Received: 7 May 2017 | Revised: 13 July 2017 | Accepted: 30 July 2017
DOI: 10.1111/tgis.12300

RESEARCH ARTICLE

WILEY Transactions in GIS

Validity of historical volunteered geographic information: Evaluating citizen data for mapping historical geographic phenomena

Guiming Zhang[1] | A-Xing Zhu[1,2,3,4,5] | Zhi-Pang Huang[6,7] | Guopeng Ren[6] | Cheng-Zhi Qin[4,5] | Wen Xiao[6,7]

[1] Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin
[2] Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing, China
[3] State Key Laboratory Cultivation Base of Geographical Environment Evolution, Nanjing Normal University, Nanjing, China
[4] State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[5] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Jiangsu, China
[6] Institute of Eastern-Himalaya Biodiversity Research, Dali University, Dali, China
[7] Collaborative Innovation Center for Biodiversity and Conservation in the Three Parallel Rivers Region of China, Dali, China

Correspondence
Wen Xiao, Institute of Eastern-Himalaya Biodiversity Research, Dali University, 2 Daligucheng Hongsheng Road, Dali 671003, China.
Email: xiaow@eastern-himalaya.cn
A-Xing Zhu, Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, School of Geography, Nanjing Normal University, Nanjing 210023, China.
Email: azhu@wisc.edu

Funding information
National Natural Science Foundation of China, Grant/Award Nos. 41431177, 31560118, 30960085; National Basic Research Program of China, Grant/Award No. 2015CB954102; Natural Science Research Program of Jiangsu, Grant/Award No. 14KJA170001; PAPD; National Key Technology Innovation Project for Water Pollution Control and Remediation, Grant/Award No. 2013ZX07103006

Abstract

Studies on volunteered geographic information (VGI) have focused on examining its validity to reveal geographic phenomena in relatively recent periods. Empirical evaluation of the validity of VGI to reveal geographic phenomena in historical periods (e.g., decades ago) is lacking, although such evaluation is desirable for assessing the possibility of broadening the temporal scope of VGI applications. This article presents an evaluation of the validity of VGI to reveal historical geographic phenomena through a citizen data-based habitat suitability mapping case study. Citizen data (i.e., sightings) of the black-and-white snub-nosed monkey (*Rhinopithecus bieti*) were elicited from local residents through three-dimensional (3D) geovisualization interviews in Yunnan, China. The validity of the elicited sightings to reveal the historical *R. bieti* distribution was evaluated through habitat suitability mapping using the citizen data in historical periods. The results of controlled experiments demonstrated that suitability maps predicted using the historical citizen data had a consistent spatial pattern (correlation above 0.60) that reflects the *R. bieti* distribution (Boyce index around 0.90) in areas free of significant environmental change across historical periods. This in turn suggests that citizen data have validity for mapping historical geographic phenomena. It provides supporting empirical evidence for potentially broadening the temporal scope of VGI applications.

1 | INTRODUCTION

Volunteered geographic information (VGI), referring to geographic information created by volunteer citizens, is an emerging phenomenon that has tremendous influence on GIScience (Goodchild, 2007). VGI has flourished with the rapid advancement of various enabling technologies in recent years (Web 2.0, virtual globe, location-based services, social media, etc.) (Brovelli, Minghini, & Zamboni, 2015; See et al., 2016). It serves as the backbone of many applications by providing data to reveal geographic phenomena in relatively recent periods (Fonte, Bastin, See, Foody, & Lupia, 2015; Gao et al., 2017; Goodchild & Glennon, 2010; Haklay & Weber, 2008; Rossiter, Liu, Carlisle, & Zhu, 2015; Sullivan et al., 2009; Zook, Graham, Shelton, & Gorman, 2010).

VGI also has the potential to provide data to reveal geographic phenomena in historical periods (*historical VGI*). Citizens as sensors (Goodchild, 2007) have long been observing the world and accumulating information, even before advanced technologies emerged. For example, local residents in remote rural areas whose livelihoods are closely linked to ecosystem services are valuable information sources for obtaining wildlife habitat use data (i.e., presences). Subsistence farmers, shepherds, hunters, and forest rangers have spent a great deal of time in the field and have encountered wildlife in its natural habitats. They have accumulated rich local ecological knowledge about the wildlife presences in their local areas (Anadón, Giménez, Ballestar, & Pérez, 2009; Huntington, 2000; Mackinson, 2001; Mackinson & Nottestad, 1998).

As a result, "citizen data," as a form of VGI that is either actively contributed by or passively elicited from volunteer citizens (Zhu et al., 2015), can be obtained to reveal geographic phenomena in the past. Zhu et al. (2015) proposed a citizen data-based approach for mapping natural geographic phenomena. Citizen data are elicited by interviewing citizen volunteers with the aid of three-dimensional (3D) geovisualization tools and are then used to build predictive models for mapping geographic phenomena (e.g., habitat suitability). Citizen data could have many practical applications. For example, it offers a cost-effective source for obtaining wildlife data to sustain conservation practices that have limited budget support (Anadón et al., 2009; Anadón, Giménez, & Ballestar, 2010; Danielsen et al., 2003). It is particularly desirable for conservation programs in poor and remote areas, where most biodiversity hotspots occur (Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000).

Using historical citizen data to understand geographic phenomena in the past is crucial for many applications. For example, species historical distributions are useful for understanding species niche conservatism and evolution (Soberón & Nakamura, 2009; Wiens & Donoghue, 2004), forming and assessing ecological hypotheses (Carnaval & Moritz, 2008; Peterson & Anamza, 2015; Werneck, Costa, Colli, Prado, & Sites, 2011), improving species delimitation (Pelletier, Crisafulli, Wagner, Zellmer, & Carstens, 2015), assessing climatic or anthropogenic effects on species persistence (Hanberry, He, Palik, & He, 2012; Nogués-Bravo, 2009), and informing species reintroduction conservation (Franklin & Miller, 2009; Graham, Ferrier, Huettman, Moritz, & Peterson, 2004).

However, special attention should be paid to the validity of citizen data when they are used to reveal geographic phenomena in historical periods. VGI quality is the most pressing concern for VGI applications, and methods have been developed to assess or assure VGI quality (Goodchild & Li, 2012; Senaratne, Mobasheri, Ali, Capineri, & Haklay, 2017). VGI quality is often assessed by comparing VGI against authoritative reference data sets from the perspectives of the fundamental dimensions of spatial data quality (positional, attribute, temporal, and semantic accuracy, logical consistency, completeness, lineage, etc.) (Antoniou & Skopeliti, 2015; Jackson et al., 2013; Senaratne et al., 2017). VGI quality assurance methods are mostly for VGI generated on digital platforms involving a large network of contributors, where observations from different contributors can be used to validate each other or converge to the "truth" (Ali & Schmid, 2014; Foody et al., 2013; Goodchild & Li, 2012). However, historical citizen data in some cases is the only source of information about many types of historical geographic phenomena, and there are simply no reference data sets available for comparing with VGI. Moreover, evaluation of VGI quality focuses on the raw data with little consideration of the validity of using VGI for spatial analysis and modeling (e.g., modeling species distribution). Besides, historical VGI contributors (e.g., residents in remote regions) often do not have access to networked digital platforms due to their lack of technical skills and limited accessibility to information infrastructures. Finally, VGI quality assessment and assurance focuses on VGI reflecting geographic phenomena in relatively recent periods after the enabling technologies (location-based services, social media, etc.) emerged. Treatment of VGI regarding geographic phenomena in historical periods is lacking.

Unlike VGI generated in the digital era, the validity of historical citizen data depends largely upon the accuracy of the recalled memories of volunteer citizens (i.e., no records were kept at the time of observation). Memories could become vague as time passes. Thus, the validity of historical citizen data might decay when being used to reveal geographic phenomena in earlier periods. For example, one should be cautious when using wildlife sightings elicited from volunteer citizens for modeling species historical distributions, because the quality of species occurrence data has a profound impact on the performance of species distribution models (Moudrý & Šímová, 2012; Osborne & Leitão, 2009). Are citizen data regarding geographic events or phenomena which happened decades ago still valid and useful to reveal the events or phenomena? If not, how far back in time can historical citizen data be considered trustworthy? These are legitimate questions regarding VGI usability (Antoniou & Skopeliti, 2015; Fonte et al., 2015) that one should ask before using historical citizen data in applications.

There are few studies involving historical citizen data collection and analysis. While collected citizen data might cover a long span back in time, studies have focused on examining the validity of citizen data to reveal geographic phenomena in relatively recent periods. For instance, Anadón et al. (2009) collected species distribution and abundance data from local shepherds. Data in a 10-year period prior to data collection were used for abundance modeling (Anadón et al., 2010). Zhu et al. (2015) collected wildlife sightings by interviewing local residents using a 3D geovisualization tool. Elicited wildlife sightings in a 5-year period prior to data collection were used for habitat suitability mapping.

There is a lack of evaluation of the validity of citizen data to reveal geographic phenomena in historical periods (e.g., decades ago). Evaluation of the validity of historical citizen data is desirable for assessing the possibility of broadening the temporal scope of VGI applications (e.g., using VGI to study species historical distributions).

This article presents an evaluation of the validity of historical VGI by evaluating citizen data for mapping geographic phenomena in historical periods. Citizen data on sightings of the black-and-white snub-nosed monkey (*Rhinopithecus bieti*) were elicited from local residents through geovisualization interviews (Zhu et al., 2015) at the Mt. Lasha area in Yunnan, China. This is the first time the geovisualization interview approach has been implemented to collect historical citizen data. The decades-long temporal coverage of the elicited citizen data was divided into three shorter historical periods. The validity of citizen data to reveal the *R. bieti* distribution in each historical period was then evaluated through habitat suitability mapping using citizen data in that period. This article is organized as follows. Section 2 presents data collection (citizen data elicitation and validation data collection) and experiment design (historical periods division, suitability mapping method, evaluation and assessment). Additional supporting materials are provided in the Supporting Information. Section 3 presents the evaluation results. Sections 4 and 5 present the discussion and conclusions, respectively.

2 | MATERIALS AND METHODS

2.1 | Study area

The study area is the Mt. Lasha area in Yunling nature reserve, northwest Yunnan Province, China (Figure 1). *R. bieti* is a significant species in this area. It is classified as "Endangered" on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (IUCN, 2016). The geographic distribution of *R. bieti* is limited to the area between the upper Mekong and Yangtze Rivers, mostly in northwest Yunnan and southeast Tibet (Long, Kirkpatrick, Zhong, & Xiao, 1994; Xiao, Ding, Cui, Zhou, & Zhao, 2003). The Mt. Lasha area is within the southern portion of the geographic range of *R. bieti*. It is a habitat for an isolated group of *R. bieti* composed of about 100 individuals (Huang, 2009; Huang, Cui, Scott, Wang, & Xiao, 2012). The Mt. Lasha area has been part of a protected area since the establishment of the Yunling Nature Reserve in 2006.

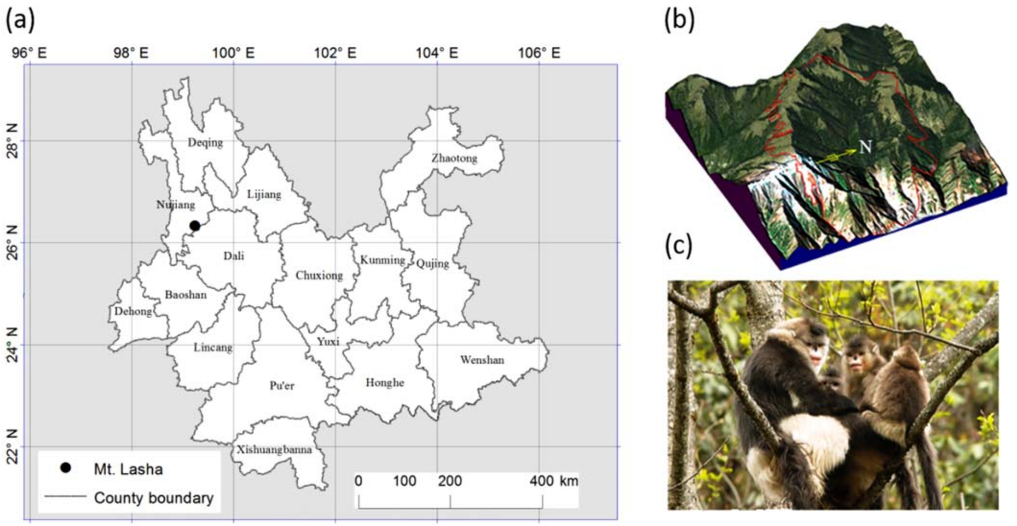

FIGURE 1 Location of the study area: (a) Mt. Lasha in Yunling Province Nature Reserve, Lanping county, Yunnan, China; (b) a 3D perspective image of the Mt. Lasha area; and (c) a family of *R. bieti* in their natural habitat. Photoby Z. F. Xiang

2.2 | Data collection

2.2.1 | Citizen data

R. bieti presence data were collected through geovisualization interviews with local residents who had intensive field experience in the local area. The interviews were conducted using 3dMapper (Burt & Zhu, 2004), a geovisualization GIS (geographic information system) software package that is capable of integrating a DEM (digital elevation model) with high-resolution satellite imagery to produce an intuitive 3D view of the study area. 3dMapper (http://solim.geography.wisc.edu) provides an intuitive and user-friendly interface through which the user can pan, zoom in or out, and rotate in the 3D scene (Figure 2). The user can report data by directly drawing points, lines, or polygons over the 3D scene and then fill in auxiliary information in the associated attribute tables.

The interviews were conducted by one biologist and one field assistant, who were very familiar with the study area (i.e., a local forest ranger). The interviewees were familiarized with the 3D visualization of the study area in 3dMapper by exploring the tool under the guidance of the interviewers. For most interviewees, it took about 20–30 min to learn how to accurately locate places named by the field assistant in the 3D scene. The interviewees were then

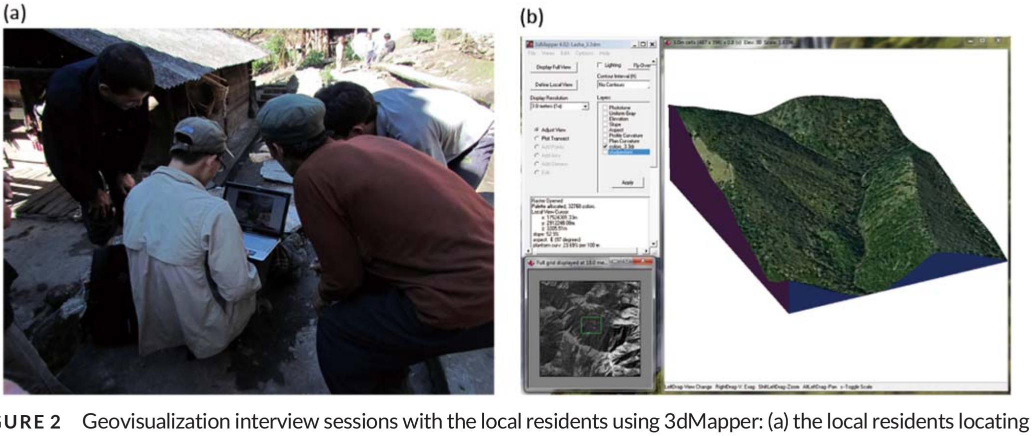

FIGURE 2 Geovisualization interview sessions with the local residents using 3dMapper: (a) the local residents locating monkey sightings and activity routes; and (b) a 3D scene of a small portion of the study area on which the local residents outlined monkey sightings and routes

TABLE 1 Age composition of the interviewees in years

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

asked to recall the areas where they had sighted the monkeys, along with the routes they took in the field. Information on where and when they sighted the monkeys was recorded as polygons and attribute tables tied to the polygons. Information on the activity routes, including the location and the timing and frequency of use of the routes, was recorded as polylines and attribute tables tied to the polylines. The spatial locations of sightings and routes were directly located and drawn on the 3D scene. Some interviewees, for example the elders, had difficulties using 3dMapper to directly locate monkey sightings and/or activity routes due to vision or map-reading problems. Instead, they provided sightings and routes through place names and/or descriptions of micro-terrain features. Sightings and routes were then located and drawn in 3dMapper by the field assistant based on their descriptions. The temporal information recalled by the interviewees (year, month) was refined by cross-checking the year against the timing of major events (e.g., national policy implementation, marriage, child born, etc.) and the month against seasonal activity patterns in the area (e.g., farming).

Geovisualization interview sessions with the local residents in the Mt. Lasha area were conducted in July and August 2010 (Figure 2). In total, 68 local residents from all five nearby villages who had extensive experience in the field were interviewed. Field activities regularly conducted by the interviewees included hunting, pasturing, medicinal herbs collection, logging, and farming in the area. Interviewees who went for hunting and pasturing usually spent more than six months per year in the field. Interviewees who went for the other activities often spent three or four months per year in the field. The age composition of the interviewees is shown in Table 1. The *R. bieti* sightings and activity routes collected through the interviews cover a temporal span from the 1950s through to 2010. Limited by the availability of environmental data needed for experiments (Section 2.3.1; Supporting Information S1), only *R. bieti* sightings in three historical periods (i.e., 1973–1981, 1987–2005, and 2006–2010) were examined in this study (Figure 3). The spatial extent of the study area (Figure 3) was determined by considering the spatial coverage of the elicited citizen data (i.e., sightings and routes) and the historical home range of the Mt. Lasha *R. bieti* group.

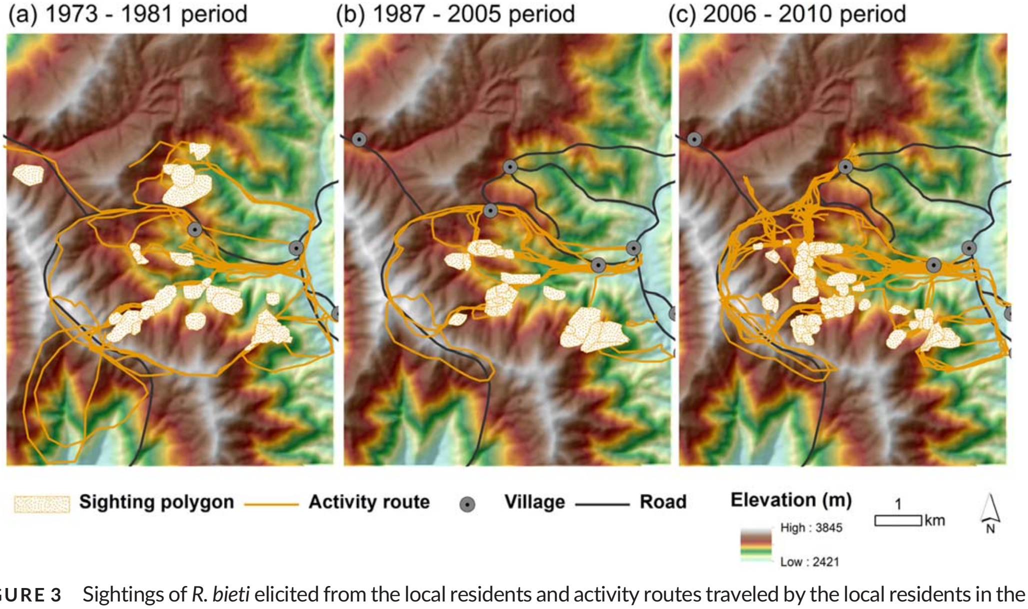

FIGURE 3 Sightings of *R. bieti* elicited from the local residents and activity routes traveled by the local residents in the three historical periods (Section 2.3.1): (a) 26 sightings and 16 routes in 1973–1981 elicited from 20 interviewees aged between 41 and 76; (b) 30 sightings and 20 routes in 1987–2005 elicited from 18 interviewees aged between 29 and 78; and (c) 59 sightings and 45 routes in 2006–2010 elicited from 32 interviewees aged between 19 and 64

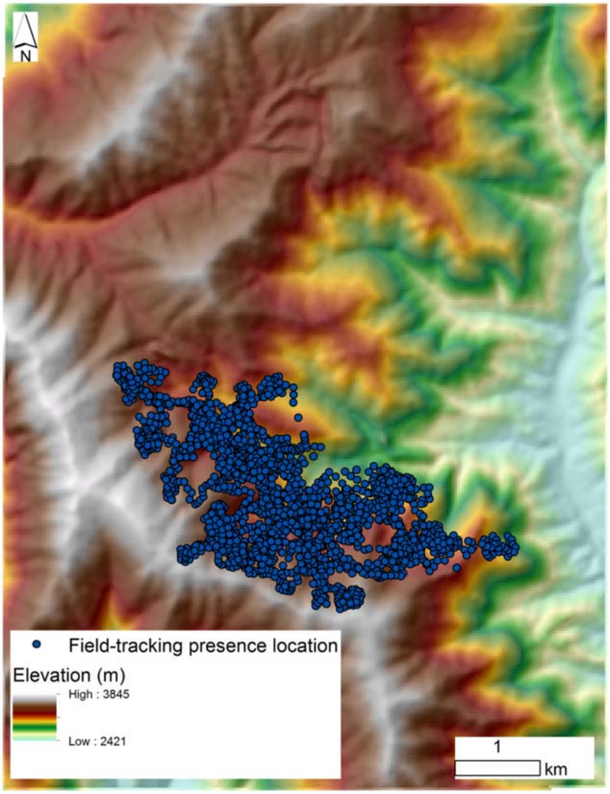

FIGURE 4 Field-tracked *R. bieti* presence locations recorded during field work conducted in 2008 and 2009 (2,707 presence locations)

2.2.2 | Validation data

The presence locations of *R. bieti* were recorded during a field-tracking campaign of the *R. bieti* group in the Mt. Lasha area. These field-tracked *R. bieti* presence locations were used as independent validation data to assess the validity of presence data elicited from the local residents (Section 2.3).

One biologist and two field assistants first spent nearly a year in the field to habituate the monkeys and to familiarize themselves with the surrounding terrain. Field tracking and direct observation of the monkeys were then conducted in 2008 and 2009, with the primary purpose of collecting detailed observational data for behavioral studies (Huang, 2009; Huang, Cui, Scott, Wang, & Xiao, 2012; Huang et al., 2017). The field tracking covered a five-month period in 2008 (May, Jun, Jul, Aug, Oct) and nine months in 2009 (Jan, Feb, Mar, Apr, Jun, Jul, Aug, Sep, Dec). During the field-tracking periods, the locations of the monkeys were recorded every 30 min in the daytime. These field-tracked presence locations consisted of the most detailed data available that reflected the distribution of *R. bieti* in the study area (Figure 4).

2.3 | Evaluation of historical citizen data

2.3.1 | Experiment design

The validity of the *R. bieti* presence data elicited from local residents to reveal the historical *R. bieti* distribution was evaluated by examining the quality of habitat suitability mapping using these presence data. Habitat suitability is mapped by coupling the environmental data layers on factors impacting wildlife habitat use and the relationships between habitat suitability and these factors (*environmental factors* hereafter) (Guisan & Zimmerman, 2000; Hirzel & Lay, 2008). The suitability–environment relationships are first derived from the environmental data values at wildlife presence locations and then applied to the entire environmental data layers to predict a habitat suitability map (Elith et al., 2006; Pearce & Boyce, 2006). The quality of the so-predicted suitability map thus depends on the validity of the wildlife presence data. It is thus fair to say that a high quality of the predicted suitability map indicates a high validity of the presence data.

The decades-long coverage of the elicited *R. bieti* presence data was divided into three shorter historical periods and the validity of the presence data was evaluated through habitat suitability mapping using presence data in each period. To solely examine the impact of the presence data on suitability mapping, temporal variabilities in the environmental data layers across the historical periods were controlled for through two means. *First*, division of the historical periods (Supporting Information S1) was determined by historical events that induced major environmental change in the study area, so that in each period the environmental conditions in the study area were relatively stable. *Second*, areas where there was no significant environmental change across the historical periods (*non-change areas* hereafter) were identified. Non-change areas were areas with neither transition between forest and non-forest nor significant change in human-posed disturbance across the periods (Supporting Information S2). The identified non-change areas take up approximately 60% of the entire study area. Habitat suitability was predicted at locations (pixels) in the non-change areas based on the suitability–environment relationships derived from *R. bieti* presence data falling in these areas.

It is assumed in this study that the underlying habitat requirements of *R. bieti* (i.e., fundamental niche) did not change significantly over the historical periods. This is a reasonable assumption given that the decades-long period is very short from an evolutionary perspective. Under this assumption, it is reasonable to expect that the distribution of *R. bieti* over the non-change areas did not change much across the historical periods. It follows that the independent validation data collected in 2008 and 2009 (Section 2.2.2) should also be able to reflect the distribution of the monkeys in earlier historical periods over the non-change areas. The quality of the suitability map predicted based on the presence data in each period was thus assessed by comparing the suitability map against the independent validation data (Section 2.3.3). If the *R. bieti* presence data elicited from the local residents were of high validity, then the suitability maps across the historical periods are expected to show a consistent spatial pattern that reflects the distribution of *R. bieti* over the non-change areas.

2.3.2 | Habitat suitability mapping

Environmental data

Habitat suitability mapping requires environmental data characterizing the factors impacting habitat use of *R. bieti*. These environmental data are needed to derive the suitability–environment relationships from *R. bieti* presence data and to predict habitat suitability based on the derived relationships. For *R. bieti*, the environmental factors impacting its habitat use include terrain, water source, shelter and food, and human-posed disturbance (Huang, 2009; Long, Kirkpatrick, Zhong, & Xiao, 1996; Xiao et al., 2003; Zhu et al., 2015).

Accordingly, the following environmental data layers were used for habitat suitability mapping: elevation, slope gradient, slope aspect, least-cost distance to rivers, least-cost distance to villages or roads, and the shelter and food factor represented by vegetation type maps. Elevation, slope gradient (%), and slope aspect (discretizing 0–360° evenly into eight categories) were derived from a 30 m resolution DEM of the study area (Supporting Information S2.2). Cost distances were computed based on the DEM and spatial distribution of rivers, roads, and villages. The same set of data layers related terrain condition, and distance to rivers was used throughout as terrain and rivers were stable across the historical periods. Distance to villages or roads was computed based on the DEM and villages/roads that were current to each period. Land-use maps obtained by classifying Landsat images (Supporting Information S2.1) on dates in the same period were aggregated by applying majority cell statistics to obtain an “average” land use/cover map for that period. A vegetation type map was extracted from the land use/cover map to represent the shelter and food factor in each period. All environmental data layers were at 30 m spatial resolution.

Mapping method

The predictive mapping method developed in Zhu et al. (2015) was adopted for mapping *R. bieti* habitat suitability using the elicited presence data. This mapping method was adopted here because of its capability to extract representative presence locations from the imprecise polygonal sighting data and its flexibility to incorporate observation effort information to compensate for spatial bias in the presence data. Details of the mapping method are beyond the scope of this article, and can be found in Zhu et al. (2015), but a brief overview of the mapping method is provided below.

Wildlife sightings elicited from local residents had only imprecise location information as they were depicted using sighting polygons (Section 2.2.1). But this does not mean that the wildlife showed up at every location in the polygon area. It is assumed that the typical environmental condition over the polygon area best approximates the environmental condition where the wildlife would occur. Under this assumption, the mapping method uses a frequency sampling strategy to extract representative presence locations from sighting polygons based on the frequency distributions of environmental data. The frequency sampling strategy includes three operational steps. First, for pixels within each sighting polygon, the frequency distributions (histograms) with respect to each of the environmental variables (e.g., elevation, slope, etc.) were constructed, and the pixels at which the environmental value falls into the modal interval (the interval of environmental values with the highest frequency) of each frequency distribution were located. Second, the pixels whose environmental values simultaneously fall into the respective modal intervals were identified as the representative presence locations of this sighting polygon. Third, representative locations identified in each sighting polygon were pooled together to form a full set of representative presence locations, which were used in combination with the environmental data to derive the suitability–environment relationships.

Sightings elicited from the local residents are also likely to be spatially biased. This is because not every location in the landscape can be equally observed from the routes taken by the local residents, given the irregular and non-random distribution of their routes and the variability of the terrain conditions. The mapping method computes cumulative visibility based on a DEM and the elicited activity routes as an approximation of observation effort. The mapping method then compensates for the spatial bias in the presence data through inversely weighting the representative presence locations by cumulative visibility at that location in deriving suitability–environment relationships from presence data.

The mapping method derives suitability–environment relationships using probability density functions (PDFs) of *R. bieti* presence locations over environmental conditions. The PDF of *R. bieti* presence over each environmental variable was estimated using kernel density estimation (Silverman, 1986) based on the environmental data values at the representative *R. bieti* presence locations (extracted from sighting polygons used through the frequency sampling strategy). Each representative presence location was inversely weighted by the cumulative visibility at that location to compensate for spatial bias when estimating the PDFs. Specifically, the following equation was used to estimate the PDF with respect to each environmental variable:

$$f(x)=\frac{1}{h_x}\sum_{i=1}^{n}\left[\frac{w_i}{\sum w_i}K\left(\frac{x-x_i}{h_x}\right)\right] \tag{1}$$

where $f(x)$ is the estimated PDF for environmental variable x, x_i is the value of environmental variable x at presence location i (n presence locations in total), w_i is the weight of presence location i and it was used to compensate for the spatial bias. w_i is inversely proportional to the cumulative visibility at location i. K is a kernel function for which the Gaussian kernel was adopted, and h_x is the bandwidth determined using the "rule-of-thumb" algorithm (Silverman, 1986).

Based on the presence PDFs on individual environmental variables, habitat suitability at each location (i.e., pixel) over the non-change areas was calculated in two steps in a rule-based fashion (Zhu, 2008). First, habitat suitability with respect to each environmental variable was computed as the normalized density value on the presence PDF curve corresponding to that variable, given the environmental value at this location, using the following equation:

$$S_j^x=\frac{f(x_j)}{\max(f(x))} \tag{2}$$

where x_j refers to the value of environmental variable x at location j, $\max(f(x))$ refers to the maximum value of the PDF, and S_j^x refers to the habitat suitability with respect to environmental variable x at location j. Second, following the limiting factor theory in ecology, the overall habitat suitability at this location was computed as the minimum value among the suitabilities to all environmental variables using the following equation:

$$S_j = \min\left(S_j^1, S_j^2, \ldots, S_j^m\right) \tag{3}$$

where m refers to the number of environmental variables.

As this study analyzes the same types of data from the same study area, the parameter setting suggestions of Zhu et al. (2015) were followed to set parameters for the mapping method (e.g., distance threshold used for computing cumulative visibility, number of representative presence locations selected in each sighting polygon, bandwidth determination in kernel density estimation, etc.).

2.3.3 | Evaluation and assessment

As discussed in Section 2.3.1, if the *R. bieti* presence data elicited from local residents were of high validity, then the predicted suitability maps in the three periods are expected to show a consistent spatial pattern (consistency) that reflects the distribution of *R. bieti* over the non-change areas (accuracy). Spearman's rank correlation coefficient ρ between pixel-wise suitability values on two maps was computed as a quantitative indicator of the consistency between suitability maps. A positive ρ indicates that the spatial patterns on two suitability maps are consistent.

The accuracy of each suitability map was evaluated through associating the field-tracked *R. bieti* presence locations with the suitability map by calculating the Boyce index (Boyce, Vernier, Nielsen, & Schmiegelow, 2002). Habitat suitability values on the map were divided into 10 bins using an "equal area" approach: suitability bins were determined in such a way that pixels falling into the 10 suitability bins were of an approximately equal total area (Boyce et al., 2002). The Boyce index is Spearman's rank correlation between the area-adjusted frequency of observed *R. bieti* presence locations within individual bins and the bin rank. Area-adjusted frequencies are the frequency of observed *R. bieti* presence locations within a bin divided by the area of that range of suitability values available across the study area. A suitability map that agrees with the distribution reflected in the validation data should have a positive Boyce index close to 1.0, as more observed presence locations (area-adjusted) would continually be falling within higher suitability bins (Boyce et al., 2002). The field-tracked *R. bieti* presence data that fell within the non-change areas were used to compute the Boyce index for suitability maps in all three historical periods.

3 | RESULTS

Habitat suitability maps predicted over the non-change areas using citizen data falling in the non-change areas in each historical period are shown in Figure 5. The three suitability maps show a consistent pattern of spatial variation of *R. bieti* habitat suitability: areas located within mid- to high elevation range and on the northeast hillslopes were predicted to be of relatively high suitability; areas over ridges, valleys, and southwest hillslopes were predicted to be of low suitability. Spearman's rank correlation coefficient ρ between any two of the three suitability maps was above 0.60 (0.62–0.66), which confirms that the suitability maps predicted in all three periods show a consistent spatial pattern. This in turn suggests that the elicited *R. bieti* presence data indeed captures a consistent spatial pattern of *R. bieti* habitat suitability across the three periods.

The accuracy of the three suitability maps was evaluated by calculating the Boyce index using the field-tracked *R. bieti* presence locations. The Boyce index of suitability maps predicted in the three periods was 0.89, 0.93, and 0.86, respectively. This indicates that the predicted suitability maps are highly consistent with the field-tracked *R. bieti* presence locations that reflect the distribution of the monkeys over the mapped areas. This demonstrates that the *R. bieti* presence data elicited from local residents has validity to reveal the *R. bieti* distribution in historical periods.

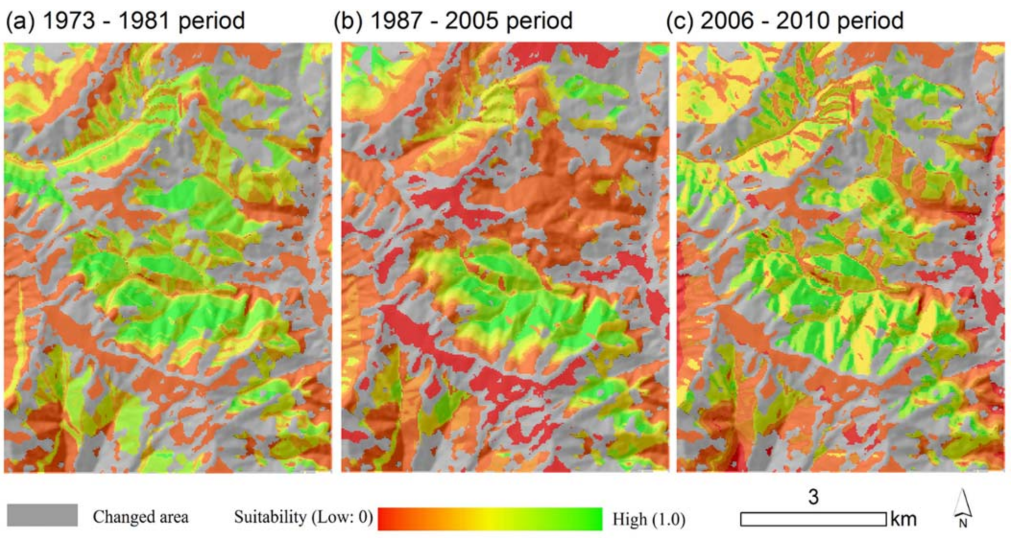

FIGURE 5 Habitat suitability maps predicted over the non-change areas using *R. bieti* presence data in each period falling in the non-change areas (based on 892, 666, and 1,000 representative *R. bieti* presence locations for the three periods, respectively)

4 | DISCUSSION

4.1 | Validity of historical citizen data

Citizens possess abundant knowledge of geographic phenomena accumulated throughout their living history in local areas (Goodchild, 2007). As already shown in other studies, citizen data are valid to reveal geographic phenomena in recent periods and can be very useful in various applications (Anadón et al., 2009, 2010; Ma et al., 2014; Zhu et al., 2015). This study presents an evaluation of the validity of citizen data to reveal geographic phenomena in historical periods. As the *R. bieti* habitat suitability mapping case study reported in this study has shown, citizen data are also valid to reveal historical geographic phenomena that occurred decades ago.

Citizen data are not expected to be as accurate and comprehensive as data collected through rigorous protocols. For example, detailed animal behavior data can be recorded by specialized biologists in field tracking. Sightings elicited from local residents often cannot provide such specifics. But even citizen data at this level of detail could be of great use for conservation practices (Anadón et al., 2009, 2010; Ma et al., 2014). Moreover, citizen data could potentially cover large areas over long periods, while rigorous protocols (e.g., biological survey, tracking) are only practicable in geographic areas and within very short time frames.

One should be fully aware of the limitations of citizen data and the necessary analytical treatments for these limitations when analyzing citizen data. For example, citizen data on wildlife sightings elicited from the local residents have only imprecise location information and biased spatial coverage. Therefore, appropriate analytical procedures need to be applied to increase the location precision and to minimize the effects of spatial bias when using citizen data in applications (e.g., Zhu et al., 2015).

4.2 | *R. bieti* habitat change across historical periods

The evaluation conducted in this study has shown that the elicited *R. bieti* presence data were valid to reveal the *R. bieti* distribution in historical periods over the non-change areas. Built upon this observation, the presence data were used to derive the relationships between *R. bieti* habitat suitability (i.e., realized niche) and environmental conditions (Figure 6) and to map *R. bieti* habitat suitability over the entire study area in each period (Figure 7). The *R. bieti* habitat change across the historical periods was examined.

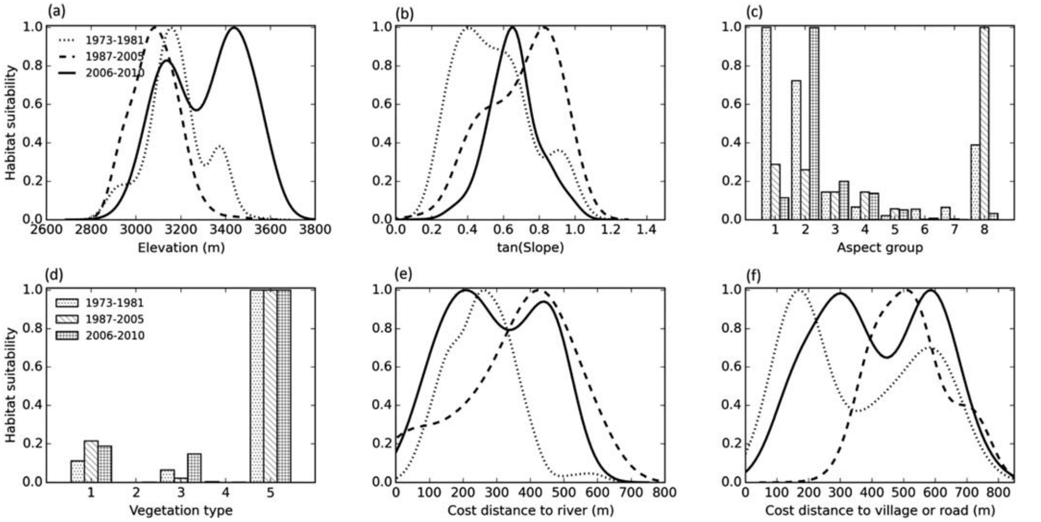

FIGURE 6 Suitability–environment relationships derived from *R. bieti* presence data in each period. Aspect group 1: 0–45° (starting from North), 2: 45–90°, 3: 90–135°, 4: 135–180°, 5: 180–225°, 6: 225–270°, 7: 270–315°, 8: 315–360°. Vegetation type 1: Evergreen Coniferous, 2: Pasture, 3: Yunnan Pine, 4: Farmland, 5: Deciduous Broadleaf

The spatial distribution pattern of high-suitability patches was generally consistent across the three periods. High-suitability patches were found in forests (Figure 6d) within the mid- to high elevation range (Figure 6a) on the north-east hillslopes (Figure 6c). Overall, high-suitability patches were reduced in the 1987–2005 period compared with the 1973–1981 period. For example, there was an area (outlined on Figure 7b) in the 1987–2005 period that has much lower suitability values compared with the previous period. This was mainly due to the establishment of new village settlements and roads, which induced significant human disturbance in this area in the 1987–2005 period (Figure 3). But there was a tendency for high-suitability patches to recover in the 2006–2010 period. The outlined area recovered to relatively high suitability in the 2006–2010 period, as the monkeys became more tolerant of proximity to villages and roads (Figure 6f).

The recovering trend found in the 2006–2010 period was confirmed by examining the relationships between *R. bieti* habitat suitability and individual environmental variables (Figure 6). In the 2006–2010 period, the evaluations of

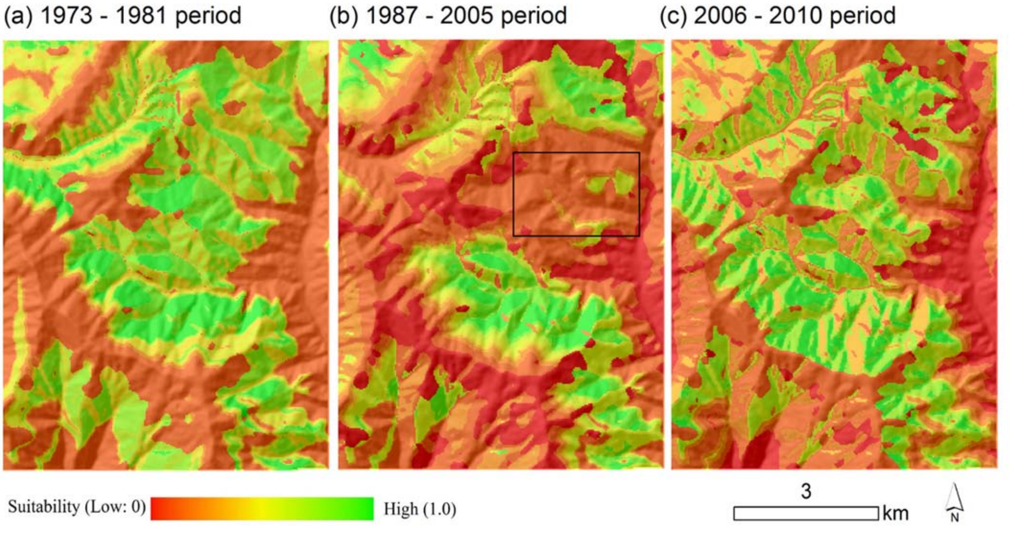

FIGURE 7 Habitat suitability maps predicted for the entire study area using *R. bieti* presence data in each period (based on 966, 696, and 1,061 representative *R. bieti* presence locations for the three periods, respectively)

high suitability shifted back to higher ranges that were comparable with (and even higher than) those in the 1973–1981 period (Figure 6a). The distance ranges (distance to rivers, distance to village or road) of high suitability shifted back to similar ranges in the 1973–1981 period (Figures 6e and f). These might be indicators that conservation practices initiated by the protected area were effective at restoring *R. bieti* habitat in the study area.

4.3 | Considerations on citizen data elicitation

Several factors should be taken into consideration when eliciting citizen data using the 3D geovisualization interview approach. 3dMapper demands high-resolution DEM and satellite imagery to help georeference citizen data as accurately as possible. High-resolution DEM and satellite imagery are now increasingly available in digital format for most parts of the world (Ma et al., 2014; Toutin, 2004; van Zyl, 2001). For areas where such data are not readily available, additional efforts on data preparation are needed. In this study, a high-resolution DEM for the Mt. Lasha study area was created based on contours digitized from hardcopy topographic maps and high-resolution satellite imagery was obtained from Google Earth.

Compared with 2D topographic maps or aerial photos, 3dMapper can better aid the interviewees to recall and locate wildlife sightings. Given a 2D representation, relief interpretation skills are needed to construct the 3D topography of the landscape based on which orientation and localization can be achieved. However, the local residents often lack such skills, as they do not have much training in map reading. 3D geovisualization facilitates relief interpretation by offering a more realistic and intuitive representation of the terrain (Carbonell Carrera & Bermejo Asensio, 2017) and improves the efficiency of visual search and navigation performance (Liao, Dong, Peng, & Liu 2016).

3dMapper also has some limitations, as learned from our field experience. First, it takes time for the interviewees to become familiar with 3dMapper. It took 20–30 min to train most interviewees (~65%) before they could orientate and accurately locate places using 3dMapper. Second, some interviewees (~35%), such as the elders, had difficulties using 3dMapper due to vision or map-reading problems. In such cases, the field assistant had to locate *R. bieti* sightings and activity routes in 3dMapper based on the place names and micro-terrain feature descriptions provided by the interviewee. This extra step may introduce uncertainty, depending on the accuracy of communication. Finally, walking around the area with the interviewees might be a better way of eliciting memories than conducting sit-down interviews using 3dMapper. We did not adopt this option due to the following considerations. The local residents are already very familiar with the terrain of the area as a result of their frequent field activities. Moreover, walking around the area with the interviewees would be logistically challenging and much more time-consuming.

The validity of the elicited citizen data depends on the availability and trustworthiness of the informants. Local residents whose livelihoods are closely related to ecosystem services are ideal informants for eliciting citizen data in their respective areas. Some local residents might be reluctant to share their knowledge, due to reasons such as conflict of economic interests. In such cases sociological tools, such as semi-direct or in-depth interviews, can be used to partially overcome this issue (Anadón et al., 2009).

The validity of the elicited citizen data also depends on the characteristics of the target geographic phenomenon under observation. For instance, local residents can often only observe diurnally active wildlife. The target wildlife should be easily recognizable to reduce misidentification, as local residents often have little training on species identification (Anadón et al., 2009). In this study, *R. bieti* is an unmistakable species with a strong historic dimension in local communities (Long et al., 1994). Only information on *R. bieti* presence was obtained. If more involved information (e.g., abundance) is to be obtained from citizen data, then additional measures are needed to ensure reliability of the elicited data and to reduce uncertainty of the inferences drawn from the data (Anadón et al., 2009, 2010).

Efforts need to be made in the elicitation process to reduce the temporal imprecision of citizen data. It could be challenging for interviewees to recall the exact timing of geographic phenomena observed decades ago. In this study, the year of *R. bieti* sightings recalled by the interviewees was cross-checked and refined with reference to the timing of major events (e.g., national policy implementation, marriage, child born, etc.) and the month to seasonal activity patterns in the area (e.g., farming). Temporal information on elicited sightings should be accurate enough to support the coarse-gain analyses performed across historical periods.

The elicitation of historical citizen data relies heavily on the memories of the interviewees, because they kept no records at the time of observation. In this regard, historical citizen data differs greatly from VGI generated using the enabling technologies of recent periods. For example, while geotagged tweets and photos can be archived and their quality does not decay even after many years, local residents may easily forget about their encounters with commonly seen wildlife in the field as time elapses. As a result, historical citizen data might only be useful to reveal geographic phenomena that are significant enough to “impress” people and maintain long-lasting memories. Thus, before historical citizen data elicitation, one should first assess the “significance” of the geographic phenomenon of interest. In this study, while other common species may not be as memorable to the local residents, sightings of *R. bieti* certainly would put a dent in their memories because of the strong historic dimension of the monkeys in local communities (Long et al., 1994).

5 | CONCLUSIONS

This article presents an evaluation of the validity of citizen data to reveal geographic phenomena in historical periods. The evaluation was demonstrated through a case study of citizen data-based habitat suitability mapping in historical periods. *R. bieti* sightings (i.e., citizen data) were elicited from the local residents through 3D geovisualization interviews in the Mt. Lasha study area in Yunnan, China. The validity of the citizen data to reveal the historical *R. bieti* distribution was evaluated through habitat suitability mapping using citizen data in historical periods. The results of controlled experiments showed that suitability maps predicted using citizen data had a consistent spatial pattern that reflects the distribution of *R. bieti* across historical periods. This in turn suggests that citizen data have validity for mapping historical geographic phenomena.

The evaluation provides supporting empirical evidence for broadening the temporal scope of VGI applications. With the validity of historical VGI evaluated, VGI can be used to support a wider range of applications requiring insights on geographic phenomena in historical periods (e.g., examining historical wildlife distributions).

ACKNOWLEDGMENTS

The work reported here was supported by grants from the National Natural Science Foundation of China (Project Nos. 41431177, 31560118, 30960085), the National Basic Research Program of China (Project No. 2015CB954102), the Natural Science Research Program of Jiangsu (14KJA170001), PAPD, and the National Key Technology Innovation Project for Water Pollution Control and Remediation (Project No. 2013ZX07103006). Support for Guiming Zhang through the Whitbeck Graduate Dissertator Award from the Department of Geography, University of Wisconsin-Madison is appreciated. Support for A-Xing Zhu through the Vilas Associate Award, the Hammel Faculty Fellow Award, the Manasse Chair Professorship from the University of Wisconsin-Madison, and the “One-Thousand Talents” Program of China is greatly appreciated. Cheng-Zhi Qin is grateful for support through the “Excellent Young Scholars” project from National Natural Science Foundation of China. We thank Mr. Jack (John) Keel for proofreading the manuscript.

ORCID

Guiming Zhang http://orcid.org/0000-0001-7064-2138

REFERENCES

Ali, A. L., & Schmid, F. (2014). Data quality assurance for volunteered geographic information. In M. Duckham, E. Pebesma, K. Stewart, & A.U. Frank, (Eds.), *Proceedings of the Eighth International Conference on Geographic Information Science* (pp. 126–141). Vienna, Austria: Springer Lecture Notes in Computer Science Vol. 8728.

Anadón, J. D., Giménez, A., & Ballestar, R. (2010). Linking local ecological knowledge and habitat modelling to predict absolute species abundance on large scales. *Biodiversity & Conservation*, *19*, 1443–1454.

Anadón, J. D., Giménez, A., Ballestar, R., & Pérez, I. (2009). Evaluation of local ecological knowledge as a method for collecting extensive data on animal abundance. *Conservation Biology*, *23*, 617–625.

Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: An overview. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *II-3/W5*, 345–351.

Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, *157*, 281–300.

Brovelli, M. A., Minghini, M., & Zamboni, G. (2015). Public participation in GIS via mobile applications. *ISPRS Journal of Photogrammetry & Remote Sensing*, *114*, 306–315.

Burt, J. E., & Zhu, A.-X. (2004). *3dMapper 4.02*. Retrieved from http://solim.geography.wisc.edu

Carbonell Carrera, C., & Bermejo Asensio, L. A. (2017). Augmented reality as a digital teaching environment to develop spatial thinking. *Cartography & Geographic Information Science*, *44*, 259–270.

Carnaval, A. C., & Moritz, C. (2008). Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, *35*, 1187–1201.

Danielsen, F., Mendoza, M. M., Alviola, P., Balete, D. S., Enghoff, M., Poulsen, M. K., & Jensen, A. E. (2003). Biodiversity monitoring in developing countries: What are we trying to achieve?. *Oryx*, *37*, 407–409.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., . . . Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129–151.

Fonte, C. C., Bastin, L., See, L., Foody, G., & Lupia, F. (2015). Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science*, *29*, 1269–1291.

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an Internet based collaborative project. *Transactions in GIS*, *17*, 847–860.

Franklin, J., & Miller, J. A. (2009). *Mapping species distributions: Spatial inference and prediction*. Cambridge, UK: Cambridge University Press.

Gao, S., Janowicz, K., Montello, D. R., Hu, Y., Yang, J. A., McKenzie, G., . . . Yan, B. (2017). A data-synthesis-driven method for detecting and extracting vague cognitive regions. *International Journal of Geographical Information Science*, *31*, 1245–1271.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*, 211–221.

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth*, *3*, 231–241.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, *19*, 497–503.

Guisan, A., & Zimmerman, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186.

Haklay, M., & Weber, P. (2008). OpenStreetMap: User-generated street maps. *IEEE Pervasive Computing*, *7*, 12–18.

Hanberry, B. B., He, H. S., Palik, B. J., & He, H. S. (2012). Comparing predicted historical distributions of tree species using two tree-based ensemble classification methods. *American Midland Naturalist*, *168*, 443–455.

Hirzel, A. H., & Lay, G. L. (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, *45*, 1372–1381.

Huang, Z. P. (2009). *Foraging, reproduction and sleeping site selection of black-and-white snub-nosed monkey (Rhinopithecus bieti) at the southern range* (Unpublished M.S. Thesis). Southwest Forestry University, Kunming, China.

Huang, Z. P., Cui, L. W., Scott, M., Wang, S. J., & Xiao, W. (2012). Seasonality of reproduction of wild black-and-white snub-nosed monkeys (*Rhinopithecus bieti*) at Mt. Lasha, Yunnan, China. *Primates*, *53*, 237–245.

Huang, Z. P., Scott, M. B., Li, Y. P., Ren, G. P., Xiang, Z. F., Cui, L. W., & Xiao, W. (2017). Black-and-white snub-nosed monkey (*Rhinopithecus bieti*) feeding behavior in a degraded forest fragment: Clues to a stressed population. *Primates*, *58*, in press.

Huntington, H. P. (2000). Using traditional ecological knowledge in science: Methods and applications. *Ecological Applications*, *10*, 1270–1274.

IUCN. (2016). *The IUCN red list of threatened species version 2016–3*. Retrieved from http://www.iucnredlist.org

Jackson, S. P., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., & Stefanidis, A. (2013). Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information*, *2*, 507–530.

Liao, H., Dong, W., Peng, C., & Liu, H. (2016). Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography & Geographic Information Science*, *44*, 474–490.

Long, Y. C., Kirkpatrick, C. R., Zhong, T., & Xiao, L. (1994). Report on the distribution, population, and ecology of the Yunnan snub-nosed monkey (*Rhinopithecus bieti*). *Primates*, *35*, 241–250.

Long, Y. C., Kirkpatrick, C. R., Zhong, T., & Xiao, L. (1996). Status and conservation strategy of the Yunnan snub-nosed monkey. *Chinese Biodiversity*, *4*, 145–152.

Ma, C., Huang, Z. P., Zhao, X. F., Zhang, L. X., Sun, W. M., Matthew, B. S., ... Xiao, W. (2014). Distribution and conservation status of *Rhinopithecus strykeri* in China. *Primates*, *55*, 377–382.

Mackinson, S. (2001). Integrating local and scientific knowledge: An example in fisheries science. *Environmental Management*, *27*, 533–545.

Mackinson, S., & Nottestad, L. (1998). Combining local and scientific knowledge. *Reviews in Fish Biology & Fisheries*, *8*, 481–490.

Moudrý, V., & Šímová, P. (2012). Influence of positional accuracy, sample size and scale on modelling species distributions: A review. *International Journal of Geographical Information Science*, *26*, 2083–2095.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853–858.

Nogués-Bravo, D. (2009). Predicting the past distribution of species climatic niches. *Global Ecology & Biogeography*, *18*, 521–531.

Osborne, P. E., & Leitão, P. J. (2009). Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity & Distributions*, *15*, 671–681.

Pearce, J. L., & Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, *43*, 405–412.

Pelletier, T. A., Crisafulli, C., Wagner, S., Zellmer, A. J., & Carstens, B. C. (2015). Historical species distribution models predict species limits in western Plethodon salamanders. *Systematic Biology*, *64*, 909–925.

Peterson, A. T., & Anamza, T. (2015). Ecological niches and present and historical geographic distributions of species: A 15-year review of frameworks, results, pitfalls, and promises. *Folia Zoologica*, *64*, 207–217.

Rossiter, D. G., Liu, J., Carlisle, S., & Zhu, A.-X. (2015). Can citizen science assist digital soil mapping? *Geoderma*, *259&260*, 71–80.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., ... Rutzinger, M. (2016). Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information*, *5*, 55.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*, 139–167.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London, UK: Chapman & Hall.

Soberón, J., & Nakamura, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences, USA*, *106*, 19644–19650.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*, 2282–2292.

Toutin, T. (2004). Comparison of stereo-extracted DTM from different high-resolution sensors: SPOT-5, EROS-A, IKONOS-II, and QuickBird. *IEEE Transactions on Geoscience & Remote Sensing*, *42*, 2121–2129.

van Zyl, J. J. (2001). The shuttle radar topography mission (SRTM): A breakthrough in remote sensing of topography. *Acta Astronautica*, *48*, 559–565.

Werneck, F. P., Costa, G. C., Colli, G. R., Prado, D. E., & Sites, J. W. (2011). Revisiting the historical distribution of seasonally dry tropical forests: New insights based on palaeodistribution modelling and palynological evidence. *Global Ecology & Biogeography*, *20*, 272–288.

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, *19*, 639–644.

Xiao, W., Ding, W., Cui, L. W., Zhou, R. L., & Zhao, Q. K. (2003). Habitat degradation of *Rhinopithecus bieti* in Yunnan, China. *International Journal of Primatology*, *24*, 389–398.

Zhu, A.-X. (2008). Rule-based mapping. In J. P. Wilson & A. S. Fotheringham (Eds.), *Handbook of geographic information science* (pp. 273–291). Hoboken, NJ: Wiley-Blackwell.

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

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy*, *2*, 6–32.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: 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;22:149–164. https://doi.org/10.1111/tgis.12300