Chapter

Integrating Citizen Science and GIS for Wildlife Habitat Assessment

Guiming Zhang

Abstract

With the rapid advancement and popularity of geospatial technologies such as location-aware smartphones, mobile maps, etc., average citizens nowadays can easily contribute georeferenced wildlife data (e.g., wildlife sightings). Due to the wide spread of human settlements and lengthy living histories of citizens in their local areas, citizen-contributed wildlife data could cover large geographic areas over long time spans. Citizen science thus provides great opportunities for collecting wildlife data of extensive spatiotemporal coverage for wildlife habitat assessment. However, citizen-contributed wildlife data may be subject to data quality issues, for example, imprecise spatial position and biased spatial coverage. These issues need to be accounted for when using citizen-contributed data for wildlife habitat assessment. Geovisualization and geospatial analysis capabilities provisioned by geographic information systems (GISs) can be adopted to tackle such data quality issues. This chapter offers an overview of citizen science as a means of collecting wildlife data, the roles of GIS to tackle the data quality issues, and the integration of citizen science and GIS for wildlife habitat assessment. A case study of habitat assessment for the black-and-white snub-nosed monkey (*Rhinopithecus bieti*) using *R. bieti* sightings elicited from local villagers in Yunnan, China, is presented as a demonstration.

Keywords: citizen observers, local villagers, wildlife sightings, geovisualization interview, habitat suitability mapping, data quality, black-and-white snub-nosed monkey (*Rhinopithecus bieti*)

1. Introduction

Habitats provide resources such as food, shelter, potential nesting sites, and mates for wildlife to achieve survival and reproduction [1]. Understanding the requirements or preferences of wildlife on their habitats and assessing the quality of wildlife habitat is of great importance for conservation biologists and conservation managers [2]. For example, wildlife habitat assessment supports conservation practices such as ex situ or reintroduction and restoration conservation, predicting risk of invasive species, systematic conservation planning, assessing threats, and setting conservation priorities [3–6].

One approach to assessing wildlife habitat quality is to predict wildlife habitat suitability maps indicating the spatial variation of habitat suitability [7]. Habitat suitability mapping is often carried out in a geographic information system (GIS) [8].

Conceptually, spatial prediction of wildlife habitat suitability requires GIS data layers characterizing the environmental conditions (environmental data) and knowledge on the relationship between wildlife habitat suitability and environmental conditions. Based on the relationship and the environmental conditions at a location (e.g., a pixel), the in situ habitat suitability can be inferred. Inferring habitat suitability at every location in the study area of interest results in a suitability map [7]. Such a habitat suitability map can then be used to assess the spatial variation of wildlife habitat quality and to support conservation. With the rapid development of geospatial technologies, environmental data for characterizing environmental conditions are becoming abundant and increasingly available [9, 10]. The key for wildlife habitat assessment through habitat suitability mapping therefore lies in obtaining knowledge on the relationship between wildlife habitat suitability and environmental conditions (environmental niche).

Data-driven approaches are most commonly adopted in deriving the relationship between wildlife habitat suitability and environmental conditions (environmental niche modeling) [7]. Data-driven approaches for environmental niche modeling require wildlife data indicating habitat use, for example, abundance data, presence and/or absence data. Wildlife data are overlaid with environmental data layers to extract the environmental conditions at locations where habitat use occurs. The relationship between wildlife habitat suitability and environmental conditions can then be derived through statistical analysis, machine learning, data mining, or other modeling techniques [7]. Thus, wildlife data become the key to deriving the relationship between habitat suitability and environmental conditions for mapping wildlife habitat suitability for wildlife habitat assessment.

Traditionally, wildlife data are collected using various techniques such as field observation, radio telemetry, infrared trapping cameras, and global positioning system (GPS) collars [11, 12]. Accurate wildlife data can be collected through these techniques, but admittedly these techniques are also somewhat expensive to deploy [13]. The high cost may prevent these techniques from being used in wildlife data collection, particularly for areas and projects with limited budget support. Besides, some of these techniques (e.g., field observation and GPS collars) are logistically difficult for areas with rugged terrains and limited accessibility [13]. Low-cost techniques such as trailing wildlife markings and interviewing local people about wildlife sightings through questionnaires are also used for wildlife data collection, but wildlife data collected with these techniques can be of low quality (e.g., inaccurate spatial location and/or time) [14, 15]. Cost-effective methods for collecting wildlife data of satisfactory quality are ideal for wildlife habitat assessment and sustainable conservation given that much of the world's biodiversity occurs in the world's poorest and remote countries [16].

Local residents were proven to be a cost-effective source of obtaining wildlife data [17, 18]. Many local residents, such as those living in remote rural areas and particularly those whose livelihoods are closely linked to ecosystem services (e.g., subsistence farmers, shepherds, and hunters), spend a great deal of time in the field. They encounter wildlife in its natural environment and, as a result, accumulate a rich knowledge about the wildlife habitat use. Wildlife data elicited from local residents at relatively low cost, although may be subject to data quality issues (e.g., data credibility, positional accuracy, spatial bias, etc.), could be used to support and sustain conservation programs with limited budget.

From a broader perspective, the increasing availability of citizen-contributed data accompanied by the advancements in GIS has created the opportunity to make full use of citizen science to address many real-world problems. On the one hand, citizen-contributed data have become increasingly available with the resurrected popularity of citizen science [19] and the emerging phenomenon of volunteered geographic information (VGI) [20]. One prominent example is the eBird citizen science project that is driven by bird watchers and documenting bird species across the globe [21]. On the other hand, the advancements in GIS capabilities (e.g., geovisualization and spatial analysis) have made it possible to accommodate the data quality issues associated with citizen-contributed data to make use of such data for scientific inquires [22, 23].

This chapter offers an overview of citizen science for wildlife data collection and its integration with GIS for wildlife habitat assessment. A case study of habitat assessment for the black-and-white snub-nosed monkey (*Rhinopithecus bieti*) using data contributed by local residents in Yunnan, China, is presented as an illustration.

2. Citizen science for wildlife data collection

2.1 Citizen science

The term *citizen science* was formally added to the Oxford English Dictionary only recently in 2014 [24], referring to "Scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions" [25]. Nonetheless, citizen science has been practiced for centuries, long before *scientist* slowly became a profession throughout the seventeenth to nineteenth centuries [24]. For example [26], Benjamin Franklin (1706–1790), as a physicist, was famous for his discoveries and theories regarding electricity while he was also a printer, diplomat, and politician; Charles Darwin (1809–1888) as a biologist was best known for his contributions to the theories of evolution, but on the Beagle voyage, he was sailing as an unpaid companion, not as a professional scientist. Even after the *scientist-as-profession* paradigm has been well established, average citizens continue to engage in scientific work at various levels of involvement [27]: *contributory* where citizens mostly contribute to data collection, *collaborative* where citizens also participate in data analysis, and *co-created* where citizens get involved at all stages of the project including conceiving and designing the research. Exemplary long-running citizen science projects related to wildlife population monitoring are the Christmas Bird Count (CBC) established in 1900 [28] and the Breeding Bird Survey (BBS) established in 1965 [29] for monitoring bird species in North America. Data contributed by participants in such citizen projects are now supporting wildlife population trends monitoring [30] and decision-making in conservation [31].

The rapid advancements of geospatial information technologies in the last decade have greatly prompted the flourish of citizen science. Location-aware portable devices constantly connected to the Internet (e.g., GPS-enabled smart phones) are now commonplace. Average citizens thus can conveniently contribute georeferenced wildlife observations using such devices via social media, mobile map, citizen science project mobile apps, etc. [26, 32, 33]. From a geographic and GIS perspective, citizen science involving geospatial data generation (e.g., wildlife sightings with location information) is called "geographic citizen science" [34] and the georeferenced wildlife observations are a form of VGI [20, 34]. Due to the increasing availability of enabling technologies, millions of citizens across the world are participating in citizen science projects and many of them are contributing large volumes of wildlife observations on a daily basis. Interested readers can check out a wide range of ongoing citizen science projects (not limited to wildlife-related projects) at scistarter.com and search for specific projects by topic and/or location. As of the time of writing, searching projects at scistarter.com by the topic "Animals," "Birds," and "Insects & Pollinators" returned 382, 162, and 190 projects, respectively. As a prominent example, the eBird project [21, 35], launched in 2002 by the Ornithology Lab at Cornell University and the National Audubon Society, as of November 2016, has engaged over 330,000 bird watchers from more than 250 countries who have reported observations of over 10,300 bird species. As of June 2018, eBird has accumulated over 500 million bird observations in its global database; in recent years, there have been more than 100 million bird observations added to the database each year.

Wildlife data contributed by participants in such citizen science projects are a form of geospatial big data [36, 37]. Complex patterns can be discovered from such intensive data through visualizations, simulations, data mining, and various modeling techniques to provide valuable insight for forming concrete hypotheses about the underlying ecological, biological, and geographical processes that generated the observed data [37]. Thus, the abundance of citizen-contributed wildlife data has the potential of shifting research paradigm in biological, ecological, and geographical studies from the traditional hypothesis-driven approach to the emerging data-driven approach; for instance, scholars are promoting the idea of "data-intensive science" for biodiversity studies and "data-driven geography" [36–38].

2.2 The (dis)advantages of citizen science for collecting wildlife data

Citizen science has several advantages as an alternative mechanism for collecting wildlife data. Citizen-contributed data contain rich local information that spans a wide temporal spectrum because citizens, as local experts and sensors [20], have long been sensing and accumulating knowledge of their respective areas. Citizen science also has the potential to provide wildlife data over large areas, given that billions of networked human sensors are distributed across the globe. In addition, citizen science can provide timely updated wildlife data that are difficult to obtain and maintain through other techniques but can be easily elicited from citizens living in the local areas. Moreover, citizen-contributed data are much less expensive than traditional scientific data collection protocols (e.g., biological survey). In many cases, citizens contribute data purely voluntarily [20]. This low cost is of great practical significance in many real-world programs falling short of funding support.

Due to the above advantages of citizen science, it is possible to obtain timely updated wildlife data using citizen science over large areas. Citizen science thus has a great potential to support and sustain long-time wildlife population monitoring at large spatial scale (e.g., eBird) and provide wildlife data for wildlife habitat assessment.

In spite of the strengths, one should be aware of the shortcomings of the "citizen science" approach to wildlife data collection. For example, this approach cannot be used in areas with low population where sufficient local citizen observers/informants are lacking. It is also not good for collecting data on evasive animals with little contact with humans. Most importantly, there can be data quality issues associated with wildlife data contributed by volunteer citizens (i.e., non-professionals) which make the data challenging to standardize and analyze [17, 18]. The following sections detail some of the data quality issues, their implications for wildlife habitat assessment, and how GIS techniques (geovisualization, geospatial analysis, geocomputation, etc.) can be adopted to tackle the issues toward reducing the impact of such issues on wildlife habitat assessment.

2.3 The data quality issues of citizen-contributed wildlife data

The quality of citizen-contributed wildlife data is the major concern when using such data for wildlife habitat assessment. The average citizens engaged in citizen science projects are not well-trained professionals; their voluntary data collection actions are mostly constrained by internal commitment. Thus, citizen-contributed wildlife data may or may not be accurate [20, 39]. Three aspects of data quality are particularly relevant to the use of citizen-contributed wildlife data for wildlife habitat assessment: data creditability, positional accuracy, and spatial bias.

2.3.1 Data credibility

In order to be useful for wildlife habitat assessment, wildlife data (e.g., sightings) reported by citizen participants need to be credible, that is, provide ground truth wildlife observations. Data credibility is affected by the characteristics of both the wildlife and the citizen observers (e.g., local residents). On the one hand, local residents often only observe wildlife that is active in the daytime. The target wildlife should be easily recognizable to reduce misidentification given that local residents usually have no training on species identification [17, 40]. On the other hand, local resident knowledge of the target wildlife, age, length of residence, and formal education also influence data credibility [41]. For instance, performance in georeferencing tasks differs between novice and expert citizen participants [42]; there exists both between-observer differences [43] and within-observer differences (over time) [44] in BBS participant bird-counting skills.

Various methods have been developed for increasing the credibility of citizen-contributed wildlife data. Ref. [45] identified a total of 12 strategies that have been adopted by citizen science programs to increase their data credibility across different program stages including training and planning, data collection, and data analysis and program evaluation. As an example, eBird uses a two-part approach to assure data credibility during data entry [39]: automated data quality filters flag records for review based on observation date and geographic location; a flagged entry, once confirmed as legitimate by the observer, is then reviewed by a regional expert reviewer again.

2.3.2 Positional accuracy

Position of the wildlife data used for habitat suitability mapping needs to be accurate so that the locations can be used to accurately obtain the corresponding environmental conditions at these locations from environmental data layers. Insufficient positional accuracy of wildlife data leads to mismatch between the locations of wildlife habitat use and the corresponding environmental conditions, and thus degrades the accuracy of environmental niche modeling and habitat suitability mapping [46].

Nonetheless, it is also important to note that the impact of positional accuracy of wildlife data on habitat suitability mapping depends on the spatial resolution at which suitability mapping is conducted. Mapping at high spatial resolution (e.g., using environmental data of 30 m × 30 m grids) definitely requires wildlife data of high positional accuracy that is comparable to the spatial resolution of the environmental data so that values of the environmental conditions at these locations can be accurately extracted from environmental data layers. In contrast, for mapping at coarse spatial resolution (e.g., 1000 m × 1000 m grids), the absolute positional accuracy of wildlife data does not have to be very high as long as it is accurate enough relative to the spatial resolution of environmental data in use.

2.3.3 Spatial bias

Wildlife observations contributed by citizens are often concentrated more in some geographic areas than others (i.e., spatial bias) because observations made by citizens are opportunistic in nature [23]. Unlike well-designed sampling or survey schemes which allocate observation sites in a way such that the geographic space and/or the environmental space are well covered by the observation sites, spatial distribution of the observation efforts of citizen volunteers would be considered neither random nor regular in the sense of sampling or survey design. One example to demonstrate this is wildlife sightings elicited from local residents. Local residents are not intentionally tracking wildlife of interest. Instead, they typically spot the wildlife en route to doing something else. The routes on which local citizens spot wildlife would be considered neither random nor regular but "ad hoc" [23]. As a result, wildlife sightings elicited from local residents are usually concentrated in areas with higher route accessibility.

Such spatial bias in wildlife data has a significant impact on environmental niche modeling and habitat suitability mapping for wildlife habitat assessment. Due to the spatial bias, citizen-contributed wildlife data might not be representative of the actual wildlife habitat use. The relationship derived based on the wildlife data thus might not well represent the underlying environmental niche. Spatial bias in citizen-contributed wildlife data, if not appropriately accounted for, would adversely affect the accuracy of wildlife habitat suitability mapping [47–49].

3. The roles of GIS

GIS is the ideal tool for conducting wildlife habitat assessment as it involves geospatial data. Besides providing an integrated environment for managing and manipulating environmental data layers and georeferenced wildlife data, GIS can also offer capabilities to remedy or address some of the data quality issues associated with citizen contribute wildlife data. Firstly, geovisualization can be used to facilitate wildlife data elicitation from citizen participants and improve positional accuracy. Secondly, based on the cause of spatial bias, spatial analysis can be used to compensate for the biased coverage in observation efforts. Lastly, geospatial computation techniques can be employed to address the computational challenges arising from analyzing very large volumes of citizen-contributed wildlife data.

3.1 Geovisualization to improve positional accuracy

In general, positional accuracy of wildlife data largely depends on the availability of positioning technology. Wildlife sightings can be accurately georeferenced with the aid of high-accuracy positioning techniques. For example, smartphones equipped with high-accuracy GPS units ensure generated data record is associated with accurate geographic coordinates. Nevertheless, the above observations hold only for citizen observers who are reporting or recording data at the time of sighting wildlife occurrences in the field. In many cases, local residents (e.g., farmers) do not keep records of daily wildlife sightings or they simply do not have access to GPS units or smartphones. Most often, wildlife data are elicited from their memories long after the time of sighting [17, 18, 23, 40].

Wildlife data (e.g., sightings) collected from local citizens through interviews or questionnaire surveys often have position information with unsatisfactory accuracy [14, 15]. Descriptions of the locations of wildlife sightings are often imprecise or vague, particularly if a long time has lapsed since the actual sightings. Such incapability partly results from the absence of an effective interviewing media (e.g., an intuitive and interactive representation of the natural environment where local citizens are active) that facilitates local citizens to recall and locate their sightings of wildlife. Ref. [17] collected distribution and abundance data of terrestrial tortoises from local shepherds over 1 km × 1 km grid cells with the aid of topographic maps. However, it is difficult to accurately locate wildlife sightings on topographic maps for the local residents who had no training in map reading.

Geospatially enabled and user-friendly geovisualization interfaces could help improve positional accuracy of the wildlife data elicited from local residents [50, 51]. Geovisualization, particularly 3D geovisualization techniques, can be adopted to help local residents to recall and locate their sightings of wildlife and obtain wildlife data with more accurate positional information [40]. Given the flat 2D topographic maps, the local residents need relief interpretation skills to re-construct the 3D topography of the landscape; local residents can then orientate themselves and locate places on the landscape. But they often do not have much training in basic map reading skills, not to mention relief interpretation. 3D geovisualization can facilitate relief interpretation by producing a realistic and intuitive terrain representation [52] and improves visual search efficiency and navigation performance [53].

Geovisualization techniques as discussed above help improve positional accuracy of wildlife data at the very beginning of data generation. Sometimes, in cases where positional uncertainty does exist in wildlife data and is indeed of concern for wildlife habitat assessment, GIS-based methods have been developed to reduce its impact on the accuracy of wildlife habitat assessment. As an example, [54] proposed a spatial sampling method for deriving probable wildlife occurrence locations from patrol records using heuristics based on data recording context and species ecology to increase the accuracy of habitat suitability mapping.

3.2 Geospatial analysis to tackle spatial bias

Many geospatial analytical methods have been proposed to account for the spatial bias in wildlife data. An *AdaSTEM* approach that exploits variation in the density of wildlife observations was proposed to accommodate spatial bias in citizen-contributed wildlife observations [22, 55]. The continent- or hemisphere-wide study area is partitioned into rectangular spatial units (i.e., sub-areas) of size dependent upon density of wildlife observations. Environmental niches are modeled with only observations in each sub-area. By training local models in sub-areas, instead of training a global model using observations over the whole area, this approach mitigates spatial bias in the overall data set to a certain degree.

Filtering samples in the geographic or environmental space (i.e., remove observations that are within certain distance of one another) is also applied to reduce spatial bias [56, 57]. This method is based on the heuristic that removing localities (i.e., field samples) that are within certain distance of one another would somehow balance the unequal sampling or observation effort. The key of this method is to determine the distance threshold properly.

If detailed information observation effort is available, such information can then be incorporated to correct for spatial bias. Spatial bias in wildlife observations was compensated for by weighting the observations with weights inversely proportional to the cumulative visibility at the observation sites, given that cumulative visibility is a good proxy of the underlying observation effort [23]. Here, cumulative visibility is the frequency at which a given location can be seen by observers from the routes the observers take. It can be computed based on a digital elevation model (DEM) representing the terrain and the routes using viewshed analysis, a common GIS function.

A FactorBiasOut method was developed to correct for spatial bias in species presence-only data for species distribution modeling with MAXENT [58]. This method first estimates an empirical distribution to approximate the underlying but usually unknown sampling distribution that generated the presence-only data. This approximate sampling distribution is then used to factor out the spatial bias in presence-only data. This is achieved by feeding MAXENT with background data that have the same spatial bias as the presence data. For example, occurrence data of a target group of species that are observed by similar methods can be taken as the estimate of the effort information and thus are used as the background data.

Recently, a general representativeness-directed approach was proposed to spatial bias mitigation in citizen-contributed wildlife observations (i.e., samples) for habitat suitability mapping [59]. The key idea is to define and quantify the representativeness of samples and then properly weigh the samples to improve representativeness. Sample representativeness is defined as the "goodness-of-coverage" of the samples in the environmental covariate space, which in turn is measured by the similarity between the probability distribution of the samples in the covariate space and the probability distribution of all mapping units (e.g., pixels) within the study area. Spatial bias is then mitigated by weighting the samples toward increasing sample representativeness. The optimal weights that maximize sample representativeness are determined through an optimization procedure using a genetic algorithm.

3.3 Geocomputation to enable big data analysis

Citizen-contributed wildlife data are an important source of geospatial big data. In spatial analysis or modeling of such large volume of data (e.g., point pattern analysis, wildlife habitat assessment, and species distribution modeling), it is urgent to address the associated computational challenges. Geocomputation technologies could be utilized to address such computational challenges.

For example, over 100 million bird observations were added to the eBird database each year. Point pattern analysis is commonly used to discover patterns from such data. Existing point pattern analysis software tools are not able to handle geospatial big data efficiently. Cutting-edge geocomputation technologies such as cloud computing and GPU (graphics processing units) computing can be leveraged to accelerate point pattern analysis algorithms. The massively parallel computing powers of cloud computing and GPU computing effectively sped up point pattern analysis tasks on big data by a factor of hundreds [60, 61]. Given the significant acceleration brought by the geocomputation technologies, geospatial big data analysis tasks that once were computationally prohibitive can now be conducted in a timely manner.

4. Integrating citizen science and GIS for wildlife habitat assessment: a *Rhinopithecus bieti* case study

A case study of mapping black-and-white snub-nosed monkeys' (*Rhinopithecus bieti*) habitat suitability using *R. bieti* sighting data elicited from local villagers at Mt. Lasha in Yunnan, China, was presented to demonstrate the integration of citizen science and GIS for wildlife habitat assessment.

4.1 Species and study site

R. bieti is an "endangered" species on the IUCN (International Union for Conservation of Nature) Red List of Threatened Species [62]. *R. bieti* is endemic to the eastern Himalayas in northwest Yunnan and southeast Tibet, China, between the upper Mekong and Yangtze Rivers with 19 relatively isolated groups [63]. The monkeys use lichens (e.g., arboreal fruticose *Bryoria* and *Usnea* spp.) as their main food [64]. They prefer fir-larch forest at higher elevations in the northern part of the range but also stay in mixed coniferous and broad-leaf forest at lower elevations (above 2600 m) in the southern range [65]. Across its geographic distribution areas, the habitat of *R. bieti* has undergone degradation (e.g., habitat reduction and fragmentation) in the past decades due to the growth of human population that mostly employed traditional modes of production in its distribution area (e.g., clear-cutting forests for farming, grazing and firewood consumption; hunting) [65].

The study site is Mt. Lasha area located in northwest Yunnan Province, southwest China (**Figure 1**). Mt. Lasha is near the southern-most part of its geographic range [63, 65]. The 20.31 km^2 study area is an important habitat for a group of about 100 *R. bieti* individuals in 11 one-male multi-female units and two all-male units [66]. *R. bieti* is a significant species with a strong historic dimension in local communities [63]. On the one hand, hunting poses the greatest threat to the monkeys [63]. Local residents had long been hunting the monkeys for various purposes. Even after the Chinese government has designated the species in the first class of protected animals since 1977, illegal hunting had not been stopped completely. On the other hand, *R. bieti* habitat use is closely related to forest-provisioned resources including food and shelter [63, 65]. In the study area, two historical events had significant impacts on these forest-provisioned resources: in 1979 the China Environmental Protection Act was enacted; in 2006, the Mt. Lasha area became a protected area as part of the Yunling Nature Reserve. The forestry policy implementations associated with these events in the study area directly affected local residents' exploitation of the forests.

4.2 Data collection

4.2.1 Wildlife data elicited from local villagers

Sightings of *R. bieti* were elicited from local villagers for assessing the habitat of the monkeys in the study area across historical periods. Local villagers whose livelihoods are closely dependent on ecosystem services have long been living in the local area and accumulated information about *R. bieti* habitat use. Sightings of *R. bieti*

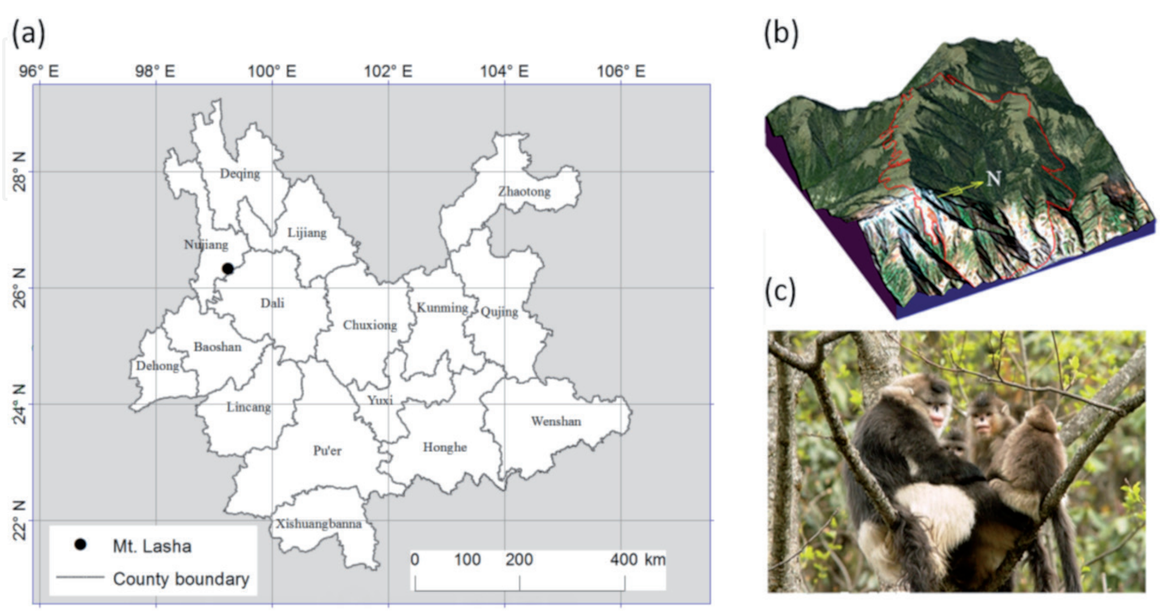

Figure 1.
Location of the study area: (a) Mt. Lasha in northwest Yunnan, China; (b) a 3D perspective image of the Mt. Lasha area; and (c) a family of R. bieti in their natural habitat (extracted from [40] with permission from John Wiley and Sons).

were elicited from local villagers through interviews. A 3D geovisualization tool was adopted to aid the interviews by using it to help the local villagers recall and locate where they sighted the monkeys more accurately.

R. bieti sightings were collected through structured interviews with local villagers (**Figure 2**). Sightings of the monkeys and activity routes of the villagers were elicited. The interviews were conducted using 3dMapper, a 3D geovisualization GIS tool that uses high-resolution DEM and satellite imagery to produce an intuitive 3D view of the study area [67] (freely available from solim.geography.wisc.edu). The user can zoom, pan, and easily draw points, lines, or polygons over the 3D scene. We introduced this geovisualization tool to the villagers to help them identify locations where they had sighted the monkeys and the daily routes they took in the area. The villagers also recalled the year and month when they sighted the monkeys or took the routes. The year of *R. bieti* sightings recalled by the interviewees was cross-checked with and refined with reference to timing of major events such as national policy implementations, date of marriage and child born, etc. and the month to seasonal activity patterns in the area such as farming and grazing. Information on where and when they sighted the monkeys was recorded as polygons. Information on the routes they took and the frequency with which they took each route was recorded as lines.

Geovisualization interview sessions were carried out by one biologist and one field assistant who were very familiar with the study area during July and August 2010. Sixty-eight local residents including herdsmen, hunters, and farmers who had extensive experience in the mountains from all five nearby villages were interviewed. The majority of them are aged between 30 and 60 (**Table 1**). The elicited *R. bieti* sightings and activity routes of the villagers cover a temporal span from the 1950s through 2010. Constrained by the availability of environmental data needed for habitat assessment, only *R. bieti* sightings in three historical periods (1973–1981, 1987–2005, and 2006–2010) were used for habitat assessment (see [40] for details) (**Figure 3**).

4.2.2 Environmental data

Environmental factors impacting *R. bieti* habitat use include terrain, water source, shelter and food, and human-posed disturbance [65, 68, 69]. Accordingly, the following environmental data layers were used in habitat assessment (habitat suitability mapping) for *R. bieti* in the study area [23, 40]: elevation, slope gradient, slope

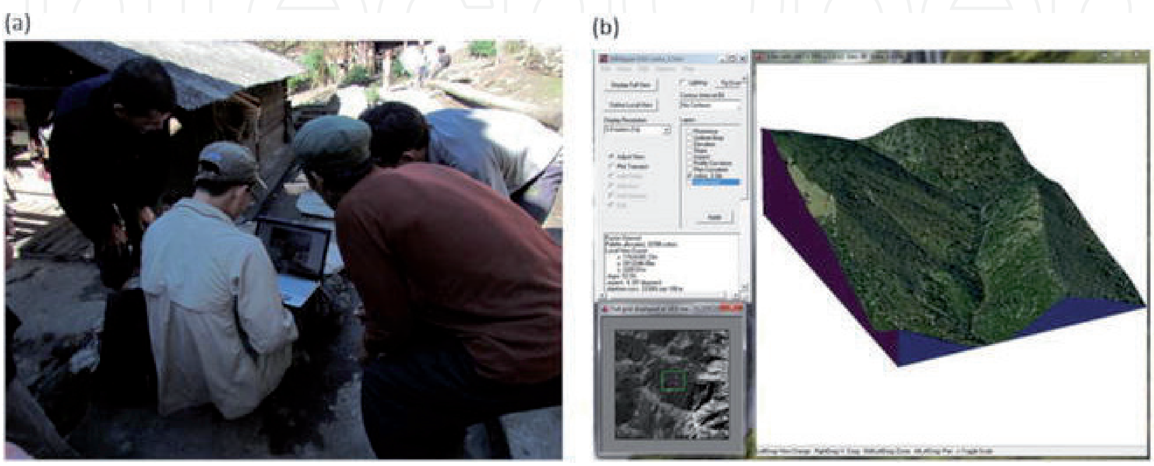

Figure 2.
Geovisualization interview sessions with the local residents using 3dMapper: (a) the local residents locating monkey sightings and activity routes and (b) a 3D scene of a small portion of the study area on which the local residents outlined monkey sightings and routes (extracted from [40] with permission from John Wiley and Sons).

Age	19–30	31–40	41–50	51–60	61–70	71–78
Count	7	12	16	18	10	5

Table 1.
Age composition of the interviewed local villagers.

aspect, distance to river, distance to village or road, and vegetation type. Interested readers can refer to [40] for details on how to obtain these environmental data layers.

4.3 Accounting for positional uncertainty and spatial bias

Data elicited from local villagers impose two challenges, namely positional uncertainty and spatial bias. First, local villagers often recall *R. bieti* sightings in the form of "I saw the monkeys over this area." Clearly, "over this area" can be depicted using a polygon, but this does not mean that the monkeys showed up at every location in the polygon area and certainly not at an equal probability within the polygon. Thus, taking all locations in the polygon as sightings is not appropriate. The question then is how to obtain locations that are representative of the actual presence of wildlife in the area outlined by the local villagers. The second challenge is the spatial bias in the elicited *R. bieti* sightings due to local villagers' opportunistic observation effort. For example, multiple sightings of monkeys at one location by many villagers do not necessarily mean that the location is highly preferred by the monkeys; it might be that the location is easily visible from multiple activity routes. Thus, every time a monkey shows up at this location, it is spotted by some villager(s). On the other hand, a monkey that shows up at locations that are preferred by monkeys but less visible to the villagers will have a lesser chance of being spotted. This spatial bias must be compensated for when using the elicited *R. bieti* sightings for wildlife habitat assessment.

Geospatial analysis methods provisioned by GIS were adopted to address the two challenges. First, a frequency sampling strategy [23, 70] was applied to reduce the

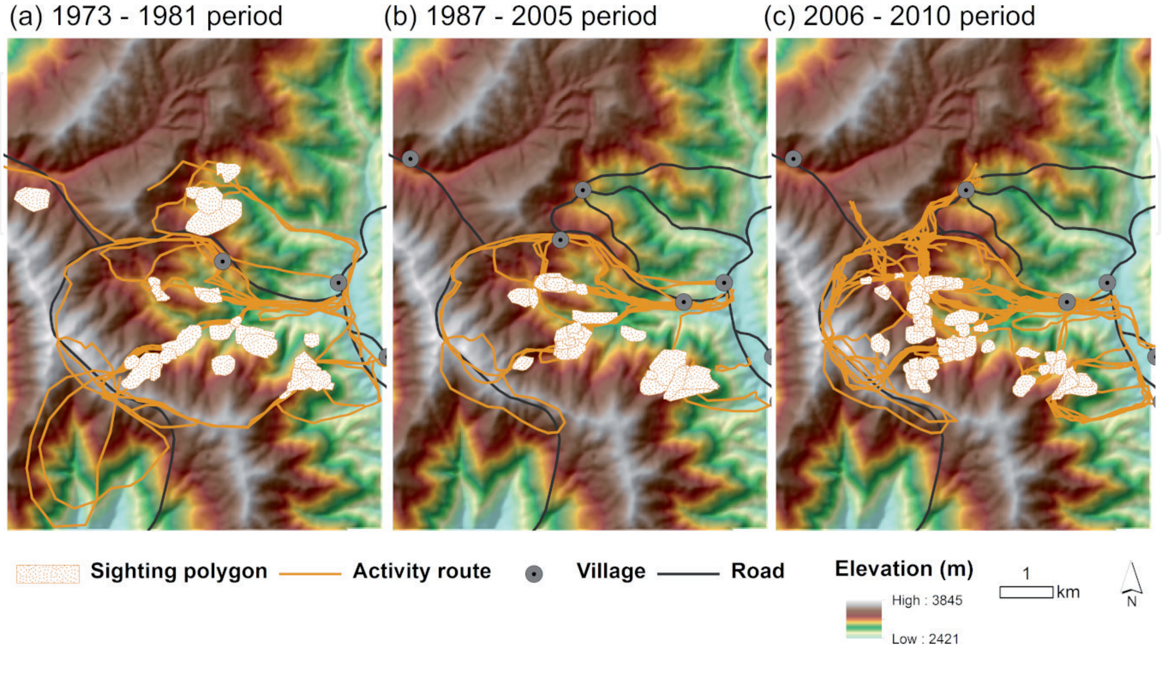

Figure 3.
Sightings of R. bieti and activity routes elicited from the local residents through geovisualization interviews: (a) 1973–1981 period; (b) 1987–2005 period; (c) 2006–2010 period (extracted from [40] with permission from John Wiley and Sons).

position uncertainty in sighting polygons provided by local villagers and to identify the representative locations for *R. bieti* presence within each polygon. It is assumed that locations at which values of environmental conditions are most frequent over the polygon area would approximate the locations of actual presence best. Under this assumption, the frequency sampling strategy implemented in GIS was applied to locate the representative locations within a polygon. Here, only the general idea was outlined as above; full details on implementing the sampling strategy in GIS can be found in [23, 70].

Second, the spatial bias was compensated for by inversely weighting each representative presence location with cumulative visibility of the location from the routes taken by local villagers [23]. In this particular case study, spatial bias in the elicited *R. bieti* sightings was a result of the non-random and uneven distribution of local villagers' observation efforts constrained by activity routes and terrain (as discussed in depth in Section 2.3.3). Thus, cumulative visibility was treated as a proxy of the underlying observation effort and can be incorporated to compensate for spatial bias. The cumulative visibility of a location can be computed in GIS based on a DEM and the activity routes of local villagers [23]. The efficacy of the frequency sampling strategy to reduce positional uncertainty and the visibility-weighting scheme to compensate for spatial has been demonstrated in [23].

4.4 Habitat assessment

A kernel density estimation-based method [23, 71] was applied to derive the relationship between *R. bieti* habitat suitability and environmental conditions. This method estimates a probability density function representing the probability distribution of wildlife presence over the gradient of each environmental factor based on the values of the environmental factors over the presence locations. In estimating the probability density functions, presence locations are weighted by the in situ cumulative visibility from activity routes of the local villagers to compensate for spatial bias. The probability density functions are then normalized to the range of [0, 1] to represent the relationships between habitat suitability and individual environmental factors (**Figure 4**). The overall habitat suitability considering all

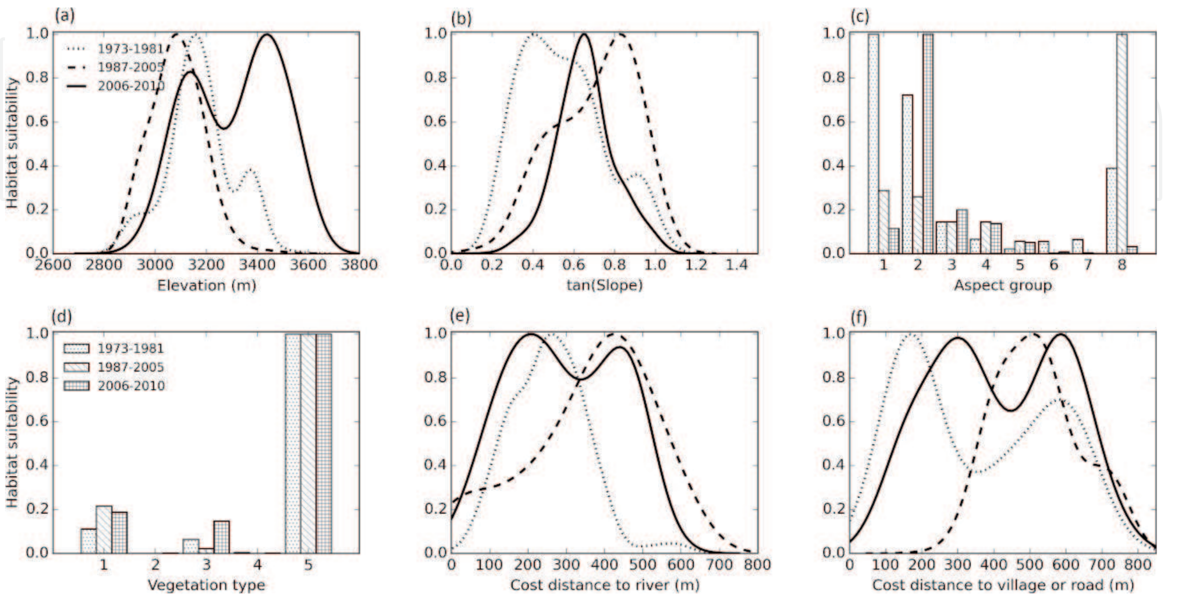

Figure 4.
Suitability-environment relationships derived from elicited R. bieti sightings in each historical period. Aspect group 1: 0–45° (starting from north), 2: 45–90°, 3: 90–135°, 4: 135–180°, 5: 180–225°, 6: 225–270°, 7: 270–315°, 8: 315–360°. Vegetation type 1: evergreen coniferous, 2: pasture, 3: yunnan pine, 4: farmland, 5: deciduous broadleaf (extracted from [40] with permission from John Wiley and Sons).

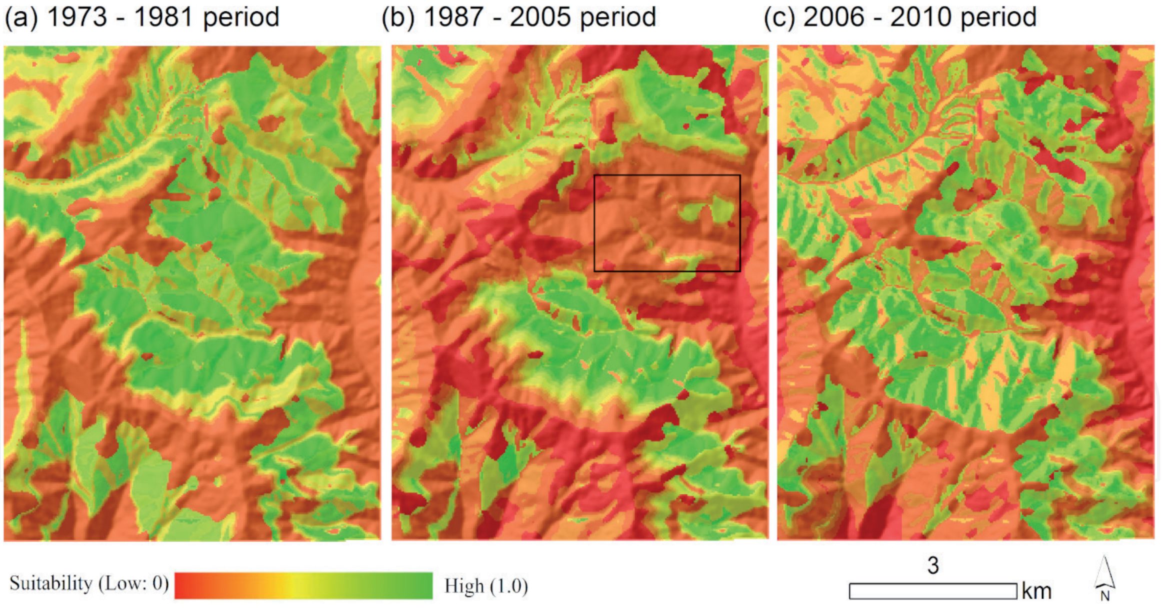

Figure 5.
Habitat suitability maps predicted for the study area using elicited R. bieti sightings in each historical period (a) 1973–1981 period; (b) 1987–2005 period; (c) 2006–2010 period (extracted from [40] with permission from John Wiley and Sons).

environmental factors is determined by integrating the relationships based on a "limiting factor" principle (see [23] for full details of the method). Computing the overall habitat suitability at every location (pixel) in the study area resulted in habitat suitability maps as shown in **Figure 5**.

Across the three historical periods, high suitability habitats were in forests (**Figure 4d**) at mid-to-high elevation range (**Figure 4a**) on the northeast hill slopes (**Figure 4c**). Overall, high suitability habitats shrank in the 1987–2005 period compared to the previous period. As an example, the area outlined on **Figure 5b** in the 1987–2005 period is of much lower suitability compared to the 1973–1981 period. This might be a result of the introduction of new village settlements and roads in that area in the 1987–2005 period which induced significant human disturbance. *R. bieti* habitats were recovering in the 2005–2010 period. The outlined area recovered to higher suitability in that period; this might be attributed to the monkeys getting used to proximity to villages and roads (**Figure 4f**).

The derived relationships between *R. bieti* habitat suitability and individual environmental factors (**Figure 4**) confirmed the recovering trend in the 2006–2010 period. The elevation range of high suitability habitats in the 2006–2010 period shifted back to higher ranges close to those in the 1973–1981 period (**Figure 4a**). The ranges of distance to rivers and distance to village or road corresponding to high suitability habitats also shifted back to similar ranges as in the 1973–1981 period (**Figure 4e, f**). These were potential evidences that conservation practices initiated by the Yunling Nature Reserve have restored *R. bieti* habitat in the area.

5. Conclusions

Wildlife data required for wildlife habitat assessment can be difficult and expensive to obtain with traditional data collection methods (e.g., biological survey, geographic sampling), especially for conservation programs with limited budget support in remote and poor areas. Citizen science offers a cost-effective way of collecting wildlife data to sustain such programs. Nevertheless, average citizens are non-professionals and their wildlife observation efforts are un-coordinated. Thus, wildlife data contributed by citizens may be subject to data quality issues such as positional uncertainty and spatial bias. This chapter provides an overview of citizen science as a means of collecting wildlife data, GIS-provisioned geovisualization, and geospatial analysis techniques for tackling the data quality issues of citizen-contributed wildlife data, and the integration of citizen science and GIS for wildlife habitat assessment. A case study of mapping *R. bieti* habitat suitability using *R. bieti* sightings elicited from local villagers in Yunnan, China, was presented as an example to demonstrate the usefulness of integrating citizen science and GIS for wildlife habitat assessment.

Acknowledgements

Support to the author from the University of Denver through the Faculty Startup Funds is greatly appreciated. The author would like to thank Dr. A-Xing Zhu in the Department of Geography at the University of Wisconsin-Madison for his invaluable inputs on designing methods, and Dr. Wen Xiao and Dr. Zhi-Pang Huang in the Institute of Eastern-Himalaya Biodiversity Research at Dali University and officials in the Yunling Nature Reserve administration for their generous support for field work of the reported *R. bieti* case study.

Conflict of interest

The author declares no conflict of interest.

Author details

Guiming Zhang
Department of Geography and the Environment, University of Denver, Denver, United States

*Address all correspondence to: guiming.zhang@du.edu

References

[1] Reunanen P, Monkkonen M, Nikula A. Habitat requirements of the Siberian flying squirrel in northern Finland: Comparing field survey and remote sensing data. Annales Zoologici Fennici. 2002;**39**(1):7-20

[2] Pulliam HR, Danielson BJ. Sources, sinks, and habitat selection: A landscape perspective on population dynamics. The American Naturalist. 1991;**137**(s1):S50-S66

[3] Araújo MB, Williams PH. Selecting areas for species persistence using occurrence data. Biological Conservation. 2000;**96**(3):331-345

[4] Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, et al. Predicting species distributions for conservation decisions. Ecology Letters. 2013;**16**(12):1424-1435

[5] Thorn JS, Nijman V, Smith D, Nekaris K a I. Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (primates: Nycticebus). Diversity and Distributions. 2009;**15**(2):289-298

[6] Thuiller W, Richardson DM, Pyšek P, Midgley GUYF, Hughes GO, Rouget M. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology. 2005;**11**(12):2234-2250

[7] Franklin J, Miller JA. Mapping Species Distributions: Spatial Inference and Prediction. Vol. 338. Cambridge: Cambridge University Press; 2009

[8] Rotenberry JT, Preston KL, Knick ST. GIS-based niche modeling for mapping species' habitat. Ecology. 2006;**87**(6):1458-1464

[9] Viña A, Bearer S, Zhang H, Ouyang Z, Liu J. Evaluating MODIS data for mapping wildlife habitat distribution. Remote Sensing of Environment. 2008;**112**(5):2160-2169

[10] Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology. 2005;**25**(15):1965-1978

[11] Campbell AF, Sussman RW. The value of radio tracking in the study of neotropical rain forest monkeys. American Journal of Primatology. 1994;**32**(4):291-301

[12] Trolle M, Kéry M. Estimation of ocelot density in the pantanal using capture-recapture analysis of camera-trapping data. Journal of Mammalogy. 2003;**84**(2):607-614

[13] Hulbert IAR, French J. The accuracy of GPS for wildlife telemetry and habitat mapping. Journal of Applied Ecology. 2001;**38**(4):869-878

[14] Msoffe F, Mturi F, Galanti V, Tosi W, Wauters L, Tosi G. Comparing data of different survey methods for sustainable wildlife management in hunting areas: The case of Tarangire–Manyara ecosystem, northern Tanzania. European Journal of Wildlife Research. 2007;**53**(2):112-124

[15] Xu F, Ma M, Wu YQ, Chundawat RS. Distribution of the ibex (Capra ibex) in Tomur National Nature Reserve of Xinjiang, China. Zoological Research. 2007;**28**(6):670-672

[16] Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000;**403**(6772):853-858

[17] Anadón JD, Giménez A, Ballestar R, Pérez I. Evaluation of local ecological knowledge as a method for collecting extensive data on animal abundance. Conservation Biology. 2009;**23**(3):617-625

[18] Anadón JD, Giménez A, Ballestar R. Linking local ecological knowledge and habitat modelling to predict absolute species abundance on large scales. Biodiversity and Conservation. 2010;**19**(5):1443-1454

[19] Dickinson JL, Shirk J, Bonter D, Bonney R, Crain RL, Martin J, et al. The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment. 2012;**10**(6):291-297

[20] Goodchild MF. Citizens as sensors: The world of volunteered geography. GeoJournal. 2007;**69**(4):211-221

[21] Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S. eBird: A citizen-based bird observation network in the biological sciences. Biological Conservation. 2009;**142**(10):2282-2292

[22] Fink D, Damoulas T, Dave J, Damoulas T, Dave J. Adaptive spatio-temporal exploratory models: Hemisphere-wide species distributions from massively crowdsourced eBird data. In: Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13). 2013. pp. 1284-1290

[23] Zhu A-X, Zhang G, Wang W, Xiao W, Huang Z-P, Dunzhu G-S, et al. A citizen data-based approach to predictive mapping of spatial variation of natural phenomena. International Journal of Geographical Information Science. 2015;**29**(10):1864-1886

[24] Eitzel MV, Cappadonna JL, Santos-Lang C, Duerr RE, Virapongse A, West SE, et al. Citizen science terminology matters: Exploring key terms. Citizen Science: Theory and Practice. 2017;**2**(1):1

[25] OED. Citizen science [Internet]. Oxford English Dictionary. 2018. Available from: http://www.oed.com/view/Entry/33513?redirectedFrom=citizen+science#eid316619123 [Accessed: November 14, 2018]

[26] Silvertown J. A new dawn for citizen science. Trends in Ecology & Evolution. 2009;**24**(9):467-471

[27] Follett R, Strezov V. An analysis of citizen science based research: Usage and publication patterns. PLoS One. 2015;**10**(11):1-14

[28] Butcher GS, Fuller MR, McAllister LS, Geissler PH. An evaluation of the Christmas bird count for monitoring population trends of selected species. Wildlife Society Bulletin. 1990;**18**(2):129-134

[29] Sauer JR, Hines JE, Fallon JE, Link WA, Fallon JE, Pardieck KL, et al. The North American breeding bird survey 1966-2011: Summary analysis and species accounts. North American Fauna. 2013;**79**(79):1-32

[30] Snäll T, Kindvall O, Nilsson J, Pärt T. Evaluating citizen-based presence data for bird monitoring. Biological Conservation. 2011;**144**(2):804-810

[31] Sullivan BL, Phillips T, Dayer AA, Wood CL, Farnsworth A, Iliff MJ, et al. Using open access observational data for conservation action: A case study for birds. Biological Conservation. 2017;**208**:5-14

[32] Catlin-Groves CL. The citizen science landscape: From volunteers to citizen sensors and beyond. International Journal of Zoology. 2012;**2012**:1-14

[33] Newman G, Wiggins A, Crall A, Graham E, Newman S, Crowston K. The future of citizen science: Emerging technologies and shifting paradigms. Frontiers in Ecology and the Environment. 2012;**10**(6):298-304

[34] Haklay M. Citizen science and volunteered geographic information: Overview and typology of participation. In: Sui D, Elwood S, Goodchild M, editors. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. Dordrecht: SpringerNetherlands; 2013. pp. 105-122

[35] Wikipedia Contributors. eBird [Internet]. Wikipedia, The Free Encyclopedia. 2018. Available from: https://en.wikipedia.org/w/index.php?title=EBird&oldid=856114315 [Accessed: November 25, 2018]

[36] Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong W-K, Kelling S. Data-intensive science applied to broad-scale citizen science. Trends in Ecology & Evolution. 2012;**27**(2):130-137

[37] Kelling S, Hochachka WM, Fink D, Riedewald M, Caruana R, Ballard G, et al. Data-intensive science: A new paradigm for biodiversity studies. Bioscience. 2009;**59**(7):613-620

[38] Miller HJ, Goodchild MF. Data-driven geography. GeoJournal. 2014;**80**(4):449-461

[39] Kelling S, Lagoze C, Wong W-K, Yu J, Damoulas T, Gerbracht J, et al. eBird: A human/computer learning network to improve biodiversity conservation and research. AI Magazine. 2013;**34**(1):10-20

[40] Zhang G, Zhu A-X, Huang Z-P, Ren G, Qin C-Z, Xiao W. Validity of historical volunteered geographic information: Evaluating citizen data for mapping historical geographic phenomena. Transactions in GIS. 2018;**22**(1):149-164

[41] Brown G, McAlpine C, Rhodes J, Lunney D, Goldingay R, Fielding K, et al. Assessing the validity of crowdsourced wildlife observations for conservation using public participatory mapping methods. Biological Conservation. 2018;**227**(September):141-151

[42] Ellwood ER, Bart HL, Doosey MH, Jue DK, Mann JG, Nelson G, et al. Mapping life—Quality assessment of novice vs. expert georeferencers. Citizen Science: Theory and Practice. 2016;**1**(1):1-12

[43] Sauer JR, Peterjohn BG, Link WA. Observer differences in the North American breeding bird survey. Auk. 1994;**111**(1):50-62

[44] Kendall WL, Peterjohn BG, Sauer JR, Url S. First-time observer effects in the North American breeding bird survey. Auk. 1996;**113**(4):823-829

[45] Freitag A, Meyer R, Whiteman L. Strategies employed by citizen science programs to increase the credibility of their data. Citizen Science: Theory and Practice. 2016;**1**(2):1-11

[46] Osborne PE, Leitão PJ. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. Diversity and Distributions. 2009;**15**(4):671-681

[47] Kadmon R, Farber O, Danin A. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications. 2004;**14**(2):401-413

[48] Boakes EH, McGowan PJK, Fuller RA, Ding C, Clark NE, O'Connor K, et al. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. PLoS Biology. 2010;**8**(6):e1000385

[49] Leitão PJ, Moreira F, Osborne PE. Effects of geographical data sampling bias on habitat models of species distributions: A case study with steppe birds in southern Portugal. International

Journal of Geographical Information Science. 2011;**25**(3):439-454

[50] Seeger CJ. The role of facilitated volunteered geographic information in the landscape planning and site design process. GeoJournal. 2008;**72**(3-4):199-213

[51] Newman G, Zimmerman D, Crall A, Laituri M, Graham J, Stapel L. User-friendly web mapping: Lessons from a citizen science website. International Journal of Geographical Information Science. 2010;**24**(12):1851-1869

[52] Carbonell Carrera C, Bermejo Asensio LA. Augmented reality as a digital teaching environment to develop spatial thinking. Cartography and Geographic Information Science. 2017;**44**(3):259-270

[53] Liao H, Dong W, Peng C, Liu H. Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. Cartography and Geographic Information Science. 2017;**44**(6):474-490

[54] Zhang G, Zhu A-X, Huang Z-P, Xiao W. A heuristic-based approach to mitigating positional errors in patrol data for species distribution modeling. Transactions in GIS. 2018;**22**(1):202-216

[55] Fink D, Hochachka WM, Zuckerberg B, Winkler DW, Shaby B, Munson MA, et al. Spatiotemporal exploratory models for broad-scale survey data. Ecological Applications. 2010;**20**(8):2131-2147

[56] Boria RA, Olson LE, Goodman SM, Anderson RP. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling. 2014;**275**:73-77

[57] Varela S, Anderson RP, García-Valdés R, Fernández-González F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography. 2014;**37**(11):1084-1091

[58] Dudík M, Phillips SJ, Schapire RE, Dudik M, Schapire RE, Phillips SJ, et al. Correcting sample selection bias in maximum entropy density estimation. Advances in neural information processing systems. 2005;**17**:323-330

[59] Zhang G. A Representativeness Directed Approach to Spatial Bias Mitigation in VGI for Predictive Mapping [Thesis]. Madison: University of Wisconsin-Madison; 2018

[60] Zhang G, Huang Q, Zhu A-X, Keel J. Enabling point pattern analysis on spatial big data using cloud computing: Optimizing and accelerating Ripley's K function. International Journal of Geographical Information Science. 2016;**30**(11):2230-2252

[61] Zhang G, Zhu A-X, Huang Q. A GPU-accelerated adaptive kernel density estimation approach for efficient point pattern analysis on spatial big data. International Journal of Geographical Information Science. 2017;**31**(10):2068-2097

[62] IUCN. The IUCN red list of threatened species version 2018-2 [Internet]. 2018. Available from: https://www.iucnredlist.org [Accessed: December 05, 2018]

[63] Long YC, Kirkpatrick CR, Zhong T, Xiao L. Report on the distribution, population, and ecology of the Yunnan snub-nosed monkey (*Rhinopithecus bieti*). Primates. 1994;**35**(2):241-250

[64] Kirkpatrick RC, Long YC, Zhong T, Xiao L. Social organization and range use in the Yunnan snub-nosed monkey *Rhinopithecus bieti*. International Journal of Primatology. 1998;**19**(1):13-51

[65] Xiao W, Ding W, Cui LW, Zhou RL, Zhao QK. Habitat degradation of *Rhinopithecus bieti* in Yunnan, China.

International Journal of Primatology. 2003;**24**(2):389-398

[66] Huang ZP, Cui LW, Scott M, Wang SJ, Xiao W. Seasonality of reproduction of wild black-and-white snub-nosed monkeys (*Rhinopithecus bieti*) at Mt. Lasha, Yunnan, China. Primates. 2012;**53**(3):237-245

[67] Burt JE, Zhu AX. 3dMapper 4.02. 4.02. In: Terrain Analytics. Madison, WI: LLC; 2004

[68] Long Y, Zhong T, Xiao L. Study on geographical distribution and population of the Yunnan snub-nosed monkey. Zoological Research. 1996;**17**(4):437-441

[69] Huang ZP. Foraging, reproduction and sleeping site selection of black-and-white snub-nosed monkey (*Rhinopithecus bieti*) at the southern range. [master's dissertation]. Kunming: Faculty of Conservation Biology, Southwest Forestry University; 2009

[70] Qi F, Zhu A-X. Knowledge discovery from soil maps using inductive learning. International Journal of Geographical Information Science. 2003;**17**(8):771-795

[71] Zhang G, Zhu A-X, Windels SK, Qin C-Z. Modelling species habitat suitability from presence-only data using kernel density estimation. Ecological Indicators. 2018;**93**:387-396