Baker, E.R., Jeger, M.J., Mumford, J.D., Brown, N. (2019) Enhancing plant biosecurity with citizen science monitoring: comparing methodologies using reports of Acute Oak Decline. *J Geographical Systems*, 21:111–131. https://doi.org/10.1007/s10109-018-0285-2

Accepted manuscript

1. Introduction

Plant pests and diseases, both invasive and native, represent a major threat to plant biosecurity globally (Mumford *et al.*, 2017; Mumford, Macarthur and Boonham, 2016; Brasier, 2008), and present policy and management challenges to the international community. These organisms have degraded ecosystems and impacted endemic species (Mumford, Macarthur and Boonham, 2016). Acute Oak Decline (AOD) threatens native oaks in Great Britain (Denman *et al*., 2014). AOD affects both *Quercus robur* and *Q. petrea,* , and causes a rapid decline over 3 to 5 years (Denman *et al*., 2014, Denman *et al*., 2017). The distribution of AOD has been investigated using both designed surveys and reports from members of the public (Brown *et al*., 2017a). Volunteer contributions to science, referred to as citizen science, have increased in recent years and provide a valuable means to combine environmental research with educational programmes (Roy *et al*., 2012). In certain fields, the use of citizen science datasets has increased in research (Dickinson *et al*., 2012; Dickinson *et al*., 2010; Jordan *et al*., 2012) and to inform policy (POST Report 476, 2014). In the field of plant biosecurity there are some examples of the use of citizen science in research (Bates *et* al., 2015; Brown *et al*., 2017a Meentemeyer *et al*., 2015; Pocock & Evans, 2014), however there is potential for the greater use of citizen science, particularly when generating policy and management responses to plant pests and diseases. The use of citizen science in these areas could be limited by potential bias, inaccuracy and incompleteness within datasets (Pescott *et al*., 2015). Therefore, it is important to increase understanding of biases within citizen science datasets and improve volunteer monitoring programs. Furthermore, there is a need to develop ways to make use of such data with appropriate statistical designs, and also to influence participants in improving geographic coverage.

1.1 The Current Role of Citizen Science in Plant Pest and Disease Surveys

Citizen science initiatives within biological recording programmes have increased over the past decade (Roy *et al.*, 2012), and have risen in prominence (Blaney *et al.*, 2016). Contributions by citizen scientists can come from structured population monitoring (Kamp *et al*., 2016). Casual observations also feed into these biological records (Kamp *et al*., 2016), as well as data from crowdsourcing campaigns (Sullivan *et al*., 2009; Sullivan *et al*., 2014). This can create sources of observer bias (Kamp *et al*., 2016) such as spatial bias towards areas with a higher population density, conservation areas and areas with higher biodiversity and endangered species (Kamp *et al*., 2016). A further limitation of citizen science data is that the data often only indicates presence of observed species, and an absence of records does not necessarily mean an absence of a species (Pocock *et al*., 2017).

Citizen science datasets from sources such as National Biodiversity Network (NBN) Atlas have been used to effectively map and understand species distributions (Pocock *et al.,* 2015; Roy *et al.,* 2015; van der Wal *et al*., 2015) and can play a role in the early detection of invasive species, which is valuable when generating policy and

management responses (Pocock *et al*., 2017). Volunteer involvement increases awareness and knowledge, engaging the public as stakeholders to communicate environmental policies (Roy *et al*., 2012; Devictor, Whittaker and Beltrame, 2010). There are a range of initiatives, projects and resources available to promote the role of citizen science to monitor tree pests in the UK. These include the Open-Air Laboratories (OPAL) (OPAL, 2016) and Observatree (2017). The data and information from which is available online via Opal Data Explorer (OPAL, 2016) and UK Plant Health Portal (Observatree, 2017).

The ubiquity of smartphone technologies mean that citizen science programs are more accessible to the wider population (Dehnen-Schmutz *et al*., 2016), as well as allow for timely reporting of results (Newman *et al*., 2012). Also, such applications and technologies could notify users of new and emerging threats and have built in validation or verification of records, thus increasing the quality of data and reducing the processing time for new records (Newman *et al*., 2012).

1.2 Acute Oak Decline and Agrilus biguttatus

AOD is a novel and distinctive form of oak decline which has increased in occurrence over recent decades in the UK (Denman and Webber, 2009; Denman *et al*., 2014). Affected oak trees are distributed across much of southern England, from the East Anglian coast as far west as the Welsh borders (Brown 2017a and 2018). AOD affected trees can be identified by bleedings from oak stems (Denman *et al*., 2014; Denman and Webber, 2009) associated with necrogenic bacteria (Denman *et al.*, 2016, Denman *et al*., 2017). A further internal symptom of AOD is galleries of *Agrilus biguttatus* (Denman *et al.*, 2014; Brown *et al*., 2014). *A. biguttatus* has been considered a factor that has contributed to oak decline within Europe (Moraal and Hilszczanski, 2000; Vansteenkiste *et al*., 2004). In the UK, however, it has previously been considered a vulnerable species that inhabits dying or recently dead oak trees (Brown *et al*, 2014).

1.3 Aims and hypotheses

This study aims to compare the results of a structured survey for AOD with two citizen science approaches. The latter two datasets are both unstructured. One dataset collated by the National Biodiversity Network contained sightings of *A. biguttatus*. The second collected by Tree Alert contains reports of AOD occurrence. Both tree alert and NBN data contain only positive detections, but by resampling the negative locations from the designed survey we will test for differences between the methodologies in factors that may affect reporter effort. We hypothesize that the locations of positive detections in the designed survey will not be influenced by spatial trends in factors that may bias reports and that the two citizen science datasets will contain correlations with these factors.

2. Methodology

2.1 Program Background

In order to examine reporter biases in citizen science data for AOD, a range of geographical co-factors that may explain biases in the distribution of data were compared between two citizen science datasets and a structured

tree health survey. The citizen science datasets were the NBN atlas records for *A. biguttatus* and Tree Alert records for AOD. The NBN atlas is the UK's largest repository of biodiversity information, holding 217 million records for species across the UK from 160 different partners, including regional wildlife trusts, biological record centres and species recording websites such as iRecord (National Biodiversity Network, 2016). The reports are often *ad hoc* and opportunistic, and due to the range of data sources the level of expertise of recorders is generally unknown. Although NBN Atlas does provide resources to assist in species identification, and encourage species recording (National Biodiversity Network, 2016), they do not spearhead specific campaigns to encourage data submissions. Conversely, Tree Alert reports have been recorded following efforts to raise awareness and encourage the public and especially forest managers and private landowners to specifically look for tree pests and disease (Forestry Commission, 2017). This process involved Forest Health meetings, dedicated web pages and a field identification guide. Volunteer data was compared with from the results of a structured AOD survey for a study by Brown *et al*. (2017a). The Brown *et al*. (2017a) dataset included both positive and negative survey locations and could be used as a control allowing for the two citizen science datasets to be systematically compared, for the biases to be assessed, and the factors that affected the probability of detection could be identified.

2.2 Datasets

The NBN dataset records species presence using a range of locational accuracies (dependent on the individual recorder and project contributing observations). In order to use as many of the available reports as possible in subsequent analyses all data were aggregated into 2 km x 2 km squares. A total of 414 records for *A. biguttatus* were extracted from NBN Atlas. For records whose GPS coordinates had an uncertainty of more than 2 km x 2 km, the respective recording bodies were contacted to see if accuracy could be improved. Records that could not be verified to an accuracy of 2 km or less were removed, leaving 342 records (that fell within 185 squares on a 2 km x 2 km grid).

Tree Alert contained 231 AOD records that fell within 199 2-km grid squares, and the AOD survey data consisted of 40 records that fell within 38 2-km squares. The negative survey locations fell within 135 grid squares. The spatial extent of the three positive detection datasets with positive detections is shown in Figure 1.

For all 2 km x 2 km squares, existing GIS datasets were used to collect information regarding land type, Coleopteran abundance and population density. The data were collected from a range of sources and represent a range of co-factors that are described in Table 1. This then allowed for the co-factors that influence the probability of *A. biguttatus* being detected to be compared between datasets. Some of the co-factor data had a non-normal distribution and so it was transformed prior to analysis (Table 1). In addition to factors that may cause bias in the data sets all analyses also included a temperature variable, day degrees above 11.5°C (Reed *et al*., 2017; Brown *et al*., 2018), to control for variations in the underlying abundance of beetles.

2.3 Analysis

Three separate General Linear Models (GLMs) were run for each of the three datasets, NBN, Tree Alert and structured survey. Each of the GLMs tested for variation of each co-factor between the positive detection

datasets and the negative detection survey locations, but did not test for interactions between co-factors. The initial model was run using a maximal model that included all the co-factors. Once the initial model had been run it was modified by removing the least significant co-factor, and then re-run with the remaining co-factors. Following this, model comparison was made using a chi squared test on the deviance. This process was repeated until the ANOVA results showed significant differences between the two models being compared. Factors were retained in the model only if there was a significant difference in the model's explanatory power. This process of analysis simplified the model by removing non-significant co-factors, and provided three minimum adequate models: one for each dataset. These models only contained the co-factors that were significant in explaining the differences between each dataset and the negative survey locations.

3. Results

The analysis was based on the assumption that there were potential biases to be found in the citizen science datasets. Furthermore, the GLM analysis assumed that there was no co-linearity between co-factors. This assumption was based on a scatterplot matrix that was created as part of the preliminary analysis to see any linear relationships between co-factors. There were still weak linear relationships between ancient woodland and National Forest Inventory (NFI) areas, as well as NBN beetle records and saproxylic beetle records, but this is to be expected as the NFI area included ancient woodland, and NBN beetle records included saproxylic beetle records. A further limitation of this analysis was the use of the same dataset of negative AOD survey locations in all three GLMs. This was unavoidable, as the two citizen science datasets were presence only datasets.

3.1 Preliminary data assessment

For the Coleopteran records, the mean for the NBN dataset was much higher than for the other datasets (see Figure 2a). The mean number of coleopteran records within each 2-km square for NBN records was 279.3 (*sd=614.8, N=185*), followed by the structured survey dataset at 92.3 (*sd=231.5, N=38*) and the Tree Alert dataset at 47.1 (*sd=119, N=199*). The negative AOD survey squares had a mean number of records of 34.7 (*sd=136.2, N=135*). Similarly, the mean number of saproxylic records for the NBN dataset was greater than for the rest of the datasets (see Figure 2b). The mean number of saproxylic beetle records for the NBN dataset was 23.1 (*sd=52.4, N=185*), followed by negative AOD survey control squares at 3.1 (*sd=17.6, N=135*), then the structured survey data at 2.4 (*sd=4.7, N=38*), and the Tree Alert dataset at 1.4 (*sd=4.3, N=199*).

There was less difference in the mean National Forest Inventory area between datasets (see Figure 2c), however the structured survey dataset had the highest mean area overall at 134 ha (*sd=91.7, N=38*), followed by the NBN dataset at 98 ha (*sd=75.9, N=185*), the negative AOD survey control sites at 93.4 ha (*sd=67.5, N=135*), and the Tree Alert dataset at 73.1 ha (*sd=73.1, N=199*). The mean area of ancient woodland was highest for the structured survey dataset (see Figure 2d) at 64.2 ha (*sd=78.5, N=38*), followed by the negative survey locations at 46.3 ha (*sd=54.3, N=135*), the NBN dataset at 34.6 ha (*sd=49.7, N=185*) and then the Tree Alert at 30.6 ha (*sd=57, N=199*). For the area of parks and gardens, the mean was greatest for the NBN (see Figure 2e) at 51.7 ha (*sd=89.2, N=185*), followed by the Tree Alert dataset at 17.3 ha (*sd=45.9, N=199*), the structured survey at

15.5 ha (*sd=37.8, N=38*) and the negative controls at 12.3 ha (*sd=38.7, N=135*). The average area of open access land within each square was highest for the NBN dataset (see Figure 2f) at 61.5 ha (*sd=85.2, N=185*), followed by the structured survey dataset at 59.1 ha (*sd=83.4, N=38*), the negative control dataset at 48.3 ha (*sd=76, N=135*), and the Tree Alert dataset at 30.7 ha (*sd=71.7, N=199*).

The average number of residents per hectare within each 2-km square was greatest for the NBN dataset (see Figure 2g) at 14.7 (*sd=23.2, N=185*), followed by the Tree Alert dataset at 4.8 (*sd=7.9, N=199*), the negative control sites at 3.5 (*sd=6.1, N=135*), and the structured survey dataset at 3 (*sd=4.04, N=38*). For the average day degrees, the greatest was in the NBN (see Figure 2h) at 552.1 (*sd=58.6, N=185*), followed by the Tree Alert dataset at 529.3 (*sd=52.4, N=199*), the structured survey dataset at 522.5 (*sd=58.6, N=38*), and the negative control at 450.5 (*sd=90.3, N=135*).

3.2 Generalized Linear Model Results

The initial generalized linear model run using the NBN records data found that NFI area was the least significant factor in the model. This was followed by ancient woodland area, saproxylic beetle records, open access land area and then human population density. Model 6 was the minimum adequate model for these data (see Table 2), as once parks and gardens were removed from the model, it was found to be significantly different from the previous general linear model (see Table 3). The final minimum adequate model included the following co-factors: Coleopteran records, parks and gardens area, and day degrees.

For the Tree Alert data, the least significant factor in the initial model was NFI area, followed by ancient woodland area, human population density, saproxylic beetle NBN records, and then parks and gardens area. Model 6 was found to be the minimum adequate model (see Table 2), as once open access land area was removed the models were significantly different (see Table 3). The final model included the following co-factors: Coleopteran records, open access land area and day degrees.

For the structured survey data, the least significant factor in the initial model was found to be ancient woodland area. This was followed by NBN saproxylic beetle records, open access land area, NBN Coleopteran records, human population density, and parks and gardens area. Model seven was found to be the minimum adequate model, which included the co-factors NFI area and day degree accumulation. The Akaike Information Criterion (AIC) for model 7 was marginally higher than for model 6, however the models were not found to be significantly different (see Table 3). See Table 2 for the co-factors included in each model that was run.

In the minimum adequate model for the NBN dataset, the most significant factors were NBN Coleopteran records and day degree accumulation, followed by parks and gardens designation. NBN Coleopteran records and day degrees were found to be positively associated with the number of NBN *A. biguttatus* records. The minimum adequate model for the Tree Alert dataset showed that NBN Coleopteran records and day degrees were the most significant factors, followed by open access land area. NBN Coleopteran records and day degrees were positively associated with the Tree Alert AOD records. The model found open access land area to have a negative

association with Tree Alert records of AOD. The minimum adequate model for the structured survey dataset found that day degrees was the most significant co-factor, followed by NFI area. Both these factors had a positive association with the occurrence of AOD in the structured survey records. See Table 4 for the full description of each of the minimum adequate models.

4. Discussion

The aims of this study were to investigate the role citizen science can play in surveying plant pests and diseases, to assist in understanding and managing threats to plant biosecurity in the UK. *Agrilus biguttatus* and AOD were used as a case study to explore factors that may affect reporter bias. The results highlight some of the challenges of using citizen science data in pest and disease surveys, and show how data collection method influence biases. Furthermore, identifying such biases allows for appropriate statistical approaches to be used in analysis of existing citizen science datasets to account for such limitations in data.

4.1 Biases in data

The results from both datasets indicate that there are hotspots of recorder activity, as they correlate with the wider abundance of Coleptera records (although this pattern is most pronounced in the NBN data). This is expected as opportunistic data recording is often characterised by uneven sampling effort (Isaac *et al.*, 2014). This could be affected by many underlying mechanisms, for example: visiting areas where there is a high abundance of Coleoptera; areas that are easily accessible; or areas where the public are already aware of high beetle abundance (Isaac and Pocock, 2015). This pattern of recorder behaviour has been found in other NBN species datasets. For example, NBN datasets for two bee species, *Bombus pascuorum* and *Bombus hypnorum*, were compared with BeeWatch, a citizen science program encouraging members of the public to be involved in recording bee species. It was found that the NBN data was patchy and clustered, both temporally and spatially, which reflected when record centres were active (Van der Wal *et al.*, 2015). In the NBN dataset, *A. biguttatus* was more likely to be found on or near areas that were registered historic parks and gardens. Vegetation structure may also influence the ease of detection for *A. biguttatus* (Isaac and Pocock, 2015), for example, ancient woodland may be a more challenging environment than parkland due to the denser vegetation. Vegetation structure has been shown to influence the detectability of other invertebrates such as butterflies (Isaac *et al.*, 2011). A further factor influencing the probability of *A. biguttatus* being detected in parks and gardens is accessibility of sites (Mair and Ruete, 2016).

Analysis of the Tree Alert dataset found that AOD (and therefore *A. biguttatus*) was less likely to be found on or near areas of open access land. This contrasts with the Coleopteran hotspots discussed above, and perhaps indicates that ease of access is more complicated than having permission to visit a site. However, it should not be overlooked that the Tree Alert dataset may have its own bias and be missing records from open access land.

4.2 The Challenges of Using Citizen Science Data in Plant Pest and Disease Surveys

Effective management and control of pests and disease are reliant on efficient and timely responses (Tomlinson, Potter and Bayliss, 2015). However, without the best available data on the incidence of a given pest or disease, the response may not be appropriate or effective in controlling the spread due to a lack of understanding of the distribution, virulence or potential impacts of the pest or disease (Potter *et al.*, 2011). To use citizen science in this context reports must be validated and unreliable data removed (Freitag and Pfeffer, 2013). The reliability and credibility of citizen science datasets is reduced through bias and misidentification (Bonter and Cooper, 2012; Crall *et al*., 2011), an issue that is most pronounced when pests and diseases are more cryptic (Crall *et al*., 2011). However, with sufficient training volunteers can achieve accuracies similar to professional surveyors (Meentemeyer et al., 2015).

4.3 How Existing Citizen Science Data Can Be Used

There is a substantial amount of citizen science data that has already been collected and stored across various platforms (Roy *et al.*, 2012). Although some of this data is biased or incomplete, these existing datasets provide important baseline data, and have the potential to be used to answer a diverse range of research questions (Dickinson *et al*., 2012). For example, existing citizen science projects and datasets may be suited to answering large-scale broad hypotheses (Dickinson *et al*., 2012) or explore and scope hypotheses for targeted research (Bates *et al*., 2015). This way, the citizen science data can potentially be integrated into the design of surveys, which was the case for Brown *et al*. (2017a). This approach could be applied to a range of citizen science datasets for differing pests and diseases, improving coverage of surveying effort and providing a structure to assess any potential biases.

For some species, including *A. biguttatus*, there are multiple datasets from different citizen science programs. One way to have a more complete dataset, or a more comprehensive understanding of the distribution of pests and diseases, would be to merge these datasets (Van der Wal *et al*., 2015). To overcome the issue of unreliable records, data could be weighted depending on perceived accuracy (Bates *et al*., 2015). To achieve this there is a need to establish a way of estimating the degree of unreliability by assessing individual reporter competency. There are statistical and modelling approaches that can be used across a range of contexts to account for error and biases in large datasets, ensuring that citizen science datasets are used to their full potential (Bird *et al*., 2014). In order to identify the correct statistical approach, an understanding of the sampling constraints, survey design and biases is required (Bird *et al*., 2014). Such an understanding can be developed through communication between volunteers and statisticians (Bird *et al*., 2014), however with data repositories, such as NBN Atlas, where this this is not feasible due to anonymity of recorders, approaches such as the methodology used in this study can be applied to assist in identifying and characterising biases in datasets.

4.4 Developing New Citizen Science Programs to Survey Plant Pests and Diseases

Often existing citizen science data can only address a limited set of questions, due to the spatial extent or nature in which the data was collected, but these limits can be overcome with suitable survey design (Tulloch *et al*., 2013a; Wiggins & Crowston, 2015). For example, encouraging volunteers to report sites they visit that are

disease free would provide additional information on the presence / absence of outbreaks and enable bias in reporter activity to be better quantified (Brotons *et al*., 2004). Horizon scanning could be used to foresee potential future threats to plant biosecurity, thus allowing scientists and policymakers to focus citizen science projects (Parrott *et al.*, 2009).

The challenges of analysing citizen science data could be overcome by structuring programs to standardise sampling effort, both spatially and temporally (Dickinson, Zuckerberg and Bonter, 2010). Standardising sampling effort temporally could be straightforward to implement, by advising volunteers on how much time, or what time they should be surveying (Dickinson, Zuckeberg and Bonter, 2010). Standardising sampling effort spatially may be more challenging, particularly if the sites of interest are in areas with a low population density (Tulloch *et al*., 2013b). Structured citizen science surveys have been shown to be more effective in detecting trends in populations (Kamp *et al*., 2016), and recording effort is more even (Pocock *et al*., 2014). However, a structured survey places additional demands on citizen scientists as reports are no longer at times and place that are most convenient to them. As such, data quality must be balanced with volunteer engagement, especially as disengagement can create its own biases and decrease data quality (Turnhout, Lawrence and Turnhout, 2016). By integrating citizen science efforts with official surveys more challenging environments and locations could be assessed by professionals to limit the potential for volunteer disengagement (Tulloch *et al*., 2013b; Brown *et al*., 2017a).

4.5 The Future of Citizen Science and Plant Biosecurity

Plant biosecurity is a persistent, widespread and chronic threat to the UK and other countries. With increased movement and global connectivity, the threats to plant health are accelerating, and there is a need to increase the rate and scale at which data regarding pest and disease incidence are collected. Citizen science is one of the most logical and effective ways to meet this demand, whilst also providing other benefits such as increased public awareness, engagement with stakeholders and reduced research costs. Although there are large volumes of existing citizen science data, the biases and incompleteness in the existing data mean that these datasets are often overlooked when carrying out research or developing policy and management plans. This study builds on existing literature and examines the biases in differing citizen science datasets for *A. biguttatus* and AOD. The results from the study have contributed to an understanding of bias in citizen science data, and highlights areas for improvement when developing new citizen science programs. There are several ways in which the challenges of working with volunteer data can be overcome and biases in data can be reduced. The future progress of citizen science will not only assist in mitigating the threats to plant health, but will aid in bridging the gap between science and policy in this area, and work towards developing a cohesive community of citizen scientists that are conscious of, and strive to protect, plant biosecurity.

References

Alexander KN. (2004) Revision of the Index of Ecological Continuity as Used for Saproxylic Beetle. Eng. Nat. http://publications.naturalengland.org.uk/publication/133006. Accessed 26 June 2017

281 Bates AJ. Fraser PL. Robinson L. Tweddle JC. Sadler JP. West SE. Norman S. Batson M. Davies L. (2015) The
282 OPAL bugs count survey: exploring the effects of urbanisation and habitat characteristics using citizen science.
283 Urban Ecosyst. 18 (4): 1477-1497. http://doi.org/10.1007/s11252-015-0470-8

284 Bird TJ. Bates AE. Lefcheck JS. Hill NA. Thomson RJ. Edgar GJ. Stuart-Smith RD. Wotherspoon S. Krkosek M.
285 Stuart-Smith JF. Pecl GT. Barrett N. Frusher S. (2014) Statistical solutions for error and bias in global citizen
286 science datasets, Biological Conservation 173:144-154 https://doi.org/10.1016/j.biocon.2013.07.037

287 Blaney R. Jones G. Philippe A. Pocock M. (2016) Citizen Science and Environmental Monitoring: Towards a
288 Methodology for Evaluating Opportunities, Costs and Benefits. Final Report on Behalf of UKEOF.
289 https://www.ukeof.org.uk/documents/guide-to-citizen-science. Accessed 7 June 2017

290 Bonter DN. Cooper CB. (2012) Data validation in citizen science: a case study from Project FeederWatch. Front.
291 in Ecology and the Environ. 10 (6): 305-307. https://doi.org/10.1890/110273

292 Brasier C. (2008) The biosecurity threat to the UK and global environment from international trade in plants.
293 Plant Pathology 57 (5): 792-808. https://doi.org/10.1111/j.1365-3059.2008.01886.x

294 Brotons L. Thuiller W. Araújo MB. Hirzel AH. (2004) Presence-absence versus presence-only modelling methods
295 for predicting bird habitat suitability. Ecography 27 (4): 437-448. https://doi.org/10.1111/j.0906-
296 7590.2004.03764.x

297 Brown N. Inward DJ. Jeger M. Denman S. (2014) A review of Agrilus biguttatus in UK forests and its relationship
298 with acute oak decline. Forestry 8 (1): 53-63. https://doi.org/10.1093/forestry/cpu039

299 Brown N. van den Bosch F. Parnell S. Denman S. (2017a) Integrating regulatory surveys and citizen science to
300 map outbreaks of forest diseases: acute oak decline in England and Wales. Proc. of the R. Society B – Biological
301 Sciences http://doi.org/10.1098/rspb.2017.0547

302 Brown N. Vanguelova E. Parnell S. Broadmeadow S. Denman S. (2018) Predisposition of forests to biotic
303 disturbance: Predicting the distribution of Acute Oak Decline using environmental factors. For. Ecology and
304 Management, 407: 145–154. https://doi.org/10.1016/j.foreco.2017.10.054

305 Crall AW. Newman GJ. Stohlgren TJ. Holfelder KA. Graham J. Waller DM. (2011) Assessing citizen science data
306 quality: an invasive species case study. Conservation Lett. 4 (6):433-442. https://doi.org/10.1111/j.1755-
307 263X.2011.00196.x

308 Dehnen-Schmutz K. Foster GL. Owen L. Persello S. (2016) Exploring the role of smartphone technology for
309 citizen science in agriculture. Agronomy for Sustainable Development. 36(25) https://doi.org/10.1007/s13593-
310 016-0359-9

311 Denman S. Brown N. Kirk S. Jeger M. Webber J. (2014) A description of the symptoms of Acute Oak Decline in
312 Britain and a comparative review on causes of similar disorders on oak in Europe. Forestry 87 (4): 535-551.
313 https://doi.org/10.1093/forestry/cpu010

314 Denman S. Doonan J. Ransom-Jones E. Brober M. Plummer S. Kirk S. Scarlett K. Griffiths AR. Kaczmarek M.
315 Froster J. Peace A. Golyshin PN. Hassard F. Brown N. Kenny JG. McDonald JE. (2017) Microbiome and infectivity
316 studies reveal complex polyspecies tree disease in Acute Oak Decline. The ISME J. 12: 386-399
317 http://doi.org/10.1038/ismej.2017.170

318 Denman S. Plummer S. Kirk S. Peace A. McDonald JE. (2016) Isolation studies reveal a shift in the cultivable
319 microbiome of oak affected with Acute Oak Decline. Systematic and Appl. Microbiology 39 (7): 484-490.
320 https://doi.org/10.1016/j.syapm.2016.07.002

Denman S. Webber J. (2009) Oak declines: new definitions and new episodes in Britain. Q. J. of Forestry 103: 285-290.

Devictor V. Whittaker RJ. Beltrame C. (2010) Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. Diversity and Distributions 16 (3): 354-362. https://doi.org/10.1111/j.1472-4642.2009.00615.x

Dickinson JL. Shirk J. Bonter D. Bonney R. Crain RL. 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 Environ. 10 (6): 291-297. https://doi.org/10.1890/110236

Dickinson JL. Zuckerberg B. Bonter DN. (2010) Citizen science as an ecological research tool: challenges and benefits. Annual Rev of Ecology, Evolution, and Systematics 41: 149-172. https:// doi.org/10.1146/annurev-ecolsys-102209-144636

Forestry Commission. (2017) Tree Alert. Available from: https://www.forestry.gov.uk/treealert. Accessed 7 August 2017.

Freitag A. Pfeffer MJ. (2013) Process, not product: investigating recommendations for improving citizen science "success". PLOS One https://doi.org/10.1371/journal.pone.0064079

Isaac NJ. Cruickshanks KL. Weddle AM. Marcus Rowcliffe J. Brereton TM. Dennis RL. Shuker DM. Thomas CD. (2011) Distance sampling and the challenge of monitoring butterfly populations. Methods in Ecology and Evolution 2 (6): 585-594. https://doi.org/10.1111/j.2041-210X.2011.00109.x

Isaac NJ. Pocock MJ. (2015) Bias and information in biological records. Biological J. of the Linn. Society. 115 (3): 522-531. https://doi.org/10.1111/bij.12532

Isaac NJ. Strien AJ. August TA. Zeeuw MP. Roy DB. (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. Methods in Ecology and Evolution 5 (10): 1052-1060. https://doi.org/10.1111/2041-210X.12254

Jordan RC, Ballard HL. Phillips TB. (2012) Key Issues and new approaches for evaluating citizen science learning outcomes. Frontiers in Ecology and Environment 10 (6): 307-309, https://doi.org/10.1890/110280

Kamp J. Oppel S. Heldbjerg H. Nyegaard T. Donald PF. (2016) Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. Diversity and Distributions 22 (10): 1024-1035. https://doi.org/10.1111/ddi.12463

Mair L. Ruete A. (2016) Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PloS One https://doi.org/10.1371/journal.pone.0147796

Meentemeyer RK. Dorning MA. Vogler JB. Schmidt D. Garbelotto M. (2015) Citizen science helps predict risk of emerging infectious disease. Frontiers in Ecology and the Environment 13 (4):189-194. https://doi.org/10.1890/140299

Moraal L. Hilszczanski J. (2000) The oak buprestid beetle (Agrilus biguttatus) (F.) (Col., buprestidae), a recent factor in oak decline in Europe. Journal of Pest Science 73 (5):134–138. https://doi.org/10.1007/BF02956447

Mumford JD. Gullino ML. Stack JP. Fletcher J. Quinlan MM. (2017) The Need for International Perspectives to Solve Global Biosecurity Challenges. Practical Tools for Plant and Food Biosecurity. Cham: Springer International 363-384. https://doi.org/10.1007/978-3-319-46897-6_18

359 Mumford R. Macarthur R. Boonham N. (2016) The role and challenges of new diagnostic technology in plant
360 biosecurity. Food Security 8 (1): 103-109. https://doi.org/10.1007/s12571-015-0533-y

361 National Biodiversity Network. (2016) NBN Action Plan. Available from: https://nbn.org.uk/about-us/what-we-
362 do/nbn-action-plan/. Accessed 10 May 2017.

363 Newman G. Wiggins A. Crall A. Graham E. Newman S. Crowston K. (2012) The future of citizen science:
364 emerging technologies and shifting paradigms. Frontiers in Ecology and the Environment 10 (6): 298-304.
365 https://doi.org/10.1890/110294

366 Observatree. (2017) Observatree: monitoring tree health. Available from:
367 http://www.observatree.org.uk/about/. Accessed 6 July 2017

368 OPAL. (2016) Available from: https://www.opalexplorenature.org/. Accessed 6 July 2017

369 Parliamentary Office of Science and Technology, Environmental Citizen Science, August 2014, Post Report 476

370 Parrott D. Roy S. Baker R. Cannon R. Eyre D. Hill M. Wagner M. Preston C. Roy H. Beckmann B. (2009) Horizon
371 scanning for new invasive non-native animal species in England. Natural England
372 http://publications.naturalengland.org.uk/publication/43005. Accessed 19th June 2017

373 Pescott OL. Walker KJ. Pocock MJ. Jitlal M. Outhwaite CL. Cheffings CM. Harris F. Roy DB. (2015) Ecological
374 monitoring with citizen science: the design and implementation of schemes for recording plants in Britain and
375 Ireland. Biological Journal of the Linn. Society 115 (3): 505-521. https://doi.org/10.1111/bij.12581

376 Pocock MJ. Evans DM. (2014) The success of the horse-chestnut leaf-miner, Cameraria ohridella, in the UK
377 revealed with hypothesis-led citizen science. PloS One 9. https://doi.org/10.1371/journal.pone.0086226

378 Pocock MJ. Chapman D. Sheppard L. Roy H. (2014) A strategic framework to support the implementation of
379 citizen science for environmental monitoring. Final Report to SEPA. Wallingford, Oxfordshire: Centre for
380 Ecology and Hydrology. https://www.ceh.ac.uk/sites/default/files/hp1114final_5_complete.pdf. Accessed 7th
381 June 2017.

382 Pocock MJ. Roy HE. Fox R. Ellis WN. Botham M. (2017) Citizen science and invasive alien species: Predicting the
383 detection of the oak processionary moth Thaumetopoea processionea by moth recorders. Biological
384 Conservation 208: 146-154. https://doi.org/10.1016/j.biocon.2016.04.010

385 Pocock MJ. Roy HE. Preston CD. Roy DB. (2015) The Biological Records Centre: a pioneer of citizen science.
386 Biological Journal of the Linn. Society 115 (3): 475-493. https://doi.org/10.1111/bij.12548

387 Potter C. Harwood T. Knight J. Tomlinson I. (2011) Learning from history, predicting the future: the UK Dutch
388 elm disease outbreak in relation to contemporary tree disease threats. Philosophical Transactions of the R.
389 Society B 366: 1966-1974. https://doi.org/10.1098/rstb.2010.0395

390 Reed K. Denman S. Leather SR. Forster J. Inward DJG. (2017) The lifecycle of Agrilus biguttatus : the role of
391 temperature in its development and distribution, and implications for Acute Oak Decline. Agricultural and For.
392 Entomology https://doi.org/10.1111/afe.12266

393 Roy H. Pocock M. Preston C. Roy D. Savage J. Tweddle J. Robinson L. (2012) Understanding citizen science and
394 environmental monitoring: final report on behalf of UK Environmental Observation Framework, final Report on
395 behalf of UK-EOF. NERC Centre for Ecology and Hydrology and Natural History Museum
396 http://nora.nerc.ac.uk/20679/. Accessed 6 May 2017.

Roy HE. Rorke SL. Beckmann B. Booy O. Botham MS. Brown PMJ. Harrower C. Nobel D. Sewell J. Walker K. (2015) The contribution of volunteer record to our understanding of biological invasions. Biological Journal of the Linn. Society 115 (3):678-689. https://doi.org/10.1111/bij.12518

Sullivan BL. Aycrigg JL. Barry JH. Bonney RE. Bruns N. Cooper CB. Damoulas T. Dhondt AA. Dietterich T. Farnsworth A. Fink D. Fitzpatrick JW. Fredericks T. Gerbracht J. Gomes C. Hochachka WM. Iliff M. Lagoze C. La Sorte FA Merrifield M. Morris W. Phillips TB. Reynolds M. Rodewald AD. Rosenberg KV. Trautmann NM. Wiggins A. Winkler DW. Wong W. Wood CL. Yu J. Kelling S. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation 169: 31-40, https://doi.org/10.1016/j.biocon.2013.11.003

Sullivan BL. Wood CL. Iliff MJ. Bonney RE. Fink D. Kelling S. (2009), eBird: A citizen-based bird observation network in the biological sciences. Biological Conservation 142 (10):2282-2292, https://doi.org/10.1016/j.biocon.2009.05.006

Tomlinson I. Potter C. Bayliss H. (2015) Managing tree pests and diseases in urban settings: The case of Oak Processionary Moth in London, 2006-2012. Urban Forestry and Urban Green 14 (2): 286-292. https://doi.org/10.1016/j.ufug.2015.02.009

Tulloch AI. Mustin K. Possingham HP. Szabo JK. Wilson KA. (2013b) To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale. Diversity and Distributions 19 (4): 465-480. https://doi:10.1111/j.1472-4642.2012.00947.x

Tulloch AI. Possingham HP. Joseph LN. Szabo J. Martin TG. (2013a) Realising the full potential of citizen science monitoring programs. Biological Conservation 165: 128-138. https://doi.org/10.1016/j.biocon.2013.05.025

Turnhout E. Lawrence A. Turnhout S. (2016) Citizen science networks in natural history and the collective validation of biodiversity data. Conservation Biology 30 (3): 532-539. https://doi.org/10.1111/cobi.12696

Van der Wal R. Anderson H. Robinson A. Sharma N. Mellish C. Roberts S. Darvill B. Siddharthan A. (2015) Mapping species distributions: A comparison of skilled naturalist and lay citizen science recording. Ambio 44 (4): 584-600. https://doi.org/10.1007/s13280-015-0709-x

Vansteenkiste D. Tirry L. Van Acker J. Stevens M. 2004. Predispositions and symptoms of Agrilus borer attack in declining oak trees. Annals of Forest Science 61 (8):815–823. https://doi.org/10.1051/forest:2004076

Wiggins A. Crowston K. 2015. Surveying the citizen science landscape. First Monday. 20 (1-5) http://dx.doi.org/10.5210/fm.v20i1.5520.

Table 1 – Description, process of collection and analysis of each of the co-factors used in the analysis.

Co-factor	Description	Collection	Analysis	Data Sources
NBN coleopteran records	This data provided a count of the total number of NBN coleopteran records for each cell.	Records were downloaded from NBN atlas, those which had a 10km uncertainty were removed and the remaining records were spatially joined	Due to the non-normal distribution of the data, prior to analysis this data was log transformed.	NBN Atlas occurrence download at http://nbnatlas.org. Accessed 3rd June 2017.

		with the 2km cells, which provided a count of the number of records within each cell.		
NBN saproxylic beetle records	This data provided a count of the total number of records for a sample of 180 saproxylic beetles within each cell. This co-factor was included as *A. biguttatus* is a saproxylic beetle, which are defined as invertebrates that are dependent on decaying or dead wood (Alexander, 2004).	NBN records of a sample of 180 saproxylic beetles. This sample was selected using a report by Alexander (2004) examining the Index of Ecological Continuity for saproxylic beetles. The 180 beetles were selected from the saproxylic beetles that were used to calculate the index of ecological continuity. The NBN records for these species were then obtained, 10km uncertainty records were removed and then spatial join was used to count the records within each cell.	Due to the non-normal distribution of the data, this was log transformed.	NBN Atlas occurrence download at http://nbnatlas.org. Accessed 20th June 2017 The distribution map is shown in online resource 1
National Forest Inventory (NFI) area	Extensive records of woodland and trees for each cell. This includes public,	Obtained from the Forestry Commission. Geoprocessing was used to calculate the	The area was calculated using ArcGIS and converted to hectares before analysis.	© Forestry commission, all rights reserved

	limited access and private land.	total area of forest within each cell.		
Ancient Woodland area	The area of ancient woodland within each cell, this includes public, limited access and private land.	Ancient woodland inventory for England was obtained from the government open data website and published by Natural England. The ancient woodland inventory for wales was obtained from Natural Resources Wales. Geoprocessing was used to calculate the total area of forest within each cell.	The total area within each 2km cell was calculated and converted to hectares before analysis.	© Woodland Trust, all right reserved
Parks and Gardens	The area of approximately 1600 sites including grounds, gardens and various planned open spaces. The dataset was obtained from a list of registers parks and gardens that are of historic importance in England. These are often limited access sites.	This data was obtained from Historic England's open datasets. Geoprocessing tools were then used to calculate the total area within each cell.	The total area within each 2km cell was calculated and converted to hectares before analysis. Due to the large number of zero values, this data was analysed using a presence/absence record and transformed to a factor prior to analysis.	© Historic England, all rights reserved

Open Access Land area	This layer combined a variety of layers that represent a range of open access areas across the UK. This layer represents the combined area of: Local Nature Reserves, National Nature Reserves, Countryside and Rights of Way access land, National trust Open Land, RSPB reserves, Woodland trust sites, National trust limited access, Local Nature reserves Wales, country parks, and forestry commission sites. Each of these layers were combined and geoprocessing tools were used to remove any overlapping areas and merge them to form a single layer from which the area was calculated for each 2km cell.	National Nature Reserves, Local Nature Reserves, Countryside and Rights of Way access land, country parks and National Trust land were obtained from the UK Government Open Data site. Woodland Trust data was obtained from the Woodland Trust Inventory. RSPB data was obtained from the RSPB website. Local Nature Reserves Wales was obtained from the geo-portal for Wales. Forestry commission sites were obtained from the Forestry Commission.	Due to the non-normal distribution, this data was log transformed.	© Office for National Statistics, all rights reserved © Woodland Trust, all right reserved © RSPB, all rights reserved © National Trust, all rights reserved © Natural Resource Wales, all rights reserved © Forestry commission, all rights reserved © Tunbridge Wells Borough Council, all rights reserved © Natural England, all rights reserved
Population density	This data describes the residents and people per hectare	This data was obtained from the 2011 census data on	Due to the skewed data, these values	© Office for National Statistics, all rights reserved

	of the county that each 2km cell falls within.	the government open data website. The data describes the people per hectare within the county that each 2km cell falls into.	were log transformed.	
Day degrees above 11.5°C	This data describes the day degrees at each recording location above 11.5°C.	This value is calculated using the amount of time and number of degrees that each location is above the threshold of 11.5°C. The average maximum and minimum monthly temperatures from the Met Office (1971-2000) in CLIMEX. The temperature of 11.5°C is an estimate of the development thresholds for *A. biguttatus* (Reed *et al*., 2017; Brown *et al*., 2018).	The raw data for this co-factor was used for the analysis.	© Met Office, all rights reserved

427 **Table 2** – Co-factors included and AIC in each GLM that was run for each dataset. The minimum adequate
428 model for each dataset is highlighted.

Model		Co-factor								AIC
		NBN Coleoptera records	NBN Saproxylic records	NFI	Ancient woodland	Parks and gardens	Open access land	Population density	Day degrees above 11.5°C	
NBN	1	x	x	x	x	x	x	x	x	254.25
	2	x	x		x	x	x	x	x	252.25

	3	x	x			x	x	x	x	250.58
	4	x				x	x	x	x	249.97
	5	x				x		x	x	248.83
	6	x				x			x	247.89
	7	x							x	253.27
Tree alert	1	x	x	x	x	x	x	x	x	358.96
	2	x	x		x	x	x	x	x	356.96
	3	x	x			x	x	x	x	354.97
	4	x	x			x	x		x	353.12
	5	x				x	x		x	351.37
	6	x					x		x	350.55
	7	x							x	355.86
Survey	1	x	x	x	x	x	x	x	x	156.91
	2	x	x	x		x	x	x	x	154.92
	3	x		x		x	x	x	x	154.29
	4	x		x		x		x	x	154.09
	5			x		x		x	x	153.56
	6			x		x			x	153.84
	7			x					x	155.64
	8								x	158.62

429 **Table 3** – Results of from the comparing each GLM using an ANOVA.

	Models compared	Deviance	Degrees of Freedom	P
NBN	1 and2	0.0031	1	>0.05
	2 and 3	0.3349	1	>0.05
	3 and 4	1.3821	1	>0.05
	4 and 5	0.861	1	>0.05
	5 and 6	1.0637	1	>0.05
	6 and 7	7.3743	1	<0.01**
Tree Alert	1 and 2	0.0079	1	>0.05
	2 and 3	0.0027	1	>0.05
	3 and 4	0.1572	1	>0.05
	4 and 5	0.2243	1	>0.05

	5 and 6	1.1841	1	>0.05
	6 and 7	7.3085	1	<0.01**
Survey	1 and 2	0.0079	1	>0.05
	2 and 3	1.367	1	>0.05
	3 and 4	1.8062	1	>0.05
	4 and 5	1.4672	1	>0.05
	5 and 6	2.2805	1	>0.05
	6 and 7	3.8001	1	>0.05
	7 and 8	4.9752	1	<0.05*

430 **Table 4** – Co-factors included in each of the minimum adequate models for each dataset.

	Factor	Slope	Mean	Deviance	Degrees of Freedom	p
NBN	NBN Coleopteran records	0.6359 (0.0998)		322.52	1	<0.001***
	Parks and Gardens		1.0093 (0.377)	314.06	1	<0.01**
	Day degrees above 11.5°C	0.0178 (0.0026)		239.89	1	<0.001***
Tree alert	NBN Coleopteran records	0.3199 (0.0851)		439.6	1	<0.001***
	Open access land area	-0.1935 (0.0722)		420.63	1	<0.01**
	Day degrees above 11.5°C	0.01713 (0.0024)		342.55	1	<0.001***
Survey	NFI	0.0056 (0.0024)		172.13	1	<0.01**
	Day degrees above 11.5°C	0.0134 (0.0031)		149.64	1	<0.001***

431