NERC Open Research Archive
sharing our research

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

Article (refereed) - postprint

Pocock, Michael J.O.; Roy, Helen E.; Fox, Richard; Ellis, Willem N.; Botham, Marc. 2017. **Citizen science and invasive alien species: predicting the detection of the oak processionary moth Thaumetopoea processionea by moth recorders**. *Biological Conservation*, 208. 146-154.
[10.1016/j.biocon.2016.04.010](https://doi.org/10.1016/j.biocon.2016.04.010)

© 2016 Elsevier Ltd

This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
(cc) BY-NC-ND

This version available http://nora.nerc.ac.uk/514996/

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at
http://nora.nerc.ac.uk/policies.html#access

NOTICE: this is the author’s version of a work that was accepted for publication in *Biological Conservation.* Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Biological Conservation*, 208. 146-154. 10.1016/j.biocon.2016.04.010

www.elsevier.com/

Contact CEH NORA team at
noraceh@ceh.ac.uk

The NERC and CEH trademarks and logos (‘the Trademarks’) are registered trademarks of NERC in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.

Citizen science and invasive alien species: predicting the detection of the oak processionary moth *Thaumetopoea processionea* by moth recorders

Michael J.O. Pocock[a]*, Helen E. Roy[a], Richard Fox[b], Willem N. Ellis[c], & Marc Botham[a]

[a] Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK

[b] Butterfly Conservation, Manor Yard, Wareham, Dorset, BH20 5QP, UK

[c] Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands

Email addresses: MJOP: Michael.pocock@ceh.ac.uk

HER: hele@ceh.ac.uk

RF: rfox@butterfly-conservation.org

WNE: wnellis@bladmineerders.nl

MB: math2@ceh.ac.uk

* Correspondence: michael.pocock@ceh.ac.uk; +(0)1491 652807

Running title: citizen science and detection of rare events

Word count: 7084 words

Summary: 250 words

Main text: 4154 words

Acknowledgements: 112 words

References: 1290 words

Tables: 457 words

Figure legends: 344 words

Number of tables: 2

Number of figures: 4

Number of references: 43

Summary

Invasive alien species are a major cause of biodiversity change and may impact upon human well-being and the economy. If new, potentially invasive, taxa arrive then it is most cost-effective to respond as early in their establishment as possible. Information to support this can be gained from volunteers, i.e. via citizen science. However, it is vital to develop ways of quantifying volunteer recorder effort to assess its contribution to the detection of rare events, such as the arrival of invasive alien species. We considered the potential to detect adult oak processionary moths (*Thaumetopoea processionea*) by amateur naturalists recording moths at light traps. We calculated detection rates from the Netherlands, where *T. processionea* is widely established, and applied these to the spatial pattern of moth recording effort in the UK. The probability of recording *T. processionea* in the Netherlands varied across provinces from 0.05-2.4% per species of macro-moth recorded on a list of species (so equalling 1-52% for a list of 30 species). Applying these rates to the pattern of moth recording in the UK: *T. processionea* could be detected (detection >0%), if it were present, in 69% and 4.7% of 10km and 1km squares, respectively. However, in most squares detection probability is low (<1% of 1km squares have annual detection probability of >10%). Our study provides a means to objectively assess the use of citizen science as a monitoring tool in the detection of rare events, e.g. the arrival of invasive alien species, occurrence of rare species and natural colonisation.

Key words: list length analysis, monitoring, volunteer, naturalist, citizen scientist, alien invasive species

Highlights

- Outbreaks of *Thaumetopoea processionea* could be detected by amateur moth recorders
- We analysed moth trapping from the Netherlands and applied results to the UK
- *T. processionea* could be detected, if present, but mostly with low probability
- This citizen science is valuable for, but insufficient to guarantee, early detection
- It is important to quantify recorder effort in citizen science

Introduction

Globally, invasive alien species are one of the major threats to biodiversity, and they may also impact negatively upon human well-being by affecting ecosystem services and human health (Millennium Ecosystem Assessment 2005; Pejchar & Mooney 2009; Pyšek & Richardson 2010). These impacts can be costly to society, but managing invasive alien species also incurs a cost, which becomes increasing high as the species become established. Therefore, if a potentially-invasive alien species is introduced to an area, early detection is important for effective (and cost-effective) control and eradication (Hobbs & Humphries 1995; Pyšek & Richardson 2010; Blackburn *et al.* 2011). The cost of detecting nascent invasions of alien species can be high (Mehta *et al.* 2007) and is an important consideration when developing optimal strategies for responding to these species (Epanchin-Niell *et al.* 2012). Thus establishing low-cost methods to provide large-scale and long-term surveillance for invasive alien species is important.

Citizen science, that is the involvement of volunteers in the process of scientific research, including making records of species' occurrences, has great potential for the detection of invasive alien species because it can be an effective method for gaining reports of rare events, including new occurrences of invasive alien species, at a relatively low cost (Dickinson, Zuckerberg & Bonter 2010). One approach is for citizen science participants to monitor fixed plots for the presence of invasive alien species (Maxwell, Lehnhoff & Rew 2009; Crall *et al.* 2011). Success depends on volunteers being effective at detecting and identifying invasive alien species; something that has been tested and repeatedly found to be true (Delaney *et al.* 2008; Gallo & Waitt 2011; Crall *et al.* 2011). This approach requires substantial resources for coordination and volunteer recruitment but, providing all the plot data are submitted, it generates information on the absence of invasive alien species as well as their presence at these locations. However, systematically monitoring pre-defined plots does not address the need for early detection of invasive alien species at large spatial or temporal extents, such as is necessary for those species that are predicted to arrive, but precisely where and when is unknown (e.g. Roy *et al.* 2014).

An alternative citizen science approach for detecting potential invasive alien species is the opportunistic reporting of observations by the general public. While the probability of arrival of invasive alien species can be modelled (Ibáñez *et al.* 2009), actual arrivals are rare stochastic events. So, while the likelihood of a particular invasive alien species occurring in a particular place at a particular time is almost negligible, when considering a large area over a long-enough time period the overall probability of arrival is much greater. Engaging with the general public and providing tools for data submission is therefore a potentially cost-efficient method for early detection across large spatio-temporal scales (Roy *et al.* 2015).

Currently, there are several examples of citizen science in which anyone can record invasive alien species, e.g. Recording Invasive Species Counts (Roy *et al.* 2012), Invaders of Texas (Gallo & Waitt 2011) or EDDMapS (Bargeron & Moorhead 2007). These types of projects have the potential to provide good spatial coverage through promotion via the media. However, one of the serious limitations is that typically the data gathered are 'presence only' records: an absence of records provides no information on the absence of the species (i.e. the situation with no observers is indistinguishable from the situation with many observers and the species absent). In order to draw inference from the absence of records (e.g. see Isaac *et al.* 2014) it would be extremely valuable to have an assessment of recorder effort, but this is very difficult to quantify. An alternative approach is to rely upon natural history enthusiasts who are already making and submitting records (an activity that falls within the definition of citizen science; Pocock *et al.* 2015), to report sightings of new invasive alien species belonging to their taxon of interest.

As a case study, we consider one approach for the detection of the oak processionary moth *Thaumetopoea processionea* (Lepidoptera: Notodontidae) in the UK. *T. processionea* is of current concern to policy makers in the UK because it has become established in west London, following its recent spread in Belgium and the Netherlands (Groenen & Meurisse 2012). *T. processionea* can impact upon human health because the larvae shed urticating setae that can cause allergic reactions such as urticaria, conjuctivitis and respiratory difficulties (Gottschling & Meyer 2006; Fenk, Vogel & Horvath 2007; Mindlin *et al.* 2012). In some parts of the species' range and at high population densities it can be a defoliator of oak trees (Wagenhoff & Veit 2011) and so potentially could impact upon oak health and biodiversity as well (although this has not occurred in the UK to date).

T. processionea was accidentally introduced to the UK on imported oak trees (*Quercus* sp.); it was first recorded in west London in 2006 and had expanded its range by about 10km radius by 2011 despite control measures, probably mostly by natural dispersal, although human-mediated dispersal is also possible (Townsend 2013). Its gradual spread from its current range is currently monitored by professionals and trained volunteers who undertake visual surveys of the silk nests built by the communal larvae and pheromone trapping for adult male moths (Mindlin *et al.* 2012; Williams *et al.* 2013). However, this approach is not suitable for detecting occurrences of the species away from the slowly-expanding distribution in west London (e.g. new introductions to the UK or human-mediated dispersal within the UK) because any such occurrences are unpredictable, requiring the long-term surveillance of very large geographical areas with extremely high financial cost if undertaken by paid surveyors. However, other approaches such as pheromone traps have proved useful to assess spread of similar species (Sharov *et al.* 2002) and could be run by volunteers. In addition, observing larval nests in low density populations is unreliable because they typically occur in the oak canopy and are often hidden by foliage (Townsend 2013), although such biases in detection can be taken into account in data from monitoring schemes (Fitzpatrick *et al.* 2009).

In the UK, the Netherlands and elsewhere many thousands of people record moths as a hobby, submitting records to national databases. The use of light traps is an especially popular form of moth recording, partly due to its convenience, e.g. traps can be left running overnight in gardens and catches recorded the following morning (Fry & Waring 2001). These enthusiasts usually record lists of species captured, in particular all the macro-moths captured, similar to the 'checklist' approach for opportunistic recording of birds (Sullivan *et al.* 2014). This allows changes in moth prevalence over time to be quantified (Groenendijk & Ellis 2010; Fox *et al.* 2014), but also means that the absence of a species from a list can be considered a non-detection (Isaac *et al.* 2014), i.e. the non-detections can be distinguished from a lack of recording effort. This is not the case for most mass participation citizen science projects where presence-only data are collected and recording effort (including recording absences) is not known. Interpretation of such data becomes increasingly difficult as the species of interest becomes less frequently recorded and often requires recording effort to be inferred, by the recording of related species (Snäll *et al.* 2011; Isaac *et al.* 2014).

Our aim in the current project was to use data from a region where *T. processionea* is established (the Netherlands) to calculate the probability that moth recorders detect *T. processionea* when it is present, and then to apply these detection probabilities to the current pattern of citizen science moth recording in the UK. From this we could estimate the probability that moth recorders would provide early detection of *T. processionea* across the UK.

Methods

The Noctua database holds data from volunteer moth recorders in The Netherlands and currently holds 4.5 million records (Groenendijk & Ellis 2010). We extracted data from the Noctua database on moth records during the flight period of *T. processionea* in 2002-2013. *T. processionea* was established in the Netherlands over this period. The flight period was 25 July- 30 August, which was defined as the range of dates where the number of records of *T. processionea* was at least 10% of the maximum number of records per day for the years 2002-2010 and 2012-2013 (the year 2011 was removed due to an apparent artefact in the data; Fig S1). The records in the Noctua database comprise species identity, grid reference, date and recorder name. We aggregated the moth records by 'species lists' (Szabo *et al.* 2010), where a species list comprises the moths recorded during one night of moth trapping; specifically we defined a 'species list' as a unique combination of 1km grid square and date. We did not use recorder name to distinguish between samples because names are not unique and can be recorded in multiple ways within the database (e.g. with or without initials and first names) and multiple recorders could have submitted the same record (e.g. when they all took part in a group moth trapping event). Considering the unique combination of 1km square and date may occasionally lead to aggregation of separate species lists (where they occurred in the same 1km grid square on the same night), but our experience suggests that this occurs only rarely at the 1km resolution.

We then calculated the probability of recording *T. processionea* (OPM) while taking account of the list length (i.e. the average ‘per-species recording probability’: $\bar{S}_{\mathrm{OPM}}$). There is spatial variation in the prevalence of *T. processionea* across the Netherlands, so throughout we undertook analyses separately in each province.

To calculate the probability that *T. processionea* had been recorded in a species list we firstly calculated the total probability that *T. processionea* was recorded on a list of length L ($P_{\mathrm{OPM},L}$; eqn 1).

$$P_{\mathrm{OPM},L} = N_{\mathrm{OPM},L}/N_{\mathrm{total},L} \quad \text{[eqn 1]}$$

where, for a given list of length L, $N_{\mathrm{total},L}$ is the total number of lists and $N_{\mathrm{OPM},L}$ is the number of lists in which *T. processionea* was present.

Following Szabo *et al.* (2010), we expected that the probability of detecting *T. processionea* ($P_{\mathrm{OPM},L}$) on a list would increase with increasing list length (L). This is because list length gives an indication of recording effort, assuming that all recorders record every macro-moth species they identify, which is typical behaviour among moth recorders in north-western Europe. It could be possible to test this assumption quantitatively in the future because biased recording of some species would result in them being more likely to be recorded on shorter lists. In the case of light traps running overnight, ‘effort’ is a function of factors including the effectiveness of the moth trap, duration of trapping, number of traps used, weather conditions, moon phase and local habitat. Calculating the per-species probability of recording *T. processionea* ($S_{\mathrm{OPM},L}$) for each category of list length L in each province takes the list length into account (eqn 2).

$$S_{\mathrm{OPM},L} = 1 - \exp(\ln(1 - P_{\mathrm{OPM},L}) / L) \quad \text{[eqn 2]}$$

Therefore, $S_{\mathrm{OPM},L}$ was calculated for each value of the list length L. We calculated the average $S_{\mathrm{OPM},L}$ (eqn 3) across a set of these values of L (i.e. treating each list length category, not the lists themselves, as the data) which met the criteria that: (i) the value of the list length was at least six (i.e. $L>5$), (ii) there were at least six lists of that list length (i.e. $N_{\mathrm{OPM},L}>5$ for each value of L), and (iii) there were some/all lists of that list length in which *T. processionea* was absent (i.e. $P_{\mathrm{OPM},L}<1$). We excluded these three cases because (i, ii) observation of the results (Fig. S1) suggested that estimates of $S_{\mathrm{OPM},L}$ tended to be lower than expected when the list lengths were very short or few lists were included in the category of length L, and (iii) in these cases $S_{\mathrm{OPM},L}$ was constrained to be one and appeared to be biased high. From $\bar{S}_{\mathrm{OPM}}$ for each province, we could back-calculate the estimated probability of recording *T. processionea* for a list of length L ($\hat{P}_{\mathrm{OPM},L}$) as one minus the probability of not detecting *T. processionea* (eqn 4).

$$\bar{S}_{\mathrm{OPM}} = \frac{1}{M}\sum_{i=1}^{M} S_{\mathrm{OPM},L} \quad \text{[eqn 3]}$$

where M is the subset of values of the list length as described in the text

$$\widehat{\mathrm{P}}_{\mathrm{OPM},L} = 1 - (1 - \bar{\mathrm{S}}_{\mathrm{OPM}})^{L} \quad \text{[eqn 4]}$$

We then applied the values of $\bar{\mathrm{S}}_{\mathrm{OPM}}$ obtained from data from the Netherlands to the pattern of moth recording across the UK. Specifically, we calculated estimated detection rate in the UK ($\widehat{D}$: eqn 5), by combining (1) the probability of recording *T. processionea* per recording event ($\bar{\mathrm{S}}_{\mathrm{OPM}}$) for Netherlands providences, with (2) the recording effort in the UK (i.e. the list length and frequency of recording). We extracted information on all recording events between 25 July and 30 August from the UK National Moth Recording Scheme database (Fox *et al.* 2010), which currently holds over 20 million records. We therefore assumed that the flight period of *T. processionea* was the same in the UK as the Netherlands. There can be a lag in the UK from record submission and verification by county recorders to acceptance into the database, so to minimise this effect we considered the records for the ten-year period 2000-2009. As for the Netherlands dataset, a recording event was defined as the list of species recorded in a unique combination of 1km grid square and date. Therefore, for any region (e.g. a 1km square) and any year, we knew the length (L) of each list (n = 1 to the total N lists in that region) and so could calculate, across all lists and for a given value of $\bar{\mathrm{S}}_{\mathrm{OPM}}$, the estimated probability of detecting *T. processionea* ($\widehat{D}$; eqn 5). Note that $\widehat{D}$ is scale-free, so it can be calculated at any extent. However, it does assume that the selected value of $\bar{\mathrm{S}}_{\mathrm{OPM}}$ is appropriate over the whole of each region (e.g. a whole 1km or 10km square). For the results presented here we calculated the average $\widehat{D}$ across the years 2000-2009.

$$\widehat{D} = 1 - \prod_{n=1}^{N}[(1 - \bar{\mathrm{S}}_{\mathrm{OPM}})^{L_n}] \quad \text{[eqn 5]}$$

Results

The probability of recording *T. processionea* in the Netherlands

Our dataset for moth recording in the Netherlands between 25 July and 30 August in 2002-2013 comprised 53 781 lists (i.e. unique combinations of 1km grid square and date) of 417 614 individual species records. *T. processionea* was recorded 2 640 times (i.e. it comprised 0.6% of species records and occurred on 4.9% of lists).

The probability of recording *T. processionea* per recording event ($\mathrm{P}_{\mathrm{OPM},L}$) increased with increasing list length (L), as we expected (Fig. 1 a-l). The average per-species detection probability ($\bar{\mathrm{S}}_{\mathrm{OPM}}$), calculated from a subset of all the list lengths (Fig. 1 and S2) was back-calculated to the observed list length ($\widehat{\mathrm{P}}_{\mathrm{OPM},L}$) and showed a good fit to the observed data (Fig. 1).

We found that provinces varied in the average per-species probability of recording *T. processionea* (Fig. 1 m and n). The two provinces in the south-east of the Netherlands, where *T. processionea* had been established longest, had per-species detection probabilities of 2.1-2.4% (i.e. this was the chance that a new species on a list at a recording event would be *T. processionea*; Fig. 1k-l). This equates to

47-52% chance of recording *T. processionea* when a recording event obtained a list of 30 species. The four provinces with medium detection rates had an average per-species probability of recording of about 1.4% (Fig. 1 g-j), equating to a 34% chance of recording *T. processionea* for a list of 30 species. Finally those provinces with the lowest detection rate, the per-species detection rate varied from 0.05 to 0.4% (Fig. 1 a-f), so for a list of 30 species there was a 1-11% chance of detecting *T. processionea*.

The probability of recording *T. processionea*, if it was present, in the UK

The number of species lists recorded in the UK during the flight period of *T. processionea* (25 July-30 August, i.e. assumed to be the same as in the Netherlands) between 2000 and 2009 was 136 344 (range per year: 9 753-15 369) with a total of 1 618 661 individual species records. *T. processionea* was not recorded on any list in this dataset, even though it was present in western London from 2006 and had been recorded at various sites on the south coast of England as a presumed immigrant from continental Europe. There were lists from 2 119 (69%) of the 3 055 10km squares in the UK during 25 Jul-30 August 2000-2009 (Fig. 2) and 12 190 (4.7%) of 256 663 1km grid squares in the UK, i.e. for each 10km square, on average only five of the 100 1km squares had records. Squares with lists were distributed across the UK although parts of Scotland and Northern Ireland were relatively sparsely covered (Fig. 2).

Applying the per-species recording probabilities from the Netherlands to the UK showed the coverage of squares at different detection thresholds (Table 1; Fig. 1). There was a greater than 0% chance of moth-recorders detecting *T. processionea*, if it had been present, in 69% of 10km squares, but only 4.7% of 1km squares, in the UK (Table 1). However, considering the situation with higher detection thresholds, the overall coverage is lower and patchy (Table 1; Fig. 1); when considering the threshold of $\widehat{D} > 50\%$ (i.e. chances are *T. processionea* would be recorded, if it was present, in any year with the pattern of recording effort during 2000-2009) then only 5.5% of 10km squares and <0.1% of 1km squares meet this criteria (Table 1; Fig. 2).

However, for the outbreaks in their earliest stages, occurrence will be at a much smaller spatial extent than the 10km square. The range (area of the minimum convex polygon) of *T. processionea* in west London in 2009 was just 58km^2 (Fig. 3). Finer resolution analysis of the data within a 50km square covering west London where *T. processionea* is established, shows how recording effort is distributed. At the resolution of 10km squares, most squares have a 10-50% annual probability of detecting *T. processionea*. However, actual recording occurs at a much finer resolution (i.e. within 1km squares, by the definition of a recording event used in the current study). Within the 50km square, most of the 1km squares have a 0% probability of detecting *T. processionea* showing the importance of considering spatial resolution of recording effort relative to invasive species range size.

Discussion

Currently citizen science is promoted as a potential method for conducting cost-effective environmental monitoring, including the early detection of invasive alien species and disease (Tree Health and Plant Biosecurity Expert Taskforce 2012; Dickinson *et al.* 2012; Roy *et al.* 2015). ‘Opportunistic’ recording can produce data which is suitable to monitor many species when recording is via a ‘checklist’ approach or when non-detections can be inferred (Snäll *et al.* 2011; Sullivan *et al.* 2014; Isaac *et al.* 2014), but is less useful as the focal species becomes less frequently recorded. Interpreting the results of projects in which people submit records of potentially invasive alien species (i.e. presence-only data from mass participation citizen science) is difficult because recorder effort cannot usually be quantified. It is important to distinguish lack of records due to the species being absent from a lack of recorders. In this study, by considering volunteers who record the target species as a by-product of general recording, we were able to estimate the probability that volunteers recording macro-moths would detect the moth oak processionary, *T. processionea*.

From our findings in this study we draw two conclusions. Firstly, across much of the UK there is a greater than zero probability that moth recorders will detect *T. processionea* if it is present; therefore this form of ‘citizen science’ could be useful for its early detection. Secondly, the actual probability of detecting *T. processionea* is low and patchy across the UK, especially at fine spatial resolutions (i.e. within 1km grid squares), so this form of monitoring is unlikely to be sufficient in providing early detection of *T. processionea*. The environment in the Netherlands (where we parameterised the model) is not a perfect match to the UK (where we applied the model), but we are confident that it is similar enough for our results to provide a good indication of the likely detection of *T. processionea* by moth recorders in the UK. Given the way naturalists record moths at light traps, it is unlikely that this distinctive species would be missed or mis-identified, if present, but lack of awareness could contribute to mis-identifications leading to non-detections for more cryptic or less distinctive species. Overall, maps of quantified recording effort (e.g. Fig. 2 for the amateur naturalists considered in this study) could be combined with maps of hazard, e.g. *T. processionea* arrival or spread (Cowley, Johnson & Pocock 2015), if such maps were available, to optimise the targeting of additional recording effort, e.g. professional monitoring or targeted advertising.

Volunteers who record moths do so for a range of motivations, including their own enjoyment, connection with nature and wanting to contribute to scientific knowledge (e.g. Fox *et al*. 2014). The early detection of invasive alien species is a by-product of this recording rather than an intended aim. Other people may have different motivations for taking part in the search for and reporting of *T. processionea*, e.g. arboriculturists, land managers, local council staff and householders concerned about human health impacts. These will all contribute to reporting, so the overall situation for effective early detection is not as pessimistic as it might seem from our analysis. However, as we have stressed, this additional recording effort cannot be easily quantified, meaning that it is not possible to predict detection probability, and so it is difficult to effectively manage resources to strategically optimize detection (Hauser & McCarthy 2009).

Asymmetry of information and data flow

If *T. processionea* is not detected then, as we have discussed, it is important to assess the probability that it was present but not detected. However, the converse is very different. If *T. processionea* is detected, then it is important for decision makers that the information is available as quickly as possible in order to determine appropriate action. Currently in Great Britain (GB) there is an alert system for early detection of invasive alien species (Roy *et al.* 2012, 2015), which has an organized structure to support rapid data flow (Fig. 4). There are three potential bottlenecks to data flow. The first is the submission of a record by the observer. Websites and especially smartphone apps facilitate the reporting of potential target species (August *et al.* 2015), but rely on people being aware of and utilising them: communication is important. The second potential bottleneck is the verification of records by experts (volunteers or professionals). A successful public awareness campaign can result in a large number of misidentified records and, even if supporting information (e.g. photographs) are submitted, resources are still needed to support this (Roy *et al.* 2015). The third potential bottleneck is the onward flow of data to those who are able to mount an appropriate response. Inter-operable data systems are an ambition (Graham *et al.* 2008) but the proliferation of individual citizen science projects can put efficient data flow under risk, and so it is incumbent upon project organizers to consider this as utmost importance.

Using citizen science as a tool for detection of rare events

In the current study we have specifically considered the effectiveness of volunteers to provide information on the presence and absence of a target species, in this case *T. processionea*, which can be compared to other methods for the detection of rare events (Table 2). Typically, active surveillance (which could be by professionals or volunteers) is considered when seeking to model the optimal monitoring strategies for early detection of rare events (Maxwell, Lehnhoff & Rew 2009). However, passive surveillance by the general public (or a trained subset thereof) has the potential to permit the long-term, large-scale surveillance of rare events at relatively little cost (Pocock *et al.* 2013); the public are potentially a resource “ready to act as the need arises” (Cooper *et al.* 2007). It is most likely to be successful when the rare events are very noticeable or directly impact people, and is dependent upon having a high public profile, e.g. extensive media coverage. This approach has been deemed successful in the past (Aitkenhead 1981; Hesterberg *et al.* 2009) even though it is not possible to directly assess the recorder effort. Alternatively, people can become involved with focussed monitoring, e.g. by deploying and checking pheromone traps (Sharov *et al.* 2002) although, as with other approaches, detection probability still needs to be considered (Fitzpatrick *et al.* 2009) and the issue of people not reporting absences remains problematic. Also, as citizen science continues to develop, further research on participants' motivations (Rotman *et al.* 2012; Nov, Arazy & Anderson 2014) will enhance our ability to effectively use citizen science as a tool for the detection of rare events (Pocock *et al.* 2013).

Conclusion

There is great enthusiasm for citizen science and its role in environmental monitoring. Citizen science clearly does have a role to play in the early detection of invasive alien species, and can also be applied to other rare events such as occurrence of wildlife disease (Kulasekera *et al.* 2000; Hesterberg *et al.* 2009), unusual weather (http://www.cocorahs.org) and landslips (https://britishgeologicalsurvey.crowdmap.com/). When assessing results from such projects it is important to quantify the recorder effort in order to distinguish the absence of records (because there are no recorders) from the absence of the event (even though potential recorders were present). However with presence-only data this is often hard to achieve. The approach in this study was to quantify recording effort by moth recorders and use this to estimate the probability of detecting an invasive alien moth, *T. processionea*, if it was present. Although moth recorders are just one subset of the potential recorders, it shows that there is a chance of recording *T. processionea* across much of the UK, but that the chance is often quite small, making records from moth recorders a valuable, but not sufficiently effective, component of an early detection network for *T. processionea*, This result is relevant to other 'rare events' including the detection of rare or highly threatened resident species and newly-colonising species. Citizen science in all its forms is bound to play an increasing role in detection of rare events but it requires thoughtful enthusiasm rather than hype to ensure that it provides many opportunities for excellent cost-effective science.

Acknowledgements

The research was funded by the Department for Environment, Food and Rural Affairs (Defra) [grant number TH0101]. We gratefully acknowledge the contribution of thousands of moth recorders in the Netherlands and the UK. The UK National Moth Recording Scheme was funded by the Heritage Lottery Fund, Butterfly Conservation, Environment Agency, Redwing Trust, Natural England, Natural Resources Wales, Northern Ireland Environment Agency, Royal Entomological Society and Scottish Natural Heritage. The Noctua database is owned by Dutch Butterfly Conservation and the Working Group Lepidoptera Faunistics. The Biological Record Centre receives support from the Joint Nature Conservation Committee and the Natural Environment Research Council (via National Capability funding to the Centre for Ecology & Hydrology). The GB non-native species Alert System is co-funded through Defra in partnership with the Joint Nature Conservation Committee and the Natural Environment Research Council. The Non-Native Species Secretariat provided invaluable support in the development of the Alert system.

Appendix

Figure S1. The phenology of *Thaumetopoea processionea* in the Netherlands, based on the number of records in the Noctua database.

Figure S2. The per-species recording probability for *Thaumetopoea processionea* (S_{OPM}) in each province in the Netherlands.

References

Aitkenhead, P. (1981) Colorado beetle - recent work in preventing its establishment in Britain. *EPPO Bulletin*, **11**, 225–234.

August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T. & Jepson, P. (2015) Emerging technologies for biological recording. *Biological Journal of the Linnean Society*, **115**, 731–749.

Bargeron, C.T. & Moorhead, D.J. (2007) EDDMapS: Early detection and distribution mapping system for the South-east Exotic Pest Plant Council. *Wildland Weeds*, **Fall 2007**, 4–8.

Blackburn, T.M., Pyšek, P., Bacher, S., Carlton, J.T., Duncan, R.P., Jarošík, V., Wilson, J.R.U. & Richardson, D.M. (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution*, **26**, 333–9.

Cooper, C.B., Dickinson, J.L., Phillips, T. & Bonney, R. (2007) Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society*, **12**, 11.

Cowley, D.J., Johnson, O. & Pocock, M.J.O. (2015) Using electric network theory to model the spread of oak processionary moth, Thaumetopoea processionea, in urban woodland patches. *Landscape Ecology*, **30**, 905–918.

Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K. a., Graham, J. & Waller, D.M. (2011) Assessing citizen science data quality: an invasive species case study. *Conservation Letters*, **4**, 433–442.

Delaney, D.G., Sperling, C.D., Adams, C.S. & Leung, B. (2008) Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions*, **10**, 117–128.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., L Crain, R., Martin, J., Phillips, T. & Purcell, K. (2012) The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, **10**, 291–297.

Dickinson, J.L., Zuckerberg, B. & Bonter, D.N. (2010) Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 149–172.

Epanchin-Niell, R.S., Haight, R.G., Berec, L., Kean, J.M. & Liebhold, A.M. (2012) Optimal surveillance and eradication of invasive species in heterogeneous landscapes. *Ecology Letters*, **15**, 803–812.

Fenk, L., Vogel, B. & Horvath, H. (2007) Dispersion of the bio-aerosol produced by the oak processionary moth. *Aerobiologia*, **23**, 79–87.

Fitzpatrick, M., Preisser, E., Ellison, A. & Elkinton, J.S. (2009) Observer bias and the detection of low-density populations. *Ecological Applications*, **19**, 1673–1679.

Fox, R., Oliver, T.H., Harrower, C., Parsons, M.S., Thomas, C.D. & Roy, D.B. (2014) Long-term changes to the frequency of occurrence of British moths are consistent with opposing and synergistic effects of climate and land-use changes. *Journal of Applied Ecology*, **51**, 949–957.

Fox, R., Randle, Z., Hill, L., Anders, S., Wiffen, L. & Parsons, M.S. (2010) Moths count: recording moths for conservation in the UK. *Journal of Insect Conservation*, **15**, 55–68.

Fry, R. & Waring, P. (2001) *A Guide to Moth Traps and Their Use*. Amateur Entomologists' Society, London, UK.

Gallo, T. & Waitt, D. (2011) Creating a successful citizen science model to detect and report invasive species. *BioScience*, **61**, 459–465.

Gottschling, S. & Meyer, S. (2006) An epidemic airborne disease caused by the oak processionary caterpillar. *Pediatric Dermatology*, **23**, 64–6.

Graham, J., Simpson, A., Crall, A., Jarnevich, C., Newman, G. & Stohlgren, T.J. (2008) Vision of a cyberinfrastructure for nonnative, invasive species management. *BioScience*, **58**, 263–268.

Groenen, F. & Meurisse, N. (2012) Historical distribution of the oak processionary moth Thaumetopoea processionea in Europe suggests recolonization instead of expansion. *Agricultural and Forest Entomology*, **14**, 147–155.

Groenendijk, D. & Ellis, W.N. (2010) The state of the Dutch larger moth fauna. *Journal of Insect Conservation*, **15**, 95–101.

Hauser, C.E. & McCarthy, M.A. (2009) Streamlining 'search and destroy': cost-effective surveillance for invasive species management. *Ecology Letters*, **12**, 683–92.

Hesterberg, U., Harris, K., Stroud, D., Guberti, V., Busani, L., Pittman, M., Piazza, V., Cook, A. & Brown, I. (2009) Avian influenza surveillance in wild birds in the European Union in 2006. *Influenza and Other Respiratory Viruses*, **3**, 1–14.

Hobbs, R.J. & Humphries, S.E. (1995) An integrated approach to the ecology and management of plant invasions. *Conservation Biology*, **9**, 761–770.

Ibáñez, I., Silander Jr, J.A., Allen, J.M., Treanor, S.A. & Wilson, A. (2009) Identifying hotspots for plant invasions and forecasting focal points of further spread. *Journal of Applied Ecology*, **46**, 1219–1228.

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P. & Roy, D.B. (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, **5**, 1052–1060.

Kulasekera, V.L., Kramer, L., Nasci, R.S., Mostashari, F., Cherry, B., Trock, S.C., Glaser, C. & Miller, J.R. (2000) West Nile virus infection in mosquitoes, birds, horses, and humans, Staten Island, New York, 2000. *Emerging Infectious Diseases*, **7**, 722–5.

Maxwell, B.D., Lehnhoff, E. & Rew, L.J. (2009) The rationale for monitoring invasive plant populations as a crucial step for management. *Invasive Plant Science and Management*, **2**, 1–9.

Mehta, S. V., Haight, R.G., Homans, F.R., Polasky, S. & Venette, R.C. (2007) Optimal detection and control strategies for invasive species management. *Ecological Economics*, **61**, 237–245.

Millennium Ecosystem Assessment. (2005) *Ecosystems and Human Well-Being: Biodiversity Synthesis*. World Resources Institute, Washington DC.

Mindlin, M.J., le Polain de Waroux, O., Case, S. & Walsh, B. (2012) The arrival of oak processionary moth, a novel cause of itchy dermatitis, in the UK: experience, lessons and recommendations. *Public Health*, **126**, 778–81.

Nov, O., Arazy, O. & Anderson, D. (2014) Scientists@Home: what drives the quantity and quality of online citizen science participation? *PloS ONE*, **9**, e90375.

Pejchar, L. & Mooney, H.A. (2009) Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution*, **24**, 497–504.

Pocock, M.J.O., Chapman, D.S., Sheppard, L.J. & Roy, H.E. (2013) *Developing a Strategic Framework to Support Citizen Science Implementation. Final Report to SEPA.* Centre for Ecology & Hydrology, Wallingford, Oxon.

Pocock, M.J.O., Roy, H.E., Preston, C.D. & Roy, D.B. (2015) The Biological Records Centre: a pioneer of citizen science. *Biological Journal of the Linnean Society*, **115**, 475–493.

Pyšek, P. & Richardson, D.M. (2010) Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources*, **35**, 25–55.

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D. & Jacobs, D. (2012) Dynamic changes in motivation in collaborative citizen-science projects. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 217.

Roy, H.E., Bacon, J., Beckmann, B., Harrower, C.A., Hill, M.O., Isaac, N.J.B., Preston, C.D., Rathod, B., Rorke, S.L., Marchant, J.H., Musgrove, A., Noble, D., Sewell, J., Seeley, B., Sweet, N., Adams, L., Bishop, J., Jukes, A.R., Walker, K.J. & Pearman, D. (2012) *Non-Native Species in Great Britain: Establishment, Detection and Reporting to Inform Effective Decision Making*. Defra, London, UK.

Roy, H.E., Peyton, J., Aldridge, D.C., Bantock, T., Blackburn, T.M., Britton, R., Clark, P., Cook, E., Dehnen-

Schmutz, K., Dines, T., Dobson, M., Edwards, F., Harrower, C., Harvey, M.C., Minchin, D., Noble, D.G., Parrott, D., Pocock, M.J.O., Preston, C.D., Roy, S., Salisbury, A., Schönrogge, K., Sewell, J., Shaw, R.H., Stebbing, P., Stewart, A.J.A. & Walker, K.J. (2014) Horizon scanning for invasive alien species with the potential to threaten biodiversity in Great Britain. *Global Change Biology*, **20**, 3859–3871.

Roy, H.E., Rorke, S.L., Beckmann, B., Booy, O., Botham, M.S., Brown, P.M.J., Harrower, C., Noble, D., Sewell, J. & Walker, K. (2015) The contribution of volunteer recorders to our understanding of biological invasions. *Biological Journal of the Linnean Society*, **115**, 678–689.

Sharov, A.A., Leonard, D., Liebhold, A.M., Roberts, E.A. & Dickerson, W. (2002) 'Slow The Spread': A national program to contain the gypsy moth. *Journal of Forestry*, **100**, 30–36.

Snäll, T., Kindvall, O., Nilsson, J. & Pärt, T. (2011) Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*, **144**, 804–810.

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J.W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W.M., Iliff, M.J., Lagoze, C., La Sorte, F.A., Merrifield, M., Morris, W., Phillips, T.B., Reynolds, M., Rodewald, A.D., Rosenberg, K. V., Trautmann, N.M., Wiggins, A., Winkler, D.W., Wong, W.-K., Wood, C.L., Yu, J. & Kelling, S. (2014) The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation*, **169**, 31–40.

Szabo, J.K., Vesk, P.A., Baxter, P.W.J. & Possingham, H.P. (2010) Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. *Ecological Applications*, **20**, 2157–2169.

Townsend, M. (2013) Oak Processionary Moth in the United Kingdom. *Outlooks on Pest Management*, **24**, 32–38.

Tree Health and Plant Biosecurity Expert Taskforce. (2012) *Interim Report*.

Wagenhoff, E. & Veit, H. (2011) Five Years of Continuous Thaumetopoea processionea Monitoring: Tracing Population Dynamics in an Arable Landscape of South-Western Germany. *Gesunde Pflanzen*, **63**, 51–61.

Williams, D.T., Straw, N., Townsend, M., Wilkinson, A.S. & Mullins, A. (2013) Monitoring oak processionary moth *Thaumetopoea processionea* L. using pheromone traps: the influence of pheromone lure source, trap design and height above the ground on capture rates. *Agricultural and Forest Entomology*, **15**, 126–134.

Table 1. The percentage of total 10km and 1km grid squares in the UK which meet the criteria for the annual probability of detecting *T. processionea* if it was present ($\widehat{D}$), based on the per species probability of recording *T. processionea* ($\bar{S}$) in the Netherlands (2002-2013) and the pattern of moth-recording in the UK (2000-2009). The different values of $\bar{S}$ are taken from the different providences in the Netherlands and are assumed to be a function of the local density of *T. processionea*, with very low to low values considered to be most relevant to situations where *T. processionea* is in the early stages of establishment

	Percentage of 10km grid squares				Percentage of 1km grid squares			
Per-species probability of recording ($\bar{S}$)	Very low (0.05%)	Low (0.39%)	Medium (1.4%)	High (2.4%)	Very low (0.05%)	Low (0.39%)	Medium (1.4%)	High (2.4%)
Threshold for predicted detection probability ($\widehat{D}$)								
>0%	69.4	69.4	69.4	69.4	4.7	4.7	4.7	4.7
>1%	30.0	51.1	57.5	59.7	0.5	1.8	2.5	2.7
>10%	6.5	24.9	36.8	42.3	0.1	0.3	0.7	0.9
>50%	0.2	5.5	12.4	15.4	<0.1	<0.1	0.1	0.1

Table 2. A framework for considering the role of citizen science in the detection of rare events, such as invasive or rare species

Type of recording	Opportunistic surveillance (presence only records of target species)	Opportunistic surveillance (as a byproduct of recording other events, e.g. other species occurrences)	Systematic surveillance (monitoring by volunteers)	Active surveillance (by professionals)
Participants	General public = mass participation citizen science	Volunteers already (recording the other events)	Participants undertaking regular monitoring at known locations and known times	Contracted surveyors; they may be actively searching an area or undertaking regular monitoring at fixed sites
Recording effort	Presence-only records, so recording effort is very difficult to assess	Can be assessed by current recording of species that are not the intended target	Protocols mean that efforts can be prescribed and known	Surveyors are under contract so (in theory) their effort can be quantified and managed
Opportunities	The potential for large-scale long-term monitoring at low cost	It is supported by the enthusiasm and motivation of those already engaged in recording other events	Volunteers can be as accurate as professionals (and this can be tested) and provide cost-efficient long-term monitoring	Surveyors are under contract so they are instructed where to survey
Challenges	Sustaining interest; Regular promotion; Feedback essential but time-consuming Responding to mis-identifications; recording effort is difficult to quantify	Promoting rapid submission of records of target events; ensuring that records are dealt with efficiently and passed on to stakeholders	Requires resources to recruit and retain participants; unlikely to detect first occurrence of a rare event unless the location of such events are predictable and locations selected to match	Incurs a direct (often large) on-going cost to employ people

Figure 1. The probability of recording *T. processionea* depends on the number of species per recording event and varies by the province in the Netherlands. In a-l the circles show the proportion of recording events of each list length in which *T. processionea* was recorded. The line shows the estimate that was back-calculated from the average per-species recording probability (given in the title of each graph along with the two-letter code for the province name) calculated as the average from a subset of the data (shown as the points that are filled (see text for details). For completeness the remaining data not used in the calculation are showed as open circles). Provinces are ordered by increasing per-species probability of recording *T. processionea*. The average per-species recording probability in the provinces occurs in three bands (m), which are distributed as shown in (n).

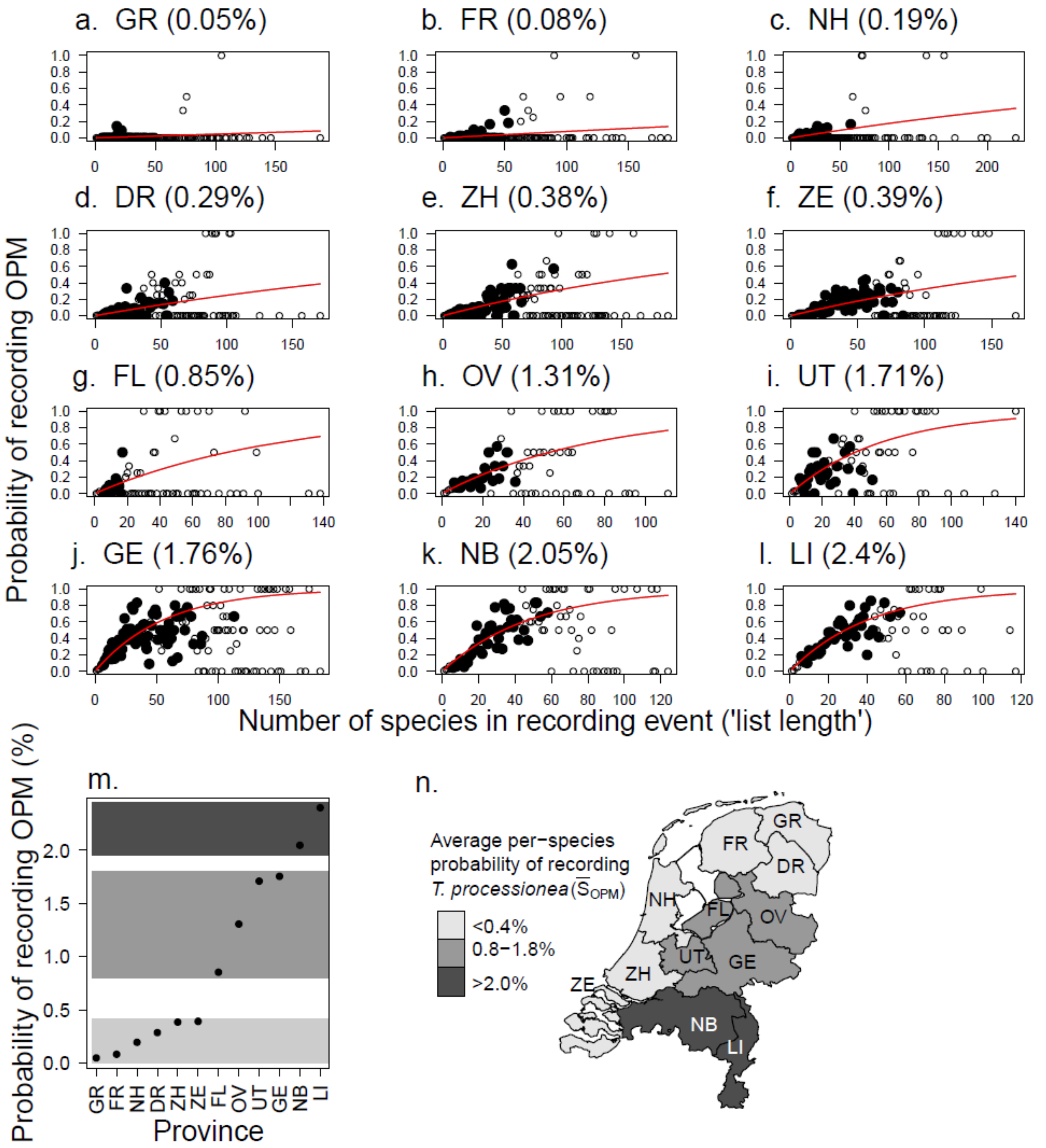

Fig. 2. The average annual probability of detecting *T. processionea* ($\widehat{D}$), if it were present, in 10km grid squares in the UK based on the observed recording effort during 25 July-30 August in 2000-2009. The results are shown when considering a low per-species probability of recording *T. processionea* ($\bar{S}_{\mathrm{OPM}}$=0.0039), based on modelling from the Netherlands (Fig. 1).

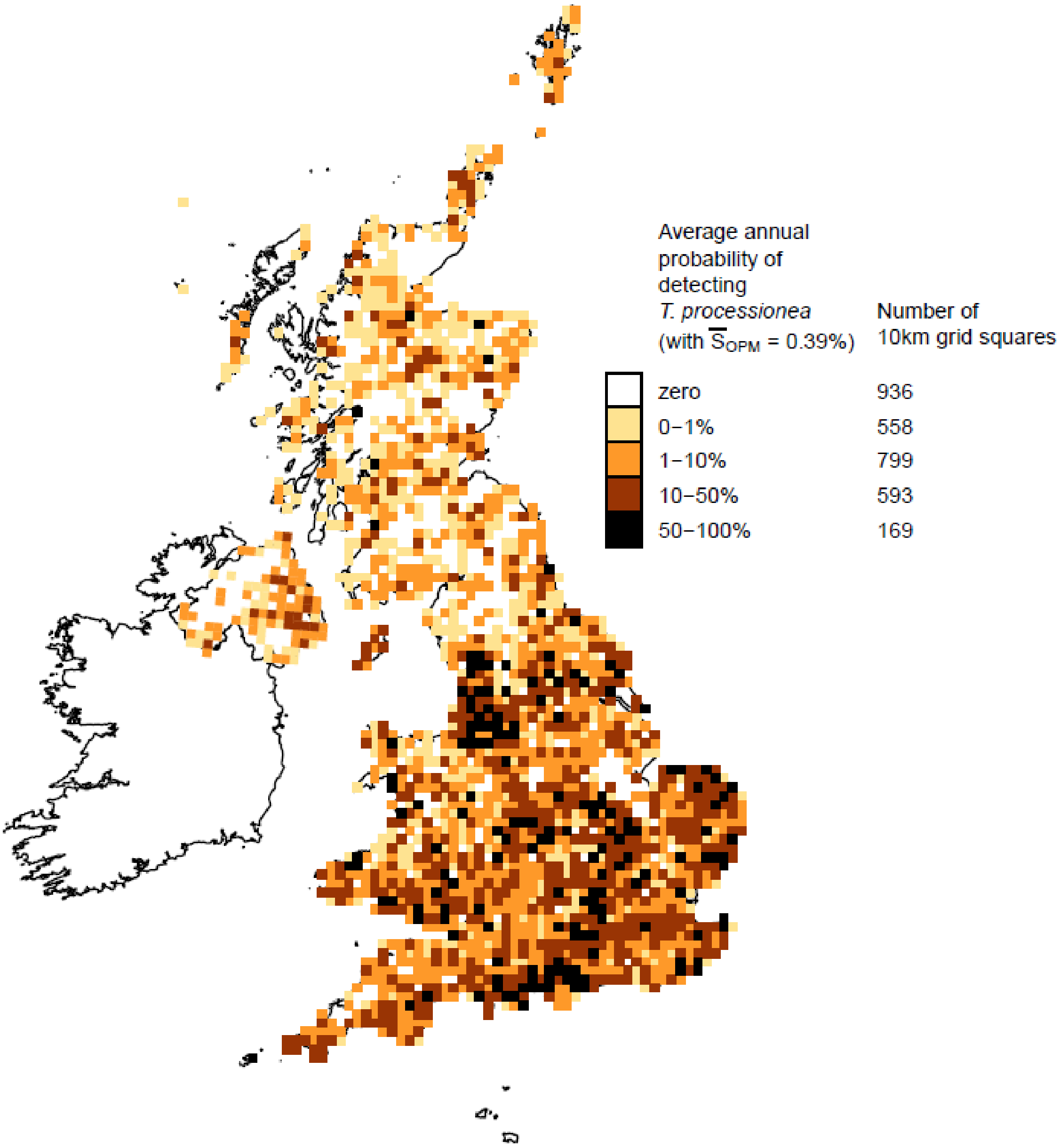

Figure 3. The probability of detecting *T. processionea*, if it was present, in (a) 10km and (b) 1km grid squares in a 50km square containing the current range of *T. processionea* in west London (thick black outline is the minimum convex polygon of the range of *T. processionea* in 2009) based on the average recording effort by moth recorders during 25 July-30 August in 2000-2009 and a low probability of recording *T. processionea* in the Netherlands (Fig. 1). (c) The box indicates the area magnified in a and b.

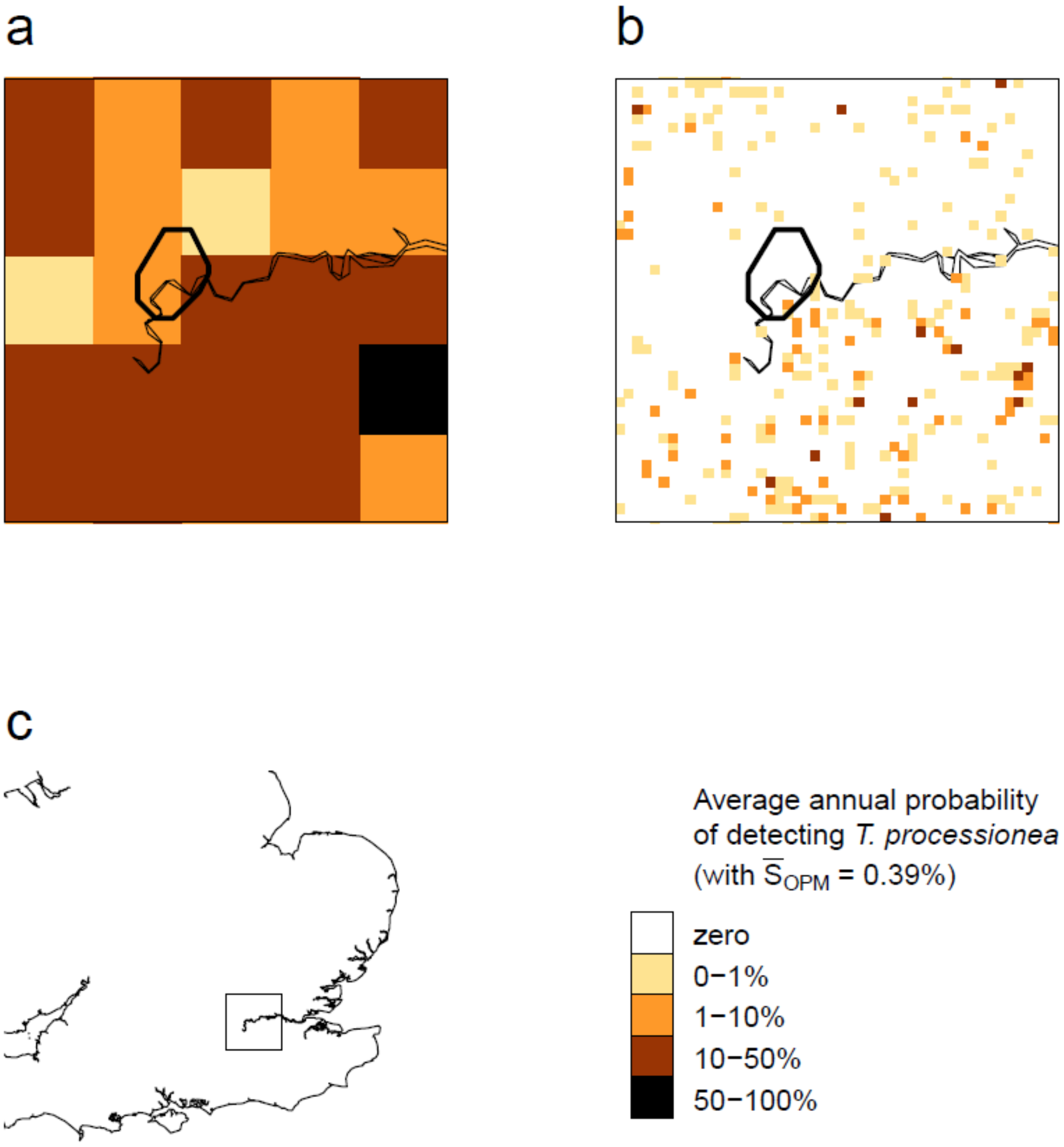

Figure 4. Summary of the Great Britain (GB) Alert system for early detection of invasive alien species. (1) After a suspected observation is submitted via a website, smartphone app or email, (2) an automatic alert allows a data checker to (3) initially review the record and (4) update the database if it is incorrect. Otherwise, suspect records are (5) submitted for rapid verification by a species expert and, if verified as correct, (6) stakeholders are alerted to take appropriate action.

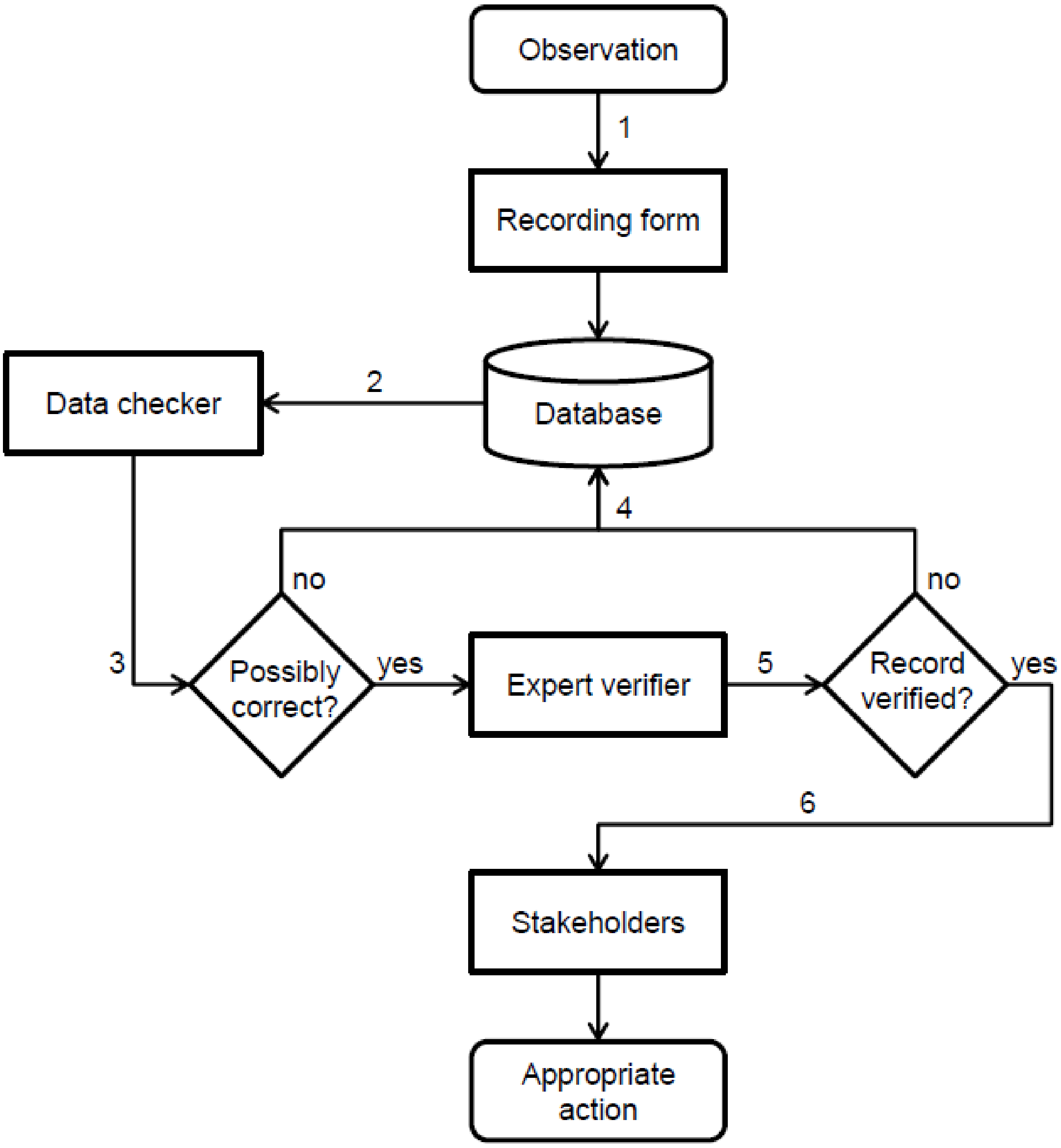

Fig. 4.