Received: 10 April 2017 | Revised: 20 August 2017 | Accepted: 21 August 2017
DOI: 10.1111/tgis.12303

RESEARCH ARTICLE

WILEY Transactions in GIS

A heuristic-based approach to mitigating positional errors in patrol data for species distribution modeling

Guiming Zhang[1] | **A-Xing Zhu**[1,2,3,4,5] | **Zhi-Pang Huang**[6,7] | **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, 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, Nanjing, 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
Zhi-Pang Huang and Wen Xiao, Institute of Eastern-Himalaya Biodiversity Research, Dali University, Dali 671003, China.
Emails: huangzp@eastern-himalaya.cn; xiaow@eastern-himalaya.cn

Funding information
National Natural Science Foundation of China, Grant/Award Nos. 31560118, 30960085, 41431177; 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

Species distribution modeling (SDM) at fine spatial resolutions requires species occurrence data of high positional accuracy to achieve good model performance. However, wildlife occurrences recorded by patrols in ranger-based monitoring programs suffer from positional errors, because recorded locations represent the positions of the ranger and differ from the actual occurrence locations of wildlife (hereinafter referred to as positional errors in patrol data). This study presented an evaluation of the impact of such positional errors in patrol data on SDM and developed a heuristic-based approach to mitigating the positional errors. The approach derives probable wildlife occurrence locations from ranger positions, utilizing heuristics based on species preferred habitat and the observer's field of view. The evaluations were conducted through a case study of SDM using patrol records of the black-and-white snub-nosed monkey (*Rhinopithecus bieti*) in Yunnan, China. The performance of the approach was also compared against alternative sampling methods. The results showed that the positional errors in *R. bieti* patrol data had an adverse effect on SDM performance, and that the proposed approach can effectively mitigate the impact of the positional errors to greatly improve SDM performance.

1 | INTRODUCTION

Species distribution modeling (SDM) (Franklin & Miller, 2009) is widely used to predict the geographic distribution of species to support decision making in conservation (e.g., managing biological invasions, identifying and protecting critical habitats, reserve selection, translocation, etc.) (Guisan et al., 2013). SDM requires two inputs: environmental data layers characterizing the spatial variation of environmental conditions (i.e., maps of environmental variables) and species data indicating the occurrence or abundance of species at sampled sites. With the rapid development of geospatial technologies such as global positioning systems (GPS), geographic information systems (GIS), and remote sensing, environmental data can now be obtained easily and in a timely manner (Gillespie, Foody, Rocchini, Giorgi, & Saatchi, 2008; Kerr & Ostrovsky, 2003; van Zyl, 2001; Viña, Bearer, Zhang, Ouyang, & Liu, 2008).

The collection of species data, particularly wildlife occurrence data, often requires much more effort than environmental data. Generally, wildlife occurrences are collected using techniques such as transects and distance sampling (Anderson, Laake, Crain, & Burnham, 1979; Buckland et al., 2001), radio telemetry (Campbell & Sussman, 1994), infrared trapping cameras (Burton, Sam, Balangtaa, & Brashares, 2012; Trolle & Kéry, 2003), and GPS collars (Hemson et al., 2005). Wildlife occurrence data collected through these techniques can be of high positional accuracy. However, these techniques are often prohibitively expensive for conservation programs with limited budgets. The high cost renders them impractical and unsustainable for conservation programs in poorer and remote regions of the world (Danielsen et al., 2003), where most of the world's biodiversity hotspots occur (Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000).

Ranger-based monitoring has been recognized as a cost-effective practical alternative to sustain conservation with limited supporting funds (Danielsen et al., 2003, 2005; Gray & Kalpers, 2005; Hilborn et al., 2006; Jachmann, 2008; Plumptre et al., 2014; Poulsen & Luanglath, 2005). Many protected areas (e.g., nature reserves and national parks) have set up routine patrols to monitor wildlife population, poaching, and illegal deforesting, and record encounters of wildlife during the patrol (Burton, 2010; Gray & Kalpers, 2005; Hilborn et al., 2006; Walpole, 2002). Ranger-based monitoring data, although subject to certain pitfalls such as nonrandom sampling (Keane, Jones, & Milner-Gulland, 2011), have been widely used to support management decisions (Danielsen et al., 2005; Gray & Kalpers, 2005), including estimating wildlife population size (Plumptre et al., 2016), evaluating patrol effectiveness (Nyirenda, 2012), and improving law-enforcement effectiveness and efficiency in protected areas (Critchlow et al., 2016; Plumptre et al., 2014). Patrol records containing wildlife encounters could potentially provide wildlife occurrence data for species distribution modeling. Yet, to the best of our knowledge, there are few attempts to use ranger-based data for species distribution modeling—except for Plumptre et al. (2016), where patrol data were used to estimate wildlife presence probability through occupancy analysis, but at a rather coarse spatial resolution (5 km).

One key factor that affects SDM accuracy is the positional accuracy of species occurrence data. Species occurrence locations are used to extract in-situ environment conditions from environmental data layers for SDM. Positional errors in species occurrence data would result in mismatching in-situ environmental conditions and thus degrade the accuracy of SDM (Fernandez, Blum, Reichle, Holzman, & Hamilton, 2009; Graham et al., 2008; Johnson & Gillingham, 2008; Moudrý & Šímová, 2012; Osborne & Leitão, 2009).

Wildlife occurrence data from ranger-based monitoring programs (e.g., patrol records) suffer from positional errors. Encountering wildlife, the patrol can often only record their own position using a handheld GPS receiver. The recorded location is not the actual wildlife occurrence location, as the wildlife would not allow the patrol to approach that closely. The difference between the recorded location and the actual wildlife occurrence location is referred to as the positional error in patrol data in this study. The distance and direction relating the two locations were often not documented, thus it is difficult to correct for the positional errors to recover the actual wildlife occurrence location.

Although some SDM could achieve good predicting performance in the presence of positional errors in species occurrences, such "robustness" only exists to certain levels of positional errors and for SDM at coarser spatial scales (e.g., 1 km spatial resolution) (Graham et al., 2008; Johnson & Gillingham, 2008). Many conservation actions require SDM at finer spatial scales (e.g., 30 m spatial resolution) to support management decisions, for example, translocation

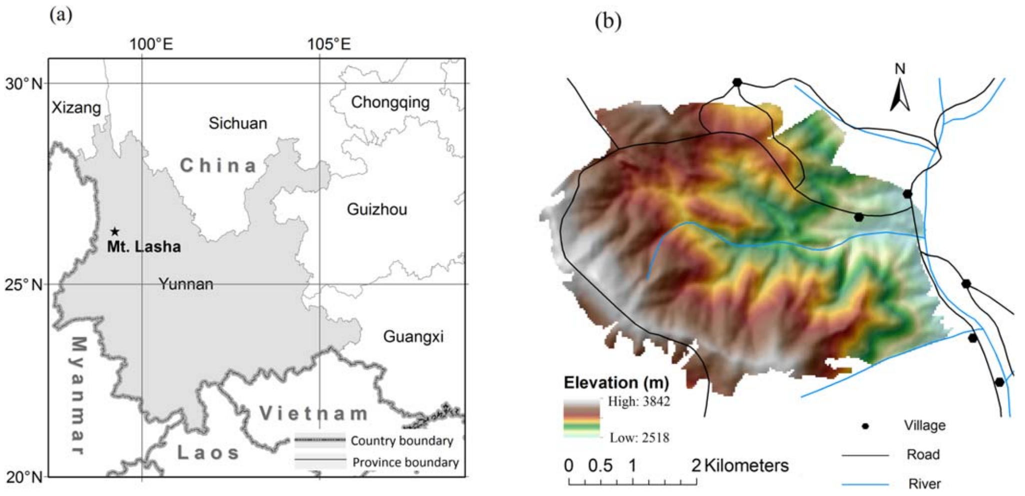

FIGURE 1 (a) Location of the Mt. Lasha study area in Yunnan, China; (b) map of the Mt. Lasha study area

of threatened populations (Guisan et al., 2013) and habitat management/restoration in parks or nature reserves. Moreover, many SDM efforts are conducted at finer spatial scales (Franklin & Miller, 2009), as high-resolution environmental data are increasingly available (Kerr & Ostrovsky, 2003; van Zyl, 2001). At such fine spatial scales, the positional errors in wildlife occurrences from ranger-based monitoring might be far too significant to neglect.

This study presents a heuristic-based approach for mitigating the positional errors in patrol data for SDM. The approach derives probable wildlife occurrence locations from patrol positions utilizing heuristics based on species preferred habitat and the observer's field of view. The aims of the study were to examine the impact of the positional errors in patrol data on SDM and to evaluate the effectiveness of the proposed approach. The evaluations were conducted through a case study using 3-year patrol data on the black-and-white snub-nosed monkey (*Rhinopithecus bieti*) in Yunnan, China. The accuracy of SDM using probable *R. bieti* occurrences derived from the approach was compared to the accuracy of SDM using the raw *R. bieti* patrol records. The performance of the approach was also compared against alternative sampling methods. Experiments were run on various partitions of the patrol data (i.e., year, season) to investigate potential annual and seasonal variability of the positional errors.

2 | MATERIALS AND METHODS

2.1 | Study area and species

R. bieti is categorized as *Endangered* on the International Union for Conservation of Nature Red List (IUCN, 2016). *R. bieti* is endemic to the eastern Himalayas in northwest Yunnan and southeast Tibet, China, between the upper Mekong and Yangtze Rivers (Long, Kirkpatrick, Zhong, & Xiao, 1994; Xiao, Ding, Cui, Zhou, & Zhao, 2003). *R. bieti* has two levels of social organization, where the monkeys form one-male multi-female units (OMUs) and all-male units (AMUs), and multiple OMUs and AMUs travel together in a cohesive band (Kirkpatrick, Long, Zhong, & Xiao, 1998).

The study area was located at Mt. Lasha (99°15′E, 26°20′N) in northwest Yunnan, China (Figure 1). Mt. Lasha is near the southernmost part of its geographic range (Long et al., 1994; Xiao et al., 2003). The 20.3 km^2 study area is part of the Yunling Nature Reserve, and it is an important habitat for a group of *R. bieti* consisting of approximately 100 individuals in 11 OMUs and 2 AMUs (Huang, Cui, Scott, Wang, & Xiao, 2012). The elevation ranges from about 2,500 to 4,000 m in the study area. The high-elevation ridgeline surrounding the area is largely deforested and used as pasture by the local villagers. The low-elevation areas in the east are mainly farmland near villages. From lower elevation to higher elevation, the vegetation transitions from deciduous broadleaved forest to dark conifer forest with mixed deciduous conifer forest in between. Several fire-induced forest clearings were present throughout the area (Huang, 2009; Huang et al., 2012, 2017).

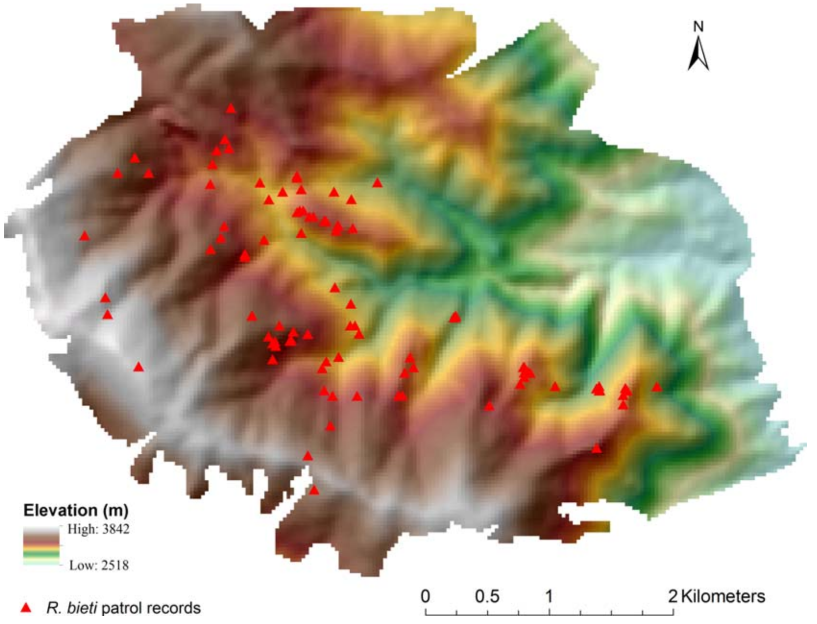

FIGURE 2 *R. bieti* occurrence locations extracted from patrol records in the 2007–2009 period

2.2 | Patrol data

Yunling Nature Reserve has initialized a ranger-based monitoring program since its establishment in 2006. Local villagers are employed and trained as forest rangers to carry out routine patrols in the protected area. The primary goals of patrols are to monitor illegal activities (e.g., deforestation and poaching) in the reserve. In the Mt. Lasha area, one forest ranger has been responsible for the routine patrol since August 2006. He regularly carried out patrols 5 days a month on three main routes and a series of secondary routes covering the whole area. Patrol forms were used to record observed forest conditions on the routes and encounters of wildlife (mostly *R. bieti*). Encounters with wildlife were visually confirmed with the naked eye or using a telescope where necessary. A handheld GPS receiver was used to obtain geographic coordinates (latitude and longitude) of the location at which the wildlife was sighted. The patrol date, geographic coordinates, species name, number of wildlife encountered, behaviors of the wildlife, and habitat type (e.g., bare rocks, grassland, forest type with dominant tree species, etc.) were recorded on a patrol form for each wildlife encounter. No information was recorded on patrol forms indicating the distance and direction between the patrol position and the wildlife occurrence location.

Patrol records in 2006 did not cover a full year, and thus were excluded from analysis. Records in 2010 and later were not examined either, because only validation data (Section 2.4.3) prior to 2010 were available. Recorded sightings of *R. bieti* were extracted from patrol forms during the period of 2007 to 2009. A total of 103 *R. bieti* occurrence locations over the 3 years were obtained (Figure 2). Year-wise, there were 25, 48, and 30 *R. bieti* occurrences in 2007, 2008, and 2009, respectively. Season-wise, there were 30, 23, 19, and 31 *R. bieti* occurrences in the spring (Mar, Apr, May), summer (Jun, Jul, Aug), autumn (Sep, Oct, Nov), and winter (Dec, Jan, Feb) seasons, respectively, across the 3 years.

2.3 | A heuristic-based approach to mitigating positional errors

A heuristic-based approach was developed to derive the probable *R. bieti* occurrence locations corresponding to *R. bieti* sightings recorded by the patrol for mitigating the impact of the positional errors on SDM using patrol data.

2.3.1 | Assumption

Positional errors exist in the *R. bieti* sightings for two reasons. On the one hand, *R. bieti* are afraid of humans and would not allow the patrol to approach them too closely. On the other hand, the ability of the patrol to see the monkeys was limited by the surrounding terrain. But since the patrol always visually confirmed *R. bieti* sightings, it is implied that the actual location of *R. bieti* occurrence must be visible from the recorded patrol position.

We assumed that the monkeys are more likely to present in forest habitat they are known to prefer (i.e., forest heuristic), that there was some minimum distance (D_{min}) the monkeys would keep from the patrol, and that there was some maximum distance (D_{max}) at which the patrol could see the monkeys, depending on the terrain (i.e., visibility heuristic). The actual *R. bieti* occurrence location corresponding to each sighting should be subject to these constraints. Accordingly, the probable *R. bieti* occurrence location was selected based on the forest and visibility heuristics.

The proposed approach uses the same logic of binary and multivariate dasymetric areal interpolation methods (see Eicher & Brewer, 2001) to first apply restrictions to narrow down the range within which phenomena occur. Once the restrictions are imposed, our problem becomes how to select one probable occurrence location within the constrained areas (whereas areal interpolation is concerned with allocating a raw areal total over the constrained areas).

2.3.2 | Implementation

Based on the above assumption, a heuristic-based approach was implemented to locate the probable *R. bieti* occurrence location corresponding to each recorded *R. bieti* sighting. *R. bieti* prefers to use coniferous, broadleaved, and the mixture of coniferous and broadleaved forests in the study area (Huang et al., 2012). *R. bieti* patrol records confirmed this, as the recorded habitat type information indicated that all encountered *R. bieti* occurred in these types of forest. The preferred forests were extracted from a vegetation type map delineated by a field biologist during a field inventory of the study area in 2009 (Huang, 2009). The spatial extent of the study area was determined by the boundary of the vegetation type map.

Viewshed analysis was performed on a digital elevation model (DEM) of the study area to locate areas that are visible from the recorded location (patrol position) within the [D_{min}, D_{max}] distance range. A DEM (30 m resolution) was created from the contours digitized from 1:50,000 topographic maps of the area (20 m contour interval). There are other freely available data sources for DEMs, for example, SRTM DEM (van Zyl, 2001) and ASTER DEM (Toutin, 2008). The mountainous study area has dramatic elevation change and dense vegetation cover. The DEM generated from topographic maps was expected to be more accurate than other space-borne DEM products. The algorithm developed by Wang, Robinson, and White (2000) was adopted to compute the viewshed at each recorded location. Based on the field experiences of the patrol in the study area, D_{min} was estimated as 50 m (the average closest distance the monkeys would allow the patrol to approach) and D_{max} as 150 m (the average maximum distance at which the monkeys were sighted).

The areas visible from a recorded location were spatially overlaid with the preferred forests. The intersection of the two layers represents the extent of probable *R. bieti* occurrence locations. One location was randomly selected from the intersection and treated as the probable *R. bieti* occurrence location corresponding to the recorded location (Figure 3). One probable *R. bieti* occurrence location was selected for each of the recorded locations following this procedure.

2.4 | Species distribution modeling

2.4.1 | Environmental data

SDM requires environmental data (predictors). Based on knowledge on the habitat use of *R. bieti* (Huang et al., 2012; Kirkpatrick et al., 1998; Long et al., 1994; Zhang et al., 2017; Zhu et al., 2015), terrain conditions, water source, shelter or food, and human impact were considered as the key factors that influence the habitat use and distribution of *R. bieti* in the study area. Accordingly, data characterizing environmental variables related to these factors (Table 1) were used in SDM. The derivation of these environmental data layers was performed using ArcMap (Esri, 2016).

2.4.2 | Modeling technique

The widely used Maxent method (Phillips, Anderson, & Schapire, 2006) was adopted for SDM. Maxent is statistically rigorous (Elith et al., 2011), well suited for SDM with species presence-only data, and generally can achieve better

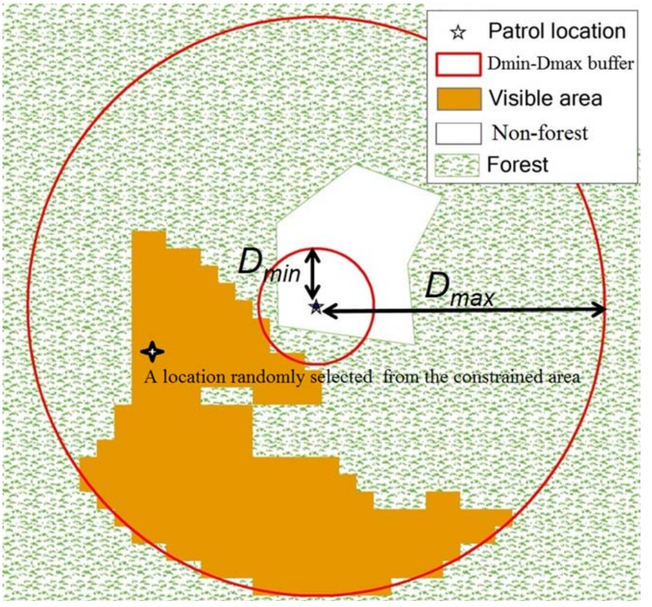

FIGURE 3 A heuristic-based approach for selecting the probable *R. bieti* occurrence location corresponding to a recorded patrol location

performance amongst other alternatives (Elith & Graham, 2009; Elith et al., 2006). Maxent can incorporate both categorical and continuous predictors, and make good predictions even when the sample size is small (Phillips & Dudík, 2008; Phillips et al., 2006). Besides, the default values of modeling parameters were fine-tuned using a large data set and thus users are relieved from the difficulty of tweaking model parameters (Phillips & Dudík, 2008).

The Maxent software (version 3.3.3k) was downloaded and used for SDM. The default values of the model parameters were fine-tuned based on a large data set and are supposed to achieve good modeling performance in general (Phillips & Dudík, 2008). Thus, the default values were used throughout all experiments in this study (e.g., auto features, logistic output format, add samples to background, etc.).

2.4.3 | Model evaluation

Validation data

Occurrence locations of *R. bieti* recorded during field tracking of the Mt. Lasha *R. bieti* group were used as independent validation data to evaluate the accuracy of SDM using patrol data.

TABLE 1 Environmental variables (predictors) used in species distribution modeling (30 m resolution)

Category	Environmental variable	Data source
Terrain condition	Elevation	DEM created from contours digitized from 1:50,000 topographic maps (20 m contour). Continuous
	Slope	Derived from DEM, unit in %. Continuous
	Aspect	Derived from DEM. 0–360° discretized into eight equal 45° categories. Categorical
Water source	Cost distance to river	Least cost distances to rivers, slope (%) as cost. Continuous
Shelter or food	Vegetation type	Vegetation type map delineated by a field biologist during a field inventory of the study area in 2009 (Huang, 2009). Categorical
Human impact	Cost distance to village or road	Least cost distances to villages or roads, slope (%) as cost. Continuous

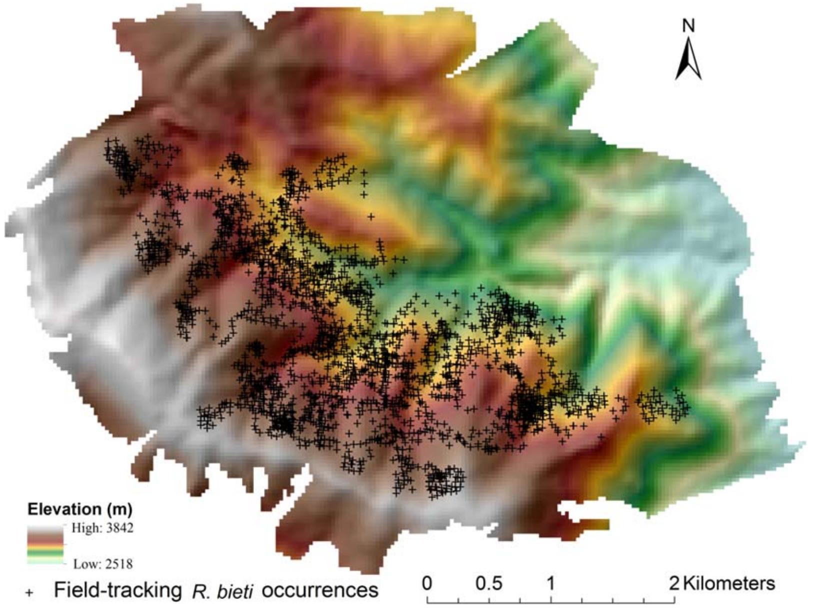

FIGURE 4 *R. bieti* occurrence locations recorded during field tracking of the monkeys in 2008 and 2009

The occurrence locations were obtained by intensively tracking the group of monkeys in the study area. The tracking was conducted by one field biologist and two assistants. They first spent nearly a year in the field to familiarize themselves with the terrain and habituate the monkeys. Tracking and direct observation of the monkeys were then conducted in 2008 and 2009 (primarily for behavioral study purposes) (Huang et al., 2012). Time spent on field tracking covered 7 months in 2008 (May, Jun, Jul, Aug, Oct, Nov, Dec) and 8 months in 2009 (Jan, Feb, Mar, Apr, Jun, Jul, Aug, Sep). During the field tracking period, the monkeys' location was plotted on a large-scale topographic map (1:50,000) and recorded every 30 min (from 7 a.m. to 8 p.m.). In total, 2,615 *R. bieti* occurrence locations were recorded (1,323 and 1,292 occurrences in 2008 and 2009, respectively; 722, 853, 584, and 466 occurrences in spring, summer, autumn, and winter, respectively) (Figure 4). These field tracking occurrences are the most accurate data available that reveal the distribution of *R. bieti* in the study area during the study period (2007 to 2009).

Evaluation metric

The area under the ROC (receiver operating characteristic) curve (AUC) was used as a measure of the accuracy of SDM. Given a predicted map of species occurrence likelihood and validation data points where the species were observed present (positive) or absent (negative), a ROC curve is obtained by plotting all true positive fraction values on the *y*-axis against their equivalent false positive fraction for all available likelihood thresholds on the *x*-axis. The AUC is the probability that the predicted occurrence likelihood at a randomly chosen presence location will be higher than that at a randomly chosen absence location. The value of the AUC is in the range [0.5, 1.0]. AUC = 0.5 indicates that the model performance is no better than random predictions. AUC = 1.0 indicates perfect model performance (Fielding & Bell, 1997). Models with AUC values greater than 0.75 are considered potentially useful (Elith, Burgman, & Regan, 2002; Phillips & Dudík, 2008). The AUC provides a single model performance measure that is independent of any choice of threshold.

If species absence data are not available in the validation data, then an AUC can be calculated using background data (i.e., pseudo-absences), chosen uniformly at random from the study area (Phillips & Dudík, 2008). In this case, the AUC is the probability that a randomly chosen present location has a higher predicted occurrence likelihood than a randomly chosen background location (Phillips et al., 2006). With presence-only data, random prediction still corresponds to an AUC of 0.5, but the maximum achievable AUC is less than 1.0 (Wiley, McNyset, Peterson, Robins, & Stewart, 2003). In this study, the AUC computed by Maxent was based on presence-only validation data (Section 2.4.3.1).

2.5 | Evaluation and assessment

To assess the impact of the positional errors, SDM was conducted using the original *R. bieti* sightings from patrol records (i.e., raw sightings). The accuracy of the modeling result was evaluated using the independent validation data (i.e., field tracking *R. bieti* occurrences).

SDM was also conducted using the probable *R. bieti* occurrence locations obtained through the heuristic-based approach (i.e., probable occurrences). The effectiveness of the approach was assessed by comparing the accuracy of SDM using the probable occurrences to that using the raw sightings. To account for the randomness in the approach (i.e., randomly selecting one location from the constrained area as probable occurrence location), the approach was repeatedly applied to the raw sightings 20 times to generate 20 sets of probable occurrences. SDM was performed on each of the 20 sets of probable occurrences. The AUC of each SDM run was computed using the independent validation data. The mean of the 20 AUCs was computed and compared to the AUC of SDM using the raw sightings through one sample *t* tests (Lehmann & Romano, 2006).

The effects of individual heuristics employed to locate the probable *R. bieti* occurrence location were examined through comparative experiments. For comparison, four alternative sampling methods were implemented to obtain the probable occurrence location corresponding to a recorded patrol position: (A) Randomly select a location that is D_{min} to D_{max} distance away and visible from the patrol position (visibility constraint only); (B) Randomly select a location that is D_{min} to D_{max} distance away from the patrol position and in the preferred forests (forest constraint only); (C) Randomly select a location that is in the preferred forest; and (D) Randomly select a location that is in the study area. Note that methods C and D totally discard the patrol position. Each sampling method was applied to the raw *R. bieti* sightings 20 times to obtain 20 sets of probable occurrence locations. SDM was conducted using each set of probable occurrence locations.

To allow for the examination of any potential annual or seasonal variations in the positional errors in patrol data, SDM was conducted on various partitions of the patrol data: all patrol data; patrol data in each year (2007, 2008, and 2009); and patrol data in each season (spring, summer, autumn, and winter). For SDM using all patrol data and using patrol data in each year, all the validation data were used to compute the AUC. For SDM using patrol data in a season, validation of the data in the corresponding season was used to compute the AUC.

3 | RESULTS

Figure 5 shows the output distribution map of SDM using raw *R. bieti* sightings in each partition of the patrol data and the distribution map averaged from the 20 output distribution maps of SDM using the corresponding 20 sets of probable *R. bieti* occurrences.

The AUC values of SDM using the raw sightings and the corresponding probable occurrences derived through the heuristic-based approach as presented in Section 2.3 are shown in Table 2. The AUC values of SDM using the raw sightings on most partitions were below 0.75 (except spring and autumn). The AUC values of SDM using the derived probable occurrences were above 0.75, and higher than those using the raw sightings on most partitions (except autumn).

The AUCs of SDM using probable *R. bieti* occurrence locations obtained using different sampling methods are shown in Figure 6. As expected, the AUCs were around 0.5 (i.e., performance comparable to random prediction) for SDM using locations randomly selected from the study area. The AUCs were around 0.75 for SDM using locations randomly selected from forests. The AUCs were generally around 0.75 for SDM using locations obtained by the visibility constraint only sampling method (except 2007, spring, and autumn). The AUCs were close to or above 0.8 for SDM using locations obtained by the forest constraint only sampling method. The AUCs for SDM using locations obtained by the proposed approach (i.e., two constraints) were slightly lower than those under the forest constraint only sampling method.

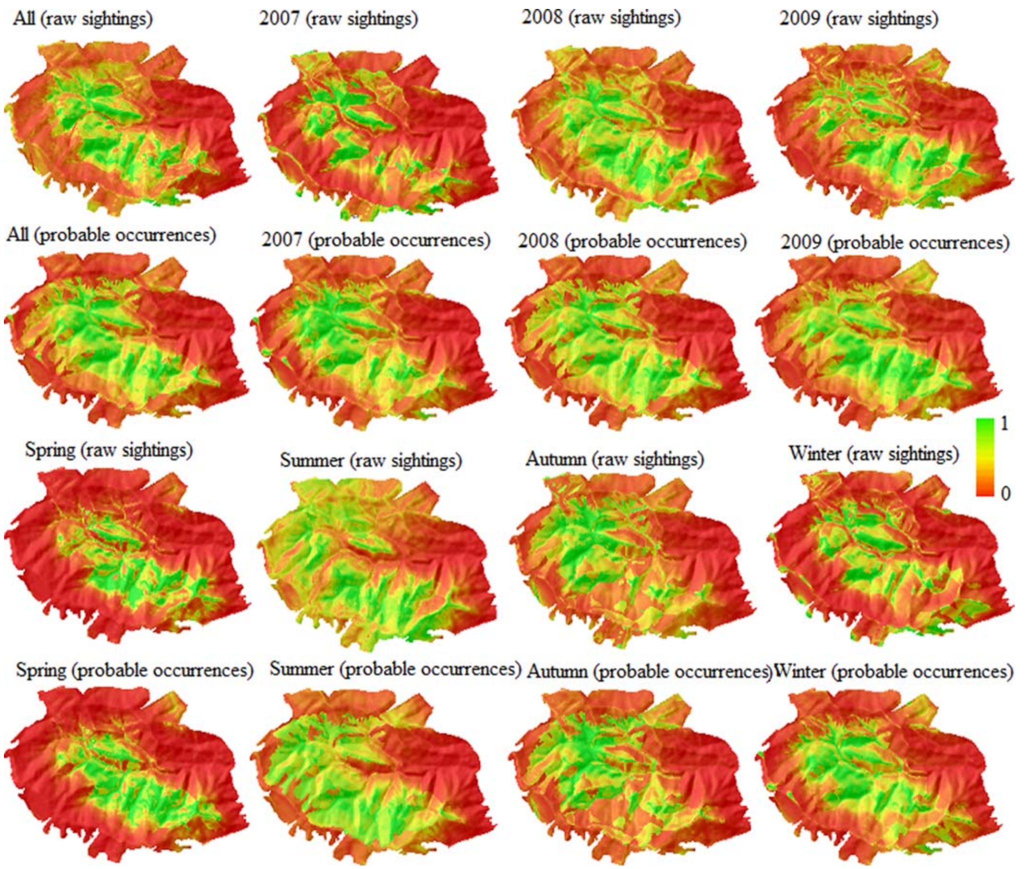

FIGURE 5 Output distribution map of SDM using raw *R. bieti* sightings and the corresponding probable occurrences obtained through the heuristic-based approach (average map of 20 output distribution maps). Values on the map indicate the relative likelihood of *R. bieti* occurrence

TABLE 2 Accuracy of SDM maps predicted using raw *R. bieti* sightings and using probable occurrences obtained through the heuristic-based approach presented in Section 2.3

		Year			Season			
Partitions	All	2007	2008	2009	Spring	Summer	Autumn	Winter
Raw sightings	0.721	0.650	0.745	0.721	0.826	0.743	0.816	0.708
Probable occurrences								
Mean	0.789	0.758	0.787	0.787	0.857	0.813	0.811	0.819
Std	0.008	0.013	0.010	0.015	0.010	0.012	0.016	0.018
t Test								
df	19	19	19	19	19	19	19	19
Critical value	37.073	35.134	17.508	18.779	13.048	26.185	−1.397	26.387
p	0.000	0.000	0.000	0.000	0.000	0.0001	0.179	0.000
Correlation	0.804	0.692	0.861	0.776	0.886	0.704	0.892	0.745

t Test: One sample *t* test of the null hypothesis that the mean of the 20 AUCs of SDM using probable occurrences is equal to the AUC of SDM using the raw sightings.
Correlation: Pearson's correlation coefficient between SDM map (pixel values) predicted using raw sightings and the average map of 20 SDM maps predicted using probable occurrences.

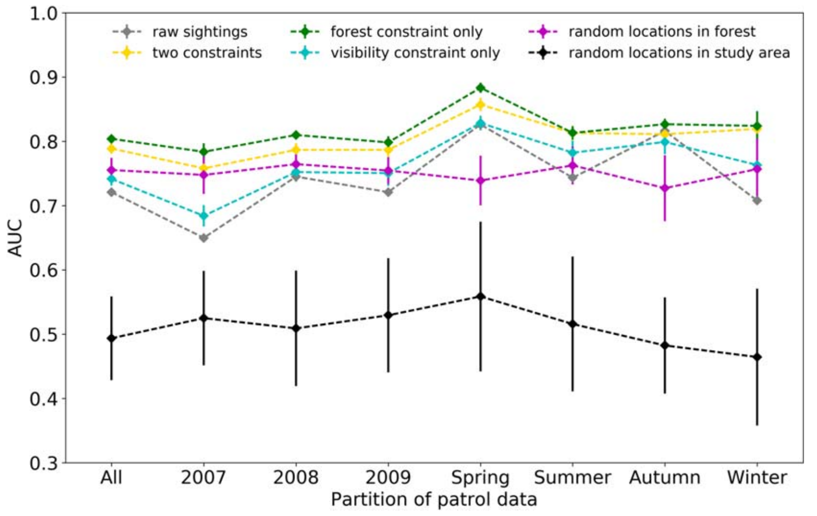

FIGURE 6 Comparison of accuracy of SDM using *R. bieti* occurrences obtained through different sampling approaches. The mean and standard deviation (shown as an error bar) of the AUC were computed based on 20 AUC values

4 | DISCUSSION

4.1 | Impact of the positional errors

SDM models with AUC values above 0.75 are considered potentially useful (Elith et al., 2002; Phillips & Dudík, 2008). It is thus fair to conclude that generally, the positional errors in the raw *R. bieti* sightings had significant adverse effects on SDM and degraded the accuracy of SDM (except spring and autumn) (Table 2).

The accuracy of SDM using the raw sightings in spring and autumn was satisfactory (AUC = 0.832 and 0.816 in spring and autumn, respectively), and higher than that in summer and winter. This indicates that patrol data collected in summer and winter were more prone to positional errors, probably due to the weather conditions in the two seasons that limited the patrol's field of view (e.g., rain and fog in summer, snow in winter).

The accuracy of SDM using the raw sightings in 2007 (AUC = 0.650) was much lower than the accuracy of SDM using the raw sightings in 2008 and 2009. This indicates that the patrol data in 2007 were more prone to positional errors. This might be attributed to the fact that the patrol was less skillful and experienced in 2007 when patrolling in the study was first initiated.

We ran the above experiments excluding *R. bieti* patrol data in 2007 (6, 5, 5, and 9 raw *R. bieti* sightings in spring, summer, autumn, and winter, respectively), and the results are shown in Table 3. Compared with SDM using the 3-year patrol data, the AUCs of SDM increased slightly on the all, spring, and winter partitions. This was a benefit of excluding the more error-prone 2007 data, but the AUCs decreased slightly on the summer and autumn partitions, which was probably due to the reduced sample sizes.

4.2 | Effectiveness of the approach

The mean AUC of SDM using the probable occurrences obtained through the heuristic-based approach (AUC_{prob}) was statistically significantly higher than that using the raw sightings (AUC_{raw}) (Table 2). The only exception was on the autumn partition, where AUC_{prob} was slightly lower than AUC_{raw}. But the difference was not statistically significant. The heuristic-based approach greatly improved the AUC of SDM using patrol data in 2007 and winter by 0.108 and 0.111, respectively. Even in spring, where SDM using the raw *R. bieti* sightings could already result in satisfactory SDM ($AUC_{raw} = 0.826$), the approach improved the AUC by 0.031. The AUC of SDM using all raw sightings was below 0.75 ($AUC_{raw} = 0.721$). By deriving probable occurrences to mitigate the positional errors, the approach improved the AUC to around 0.8 ($AUC_{prob} = 0.789$; an improvement of 0.068). A similar pattern was observed in the experiment results

TABLE 3 Accuracy of SDM maps predicted using raw *R. bieti* sightings and probable occurrences obtained through the heuristic-based approach (excluding data in 2007)

		Year		Season			
Partitions	**All**	**2008**	**2009**	**Spring**	**Summer**	**Autumn**	**Winter**
Raw sightings	0.736	0.745	0.721	0.831	0.730	0.783	0.743
Probable occurrences							
Mean	0.797	0.787	0.787	0.850	0.800	0.774	0.840
Std	0.006	0.010	0.015	0.016	0.014	0.017	0.023
t Test							
df	19	19	19	19	19	19	19
Critical value	45.860	17.508	18.779	5.020	21.138	−2.150	18.571
p	0.000	0.000	0.000	0.000	0.0001	0.045	0.000
Correlation	0.817	0.861	0.776	0.899	0.705	0.927	0.715

t Test: One sample *t* test of the null hypothesis that the mean of the 20 AUCs of SDM using probable occurrences is equal to the AUC of SDM using the raw sightings.
Correlation: Pearson's correlation coefficient between SDM map (pixel values) predicted using raw sightings and the average map of 20 SDM maps predicted using probable occurrences.

excluding data in 2007 (Table 3). A largest AUC improvement of 0.097 was brought about by the approach on the winter partition. The AUC improvement was 0.6 using all data in 2008 and 2009. Overall, the proposed heuristic-based approach effectively mitigated positional errors in patrol data and improved the accuracy of SDM using patrol data.

4.3 | Effects of individual heuristics

As shown in Figure 6, the proposed approach (i.e., two constraints) and the forest constraint only sampling method performed much better than the visibility constraint only sampling method and other sampling methods. Consistently, the forest constraint only method had a slightly better performance than the proposed approach (Figure 6). These patterns were also observed in experimental results excluding 2007 data (Figure 7).

It was clear that the forest constraint was more effective than the visibility constraint in mitigating positional errors in patrol data. However, it is recommended not to abandon the visibility constraint in the heuristic-based

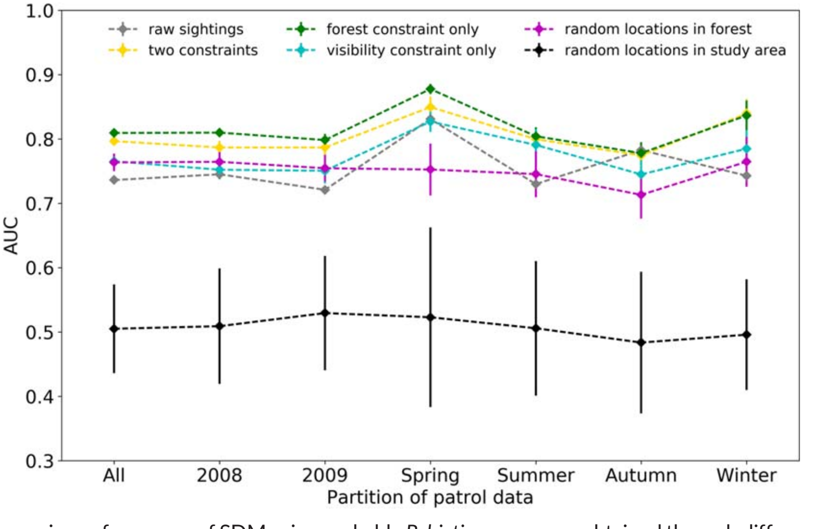

FIGURE 7 Comparison of accuracy of SDM using probable *R. bieti* occurrences obtained through different sampling approaches (excluding data in 2007). The mean and standard deviation (shown as error bar) of AUC were computed based on 20 AUC values

approach. The visibility constraint is grounded on an understanding of the patrol data generation processes. In routine patrols, visual confirmation of *R. bieti* sightings by the patrol was subject to the visibility constraint given the variety of rugged terrains in the study area. We are aware that DEM-based viewshed analysis might not be sufficient to represent the real visibility conditions in the study area. Visibility could also be constrained by other factors such as vegetation and weather conditions. Nevertheless, DEM-based viewshed analysis provides a good first approximation to the visibility constraint.

4.4 | Applicability of the approach

Positional errors in patrol data might not be an issue for SDM at coarse spatial resolutions. For example, Plumptre et al. (2016) build an occupancy model using patrol data on 5 km $\times$ 5 km grid cells to predict Gorilla presence probability. The proposed heuristic-based approach is intended to mitigate the positional errors in patrol data for SDM at fine spatial resolutions (e.g., 30 m), which could be useful to support local-level conservation practices such as habitat management and restoration in parks or nature reserves (Franklin & Miller, 2009).

The proposed approach was designed based on knowledge of the ecology of species and the data generation processes in the field. In this regard, the approach should be applicable to tackle positional errors in patrol data in general. In this study, *R. bieti* is a very localized species in the study area (Huang et al., 2012; Huang, 2009), and two simple heuristics were implemented based on *R. bieti* habitat preferences and the patrol's field of view. If other useful knowledge is available, it can be added to further narrow down the scope of the probable occurrence location. However, the specific knowledge that is incorporated to select the probable species occurrence locations may depend on the species of interest and situations. In other cases, heuristics based on a different set of knowledge on the species ecology and data generation processes might be used instead.

Detailed and accurate species data, such as the *R. bieti* field tracking data used in this study, would not be available in many other cases. Collecting such detailed data is costly, labor-intensive, and often carried out for species behavioral studies (Huang et al., 2012). But it is worth pointing out that the tracking data were only used as independent validation data to evaluate the accuracy of SDM using patrol data. The proposed approach itself does not require such tracking data; it requires only patrol data and environmental data (e.g., land cover map, DEM). Some freely available data sources for land cover map and DEM include the USGS 30 m Global Land Cover (Hansen et al., 2013), SRTM DEM (van Zyl, 2001), and the ASTER DEM (Toutin, 2008).

Cost-effective patrol data accumulated in ranger-based monitoring programs might be the only available species data to support conservation decision making in many funds-lacking scenarios (Danielsen et al., 2003). However, we should remember that patrol data may be subject to data quality issues other than positional errors (Keane et al., 2011). For example, the patrol may not record every wildlife sighting, especially for common species. This would result in biases in occurrence data among species. Moreover, due to spatially or temporally unevenly distributed observation efforts, there might be spatial or temporal sample selection bias in patrol data. In the case study, the ranger regularly patrolled on routes covering the whole study area and recorded every *R. bieti* sighting; thus we expected the abovementioned biases to be insignificant. In cases where such biases are significant, measures should be taken to account for them when using such data for SDM (Phillips et al., 2009; Zhang et al., 2017; Zhu et al., 2015).

5 | CONCLUSIONS

SDM at fine spatial resolutions using patrol data suffers from positional errors (i.e., recorded locations were patrol positions rather than actual wildlife occurrence locations). This study presented an evaluation of the impact of the positional errors in patrol data on SDM and developed a heuristic-based approach for mitigating the positional errors based on species preferred habitat and the observer's field of view. The effectiveness of the approach was thoroughly evaluated through a case study of *R. bieti* in Yunnan, China. Results showed that the positional errors in *R. bieti* patrol data had an adverse effect on SDM. The proposed heuristic-based approach effectively mitigated the positional errors and greatly improved SDM performance. The evaluations, combined with the proposed positional errors mitigation approach, could better inform and improve the appropriate uses of cost-effective ranger-based monitoring data to support conservation.

ACKNOWLEDGMENTS

The work reported here was supported by grants from the National Natural Science Foundation of China (Projects Nos. 31560118, 30960085, 41431177), the National Basic Research Program of China (Project No. 2015CB954102), the Natural Science Research Program of Jiangsu (14KJA170001), and PAPD, 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 at the 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 are greatly appreciated. We are grateful to the Yunling Nature Reserve Administration Bureau and Jin-Fu Zhang for their assistance in field data collection.

ORCID

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

REFERENCES

Anderson, D. R., Laake, J. L., Crain, B. R., & Burnham, K. P. (1979). Guidelines for line transect sampling of biological populations. *Journal of Wildlife Management*, *43*, 70–78.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). *Introduction to distance sampling estimating abundance of biological populations*. Oxford, UK: Oxford University Press.

Burton, A. C. (2010). *Wildlife monitoring and conservation in a West African protected area* (Unpublished PhD dissertation). University of California, Berkeley, CA.

Burton, A. C., Sam, M. K., Balangtaa, C., & Brashares, J. S. (2012). Hierarchical multi-species modeling of carnivore responses to hunting, habitat and prey in a West African protected area. *PLoS One*, *7*, e38007.

Campbell, A. F., & Sussman, R. W. (1994). The value of radio tracking in the study of neotropical rain forest monkeys. *American Journal of Primatology*, *32*, 291–301.

Critchlow, R., Plumptre, A. J., Alidria, B., Nsubuga, M., Driciru, M., Rwetsiba, A., . . . Beale, C. M. (2016). Improving law-enforcement effectiveness and efficiency in protected areas using ranger-collected monitoring data. *Conservation Letters*, *2016*, 12288.

Danielsen, F., Jensen, A. E., Alviola, P. A., Balete, D. S., Mendoza, M., Tagtag, A., . . . Enghoff, M. (2005). Does monitoring matter? A quantitative assessment of management decisions from locally-based monitoring of protected areas. *Biodiversity & Conservation*, *14*, 2633–2652.

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.

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography & Geographic Information Science*, *28*, 125–138.

Elith, J., Burgman, M. A., & Regan, H. M. (2002). Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, *157*, 313–329.

Elith, J., & Graham, C. H. (2009). Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, *32*, 66–77.

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.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of Maxent for ecologists. *Diversity & Distributions*, *17*, 43–57.

Esri. (2016). *ArcGIS Desktop: Release 10.1*. Redlands, CA: Author.

Fernandez, M. A., Blum, S. D., Reichle, S., Holzman, B., & Hamilton, H. (2009). Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodiversity Informatics*, *6*, 36–52.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*, 38–49.

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

Gillespie, T. W., Foody, G. M., Rocchini, D., Giorgi, A. P., & Saatchi, S. (2008). Measuring and modelling biodiversity from space. *Progress in Physical Geography*, *32*, 203–221.

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Peterson, A. T., Loiselle, B. A., & The NCEAS Predicting Species Distributions Working Group. (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, *45*, 239–247.

Gray, M., & Kalpers, J. (2005). Ranger-based monitoring in the Virunga–Bwindi region of east-central Africa: A simple data collection tool for park management. *Biodiversity & Conservation*, *14*, 2723–2741.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., ... Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Loveland, T. R. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*, 850–853.

Hemson, G., Johnson, P., South, A., Kenward, R., Ripley, R., & MacDonald, D. (2005). Are kernels the mustard? Data from global positioning system (GPS) collars suggests problems for kernel home-range analyses with least-squares cross-validation. *Journal of Animal Ecology*, *74*, 455–463.

Hilborn, R., Arcese, P., Borner, M., Hando, J., Hopcraft, G., Loibooki, M., ... Sinclair, A. R. E. (2006). Effective enforcement in a conservation area. *Science*, *314*, 1266.

Huang, Z. P. (2009). *Foraging, reproduction and sleeping site selection of black-and-white snub-nosed monkey (Rhinopithecus bieti) at the southern range* (Unpublished MS 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*, 517–524.

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

Jachmann, H. (2008). Monitoring law-enforcement performance in nine protected areas in Ghana. *Biological Conservation*, *141*, 89–99.

Johnson, C. J., & Gillingham, M. P. (2008). Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou. *Ecological Modelling*, *213*, 143–155.

Keane, A., Jones, J. P. G., & Milner-Gulland, E. J. (2011). Encounter data in resource management and ecology: Pitfalls and possibilities. *Journal of Applied Ecology*, *48*, 1164–1173.

Kerr, J. T., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends in Ecology & Evolution*, *18*, 299–305.

Kirkpatrick, R. C., Long, Y. C., Zhong, T., & Xiao, L. (1998). Social organization and range use in the Yunnan snub-nosed monkey *Rhinopithecus bieti*. *International Journal of Primatology*, *19*, 13–51.

Lehmann, E. L., & Romano, J. P. (2006). *Testing statistical hypotheses*. New York: Springer-Verlag.

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.

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.

Nyirenda, V. R. (2012). Field foot patrol effectiveness in Kafue National Park, Zambia. *Journal of Ecology & the Natural Environment*, *4*, 163–172.

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.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259.

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, *31*, 161–175.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*, 181–197.

Plumptre, A. J., Fuller, R. A., Rwetsiba, A., Wanyama, F., Kujirakwinja, D., Driciru, M., … Possingham, H. P. (2014). Efficiently targeting resources to deter illegal activities in protected areas. *Journal of Applied Ecology*, *51*, 714–725.

Plumptre, A. J., Nixon, S., Kujirakwinja, D. K., Vieilledent, G., Critchlow, R., Williamson, E. A., … Hall, J. S. (2016). Catastrophic decline of world's largest primate: 80% loss of Grauer's Gorilla (*Gorilla beringei graueri*) population justifies critically endangered status. *PLoS One*, *11*, e0162697.

Poulsen, M. K., & Luanglath, K. (2005). Projects come, projects go: Lessons from participatory monitoring in southern Laos. *Biodiversity & Conservation*, *14*, 2591–2610.

Toutin, T. (2008). ASTER DEMs for geomatic and geoscientific applications: A review. *International Journal of Remote Sensing*, *29*, 1855–1875.

Trolle, M., & Kéry, M. (2003). Estimation of ocelot density in the Pantanal using capture–recapture analysis of camera-trapping data. *Journal of Mammalogy*, *84*, 607–614.

van Zyl, J. J. (2001). The Shuttle Radar Topography Mission (SRTM): A breakthrough in remote sensing of topography. *Acta Astronautica*, *48*, 559–565.

Viña, A., Bearer, S., Zhang, H., Ouyang, Z., & Liu, J. (2008). Evaluating MODIS data for mapping wildlife habitat distribution. *Remote Sensing of Environment*, *112*, 2160–2169.

Walpole, M. J. (2002). Factors affecting black rhino monitoring in Masai Mara National Reserve, Kenya. *African Journal of Ecology*, *40*, 18–25.

Wang, J., Robinson, G. J., & White, K. (2000). Generating viewsheds without using sightlines. *Photogrammetric Engineering & Remote Sensing*, *66*, 87–90.

Wiley, E. O., McNyset, K., Peterson, T., Robins, R., & Stewart, A. (2003). Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography*, *16*, 120–127.

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.

Zhang, G., Zhu, A. X., Huang, Z. P., Ren, G., Qin, C. Z., & Xiao, W. (2017). Validity of historical volunteered geographic information: Evaluating citizen data for mapping historical geographic phenomena. *Transactions in GIS*, *21*, in press.

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.

How to cite this article: 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;22:202–216. https://doi.org/10.1111/tgis.12303