Passive citizen science: the role of social media in wildlife observations

Thomas Edwards[1]◑*, Christopher B. Jones[1]◑, Sarah E. Perkins[2], Padraig Corcoran[1]◑

1 School Of Computer Science and Informatics, Cardiff University, Cardiff, UK **2** School of Biosciences, Cardiff University, Cardiff, CF10 3AX

◑These authors contributed equally to this work.
* EdwardsTJ1@cardiff.ac.uk

Abstract

Citizen science plays an important role in observing the natural environment. While conventional citizen science consists of organized campaigns to observe a particular phenomenon or species there are also many ad hoc observations of the environment in social media. These data constitute a valuable resource for 'passive citizen science' - the use of social media that are unconnected to any particular citizen science program, but represent an untapped dataset of ecological value. We explore the value of passive citizen science, by evaluating species distributions using the photo sharing site Flickr. The data are evaluated relative to those submitted to the National Biodiversity Network (NBN) Atlas, the largest collection of species distribution data in the UK. Our study focuses on the 1500 best represented species on NBN, and common invasive species within UK, and compares the spatial and temporal distribution with NBN data. We also introduce an innovative image verification technique that uses the Google Cloud Vision API in combination with species taxonomic data to determine the likelihood that a mention of a species on Flickr represents a given species. The spatial and temporal analyses for our case studies suggest that the Flickr dataset best reflects the NBN dataset when considering a purely spatial distribution with no time constraints. The best represented species on Flickr in comparison to NBN are diurnal garden birds as around 70% of the Flickr posts for them are valid observations relative to the NBN. Passive citizen science could offer a rich source of observation data for certain taxonomic groups, and/or as a repository for dedicated projects. Our novel method of validating Flickr records is suited to verifying more extensive collections, including less well-known species, and when used in combination with citizen science projects could offer a platform for accurate identification of species and their location.

Introduction

Observations on the distribution of wildlife species have always formed a crucial part of conservation and species management ([1,2]), but are increasingly important in the face of rapid ecosystem changes that can be brought about, for example, by climate change and invasive species, the consequences of which have implications for disease emergence and spread, as well as food security [2]. High-quality species distribution data are typically collected by professionals, but such data can be time-consuming, and expensive to gather, and hence often lack broad coverage [1]. To overcome this knowledge gap, especially over a large spatial and/or temporal scale citizen scientists are often engaged; members of the public who volunteer to record the presence of a given species and associated metadata, such as time, date, and location ([2–4]). Such projects can effectively crowdsource data, so amassing large volumes of species distribution data [4]. Due to the fact of using non-professionals, however, projects frequently come under criticism in terms of the accuracy of species identification, and associated data [5].

Social network websites such as Flickr, Twitter, and Facebook have built a network of more than 2 billion users worldwide, generating millions of messages daily that are easily accessible, and reflect the observed reality of a quarter of the human population [6]. Social media websites have therefore emerged as an informal real-time information source that can contribute to the detection of trends and early warnings in critical fields such as ecological change, environmental problems, and shifts in ecosystems ([6–8]).

A quantitative review of the application of social media in environmental research, conducted by [9] suggests a very rapid growth in the field of environmental monitoring, with Twitter and Flickr being most frequently used as data sources. Among the identified strengths of social media are the large volume of available data samples which makes data collection a less labour-intensive, time-consuming and costly procedure [9–11]. Social media data allows for a timely and (near) real-time monitoring and analysis of species distribution ([9, 12, 13]). Despite the potential of social media to be used for species distribution models there are still some concerns about the quality and reliability of information mined from social media ([6, 9, 14]). There are also concerns about the data ownership and future availability of social network data ([6, 9, 15]).

Arguably, the geotags given to photos on social media sites can be more reliable than user-submitted data as they are assigned automatically by GPS location systems, and if automatic identification of species can be employed such an approach has the potential to outweigh the skills of the general populace. It is for these reasons that the photo sharing site, Flickr, has been recognised as a particularly valuable resource in ecology that could contribute to species distribution models ([2, 6, 16]).

The use of internet sources for gathering wildlife-related data in citizen science initiatives has emerged in recent years ([17], [18], [19]). An example includes urban residents reporting occurrences of tagged birds through a Facebook group, a smartphone application and email [17]. A crowdsourcing tool was employed in [18] to collect data for the creation of a land cover map, while in [19] crowdsourcing is used as a supplemental method for collecting hydrologic data. An overview of the impact of internet social networks on traditional biodiversity data collection methods in [7] is optimistic and concludes that social media can potentially play an important role in conservation science.

Here, we focus specifically on the potential of Flickr for collecting species distribution data as it hosts one of the most extensive and easily accessible collections of geo-referenced photos of its kind and, because it is photo-based, it enables the possibility to validate observations by comparing the asserted species name, as provided in a tag or caption, with the content of the image. We assess the value of species distribution data gathered from the Flickr photo-sharing website relative to existing content on the UK National Biodiversity Network portal. The National Biodiversity Network (NBN) Atlas[1] portal holds the most extensive collection of biodiversity information within the UK with over 220 million species occurrences.

Our aim is to explore the potential of social media to supplement species distribution data, and in doing so to serve as a form of passive citizen science. We conducted analyses with two case studies, one being the 1500 species that were most frequently recorded on NBN and the other being invasive species in the UK that have records on NBN. In comparing species distributions from Flickr with those of the

[1] `https://nbn.org.uk`

National Biodiversity Network Atlas we quantify the value of social media acquired distribution data on the largest number of species considered to date in such studies. Our approach uses a novel method of validating Flickr species images with the Google Cloud Vision API, that extends the method presented in [13] by automatic matching of the assigned categories to the content of a hierarchical species taxonomy.

Research analogous to our own has been carried out previously, but on a much smaller scale, and in a time-consuming manner ([2, 6, 13]). In the forementioned research, the authors evaluate social network sites (Flickr and Twitter) relative to biodiversity data portals in order to identify the potential use of ad-hoc methods for augmenting traditional citizen science data collections. This previous research was conducted on a narrow range of species (between two and four). Another similar research by [8] investigate whether an image classifier for identifying plants could facilitate the discovery of unexploited biodiversity data from Flickr. However, this approach is focused purely on species occurrence on Flickr and thus does not provide a clear evaluation of the role of social sites observations compared to more traditional approaches.

Our validation approach is similar to that used in [13] to verify Flickr data using image content recognition with the Google Reverse Image Search. Their method consists of exact matching between the species names (scientific and common name) and the labels (tags) returned by Google's Reverse Image Search, which can sometimes result in false negatives where the Google API provides only a more generic label or another name for the species. Another disadvantage of this approach is that it requires the manual upload of the images because Google's Reverse Image Search does not provide a programmable interface. In our approach, we deploy Google Cloud Vision API which allows fully automatic image verification. We reduce the incidence of missed matches by employing a species taxonomy that supports matching between alternative names for a species as well as generic matches between terms in the relevant species hierarchy that were not used in the Flickr tags.

In summary, there are several limitations of previous research on using social media to augment traditional biodiversity portals in that the analyses have been performed on very small numbers of species, the methods for accessing the social media are either manual or only partly automated, and the results are limited in the degree of verification.

Materials and Methods

We perform three types of analysis to compare species occurrence between the NBN and Flickr, consisting of a summary statistical analysis and spatial and temporal analyses. The statistical analysis compares the frequency of occurrence of species between the two data collections, performed on different taxonomic levels of species and class. The spatial analysis determines whether Flickr species observations match by location the NBN species observations. Because many species have variable distributions and abundances throughout the year we also use a temporal analysis to compare the time patterns of the NBN and Flickr data collections. We compared the locations of data occurrences for the two data sources for a time span of 3 months, 6 months and 12 months. We verify Flickr species identification through an image content verification approach using the Google Cloud Vision API to identify objects that appear in a given photo. The Google Cloud Vision API labels images with multiple taxonomic categories (i.e. labels) ranging from general to specific. Our image validation approach is based on coarse matching between all species names following down from the class of a species and the tags returned by Google Cloud API. In this way, we avoid a potentially high number of false negatives for less common species that are less likely to be identified on the API at the species level but might be identified at higher taxonomic levels. An outline of the methodology is depicted in Fig 1 and each step is detailed below.

Fig 1. Methodology overview

Data collection

NBN data collection The NBN was selected as the biodiversity data portal for our study because it holds the most extensive collection of biodiversity information within the UK. We collected the names and the number of occurrences for the top 1500 species on NBN using the NBN Atlas Occurrence Facet Search. We performed our search over all collections within the NBN and limited it for the territory of the UK.

For each of the species retrieved from the NBN we obtained, via a search on the NBN, all the alternative names associated with the species (scientific name and common names), the NBN species ID, and its taxonomic classification hierarchy. The names associated with each of the species were used for downloading data from Flickr. The taxonomic classification hierarchy is used for the verification of the Flickr data collection in combination with the Google Cloud Vision API. The NBN service does not perform an exact search and downloaded records can include irrelevant species. Further to that, some records are incomplete, lacking temporal or geo-information. To address this we filtered out irrelevant records and those with missing information. For inclusion in our dataset each record constituted record ID, geo-coordinates of the occurrence, date of the occurrence, NBN species ID.

Flickr data collection Using the Flickr API interface we used both the scientific and common names, and limited our search to geo-tagged posts within the UK. Our search was therefore based on downloading posts with tags matching at least one of the alternative names given for a species in NBN. We downloaded the following types of information from Flickr: image coordinates, date of upload of the post, post ID, the image associated with the post, title, and all the tags associated with the post.

Flickr Data Validation

Flickr images needed to be validated because the content of the photos uploaded with associated tags might not match the species name tags given. We use Google Cloud Vision API to coarse match between all names following down from species taxonomic class and the tags returned by Google Cloud Vision API. When there is an overlap between the Google Cloud API tags and the classification names, we consider the results from Google Cloud API to be correct. Google Cloud Vision API is however not trained on wildlife data and thus some of the less well-known species names might not be returned as tags, for instance, 10-spot Ladybird (*Adalia decempunctata*) and 22-spot Ladybird (*Psyllobora vigintiduopunctata*). Also, species belonging to the same class (e.g. 'cuckoo' and 'sparrowhawk') might have very similar visual appearance and thus the Google Cloud Vision API cannot be assumed to distinguish between the two species. Therefore, using exact matching between species names and Google labels will lead to a high number of false negatives.

An example of a Google Cloud Vision API result for a single photo correctly tagged on Flickr as Adder, gives the following categories: *Reptile (98%), Snake (98%), Scaled reptile (93%), Viper (91%), Serpent (89%), Terrestrial Animal (87%), Rattle Snake (84%), Sidewinder (70%), Adaptation (67%), Colubridae (65%), Eastern Diamondback Rattlesnake (56%), Elapidae (53%).* The higher the score, the more confident the assignment of the category is for the given image, where the score is given in brackets next to the tags.

The photo labels returned by Google Cloud Vision API can be organised as a taxonomy that matches the species taxonomy returned by NBN.

Fig 2. Google Cloud Vision API label taxonomy and NBN classification for Adder

Figure 2 displays the NBN classification for Adder and the labels returned by Google Cloud Vision API for this photo. We use the NBN taxonomic classification for the species to choose relevant tags to match the tags returned by Google Cloud Vision API. A bag-of-words (BoW) approach is adopted where we treat the names in the classification hierarchy for a species as a list of names ignoring the hierarchical and semantic relations between these names. We consider all names in the classification hierarchy following down from class, and we match these terms to the terms given by Google Cloud API. In the example, given in Fig. 2, the use of the classification finds an exact match between the species name 'Viper' (an alternative name for 'Adder') and the Google Cloud Vision API term Viper. Using exact matching on the Flickr label of Adder would not have found any match, resulting in a false negative for this observation. Another example is for species *Phleum pratense (Timothy Grass)*, which is from class Magnoliopsida and family *Poacae (Grass)*. Google Cloud Vision API returns for images with this species the label 'grass', rather than 'timothy grass' and thus coarse match would be successful in this case.

Note that we use both scientific and common names for matching, as both can occur within the NBN derived taxonomy and the tags returned by the Google Cloud Vision API. We performed manual verification of the Google results for a 1000 randomly selected instances.

Data analysis

The data analyses are based on two case studies: the 1500 most frequently recorded species on NBN and the invasive species in the UK that appear in both data collections. Spatial comparison between the NBN and Flickr datasets was performed using spatial grid modelling, in which geographic space is divided into regular grid cells. The cells were classified according to whether they contained observations from one or other or both of the two sources. The classification was further refined according to time windows to support both a spatial and a temporal analysis. By varying grid cell sizes, and cell aggregation (i.e. one by one vs three by three), as well as the time window, we performed a number of scale-variant spatio-temporal analyses.

There were two main methods of performing spatial analysis:

1. One by one cell comparison: We compare Flickr and NBN species occurrence data per 10km, 20km, and 40km grid square. We calculate a confusion matrix, which is used to describe the performance of a classifier on a test data set for which the true values are known, where Flickr is the test data set and NBN the true values. The cells of the confusion matrix are defined as follows:
   - 'True Positive' (TP): a cell has both NBN and Flickr data points for the species
   - 'True Negative' (TN): a cell does not have occurrences from either of the sources
   - 'False Negative' (FN): a cell has no Flickr data for the species, but it does have NBN data for the species
   - 'False Positive' (FP): a cell has Flickr data for the species but no NBN data

2. Three by three cell comparison: We compare Flickr and NBN using a three by three analysis centred on every cell. In this approach, we count a true positive if there is a Flickr posting in a cell and if there are NBN records within either the cell itself or in any of the adjacent eight cells. A false negative would be declared if a set of nine cells had at least one NBN record but no Flickr record. A false positive indicates if there is a Flickr posting in a cell, but there are no NBN records within either the cell itself or in any of the adjacent eight cells. A 'True Negative' would be no Flickr postings and no NBN records in any of the nine cells.

Based on the measures above we compute precision, recall, and F1-measure. Recall is calculated by dividing the number of True Positives by the True Positives plus the False Negatives ($\frac{TP}{TP+FN}$). Thus if there were 10 cells containing NBN data, and for each of them Flickr data were also present, then the recall would be 100% or 1.0.

Precision is calculated by dividing the number of True Positives by the False Positives plus the True Positives ($\frac{TP}{TP+FP}$). In the previous example if, in addition to the 10 cells containing both NBN and Flickr data, there were a further 5 cells that contained Flickr data but no NBN data, then the precision would be 66% or 0.66.

F1 Score is calculated using recall and precision. It is used because precision and recall alone are not an accurate representation of one data set's superiority over another, as one could have better precision and the other a better recall. The F1 Score provides a harmonic mean that gives a clearer view of a dataset's accuracy when compared with ground truth. It is computed as double the product of precision and recall divided by the sum of precision and recall($\frac{2*precision*recall}{precision+recall}$).

We look at temporal accuracy of Flickr on seasonal (3 months), half yearly (6 months) and yearly patterns (12 months).

Results and discussion

Statistical Analysis

NBN and Flickr datasets comparison

Across the 1500 most numerous species on NBN Atlas, 90% were found on Flickr and a 100% of species in the Flickr dataset were found on NBN Atlas. The NBN Atlas records, as expected, far outnumber those on Flickr, being 93,656,179 and 791,059 respectively. It is worth noting that NBN data used here covers the entire collection period; 1800-2018 while Flickr data covers only 2006-2018. It was found that 35% of the species counted on Flickr have more than 100 occurrences. Table 1 lists the top 10 most frequently recorded species on Flickr (mostly with more than 10000 occurrences).

Table 1. The top 10 species on Flickr with the highest number of records

Scientific name	Common name	Flickr count	NBN count
Hyacinthoides non-scripta	*Bluebell*	20,940	54,893
Bellis perennis	*Daisy*	20,656	28,748
Erithacus rubecula	*Continental Robin*	19,248	3,938,616
Fagus sylvatica	*Beech*	15,842	24,973
Hedera helix	*Ivy*	14,474	27,211
Anas platyrhynchos	*Mallard*	13,500	834,039
Taraxacum officinale agg.	*Dandelion*	13,443	27,269
Pteridium aquilinum	*Bracken*	12,708	30,741
Phleum pratense	*Timothy Grass*	9,000	11,903

The best represented species on Flickr (see Table 1,can be split into three main categories: pretty, ie. photogenic, flowers (Bluebell, Daisy, Dandelion), sessile green plant species (Ivy, Beech, Bracken) and garden and aquatic birds, which are also diurnal (Continental Robin, Mallard). Notably all are easily accessible. These same patterns were mirrored at the class level with the highest number of returns for Flickr being *Magnoliopsida*, a class of flowering plants, and the second highest was *Aves*. *Phleum pratense (Timothy Grass)* as a well documented species in Flickr is an interesting observation as, compared to the other commonly observed species (see Table 1), it is not a well known species that is readily identified, suggesting that it was incidental in many images and Flickr may be good at picking up species that appear as a background in a photo. Another example of such a species is *Hedera helix (Ivy)*.

NBN and Flickr datasets are similar in the diversity of classes they represent with the ten best represented classes in both collections being the same , with the same three most common classes of *Insecta* (Insects), *Magnoliopsida* (plants), and *Aves* (birds). Both data collections are representing species from a small number of classes. This suggests that the same observer bias in photos also occurs in NBN data collections.

The top 10 species on NBN are garden birds (see Table 2), and they are represented well in the Flickr dataset with occurrences in most cases above a thousand.

Table 2. The top 10 species on NBN with the highest number of records

Scientific name	Common name	NBN count	Flickr count
Turdus merula	*Blackbird*	4,609,821	3,234
Cyanistes caeruleus	*Blue Tit*	4,164,338	3,491
Erithacus rubecula	*Continental Robin*	3,938,616	2,786
Columba palumbus	*Woodpigeon*	3,584,436	1,660
Prunella modularis	*Dunnock*	3,513,651	2,179
Parus major	*Great Tit*	3,507,350	2,670
Fringilla coelebs	*Chaffinch*	3,444,776	3,474
Passer domesticus	*House Sparrow*	3,184,175	2,312
Streptopelia decaocto	*Collared Dove*	3,094,475	929
Chloris chloris	*Greenfinch*	2,900,214	2,030

NBN and Flickr datasets comparison for invasive species in the UK

There are 82 invasive species for UK that also have occurrence records on NBN. The total count of records of invasive species on NBN is 1,485,744. The total number of Flickr posts for the invasive species that are also recorded on NBN is 27,187. The number of species with occurrences above 100 for both NBN and Flickr data collections is 19 (of 82), which is 23% of the number of invasive species on NBN.

The invasive species with more than 100 occurrences for both NBN and Flickr are diurnal mammals, birds (more than 50%) along with a few ”pretty” flower species (see Table 3).

The best represented invasive species on Flickr are *Branta canadensis (Canada Goose)*, *Scirurus carolinensis (Grey Squirrel)*, *Gallinago gallinago (Snipe)*, *Oryctolagus cuniculus (Rabbit)*, *Rhododenron ponticum (Rhododendron)*, *Aix galericulata (Mandarin Duck)*, and *Cygnus atratus (Black Swan)*. The species for which NBN and Flickr have a similar number of records are *Sus scrofa (Wild boar)* and *Bubo bubo (Eurasian Eagle Owl*).

Table 3. Species occurrences for NBN and Flickr for invasive species with number of species above 100

Scientific name	Common name	NBN species count	Flickr species count
Branta canadensis	*Canada Goose*	377,111	3,328
Sciurus carolinensis	*Grey squirrel*	350,113	3,249
Gallinago gallinago	*Snipe*	325,210	1,619
Oryctolagus cuniculus	*Rabbit*	96,093	7,994
Alopochen aegyptiacus	*Egyptian Goose*	31,591	862
Rhododenron ponticum	*Rhododendron*	30,803	3,489
Branta leucopsis	*Barnacle Goose*	24,269	289
Aix galericulata	*Mandarin Duck*	19,693	1,500
Muntiacus reevesi	*Reeve's muntjac*	16,428	489
Cygnus atratus	*Black Swan*	8,761	1,148
Buddleja davidii	*Buddleia*	5,654	443
Heracleum mantegazzianum	*Giant Hogweed*	5,348	190
Anser caerulescens	*Snow Goose*	5,085	177
Anser indicus	*Bar-Headed Goose*	3475	164
Aix sponsa	*Wood Duck*	2,688	290
Cervus nippon	*Sika Deer*	2,442	226
Chrysolophus pictus	*Golden Pheasant*	1,745	167
Sus scrofa	*Wild boar*	441	373
Bubo bubo	*Eurasian Eagle Owl*	395	537

Flickr data verification

In initial exploratory work, we performed tests with the tags returned by the Google Cloud Vision API. We found that the tags with a score above 60% are more likely to imply the correct species displayed on the photos. The tags with a score lower than 60% usually describe either less relevant objects of the photo, e.g. parts of the background (*'leaf'*), characteristics of the animal ('fawn'), or are names of species that are irrelevant to the photo ('Diamondback Rattlesnake' when the species is Adder). Therefore, we used only tags with a score higher than 60%.

We evaluate the efficiency of our image verification approach (i.e. bag of words) against the exact-match approach described in [13] for a 1000 randomly selected instances spanning 10 species. The authors of [13] used Google Reverse Image Search for performing their experiments. However, we use Google Cloud Vision API as Google Reverse Image Search does not support automatic verification.

Results (presented in Table 4) show that the BoW-based approach performs better than the baseline approach (exact-match) for identifying genuine wildlife observations on Flickr with accuracy 96% versus 27%.

The precision of both approaches is nearly 100%. However, the recall of the exact-match approach is quite low with 24% while the BoW approach is 96%. These results show that our approach is an effective method for verifying wildlife observations on Flickr for large species collections.

Table 4. Comparison between BoW image and exact-match image verification

	BoW Approach	Exact-Match Approach
precision	0.99	0.99
recall	0.96	0.24
F1-measure	0.98	0.38
accuracy	0.96	0.27

The reason the BoW approach gives much higher recall is that Google Cloud Vision API performs well distinguishing between species classes, however it can struggle to distinguish between species that belong to the same genus or class. For instance, if we have a photo of *Sciurus carolinensis (Grey squirrel)*, Google Cloud Vision API will return a large number of squirrel-related tags, including the names of other species such as 'fox squirrel', and 'ground squirrels'. The most common causes of false positives for BoW are photos that include an artificial representation of a species, such as a boat with a figure of a goose (Fig 3), and hence do not represent a living species. Common cases of false negatives for BoW are photos which include the species but the focus of display is another object. In the example given in Fig 4, the main object in the photo is a building, and thus Google Cloud Vision API returns labels associated with the building and the characteristics of the building, rather than the plant (i.e. Hedera helix (Ivy)).

Fig 3. Common cases of false positive and false negative: False Positive for *Marus bassanus (Gannet)*(left) tags: bird, goose, vehicle, tall ship and False Negative for *Hedera helix (Ivy)* (right): tags: property, house, home, building, residential area, cottage, real estate, neighbourhood

Spatial and Temporal analysis

The top 1500 species on NBN

The average precision and recall across all species for each type of spatial and temporal constraint for one by one grid cell analysis is 0.38 (38%) for precision and 0.2 (20%) for recall. The recall score shows that on average 20% of all NBN data was also reflected by the Flickr data. The precision score shows that on average 38% of the Flickr cell-based identifications of a species were correct relative to the NBN (see Fig 4).

The average precision and recall across all species for each type of spatial and temporal constraint for three by three analysis is 0.6 (60%) for precision and 0.1 (10%) for recall (see Fig 4). In comparison to one by one analysis, the average precision for three by three analysis is higher ranging from 0.27 (27%) to 0.78 (78%) for the different cell sizes while recall tends to be lower and does not vary much for the different cell sizes. The number of false negatives is significantly higher for three by three analysis and thus the recall value is lower. The reason for this is can be attributed to the wider range of species recorded within the NBN in comparison to the Flickr records. Further, according to the conditions for three by three analysis comparisons, false negatives occur when a set of nine cells have at least one NBN record in the absence of any Flickr record for the given species. Therefore, for species where the number of NBN records is high and the number of Flickr records is low it is very likely that cells with no Flickr occurrences will be associated with cells containing NBN records (note however that for species that are well represented on Flickr this is less likely to be the case).

The highest precision and recall scores across both types of analysis are achieved for experiments performed with cell size 40km and no temporal constraints. The lowest results are achieved for experiments performed with a time constraint of twelve months which we attribute to lack of data on Flickr.

Precision tends to be higher than recall. This higher precision reflects the fact that most locations with Flickr occurrences also contain NBN occurrences. The low recall indicates that there are many locations with NBN observations but with no Flickr observations, leading as indicated above to false negatives. However, the recall value increases significantly as the cell size increases. The more relaxed spatial restrictions allow for what would otherwise be false negatives to become false positives as Flickr occurrences over a wider region are taken into account. Also, precision increases for bigger cell sizes for the converse reason of taking account of NBN occurrences over a wider region relative to a Flickr observation. This indicates that a grid split, consisting of 40km cells provides a better balance between precision and recall measures and thus can be regarded as more suitable for validating social network observations.

Fig 4. Average Precision and Recall comparison per cell size and temporal restriction: One by one analysis (left), Three by three analysis (right) where 'P' refers to precision and 'R' refer to recall. In the figure 'no constraints' refers to the analyses performed with no temporal constraints.

We calculated the best, worst and average F1 performance (see Fig 5), where best and worst were based on the average F1 scores for individual species for a particular cell size, while average F1 performance was across all species for the particular cell size. The average F1 measurement does not exceed 0.5. Best performing species have poorer F1 scores for cell sizes 10km and 20km and F1 score of 0.7 for the analysis performed on cell size 40km. F1-measure on average is higher when the analysis is performed with no temporal constraints. Further, the average F1 scores for the analysis conducted using a 12-month window are the lowest.

Fig 5. Comparison of average, best, worst F1 measure values per temporal restriction and cell size: One by One analysis (left),Three by three analysis(right)

The results (see Fig 4 and Fig 5) show that the Flickr dataset best reflects the NBN dataset on a purely spatial analysis with no time constraints. The comparison with a constraint that observations are within 12 months of each other gives the lowest results on all measures.

As indicated above, the overall comparison between the two datasets is notable for the highly unbalanced precision and recall scores. As these scores are averaged across all considered species, we investigated those species with precision and recall both being above 0.5 , and we found 134 distinct such species. We found the average F1 score for the top 10 of these species with a 40km grid size to be 0.68 (see Table 5). As before the best results were obtained with no temporal constraints, though with a couple of exceptions for a 6 month temporal window. The best represented species on Flickr in comparison to NBN as represented in Table 5 are, with one exception, birds, most but not all of which are diurnal.

Table 5. The top ten results with the highest f1-measure across all species

Species name	Analysis type	Cell size	Precision	Recall	F1-measure
Thymelicus sylvestris (Small Skipper)	no constraints	40	0.64	0.77	0.70
Strix aluco (Tawny Owl)	no constraints	40	0.65	0.76	0.70
Sitta europaea (Nuthatch)	no constraints	40	0.6	0.82	0.69
Primula veris (Cowslip)	no constraints	40	0.61	0.79	0.69
Aegithalos caudatus (Long-Tailed Tit)	no constraints	40	0.56	0.88	0.68
Botaurus stellaris(Bittern)	no constraints	40	0.61	0.76	0.68
Libellula depressa (Broad-Bodied Chaser)	no constraints	40	0.63	0.73	0.68
Sitta europaea (Nuthatch)	6 months	40	0.62	0.74	0.68
Aegithalos caudatus (Long-Tailed Tit)	6 months	40	0.59	0.78	0.67
Certhia familiaris (Treecreeper)	no constraints	40	0.56	0.83	0.67

Invasive species for UK

The average results for the invasive species demonstrate the same spatial and temporal patterns as the average results for the top 1500 species, presented in the previous section, i.e. best performance is for spatial analysis performed with 40 km grid cell size with no time constraints (Fig 6 and 7).

The average precision and recall across all species for each type of spatial and temporal constraint for one by one analysis is 0.4 (40%) for precision and 0.2 (20%) for recall (see Fig 6). The average precision and recall across all species for each type of spatial and temporal constraint for three by three analysis is 0.6 (60%) for precision and 0.1 (1%) for recall (see Fig 6).

Fig 6. Average Precision and Recall comparison per cell size and temporal restriction for invasive species: One by one analysis on the left, three by three analysis on the right, where 'P' refers to precision and 'R' refers to recall. 'no constraints' refers to analyses performed with no temporal constraints.

Fig 7. Comparison of average, best, worst F1 measure values per temporal restriction and cell size for invasive species: The one by one analysis is on the left, the three by three analysis on the right.

The best represented invasive species on Flickr in comparison to NBN, with precision and recall both being above 0.5, are given in Table 6. Results are promising for these species as the F1-measure is on average 61.2%, specifically for representing spatial patterns on 40km cell size with no time constraints (see Table 6). There are seven distinct species with the best performance among the invasive species (note that in Table 6 some species have multiple rows with different conditions of analysis). The species - *Branta canadensis (Canada Goose)* and *Sciurus carolinensis (Grey squirrel)* appear across the multiple categories of no temporal constraints, three months constraints and six months constraints. They are the best performing species in terms of having both precision and recall above 0.5 for multiple spatial and temporal restrictions and are the only species which have both precision and recall above 0.5 for cell size 20km. The best performance in terms of highest precision and highest F1 measure has been achieved for *Bubo bubo (Eurasian Eagle Owl)* with F1 = 0.71 and precision = 0.64 (with recall 0.79). These results are achieved with 40km cell size and no temporal constraints.

Of the top 1500 most numerous species on NBN 90% were also found on Flickr, confirming that social media data can represent a wide range of species. A comparison between the two data collections on the diversity of species shows that NBN and Flickr datasets are similar on the class of species they represent. The best represented classes in both collections are the same with the top three being Insecta (Insects) Magnoliopsida (Plant class), and Aves (birds). Flickr has a good representation of flowering plants and garden and seabirds. Many Flickr uploads represent species that look attractive on photos and are easier to capture (i.e. they are diurnal, and/or are sessile) as well as being relatively common species.

Our image verification approach proved to work well on a large collection of species. The approach by [13] of exact match between the Google tags and species names may work for a small collection of well-known species for which the Google species labels tend to be more reliable, but not for a more extensive collection, including less well-known species, for which the Google label is liable to be more generic (i.e. providing the class or genus rather than the actual species name). In our approach, we use the taxonomy

Table 6. Results for the invasive species where precision and recall are both above 0.5

Species name	Analysis type	Cell size	Precision	Recall	F1-measure
Branta canadensis (Canada Goose)	no constraints	40km	0.55	0.80	0.65
Branta canadensis (Canada Goose)	no constraints	20km	0.55	0.53	0.54
Cygnus atratus (Black Swan)	no constraints	40km	0.59	0.79	0.68
Sciurus carolinensis (Grey squirrel)	no constraints	20km	0.57	0.58	0.58
Sciurus carolinensis (Grey squirrel)	no constraints	40km	0.59	0.80	0.68
Buddleja davidii (Buddleia)	no constraints	40km	0.51	0.51	0.51
Bubo bubo (Eurasian Eagle Owl)	no constraints	40km	0.64	0.79	0.71
Aix galericulata (Mandarin Duck)	no constraints	40km	0.51	0.60	0.55
Cygnus atratus (Black Swan)	3 months	40km	0.55	0.54	0.55
Branta canadensis (Canada Goose)	3 months	40km	0.62	0.62	0.62
Sciurus carolinensis (Grey squirrel)	3 months	40km	0.59	0.63	0.62
Cygnus atratus (Black Swan)	6 months	40km	0.59	0.71	0.65
Branta canadensis (Canada Goose)	6 months	40km	0.59	0.69	0.64
Sciurus carolinensis (Grey squirrel)	6 months	40km	0.60	0.73	0.66
Bubo bubo (Eurasian Eagle Owl)	6 months	40km	0.55	0.75	0.64
Aix galericulata (Mandarin Duck)	6 months	40km	0.54	0.50	0.52

structure of the species to select relevant tags. Thus we verify images as genuine wildlife by matching the provided Flickr species name against the Google-provided class or the genus of the image content and all the tags lower down the classification hierarchy.

The spatial and temporal analyses for both case studies show that the Flickr dataset reflects the NBN dataset patterns best for experiments performed with cell size 40 km with no temporal constraints. The poorer results from the analysis performed with temporal constraints suggest that the Flickr dataset does not represent the temporal patterns for the species on NBN well. This is especially true for the yearly comparison between the two datasets (i.e. 12 month window).

The results of the precision calculations showed that there is a large number of species for which precision is higher than 60%, for cell sizes 20km and 40km. This observation suggests that Flickr posts do present a potentially useful source of wildlife observations. However, the low recall value indicates that the Flickr data collection is less able to represent the full range of wildlife species in comparison to NBN. This is emphasised in the three by three analysis that gives the highest precision values, but provides the poorest recall. It should be remarked here that our scores for precision depend upon the quality of the NBN ground data, and it is quite possible that some of the Flickr observations classed here as false positive could actually be correct due to the absence of existing citizen science observations at the respective location.)

This problem leads to our next step, which is to conduct a similar larger scale study of the potential (beyond the relatively limited studies conducted to date) of other social networks such as Twitter to determine whether they can also supplement traditional biodiversity data sources. Collecting social network data on a larger scale is a challenging task because most of the networks have restrictions on data access with thresholds on the amount of data that can be downloaded. A solution to this might be to look at how data from multiple social network sources can be combined for extracting wildlife data. It is also a strong motivation to apply and if possible improve upon methods for geocoding the many accessible social media posts that do not have GPS coordinates [20].

There is also scope to improve our image-verification method by looking for example at using a combination of inclusive and exclusive tags (i.e. tags used to consider a photo irrelevant) and through the development of more sophisticated computer vision methods for automated identification of individual species. We will also investigate methods of automatic verification that a social media posting is a genuine wildlife observation.

References

1. Amano T, Lamming JD, Sutherland WJ. Spatial gaps in global biodiversity information and the role of citizen science. Bioscience. 2016;66(5):393–400.
2. Barve V. Discovering and developing primary biodiversity data from social networking sites: A novel approach. Ecological Informatics. 2014;24:194–199.
3. Silvertown J. A new dawn for citizen science. Trends in ecology & evolution. 2009;24(9):467–471.
4. Cohn JP. Citizen science: Can volunteers do real research? BioScience. 2008;58(3):192–197.
5. Guerrini CJ, Majumder MA, Lewellyn MJ, McGuire AL. Citizen science, public policy. Science. 2018;361(6398):134–136. doi:10.1126/science.aar8379.
6. Daume S. Mining Twitter to monitor invasive alien species—An analytical framework and sample information topologies. Ecological Informatics. 2016;31:70–82.
7. Di Minin E, Tenkanen H, Toivonen T. Prospects and challenges for social media data in conservation science. Frontiers in Environmental Science. 2015;3:63.
8. August TA, Pescott OL, Joly A, Bonnet P. AI Naturalists Might Hold the Key to Unlocking Biodiversity Data in Social Media Imagery. Patterns. 2020;1(7):100116.
9. Ghermandi A, Sinclair M. Passive crowdsourcing of social media in environmental research: A systematic map. Global environmental change. 2019;55:36–47.
10. Antoniou V, Fonte C, See L, Estima J, Arsanjani J, Lupia F, et al. Investigating the feasibility of geo-tagged photographs as sources of land cover input data. ISPRS International Journal of Geo-Information. 2016;5(5):64.
11. Soliman A, Soltani K, Yin J, Padmanabhan A, Wang S. Social sensing of urban land use based on analysis of Twitter users' mobility patterns. PlOS ONE. 2017;12(7):e0181657.
12. Daume S, Albert M, von Gadow K. Forest monitoring and social media–Complementary data sources for ecosystem surveillance? Forest Ecology and Management. 2014;316:9–20.
13. ElQadi MM, Dorin A, Dyer A, Burd M, Bukovac Z, Shrestha M. Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in Australia. Ecological informatics. 2017;39:23–31.
14. Kent JD, Capello Jr HT. Spatial patterns and demographic indicators of effective social media content during the Horsethief Canyon fire of 2012. Cartography and Geographic Information Science. 2013;40(2):78–89.
15. Palomino M, Taylor T, Göker A, Isaacs J, Warber S. The online dissemination of nature–health concepts: Lessons from sentiment analysis of social media relating to "nature-deficit disorder". International Journal of Environmental Research and Public Health. 2016;13(1):142.

16. Sui D, Goodchild M. The convergence of GIS and social media: challenges for GIScience. International Journal of Geographical Information Science. 2011;25(11):1737–1748.

17. Davis A, Major RE, Taylor CE, Martin JM. Novel tracking and reporting methods for studying large birds in urban landscapes. Wildlife Biology. 2017;2017(4).

18. Fritz S, McCallum I, Schill C, Perger C, See L, Schepaschenko D, et al. Geo-Wiki: An online platform for improving global land cover. Environmental Modelling & Software. 2012;31:110–123.

19. Lowry CS, Fienen MN. CrowdHydrology: crowdsourcing hydrologic data and engaging citizen scientists. GroundWater. 2013;51(1):151–156.

20. Stock K. Mining location from social media: A systematic review. Computers, Environment and Urban Systems. 2018;71:209 – 240.