



Crowd surveillance: estimating citizen science reporting probabilities for insects of biosecurity concern

Implications for plant biosecurity surveillance

Peter Caley¹ · Marijke Welvaert¹ · Simon C. Barry¹

Received: 7 October 2018 / Revised: 24 March 2019 / Accepted: 6 April 2019 / Published online: 11 June 2019
© The Author(s) 2019

Abstract

Data streams arising from citizen reporting activities continue to grow, yet the information content within these streams remains unclear, and methods for addressing the inherent reporting biases little developed. Here, we quantify the major influence of physical insect features (colour, size, morphology, pattern) on the propensity of citizens to upload photographic sightings to online portals, and hence to contribute to biosecurity surveillance. After correcting for species availability, we show that physical features and pestiness are major predictors of reporting probability. The more distinctive the visual features, the higher the reporting probabilities—potentially providing useful surveillance should the species be an unwanted exotic. Conversely, the reporting probability for many small, nondescript high priority pest species is unlikely to be sufficient to contribute meaningfully to biosecurity surveillance, unless they are causing major harm. The lack of citizen reporting of recent incursions of small, nondescript exotic pests supports the model. By examining the types of insects of concern, industries or environmental managers can assess to what extent they can rely on citizen reporting for their surveillance needs. The citrus industry, for example, probably cannot rely on passive unstructured citizen data streams for surveillance of the Asian citrus psyllid (*Diaphorina citri*). In contrast, the forestry industry may consider that citizen detection and reporting of species of the large and colourful insects such as pine sawyers (*Monochamus spp.*) may be sufficient for their needs. Incorporating citizen surveillance into the general surveillance framework is an area for further research.

Keywords Citizen surveillance · General surveillance · Citizen science · Pest · Biosecurity · Hemiptera · Coleoptera

Key message

- The citizen reporting probabilities of insects are dramatically influenced by physical features such as size, colour, pattern and morphology.
- We demonstrate how it is possible to correct for the inherent bias in citizen reporting by modelling the effect

of physical features and species distribution and abundance using a case–control design. This enables predictions of the surveillance sensitivity of citizen reporting of exotic species of biosecurity interest.

- For highly featured and/or large exotic insect pests, citizen reporting may provide adequate surveillance for plant health needs.
- Exotic insect species for which citizen reporting is unlikely to be effective can be predicted in advance.

Communicated by S. Macfadyen.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10340-019-01115-7>) contains supplementary material, which is available to authorized users.

✉ Peter Caley
peter.caley@csiro.au

¹ CSIRO Data61, GPO Box 1700, Canberra, ACT, Australia

Introduction

Biosecurity surveillance aims to protect the natural environment, plants and animals, as well as agri- and horticulture from harm caused by pests and diseases (Froud et al. 2008). The biosecurity threat arising from the invasion of exotic

insect pests is highly diffuse in that the number of target species is very large and the potential points of entry are numerous. This presents particular logistical challenges for implementing effective surveillance—it is impossible to deploy targeted traditional surveillance (e.g. species specific traps, trained inspectors, etc.) for all threats in all locations. Proposed alternatives to such traditional surveillance include increased use of sensors, robots and citizen science. Here, we focus on the latter option. Citizens can potentially contribute to biosecurity surveillance in many ways, ranging from inadvertent references to invasive organisms on social media platforms (e.g. Twitter), to deliberate through unstructured (spatially and temporally opportunistic) reporting of species via dedicated online portals (e.g. iNaturalist, <https://www.inaturalist.org/>), to deliberate structured (designed) surveys (Welvaert and Caley 2016). The potential surveillance power of the general public is evident from a New Zealand study, where nearly half of all new exotic species detections over a 3-year period were from members of the general public (Froud et al. 2008). In a similar vein, Thomas et al. (2017) recorded that 95% of non-indigenous invertebrate species new to Barrow Island were detected by members of the local community. Such surveillance contributes to what is termed “general surveillance” (Hammond et al. 2016a).

Detecting environmental biosecurity events from human social media communications in a timely manner faces some particular challenges, some technical (Daume 2016) and others largely arising from uncertainty and bias relating to the observation process (Welvaert and Caley 2016). In comparison with self-reported syndromic human health surveillance, the spatial scale and number of events to be detected is small initially (at the time when detection is most critical), and the direct impact on individuals typically minimal. For example, the combined effects of citrus greening disease (Huanglongbing—currently causing massive economic loss to the citrus industry in the Americas), vectored by the Asian citrus psyllid (*Diaphorina citri*) (Grafton-Cardwell et al. 2013) are neither immediate nor direct on human individuals *per se*, until the pathogen has spread significantly and affected trees are showing visible symptoms. Furthermore, the impacts and/or symptoms of exotic pests and diseases may be unknown or hard to detect or difficult to distinguish from endemic pests, resulting in varying levels of detectability (Jarrad et al. 2011). The detection of small-scale biosecurity events through social media also requires that the taxonomy of organism is widely known but also unique; otherwise, the signal-to-noise ratio is too low for reliable signal retrieval (Welvaert et al. 2017). Hence, reliably detecting the arrival of exotic insects within the social media data stream is likely to be highly problematic.

Insect collecting is a worldwide contemporary and historical hobby/passion of many members of the public, with the major recent change being the move to photography

in place of physical specimen collecting, and the ability to share these images online. In comparison with social media, the uploading of photographs onto citizen science data portals is much more deliberate and has a taxonomic underpinning. Dedicated online platforms now exist to store such observations, and to crowd-source their species identification. The number of citizen-sourced record uploads goes in the tens of millions. Note, however, these data sources generally do not contain biosecurity related species information. By definition they would not contain records for invasive alien species that have not yet entered a country. These growing datasets, however, can be used to inform us about the type of species that are typically reported by citizen scientists, and whether they are likely to include exotic pests and/or pathogen species should they arrive. Indeed, in Australia for example, a wide variety of sightings of insect species are uploaded to the Atlas of Living Australia (ALA, <https://www.ala.org.au/>) which acts as a repository for most citizen science platforms in Australia along with professionally collected museum specimens, etc. The number of citizen-sourced record uploads of insect species to the ALA already number in the 100,000s, involving 1000s of species. Although these numbers may seem impressive, at face value they provide little information on whether an emergency plant pest (e.g. *D. citri*) would be detected and reported in a timely manner.

There is clearly overlap between the types of insects that are recorded, and exotic insect species of biosecurity concern, raising the possibility of using an analogue approach to estimate surveillance sensitivity. For example, the black spittlebug (*Amarusa australis*), a harmless native species in Australia, is from the same Cicadellidae family as the glassy-winged sharp shooter (*Homalodisca vitripennis*)—the key vector for the causative pathogen *Xylella fastidiosa* of Pierce’s Disease in grapevines. As of 30-06-2016, there had been two citizen sightings of *A. australis* uploaded to ALA, and notably, both were identified on the same day as they were uploaded. However, the two species differ substantially in size and colour (*H. vitripennis* is larger and more colourful) (Fig. 1a, b), calling into question the accuracy of the analogue approach. Clearly some form of model is required to infer what this sighting rate may mean for the detection and reporting of an incursion of *H. vitripennis*, as it is larger and more colourful. Answering this question requires knowing the factors that motivate people to report the insects they discover, and applying these factors to emergency plant pests of concern to estimate the likely reporting probabilities.

This study introduces a quantitative, statistical approach to estimating the citizen reporting probabilities of insects based on their physical features. In doing so it quantifies the contribution of citizen science activities to biosecurity



(a)



(b)

Fig. 1 Insect species from within the same family (Cicadellidae in this instance) can vary considerably in size and colour, rendering an analogue approach to inferring citizen reporting probabilities based on taxonomy inaccurate. **a** Black spittle bug (*Amarusa australis*). Sighting by Dianne Clarke <http://www.bowerbird.org.au/observatio>

ns/50067. Licensed CC BY 3.0 AU <https://creativecommons.org/licenses/by/3.0/au/>. **b** Glassy-winged sharp shooter (*Homalodisca vitripennis*). Sighting by Debra Hendricks <https://www.inaturalist.org/taxa/199381-Homalodisca-vitripennis>. Licensed CC BY-NC 4.0 <https://creativecommons.org/licenses/by-nc/4.0/>.

surveillance, and enables identification of invasive insects for which citizen science would not provide effective surveillance.

Methods

Experimental design

We used a case–control experimental design to assess factors that influence the probability of an insect species being uploaded to the Atlas of Living Australia through citizen science channels (ALA 2016a, b). The Atlas of Living Australia is Australia’s national biodiversity database. It is an online biodiversity data management system which links Australia’s biological knowledge with its scientific and agricultural reference collections and other custodians of biological information. The initial focus of the ALA was on assembling a comprehensive database of collections and records generated by professional taxonomists and scientists. Subsequently, it has developed (and actively encouraged) the direct recording of sightings by non-professionals (“citizen scientists”) including the ability to upload datasets, and to receive sighting data streams from stand-alone citizen science reporting platforms. The predominant citizen science sources for the insect orders of interest (see below) were Bowerbird (<http://www.bowerbird.org.au/>), iNaturalist (<https://www.inaturalist.org/>), QuestaGame (<https://questagame.com/>) and direct citizen uploads. The predominant source of uploads from professionals was from museums within Australia’s seven states and territories participating in the Online Zoological Collections of Australian Museums (<https://www.ozcam.org.au/>) and scientific collection expeditions. Uploads from professionals for the insect orders of interest out-numbered those by citizens by a factor of *c.* 50, but note that this figure is highly dynamic.

Cases ($n = 278$) were species for which at least one record by a citizen source was uploaded through the Atlas of Living Australia (ALA) portal in the two years up until 30 June 2016. Controls ($n = 196$) were a weighted (by number of observations) sample of all species within the ALA for which there were zero records by citizens over the same period. Only the orders Coleoptera and Hemiptera were considered, as these orders encompass the vast majority of emergency plant pests (EPPs). The Hemiptera in particular appear particularly difficult to prevent from invading and are typically not detected on incursion pathways (Caley et al. 2015).

For each species, we assessed the following:

- Order (Coleopteran or Hemipteran)
- Body length (mm)
- Colour—rated on a scale from 0 (no colour) to 4 (Vividly coloured or Very highly coloured)
- Pattern—rated on a scale from 0 (no pattern) to 4 (Very highly patterned or ornate)
- Morphology—rated on a scale from 0 (no morphology of interest) to 4 (Unique or spectacular morphology)
- Range size (km²)—minimum convex polygon of all ALA records
- Observation intensity (km⁻²)—Density (intensity) of all citizen science reports for all insect species within the range over the 2-year period ($\bar{x} = 0.26 \text{ km}^{-2}$, $\bar{x} = 0.7 \text{ km}^{-2}$, 95% C.I. = 0.001–2.4 km⁻²)
- Pest status (Logical)—Result of naïve internet search for evidence of the species being a pest (see below for more details).

Examples of scoring for colour, pattern and morphology are provided in the Supplementary Information (Figures S1, S2 and S3). Scoring was undertaken by two of the authors (PC & MW). The internet search for evidence of being a pest included three searches. First, a direct Google search

including the terms “Genus species” AND “Pest”, a second search within Google Scholar using the same search terms and finally a search within the Pests and Diseases Information Library (PaDIL, <http://www.padil.gov.au/>) using the taxon name only. Hits were checked for relevance, with searching stopped either as soon as an article was found that clearly identified the taxon as being a pest (in any environment), or hits stopped containing both required search terms. We did not attempt to assess impact, for as the thrust of the work relates to citizen’s motivation to report a taxon, this albeit subjective definition of pest status suffices (i.e. the taxon has been recorded behaving in a way that is considered a pest).

Statistical analysis

We use two methods of analysis for classifying whether a species will be detected and reported. The first, logistic regression, produces easily interpreted coefficients (e.g. the effect of factor X is to increase the odds of reporting by Y). The second, random forests (Breiman 2001), is essentially a form of data mining whose performance (discriminatory ability) we would *a priori* expect to be close to the maximum obtainable. The downside is that interpreting the influence of the covariates from the many individual classification and regression trees within the “forest” so generated is not straightforward, although the relative contribution and importance of the covariates can be assessed.

Logistic regression models the reporting probability onto the ALA via citizen science platforms as a linear function of the covariates (the “linear predictor”) as:

$$\begin{aligned} \text{logit}(p) &= \beta_0^* + \beta_1 \text{Order} + \beta_2 \text{Size} + \beta_3 \text{Colour} + \beta_4 \text{Pattern} \\ &\quad + \beta_5 \text{Morphology} + \beta_6 \log_{10}(\text{Range}) \\ &\quad + \beta_7 \log_{10}(\text{ObsIntens}) + \beta_8 (\text{Pest}) \\ &= \boldsymbol{\beta}^{*'} \mathbf{x} \end{aligned} \quad (1)$$

where $p = Pr(\text{Reported} \mid \text{Covariates} \cap \text{Sampled})$ and $\boldsymbol{\beta}^{*'} = (\beta_0^*, \beta_1, \dots, \beta_k)$ are the coefficients for the k covariates \mathbf{x} .

Note that the asterisk(*) for $\boldsymbol{\beta}$ in Equation 1 signifies that this is a biased estimate of the intercept as a result of the case–control sampling process (see below). The logit transformation of a probability (p) is defined as the log of the odds. That is:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right). \quad (2)$$

Treating the scoring variables as continuous could be criticized; however, the purpose of the model is primarily for classification, and the approach facilitates better

communication of the covariate effects on reporting probabilities for less quantitative readers.

The random forest model was fitted to the same set of covariates using the default parameter settings and a forest size of 1000 trees.

Reporting probabilities over the 2-year period were converted to yearly reporting probabilities assuming citizen reporting effort could be approximated as constant across years.

Model evaluation

We evaluated the classification performance of the logistic regression model using 10-fold cross-fold validation, whereby the data were randomly divided into 10-folds, which were held out in turn and classification errors assessed. The ten values were then averaged to provide an overall estimate of classification errors expected during prediction. For the random forest model, the in-built out-of-the-bag (OOB) error rate was used as an estimate of the classification error when predicting.

The 10-fold cross-validation performance for the logistic regression model (Sensitivity = 89%, Specificity = 83%, Overall error rate = 13.5%) slightly outperformed the out-of-the-bag error rates of the random forest (Sensitivity = 89%, Specificity = 77%, Overall error rate = 16%). Armed with this knowledge that the logistic regression model was at least as good as the data-mining alternative, we used it for prediction and interpretation.

Model prediction

Predicting the probability of reporting given only the covariates requires explicit formulation that accounts for proportion of cases sampled (P_1) and controls sampled (P_0). The appropriate equation (Keating and Cherry 2004) is:

$$Pr(\text{Reported} \mid \text{Features}) = \frac{\exp\left(\boldsymbol{\beta}^{*'} \mathbf{x} - \ln\left(\frac{P_1}{P_0}\right)\right)}{1 + \exp\left(\boldsymbol{\beta}^{*'} \mathbf{x} - \ln\left(\frac{P_1}{P_0}\right)\right)} \quad (3)$$

where $\boldsymbol{\beta}^{*'} \mathbf{x}$ is either the linear predictor described by Equation 1, or the logit-transformed probability arising from the random forest model.

Model prediction with application to exotic insects

We used Eq. 3 to estimate the reporting probability for high priority pests of concern to Plant Health Australia of cross-sectoral concern. To do this, these species were scored for size, colour, pattern and morphology using the same criteria

as those applied to the ALA records. The incursion size was arbitrarily set at 100 km² (10 km × 10 km), the observation intensity set to the median (0.26 km⁻²), and the species was considered be present as a pest. The model estimates can be rerun for different desired combinations of observation intensity and outbreak size, depending on what size outbreak authorities consider they are capable of eradicating. The 100 km² was chosen as it is a figure bandied around by management agencies when considering the largest sized insect invasion that they have sufficient resources for there to be a reasonable chance of eradication.

Table 1 Features, associated coefficients, odds ratios (coefficients exponentiated to base *e*) and associated confidence intervals (C.I.) influencing the citizen reporting probability of insect species to the Atlas of Living Australia

Feature	Coef- ficient (β)	Odds ratio ($exp(\beta)$)	95% C.I.
Order ¹	β_1	2.2 (Beetles)	1.2–4.0
Size	β_2	1.1 (per mm)	1.06–1.14
Colour	β_3	2.0 (per unit score)	1.4–2.7
Pattern	β_4	3.3 (per unit score)	2.3–4.9
Morphology	β_5	1.7 (per unit score)	1.2–2.4
Range	β_6	2.2 (per 10-fold increase)	1.4–3.4
Observation intensity	β_7	3.0 (per 10-fold increase)	1.7–5.3
Pest	β_8	15.4 (Yes)	6.0–36.9

Values in parentheses indicate either the level of the factor associated with the coefficient, or for continuous measures the scaling of the parameter for a unit change associated with the coefficient. See Eq. (1) for model detail

¹Only species from the Hemipteran (Bugs) and Coleopteran (Beetles) orders were considered

Analyses were undertaken using the R software environment R Development Core Team (2017), including use of the “randomForest” package (Liaw and Wiener 2002).

Results

Effect of features on reporting

The features we recorded had a very large impact on the estimated reporting probabilities via citizen science platforms into the ALA (Table 1). The probability of a beetle being reported was considerably higher than a bug (odds ratio = 2.2, Table 1), possibly reflecting the popularity of beetle collecting. Species considered pests had a much higher reporting probability (odds ratio = 15.4, Table 1), possibly arising from the increased visibility that their plant damage brings, but also probably arising from their higher abundance and range. The estimated range of the species and the estimated activity of citizen reports also had a significant positive effect on the probability of reporting (Table 1). In terms of the features of the beetles and bugs considered, those species not reported through the ALA citizen science channels are typically smaller, less colourful, less patterned and morphologically uninteresting such as the commonly found black larder beetle (Fig. 2b), despite being a household pest, compared with those that are reported (Table 1). Indeed, despite the large number of sightings uploaded, some widespread common pest species have not been uploaded as of 30 June 2016. A further example of a widespread though unrecorded is the green peach aphid (*Myzus persicae*) (Fig. 3a), despite it causing considerable economic loss during the period of the study by vectoring beet western yellows virus during the spring of 2014.

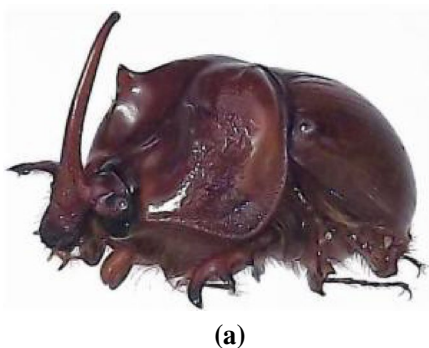


Fig. 2 Contrasts in morphology. The rarely sighted *Blackburnium acutipenne* has been uploaded to the Atlas of Living Australia via citizen science channels as of 30 June 2016, whereas the ubiquitous black larder beetle (*Dermester ater*) has not. **a** Scarab beetle (*Blackburnium acutipenne*). By Mark Golding <https://images.ala.org.au>

[image/view/136812726](https://creativecommons.org/licenses/by-nc/4.0/). Licensed CC BYNC4.0 <https://creativecommons.org/licenses/by-nc/4.0/>. **b** The black larder beetle (*Dermester ater*). By Simon Hinkley & Ken Walker, Museum Victoria <http://www.padil.gov.au/pests-and-diseases/pest/main/135708>. Licensed CC BY 3.0 AU <https://creativecommons.org/licenses/by/3.0/au/>



(a)



(b)

Fig. 3 Contrasts in colour and pattern. Despite being a widespread pest, the green peach aphid has a low citizen science reporting probability on account of its small size (1–2 mm) and “plain” looks. In contrast, *Calomela parilis* from the leaf beetle genus is a citizen observers treasure. **a** The green peach aphid (*Myszus persicae*). By David Cappaert <https://www.forestryimages.org/browse/detail.cfm?imgnum=5422731>. Licensed CC BY-NC 3.0 US <https://creativecommons.org/licenses/by-nc/3.0/us/>. **b** A scarab beetle (*Calomela parilis*). By Martin Lagerway <http://bowerbird.org.au/observations/38151>. Licensed CC BY3.0 AU <https://creativecommons.org/licenses/by/3.0/au/>.

a The green peach aphid (*Myszus persicae*). By David Cappaert <https://www.forestryimages.org/browse/detail.cfm?imgnum=5422731>. Licensed CC BY-NC 3.0 US <https://creativecommons.org/licenses/by-nc/3.0/us/>. **b** A scarab beetle (*Calomela parilis*). By Martin Lagerway <http://bowerbird.org.au/observations/38151>. Licensed CC BY3.0 AU <https://creativecommons.org/licenses/by/3.0/au/>.



(a)



(b)

Fig. 4 The Colorado potato beetle is predicted to have a high (98%) citizen reporting probability in comparison with the Russian wheat aphid (4%). **a** The Colorado potato beetle (*Leptinotarsa decemlineata*). By Scott Bauer USDA ARS [Public domain] https://commons.wikimedia.org/wiki/File:Colorado_potato_beetle.jpg, via Wikimedia-

diaCommons. **b** Russian wheat aphid (*Diuraphis noxia*). By Kansas Department of Agriculture <https://www.forestryimages.org/browse/detail.cfm?imgnum=5512060>. Licensed CC BY-NC 2.0 <https://creativecommons.org/licenses/by-nc/2.0/>.

Predicted reporting probabilities

When the model was applied to a subset of the Plant Health Australia cross-sectoral high priority pest species (HPPs) for a given set range size (100 km²) and median citizen science observation intensity, the estimated yearly reporting probability ranged from a low of 2% (e.g. sugarcane side-winder) to near 97% (Lychee longicorn beetle) (Tables S1 & S2 in Supplementary Material). Generally speaking, HPPs with very low estimated probabilities of reporting are dominated by Hemipterans (Table S2 in Supplementary Material), whilst those with high probabilities of reporting are dominated by the Coleoptera (Table S1 in Supplementary Material).

Insects that have high estimated reporting rates include the Colorado potato beetle, for which its size, colour and distinctive pattern (Fig. 4a) result in an estimated yearly

reporting probability of 0.78. In contrast, the Russian wheat aphid (Fig. 4b) has a low predicted citizen reporting probability of 3%.

Discussion

Our model has quantitatively inferred the extent to which size, colour, pattern and morphology all influence the citizen reporting probability of insects. Although this finding is unsurprising, this is the first time that the reporting probability and the factors that influence it have been quantified. This enables a more objective evaluation of the contribution of unstructured online reporting platforms to plant biosecurity surveillance. Importantly, when applied to exotic pests of biosecurity concern, we have inferred that the passive citizen reporting probabilities for many (particularly small,

nondescript bugs) would be considered insufficient for biosecurity surveillance needs.

Recent incursions of exotic pests into Australia support this estimated low reporting probability. For example, the incursion of the Russian wheat aphid went unreported by citizen scientists for possibly two years whilst spreading over a considerable area in southern Australia. It was first detected by strategic surveillance. Likewise, the tomato potato psyllid (*Bactericera cockerelli*), another small, unremarkable species was widespread, occurring on hundreds of premises in Western Australia before being detected and reported.

The citizen surveillance we have described here contributes to what is termed “general surveillance” within plant health (Hammond et al. 2016a), which is a catch-all phrase for describing surveillance that is not targeted. General surveillance activities are an important part of early detection and demonstrating area freedom (Hammond et al. 2016b). The estimated citizen reporting probabilities we have estimated here can be used to infer the likely sensitivity of general surveillance for exotic species from the “passive” citizen component. The predictions we have made here are simplistic in how they have chosen the citizen science observation intensity (simply using the median). In reality, the citizen observation intensity will vary greatly depending on the incursion location in relation to citizen science activity. The implications of this could be explored in more detail (see Pocock et al. 2017, for an example) and used as a means of directing where targeted surveillance could augment citizen surveillance. This is an area of further work.

The implication of these results for citizen reporting as a form of surveillance will vary depending on the features of the pest species of concern. Industries and environmental managers whose assets are potentially impacted by species with low reporting probabilities will clearly need to implement more structured/active surveillance if they require higher surveillance sensitivity. The citrus industry, for example, probably cannot rely on passive unstructured citizen science data streams for surveillance of *D. citri*—some form of structured surveillance will be required. In contrast, the forestry industry may consider that citizen detection and reporting of species of pine sawyers may be sufficient for their needs. Incorporating such inferred citizen surveillance reporting probabilities into the general surveillance framework is an area for further research. Targeted use of social media shows promise.

It is well known that citizen reporting rates are heavily biased in space and time (Isaac and Pocock 2015), along with the visibility of the organism in question. Here, we have demonstrated quantitatively further inter-species reporting biases relating to perceptions (pattern, colour, morphology). This finding is generalizable to most unstructured citizen science reporting platforms relating to animals and plants. Although we have demonstrated the importance of physical

features and availability for citizen reporting probability, motivations for reporting insects using online reporting portals are likely diverse and may change with time. This will be an ongoing challenge for the use of citizen surveillance.

Acknowledgements Petra Kuhnert and Chris Wikle participated in useful discussions on the paper format. Rieks van Klinken made useful comments on an earlier draft. The comments of Michael Pocock and an anonymous referee further improved the manuscript. We thank them all.

Authors contributions PC and SCB were involved in conceptualization; PC and MW undertook data curation; PC, SCB and MW undertook formal analysis; SCB helped with funding acquisition; PC and MW were involved in investigation; PC, SCB and MW contributed to the methodology; MW and PC were involved in project administration; PC and MW contributed to software; PC undertook project supervision; PC helped in validation; PC was involved in visualization; PC and MW wrote the original draft; PC and MW contributed to writing, review and editing.

Funding The study was funded by the Australian Government’s Plant Biosecurity Cooperative Research Centre Program (Plant Biosecurity CRC Project 1029).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- ALA (2016a) Atlas of Living Australia occurrence download at http://biocache.ala.org.au/occurrences/search?&q=species_group%3AInsects+country%3AAustralia+basis_of_record%3AHumanObservation+matched_name_children%3AHEMIPTERA+occurrence_date%3A%5B*+TO+2016-06-30T00%3A00%3A00Z%5D. Accessed on 30 Aug 2016
- ALA (2016b) Atlas of Living Australia occurrence download at http://biocache.ala.org.au/occurrences/search?&q=species_group%3AInsects+country%3AAustralia+matched_name_children%3ACOLEOPTERA+occurrence_date%3A%5B*+TO+2016-06-30T00%3A00%3A00Z%5D/. Accessed on 30 Aug 2016
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Caley P, Ingram R, De Barro P (2015) Entry of exotic insects into Australia: does border interception count match incursion risk? *Biol Invasions* 17:1087–1094. <https://doi.org/10.1007/s10530-014-0777-z>
- Daume S (2016) Mining twitter to monitor invasive alien species—an analytical framework and sample information topologies. *Ecol Inform* 31:70–82. <https://doi.org/10.1016/j.ecoinf.2015.11.014>
- Froud K, Oliver T, Bingham P, Flynn A, Rowswell N (2008) Passive surveillance of new exotic pests and diseases in New Zealand. In:

- Surveillance for biosecurity: pre-border to pest management New Zealand Plant Protection Society, Paihia, New Zealand, pp 97–110
- Grafton-Cardwell EE, Stelinski LL, Stansly PA (2013) Biology and management of Asian citrus psyllid, vector of the huanglong-bing pathogens. *Ann Rev Entomol* 58:413–432. <https://doi.org/10.1146/annurev-ento-120811-153542>
- Hammond NEB, Hardie D, Hauser CE, Reid SA (2016a) Can general surveillance detect high priority pests in the Western Australian Grains Industry? *Crop Prot* 79:8–14. <https://doi.org/10.1016/j.cropro.2015.10.004>
- Hammond NEB, Hardie D, Hauser CE, Reid SA (2016b) How would high priority pests be reported in the Western Australian grains industry? *Crop Prot* 79:26–33. <https://doi.org/10.1016/j.cropro.2015.10.005>
- Isaac NJB, Pocock MJO (2015) Bias and information in biological records. *Biol J Linn Soc* 115:522–531. <https://doi.org/10.1111/bij.12532>
- Jarrad FC, Barrett S, Murray J, Stoklosa R, Whittle P, Mengersen K (2011) Ecological aspects of biosecurity surveillance design for the detection of multiple invasive animal species. *Biol Invasions* 13:803–818. <https://doi.org/10.1007/s10530-010-9870-0>
- Keating KA, Cherry S (2004) Use and interpretation of logistic regression in habitat-selection studies. *J Wild Man* 68:774–789. [https://doi.org/10.2193/0022-541X\(2004\)0682.0.CO;2](https://doi.org/10.2193/0022-541X(2004)0682.0.CO;2)
- Liaw A, Wiener M (2002) Classification and regression by random forest. *R News* 2:18–22
- Pocock MJ, Roy HE, Fox R, Ellis WN, Botham M (2017) Citizen science and invasive alien species: predicting the detection of the oak processionary moth *Thaumetopoea processionea* by moth recorders. *Biol Cons* 208:146–154. <https://doi.org/10.1016/j.biocon.2016.04.010>
- R Development Core Team (2017) R: a language and environment for statistical computing. Vienna, Austria <https://www.r-project.org/>
- Thomas ML, Gunawardene N, Horton K, Williams A, OConnor S, McKirdy S, van der Merwe J (2017) Many eyes on the ground: citizen science is an effective early detection tool for biosecurity. *Biol Invasions* 9:2751–2765. <https://doi.org/10.1007/s10530-017-1481-6>
- Welvaert M, Caley P (2016) Citizen surveillance for environmental monitoring: combining the efforts of citizen science and crowdsourcing in a quantitative data framework. *Springer Plus* 5:1–14. <https://doi.org/10.1186/s40064-016-3583-5>
- Welvaert M, Al-Ghattas O, Cameron M, Caley P (2017) Limits of use of social media for monitoring biosecurity events. *PLOS ONE* 12:e0172457. <https://doi.org/10.1371/journal.pone.0172457>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.