in a forecasting setting the parameters, inputs or states (water levels) can be updated while the model is in use,
20 using data assimilation.

16

Figure 7: Flood models data requirements. Orange coloured tiles correspond to data that citizens have contributed to in a flood modelling context and gridded tiles correspond to data citizens cannot contribute to (forecasted water levels and discharges).

. From the studies analysed (Table 5), three consider 1D channels and the others worked in a 2D setting. Most of them analyse only one variable, except Smith et al. (2015) that evaluate water level and velocity. Moreover, most of them model urban floods, some in a pluvial and others in a fluvial context.

Table 5: Scientific literature on crowdsourced data used in flood modelling

Use in modelling	Study	Measurement method	Type	Variable	Flood type	Location

17

Model setup	Dorn et al. (2014)	Use of Open Street Maps	2D	Land cover	Fluvial flood	Austria
	Shaad et al. (2016)	Analysis of pictures captured by volunteers at selected location	2D	Topography	Fluvial flood	Indonesia
Calibration	Smith et al. (2015)*	Analysis of pictures and tweets collected from social media (Twitter)	2D	Water level and velocity	Pluvial and drainage flood	UK
	Le Coz et al. (2016)	LSPIV analysis of videos sent through webpage	1D	Velocity	Fluvial flood	Argentina
	Yu et al. (2016)	Citizen's visual identification of flooded location provided through Chinese website	2D	Flood extent	Pluvial and drainage flood	China
Validation	Kutija et al. (2014)	Analysis of pictures collected from the University and City Council	2D	Water level	Pluvial and drainage flood	UK
	Yu et al. (2016)	Citizen's visual identification of flooded location provided through Chinese website	2D	Flood extent	Pluvial and drainage flood	China
Data assimilation	Aulov et al. (2014)	Visual analysis of texts and pictures collected from social media (Twitter and Instagram)	2D	Water level and flood extent	Coastal flood	USA
	Mazzoleni et al. (2015, 2017)	Simulated citizen reading of water level gauge sent through app	1D	Water level	Flood forecasting without flood model	Italy and USA
	Fava et al. (2014)	Citizen's reading of a water level gauge sent through app or webpage	1D	Water level	Flood forecasting without flood model	Brazil

* It is classified as calibration because, in the classical sense, it improves the model according to observations. However, what actually is done is the fine-tuning selection of the precipitation field that fits the observations better.

Considering model building, specifically the model setup, citizens contributed to improving/updating land cover (and consequently roughness) and topography information. Dorn et al. (2014) used the land cover information contained in Open Street Maps[3] for modelling a fluvial flood.. They do not analyse how much contribution was made by the citizens and data processing is restricted to attributing land cover classes to the features displayed in the maps. In the study of Shaad et al.

---

[3] Open Street Maps (OSM) is an online platform that provides street maps and other information. The maps provided can be edited by the users at any time

(2016), which addresses topography, there is only one citizen contribution (low-cost alternative) in one selected location that is merged with an existing DEM and then used in the model. In both cases, the objective was to compare the performance of this low-cost alternative against the performance of consolidated technologies when used for hydrodynamic simulations.

5   Crowdsourced data has also been used to calibrate and validate flood models in four studies. One study gathered such data through social media and public image repositories mining and the others through data uploaded by citizens on specific platforms. Smith et al. (2015) identified storm events through social media, triggering shock-capturing hydrodynamic model runs with various rainfall intensities. The results were compared with social media data on water level/velocity. The comparison consisted of defining a buffer zone around the crowdsourced observation location, built a histogram of simulated
10   cell values within it and evaluating the overlap of crowdsourced value/range and the histogram 70-95th percentile range. As most citizen contributions did not have a water level.velocity value, they received a minimum water level value. Because of that, the selected simulation was the one with more 'overlaps' and that would not perform better than a simulation with rainfall slightly higher. Yu et al. (2016) collected flooded data through a Chinese website and divided it into calibration and validation data sets for a pluvial flood model verification. There is no mentioning on how this data is provided (e.g. text or
15   image). Le Coz et al. (2016) obtained a discharge value for calibration of a hydraulic model based on the surface velocity data obtained by a video uploaded to a specific website. Kutija et al. (2014) collected pictures uploaded by citizens and extract from them water levels by comparison with reference objects, such as cars (no further detailing on the method of extraction is made). Water level data is then used to validate a pluvial flood model.

20   The described approaches so far consider citizen data for model building and its possible extension for recalibration and revalidation. Four studies went one step further, integrating crowdsourced data in model usage. Mazzoleni et al. (2015, 2017) used synthetically generated data to represent citizen observations, which were incorporated in the model through data assimilation algorithms, adapted to deal with the intermittent nature of crowdsourced data. Aulov et al. (2014) and Fava et al. (2014) also used the data for simulation/data assimilation, but the methods used are not detailed in the studies. However, the
25   studies of Mazzoleni et al. (2015, 2017) and Fava et al. (2014) were made for flood forecasting through hydrological models and not using hydrodynamic models.

3.1 Crowdsourced data information content

If we aim at integrating data into model, data accuracy, volume and temporal and spatial coverage should be at a certain level. When these data properties are inadequate, data integration would not provide useful results (i.e. the model
30   performance can be low). Although most modelling variables vary in time and space, the data does not need to cover all dimensions in all parts of the modelling process. For instance, in model setup, topographic data is not needed every 15 minutes, hourly or daily; it can be provided in a discrete time coverage, from months to years. We analyse four data properties: temporal coverage, spatial coverage, volume and uncertainty (Table 6). Although same for all parts, the last two

properties vary significantly when analysing the information content of crowdsourced data and that is why these properties are included (Table 6).

Table 6: Data properties currently required in the modelling process

	Setup	**Calibration & Validation**[1]	**Simulation**	**Data assimilation**	**Data assimilation**
	Topography Land Cover	*Water Level Velocity Flood Extent*	*Water Level Velocity*	*Water Level Velocity*	*Flood Extent*
Temporal coverage	Discrete	Discrete/Continuous	Continuous	Variable	Variable
Spatial coverage	Distributed	Discrete/Distributed	Discrete	Discrete	Unknown
Uncertainty	The lower the better				
Volume	The higher the better				

5   [1] Dependent on purpose of the model

Analysing crowdsourcing studies by their information content, it is possible to draw the following conclusions:

10
- Model setup: for integration of topographic and land cover data, it is necessary to have spatially distributed data. While this has been achieved within land cover studies, there is only one study involving topography and the data obtained so far have discrete spatial coverage.

- Calibration and validation: through mining and crowdsourcing of water level and flood extent estimates, spatially distributed crowdsourced data have already been obtained for calibration/validation of simulation models. The accuracy of the time stamp was considered vital (Kutija et al., 2015) and results in time have a preliminary good
15   level of agreement with citizen observations (Yu et al., 2016). However, even though these studies compare the results with citizen observations in time, this is done qualitatively and there is no focus on reporting and evaluating the temporal coverage.

- Simulation: traditional modelling efforts require time series of data at specific frequencies, which has only been achieved through crowdsourcing in the realm of community-based approaches, in which water levels are measured
20   at 6 a.m. and 6 p.m. in agreement with the community (Walker et al., 2016). However, this type of data has been only monitored and not used in a modelling context so far.

- Data assimilation: it generally assimilates data provided with a fixed time frequency, but there are a few studies that consider intermittent data to be assimilated (Mazzoleni et al., 2015, 2017). However, similarly to simulation, the temporal coverage of crowdsourced data is insufficient for data assimilation efforts.

25

Considering uncertainty, this is highly dependent on the collection/analysis method. For example, obtaining water level values from pictures of flooded areas (2D) is uncertain, as it mostly involves the selection of what constitutes a good

reference point to be made by the citizen. Flood extent, on the other hand, tends to be less uncertain to measure, due to its binary nature. The collection through data mining (and sometimes crowdsourcing) has, in general, more sources of uncertainty: from geotagging, timestamping and the observed value. To deal with the first two, Aulov et al. (2014) used only data that contained proper geotag and time stamp. Kutija et al. (2014) classified non-timestamped data as during or after the event, based on picture visual inspection, defining an observation time range. Smith et al. (2015) dealt with uncertainty in location by generating a histogram of simulated values around the observed point. Yu et al. (2016) acknowledged these sources of uncertainty. Regarding uncertainty in value, existent in all sources of crowdsourced data, most studies used the (processed) observations as were, without indication of uncertainty. Smith et al. (2015) defined ranges, although these are not discussed. Mazzoleni et al. (2015, 2017), used uncertain synthetic crowdsourced data with variable uncertainty.

Regarding volume of data collected, this is an issue for all modelling processes, although data mining has again been able to provide a better coverage. Besides the challenge of uncertainty, data mining has also the challenge in providing data in conditions that are not extreme, as most of the contributions are done in floods situations and it is limited to certain variables (water level, flood extent and velocity). Some of the studies were proof of concepts and integrated up to 3 crowdsourced observations each (Le Coz et al., 2016; Fava et al.; 2014; Shaad et al., 2016). Others ranged from 12 to 298 observations (Kutija et al., 2014; Smith et al., 2015; Yu et al., 2016) and in some cases it was not possible to define the exact number (Aulov et al., 2014; Dorn et al., 2014).

4 Opportunities and challenges

In the last years, the interest in citizen science and the number of citizen science studies in the water resources context has risen considerably. The main factors affecting its use in flood modelling are the degree of how difficult it is to acquire and evaluate these data and their integration into the models. Our analysis of the existing literature allows for pointing out a number of positive experiences from which we can derive opportunities to:

- ▪ Explore and improve the existing methods to obtain water velocity and topography from videos
- ▪ Explore calibration and validation employing data collected through social media in urban environments
- ▪ Explore the possibilities of setting up the models with the use of land cover maps validated with citizen science
- ▪ Make use of apps/websites already developed for citizen science

The first one is based on small scale but successful studies related to using well-developed techniques in a citizen science scenario. The relevant experience in data gathering and analysis can be updated to fit the needs of flood modelling. Also, social media and public image repositories mining has proved to be successful in calibration and validation in modelling studies, proving the concept and opening the opportunity to investigate how large this contribution is. As mentioned

previously, in the field of land cover map generation citizen data has been used to validate maps and this successful example could be used to obtain new roughness maps in a modelling context. Lastly, technological development of apps, websites and techniques could be shared and put to public use, to be tested further and to avoid duplicated work.

5 There are aspects of the integration of crowdsourced data into flood modelling that are still challenging. These are:

- ▪ Explore the use of citizens as data interpreters
- ▪ Improve methods to estimate water level from pictures
- ▪ Harmonise the time frequency and spatial distribution of models with the ones of crowdsourced data
10 - ▪ Quantification of uncertainty
- ▪ Increase the volume of data gathered, mainly in non-urban environments

Most of the analysed studies regard the citizen as a sensor, with the exception of studies about land cover related data, in which the citizen also acts as an interpreter. For other variables, some studies have already started evaluating the ability of
15 citizens to provide interpreted information (Degrossi et al., 2014), but these are few. Regarding water levels, readings from rulers and extraction from pictures are described differently in the literature, with varying degrees of thoroughness, indicating a need for development and testing of water level measurement methodologies in the context of citizens' contributions. The third point brings up a challenge that concerns not only citizen science but also modelling: what is the necessary temporal and spatial distribution? Is the traditional modelling approach definitive in terms of data requirements
20 and citizen science approaches should adapt to it, or, the modelling process can be adapted to receive citizen science data? The fourth challenge relates to the quality of data and, again, in the area of global land cover maps some articles have already discussed the subject (Foody et al., 2013), but still, when modelling is concerned, the crowdsourced data are treated as traditional data and the issue of quality is hardly addressed (albeit recognized as an issue). To which extent does this assumption hold? What is the uncertainty in citizen science data? Lastly, there is a challenge mentioned by many studies but
25 not really addressed in itself and it is the volume of data. Although the volume of data necessary depends on the objective of the modelling effort, the volume of crowdsourced data tends to be low, lacking temporal/spatial coverage for integration into models. This leads to the question: How to increase the volume of data? Considering this limitation, it is also natural to move towards the question: How much data is needed to improve the model significantly?

30 Application of citizen science in modelling brings an extra challenge of interdisciplinary. Among similar technical fields (e.g. geosciences and hydrodynamic modelling) there is an issue of technology transfer to be addressed, and there are discussions on underlying assumptions and uncertainties that need to be considered. Additionally, hard and soft sciences are also very linked, as the quality and value of the citizens' observations and their temporal/spatial coverage are intrinsically

related to social drivers such as why citizens engage, for how long, with which frequency and what is the role of various stakeholders.

5 Conclusions and recommendations

Citizen science has successfully made its way in many scientific domains and it is only fair that the contribution of citizens to modelling floods is also investigated, due to the related intensive data needs. Analysis of literature clearly shows an increasing number of scientific studies in this area. Successful examples of using existing measurement and analysis methods (e.g. velocity and land cover) and of modelling floods with citizen science data (e.g. social media mining) have been published and are seen as a good basis for further exploration. There is a clear need to standardise and consolidate methodologies and there are challenges involving temporal and spatial distribution of data, uncertainty and volume.

It can be observed that the role of citizen contributions is not only in providing information about the current state of the environment, in monitoring and mapping studies, but also in providing data that can be used in its modelling and forecasting. Studies reviewed in this article showed that crowdsourced data can be integrated: in model building, to improve their overall performance; and directly into models (by data assimilation), to improve immediate forecasts. These are promising studies, however still too few, and they highlight the need for further work in this direction. The integration of crowdsourced data into flood models is a viable way to help solve issues of data scarcity in both ungauged catchments and systems subject to change.

One of the challenges worth mentioning is the integration of citizen data with other more traditional data sources like gauging and remote sensing. It is also necessary to analyse cases in which citizens are involved at higher levels of engagement (e.g. participating in the problem definition, analysis of results and even in the decision-making process) and to evaluate the trade-off between model data needs and levels of engagement.

Finally, there is the challenge to make citizen contributions valuable in a time where automation in gaining increasing space. One may say that citizens are not needed because of automated sensors. At the same time, there are situations where crowdsourced data are very valuable. One of the non-technical challenges that we see here is to demonstrate such situations and increase acceptance of crowdsourced data by water managers.

References

Albrecht, F., Zussner, M., Perger, C., Dürauer, M., See, L., McCallum, I., Fritz, S. and Wolfgang, K.: Using Student Volunteers to Crowdsource Land Cover Information, in GI_Forum 2014. Geospatial Innovation for Society, pp. 314–317, Austrian Academy of Sciences Press., 2014.

5   Alfonso, L., Lobbrecht, A. and Price, R.: Using Mobile Phones To Validate Models of Extreme Events, in 9th International Conference on Hydroinformatics, pp. 1447–1454, Tianjin, China., 2010.

Alfonso, L., Chacón, J. C. and Peña-Castellanos, G.: Allowing citizens to effortlessly become rainfall sensors, in E-proceedings of the 36th IAHR World Congress, The Hague, The Netherlands., 2015.

Aspuru, I., García, I., Herranz, K. and Santander, A.: CITI-SENSE: methods and tools for empowering citizens to observe

10   acoustic comfort in outdoor public spaces, Noise Mapp., 3(1), 37–48, doi:10.1515/noise-2016-0003, 2016.

Aulov, O., Price, A. and Halem, M.: AsonMaps: A Platform for Aggregation Visualization and Analysis of Disaster Related Human Sensor Network Observations, in ISCRAM 2014 Proceedings - 11th International Conference on Information Systems for Crisis Response and Management, Pennsylvania, USA., 2014.

Di Baldassarre, G., Schumann, G., Brandimarte, L. and Bates, P.: Timely Low Resolution SAR Imagery To Support

15   Floodplain Modelling: A Case Study Review, Surv. Geophys., 32(3), 255–269, doi:10.1007/s10712-011-9111-9, 2011.

Le Boursicaud, R., Pénard, L., Hauet, A., Thollet, F. and Le Coz, J.: Gauging extreme floods on YouTube: application of LSPIV to home movies for the post-event determination of stream discharges, Hydrol. Process., 30(1), 90–105, doi:10.1002/hyp.10532, 2016.

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., De Biềvre, B., Bhusal, J., Clark, J.,

20   Dewulf, A., Foggin, M., Hannah, D. M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., Tilahun, S., Van Hecken, G. and Zhumanova, M.: Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development, Front. Earth Sci., 2, 26, doi:10.3389/feart.2014.00026, 2014.

Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J. and Waters, N.: Using Twitter for tasking remote-sensing data

25   collection and damage assessment: 2013 Boulder flood case study, Int. J. Remote Sens., 37(1), 100–124, doi:10.1080/01431161.2015.1117684, 2016.

Conrad, C. C. and Hilchey, K. G.: A review of citizen science and community-based environmental monitoring: issues and opportunities, Environ. Monit. Assess., 176(1–4), 273–291, doi:10.1007/s10661-010-1582-5, 2011.

Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R.,

30   Dramais, G. and Braud, I.: Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand, J. Hydrol., 541, 766–777, doi:10.1016/j.jhydrol.2016.07.036, 2016.

Craglia, M., Ostermann, F. and Spinsanti, L.: Digital Earth from vision to practice: making sense of citizen-generated content, Int. J. Digit. Earth, 5(5), 398–416, doi:10.1080/17538947.2012.712273, 2012.

Dashti, S., Palen, L., Heris, M. P., Anderson, K. M., Anderson, T. J. and Anderson, S.: Supporting Disaster Reconnaissance with Social Media Data : A Design-Oriented Case Study of the 2013 Colorado Floods, in Proceedings of the 11th International ISCRAM Conference, Pennsylvania, USA., 2014.

Degrossi, L. C., Porto De Albuquerque, J., Fava, M. C. and Mendiondo, E. M.: Flood Citizen Observatory : a crowdsourcing-based approach for flood risk management in Brazil, in SEKE, pp. 570–575., 2014.

Dong, J., Xiao, X., Sheldon, S., Biradar, C., Duong, N. D. and Hazarika, M.: A comparison of forest cover maps in Mainland Southeast Asia from multiple sources: PALSAR, MERIS, MODIS and FRA, Remote Sens. Environ., 127, 60–73, doi:10.1016/j.rse.2012.08.022, 2012.

Dorn, H., Vetter, M. and Höfle, B.: GIS-Based Roughness Derivation for Flood Simulations: A Comparison of Orthophotos, LiDAR and Crowdsourced Geodata, Remote Sens., 6(2), 1739–1759, doi:10.3390/rs6021739, 2014.

Evers, M., Jonoski, A., Almoradie, A. and Lange, L.: Environmental Science & Policy Collaborative decision making in sustainable flood risk management : A socio-technical approach and tools for participatory governance, Environ. Sci. Policy, 55, 335–344, doi:10.1016/j.envsci.2015.09.009, 2016.

Fava, M. C., Santana, G., Bressiani, D. A., Rosa, A., Horita, F. . E. A. and Mendiondo, E. M.: Integration of Information Technology Systems for Flood Forecasting With Hybrid Data Sources, in 6th International Conference on Flood Management, pp. 1–11, São Paulo, Brazil., 2014.

Fohringer, J., Dransch, D., Kreibich, H. and Schröter, K.: Social media as an information source for rapid flood inundation mapping, Nat. Hazards Earth Syst. Sci., 15(12), 2725–2738, doi:10.5194/nhess-15-2725-2015, 2015.

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C. and Boyd, D. S.: Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project, Trans. GIS, 17(6), 847–860, doi:10.1111/tgis.12033, 2013.

Friedrichs, A., Busch, J. A., Henkel, R., Heuermann, R., John, C. and Zielinski, O.: Measuring fluorescence by means of smartphones with the new Citclops-Application, in Ocean Optics XXII, Portland, USA., 2014.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F. and Obersteiner, M.: Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover, Remote Sens., 1(3), 345–354, doi:10.3390/rs1030345, 2009.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F. and Obersteiner, M.: Geo-Wiki: An online platform for improving global land cover, Environ. Model. Softw., 31, 110–123, doi:10.1016/j.envsoft.2011.11.015, 2012.

Gallart, F., Llorens, P., Latron, J., Cid, N., Rieradevall, M. and Prat, N.: Validating alternative methodologies to estimate the regime of temporary rivers when flow data are unavailable, Sci. Total Environ., 565, 1001–1010, doi:10.1016/j.scitotenv.2016.05.116, 2016.

Gharesifard, M. and Wehn, U.: To share or not to share: Drivers and barriers for sharing data via online amateur weather networks, J. Hydrol., 535, 181–190, doi:10.1016/j.jhydrol.2016.01.036, 2016.

Grimaldi, S., Li, Y., Pauwels, V. R. N. and Walker, J. P.: Remote Sensing-Derived Water Extent and Level to Constrain Hydraulic Flood Forecasting Models: Opportunities and Challenges, Surv. Geophys., 37(5), 977–1034, doi:10.1007/s10712-016-9378-y, 2016.

Haklay, M.: Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation, in

5 Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice, edited by D. Sui, S. Elwood, and M. Goodchild, pp. 105–122, Springer Netherlands., 2013.

Heipke, C.: Crowdsourcing geospatial data, ISPRS J. Photogramm. Remote Sens., 65(6), 550–557, doi:10.1016/j.isprsjprs.2010.06.005, 2010.

Higgins, C. I., Williams, J., Leibovici, D. G., Simonis, I., Davis, M. J., Muldoon, C., Van Genuchten, P. and O 'hare, G.:

10 Citizen OBservatory WEB (COBWEB): A generic infrastructure platform to facilitate the collection of citizen science data for environmental monitoring, Int. J. Spat. Data Infrastructures Res., 11, 20–48, doi:10.2902/1725-0463.2016.11.art3, 2016.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. F. G., Zipf, A. and Porto De Albuquerque, J.: The use of Volunteered Geographic Information and Crowdsourcing in Disaster Management: a Systematic Literature Review, in Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, USA., 2013.

15 IDEM: Volunteer Stream Monitoring Training Manual., 2015.

Iwao, K., Nishida, K., Kinoshita, T. and Yamagata, Y.: Validating land cover maps with Degree Confluence Project information, Geophys. Res. Lett., 33(23), L23404, doi:10.1029/2006GL027768, 2006.

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J. and Xiao, T.: An automated method for extracting rivers and lakes from Landsat imagery, Remote Sens., 6(6), 5067–5089, doi:10.3390/rs6065067, 2014.

20 Kim, S., Robson, C., Zimmerman, T., Pierce, J. and Haber, E. M.: Creek watch: Pairing Usefulness and Usability for Successful Citizen Science, in Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11, p. 2125, ACM Press, New York, New York, USA., 2011.

Klonner, C., Marx, S., Usón, T., Porto de Albuquerque, J. and Höfle, B.: Volunteered Geographic Information in Natural Hazard Analysis: A Systematic Literature Review of Current Approaches with a Focus on Preparedness and Mitigation,

25 ISPRS Int. J. Geo-Information, 5(7), 103, doi:10.3390/ijgi5070103, 2016.

Kotovirta, V., Toivanen, T., Järvinen, M., Lindholm, M. and Kallio, K.: Participatory surface algal bloom monitoring in Finland in 2011–2013, Environ. Syst. Res., 3(1), 24, doi:10.1186/s40068-014-0024-8, 2014.

Kutija, V., Bertsch, R., Glenis, V., Alderson, D., Parkin, G., Walsh, C. L., Robinson, J. and Kilsby, C.: Model Validation Using Crowd-Sourced Data From a Large Pluvial Flood, in 11th International Conference on Hydroinformatics, New York,

30 USA., 2014.

Leibovici, D. G., Evans, B., Hodges, C., Wiemann, S., Meek, S., Rosser, J. and Jackson, M.: On Data Quality Assurance and Conflation Entanglement in Crowdsourcing for Environmental Studies, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., II-3/W5, 195–202, doi:10.5194/isprsannals-II-3-W5-195-2015, 2015.

Lesiv, M., Moltchanova, E., Schepaschenko, D., See, L., Shvidenko, A., Comber, A. and Fritz, S.: Comparison of Data Fusion Methods Using Crowdsourced Data in Creating a Hybrid Forest Cover Map, Remote Sens., 8(3), 261, doi:10.3390/rs8030261, 2016.

Li, Z., Wang, C., Emrich, C. T. and Guo, D.: A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods, Cartogr. Geogr. Inf. Sci., 0(0), 1–14, doi:10.1080/15230406.2016.1271356, 2017.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P. and Vandenberg, J.: Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, Mon. Not. R. Astron. Soc., 389(3), 1179–1189, doi:10.1111/j.1365-2966.2008.13689.x, 2008.

Lowry, C. S. and Fienen, M. N.: CrowdHydrology: Crowdsourcing Hydrologic Data and Engaging Citizen Scientists, Ground Water, 51(1), 151–156, doi:10.1111/j.1745-6584.2012.00956.x, 2013.

Mazzoleni, M., Alfonso, L., Chacon-Hurtado, J. and Solomatine, D.: Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models, Adv. Water Resour., 83, 323–339, doi:10.1016/j.advwatres.2015.07.004, 2015.

Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M. and Solomatine, D. P.: Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?, Hydrol. Earth Syst. Sci., 21(2), 839–861, doi:10.5194/hess-21-839-2017, 2017.

McDougall, K.: Using volunteered information to map the Queensland floods, in Surveying & Spatial Science Biennial Conference, Wellington, New Zealand., 2011.

McDougall, K. and Temple-Watts, P.: The use of LidAR and Volunteered Geographic Information to map flood extents and inundation, in ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 251–256, Melbourne, Australia., 2012.

Merkuryeva, G., Merkuryev, Y., Sokolov, B. V., Potryasaev, S., Zelentsov, V. A. and Lektauers, A.: Advanced river flood monitoring, modelling and forecasting, J. Comput. Sci., 10, 77–85, doi:10.1016/j.jocs.2014.10.004, 2015.

Michelsen, N., Dirks, H., Schulz, S., Kempe, S., Al-Saud, M. and Schüth, C.: YouTube as a crowd-generated water level archive, Sci. Total Environ., 568, 189–195, doi:10.1016/j.scitotenv.2016.05.211, 2016.

Miller-Rushing, A., Primack, R. and Bonney, R.: The history of public participation in ecological research, Front. Ecol. Environ., 10(6), 285–290, doi:10.1890/110278, 2012.

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. R.: Crowdsourcing for climate and atmospheric sciences: Current status and future potential, Int. J. Climatol., 35(11), 3185–3203, doi:10.1002/joc.4210, 2015.

Oxendine, C. E. ., Schnebele, E. ., Cervone, G. . and Waters, N. .: Fusing non-authoritative data to improve situational awareness in emergencies, in Proceedings of the 11th International ISCRAM Conference, pp. 762–766, Pennsylvania, USA., 2014.

Padawangi, R., Turpin, E., Herlily, Prescott, M. F., Lee, I. and Shepherd, A.: Mapping an alternative community river: The case of the Ciliwung, Sustain. Cities Soc., 20, 147–157, doi:10.1016/j.scs.2015.09.001, 2016.

Poser, K. and Dransch, D.: Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation, Geomatica, 64(1), 89–98, 2010.

5   Rosser, J. F., Leibovici, D. G. and Jackson, M. J.: Rapid flood inundation mapping using social media, remote sensing and topographic data, Nat. Hazards, 87(1), 103–120, doi:10.1007/s11069-017-2755-0, 2017.

Salk, C. F., Sturn, T., See, L., Fritz, S. and Perger, C.: Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game, Int. J. Digit. Earth, 9(4), 410–426, doi:10.1080/17538947.2015.1039609, 2016.

Schnebele, E. and Cervone, G.: Improving remote sensing flood assessment using volunteered geographical data, Nat.

10   Hazards Earth Syst. Sci., 13(3), 669–677, doi:10.5194/nhess-13-669-2013, 2013.

Schnebele, E., Cervone, G., Kumar, S. and Waters, N.: Real Time Estimation of the Calgary Floods Using Limited Remote Sensing Data, Water, 6(2), 381–398, doi:10.3390/w6020381, 2014.

See, L., Schepaschenko, D., Lesiv, M., McCallum, I., Fritz, S., Comber, A., Perger, C., Schill, C., Zhao, Y., Maus, V., Siraj, M. A., Albrecht, F., Cipriani, A., Vakolyuk, M., Garcia, A., Rabia, A. H., Singha, K., Marcarini, A. A., Kattenborn, T.,

15   Hazarika, R., Schepaschenko, M., van der Velde, M., Kraxner, F. and Obersteiner, M.: Building a hybrid land cover map with crowdsourcing and geographically weighted regression, ISPRS J. Photogramm. Remote Sens., 103, 48–56, doi:10.1016/j.isprsjprs.2014.06.016, 2015a.

See, L., Fritz, S., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., Duerauer, M., Sturn, T., Karner, M., Kraxner, F. and Obersteiner, M.: Harnessing the power of volunteers, the internet and Google Earth to collect and validate global spatial

20   information using Geo-Wiki, Technol. Forecast. Soc. Change, 98, 324–335, doi:10.1016/j.techfore.2015.03.002, 2015b.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H. and Milˇ, G.: Crowdsourcing , Citizen Science or Volunteered Geographic Information ? The Current State of Crowdsourced Geographic Information, , doi:10.3390/ijgi5050055, 2016.

Shaad, K., Ninsalam, Y., Padawangi, R. and Burlando, P.: Towards high resolution and cost-effective terrain mapping for

25   urban hydrodynamic modelling in densely settled river-corridors, Sustain. Cities Soc., 20, 168–179, doi:10.1016/j.scs.2015.09.005, 2016.

Smith, L., Liang, Q., James, P. and Lin, W.: Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework, J. Flood Risk Manag., doi:10.1111/jfr3.12154, 2015.

Starkey, E., Parkin, G., Birkinshaw, S., Large, A., Quinn, P. and Gibson, C.: Demonstrating the value of community-based

30   ("citizen science") observations for catchment modelling and characterisation, J. Hydrol., 548, 801–817, doi:10.1016/j.jhydrol.2017.03.019, 2017.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D. and Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environ. Model. Softw., 90, 201–216, doi:10.1016/j.envsoft.2017.01.006, 2017.

28

Uhrner, U., Grosso, G., Romain, A. C., Hutsemekers, V., Delva, J., Kunz, W., De Groof, A., Arnaud, Y., Valoggia, P., Johannsen, L., Stevenot, B. and Ledent, P.: Development of an environmental information system for odour using citizen and technology innovative sensors and advanced modeling, in CEUR Workshop Proceedings, vol. 1322, Aachen, Germany., 2013.

5 De Vos, L., Leijnse, H., Overeem, A. and Uijlenhoet, R.: The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam, Hydrol. Earth Syst. Sci., 21(2), 765–777, doi:10.5194/hess-21-765-2017, 2017.

Walker, D., Forsythe, N., Parkin, G. and Gowing, J.: Filling the observational void: Scientific value and quantitative validation of hydrometeorological data from a community-based monitoring programme, J. Hydrol., 538, 713–725, doi:10.1016/j.jhydrol.2016.04.062, 2016.

10 Yang, Y. Y. and Kang, S. C.: Crowd-based velocimetry for surface flows, Adv. Eng. Informatics, 32, 275–286, doi:10.1016/j.aei.2017.03.007, 2017.

Yu, D., Yin, J. and Liu, M.: Validating city-scale surface water flood modelling using crowd-sourced data, Environ. Res. Lett., 11(12), 124011, doi:10.1088/1748-9326/11/12/124011, 2016.

15