

REVIEW

Open Access



Citizen surveillance for environmental monitoring: combining the efforts of citizen science and crowdsourcing in a quantitative data framework

Marijke Welvaert^{1,2*} and Peter Caley^{1,2}

*Correspondence: Marijke.Welvaert@canberra.edu.au
² Plant Biosecurity
Cooperative Research Centre,
Canberra, Australia
Full list of author information
is available at the end of the
article

Abstract

Citizen science and crowdsourcing have been emerging as methods to collect data for surveillance and/or monitoring activities. They could be gathered under the overarching term *citizen surveillance*. The discipline, however, still struggles to be widely accepted in the scientific community, mainly because these activities are not embedded in a quantitative framework. This results in an ongoing discussion on how to analyze and make useful inference from these data. When considering the data collection process, we illustrate how citizen surveillance can be classified according to the nature of the underlying observation process measured in two dimensions—the degree of observer reporting intention and the control in observer detection effort. By classifying the observation process in these dimensions we distinguish between crowdsourcing, unstructured citizen science and structured citizen science. This classification helps the determine data processing and statistical treatment of these data for making inference. Using our framework, it is apparent that published studies are overwhelmingly associated with structured citizen science, and there are well developed statistical methods for the resulting data. In contrast, methods for making useful inference from purely crowd-sourced data remain under development, with the challenges of accounting for the unknown observation process considerable. Our quantitative framework for citizen surveillance calls for an integration of citizen science and crowdsourcing and provides a way forward to solve the statistical challenges inherent to citizen-sourced data.

Keywords: Citizen science, Crowdsourcing, General surveillance, Environmental statistics, Data generation process

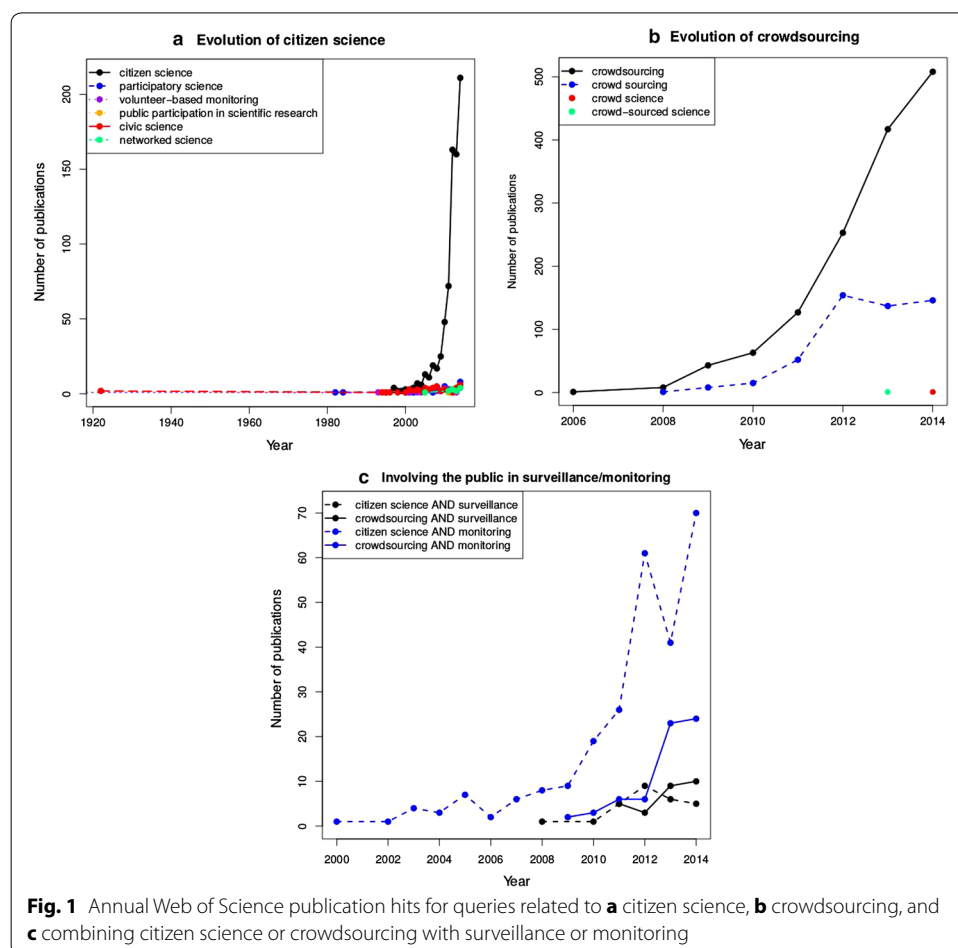
Background

Over the last decade, citizen science has become popular in many scientific areas. Involving citizens in scientific research has benefits for both parties. Scientists get access to a wealth of data that would not have been attainable by individual researchers, while citizens get the satisfaction of being involved in science and being part of a community project. In addition, crowdsourcing offers another data stream in which the general public collects or processes data that are used for scientific studies. However, citizen science and crowdsourcing seem to live parallel lives. While for citizen science the focus is

on citizens doing science (e.g. data collection, conducting experiments, etc.), in crowdsourcing the volunteers contribute resources (e.g. computing power, money, data, etc.) to science.

Figure 1 demonstrates that both disciplines have attracted increasing interest from the scientific community. Searching for publications relating to citizen science and crowdsourcing using the Web of Science database clearly shows that the number of publications has risen exponentially during the past 5 years. Interestingly, Fig. 1c indicates that citizen science and crowdsourcing has become increasingly popular for monitoring and surveillance activities. Furthermore, according to Roy et al. (2012), many citizen science projects are making use of crowdsourced data. Therefore, it seems natural to consider both disciplines as potential data sources for monitoring and surveillance activities.

In this paper, we propose a quantitative data-focused framework for characterizing citizen surveillance. We define citizen surveillance as *any type of activity conducted by volunteers, recruited or not, that results in monitoring or surveillance data*. Natural application domains for citizen surveillance are environmental surveillance, biodiversity monitoring and biosecurity surveillance, and both citizen science and crowdsourcing are logical subdisciplines of citizen surveillance in this context.



Other reviews (for example Bonney et al. 2009) have mainly focused on how to develop citizen science projects and the benefits in terms of science knowledge and scientific literacy for the communities involved. Although these frameworks include data analysis as an important aspect of the citizen science project, data validation and analysis is still a little studied part of the possibilities of citizen science (Butt et al. 2013). Citizen science, and in this paper we will extend this to citizen surveillance, still struggles to be accepted as a valid scientific method (Bonney et al. 2014) and is often associated with some skepticism with respect to the quality of the data generated. Therefore, not surprisingly, a lot of effort has gone in assessing whether volunteer collected data are of similar quality as data collected by professionals or experts. Recently, quite a number of studies have revealed a strong agreement between volunteer-collected data and empirical or expert-collected data (e.g. Davies et al. 2012; Hoyer et al. 2012; Gollan et al. 2012). However, Snall et al. (2011) warn to be cautious when using citizen science data because these data should be handled with great care. When using citizen science data for environmental surveillance purposes, one needs to be aware of and account for the limitations of those data. For example, Gardiner et al. (2012) demonstrated an overestimation of species richness and diversity values in lady beetles based on citizen science data compared to verified data. This was attributed to under-reporting of common species combined with over-reporting of rare native species (including false positives). Accounting for such findings can clearly improve the analysis of citizen surveillance data.

By focusing on the quantitative aspects of citizen science and crowdsourcing in relation to the reporting and detection process, we hope to provide a way forward to a unified search for statistical solution for citizen-sourced data. We discuss the trade-off between data quality and data quantity in citizen surveillance data, which datasets can be expected from the different types of citizen surveillance and we highlight the analytic challenges which go hand in hand with citizen surveillance data. We hope that by focusing on the quantitative data aspects of citizen surveillance, users will be guided in how to choose an appropriate type of citizen surveillance for their environmental research questions or monitoring goals.

The remainder of this paper is organized as follows: “[The emergence of citizen science](#)” section provides a brief overview of the history of citizen-sourced data and presents the results of a extensive literature analysis demonstrating the lack of studies investigating statistical methods for analysing citizen-sourced data. “[Citizen surveillance data: a trade-off between data quality and data quantity](#)” section goes into the details of the observation process of citizen surveillance and proposes a new framework that differentiates between different types of citizen surveillance based on common elements in the observation and reporting process. This framework is used to assess the statistical challenges for citizen surveillance. Finally, in “[Conclusion](#)” section we provide some recommendations for maximizing the information retrieved from citizen surveillance data and urge for an integrated approach in developing statistical methods for citizen-sourced data.

The emergence of citizen science

Citizen science may be a rather new term, but in fact its practice dates back several centuries. Before the twentieth century, scientific work was most often conducted by *gentleman scientists* such as Isaac Newton, Abraham Franklin and Charles Darwin. It is only

since the mid-twentieth century that science has become almost completely professionalized with scientists employed by universities and research organizations. While this is still mostly the case, over recent decades, several projects have emerged that have successfully engaged volunteers without specific scientific training into posing research questions, and collecting and analyzing data (e.g. Trumbull et al. 2000).

Citizen science has been defined as “the systematic collection and analysis of data; development of technology; testing of natural phenomena; and the dissemination of these activities by researchers on a primarily avocational basis” (OpenScientist 2011). In short, citizen science is commonly known as “public participation in scientific research” (e.g. Hand 2010) and is sometimes referred to as *crowd science*, *crowd-sourced science*, *civic science*, or *networked science*. The term “citizen science” has become commonly used by both scientists (see Fig. 1) and the community from 2009 onward (see Bonney et al. 2009 for a definition of the fields). It is believed that citizen science is a cost-effective means to provide solutions to research questions, because, in general the data is cheaper than that gathered by trained professionals, while similar in quality and quantity after intelligent analyzing and processing (Barbier et al. 2012). In particular, citizen science can be helpful when researchers attempt to answer questions on a large spatial or temporal scale (Bonney et al. 2009) that cannot be sampled otherwise due to limited funding.

Citizen science projects are increasing in popularity (see Theobald et al. 2015; See et al. 2016; Fig. 1). In order to assess whether this is also the case in the scientific literature, we conducted a literature search using the Web Of Science database, indexing scientific publications going back as far as 1864, and the query “citizen science”. The search resulted in 379 publication hits (results as of December 5, 2013). After inspection, 33 of these (8.7%) were considered irrelevant because they were not related to citizen science as we defined above, another 5 publications (1.3%) were disregarded because they had no assigned author and therefore their geographical source could not be traced. After removing these results, 342 publications remained that all discussed citizen science, although in many different forms.

A content-based analysis of the database revealed that the most common publication type related to citizen science were applications (i.e. 45.6%). This included reporting on the analysis and interpretation of citizen science data (e.g. Wilson et al. 2013); the development of educational projects using citizen science was discussed (e.g. Zoellick et al. 2012); the implementation of a new citizen science project was described (e.g. Arvanitidis et al. 2011); or reporting the discovery of new species as a result of citizen science (e.g. Collins et al. 2011; Cohen et al. 2011).

Only 5.8% of the publications discuss specific analysis challenges and methods to analyze citizen science data, while another 12.2% present studies that tried to validate the citizen science data by, for example, comparing the quality of the data gathered by volunteers on the one hand and by expert professionals on the other. This illustrates that data validation and analysis is indeed understudied. Interestingly, more than half of the citizen science publications found here were published in an environmental related field, namely *Environmental Sciences & Ecology*, *Biodiversity & Conservation*, *Zoology*, and *Entomology*. These fields carry a tradition of large monitoring surveys and projects aimed to assist the conservation of species and it is obvious that citizen science is already well embedded. However, we will demonstrate that not all types of citizen surveillance will be the optimal choice to gather data for conservation monitoring.

Citizen surveillance data: a trade-off between data quality and data quantity

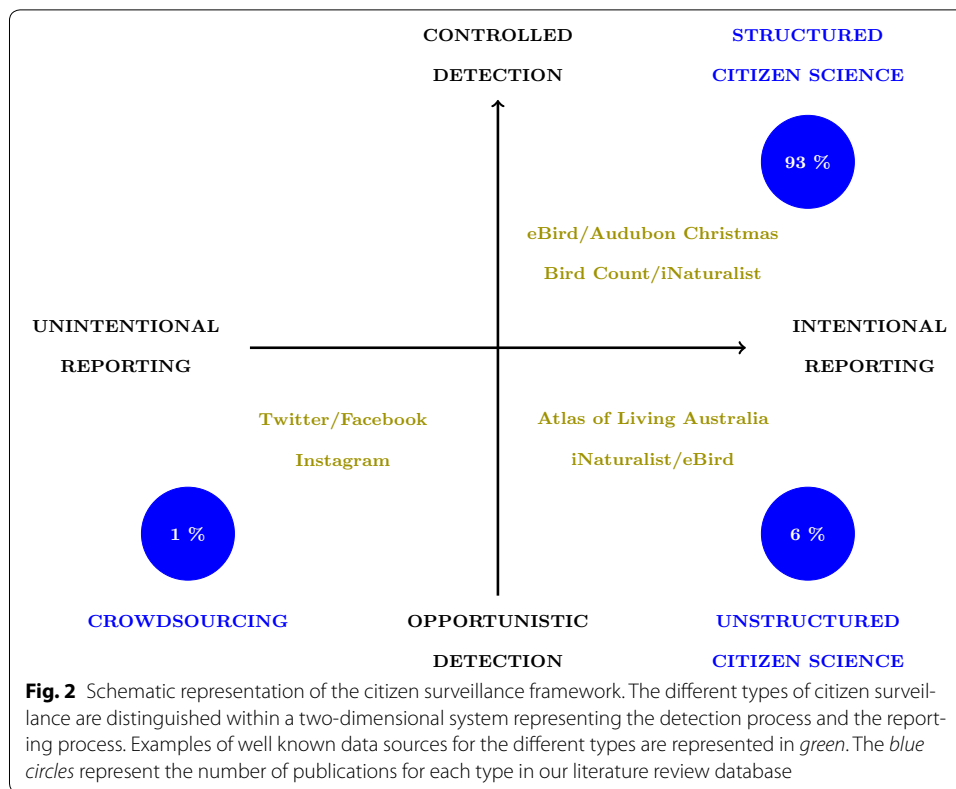
Citizen science projects can be classified into different types depending on several factors. For example, Roy et al. (2012) derived different clusters of citizen science projects based on how they scored on two dimensions. The first dimension accounted for the *degree of mass participation* (i.e. mass participation projects versus local monitoring projects), while the second dimension was interpreted as the *degree of investment* (i.e. investment of project managers, investment of background information and supporting material, and the time-investment of the participants). In this paper, we follow a similar thinking process in terms of classifying different types of citizen surveillance into a two-dimensional system. However, we specifically focus on the quantitative characteristics of the citizen-sourced data, because they are more relevant with respect to choosing an analysis method.

Structuring citizen surveillance based on observation and reporting

Citizen-sourced surveillance/monitoring data, in particular from the natural environment, will depend jointly on the processes of species detection/observation and reporting. Factors relating to detection and reporting can be used to characterize the data observation process, with consequences for data preprocessing, appropriate statistical analysis and ultimately the amount of useful information that can be retrieved from the data. To our knowledge, thus far this distinction between the two processes has not been made. Here we argue, differentiating between these processes is crucial for making inference from these data.

Figure 2 illustrates how three broad types of citizen surveillance can be situated in the quadrants of a detection–reporting axis system. On the detection axis, a distinction is made between controlled and opportunistic detections. Controlled detection implies that the time and place of the detection is part of a sampling design, while opportunistic detections are side-products from other day-to-day activities (e.g. day hiking) and therefore there is less control on where and when these detections occur. On the reporting axis, we differentiate between unintentional and intentional reporting. This dimension mainly refers to the platform to which the citizen detections are being reported. Intentional reporting is typically made on a dedicated citizen science platform (e.g. website), while an unintentional reporting is made on a medium that is not directly related to citizen science (e.g. social media).

Within this two-dimension system, we note that three types of citizen science can be differentiated: (1) Structured citizen science projects are projects that are typically organized by a scientist (or using scientific principles) and in which strong control is exercised on the observation and reporting process. These projects typically offer some form of training to the volunteers, which are purposefully recruited, and provide questionnaires and checklists. In some cases even, an experimental design is adopted. (2) In unstructured citizen science, the reporting is still done on purpose, but the observation process is opportunistic. Observations are typically spontaneous and a result of trips planned or undertaken independently with respect to the species reported. There is typically no information available on the time spent in the field, nor information on other species recorded on the same visit to a given site. This type of citizen science typically needs to be enabled by a web-platform on which volunteers can report their detections.



(3) A last category is crowdsourcing. This term captures information that is observed in an opportunistic way and reported unknowingly. For example, Tweets reporting something, sometimes supported by photographic material (see Fig. 3). Crowdsourcing is often not regarded as citizen science because the volunteers are not necessarily involved in the scientific research (Miller-Rushing et al. 2012). However, within the context of our framework, this data stream could be a natural part of citizen surveillance, keeping in mind that it will have its own statistical challenges and does not necessarily provide the same benefits to the community (e.g. scientific literacy).

The literature analysis described in “[The emergence of citizen science](#)” section showed that the scientific literature is dominated by the first type, structured citizen science (Fig. 1). The close involvement of scientists in these types of projects is the most probable explanation for this over-representation. As mentioned before, crowdsourcing is often not regarded as citizen science, and therefore, is most likely under-represented in the database presented in “[The emergence of citizen science](#)” section.

Figure 3 also illustrates one of the data quality issues inherent to citizen science (i.e. misidentification), whereby a tweet contains both “plague” and “locust”—seemingly referring to an outbreaking species that may cause extensive crop damage. The subject of the tweet (as revealed by a photo) is a harmless species. One of the benefits of the framework shown in Fig. 2 is that we can clearly identify a trade-off along the diagonal axis between data quality and data quantity. Table 1 summarizes some of the advantages and disadvantages for the different types of citizen science.



Fig. 3 Example of crowdsourcing from Twitter. The photographic material makes it possible to correct the misidentification in the tweet. In this case it is a cicada, unrelated to locusts

Table 1 Evaluation of some typical characteristics of citizen surveillance data

	Type of citizen science		
	Structured	Unstructured	Crowdsourcing
Control over observation process	++	-	--
Data quantity	-/+	-/+	+
Signal-to-noise ratio	++	-/+	--
Temporal coverage	-/+	+	+
Spatial coverage	-/+	+	+

Choosing a type of citizen surveillance will typically be a trade-off between data quality and data quantity
 Symbol legend: ++ Good; + Moderate; -/+ Variable; - Low; -- Bad

Data quality issues have been heavily investigated in the assessment of crowd-sourced data for mapping purposes (Antonioni and Skopeliti 2015; See et al. 2013; Jokar Arsanjani et al. 2015). While a lot of the issues related to this field are inherently embedded in citizen surveillance for environmental monitoring, we do believe that the data quality problems go even further due to the differences in the data generating process. For example, disaster detection like earthquakes based on georeferenced information relies on the fact that the public will be reporting about the disaster using media that are georeferenced, e.g. Twitter. This works quite well because people immediately recognise the disaster (the observation) and it affects them in such manner that they feel inclined to Tweet about it (the reporting). However, in environmental monitoring, both of these processes are challenged. The observation will be subject to bias and false positives because some species, especially rare/exotic species and/or small species, are hard to detect and it can be very difficult to differentiate between the species to the untrained eye. In addition, the mere observation of a species will not always trigger a reporting response from the observer, unless they have a vested interest in reporting that species or are highly motivated to do so.

Because of its underlying scientific principles, structured citizen science has the major advantage that there is relative control over the observation process. This might be the reason why this type is so present in the scientific literature. On the other hand, unstructured citizen science and crowdsourcing are subject to many uncontrolled processes, most of them unknown. The challenge in analyzing the data from these types of citizen science will be accounting for unknown observation processes and coming up with ways to incorporate the increased uncertainty. In addition, the differences in data type will trigger different analysis strategies. While structured citizen science most likely will provide counts or presence-absence data, unstructured citizen science and crowdsourcing will mainly provide presence-only data. This further complicates the inference that can be made from these data and is amongst statisticians a topic of ongoing debate (Royle et al. 2012; Hastie and Fithian 2013).

Another important aspect is data quantity. Typically, the more data that are available, the more likely it is that uncertainty in the data can be accounted for (i.e. partition signal from noise). However, in citizen science data this is not necessarily true. Data quantity is linked to the signal-to-noise ratio. For example, crowdsourcing provides huge datasets but the signal of interest in these data is only a small fraction and can be totally buried in the noise. Analysis of these data will require a lot of preprocessing and will be extremely hard to validate. There might be less data from structured citizen science, but the signal-to-noise ratio will be a lot higher, meaning that the information in the data is more easily accessible.

Temporal and spatial coverage are important characteristics for environmental surveillance in particular. Rather than having data for one location in a specific time period, environmental monitoring and biodiversity monitoring can benefit immensely from data that cover a lot of space and that are observed frequently. Because structured citizen science is mostly, but not always, an organized activity that brings together a lot of people at the same time at the same place, it is expected to have less spatial and temporal coverage compared to unstructured citizen science and crowdsourcing. The joint effort in these types of citizen surveillance will most likely cover more ground and time, but

are also prone to spatial and temporal biases due to opportunistic sampling (see “Challenges” section).

Challenges

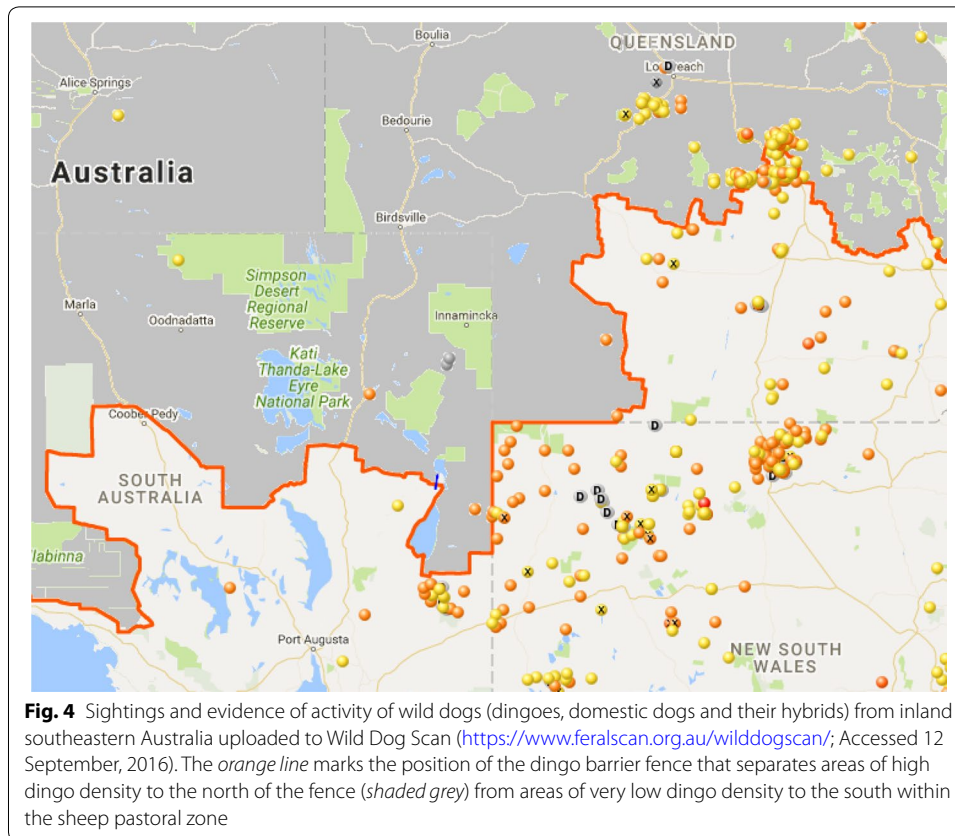
Sampling design and analysis

We have already demonstrated that citizen surveillance can vary from highly structured to completely unstructured. Structured surveillance can be found in avian monitoring, which has a long tradition of citizen science and, not surprisingly, has also been the origin of improved sampling designs for large monitoring studies (e.g. Heezik and Seddon 2012; MacLeod et al. 2012). In contrast, for unstructured citizen surveillance, a statistically sound sampling design is normally lacking and most data are collected based on convenience sampling (Boakes et al. 2010). This means that sampling locations are often near roads or other easily accessible locations, and sampling is most likely on weekends and during daylight hours. In addition, both inter- and intra-observer sampling effort is highly variable and will result in heterogeneity of detection probabilities. The previously mentioned analysis methods, while potentially useful for structured citizen science projects, do not apply to the unstructured part of citizen surveillance. In these cases, induced heterogeneity will have to be accounted for in the analysis of the data.

Developing validated methods for correcting sampling bias for such data is an active area of research in the species distribution modelling field (e.g. Fithian et al. 2015), including the use of high-quality data from planned surveys to leverage information from presence-only data (Dorazio 2014). Other bias correction mechanisms have a logical basis (e.g. using the density of roads as a covariate in the case of citizen-sighted koalas), though are applied without knowing how well the correction is performing (see Sequeira et al. 2014).

In some situations, however, no amount of correction can remedy observer reporting (c.f. sampling) bias inherent in unstructured citizen science. For example, Wild Dog Scan (<https://www.feralscan.org.au/wilddogscan/>) enables landholders in Australia to report sightings of dingoes (*Canis lupus dingo*) and their hybrids. Such sightings are of particular interest to the sheep pastoral industry, as dingoes may cause severe economic losses. The resulting dingo sightings map is dominated by sightings within the sheep pastoral zone, despite dingoes being rare in this region as a result of intensive control, in comparison with areas to the north outside the dingo barrier fence, where dingoes are comparatively abundant (Fig. 4). Recognising that these data are unstructured should alert anyone using them to research the observation process before undertaking any analysis—failure to do so in this case would result in the incorrect inference that dingo abundance is highest inside the dingo fence. In this situation the sightings probability being somewhat perversely inversely related to abundance is driven by sightings being of much greater economic importance to graziers inside the dingo fence than out. This is not to say that the reporting App and the data it generates aren't of use to users within the dingo fence.

In the case of crowdsourcing involving unintentional reporting in conjunction with opportunistic detection, there has been little development of validated methodology for retrieving stable surveillance information these type of data.



Observer effects

Because citizen surveillance often relies so heavily on inexperienced volunteers, observer effects due to heterogeneity in skills, age, background, etc. can be non-trivial and should not be ignored. Not only do they affect species detectability, they also will influence the adopted sampling design and ultimately the error probability. In the case of citizen science, probably most forms of bias in the data can be traced back to the observers, either through their observing skills, or their choice of sampling locations and time. However, reducing these effects through, for example, standardization may not be sufficient to eliminate the heterogeneity as there are other variables that will influence species detectability (Yoccoz et al. 2001). Sauer et al. (1994) illustrate that long-term trends in observer quality can substantially bias estimates of population trend.

In general, it is known that novices and non-experts are more likely to over-report rare species that are easy to detect (Sullivan et al. 2009). On the other hand, self-assessed experts may suffer more from overconfidence (Larrick et al. 2007), making them more prone to false-positive errors (Moore and Healy 2008).

However, these observer effects can be either controlled through training with structured citizen science, or accounted for provided information is available on, for example, survey effort, experience, etc.

A way forward

Merely pointing out the challenges arising from using citizen surveillance data is obviously not enough. When making inference from those data, sampling bias and observer

effects will have to be accounted for. In our proposed framework (Fig. 2) we made a clear distinction between the observation and reporting process. Most of the challenges discussed above were looked at in the light of the observation process simply because little is known on how the reporting process is affected. We can speculate that there will be similarities in the types of bias arising in both processes, but there might also be differences. From a statistical perspective, we would like to account for as much bias as possible to ensure valid inference and it would be particularly useful if we could even distinguish between the different sources of potential bias.

Given the rise of web-platforms enabling unstructured citizen science, we believe that effort should be targeted to developing statistical methods for analysing these data. Because of their less controlled observation and reporting process, compared to structured citizen science, sampling and reporting bias will play a major role (as illustrated by the Wild Dog example discussed earlier) and will need to be accounted for. However, this will only be possible, if information around sampling effort or knowledge around what drives the reporting is available. Only then statisticians can try to develop methods that incorporate such information to improve inference from unstructured citizen science.

Conclusion

In this paper, we advocate an integrated view on citizen surveillance that combines the existing efforts of citizen science and crowdsourcing. Based on the characteristics of the observation process, both disciplines can be naturally combined in one quantitative framework that allows for an integrated approach to solve the statistical challenges inherent to citizen-sourced data.

See et al. (2016) also present a two-dimensional framework to distinguish types of crowdsourced geographic information based on who collects the data (government agencies versus general public) and the contribution characteristic of the data (active versus passive). Their framework no doubt contributes to structuring the thinking around crowdsourced geographic information, however, it does not cover off on the specifics of the data collection process as practiced in citizen surveillance for environmental monitoring. In particular, no distinction is being made between the actual observation process of a species versus the reporting process of a species. In this review, we have highlighted the importance of this distinction as knowledge of those processes, or the absence of such knowledge, will have to inform the inference methods applied to these data.

Our framework distinguished between three major types of citizen surveillance:

1. *Structured citizen science*

Currently the most published form of citizen surveillance and also the type that aligns most with the current understanding of citizen science. The use of scientific principles, the potential for training of participants and the overall high level of involvement of the volunteers creates a data stream that has the potential to provide high-quality data that are in a format (e.g. counts) that can be analyzed with readily available tools. However, because of its reliance on funding, the temporal and spatial scope can be limited.

2. *Unstructured citizen science*

Enabled by online platforms like iNaturalist (<http://www.inaturalist.org>), unstructured citizen science provides access to an audience that is intrinsically motivated to report observations. These datasets are more prone to sampling and observer bias, but additional data fields in the recording process can allow for the collection of meta-data that can be incorporated in the analysis process. However, statistical techniques to make useful inference of these data are not readily available at this point in time and more effort will have to be targeted to the validation of these techniques if we want to rely more on unstructured citizen science for environmental monitoring.

3. *Crowdsourcing*

Crowdsourcing offers an untapped resource of surveillance data with contributions to, for example, social media increasing constantly. Successful examples include, but are not limited to, the tracking of human disease (Culotta 2010) and resource allocation after disasters (Zhu et al. 2011). However, analyzing these data for environmental surveillance purposes will be a huge challenge because the information of interest can be buried amongst numerous irrelevant posts. Methods to reliably retrieve this information are still being developed. In addition, both unstructured citizen science and crowdsourcing will most commonly provide presence-only data and the analysis of these data is still an active topic of discussion amongst statisticians (Royle et al. 2012; Hastie and Fithian 2013).

This framework that classifies citizen surveillance data is a way forward to achieve the full potential of those data. Indeed, such classification models are a necessity because they point to the importance of describing the sampling process and the observation process (Bird et al. 2014). This information is crucial during the analysis process as it can be used to improve statistical models for citizen surveillance data.

Initiatives like eBird (<http://www.ebird.org>) make it possible to collect environmental data world-wide. While current efforts go mainly in engaging volunteers, data storage and data sharing, there is still a major role for statisticians to guide citizen surveillance projects with better ways to collect their data and to improve statistical models for citizen surveillance data.

Authors' contributions

MW collected the data, designed the study and wrote the manuscript. PC participated in the design of the study and helped draft the manuscript. Both authors read and approved the final manuscript.

Author details

¹ Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia. ² Plant Biosecurity Cooperative Research Centre, Canberra, Australia.

Acknowledgements

This work was supported by the Australian Government's Cooperative Research Centres Program (Project CRC 1029, Plant Biosecurity CRC). Many thanks to Sorada Tapsuwan, Ross Sparks, Rieks Van Klinken and an anonymous reviewer for their insightful comments on earlier drafts of this paper. We would also like to thank Roger Pech, Peter West and Jennifer Cruz-Bernall for their input on the topic of citizen surveillance.

Competing interests

Both authors declare that they have no competing interests.

Received: 24 November 2015 Accepted: 19 October 2016

Published online: 28 October 2016

References

- Antoniou V, Skopeliti A (2015) Measures and indicators of VGI quality: an overview. In: ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, La Grande Motte, France, vol II-3/W5, pp 345–351. doi:[10.5194/isprannals-II-3-W5-345-2015](https://doi.org/10.5194/isprannals-II-3-W5-345-2015)
- Arsanjani JJ, Mooney P, Zipf A, Schauss A (2015) Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In: Arsanjani JJ, Zipf A, Mooney P, Helbich M (eds) OpenStreetMap in GIScience. Lecture notes in geoinformation and cartography. Springer International Publishing, Cham, Switzerland, pp 37–58
- Arvanitidis C, Faulwetter S, Chatzigeorgiou G, Penev L, Banki O, Dailianis T, Pafilis E, Kouratoras M, Chatzinikolaou E, Fanini L, Vasileiadou A, Pavludi C, Vavilis P, Koulouri P, Dounas C (2011) Engaging the broader community in biodiversity research: the concept of the comber pilot project for divers in vibrant. *ZooKeys* 1(150):211–229. doi:[10.3897/zookeys.150.2149](https://doi.org/10.3897/zookeys.150.2149)
- Barbier G, Zafarani R, Gao HJ, Fung G, Liu H (2012) Maximizing benefits from crowdsourced data. *Comput Math Organ Theory* 18(3):257–279. doi:[10.1007/s10588-012-9121-2](https://doi.org/10.1007/s10588-012-9121-2)
- Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, Edgar GJ, Stuart-Smith RD, Wotherspoon S, Krkosek M, Stuart-Smith JF, Pecl GT, Barrett N, Frusher S (2014) Statistical solutions for error and bias in global citizen science datasets. *Biol Conserv* 173:144–154
- Boakes EH, McGowan PJK, Fuller RA, Ding CQ, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol* 8(6):11. doi:[10.1371/journal.pbio.1000385](https://doi.org/10.1371/journal.pbio.1000385)
- Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, Shirk J (2009) Citizen science: a developing tool for expanding science knowledge and scientific literacy. *Bioscience* 59(11):977–984. doi:[10.1525/bio.2009.59.11.9](https://doi.org/10.1525/bio.2009.59.11.9)
- Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK (2014) Next steps for citizen science. *Science* 343:1436–1437
- Butt N, Slade E, Thompson J, Malhi Y, Riutta T (2013) Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. *Ecol Appl* 23(4):936–943
- Cohen CS, McCann L, Davis T, Shaw L, Ruiz G (2011) Discovery and significance of the colonial tunicate *Didemnum Vexillum* in Alaska. *Aquat Invasions* 6(3):263–271. doi:[10.3391/ai.2011.6.3.03](https://doi.org/10.3391/ai.2011.6.3.03)
- Collins AG, Bentlage B, Gillan WB, Lynn TH, Morandini AC, Marques AC (2011) Naming the Bonaire banded box jelly, *Tamoya ohboya*, n. sp. (Cnidaria: Cubozoa: Carybdeida: Tamoyidae). *Zootaxa* 2753:53–68
- Culotta A (2010) Towards detecting influenza epidemics by analyzing twitter messages. In: Proceedings of the first workshop on social media analytics, ACM, New York, NY, USA, SOMA '10, pp 115–122. doi:[10.1145/1964858.1964874](https://doi.org/10.1145/1964858.1964874)
- Davies TK, Stevens G, Meekan MG, Struve J, Rowcliffe JM (2012) Can citizen science monitor whale-shark aggregations? Investigating bias in mark-recapture modelling using identification photographs sourced from the public. *Wildl Res* 39(8):696–704. doi:[10.1071/WR12092](https://doi.org/10.1071/WR12092)
- Dorazio RM (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Glob Ecol Biogeogr* 23(12):1472–1484. doi:[10.1111/geb.12216](https://doi.org/10.1111/geb.12216)
- Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods Ecol Evol* 6:424–438. doi:[10.1111/2041-210x.12242](https://doi.org/10.1111/2041-210x.12242)
- Gardiner MM, Allee LL, Brown PMJ, Losey JE, Roy HE, Smyth RR (2012) Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Front Ecol Environ* 10(9):471–476. doi:[10.1890/110185](https://doi.org/10.1890/110185)
- Gollan J, de Bruyn LL, Reid N, Wilkie L (2012) Can volunteers collect data that are comparable to professional scientists? A study of variables used in monitoring the outcomes of ecosystem rehabilitation. *Environ Manag* 50(5):969–978. doi:[10.1007/s00267-012-9924-4](https://doi.org/10.1007/s00267-012-9924-4)
- Hand E (2010) People power. *Nature* 466(7307):685–687. doi:[10.1038/466685a](https://doi.org/10.1038/466685a)
- Hastie T, Fithian W (2013) Inference from presence-only data: the ongoing controversy. *Ecography* 36(8):864–867. doi:[10.1111/j.1600-0587.2013.00321.x](https://doi.org/10.1111/j.1600-0587.2013.00321.x)
- Hoyer MV, Wellendorf N, Frydenborg R, Bartlett D, Canfield DE Jr (2012) A comparison between professionally (Florida Department of Environmental Protection) and volunteer (Florida LAKEWATCH) collected trophic state chemistry data in Florida. *Lake Reserv Manag* 28(4):277–281. doi:[10.1080/07438141.2012.736016](https://doi.org/10.1080/07438141.2012.736016)
- Larrick RP, Burson KA, Soll JB (2007) Social comparison and confidence: when thinking you're better than average predicts overconfidence (and when it does not). *Organ Behav Hum Decis Process* 102(1):76–94. doi:[10.1016/j.obhdp.2006.10.002](https://doi.org/10.1016/j.obhdp.2006.10.002)
- MacLeod CJ, Tinkler G, Gormley AM, Spurr EB (2012) Measuring occupancy for an iconic bird species in urban parks. *N Z J Ecol* 36(3, SI):398–407
- Miller-Rushing A, Primack R, Bonney R (2012) The history of public participation in ecological research. *Front Ecol Environ* 10(6):285–290. doi:[10.1890/110278](https://doi.org/10.1890/110278)
- Moore DA, Healy PJ (2008) The trouble with overconfidence. *Psychol Rev* 115(2):502–517. doi:[10.1037/0033-295X.115.2.502](https://doi.org/10.1037/0033-295X.115.2.502)
- OpenScientist (2011) Finalizing a definition of "Citizen Science" and "Citizen Scientist". <http://www.openscientist.org/2011/09/finalizing-definition-of-citizen.html>. Accessed 17 Dec 2013
- Roy H, Pocock M, Preston C, Roy D, Savage J, Tweddle J, Robinson L (2012) Understanding citizen science and environmental monitoring. Final report on behalf of UK environmental observation framework, Nov 1–67
- Royle JA, Chandler RB, Yackulic C, Nichols JD (2012) Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods Ecol Evol* 3(3):545–554. doi:[10.1111/j.2041-210X.2011.00182.x](https://doi.org/10.1111/j.2041-210X.2011.00182.x)
- Sauer JR, Peterjohn BG, Link WA (1994) Observer differences in the North American breeding bird survey. *Auk* 111(1):50–62
- See L, Comber A, Salk C, Fritz S, van der Velde M, Perger C, Schill C, McCallum I, Kraxner F, Obersteiner M (2013) Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLoS ONE* 8(7):11. doi:[10.1371/journal.pone.0069958](https://doi.org/10.1371/journal.pone.0069958)

- See L, Mooney P, Foody G, Bastin L, Comber A, Estima J, Fritz S, Kerle N, Jiang B, Laakso M, Liu HY, Milčinski G, Nikšić M, Painho M, Pödör A, Olteanu-Raimond AM, Rutsinger M, (2016) Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *Int J Geoinf*. doi:[10.3390/ijgi5050055](https://doi.org/10.3390/ijgi5050055)
- Sequeira AM, Roetman PE, Daniels CB, Baker AK, Bradshaw CJ (2014) Distribution models for koalas in South Australia using citizen science-collected data. *Ecol Evol* 4(11):2103–2114
- Snall T, Kindvall O, Nilsson J, Part T (2011) Evaluating citizen-based presence data for bird monitoring. *Biol Conserv* 144(2):804–810. doi:[10.1016/j.biocon.2010.11.010](https://doi.org/10.1016/j.biocon.2010.11.010)
- Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S (2009) eBird: a citizen-based bird observation network in the biological sciences. *Biol Conserv* 142(10):2282–2292. doi:[10.1016/j.biocon.2009.05.006](https://doi.org/10.1016/j.biocon.2009.05.006)
- Theobald EJ, Ettinger AK, Burgess HK, DeBey LB, Schmidt NR, Froehlich HE, Wagner C, HilleRisLambers J, Tewksbury J, Harsch MA, Parrish JK (2015) Global change and local solutions: tapping the unrealized potential of citizen science for biodiversity research. *Biol Conserv* 181:236–244. doi:[10.1016/j.biocon.2014.10.021](https://doi.org/10.1016/j.biocon.2014.10.021)
- Trumbull DJ, Bonney R, Bascom D, Cabral A (2000) Thinking scientifically during participation in a citizen-science project. *Sci Educ* 84(2):265–275. doi:[10.1002/\(SICI\)1098-237X\(200003\)84:2<265::AID-SCE7>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<265::AID-SCE7>3.0.CO;2-5)
- van Heezik Y, Seddon PJ (2012) Accounting for detectability when estimating avian abundance in an urban area. *N Z J Ecol* 36(3):391–397
- Wilson S, Anderson EM, Wilson ASG, Bertram DF, Arcese P (2013) Citizen science reveals an extensive shift in the winter distribution of migratory western grebes. *PLoS ONE*. doi:[10.1371/journal.pone.0065408](https://doi.org/10.1371/journal.pone.0065408)
- Yoccoz N, Nichols J, Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends Ecol Evol* 16(8):446–453. doi:[10.1016/S0169-5347\(01\)02205-4](https://doi.org/10.1016/S0169-5347(01)02205-4)
- Zhu J, Xiong F, Piao D, Liu Y, Zhang Y (2011) Statistically modeling the effectiveness of disaster information in social media. In: *Global humanitarian technology conference (GHTC), 2011 IEEE*, pp 431–436. doi:[10.1109/GHTC.2011.48](https://doi.org/10.1109/GHTC.2011.48)
- Zoellick B, Nelson SJ, Schauffler M (2012) Participatory science and education: bringing both views into focus. *Front Ecol Environ* 10(6, 5):310–313. doi:[10.1890/110277](https://doi.org/10.1890/110277)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
