Zhang et al. - Validity of historical volunteered geographic information: Evaluating citizen data for mapping historical geographic phenomena

Supporting Information

S1 Division of historical periods

It is well known that the habitat use of *R. bieti* is closely related to forest-provisioned resources such as food and shelter (Long et al. 1994; Xiao et al. 2003). In the study area two historical events had significant impacts on these forest-provisioned resources: in 1979 the China Environmental Protection Act was enacted; in 2006 the Mt. Lasha area became part of the Yunling Nature Reserve. The forestry policy implementations associated with these events in the study area directly affected local residents' exploitation of the forests. Thus, these two events can be naturally taken as the temporal breakpoints to divide the decades-long citizen data coverage into shorter historical periods in which the forest-provisioned resources to support the monkeys were relatively stable.

Multi-temporal Landsat multispectral images (landsat.usgs.gov) were used to detect the transitions between forest land and non-forest land in the study area and to identify non-change areas (Section S2.1). Given the timing of the two historical events and limited by the availability of Landsat images (Section S2.1), the *R. bieti* presence data elicited from local residents were separated into three historical periods for evaluation through habitat suitability mapping: 1973–1981, 1987–2005, and 2006–2010.

S2 Identifying non-change areas

Studies (Long et al. 1994; Xiao et al. 2003) have found that in the last sixty years habitat reduction and fragmentation caused by human-induced deforestation is the major threat to

R. bieti. The effects of climate change are negligible. Across the three historical periods, two major drivers induced environmental change that impacted *R. bieti* habitat use in the study area (Huang 2009). *First*, the transition between forest land and non-forest land affected forest-provisioned resources for the monkeys. *Second*, the establishment and disappearance of country roads and small village settlements introduced variabilities in human-posed disturbance for the monkeys. The terrain conditions influencing *R. bieti* habitat use (e.g., elevation, slope gradient, slope aspect, etc.) did not change significantly over the period. Thus, the non-change areas can be identified by locating areas free of transitions between forest and non-forest, and free of significant change of human-posed disturbance across the three historical periods.

S2.1 Identifying areas free of transitions between forest and non-forest

Areas free of transitions between forest land and non-forest land across the three historical periods were identified based on land use/cover maps classified from multi-temporal Landsat images. United States Geological Survey provides access to Landsat raw images (digital numbers) and Landsat surface reflectance data products (Masek et al. 2006). Landsat raw images are available for the study area as early as 1973 (MSS sensor). During 1982–1986, either no Landsat images or no cloud-free Landsat images are available for the study area. Landsat surface reflectance data products are available since 1987 (TM, ETM, or ETM+ sensor). Limited by the availability of Landsat images and the timing of the two historical events (Section S1), MSS raw images were used to examine land use/cover change in the 1973–1981 period, and TM surface reflectance images were used for the 1987–2005 period and the 2006–2010 period.

One scene of cloud-free Landsat image for the study area in winter months was selected for each year wherever possible, as forest land and non-forest land are most discernable in winter months in the study area (e.g., harvested farmland in winter has a very different spectral signature from forest), and there are more cloud-free Landsat images for the study area in winter months. Eighteen scenes of Landsat images in total were selected: six MSS raw images (dates: 18-Dec-1973, 5-Jan-1974, 10-Dec-1976, 23-Dec-1979, 10-Jan-1980, and 19-Oct-1981) and twelve TM land surface reflectance images (dates: 30-Dec-1987, 25-Jan-1989, 7-Nov-1991, 30-Dec-1993, 6-Dec-1996, 17-Dec-2000, 20-Dec-2001, 26-Dec-2003, 28-Dec-2004, 1-Feb-2006, 8-Jan-2009 and 11-Jan-2010). All selected images were downloaded through the official Landsat website (landsat.usgs.gov).

Supervised classification of the selected Landsat images was performed to derive land use/cover maps. MSS raw images were first calibrated using the Landsat MSS Calibration tool in ENVI (Exelis Visual Information Solutions, Boulder, Colorado) to derive surface reflectance from digital numbers (Chander et al. 2009). All four MSS bands (60 m spatial resolution) were then resampled to 30 m resolution (to match TM image spatial resolution) and used in classification. TM surface reflectance images need no further preprocessing. Except for band 6 (the thermal infrared band), all the other six bands (30 m spatial resolution) were used for classification.

The *classification scheme* was derived from a land use/cover map delineated by a field biologist during a field inventory of the study area in 2009 (Huang 2009) (*reference land use/cover map* hereafter). On the reference land use/cover map (Figure S1), major land use/cover types in the study area are: *Evergreen Coniferous*, *Evergreen Broadleaf and Coniferous*, *Deciduous Broadleaf and Coniferous*, *Deciduous Broadleaf*, *Yunnan Pine*, *Pasture*, and *Farmland*. Other minor types (*Burning Remains*, *Charcoal Remains*, and *Shrubberies*) were omitted from the classification scheme because they have very limited coverage in the study area. Among the major types, the difference between *Evergreen Coniferous* forests and *Evergreen Broadleaf and Coniferous* forests might be indistinguishable on Landsat images in winter months. Moreover, the monkeys do not differentiate much between these two types of forests in their habitat use. Therefore, *Evergreen Broadleaf and Coniferous* was combined into *Evergreen Coniferous* to simplify the classification scheme. Similarly, *Deciduous Broadleaf and Coniferous* was combined into *Deciduous Broadleaf.* Thus, there were five land use/cover types in the final classification scheme: *Evergreen Coniferous, Deciduous Broadleaf, Yunnan Pine, Pasture,* and *Farmland.*

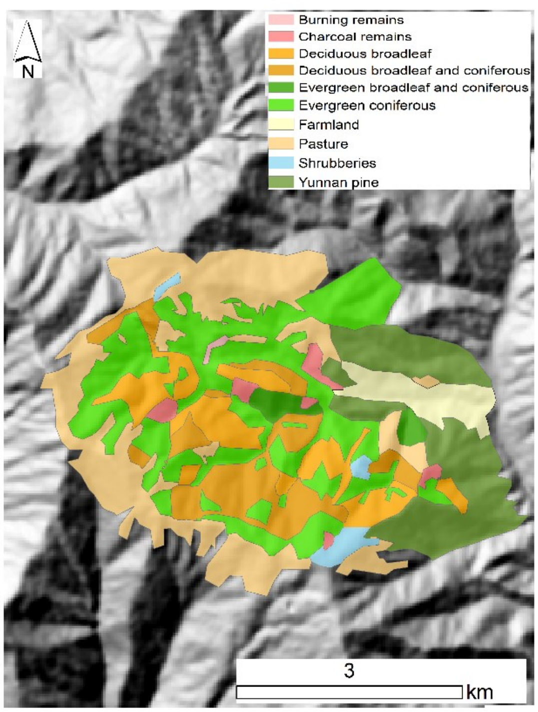

Figure S1. The reference land use/cover map of the study area delineated by a field biologist during a field inventory of the study area in 2009.

Training examples required by supervised classification were constructed based on the reference land use/cover map and multi-temporal Landsat images. For each type in the classification scheme, pixels were located where there was little land use/cover change over the years. Specifically, the reference map was vectorized and overlaid on top of multi-temporal Landsat images in ArcMap (Environmental Systems Research Institute, Redlands, California). Landsat images on each date were displayed in a false color composition (near-infrared, red, green) that is most effective for visualizing vegetation. For each classification type, polygon(s) of that type were highlighted on the reference map. Within a highlighted polygon, areas with little hue and saturation change on multi-temporal false color composition images were visually located and examined. Based on knowledge and field experience in the study area, we were confident that the located areas could represent areas with stable land use/cover over the years. Training examples for this land use/cover type were created by drawing small polygons enclosing a small number of pixels over the located areas while avoiding being too close to polygon boundaries. These pixels were then used as training examples to classify Landsat multispectral images on all dates.

Support vector machine, as one of the most commonly used supervised remote sensing image classification method (Huang et al. 2002; Pal and Mather 2005; Mountrakis et al 2011), was adopted to classify the Landsat images in ENVI. A majority analysis in a neighborhood of 5 by 5 pixels was performed on each classified land use/cover map using the Post Classification tools in ENVI to smooth the noise on the classification map.

Post-classification comparison (Coppin et al. 2004; Hansen and Loveland 2012) was performed by conducting change analysis based on the land use/cover maps on the 18 dates to identify areas free of transitions between forest and non-forest over the years. For this purpose, the five classes in the classification scheme (*Evergreen Coniferous*, *Deciduous Broadleaf*, *Yunnan Pine*, *Pasture*, and *Farmland*) were further grouped into two broader classes: *Forest* (*Evergreen Coniferous*, *Deciduous Broadleaf*, and *Yunnan Pine*) and *Non-forest* (*Pasture* and *Farmland*). This grouping was applied to the land use/cover maps to derived a *Forest* and *Non-forest* map for each of the 18 dates. *Based on these binary classification maps, areas free of transitions between forest and non-forest across the three periods were identified* (i.e., pixels that were always classified as *Forest* or as *Non-forest* on the 18 binary classification maps) (Figure S2).

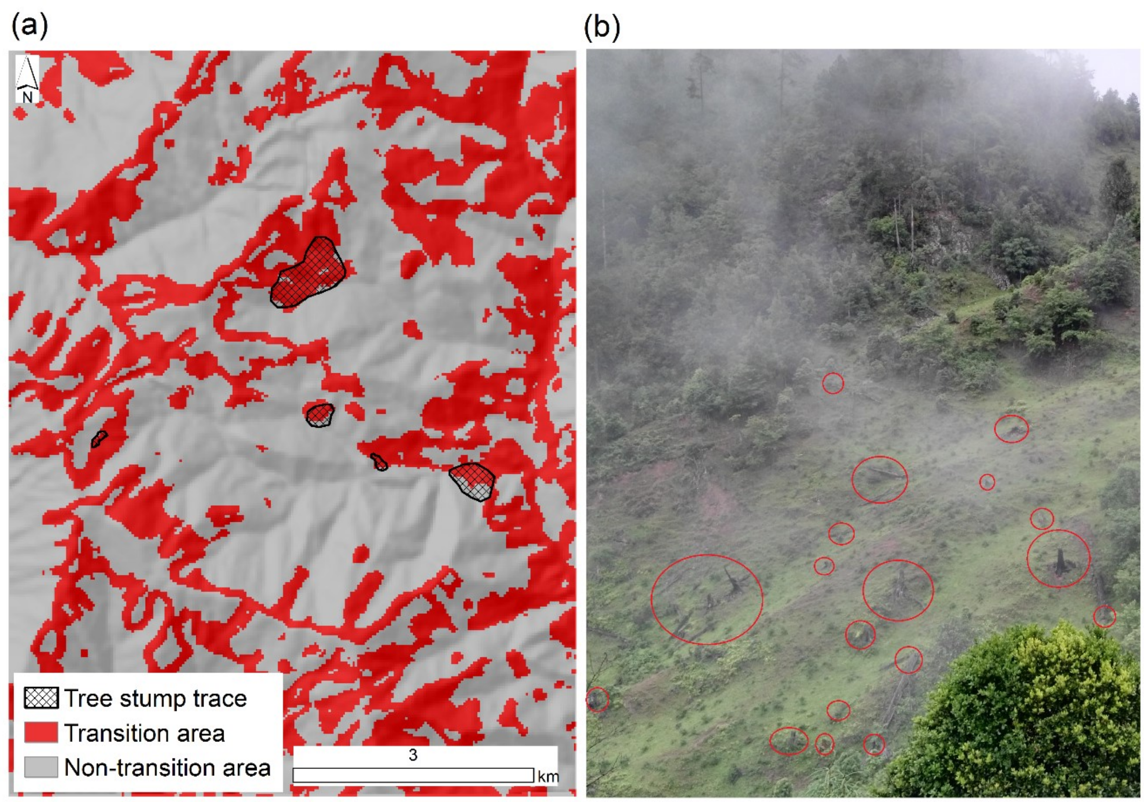

Figure S2. (a) The identified areas free of transitions between forest and non-forest across the three historical periods. (b) A typical tree stump trace area (large tree stumps circled in red; photo by Huang, ZP).

The ‘transition area’ on the change map (Figure S2) refers to areas with transitions between forest and non-forest, i.e., areas changed from forest to non-forest and/or from non-forest to forest at least once across the three periods. Multi-temporal field-observed land use/cover reference data are needed for a comprehensive accuracy evaluation of this change map. Unfortunately, such validation data are not available. Nevertheless, we do have confidence in the accuracy of the change map based on the following facts. *First*, Landsat images are proven capable of accurately differentiating between forest land and non-forest land (Kim et al. 2014). *Second*, training examples for the forest classes and non-forest classes in the supervised classifications were carefully selected (as discussed above). *Third*, support vector machine is a well-tested supervised classification method that can achieve high land cover/use classification accuracy (Huang et al. 2002; Pal and Mather 2005; Mountrakis et al 2011). *Fourth*, we delineated areas where tree stump traces indicating forest to non-forest transitions were observed during field work in the area. The tree stump trace areas align very well with the identified transition areas on the change map (Figure S2). *Finally*, the overall spatial pattern of the change map matches our knowledge of the area (Huang 2009): transitions between forest and non-forest were distributed mainly near villages in low elevation areas (e.g., forests encroached upon by farmlands) and close to the transition zones between forests and high elevation pastures (e.g., cutting forests for firewood).

S2.2 Identifying areas free of significant change in human-posed disturbance

Areas free of significant change in human-posed disturbance were identified based on cost distance to the closest county road or village where slope gradient is the cost. Hunting was also a major human-posed threat to *R. bieti* (Long et al. 1994). However, it is very hard to quantify the spatial distribution of hunting intensity because data directly reflecting hunting activities were not available. Thus, we did not consider hunting intensity in identifying areas free of significant change in human-posed disturbance. Nevertheless, we did expect that, as a proxy, the distance to road or village can represent hunting intensity to some extent. A 30-meter resolution DEM was created from the contours digitized from 1:50,000 topographic maps of the study area, and slope gradient (%) was derived from the DEM in ArcMap. For each period, the cost distance was computed based on the DEM and the roads and villages current to that period (Figure 3). Based on our knowledge of *R. bieti* in the study area, a threshold of 100 m was applied to each of the computed cost distance layers to convert a continuous distance layer into a binary map indicating *high human-posed disturbance* (< 100 m) and *low human-posed disturbance* (> 100 m). Locations (pixels) that always fell into the same disturbance category throughout the three periods constitute areas where there was no significant change of human-posed disturbance over the three historical periods.

S2.3 Identifying areas free of significant environmental change

Finally, the non-change areas free of significant environmental change across the three historical periods were identified by intersecting areas free of transitions between forest and non-forest with areas free of significant change of human-posed disturbance. The identified non-change areas took up about 60% of the whole study area (Figure S3).

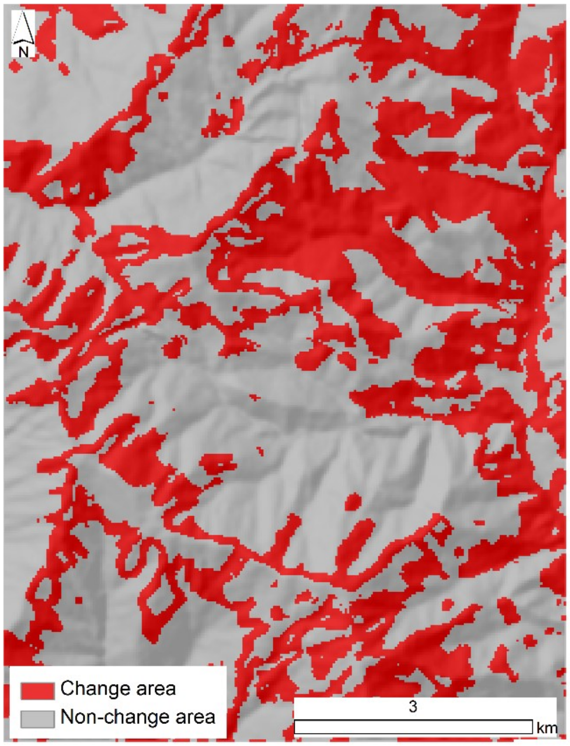

Figure S3. The identified non-change areas where there was neither transitions between forest and non-forest nor significant change of human-posed disturbance across the three historical periods.

References

Chander G, Markham B L, and Helder D L 2009 Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote sensing of environment* 113: 893–903

Coppin P, Jonckheere I, Nackaerts K, Muys B, Lambin E 2004 Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing* 25: 1565–1596

Hansen M C and Loveland T R 2012 A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment* 122: 66–74

Huang C, Davis L S, and Townshend J R G 2002 An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23: 725–749

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

Kim D H, Sexton J O, Noojipady P, Huang C, Anand A, Channan S, Feng M, and Townshend J R 2014 Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sensing of Environment* 155: 178–193

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

Masek J G, Vermote E F, Saleous N E, Wolfe R, Hall F G, Huemmrich K F, Gao F, Kutler J, and Lim T K 2006 A Landsat surface reflectance dataset for North America, 1990-2000. *Geoscience and Remote Sensing Letters, IEEE* 3: 68–72

Mountrakis G, Im J, and Ogole C 2011 Support vector machines in remote sensing: A review. ISPRS *Journal of Photogrammetry and Remote Sensing* 66: 247–259

Pal M and Mather P M 2005 Support vector machines for classification in remote sensing. *International Journal of Remote Sensing* 26: 1007–1011

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