Conservation of endemic species in China

by

Binbin Li

Nicholas School of the Environment
Duke University

Date: ______________________

Approved:

___________________________

Dr. Stuart Pimm, Supervisor

___________________________

Dr. Jeff Vincent

___________________________

Dr. Alex Pfaff

___________________________

Dr. Jennifer Swenson

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Chemistry in the Graduate School
of Duke University

2017
ABSTRACT

Conservation of endemic species in China

by

Binbin Li

Environment Program
Nicholas School of the Environment
Duke University

Date: ______________________

Approved:

___________________________
Dr. Stuart Pimm, Supervisor

___________________________
Dr. Jeff Vincent

___________________________
Dr. Alex Pfaff

___________________________
Dr. Jennifer Swenson

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2017
Abstract

China is one of the most biodiverse countries in the world, harboring more than 10% of the species in the world. Among them, 11% of the vertebrate genera and 7% plant genera are endemic to China. During its rapid social and economic development, increasing habitat loss and fragmentation have occurred. However, it wakes up to the threats of biodiversity in recent years. Protected areas, as an essential conservation tool to reduce habitat loss and species extinction have expanded dramatically in China. Protected areas with various other concepts such as umbrella species and payment for ecosystem services have been promoted to conserve the biodiversity. However, questions remain that whether they work, how they work and how we could do better. It is crucial to answer these questions with the data and technology that are more available to us now.

Thus, my dissertation divides into four chapters and tackles the following four questions. 1) Where do the most of the endemic species concentrate in China? Do umbrellas species such as giant pandas effectively protect other species? 2) With the increasing of tree plantation and available remote sensing data, how does it change the available habitat for forest species, their threat levels and priority setting? 3) Within the conservation priority areas, new threats that are hardly detected by traditional evaluation index such as forest cover emerge. How does a prevalent human disturbance
livestock grazing impact the conservation of giant pandas? What are the socio-economic drivers and solutions to this issue? 4) To better monitor the population and evaluate conservation efforts, new techniques need to be added. Can we use footprints from wild pandas to identify individuals and provide a cost-effective alternative to the current methods?

In Chapter 1, I first used detailed data on geographical ranges for endemic forest species to identify patterns of species richness. After refining each species’ range by its known elevational range and remaining forest habitats as determined from remote sensing, I identified the top 5% richest areas as the centers of endemism. Over 96% of the panda habitat overlapped the endemic centers. Thus, investing in almost any panda habitats will benefit many other endemics. Existing panda national nature reserves cover all but one of the endemic species that overlap with the panda’s distribution. For whole China, of particular interest are 14 mammal, 20 bird, and 82 amphibian species that are inadequately protected. Most of these the IUCN currently deems threatened. But 7 mammal, 3 bird, and 20 amphibian species are currently non-threatened, yet their geographical ranges are <20,000 km² which is the threshold for IUCN to consider it as threatened. There is a high concentration of these species in the east Daxiang and Xiaoxiang Mountains of Sichuan where pandas are absent and where there are no national nature reserves. The others concentrate in Yunnan, Nan Mountains and Hainan.
Here, ten prefectures might establish new protected areas or upgrade local nature reserves to national status.

In Chapter 2, I used a remote sensing product to differentiate oil palm and rubber plantations from natural forests in Southeast Asia and reevaluated the threat level of endemic forest species identified by IUCN. Tropical, mainland Southeast Asia is under exceptional threat, yet relatively poorly known. This region contains over 122, 183, and 214 endemic mammals, birds, and amphibians, respectively, of which the IUCN considers 37, 21, and 37 threatened. When corrected for the amount of remaining natural habitats, the average sizes of species ranges shrink to <40% of their published ranges and more than 42 percent of species face a much higher risk of extinction from habitat loss than previously thought. Moreover, these species are not better protected by the existing network of protected areas than are species that IUCN accepts as threatened. Furthermore, incorporating remote sensing data showing where habitat loss is prevalent changes the locations of conservation priorities.

Chapter three focuses on a specific threat - livestock grazing in the endemic center that I identified in the first chapter. With the Natural Forest Conservation Program and Grain to Green programs, the deforestation that was once the biggest threat to pandas has been halted. However, a previously unrecognized threat is emerging. Livestock grazing has become the most prevalent human disturbance throughout panda habitats. I applied wildlife sign surveys, vegetation surveys, GPS
collar tracking, and species distribution modeling to study how the livestock grazing impacts the habitat use of giant pandas. This study shows that livestock grazing especially from horses has caused a dramatic decline in bamboos and reduced its regeneration. In the past 15 years, pandas have changed their habitat use and are driven out of areas that are heavily used by livestock. About 49% of panda habitat has been lost especially in the lower elevation areas from 2004 until now due to impacts of livestock. Several natural policies and projects such as Natural Forest Conservation Project, Grain for Green projects and ICDP, dam construction as well as the 2008 earthquake, encouraged horse riding practice during the development of ICDP have contributed to the increasing dependence on livestock sector. Livestock ban with payment for ecosystem services or feedlot operation could be possible solutions for this issue.

Chapter four explores the innovative technique to identify giant panda individuals to facilitate better conservation. Two methods have been used previously to identify individuals and population for giant pandas, fecal bamboo bite size combined with home range analysis and microsatellite analysis of fecal DNA. However, the first one suffers from the lack of accuracy and the latter one is limited by the freshness of the fecal sample and high processing cost. I developed the footprint identification technique in JMP based on two multivariate methods: discriminant analysis and the canonical centroid plot method using the anatomy measurements of footprints. I used 30 captive pandas to develop the algorithm and 11 individuals for validation. The overall accuracy
of FIT for individual identification is 90% and sex discrimination is 85%. This technique is embedded in FIT as an add-in and free for conservation practitioners now. In summary, this dissertation includes the following four papers.


Chapter 3, Li et al., Emerging threat from livestock on giant panda conservation

Chapter 4, Li et al., Identifying individual and sex of giant pandas through Footprint Identification Technique.

With supporting information from the following publication during my Ph.D.:


Contents

Abstract .............................................................................................................................................. iv

List of Tables ......................................................................................................................................... ix

List of Figures ........................................................................................................................................ x

Acknowledgements .............................................................................................................................. xiii

Chapter 1. China’s endemic vertebrates sheltering under the protective umbrella of the giant panda ................................................................................................................................. 16

1.1 Introduction ..................................................................................................................................... 16

1.2 Methods ........................................................................................................................................ 20

1.2.1 Study area and species ............................................................................................................. 20

1.2.2 Patterns of biodiversity in China ............................................................................................... 21

1.2.3 Range map ............................................................................................................................... 22

1.2.4 Nature reserve and gap analysis ............................................................................................... 23

1.2.5 Role of the giant panda ........................................................................................................... 24

1.2.6 Identifying priorities for future conservation .......................................................................... 24

1.3 Results ........................................................................................................................................... 25

1.3.1 Refining the IUCN ranges by elevational range and available habitat ................................. 28

1.3.2 Centers for endemic forest species in China ........................................................................... 28

1.3.3 Gap analysis for national nature reserves .............................................................................. 31

1.3.4 Coverage by giant panda national nature reserves ............................................................... 34

1.3.5 Concerned species for future conservation ............................................................................ 34

1.3.6 Priority settings for endemism ............................................................................................... 35
1.4 Discussion.............................................................................................................................................36
1.4.1 Patterns of biodiversity in China .................................................................................................36
1.4.2 The giant panda as an umbrella .................................................................................................37
1.4.3 Future species of concern ............................................................................................................38
1.4.4 Data for systematic conservation planning ...............................................................................38
1.4.5 Protected areas and gaps ............................................................................................................39
1.5 Supporting information....................................................................................................................40
1.6 Citation................................................................................................................................................40

Chapter 2. Remotely sensed data informs red list evaluations and conservation priorities in southeast Asia .........................................................................................................................41
2.1 Introduction.........................................................................................................................................41
2.2 Methods ................................................................................................................................................46
2.2.1 Study area and species .................................................................................................................46
2.2.2 Elevational range and suitable habitat .......................................................................................47
2.2.3 Forest ..............................................................................................................................................48
  2.2.3.1 Errors of omission ...................................................................................................................51
  2.2.3.2 Errors of commission ..............................................................................................................51
2.2.4 Comparison with other global forest products ............................................................................52
2.2.5 Refining ranges and patterns of biodiversity ..............................................................................52
2.2.6 Difference in conservation priorities after incorporating more information.................53
2.3 Results ...............................................................................................................................................54
  2.3.1 Elevation, forest, and protected areas .......................................................................................54
2.3.2 Refining ranges ........................................................................................................... 54
2.3.3 Protection for different taxa ..................................................................................... 60
2.3.4 Patterns of biodiversity and conservation priority ................................................. 61
2.3.5 Species of concern ...................................................................................................... 62
  2.3.5.1 Data deficient Species ....................................................................................... 64
  2.3.5.2 Difference in conservation priorities after incorporating more information .... 64
2.4 Discussion ....................................................................................................................... 66
  2.4.1 Knowledge from remote sensing modifies conservation priorities ............... 70
  2.4.2 A modest, practical proposal .................................................................................. 72
2.5 Supporting information .................................................................................................. 73
2.6 Citation .......................................................................................................................... 73

Chapter 3. Emerging threat from livestock grazing on giant panda habitats .......... 74
  3.1 Introduction .................................................................................................................. 74
  3.2 Methods ....................................................................................................................... 76
    3.2.1 Study area ............................................................................................................... 76
    3.2.2 Data collection ........................................................................................................ 77
      3.2.2.1 Survey plot ....................................................................................................... 77
      3.2.2.2 Bamboo data .................................................................................................. 78
      3.2.2.3 Monitoring data ............................................................................................. 79
      3.2.2.4 Collar data ..................................................................................................... 79
    3.2.3 Data analysis ........................................................................................................... 80
3.2.3.1 Impacts of livestock grazing on panda habitat use over time ............... 80
3.2.3.2 Species distribution model .................................................................. 80
4.2.3.3 GPS collars .......................................................................................... 83
3.3 Results ........................................................................................................ 84
3.3.1 Impacts on bamboos .............................................................................. 84
3.3.2 Habitat use of livestock ......................................................................... 85
3.3.3 Changes in habitat use for pandas ....................................................... 91
3.3.4 Habitat use for giant panda ................................................................. 92
3.3.5 Changes between current and previous panda habitats .................... 93
3.3.6 GPS tracking ......................................................................................... 96
3.4 Discussion .................................................................................................. 97
3.4.1 Emerging threat from livestock .......................................................... 97
3.4.2 Diagnosis and possible policy changes ............................................. 100
Chapter 4. Identifying individual and sex of giant pandas through Footprint
Identification Technique .................................................................................. 103
4.1 Introduction ............................................................................................... 103
4.2 Methods ..................................................................................................... 106
4.2.1 Study population .................................................................................. 106
4.2.2 Study period ......................................................................................... 107
4.2.3 Foot anatomy and data collection ...................................................... 107
4.2.4 The footprint collection protocol ...................................................... 109
4.2.5 Extracting a geometric profile from footprint images ..................... 110
**List of Tables**

Table 1: Summaries of the numbers of species in each category and overlap of areas.. 30

Table 2: Number of species in each taxon in different categories and coverage by protected areas. .............................................................................................................. 59

Table 3: Numbers of species that have < 10% or > 80% of their refined ranges protected. .................................................................................................................................... 60

Table 4: Final variables used in the model and description.................................................. 81

Table 5: Summary for livestock distribution models. See Table 4 for variable definitions ....................................................................................................................................... 89

Table 6: Comparison between two periods of panda occurrences ........................................... 92

Table 7: Panda distribution model with livestock interaction.................................................. 93

Table 8: Model comparison for “Before” and “After” periods.................................................. 93

Table 9: Summary for 30 pandas............................................................................................... 115
List of Figures

Figure 1: Species richness maps of China. (A-C) show all the species that occur in China for mammals, birds, and amphibians. The second row (D-F) shows the endemic species (see text for definition). The third row (G-I) shows the distribution of endemic forest species, and the fourth row (J-L) shows the nonforest species. (M) shows the elevational range of China, (N) is the distribution of forests, and (O) shows the distribution of both forests and shrublands. The land covers types are from ESA 300 m Land cover map. .......................... 27

Figure 2: (A-C) show the numbers of endemic forest species after refined by elevational ranges and suitable habitats. The second row (D-F) shows the endemic centers for each taxon (see text for definition). The third row shows the distribution of national nature reserve (G), overlap of endemism centers for 3 taxa (H) and the relationship between these areas and the giant panda (Ailuropoda melanoleuca) distribution (I). .............................. 29

Figure 3: Distribution of endemic species in conservation gaps. These species are either threatened species or non-threatened species with remaining ranges smaller than 20,000 km$^2$. (A-C) shows gap mammal species, gap bird species and gap amphibian species. (D) is for all gap species. Grey polygons show the distribution of national nature reserves in these areas. (E-H) are enlargements of the high concentration area of gap species. (E): Sichuan; (F): the Nan Mountains; (G): Yunnan; (H): Hainan. ................................................. 34

Figure 4: Study area. (A) Boundary of the study area. (B) Elevation with names of major mountain ranges. (C) Protected Areas from WDPA. (D) Forest cover according to our definition. (E) Forest classified by ESA. (F) Comparison between the two forest maps. 47

Figure 5: Flow chart of the analysis performed in this paper ......................................................... 50

Figure 6: Percentage of species with different levels of remaining ranges after each step of refining.......................................................................................................................... 55

Figure 7: Number of species in each IUCN category, with their original ranges, then after refinement by elevation and then remaining forest. The scale is square root transformed. The dotted lines divide the ranges into three groups, 0-5,000 km$^2$, 5,000-20,000 km$^2$, and above. When a species range falls under 20,000 km$^2$, we treat it as a species of concern. 58

Figure 8: Patterns of biodiversity in Southeast Asia. (A-C) are using the original range maps from IUCN of BirdLife International for all the endemic species. (D-F) are using the range maps after refining by elevational range and forest for each species. (G-I) are using the refined range maps for the species of concern (threatened species and non-
threatened species with range < 20,000 km$^2$. (J-L) are using the refined range maps for data deficient species..........................................................63

Figure 9: Comparisons of endemism centers using different range maps. The two endemism centers for each taxon were identified using the original ranges and updated ranges respectively. Black areas are agreed upon by both methods. Red areas are the endemism center identified only by overlaying the updated ranges and blue areas are for original ranges only..........................................................65

Figure 10: Study area ..........................................................................................................................77

Figure 11: Impacts on Bamboos. The figure on the left shows the percentage of plot being grazed by livestock and the one on the right shows the percentage of leaves on a stem being grazed. Data were classified into four categories for both measurements, 0-24%, 25-49%, 50-74% and 75-100%. The bar in each paragraph represents the percentage of plots falling into each of the four categories. For example, about 20% of the plots with both cattle and horses had less than 24% of their area grazed........................................84

Figure 12: Height and diameter of bamboos with and without grazing. The lines on the box correspond to the 3rd quartile and 1st quartile and white bars are the mean. Whiskers drawn to the furthest point within 1.5 x IQR (3rd quartile minus 1st quartile) from..........................................................87

Figure 13: Impacts of livestock grazing on the number of shoots per clump. The lines on the box correspond to the 3rd quartile and 1st quartile and white bars are the mean. Whiskers drawn to the furthest point within 1.5 x IQR (3rd quartile minus 1st quartile) from the box..........................................................88

Figure 14: Cattle and horse distribution in summer and winter..........................................................90

Figure 15: Comparison between panda habitats before and after livestock number has increased, overlaid with GPS collar data from horses and cattle..........................................................96

Figure 16: Foot anatomy and footprints of the giant panda. A) digital pads; B) metacarpal pad; C) thumb pad; D) carpal pad; a-d, prints of corresponding pads. ........109

Figure 17: Landmark points and computed derived measurements. At the left, the 7 landmarks that are needed to input manually, 1-5 on centroids of the digital pads, 6 on centroid of the thumb pad and 7 on the distal end of the carpal pad indentation. At the
right, the 7 landmarks with the derived measurements and other variables automatically generated by FIT. ................................................................. 111

Figure 18: Two-way canonical plots. The red points are from all the female individuals in the database and blue points are for all the males. Each single point represents a footprint. Each ellipse is for the 95% confidence interval around the mean for a different trail showing trail data. In A, the green and red ellipses are from the same individual Junzhu, a female, and in B, two different individuals (blue for the male Wu Jun and red for Jun Zhu)............................................................................................................. 112

Figure 19: Dendrogram of individual identification. Each row in the dendrogram represents a trail from a known individual. The letter following the name of an individual shows the ID of different trails. This shows the results without changing the default setting for giant pandas, which correctly predict 6 individuals. The diamond shows the threshold value. Any trails branch out right to it are identified as the same individual............................................................................................................. 117

Figure 20: Result of a comprehensive holdback trial for 30 pandas. If the test set size equals 3, then the training set size is 27 to build the model. The process was iterated 10 times for each test set size. The red line represents the 1:1 line of the true values. The grey triangles are the 10 predicted values at each test set size. The black squares are the means of predicted test set size. The trend is shown by linking the points. ..................... 118

Figure 21: Cluster dendrogram for 32 individuals including trails collected in snow from two known individuals Yeye and Zhangka from a semi-enclosure. The model predicted 29 individuals with the two added individuals, identified correctly as separate individuals. The highlighted trails are from these two individuals that are semi-enclosure in the natural habitats. The diamond shows the threshold value. Any trails branch out right to it are identified as the same individual. ................................................. 120

Figure 22: Model accuracy with varying numbers of variables used in the sex discriminant analysis. The accuracy increases with the number of variables in the model until it gets to 35 variables............................................................................................................. 122

Figure 23: Canonical plot from sex discrimination. Each red point is for female footprint and blue is for male. The separation of 95% confidence interval ellipses shows the success of differentiating sex............................................................................................................. 123
Acknowledgements

This dissertation and my whole growth to become an independent researcher is not possible without the support from my major advisor Dr. Stuart Pimm. He encouraged me to apply for his Ph.D. when I was lost and doubted if females could excel in this area. Stuart values conservation, good research, hard work as well families, healthy life, public communication, extra experiences that could enrich one’s life, which I am so grateful for as I share the same values. More importantly, the freedom he gives me pushes me out of my comfort zone to search for interesting ideas, develop and complete research independently and get recognized by my colleagues.

I would also like to thank my committee members Dr. Alex Pfaff, Dr. Jennifer Swenson and Dr. Jeff Vincent. On the way of trying to do interdisciplinary research, they have offered tremendous help. Jennifer’s door has always been open and jammed by my sporadic questions. Jeff has been so patient and guided me through the mist of doing research in an area that I am not familiar with. The discussion with Alex always sparks new ideas and is very enlightening.

My five years at Duke would not be happy without the support from my husband, Jonathan Wang. From boyfriend to husband, he keeps supporting my dream to do conservation, although I was not reachable some months each year for fieldwork. We have been in long-distance throughout my Ph.D. study. However, just like what
Roger de Rabutin de Bussy said, “Absence is to love what wind is to fire; it extinguishes the small, it inflames the great.” We have become better ourselves. Thanks for not being here my dear love, I had not much to do but to finish my dissertation.

I would also to thank for my parents who always believe in me and offer me warm harbor physically and psychologically whenever I need. My amazing lab mates, Dr. Andrew Jacobson, Dr. Clinton Jenkins, Dr. Natalia Ocampo-peñuela, Alexandra Sutton, Varsha Vijay, Ryan Huang and Ruben Palacio, thanks for all your input, help, jokes and drinks that make our group really like a family to me. Thanks for my friends in Nicholas School, Anna Braswell, Danica Schaffer-Smith, Maria de Oca, Bijan Seyednasrollah, Chen Xing for their support through various stages of the Ph.D. study. I have learned a lot from my colleagues and got inspired by them, such as Dr. Lu Zhi, Dr. Li Sheng, Dr. Jack Liu, Dr. Ouyang Zhiyun, Dr. Xu Weihua, Dr. Edward Wilson, Dr. Peter Raven, Dr. Dean Urban, Dr. Randy Kramer and many others.

This dissertation is not possible without the support from Wanglang National Nature Reserve. The managers Chen Youping, Jiang Shiwei, Zhao Lianjun and my field guide Luo Chunping have provided unlimited support to make the project great. Thanks for my funding sources, Hongkong Ocean Park Conservation Foundation, Chinese Scholarship Council, State Forestry Administration, Explorer’s Club, and Duke University.
Last but not the least, I would like to thank all the failures and difficulties that I have been through. No matter they were the cheap and broken GPS collars, or the bankrupted company producing them, or a fancy model that never worked, or a sadly small interview sample size although I have tried to slave my volunteer team to get it to this big, or the roosters that I have sacrificed to pray the rain to stop, they are great treasures for my life. They make me patient, realistic, listen and observe, and understand the truth is always more complicated.

Thanks for the wonderful five years, I have no regret in any moments, and I will carry on the desire of using science to inform conservation work.
Chapter 1. China’s endemic vertebrates sheltering under the protective umbrella of the giant panda

1.1 Introduction

China is exceptionally diverse. It holds many ecosystems types (Liu et al. 2003) and harbors 15% of the world’s vertebrate species and 12% of all plant species (Kram et al. 2012). While China is confronting serious environmental problems during its rapid social and economic development (Wu et al. 2014), it has emphasized conservation of biodiversity in recent years (Wandesforde-Smith et al. 2014; Wang et al. 2007; Xie et al. 2014). Importantly, the total protected area has expanded 35-fold since 1980 (Liu et al. 2003; Liu & Diamond 2005; Xie et al. 2014; Xu & Melick 2007; MEP 2013). We ask: how well has China allocated these areas to protect its diversity?

A simple part of the answer is that like other nations, it has protected high elevation and sparsely vegetated lands disproportionately (Xie et al 2004; Jenkins & Joppa 2009). Unlike other nations, China has the global icon of endangered species, the giant panda, to which China and international organizations devote exceptional resources. How well does this emphasis protect other vertebrate species?

By 2010, China had dedicated 18% of its land to conservation in >8,000 protected areas (Xie et al. 2014). These include nature reserves, forest parks, scenic areas and national parks. They vary in the level of legal protection and effectiveness. Thus, they differ in their capability to protect effectively their species (Singh 1999; Xie et al. 2014). Nature reserves are the most important as well as the best protected class (Xu & Melick,
By 2013, China had established 2669 nature reserves covering 1,497,900 km\(^2\) (This excludes Taiwan, Macau, and Hong Kong). Of this, the Ministry of Environmental Protection (MEP) estimated that in 2013 1,433,800 km\(^2\) was terrestrial, some 14.9% of the land area. The remainder was marine. This is substantial progress towards the Convention on Biological Diversity’s Aichi Target 11, which seeks to protect 17% of a country’s area (CBD, 2011). What should China protect next to meet that target?

Some 69% of this total is in the four western and northern, arid or high (or both) and sparsely populated regions of Tibet, (413,689 km\(^2\)), Xinjiang, (214,944 km\(^2\)), Qinghai (218,222 km\(^2\)), and Inner Mongolia (136,890 km\(^2\)). The majority of the reserves are 10-100 km\(^2\) (40%) and 100-1000 km\(^2\) (34%). The largest size category (>10,000 km\(^2\)) contains only 19 nature reserves, but covers 58% of the total reserve area (MEP 2013). Nature reserves are small in central and eastern China, where human activities dominate (Wu et al. 2014).

There are two administrative levels: national and local. The latter includes provincial nature reserves, municipal nature reserves and county nature reserves. National nature reserves cover 62.5% of the protected areas and 9.8% of the total land area (MEP 2013). Moreover, while only 52.2% of the local nature reserves have established managements, all national nature reserves have established corresponding entities (MEP 2013). Compared to local nature reserves, national-level reserves usually receive more financial resources (Li et al. 2013).
In addition to this uneven geographic coverage, there is an uneven coverage for species conservation. From 2007 to 2014, the number of nature reserves targeted mainly at giant pandas doubled from 34 to 67 (State Forestry Administration, 2015). They now cover 33,600 km² — some 7.5% of the 450,055 km² of protected areas that lie outside the four less populous provinces listed above.

Charismatic and threatened species, especially giant pandas, have drawn disproportionately from other conservation resources (Lu et al. 2000). Around 45 pandas are loaned to other countries with an annual fee about $1 million per pair to support their conservation in China. In addition to these fees from zoos, in 1993, the Chinese government created the National Panda Program with US$5 million support and announced a new Wildlife Conservation and Nature Reserve Development Program (WWF, 2004). It listed the panda first among 15 targeted flagship species in 2002. This plan aims to improve the infrastructure, management, education, research and monitoring capacity of the existing 34 panda nature reserves, establish 28 more panda reserves, a monitoring station for each of the 55 counties with panda distribution, breeding and reintroduction centers and encourage research. The main budget includes US$5 million per new nature reserve and breeding center from 2001-2010 and US$7.5 million from 2011-2030 (National Development and Reform Commission, 2007). A national panda survey takes place every decade lasting 3-4 years. It covers >43,600 km² and with DNA analysis in the survey starting in 2011 (State Forestry Administration,
In addition, large, International non-governmental organizations have been actively involved in panda conservation since the early 1980s.

Does emphasizing a single species jeopardize China’s vast array of biodiversity (Kram et al. 2012)? Might the focus on the panda harm less attractive species that are under-represented in the conservation agenda, such as small mammals (Ceballos & Brown 1995; Entwistle & Stephenson 2000)? Giant pandas are widely recognized as flagship species with their possibility to protecting broader biodiversity (Lu et al. 2000). Yet, there is no study that quantifies the effectiveness of how much of the range of other species — especially other forest species — it can protect.

To assess effectiveness, we map the distributions of amphibians, birds, and mammals to ask several questions. First, what are the patterns of species diversity and endemism in China? As in previous studies, (Harris & Pimm 2008; Ocampo-Penuela & Pimm 2014; Schnell et al. 2013) we start with published range maps then see how data on elevation preferences and remaining vegetation cover modify these distributions. Second, we consider the species that IUCN defines to be threatened. As in previous studies, we assess whether other species might be added to this list in light of our updated predictions of their remaining ranges.

Third, we ask how well the giant panda protects China’s other vulnerable species. Finally, we ask how well the existing national nature reserves cover these crucial regions of endemism and how one might make improvements.
1.2 Methods

1.2.1 Study area and species

We compiled the species lists for terrestrial mammals and amphibians from the International Union for Conservation of Nature (IUCN, 2014a) and for terrestrial birds from BirdLife International (BirdLife International, 2014) for China. We include Hong Kong, Macao, Hainan and Taiwan. Some 608 mammal, 1342 bird and 394 amphibian species have part or all of their ranges in China.

To define which species are endemic to China, we created a 100km buffer around its political boundary. If a species’ range entirely falls in this area and has more than 80% of its range in China, then we defined it as endemic. This includes species that are mostly within China, but range outside, while it excludes species with very small ranges that fall mostly within the 100 km buffer. For birds, we only considered their resident or breeding ranges. So defined, our study includes 132 mammal species, 117 birds species and 250 amphibian species.

Among these species, the IUCN Red List (IUCN, 2014a) classified 65 mammal, 78 bird and 96 amphibian species as threatened species — that is, vulnerable, endangered, or critically endangered. In addition, IUCN identifies 23 mammals and 62 amphibians as data deficient.
We produced richness maps for each taxon to identify patterns of biodiversity by summing up all the available range maps. Because Oreolalax weigoldi lacks spatial distribution data, we mapped only 249 amphibian species.

1.2.2 Patterns of biodiversity in China

Although the IUCN range maps provide useful initial guidance, they encompass areas that are not suitable habitats for species. Thus, we refined the range maps first by elevational range and then by suitable vegetation type for each species. We excluded species that are data deficient from this step onwards because the data are inadequate to identify their habitats.

We collected elevational data from Birdlife International and the IUCN Red List. To supplement the missing information from these sources, we also searched for individual studies. We used elevational data from the 90m-resolution Digital Elevation Model (DEM) from the NASA Shuttle Radar Topographic Mission (SRTM).

We extracted information about preferred habitats for each bird species from Birdlife International, which lists the principal habitats under a class they call level 1. We extracted corresponding vegetation types from the global land cover map GlobCover Version 2.3 2009 with 300m resolution from the European Space Agency (ESA). Then we produced a species-specific habitat map for further trimming. We compiled similar information for mammals and amphibians from IUCN red lists. Matching habitat
descriptions to land cover classifications need not be easy for they are independent
sources with potentially different definitions.

Excluding data deficient species, data on endemic species suggests that about
half (53 mammals, 49% of total; 73 birds (62%) and 121 amphibians (64%)) are
principally forest species. Often their habitat descriptions include shrublands as well.
Inspection of where shrublands are show that they are usually in the transitions from
forest to other habitats in mountainous areas or areas with historical human disturbance
like logging (Figure 1). Species that use both forests and shrublands mainly depend on
forests: the highest percentage of shrublands in a species’ range is 35% among our target
species. We define forest species as those that exclusively use forest or use both forest
and shrubland as suitable habitats.

Species that use other habitats are not so readily matched to maps of other land
use cover classifications. For these, we cannot confidently trim their ranges by remaining
habitat.

1.2.3 Range map

After refining each species’ range by elevation and habitat, we summed the
ranges to produce a richness map for each taxon. To compare the key areas across
different taxa without the influence from differences in the number of species, we
identified the center of endemism by selecting the top 5% richest land areas of forest and
shrubland for each taxon (Jenkins et al. 2013). Thus, by simply summing up three
centers, we were able to rank the importance of different regions according to the extent of overlap for three taxa.

1.2.4 Nature reserve and gap analysis

Peking University’s Center for Nature and Society compiled the spatial boundary data for national nature reserves. Information about some national nature reserves is incomplete either because the spatial data were not available at that time or because a nature reserve was upgraded afterward. We supplemented these data by digitalizing maps from the website of Ministry of Environmental Protection. There were 395 terrestrial national nature reserves established before 2014. From this section on, we only consider mainland China and Hainan. We exclude Hong Kong, Macao and Taiwan and their endemics because they have different protected area systems. For species occur on both the mainland and these islands, we only analyzed the range on the mainland.

In a conservation context, J. M Scott developed and pioneered implementation of gap analysis in the late 1970s as a consequence of seeing the incomplete coverage of concentrations of threatened Hawaiian birds by the existent network of protected areas (Scott et al., 1987; Scott et al. 1993). It is widely used as a method to identify gaps of protected area network in representing or supporting the survival of target species or ecosystems (Margules & Pressey 2000).

There is a large body of literature on what should be the minimum level of protection. We adopted criteria from Watson et al. (2011). If a species’ geographic range
was <10,000 km$^2$, then the target of coverage was set to be the smaller of 1,000 km$^2$ or 100% of the range. If the range was >10,000 km$^2$, the target was at least 10% of the range.

If a species had less area protected than the target, then for simplicity, we called such inadequately protected species a gap species.

We also evaluated the coverage of current national protected areas in conserving the endemism centers. We calculated the percentages of one, two and three-taxon centers that fell within the national protection network.

1.2.5 Role of the giant panda

We overlapped the giant panda distribution and the endemic centers. Then, we calculated the area of giant pandas’ range that falls in each category, one-taxon, two-taxon and three-taxon center, respectively. We also calculated how the giant panda national nature reserves cover the endemic species that overlap with giant panda distribution.

1.2.6 Identifying priorities for future conservation

For all the endemic species, we focused on threatened species and non-threatened species with their remaining ranges <20,000 km$^2$. We called these endemic species of concern. (Parenthetically, the range size of 20,000 km$^2$ is the threshold to consider the species as threatened species by IUCN. This applies to the range size before being refined by elevational limits and remaining habitats. Nonetheless, we retain this threshold to flag potential threatened status.)
As many species lack information or adequate studies of their life history or population size, range size becomes a key factor in identifying their vulnerability (Harris and Pimm, 2008). We define gap species of concern as those that are both gap species and endemic species of concern. To direct future conservation efforts more effectively, we mapped the distribution of all the gap species of concern and summed up their remaining ranges. This process produced a map to reveal the high concentration of gap species, where future conservation priorities should be.

We applied one-way ANOVA to test the statistical differences between groups. Post-hoc test, Turkey-Kramer HSD was implemented in conjunction with ANOVA for the following three-way comparison. We used Pearson’s chi-squared to compare the composition difference between two samples.

1.3 Results

The greatest numbers of bird, mammal, and amphibian species are in the tropical province of Yunnan, along the border with Myanmar, Laos and Vietnam (Figure 1, A-C). While all three taxa share general trends, they differ in details. Mammal richness reaches farther north to the Qinling (Figure 1A, Supplemental Figure 1 is a map with the places names used in the text). While birds are similar to mammals, high numbers extend more to the east, along the southern coastal lowlands (Figure 1B). Amphibian richness differs from the other taxa and is highest in southeast China (Figure 1C).
Figure 1: Species richness maps of China. (A-C) show all the species that occur in China for mammals, birds, and amphibians. The second row (D-F) shows the endemic species (see text for definition). The third row (G-I) shows the distribution of endemic forest species, and the fourth row (J-L) shows the nonforest species. (M) shows the elevational range of China, (N) is the distribution of forests, and (O) shows the distribution of both forests and shrublands. The land covers types are from ESA 300 m Land cover map.

The highest endemism distributes along easternmost edge of the Tibetan Plateau (Figure 1, D-F). This area is perpendicular to the main Himalayan chain and mainly constitutes of the eastern Hengduan Mountains (which includes the Min Mountains, Qionglai Mountains, Daxue Mountains, Gaoligong mountains). Total richness and endemism differ because the former include species that occur in adjacent countries, while the latter do not (Orme et al. 2005). Endemic mammals concentrate south to the Gaoligong Mountains and north to the Qinling and the Daba Mountains (Figure 1D). Endemic birds concentrate in the central Hengduan Mountains (Min, Qionglai and Daxue Mountains) and Qinling (Figure 1E). Because most of the amphibians that occur in China are endemic species, the pattern of total richness matches that of the endemism. Except for the southern tip of Yunnan Province, amphibians still show a high concentration in southeast China with two distinctive areas (Figure 1F). One is the Wuyi Mountains in Fujian and Jiangxi Province. The other is the Nan Mountains along the boundaries of Guangdong, Guangxi and Hunan Province.

The highest concentrations of forest endemics for all three taxa are in the forest biomes of central and southeast China, especially its montane areas (Figure 1 D-F). The
patterns of total endemism are mainly driven by forest endemic species (Figure 1 H-G). Non-forest mammals and birds have a clear high concentration on the eastern Tibetan Plateau where the most important three rivers of Asia (Yangtze River, Yellow River and Mekong River) originate.

1.3.1 Refining the IUCN ranges by elevational range and available habitat

As one trims the published ranges by elevational limits and available habitats, range sizes shrink (Harris and Pimm, 2008, Schnell et al. 2013, Ocampo-Peñuela and Pimm 2014). Moreover, they sometimes do substantially, especially for species in mountainous regions and where habitat destruction is extensive. Furthermore, some species that IUCN designates as non-threatened sometimes show refined ranges smaller than those it deems as threatened. Appendix 4 in the Supporting Information lists each forest species, its original range, the range trimmed by elevation, then by habitat, how much of its protected, and how much of it is protected by reserves established for pandas.

1.3.2 Centers for endemic forest species in China

Figures 2 A-C show the patterns of forest endemism in China. After trimmed by elevational range and remaining vegetation type, the species' range and the distribution of endemism are obviously highly fragmented. We set the top 5% forest and shrubland area with the highest richness as the endemic center for each taxon (Figure 2, D-F). In addition to the arch of the area formed by the Hengduan, Qinling and Daba mountains,
Yunnan and Taiwan are important areas for endemic mammals (Figure 2D). Similar to mammals overall, the differences between birds and mammals are that instead of Yunnan, the Nan mountains in southeast China and Taiwan are rich in endemic birds (Figure 2E). Amphibian richness is high in the arch but absent from Qinling (Figure 2F). The Nan Mountains, the Wuyi Mountains in Southeast China and Hainan are centers of endemism. The eastern edge of Himalaya in central China, mountainous areas in the southeast and Taiwan are crucial regions for endemics of all three taxa (Figure 2H).

Figure 2: (A-C) show the numbers of endemic forest species after refined by elevational ranges and suitable habitats. The second row (D-F) shows the endemic centers for each taxon (see text for definition). The third row shows the distribution of
national nature reserve (G), overlap of endemism centers for 3 taxa (H) and the relationship between these areas and the giant panda (Ailuropoda melanoleuca) distribution (I).

The giant panda range overlaps with 70% of the forest bird species, 70% of forest mammals and 31% of forest amphibians that are endemic to mainland China (Table 1). Ninety-six percent of its range falls in the endemic centers (Figure 2). Some 13% of its range falls in the center for one endemic taxon, 58% falls in the center for two taxa, and 25% falls in three-taxon center. Thus, by directing conservation efforts into panda habitats could protect 28.6% of the three-taxon center and 17.1% of the two-taxon center.

Table 1: Summaries of the numbers of species in each category and overlap of areas.

<table>
<thead>
<tr>
<th>Number of Species</th>
<th>Mammal</th>
<th>Bird</th>
<th>Amphibian</th>
</tr>
</thead>
<tbody>
<tr>
<td>All species occur in China</td>
<td>608</td>
<td>1342</td>
<td>394</td>
</tr>
<tr>
<td>Endemic species</td>
<td>132 (22%)</td>
<td>117 (9%)</td>
<td>249 (65%)</td>
</tr>
<tr>
<td>Endemic forest species</td>
<td>53</td>
<td>73</td>
<td>121</td>
</tr>
<tr>
<td>Species for gap analysis</td>
<td>45</td>
<td>53</td>
<td>112</td>
</tr>
<tr>
<td>Occur on mainland</td>
<td>43</td>
<td>50</td>
<td>101</td>
</tr>
<tr>
<td>Overlap with giant panda distribution</td>
<td>30</td>
<td>35</td>
<td>31</td>
</tr>
<tr>
<td>Small-range species gap species</td>
<td>11</td>
<td>13</td>
<td>72</td>
</tr>
<tr>
<td>&lt;10,000 km²</td>
<td>2</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>14</td>
<td>73</td>
</tr>
<tr>
<td>Large-range species gap species</td>
<td>18</td>
<td>23</td>
<td>33</td>
</tr>
<tr>
<td>&gt;10,000 km²</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>32</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>Gap species of concern</td>
<td>14</td>
<td>20</td>
<td>82</td>
</tr>
<tr>
<td><strong>Coverage from panda NNR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overlap with panda NNR</td>
<td>31</td>
<td>34</td>
<td>32</td>
</tr>
<tr>
<td>Average coverage</td>
<td>7%</td>
<td>6%</td>
<td>11%</td>
</tr>
<tr>
<td>Single species priority areas</td>
<td>8%</td>
<td>9%</td>
<td>1%</td>
</tr>
<tr>
<td>Any two taxa</td>
<td>13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All three taxa</td>
<td>10%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.3.3 Gap analysis for national nature reserves

Figure 2G shows the distribution of 395 national nature reserves established up to 2014. The concentrations of endemic species lie elsewhere, however. Overall, national nature reserves cover 6.9% of the area that are important for one taxon, 19.2% for two and 13.0% for three on mainland and Hainan. Amphibians have a higher percentage of gap species — those without adequate protection — than do birds and mammals (Table 1). Small-range species are less protected in terms of proportion of gap species as well as the number of species without any protection (Table 1). The nature of small-ranges makes it easier to miss the coverage of these species in the national nature reserve network. Amphibians receive the least protection compared to birds and mammals: 99% of small-ranged amphibian species, and 85% of large-ranged species are not adequately protected. Mammals are the most effectively protected taxon in terms of percentage of species that reach the target coverage.

The national nature reserve system has significant higher coverage for threatened species than for non-threatened species (One-way ANOVA, p<0.003). While non-threatened endemic species have an average of 8.9% ± 6.2% of the range under protection, threatened species are 6.7% higher (15.6% ± 22.5%). Nonetheless, the threatened species are usually small-ranged species. Compared to 22% of the non-threatened species that have ranges <10,000 km², about 77% of threatened species are small-ranged species. As a result, though the coverage of protected areas is higher for
threatened species, yet it is far from the target coverage. This leads to a higher proportion (91%) of gap species for threatened species, compared to 73% for non-threatened species ($\chi^2=10.73$, $p<0.01$).

The gap species of concern include 14 mammal, 20 bird and 82 amphibian species. Interestingly, where these species concentrate strongly overlaps with the richest area for endemism (Figure 3). This is similar to the global pattern (Rodrigues et al. 2004). Most of the concerned gap species of mammals are in the area east to Mount Gongga of central Sichuan and southern Hengduan Mountains in northern Yunnan Province (Figure 3A). Gap species of birds mainly distribute along the edge of Sichuan basin, Hainan and southeastern provinces. The highest concentration of gap species is primarily in Daxiang Ling and Xiaoxiang Ling (south to Qionglai Mountains), and Nan Mountains along the boundaries of Guangxi, Hunan and Guangdong Province (Figure 3B). Gap amphibian species are mainly in the mountainous area to the west of Sichuan basin, central Yunnan and southeastern provinces, a pattern similar to that for birds (Figure 3C). Hainan province shows a high concentration of gap amphibian species.

National nature reserves only cover a small portion of these gap species ranges or miss the whole range of some species (Figure 3D). Local nature reserves cover some of the areas. As many of them lack administrative entities, management and clear boundaries, it is hard to evaluate whether they provide effective protection.
1.3.4 Coverage by giant panda national nature reserves

Except for one bird species, *Phylloscopus emeiensis*, the national nature reserves devoted for the giant panda protection include all the endemic forest species that overlap with the panda distribution. They also cover 25% of the Scutiger ningshanensis’s range and 1% of the Cansumys canus’s range. Neither overlaps with the giant panda distribution. On average, panda reserves protect 7%±7% of these mammals’ range, 6%±5% of birds’ range and 11%±19% of the amphibians’ range. These are about half of average coverage from all national nature reserves for mammals (14%) and bird (11%), but just slightly lower for amphibians (12%).

Panda reserves cover 8% of endemic center for mammals, 9% for birds and 1% for amphibians. They also protect 13% of the two-taxa center and 10% of the three-taxa center.

1.3.5 Concerned species for future conservation

Some species that IUCN identifies as non-threatened and least concerned are likely more vulnerable than expected. When ranges are trimmed by elevation limits and available habitats, four non-threatened mammal species (*Cansumys canus, Chodsigoa
lamula, Chodsigoa smithii, Lepus yarkandensis), three bird species (Certhia tianquanensis, Phoenicurus alaschanicus and Phylloscopus emeiensis) and 10 amphibian species (Amolops daiyunensis, Amolops lifanensis, Oreolalax lichuanensis, Oreolalax schmidti, Oreolalax xiangchengensis, Paramesotriton hongkongensis, Rhacophorus zhaojuensis, Theloderma rhododiscus, Xenophrys binchuanensis and Xenophrys omeimontis) have their remaining ranges falling below 5,000 km$^2$. Only C. tianquanensis is under adequate protection from the national nature reserves.

Six more non-threatened mammal species, three bird species, and 12 amphibian species have their remaining range size less than 20,000 km$^2$. One bird species (Babax waddelli), three mammal species (Blarinella quadratica, Myotis davidii and Niviventer excelsior) and 10 amphibian species (Bufo cryptotympanicus, Ichthyophis bannanicus, Leptobrachium ailaonicum, Odorrana lungshengensis, Oreolalax popei, Paramesotriton caudopunctatus, Tylototritontaliangensis, Xenophrys glandulosa, Xenophrys jingdongensis and Xenophrys mangshanensis) are inadequately protected by the criteria defined in our paper.

1.3.6 Priority settings for endemism

Interestingly, where most gap species concentrate are also the areas of the richest endemism (Figure 3). Setting future priorities to fill in the conservation gaps could also enlarge the protection for endemic species in general. For example, Sichuan has 18.5% area protected (6.0% under national protection), ranking the third in China. (It is just behind Tibet and Qinghai where vast areas have very low human densities.) Because of
the high concentration of endemism, many species are unprotected despite the large percentage of current protection.

There are four major gap species concentrations: central Sichuan, central Yunnan, Hainan and Nan Mountains along the borders of Guangxi, Guangdong, Guizhou and Hunan Province (Figure 3E-H). For Sichuan, the prefecture cities Ya’an, Leshan and Meishan have the most gap species, where pandas are mostly absent (Figure 2I). Some of the areas are protected by provincial nature reserves. These need more resources for their management. Hainan also has a high concentration of gap species especially in the central and western part. (Appendix C lists specific areas).

1.4 Discussion

1.4.1 Patterns of biodiversity in China

China’s endemic species are now largely confined to mountainous areas. These areas suffer less from anthropogenic influences (Korner & Spehn 2002; Tang et al. 2006). Even the road network expansion has been slower there than other areas because of the huge cost (Li et al. 2010). Since the late 1990s, forests have become an important target for protection under the large national level policies such as the National Forest Protection Program. This includes a national logging ban and reforestation projects, and Sloping Land Conversion Program (Xu & Melick 2007). Although the actual effects need to be scrutinized continuously, these efforts have nonetheless reduced large-scale
deforestation (Li et al. 2013). Thus, the topographic complexity coupled with national level policy for forests harbors the remaining habitats for endemic forest species.

Our study shows differences in the centers of endemism from a previous study (Lei et al. 2003). This discrepancy is due to the selection of forest endemic species and the use of different data. We point out that another important area for birds, Nan Mountains that has not been mentioned (Lei et al. 2003). While different taxa share common areas for high species richness, there is an obvious difference in their patterns. Amphibians are distinctive compared to birds and mammals. This raises a concern of using single taxon as a surrogate for all taxa (Rodrigues et al. 2004). If the intention is to inform crucial areas for different groups of wildlife, one should take into account the differences and weigh the importance according to richness and threats for each taxon with concern.

1.4.2 The giant panda as an umbrella

The giant panda serves as an umbrella species: 96% of its range overlaps with the centers for at least one endemic taxon. This overlap means that directing resources to almost any panda distribution area or restoration of forest to connect habitat fragments could lead to the protection of the richest forests for endemics. We caution that although it shows the important role of giant pandas and the existing panda reserves, one should look beyond them. There is a significant gap in Daxiang Ling and Xiaoxiang Ling in Sichuan, a center for all three taxa, but with few pandas.
1.4.3 Future species of concern

Many species currently not listed as threatened likely have smaller ranges than previously thought. These small-ranged species should receive immediate scrutiny. IUCN requires information on population or geographic range to list a species as threatened (IUCN 2012). Compared to range size, population estimates are usually hard to get (Harris & Pimm 2008). Of course, the conservation community should regularly evaluate species ranges and estimate populations sizes. Our methods have limitations, but they shed light on the need for exploration and focus on where assessments need particular care. With the rapid development of land cover products at finer scale with free access and large platforms for sharing species occurrence data regionally and globally, we believe it could promote the conservation work on these under-studied species.

1.4.4 Data for systematic conservation planning

Incorporating more information into IUCN range maps can change priority settings. While the general patterns are consistent, issues may appear when come down to the practical work. From the global Aichi target to the national commitment of 20% of land area protected by China (IUCN, 2014b), conservationists usually set a percentage of land that will be put aside for conservation. The priorities may change, and some key areas will be missing if they do not incorporate necessary information to refine the relatively coarse ranges to inform national strategy.
1.4.5 Protected areas and gaps

Compared to charismatic mammals and birds, amphibians are less attractive to the public as well as to previous conservation planners. The taxonomic bias lead to fewer nature reserves specifically designed to protect them. Conservation planning should pay more attention to amphibians and other neglected taxa. Although the national nature reserve network does a better job in covering threatened species, it is still inadequate to support long-term survival considering their small ranges. Besides continuing work on these threatened species, small-range non-threatened species should be another focus as we have discussed above.

The gaps identified in our analysis not only show a lack of coverage in certain areas, but also reveal a problem with the existing national nature reserves. Even though they are in the right place, many reserves are too small to provide sufficient geographic protection for the surrounding endemic rich areas (Figure 3F). Admittedly, other gaps may be covered with local nature reserves or other forms of protected areas. These local level nature reserves can be the candidates for future national nature reserves. They can also turn into opportunities for the establishment of other forms of protected areas such as national parks (Wang et al. 2012) or private protected areas (Stolton et al. 2014) that China is exploring.
1.5 Supporting information

Detailed Map of China (Appendix A), refining ranges of IUCN (Appendix B), and subregions for future conservation (Appendix C) and species lists and the estimates of remaining habitat and protection (Appendix D) are at the end of this document.

1.6 Citation

This chapter was published in 2016 in the journal Conservation Biology. The complete citation is:

Chapter 2. Remotely sensed data informs red list evaluations and conservation priorities in southeast Asia

2.1 Introduction

The IUCN Red List (IUCN 2015) classifies species into various categories that reflect their risk of extinction. It employs rigorously objective criteria, is transparent, and democratic in soliciting comments on individual species decisions. It aspires to provide a global analysis of the status and distributions of all species based on the best available data and expert analysis for each species. The process updates many species regularly, and the associated data are publicly accessible. As individuals who contribute to these assessments and encourage species-specific research to address limited data, we nonetheless worry that important data do not enter this process as effectively as they might. In particular, the rapid growth in remote sensing provides geographically fine-scale data on elevations and increasingly sophisticated characterisations of land use and land use change (Pimm et al. 2014). Moreover, there are global databases on which areas are protected — the first line of defense at ensuring species survival (Pimm et al. 2001, Joppa et al. 2008).

There are two specific challenges. First, there are taxa and areas that are poorly known, where species distributions are incomplete and out of date, and for which the Red List considers many assessments as data deficient. Second, in regions where human
actions are rapidly destroying natural habitats, conventional species’ assessments might be too slow to respond.

What quantitative measures can we derive from widely available, remotely sensed data that sensibly inform a species’ risk of extinction? Some of these inputs may informally contribute to existing Red List determinations, but we show that we can succinctly and easily include additional quantitative data for each species in a standardised way. Our goal is not to replace the existing process, nor even subject it to unwarranted criticism, but to add additional tools to the process. These add precision and accuracy that until recently were impossible, and which help accurately assess species true distribution and status. Secondly, will our results suggest species of concern — those at considerably greater risk than hitherto appreciated? These assessments are not only important on a species-by-species basis, but by combining distributions of threatened species, we create maps of conservation priorities and assay to what degree these areas are currently protected. Thirdly, we ask how might knowledge from remote sensing modify the priorities? Finally, can we develop a quick and simple method to identify and modify the priority setting in a landscape where human actions are rapidly expanding?

We wish to ask these questions in a region that is under exceptional threat, yet relatively poorly known. For such a region, additional quantitative measures may be particularly helpful. Thus, we consider the birds, mammals, and amphibians of tropical,
mainland Southeast Asia. We define this region to be from the province of Yunnan, China, to eastern India, and south to Singapore. The study area encompasses 2,691,773 km\(^2\) and is in the heart of the Indo-Burma Biodiversity Hotspot and northern Sundaland Hotspot (Myers et al. 2000). We further restricted consideration to those species that are broadly endemic to this region, meaning that >80% of each species’ range occurred there. Endemism has always been one of the key targets for setting priorities, (Brooks et al. 2006). Small ranged species are much more likely to be at risk of extinction (Manne et al. 1999). Furthermore, restricted range species tend to be locally rare within those ranges (Pimm and Jenkins 2010), further increasing a species’ risk of extinction.

There are several reasons to choose these taxa in this region. First, these species form a substantial sample and so a case study for tropical areas globally. This area also has many endemic species (Bickford et al. 2010), and such species are more at risk of extinction than widespread ones (Manne et al. 1999, Myers et al. 2000, Lamoreux et al. 2006, Cardillo et al. 2008, Pimm et al. 2014). This region also has many species that IUCN considers to be data deficient (Koh et al. 2013). We must not only ask if we can identify species with unappreciated risks of extinction but assay if remotely sensed data can aid the threat assessments of species for which we otherwise know too little.

This Southeast Asian region suffers rapid deforestation — the major driving force behind global biodiversity loss (Pimm and Raven 2000). It had 0.48 million ha per year net deforestation from 2000-2010, twice the rate from 1990-2000 (Achard et al. 2014),
and is increasing (Sodhi et al. 2010). Thus, traditional species-by-species determinations of risk may be too slow to recognise constantly changing threats. Furthermore, the deforestation often occurs as conversion of natural forests to commercial tree plantations of rubber, oil palm, and other trees crops (Wilcove et al. 2013, Ahrends et al. 2015). Mainland Southeast Asian countries account for 56% of the global rubber production and 39% of the palm oil production (FAO 2013). Much of the expansion of these plantations occurs at the cost of natural forests and peatlands (Gibbs et al. 2010, Wilcove et al. 2013, Vijay et al. 2016).

The majority of agricultural and forestry expansions are monocultures rather than the traditional mixed agroforestry systems (van Noordwijk et al. 2012). These plantations support few species (Fitzherbert et al. 2008, Gibson et al. 2011) and so have potentially catastrophic biodiversity impacts (Koh and Wilcove 2008, Fox et al. 2012, Warren-Thomas et al. 2015). Unfortunately, rubber and oil palm plantations pose a particular challenge to identify and map at broad geographical scales from remote sensing. In traditional forms of classification, they are almost indistinguishable from natural forests in their spectral reflectance characteristics (Li and Fox 2012), and are hard to differentiate from sparse or degraded forest when they are young (Li and Fox 2012). Although regional efforts have identified these tree plantations (Koh et al. 2011, Li and Fox 2012), most of the global datasets do not differentiate them in forest classifications (Dong et al. 2014).
Finally, this is a fortuitous time to ask these questions. Data on vertebrate species ranges, fine-scale topography, and protected areas have been readily available for some years. Fine-scale tree cover and plantation data, both essential to our estimates of remaining habitat, have only been available since 2014 and 2015 respectively. We expect these datasets to be improved and updated periodically, as satellite technology and classification algorithms continue to be enhanced. The approaches outlined here are designed to adapt to such changes and integrate them into further analyses.

Our methods follow a sequential process that starts from the available species range maps (IUCN 2015). We have applied similar methods to address related questions about other priority areas for biodiversity — Central America (Harris and Pimm 2008), coastal Brazil (Jenkins et al. 2011), the Western Andes of Colombia (Ocampo-Peñauela and Pimm 2014), the USA (Jenkins et al. 2015), and China (Li and Pimm 2016). The first step recognises that in mountainous regions, such as here, the elevational limits of species reduce the actual habitat below what the typical range-wide maps show — and often by an order of magnitude. Moreover, the nature of montane landscapes is such that actual ranges are also substantially fragmented. Next, to the extent possible, we further refine species’ ranges by the amount of remaining habitat. We consider species in this region that are forest-dependent, because they are the majority of species and forests are relatively easy to identify compared to other land cover types like shrublands or grasslands.
2.2 Methods

2.2.1 Study area and species

The study area includes Yunnan Province of China, Arunachal Pradesh, six eastern states of India (Assam, Manipur, Meghalaya, Mizoram, Nagaland, and Tripura), Laos, Vietnam, Thailand, Cambodia, Myanmar, Peninsular Malaysia, and Singapore (Figure 4A). We defined the species endemic to this area as those that have more than 80% of their range (breeding range for birds) within it. We compiled the species lists and range maps for terrestrial mammals and amphibians from the IUCN (IUCN 2015), and for birds from Birdlife International (BirdLife International 2015). In total, 209 birds, 165 mammals, and 286 amphibians were endemic to this region. Among them, the IUCN has classified 30 birds, 43 mammals, and 40 amphibians as being threatened: critically endangered, endangered, or vulnerable.
Figure 4: Study area. (A) Boundary of the study area. (B) Elevation with names of major mountain ranges. (C) Protected Areas from WDPA. (D) Forest cover according to our definition. (E) Forest classified by ESA. (F) Comparison between the two forest maps.

2.2.2 Elevational range and suitable habitat

We compiled elevational ranges and habitat preferences from IUCN for mammals and amphibians, and from (del Hoyo et al. 2015) for birds. To be conservative
and reduce bias caused by the few records of some species, we applied a minimum of
500m elevational width. If a species’ elevational range fell below this width or only one
listed elevation was available, we took the midpoint and extended 250m on both sides.
As the lower end does not extend below sea level, the lowest elevational range was set
to 0-500m. Some species did not have enough information for their elevational ranges. If
the only available information was that a species occurred in “lowland” or “montane,”
we used 0-1000m and >1000m respectively. If there was no information, we did not put a
limitation on its elevational range.

We compiled habitat information from the classification schemes in individual
species accounts in available databases (del Hoyo et al. 2015, IUCN 2015). Only habitats
deemed suitable in published accounts were included. We defined forest species as
those that only depend on forests, as well as mammal species that depend on both forest
and caves, and amphibians that depend on both forest and wetland. In total, we
included 183 birds, 122 mammals, and 213 amphibians as forest species. Among them,
21 birds, 37 mammals, and 37 amphibians are currently listed as threatened species and
a further 39 mammals and 111 amphibians are listed as data deficient.

2.2.3 Forest

Southeast Asia has a range of forest types, ranging from relatively sparse dry
forests to dense rainforests. We obtained data on continuous tree-cover for 2005 (Sexton
et al. 2013) for the entire region. We then reclassified the percent tree cover to produce
four forest classifications that used 30, 40, 50, and 60 percent cover as the cut-off for whether we consider an area forested or not.

Sexton et al. (2013) show the forest cover of ten years ago, and since then there have been major changes. We updated these forest cover estimates using Hansen’s forest loss data (Hansen et al. 2013) that shows deforestation since 2005. We further excluded the areas that Li and Fox (Li and Fox 2012) identified as rubber plantations and areas classified as oil palm plantations from Miettinen et al.’s study (Miettinen et al. 2012). Outputs from these steps are forest layers showing 30%, 40%, 50%, and 60% forest cover.

The minimum tree cover needed to define “natural forest” differs widely and the discrepancies are mainly in regions of relatively sparse tree cover (Sexton et al. 2015). In our study area, the forests in dry areas tend to be more open because of their different species compositions (Wohlfart et al. 2014). Lower tree cover thresholds are applied to identify forests in drier regions. To select forested areas correctly, we complemented our layers by testing our four forest layers with different tree cover thresholds against annual rainfall from WorldClim (Hijmans et al. 2005). For regions with different average annual rainfall, we chose the best performing tree cover thresholds to identify forests (Figure 5).
Figure 5: Flow chart of the analysis performed in this paper.

To assess the resultant map of forest cover, we need to address two kinds of errors. The first are errors of omission — places where we exclude natural forest areas...
from a species’ range. The other are errors of commission, where we include as forests land covers such as plantations that are likely unsuitable habitats.

### 2.2.3.1 Errors of omission

To address these, we chose points within protected areas across a 1° latitude-longitude grid across the entire area. We then examined each point using high-resolution Google Earth imagery to ensure that all areas within 90m of the point were continuous, natural forest, using imagery from 2012 onwards. At this scale, human encroachments, including tree plantations, were obvious. If present, we selected other points. The final dataset contained 195 natural forest points.

We considered whether we classified these points correctly at the 60% tree cover threshold. The initial results showed that the classification often failed for points where the rainfall was <1800 mm annually, where forests naturally tend to be more open. For these areas, we used a 30% threshold, which classified the points correctly. Details of this process and results are in the Appendix E.

### 2.2.3.2 Errors of commission

Land cover products and forest layers are often unable to differentiate natural forests from tree plantations (Dong et al. 2014). Two major tree crops in Southeast Asia are rubber and oil palm (Wilcove et al. 2013), followed by relatively small tracts of other types, such as teak, acacia and fruit trees. Although rubber plantations have extended to the north far beyond their optimal range (Ahrends et al. 2015), rubber usually occurs in
the region where there is >2000mm annual precipitation (Rao and Vijayakumar 1992, Lemmens et al. 1995) and most consider 1500mm annual rainfall to be the lower limit for commercial production. We randomly selected the locations of 200 oil palm mills from the Global Forest Watch (FoodReg and WRI 2016) dataset and manually checked each using high-resolution Google Earth imagery to find nearby oil palm plantations. For rubber, we chose three regions, Yunnan from China, Central Highlands from Vietnam, and Ubon Ratchathani from Thailand. We identified rubber plantations, other than those Li and Fox identified in these regions, using high resolution Google Earth imagery and randomly chose 200 points with a minimum distance of 1km between each. We used these as verification points to estimate the error of commission for our forest layer.

2.2.4 Comparison with other global forest products

To assess the effect of using alternative forest maps, we looked at the Land Cover 2010 map that built upon ESA-CCI from the European Space Agency (ESA)-Climate Change Initiative (CCI). We extracted all the forest-related categories (50, 60, 61, 62, 70, 160, and 170) to make a forest layer for this product (ESA-CCI forest). ESA-CCI is a more conservative estimate of forests when compared to other global land cover products such as MCD12Q1 and the PALSAR-based forest map (Dong et al. 2014).

2.2.5 Refining ranges and patterns of biodiversity

We refined the species range maps by clipping them to the species' elevational limits and remaining natural forests. We used elevational data from the 90m-resolution
NASA Shuttle Radar Topographic Mission (SRTM) downloaded from http://earthexplorer.usgs.gov/ and the forest layers that we updated (see above). We calculated range size after each refining step. We then summed all species ranges for each step to understand the patterns of biodiversity and how these patterns change after range refinement (Figure 5).

In addition to threatened species listed by IUCN, we also consider non-threatened species that have limited ranges (< 20,000 km²) after refinement using elevation and habitat to be species of concern. To assess how well these species are protected, we downloaded protected area boundaries from the World Database of Protected Areas (IUCN and UNEP-WCMC 2015), excluding those listed as only “proposed.” For each species, we calculated how much of its range intersects with protected areas (Figure 5).

2.2.6 Difference in conservation priorities after incorporating more information

To identify the differences in conservation priorities between using the original ranges and those incorporating species’ elevational preferences and remaining forests, we selected the most biodiverse areas until our sample contained 10% of the forest area. The selection first added the pixels with the highest number of target species, then those with the next highest until the process reached the 10% limit. We did this for both the original ranges and those after refinement by elevation and by forest. We refer to the selected areas as “endemism centres” (Figure 5).
2.3 Results

2.3.1 Elevation, forest, and protected areas

Figure 4 shows the region, its political boundaries, and the data layers used in our analyses. There are two maps for forest cover with a third comparing the two. Appendix E detail how we calibrated the forest layer and the accuracy matrix. The major differences between our forest map and the ESA-CCI forest map are in the dry areas with annual precipitation lower than 1800mm (Figure A in S1 File). These areas in eastern Yunnan and northern Vietnam and Laos are the same areas where Sexton et al. (Sexton et al. 2015) find considerable disagreement in various global maps of forest cover.

2.3.2 Refining ranges

Figure 6 shows the fraction of original ranges after accounting for elevational range and remaining natural forests. The average refined ranges are 13%, 36%, and 39% of the original ranges for amphibians, mammals, and birds respectively. The majority of the species still had more than 80% of their range within their elevational preference (Figure 6). The major reduction in range size happened during the refinement by forest cover. Appendix E tabulates the changes for each species.
Figure 6: Percentage of species with different levels of remaining ranges after each step of refining.

Figure 7 shows the distributions of original range sizes, then ranges refined by elevation, then by remaining natural forest, and by species’ category of threat. As one expects, species with smaller ranges are more likely to be classified into a higher level of threat than those with large ranges. Some species with relatively large ranges have high
levels of threat and these include large-bodied birds and mammals, which often exist at naturally low population densities and are heavily hunted. There are large numbers of data deficient species, especially of amphibians, which have small geographical ranges. While it would be wrong to assume that all data deficient species are classified as such because they have small ranges — and thus are likely to be rare within the small ranges (Pimm et al. 2014) — clearly that is a strong possibility.
Figure 7: Number of species in each IUCN category, with their original ranges, then after refinement by elevation and then remaining forest. The scale is square root transformed. The dotted lines divide the ranges into three groups, 0-5,000 km$^2$, 5,000-20,000 km$^2$, and above. When a species range falls under 20,000 km$^2$, we treat it as a species of concern.

Range sizes inevitably shrink when refined by elevation and by remaining forest. The IUCN criteria includes range size, but does not refined range size directly using high precision variables. They may well do so indirectly, via the expert opinions of those who assess the species. Nonetheless, many species move into range size classes where the majority of the comparable species have a higher level of threat than IUCN attributes to them. For mammals, there are respectively two (*Biswamoyopterus biswasi*, *Rhinopithecus strykeri*), three (*Hadromys humei*, *Hapalomys longicaudatus*, *Laonastes aenigmamus*), and one species (*Arielulus aureocolaris*) of critically endangered, endangered, and least concerned species with ranges <5,000 km$^2$. When refined by elevation and remaining forest, the numbers increase to two additional critically endangered (*Trachypithecus delacouri*, *Rhinopithecus avunculus*), two more endangered (*Hipposideros halophyllus*, *Trachypithecus shortridgei*), five more vulnerable (*Craseonycteris thonglongyai*, *Trachypithecus laotum*, *Niviventer cameroni*, *Hipposideros khaokhouayensis*, *Arielulus societatis*), and three more least concern (*Dremomys gularis*, *Anourosorex assamensis*, *Leopoldamys milleti*).

Overall, after accounting for elevation and remaining forest, 61 mammals, 42 birds, and 182 amphibians have remaining ranges <20,000 km$^2$. Most of these species have their remaining range <5,000 km$^2$ (Figure 6; Table 2). We created a final list for
species of concern that included the threatened species identified by IUCN (37 mammal, 21 bird, and 37 amphibians) and additionally the non-threatened species (as well as data deficient species) with a refined range <20,000 km² (42 mammal, 28 bird, and 147 amphibians). More species currently deemed non-threatened (46 species plus 7 data deficient species), than threatened species (9 species), have their ranges shifted from above to under 20,000 km².

Table 2: Number of species in each taxon in different categories and coverage by protected areas.

<table>
<thead>
<tr>
<th></th>
<th>Mammal</th>
<th>Bird</th>
<th>Amphibian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endemic Species</td>
<td>122</td>
<td>183</td>
<td>213</td>
</tr>
<tr>
<td>Threatened</td>
<td>37</td>
<td>21</td>
<td>37</td>
</tr>
<tr>
<td>Data Deficient</td>
<td>39</td>
<td>0</td>
<td>111</td>
</tr>
<tr>
<td>Refined Range &lt; 20,000 km²</td>
<td>61</td>
<td>42</td>
<td>182</td>
</tr>
<tr>
<td>Refined Range &lt; 5,000 km²</td>
<td>44</td>
<td>24</td>
<td>144</td>
</tr>
<tr>
<td>Species of Concern</td>
<td>79</td>
<td>49</td>
<td>184</td>
</tr>
<tr>
<td>Average % Protection</td>
<td>24 ± 22</td>
<td>20 ± 14</td>
<td>32 ± 24</td>
</tr>
<tr>
<td>No Protection</td>
<td>14</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td>Data Deficient with no Protection</td>
<td>10</td>
<td>0</td>
<td>34</td>
</tr>
</tbody>
</table>

Our species of concern include 79 mammals, 49 birds, and 184 amphibians. More than 50% of them IUCN currently designates as non-threatened. Among all the non-threatened species identified by IUCN, 84% of amphibians, 49% of mammals, and 17% of birds are considered species of concern. This increases the current IUCN estimates significantly. Especially for data deficient species, 93% of them have their ranges <20,000 km², and 89% are <5,000 km².
2.3.3 Protection for different taxa

Figure 4C shows the distribution of protected areas in Southeast Asia, and Table 3 shows how well different groups are protected. (Appendix F tabulate the details for each species.) Some 22% of mammals, 14% of birds, and 35% of amphibians have less than 10% of their refined ranges protected. Only two bird species (Pitta gurneyi, Spelaeornis longicaudatus) have no part of their ranges within protected areas, while 14 mammals and 39 amphibians do. Only six mammals, four birds, and six amphibians have more than 80% of their ranges protected.

Table 3: Numbers of species that have < 10% or > 80% of their refined ranges protected.

| PA Coverage | Species of Concern | | | Other | |
|-------------|--------------------|------------------|------------------|------------------|------------------|------------------|
|              | Threatened < 20,000 km² | Non-Threatened < 20,000 km² | Data Deficient < 20,000 km² | Total | Non-Threatened > 20,000 km² | Data Deficient > 20,000 km² | Total |
| Mammals     | 6 | 2 | 15 | 23 | 3 | 1 | 4 |
| < 10%       | 1 | 1 | 4 | 6 | 0 | 0 | 0 |
| > 80%       | 37 | 12 | 30 | 79 | 34 | 9 | 43 |
| Total       | 37 | 38 | 109 | 184 | 27 | 2 | 29 |
| Birds       | 7 | 7 | 14 | 12 | 12 | 12 | 12 |
| < 10%       | 1 | 3 | 4 | 0 | 0 | 0 | 0 |
| > 80%       | 21 | 28 | 49 | 134 | 134 | 134 | 134 |
| Total       | 37 | 38 | 109 | 184 | 27 | 2 | 29 |
| Amphibians  | 14 | 11 | 47 | 72 | 2 | 0 | 2 |
| < 10%       | 3 | 3 | 29 | 35 | 0 | 0 | 0 |
| > 80%       | 37 | 38 | 109 | 184 | 27 | 2 | 29 |
Table 3 also shows how protected areas cover different subsets of each taxon. There is no significant difference in protection rates between threatened and non-threatened species no matter the size of the remaining ranges (t-test, p>0.05). For species that have refined ranges < 20,000 km$^2$, two features are of note. First, consider the fractions of species with protected range <10% between what IUCN deems non-threatened with those threatened: 2 of 12 versus 6 of 37 mammals, 7 of 28 versus 7 of 21 birds, and 11 of 38 versus 14 of 38 amphibians. In each case, the fractions are very similar — a pattern confirmed when one calculates the average fractions of ranges protected (t-test, p>0.05). Simply, putatively non-threatened species do not have more of their ranges protected that do comparable species that are threatened. The existing protected area network fails to emphasize the coverage of either threatened species listed by IUCN or the non-threatened species with higher threat because of their limited remaining ranges.

Second, consider the fractions of data deficient species. For both amphibians and mammals, about half the data deficient species have <10% of their estimated ranges protected. There are no data deficient birds.

2.3.4 Patterns of biodiversity and conservation priority

The mountainous areas of Southeast Asia have high numbers of endemic mammals, birds, and amphibians (Figure 8). The Hoang Lien Son Range in northern Vietnam and the Annamite Range, the principal mountain range of Indochina, and the boundary between Laos and Vietnam are the richest areas for all three taxa. Beyond
these mountain ranges, endemic mammals also concentrate in the northern regions of Laos, Vietnam, and Thailand. They extend south to Malay Peninsula through Bilauktaung Range and have another concentration in the Da Lat Plateau. Birds are similar to mammals, but extend further south and north. They have a concentration in the mountain ranges of the Arakan Yoma, the Chin Hills, and Mizo and Naga Hills along the border with Myanmar and India. This is the Mizoram-Manipur-Kachin rainforest ecoregion (Olson et al. 2001).

The pattern for amphibians is substantially different from birds and mammals. The highest richness areas are relatively concentrated and isolated from each other. The south and central Annamite Range between Ban Vangchang in Laos and Thanh Lang Xa in Vietnam have the most endemic amphibians, followed by the Central Malay Peninsula. Many regions have not been well-surveyed for amphibians. High endemism may also exist in less well explored and surveyed areas.

2.3.5 Species of concern

The species of concern concentrate in different places (Figure 8 G-I) than for the endemic species overall (Figure 8 D-F). Species of concern for the three taxa all have their highest concentration along the Annamite Range. Mammals have the highest number of species of concern in the north of this range while birds and amphibians have the highest richness in the southern tip of this mountain range and Da Lat Plateau.
Figure 8: Patterns of biodiversity in Southeast Asia. (A-C) are using the original range maps from IUCN of BirdLife International for all the endemic species. (D-F) are using the range maps after refining by elevational range and forest for each species. (G-I) are using the refined range maps for the species of concern (threatened species and non-threatened species with range < 20,000 km²). (J-L) are using the refined range maps for data deficient species.
2.3.5.1 Data deficient Species

About 32% of mammals and 52% of amphibians (Table 2) are data deficient. Such mammals mainly concentrate near the border between China, Vietnam and Laos, northern Annamite Range, Malay Peninsula, and southern Myanmar (Figure 8 J). Data deficient amphibians mainly are in the Malay Peninsula, the southern tip of Thailand, and southern Myanmar (Figure 8 L). It is possible that many undescribed species exist in un-surveyed parts of the region.

2.3.5.2 Difference in conservation priorities after incorporating more information

We calculated endemism centres from the original ranges (Figure 8 A-C) and after refinement by elevation and forest (Figure 8 D-F). Figure 9 shows the areas in agreement or otherwise for the three taxa. Green and red areas are the endemic identified using updated range maps. All three taxa have part of the endemic centre in the Annamite Range of Laos and Cambodia. Mammals extend more to north Thailand while birds also have another high concentration on the border between India and Myanmar. The Dawna range of eastern Myanmar and adjacent Thailand and the Malaysia peninsular are only identified as endemism centres for amphibians.
Figure 9: Comparisons of endemism centers using different range maps. The two endemism centers for each taxon were identified using the original ranges and updated ranges respectively. Black areas are agreed upon by both methods. Red areas are the endemism center identified only by overlaying the updated ranges and blue areas are for original ranges only.

There are broad similarities between the two approaches and among the three maps. Simply using the original IUCN ranges overlooks many regions of importance, which are shown in red. For mammals, this mainly concentrates in north Thailand, Laos and Vietnam, as well as southern Thailand. For birds, the overlooked areas are in Dawna range in western Thailand and the southern tip of Annamite Range. For amphibians, they are in northern and southern tips of Annamite Range, as well as the Dawna range in Thailand. The blue areas were considered unsuitable for species because they were either outside the elevational range or devoid of natural forests. The total area of the centres of endemism that is protected is 11326 km². Protected areas
cover 2982 km\(^2\), 4606 km\(^2\), 6036 km\(^2\) of the priority areas of mammals, birds, and amphibians respectively.

2.4 Discussion

The expert opinion system that the IUCN Red List employs factors in specific knowledge of diverse threats, which could involve whether the species is hunted, how specific are its habitat requirements, how much habitat remains within its broad range, and whether habitat is being lost or numbers are decreasing for other reasons. IUCN uses an original range of 20,000 km\(^2\) as an important benchmark for endangerment. The majority of threatened species have ranges smaller than this and are in places where there is continuing loss of habitat.

Small range size alone is not sufficient for a “threatened” listing. Nor is it necessary. For example, IUCN classifies some birds and mammals as critically endangered when their original ranges are >20,000 km\(^2\) and some as vulnerable when their original ranges are >100,000 km\(^2\). These species are often large-bodied ones that are hunted or persecuted in other ways. For example: the vulture, \textit{Gyps bengalensis}, is critically endangered, had a range of > 4,000,000 km\(^2\) — and was common within it until recently, when poisoning massively depleted its numbers (Cuthbert et al. 2016). Remote sensing cannot assess such threats.

Nevertheless, many species are endangered due to habitat loss and degradation. This is simpler to quantify than hunting pressure, though that can be a significant threat
to many taxa even within intact forest regions. Thus, the lack of explicit information on elevational range, habitat remaining within the range of a species, how fragmented is that range, and how much of that range is protected, is a significant limitation that remote sensing can help overcome.

Our threshold for designating species of concern is also 20,000 km². We understand that the criteria for endangerment that IUCN employs are based on the original ranges, not our refined ones, which will inevitably be smaller. Our use of the same extent is not to confuse it, but merely to identify species that IUCN does not deem threatened and yet which have small enough geographical ranges that they might qualify, or be classified at a higher risk than at present.

There are substantial numbers of species (Figure 7) that, once their range is refined, warrant a re-examination of their IUCN listing. Indeed, there are four mammals, nine birds, and seven amphibians that IUCN deems to be species of least concern yet have refined ranges <5,000 km² (mammals: *Dremomys gularis, Anourosorex assamensis, Leopoldamys milleti, Arielulus aureocollaris*; birds: *Alcippe danisi, Alcippe klossi, Arborophila campbelli, Arborophila cambodiana, Carduelis monguilloti, Garrulax annamensis, Garrulax peninsulae, Psilopogon chersonesus, Spelaeornis oatesi*; amphibians: *Amolops archotaphus, Amolops mengyangensis, Hylarana milleti, Kurixalus bisacculus, Leptolalax melanoleucus, Microhyla marmorata, Theloderma andersoni*).
Our results are double-edged. First, the actual ranges are small, indeed small enough that IUCN classifies most other species with comparable, refined range sizes as threatened. Second, as Figure 6 shows, only a small part (13-39% on average) of their original range still has potential habitat. Likely, human actions have destroyed the remainder. Habitat is likely to be continuing to be lost — something that remote sensing can best confirm.

Some may assert that decisions on the risks experienced by individual species may be correct even if the assessors did not use geographical data explicitly. Even were this assertion correct, we would argue that it is better to use explicit standardised quantitative criteria whenever possible rather than rely on indirect judgements, which may be less precise and more vulnerable to bias. Transparency in the process is important. In any case, we doubt the assertion for three reasons.

First, refining ranges by elevation limits and remaining habitat cover greatly alters the estimates of remaining range. After accounting for suitable habitats, most species (60% of mammals, 51% of birds, and 60% of amphibians) have their refined ranges <40% of those originally published (Figure 6). Most of the range reduction happened during the refinement by remaining forests rather than by elevational ranges (Figure 6). Thus, the difference between the range as IUCN reports it and the remaining range mainly relies on the knowledge of where forest remains. We recognise, of course,
that even these estimates of remaining habitat may be too optimistic: a species may have particular habitats that further restrict its range.

Second, our results show that refining ranges by elevation and habitat differentially alters how much range remains. Some species have proportionately considerably smaller ranges than do others following the same process. Consequently, they likely have a considerably greater risk of extinction than hitherto appreciated.

Third, one might answer that IUCN is making consistent decisions by arguing that those species that it does not consider threatened, but that have small refined ranges, might be disproportionately protected by the network of protected areas. Table 3 rejects that possibility.

Finally, data deficient is not a category to ignore, but one that requires extra attention in conservation planning. Many such species have small known ranges and are overlooked by the protected areas. These may be at greater risk than the officially threatened species designated by IUCN, especially so since their data deficient status already suggests that these species are likely to be exceedingly rare.

In sum, the most parsimonious interpretation of our results is that more species are at risk than initially thought. These species of concern need immediate assessment and involve more than half of the mammal and bird species, and about 74% amphibian species that are currently not considered threatened (Table 2).
There are caveats. The critical step in refining species’ ranges comes from estimating how much natural forest cover remains. Across large areas, this poses no challenge — there is no tree cover of any kind. There are classes of problems where determining forest cover is difficult — tree plantations and sparse forest.

The classification of tree plantations as forests is still a problem for large regions, especially at fine geographical scales. To be conservative, we used a relatively high tree cover threshold both for dry and wet regions, and we additionally excluded tree plantations based on other studies. Our map, however, may still miss some of these areas and some recent expansions of plantations. At the same time, we may miss natural forests that have low tree cover, such as some woodlands and savannahs, especially in dry areas. The conservation community would benefit from further improvements to future forest maps, including the differentiating of natural forest and plantations.

2.4.1 Knowledge from remote sensing modifies conservation priorities

Our results show a substantial difference between the locations of centres of endemism before and after using the remote sensing data (Figure 8). About 20% of the area identified using original mammal ranges, 25% for birds, and 40% for amphibians does not appear to have natural forests (Figure 9). Remote sensing data can clearly reveal human-modified landscapes. Given the limited conservation resources, incorporating such data for a more accurate delineation of priorities can produce more practical and efficient conservation plans, especially at local scales.
Compared to the lowland forest in Southeast Asia, which is among the most diverse biomes in the world, montane forests in Southeast Asia receive considerably less attention (Koh et al. 2013). Our study shows that the mountainous areas harbour the highest biodiversity that is restricted to this region (Figure 8). These montane forests are experiencing rapid deforestation (Sodhi and Brook 2006, Koh et al. 2013). Plantations of rubber are expanding to higher altitudes and drier and colder areas (Ahrends et al. 2015) due to China’s success in researching the growth of rubber in non-traditional environments and to increasing global demand. These expansions are a great concern for the endemic species whose survival depends on the integrity of natural forests. Because many countries in the study area use mountain ranges to delineate the country borders, it will require collaboration between countries to reduce threats such as poaching, illegal logging, and transformation of natural forests to plantations. Thus, trans-boundary protected areas are essential where the richest biodiversity straddles country borders.

Habitat loss remains the most serious threat to species in Southeast Asia, to which poaching, invasive species, and climate change add further risks (Wilcove et al. 2013). Rapid land cover changes require quick and repeatable assessments to respond to these threats. Using publicly available datasets for our analysis, our results show two crucial aspects to consider in future priority-setting, 1) species at risk according to the size of their remaining habitats and coverage by protected areas, and 2) areas with high
irreplaceability and vulnerability. This quick and simple method can be widely used and replicated. It is especially useful for conservation practitioners with little access to academic resources or budgets to purchase data. With the increase in remote sensing information freely available to the public and through sharing platforms (Pimm et al. 2015, Sexton et al. 2015) like Global Forest Watch, it empowers practitioners to track the changes frequently and independently. Our methods are not a panacea to solve conservation problems but do provide an efficient way to track the changes in a species’ status that accounts for its habitat needs, land cover changes, and protected area settings.

Although basal species maps produced by the IUCN may have issues, and are produced by “expert assessment” rather than explicit empirical data, our approach improves the accuracy by incorporating relevant environmental data to help refine suitable areas. Whilst this approach is dependent upon the reliability of basal species range maps — which we do not assess here — our approach represents a considerable advance in increasing the accuracy and better representing the distributions of these species. Furthermore, through the development of an ArcToolbox it is easy to standardise the approach and allow replication in other regions, or even the possibility to become integrated into the existing range mapping process.

2.4.2 A modest, practical proposal

We argue that individual species assessments may be inconsistent and that the conservation priorities based on them are in less than optimal places. Whatever the pros
and cons of the arguments we have presented, it would behove those who evaluate species to present estimates of (1) how much of a species range is within its known elevational range, (2) for forest species, how much of that range still has forest cover, and (3) how much of the range is within protected areas. Such data would permit quantitative insights into a species’ risk of extinction, insights that are currently absent.

2.5 Supporting information

Appendix E Mapping Forest and Evaluation. Table E1) Confusion matrix and error estimate. Figure E1. Validation of forest cover. It also provides a link to download species information tables, which list the IUCN status, elevational range, original range, range after refined by elevational range, and remaining forest, coverage from the protected areas(PA), and whether it is considered as a species of concern in our paper.

2.6 Citation

This chapter was published in 2016 in the journal PLOS One. The complete citation is:

Li et al., 2016. Remotely sensed data informs Red List evaluations and conservation priorities in Southeast Asia. PLOS One, 11(8), e0160566.
Chapter 3. Emerging threat from livestock grazing on giant panda habitats

3.1 Introduction

Habitat loss and fragmentation from deforestation, agriculture expansion, road construction, and other disturbances have divided the wild population of giant pandas into 33 isolated populations in 6 mountain ranges in China (State Forestry Administration 2015). Most have fewer than ten individuals (State Forestry Administration 2015). Climate change is predicted to further threaten their survival by causing a shortage of their major food bamboos (Tuanmu et al. 2013), and up to 2/3 of their habitat loss (Songer et al. 2012, Fan et al. 2014). With the Natural Forest Conservation Program and Grain to Green programs, the deforestation that was once the biggest threat to pandas has been halted (Liu et al. 2008). Moreover, China has devoted unparalleled resources to conserve this endemic and endangered species including establishing 67 nature reserves to protect its habitats. More than 54% of panda range is under protection, which also protects the centre of forest endemism that overlaps substantially with panda range (Li and Pimm 2016).

A previously unrecognized threat is now emerging — livestock grazing. Livestock have become the most prevalent human disturbance throughout panda distribution (Hull et al. 2014, State Forestry Administration 2015). More than one third of the plots during the fourth national survey of giant panda during 2011-2014 showed
evidence of livestock grazing (State Forestry Administration 2015). Here, we ask: How does livestock grazing influence the survival of giant pandas?

Previous studies show that there is overlap as well as differences in the spatial distribution of pandas and livestock (Ran et al. 2002b, Ran et al. 2002a, Kang et al. 2011, Wang et al. 2015, Zhou et al. 2016). Moreover, livestock graze on bamboos (Hull et al. 2014). These studies fail to answer some important questions related to whether livestock grazing is a significant threat to pandas. The differences in distribution or home range between species could be a result of the interspecies interaction, but also from different habitat preferences. A way to differentiate these two possible causes relies on the study of long-term changes.

Wanglang National Nature Reserve was established in 1965 to protect giant pandas. There is no logging, agriculture or human residences inside the protected area. Since early 2000, however, local people have put livestock inside the reserve and let them free-range inside the forests. The livestock population has increased 8-9 fold since 2004. This case provides an excellent chance to study the influence of livestock on the habitat use of giant pandas.

Here, we ask several fundamental questions related to this issue. First, what are the immediate and long-term impacts of livestock grazing on bamboos? Second, has the increasing number of livestock changed the distribution of giant pandas? Third, how does the incorporation of livestock influence the accuracy of predicting of giant panda
habitat? Finally, how does the total area of giant panda habitat change after the introduction of livestock and where are the most dramatic changes?

This study represents the first effort to quantify the impacts of livestock on endangered giant pandas. By utilizing long-term monitoring data, species distribution modeling, and GPS collar tracking data, we can answer the above questions and provide suggestions for giant panda management.

3.2 Methods

3.2.1 Study area

Wanglang National Nature Reserve is in the northern Minshan, Sichuan, China (Figure 10), part of the South-Central China biodiversity hotspot (Myers et al. 2000). It lies in Pingwu County — the county that has the most pandas (State Forestry Administration 2015). It covers 323 km² with elevations ranging from 2300m-4980m.
Figure 10: Study area

3.2.2 Data collection

3.2.2.1 Survey plot

We collected data from June-July, 2013 and Jan-April in 2014 with systematic sampling, placing plots (20m×20m) at least 300m, which is the home range radius of pandas (Hu et al., 1985) apart along a transect to avoid spatial correlation. Data come from 124 plots in summer and 188 plots in winter. In each plot, we noted any signs that indicated the presence of a panda, cattle or horse including feces, feeding signs, scent marks, hairs, and footprints. We also took the vegetation measurements including
understory coverage and density, bamboo coverage and height, and the extent of livestock grazing on bamboos. The feeding signs from livestock can be easily distinguished from other wildlife as they are less selective of individual culms (Hull et al. 2014). Two measurements assessed the impacts of livestock grazing: the percentage of bamboo cover that was grazed in a plot and the percentage of bamboo leaves on a culm grazed by livestock. Other factors that may influence the use of the habitat for pandas and livestock (Liu et al. 1999) including tree cover, slope, elevation, aspect, distance to river, distance to paved road and trails we calculated from a digitized map of the nature reserve using ArcMap. We used a Digital Elevation Model (DEM) with a resolution of 30m to derive elevation, slope, and aspect for the neighboring area of a plot. We obtained a continuous tree cover map with a 30m resolution for this region (Sexton et al. 2013).

3.2.2.2 Bamboo data

We selected three locations in the nature reserve to compare the impacts of livestock on the regeneration of bamboos. In each site, there was a natural barrier, such as a river, so that livestock were only able to use part of the area and left the rest untouched. This setting provided an ideal experiment. Thus, we got three heavily grazed areas and three controls. We took field measurement from Aug 15-17th, 2015. Mid-August is the end of rapid growth period of the first year bamboo shoots (Wang et al. 1987). In each location, we measured 100 clumps of bamboos. For each clump, we
placed 1*1m plot centered on the centroid of the clump. Plots were at least 20m apart. For each clump, we counted the number of culms and shoots and calculated shooting rate as the ratio between the number of shoots and number of culms for each plot. Then, we measured the height and base diameter for five culms and five shoots.

3.2.2.3 Monitoring data

The reserve staff did regular patrols along 16 fixed routes and eight random routes from 1997 until 2014. They kept a record of all the sightings of wildlife or their signs along the routes, with a GPS location, time stamp and basic environmental measurements for each observation.

3.2.2.4 Collar data

We purchased 21 GPS collars (Anit-GSMcollor810) from Tianjin Blueoceanix Technology Co., LTD and placed on 21 herds (11 for cattle and 10 for horses). Because the dominant animal is usually in the center of a herd and tends to lead when a herd is moving (Sato 1982), we put the collars on the dominant animal for each herd on the advice of its owner. Each GPS collar recorded the location every 2 hours. Before deployment, we placed all the collars in open area and then under forest with an average canopy cover of 73% in the valleys to test for accuracy and stability. The precision was 10m±9m in open areas and 21m± 17m under forests. Due to battery failure and other technical problems, we had to change the collars or the animals on which we put them. In total, we collected data from 11 horses and 19 cattle.
3.2.3 Data analysis

3.2.3.1 Impacts of livestock grazing on panda habitat use over time

The number of livestock increased dramatically after 2005. Before that, fewer than 200 livestock were in the nature reserve. Thus, we divided the data into three periods, 1997-2004 as “before”, 2005-2008 as “transition” and 2009-2014 as “after”. These three periods correspond to increasing numbers of livestock.

We calculated the number of panda observations falling into the predicted livestock habitats for “before” and “after” periods. We divided the livestock habitat by type (cattle or horse) and by intensity. The overlap between summer habitats and winter habitats we considered to be areas of high grazing intensity.

For the monitoring data, survey efforts were not consistent through years. To avoid the bias in survey efforts, we used the number of observations for other species as a measure of survey intensity. We assumed that the number of other species directly related to the survey intensity (Phillips et al. 2009).

To avoid spatial auto correlation, we excluded observations of the same species within 100m from the same patrol, which is the average daily movement distance for all the monitoring species.

3.2.3.2 Species distribution model

We measured GIS-derived data such as tree cover at different scales – circles with 30m, 60m, 90m, 300m, or 600m diameters. The same environmental factor at
different scales may have different influences on habitat selection (Mashintonio et al. 2014). However, the highly correlated nature between these and other variables can cause the inflation of the variance of regression parameters and lead to the wrong identification of relevant predictors (Dormann et al. 2012). We used Random Forest to select the environmental factors in the most relevant geographic scale while controlling for correlation between predictors (Strobl et al. 2008). Mean decrease of accuracy is a measure generated by the Random Forest to evaluate the relative importance of variables. We select the factor with the scale that had the highest mean decrease of accuracy value (Cutler et al. 2007). We further excluded the variables with a correlation coefficient $|r| > 0.7$ (Dormann et al. 2012). Table 4 shows the final list of variables.

**Table 4: Final variables used in the model and description**

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>Elevation derived from 30m SRTM DEM (<em>Data available from the U.S. Geological Survey</em>)</td>
</tr>
<tr>
<td>Slope</td>
<td>Slope derived from 30m DEM</td>
</tr>
<tr>
<td>dis_road</td>
<td>Distance to trail</td>
</tr>
<tr>
<td>dis_pave</td>
<td>Distance to paved road</td>
</tr>
<tr>
<td>dis_river</td>
<td>Distance to river</td>
</tr>
<tr>
<td>tc_10</td>
<td>Average tree cover in 600m diameter</td>
</tr>
<tr>
<td>landcover</td>
<td>Landcover type: grassland, shrubs or forest</td>
</tr>
<tr>
<td>TPI</td>
<td>Topographic position index, calculated from 30m DEM (Weiss 2001). Range from -2 to 2, -2 stands for valley bottom, -1 lower slope, 0 for flat area, 1 for upper slope, 2 for ridge.</td>
</tr>
</tbody>
</table>
We applied GLM, which is easier to interpret the direction and effects of different environmental factors (Merow et al. 2014). We first built a GLM model for cattle and horse in each season using the software R 3.3.1. Based on our understanding of the species-environment correlation, we added the quadratic terms for elevation, slope, tree cover, and distance to river, paved road and trail to the initial full model. Then, we used stepwise selection to find the most parsimonious model with the minimum AIC. Finally, the Area under the ROC curve (AUC) and 10-fold cross-validation were used to evaluate accuracy of the models. AUC of 0.5 represents random classifier to differentiate habitat vs. non-habitat and AUC of 1 represents a perfect one. The 10-fold cross-validation randomly partitions the dataset into 10 subsamples. It repeats through the 10 subsamples with 9 subsamples for training and one for testing each time.

For giant pandas, the pretest showed that season was among the least important variables in predicting the habitat selection. Thus, we pooled data from the two seasons. We first built the model with only abiotic factors using GLM. Then we added the presence, absence, and abundance information on cattle and horses into the model. We used AIC to evaluate how different additions of biotic interactions impacted the performance of panda habitat predictions. We chose the model with the minimum AIC as the final model and calculated AUC, accuracy from 10-fold cross-validation, pseudo $R^2$ to compare with the model with only abiotic factors. These measurements could
explain whether incorporating livestock distribution could improve the accuracy and explanation power of where pandas might be. We chose the cutoff values which maximized the sum of sensitivity and specificity (Sing et al., 2005).

To compare with the previous distribution of giant pandas, we used the monitoring data from 1997-2004, the period when the livestock were few in Wanglang. To be comparable, we first thinned the data using 160m threshold, which was the average distance between plots from the field survey, resulting in 312 presence points. Pandas are mostly below 3200m, which is the upper limit for bamboo growth. The background range was set to be within the elevation range from 2300-3200m and 160m away from all the presences points. We randomly selected 724 points from this range. Then, we built the GLM model for this period. AUC and 10-fold cross validation were calculated. We compared the distribution of pandas before and after the introduction of large numbers of livestock to identify the lost habitats and new expansions.

We used ArcGIS 10.2 to map the distribution of livestock and panda. We applied t-test to examine the effects of livestock grazing on bamboos.

4.2.3.3 GPS collars

To ensure the accuracy of the data points for use, we screened the raw data and excluded the duplicates and abnormal data points. 2D location data usually have less spatial precision as they are identified with fewer satellites than 3D location data (Frair et al. 2010), so we further excluded these data points. We calculated the percentage of
GPS tracking data in each type of habitats, which included the lost and expanded habitats of pandas.

### 3.3 Results

#### 3.3.1 Grazing impacts on bamboos

Horses grazed bamboos more intensively than cattle in our sign survey. Cattle usually grazed 0-24% of the plot areas and 0-24% of the bamboo leaves. Horses grazed more of a plot area and the majority of the bamboos had more than 25-50% of their leaves grazed. Bamboos were worse off when both horses and cattle were present an area (Figure 11).

![Figure 11: Impacts on Bamboos. The figure on the left shows the percentage of plot being grazed by livestock and the one on the right shows the percentage of leaves on a stem being grazed. Data were classified into four categories for both measurements, 0-24%, 25-49%, 50-74% and 75-100%. The bar](image)
in each paragraph represents the percentage of plots falling into each of the four categories. For example, about 20% of the plots with both cattle and horses had less than 24% of their area grazed.

The bamboo shooting rate of the plots where there was livestock grazing was 4.2%, significantly lower than the ones without grazing, which was 18.9% (t-test, p<0.0001). Some 42% of our plots with livestock grazing had no shoots at all while it was only 3% without livestock impacts. Meanwhile, the livestock grazing significantly reduced the average diameter of shoots (t-test, p<0.0001) and the height (t-test, p<0.0001) of the shoots (Figure 12). Normally, the diameter and height of the shoots are similar to adult culms after for the first growing season (Figure 12) (Wei et al. 2013). In addition, livestock grazing reduces the height of mature culms (t-test, p<0.0001) and number of culms per clump (t-test, p=0.0338) (Figure 11). We found more dead culms in the grazed areas. In sum, livestock grazing on bamboos resulted in reduced number of leaves, lower height of mature culms, weaker shoots, and thus harmed bamboo growth and regeneration.

### 3.3.2 Livestock habitat use

Cattle and horses have distinctive habitat use from each other and as well in summer versus winter. Thus, we built two different models for each season for each species (Table 5). Generally, cattle and horse preferred to stay closer to the paved road and lower elevations — especially in winter (Figure 14). In summer, they usually moved to higher elevations and upper slopes. Horses prefer areas with higher tree cover and
use forests similar to other vegetation types, while cattle have a clear preference for grasslands and shrublands (Table 5). Slope and elevation were more of the limiting factors for horses.
Figure 12: Height and diameter of bamboos with and without grazing. The lines on the box correspond to the 3rd quartile and 1st quartile and white bars are the mean. Whiskers drawn to the furthest point within 1.5 x IQR (3rd quartile minus 1st quartile) from the box.
Figure 13: Impacts of livestock grazing on the number of shoots per clump. The lines on the box correspond to the 3rd quartile and 1st quartile and white bars are the mean. Whiskers drawn to the furthest point within 1.5 x IQR (3rd quartile minus 1st quartile) from the box.

Cattle summer model:

$$\text{Cattle} \sim -7.8 - 0.75 \cdot (\text{tpi}_1) - 0.37 \cdot (\text{tpi}_0) + 0.14 \cdot (\text{tpi}_1) + 2.5 \cdot (\text{tpi}_2) - 0.05 \cdot \text{tree cover} - 0.22 \cdot \text{dis_river} - 0.51 \cdot \text{dis_pave}$$

Cattle winter model:

$$\text{Cattle} \sim -0.28 - 0.01 \cdot \text{elevation} + 0.07 \cdot (\text{tpi}_1) - 0.82 \cdot (\text{tpi}_0) + 2.1 \cdot (\text{tpi}_1) + 2.0 \cdot (\text{tpi}_2) - 1.12 \cdot \text{forest} - 0.2 \cdot \text{shrub} - 0.01 \cdot \text{slope} - 0.13 \cdot \text{tree cover} - 0.60 \cdot \text{dis_pave} + 0.03 \cdot \text{dis_river}$$

Horse summer model:
Horse ~ $-8.7 + 6.36E-07 \cdot \text{elevation}^2 - 1.79 \cdot (\text{tpi}_1) + 0.19 \cdot (\text{tpi}_0) + 3.79 \cdot (\text{tpi}_1) + 1.16 \cdot (\text{tpi}_2) + 0.51 \cdot \text{treecover} - 6.05E-03 \cdot \text{treecover}^2 - 0.21 \cdot \text{dis_pave}^2 + 1.23 \cdot \text{dis_pave} + 0.06 \cdot \text{dis_river} - 0.11 \cdot \text{slope}$

**Horse winter model:**

Horse ~ $-30.3 + 1.6E-03 \cdot \text{elevation}^2 + 1.08 \cdot \text{treecover} - 0.01 \cdot \text{treecover}^2 - 0.24 \cdot \text{dis_pave} - 0.33 \cdot \text{dis_trail} - 0.15 \cdot \text{slope} + 2.9E-03 \cdot \text{Slope}^2$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cattle</th>
<th>Horse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>winter</td>
<td>summer</td>
</tr>
<tr>
<td>elevation</td>
<td>-</td>
<td>*</td>
</tr>
<tr>
<td>elevation^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dis_pave</td>
<td>-</td>
<td>- **</td>
</tr>
<tr>
<td>dis_pave^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dis_road</td>
<td></td>
<td>- ***</td>
</tr>
<tr>
<td>dis_river</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>dis_river^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>slope</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Slope^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tree Cover</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tree Cover^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>as.factor(tpi)-1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>as.factor(tpi)0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>as.factor(tpi)1</td>
<td>+ *</td>
<td>+</td>
</tr>
<tr>
<td>as.factor(tpi)2</td>
<td>+ **</td>
<td>+ **</td>
</tr>
<tr>
<td>Forest compared to grassland</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Shrubs compared to grassland</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.86</td>
<td>0.78</td>
</tr>
<tr>
<td>CV-10fold accuracy</td>
<td>82.6%</td>
<td>80.9%</td>
</tr>
</tbody>
</table>
In both seasons, cattle and especially horses heavily use the areas around the paved road in the valleys areas. Cattle tended to move down to the valley only in winter. The grey area in the map shows areas lower than 3200m — the elevation constraint for bamboos. Figure 14 reveals that most of the heavily used, year-round areas by horses were also panda habitats. For cattle, mostly the winter range was also in panda habitats. Besides major valleys, cattle used more of the ridges than valleys in the smaller watersheds, when compared to horses.

**Figure 14: Cattle and horse distribution in summer and winter**
3.3.3 Changes in habitat use for pandas

Because the river valleys are the most heavily used areas by livestock within potential panda habitats, we created a 300m buffer for the major rivers and 100m buffer for secondary rivers. These valley areas within the buffer accounted for 69% of the all-year-round habitats for horse and 36% for cattle in potential panda habitat. If counting all types of habitats, more than 62% of the valley areas were used by horse and 57% were used by cattle.

Table 6 compares sightings of pandas and other species before and after large numbers of livestock were introduced. There is a striking reduction in the number of pandas seen in the valleys — 297 before compared to 111 afterwards, i.e. 63%. (In contrast, the reduction in signs of other species fell only from 880 to 782, i.e. 11%, showing that this decrease cannot be due principally to changing sampling effort. When we consider parts of the valley only used by cattle, the drop in panda observation is from 191 to 55 (71% — compared to 11% for other species) and for parts of the valley used only by horses 167 to 59 (65%) (compared to 9% for other species.)

In valleys, 25% of all species observations were pandas in the “before” period, and it reduced to 12% for “after” period. In another words, there was a 50% reduction of observations of pandas in the valleys (Table 6) after the number of livestock increased, and 29% fewer observations of panda in cattle habitats and 8% fewer in horse habitats. Pandas were observed more in the habitats that livestock did not use. However, outside
the valley, there was higher percentage of observations of pandas within livestock habitats. About 69% of increasing use of horse habitats were within summer horse habitats, which were mostly on ridges and 49% of the increasing use of cattle habitats were within the year-round habitats, which were also on ridges.

### Table 6: Comparison between two periods of panda occurrences

<table>
<thead>
<tr>
<th></th>
<th>Valley sub-total</th>
<th>Valley cattle habitat</th>
<th>Valley horse habitat</th>
<th>Outside Valley sub-total</th>
<th>Outside Valley cattle habitat</th>
<th>Outside Valley horse habitat</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>panda</td>
<td>after</td>
<td>111</td>
<td>55</td>
<td>59</td>
<td>353</td>
<td>194</td>
<td>138</td>
</tr>
<tr>
<td>panda</td>
<td>before</td>
<td>297</td>
<td>191</td>
<td>167</td>
<td>362</td>
<td>175</td>
<td>85</td>
</tr>
<tr>
<td>other</td>
<td>after</td>
<td>782</td>
<td>531</td>
<td>620</td>
<td>421</td>
<td>269</td>
<td>165</td>
</tr>
<tr>
<td>other</td>
<td>before</td>
<td>880</td>
<td>599</td>
<td>678</td>
<td>569</td>
<td>347</td>
<td>242</td>
</tr>
</tbody>
</table>

### 3.3.4 Habitat use for giant panda

Season ranked the least in the mean decrease of accuracy, which is a measurement of variable importance from random forest. In addition, adding this variable into the GLM model did not improve AIC, thus this variable was not important in our analysis in predicting pandas' presence or absence. Consequently, we pooled data from winter and summer to construct habitat model for pandas.

The best model without considering livestock constituted elevation, slope, distance to river and paved road, shrub density, bamboo coverage and tree cover. AUC for this model was 0.916 and the accuracy of the model from 10-fold cross-validation was 90.7%. Including the species interaction horse, cattle or both improved the model performance (Table 7). This indicated that livestock provided additional explanation beyond the environmental factors in the distribution of livestock. The absence and
presence of cattle and horse better predicted the absence and presence of pandas than the number of each species. The best model for pandas was composed of the environmental factors and the presence and absence of both cattle and horse. The AUC for this model was 0.938 and the accuracy from 10-fold cross validation was 91.6%. The explained variance increased from 42.8% to 50.0% after species interaction incorporated into the model.

Table 7: Panda distribution model with livestock interaction

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bamoboo_coverage + TPI + distance to paved road + slope + distance to river + elevation + shrub + tree cover + number of cattle</td>
<td>187.2</td>
</tr>
<tr>
<td>W ~ number of horse</td>
<td>180.0</td>
</tr>
<tr>
<td>~ + cattle_10</td>
<td>179.5</td>
</tr>
<tr>
<td>~ + horse_10</td>
<td>179.4</td>
</tr>
<tr>
<td>~ + number of horse + number of cattle</td>
<td>180.6</td>
</tr>
<tr>
<td>~ + number of cattle + horse_10</td>
<td>174.0</td>
</tr>
<tr>
<td>~ + cattle_10 + horse_10</td>
<td>172.5</td>
</tr>
</tbody>
</table>

3.3.5 Changes between current and previous panda habitats

Table 8 detailed the models for “before” and “after period” of giant panda habitats. Predicted habitats differed substantially. The red area in the Figure 15 was the habitat identified by the before model, but lost for current model.

Table 8: Model comparison for “Before” and “After” periods

<table>
<thead>
<tr>
<th></th>
<th>After Estimate</th>
<th>Std. Error</th>
<th>Before Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPI-1</td>
<td>-1.6E+00</td>
<td>8.6E-01</td>
<td>*-0.5762</td>
<td>2.4E-01</td>
</tr>
</tbody>
</table>
The reduced area of giant panda habitat was about 33.0 km$^2$, which was mostly in low elevation and valley areas. The total area of the new expansion was about 7.5 km$^2$, which resulted in net loss of 34% of its original habitats. The lost habitats located closer to paved road, closer to river, closer to trails, lower elevation, gentler slope, lower tree cover areas.
Figure 15: Comparison between panda habitats before and after livestock number has increased, overlaid with GPS collar data from horses and cattle

3.3.6 GPS tracking

For horses, 46% GPS locations were in panda habitats. Within these points, 87% were in the lost habitat areas, and 8.3% were in the unchanged panda habitats. For cattle, 56% of the recorded locations were in panda habitats, 80% of which were in lost panda habitats and 16.3% were in unchanged panda habitats.
3.4 Discussion

3.4.1 Emerging threat from livestock

Degraded bamboos, changes in habitat use for pandas after the introduction of livestock, the overlap of livestock range and lost panda habitats, are all observations that reinforce our hypothesis that livestock grazing harms giant pandas.

Previously, researchers assumed there were no major competitors for giant pandas in the wild (Schaller 1985, Zhang et al. 2006, Hull et al. 2014, Wang et al. 2015). However, with the introduction of livestock and rapid increase in their numbers, they have become an invasive competitor for pandas. Similar to overgrazing in grassland areas, the overgrazing in forest ecosystems, especially on understory vegetation, is destructive.

Livestock caused considerable damage of bamboos and did so quickly. With reduced number of leaves and shorter stems, bamboos failed to generate healthy clones. The degradation further develops to the death of bamboo forests, which has been observed along the major valley in Wanglang. Some bamboo species in panda habitats are predicted to have a huge reduction in their distribution under climate change (Tuanmu et al. 2013). The impacts from livestock could further deteriorate the situation, reduce available habitats and result in a more fragmented landscape.

Pandas prefer gentle slope and were once widespread in the lower valley areas throughout their distribution (Schaller 1985, Liu et al. 1999). However, human
disturbances in lowland areas such as logging and agriculture have driven pandas to higher elevation and steeper slopes (Hull et al. 2014). In Wanglang, we found out similar trend because of livestock grazing. Although these animals are free-ranging, their owners come in by road to feed them salt 1-4 times a month and so the animals stay longer in the valleys. In winter, the animals migrate down to feed as deep snow covers the high altitude grasslands. Thus, salt and food dependence associated with valley areas has caused more intensive use.

With the conclusion that livestock does compete for food and habitat with giant panda, our results also demonstrate the necessity to incorporate livestock disturbance into panda distribution models. Species distribution modeling is widely used in conservation planning and species management (Guisan and Thuiller 2005), more recently in projecting the impacts of climate change and other human disturbance on the distribution of biodiversity (Pearson and Dawson 2003, Araujo et al. 2008). However, most of the studies only consider the relationship between a species and its abiotic environment and without taking into account the species interactions (Elith and Leathwick 2009, Wisz et al. 2013).

The livestock-panda interaction provided additional explanatory power beyond environmental factors and improved the model performance. The best model in our study is composed of the presence and absence of both types of livestock instead of the number of livestock. Thus, giant pandas may show lower tolerance of the livestock
disturbance than we expected. The existence of livestock could be enough to change the preference of habitat use for pandas no matter what the intensity is. In addition, the GPS collar tracking data showed that livestock mainly concentrate in the lost panda habitats. It further validated the competitive exclusion: livestock occupy the resources and drive pandas out of their suitable habitats.

The lost panda habitat overlaps mostly with the year-round habitats for livestock. Thus, the intensity and frequency of livestock grazing have a direct impact on panda habitat use. The expansions are in higher elevation and on the edge of the bamboo distribution limit. These areas cannot replace the role of the lower habitats that keep habitat connectivity and provide elevational range for seasonal migration.

Apart from degraded bamboos, 82% of the plots with livestock occurrence have some extent of bare ground because of the trampling from livestock. The reduction of moss, grass, and other above-ground vegetation significantly alters the microhabitats for other, smaller species, soil processes, and forest dynamics. We urge more studies for these issues. As giant pandas serve as protective umbrella for other endemic species (Li and Pimm 2016), we hope the impacts of emerging livestock grazing in this biodiverse mountane forests on pandas could draw attention to the impacts on the whole ecosystem.

Livestock could also harm panda breeding. In the northeast of Wanglang where there are predictions of habitat loss, there were fewer observations of pandas and scent
marks during the breeding season in the past years (personal communication with park manager). This area is critical for panda breeding (Liu et al. 2005). As livestock are prevalent throughout whole panda range, it is crucial to understand if livestock disturbance not only shape the distribution at local scale but also at regional or macroecological scales (Leathwick and Austin 2001, Araujo et al. 2008). Future studies should incorporate livestock information to improve prediction of panda distribution and habitat evaluation.

### 3.4.2 Diagnosis and possible policy changes

Natural Forest Conservation Project and Grain for Green have been effective in reducing soil erosion and increasing forest cover. The success in reducing logging and agriculture may have led to more livestock, which compromised the goals of protected areas and panda conservation. The average of 85% income loss from logging because of the logging ban from NFCP drove the shift of local livelihood. Livestock sector has boomed with the assistance of the payment from GFG for returning the cropland to forests in this region, encouragement from ICDP to raise horses for tourism, and the labor freed up from logging and agriculture (Uchida et al. 2007). Moreover, the weak regulation and law enforcement over livestock grazing inside state forests allow local communities to exploit forests in other ways. Especially for the relocated communities because of dam construction, the use of forest for livestock has become a major way of local income source. This is not unique to Wanglang, but common throughout China’s forests.
In 1997, the World Wildlife Fund (WWF) launched an Integrated Conservation and Development Project (ICDP) in the adjacent community to Wanglang with the claim that it was the first ICDP in giant panda habitats to develop alternative livelihoods for surrounding communities. This ambitious goal involved developing horse riding for tourism. WWF made no assessment of the likely impacts of this recommendation.

Later, road construction and relocation because of hydropower development, and a massive earthquake in Sichuan cut off the income from tourism and accelerated the growth of livestock. Moreover, the weak regulation and law enforcement over livestock grazing inside state forests allow the locals to take advantages of the forest in another way.

With the high death rate of animals because of overstocking and severe weather events, local communities face a riskier future. Moreover, there is less available labor as young people are less willing to work in rural communities. However, with the increasing demand for meat and its low cost, this traditional way of raising livestock will not die off by itself.

A conservative intervention might be to transform free-ranging livestock to feedlot operation, which could reduce the impacts on forest and the high death rate of animals. The government needs to provide initial support in facility building, training, and subsidy in animal feed. The more progressive intervention is to ban livestock with a payment for ecosystem service project. As horses are more detrimental to forests than
cattle, we suggest removing horses as the priority. The payment can come in one of two ways according to our discussions with the local community: an annual subsidy or job opportunities in the tourism sector.

With the rapid development of ecotourism in panda habitats, the avoidance of valley areas by pandas where most tourists stay and the loss of habitats will surely damage the economic interests of the local communities in the long run. Thus, apart from pursuing increasing forest cover, we encourage the local governments to implement related policies to control livestock, provide economic incentives and carry out monitoring and law enforcement to ensure the long-term success of panda conservation and socio-economic development.
Chapter 4. Identifying individual and sex of giant pandas through Footprint Identification Technique

4.1 Introduction

The giant panda (*Ailuropoda melanoleuca*) is one of the best-known threatened species, with an estimated 1864 surviving in the wild (State Forestry Administration 2015). They are isolated into 33 populations with 22 groups having fewer than 30 individuals, and 18 have fewer than 10 individuals (State Forestry Administration 2015). For the long-term survival and management, a thorough understanding of giant panda populations is crucial. To date there have been no ideal methods for identifying individuals or discriminating sex. Direct observation and counts are impossible because of its low population density, complex topography, and elusiveness of the species (Zhang et al. 2006). Unlike tigers or leopards, the similar appearance of individual pandas with no identifiable features such as strips or spots makes them difficult to differentiate from camera trap images. Here, we report the development of the giant panda Footprint Identification Technique (FIT) to sex and identify individual animals. It provides a potentially powerful tool to assist the management and conservation of this endangered species.

Currently, there are two primary methods used to identify individual giant pandas: the bite-size technique and DNA. The bite-size technique was originally used to differentiate age groups of pandas (Schaller 1985) and then was extended to identify individuals (Garshelis et al. 2008). Studies of giant pandas in both the wild and captivity
have shown individual differences in "bite size" and "chew rates" of the bamboo stems in their droppings (Schaller 1985, Yin et al. 2005). The method assumes that giant pandas are usually solitary with relatively stable home range. If two droppings are found with the same extent of freshness in a short survey period but are farther than the radius of home range apart, then they belong to two different individuals. If the two droppings are geographically close, the average bite sizes (usually measuring 100 stem/leaf fragments in droppings) are used for individual identification (Yin et al. 2005). If the average bite sizes for these two samples have no significant differences or are smaller than 2mm, then they are considered to belong to the same individual. This method has been used for the third (Year 1999-2003) and fourth (Year 2011-2014) national survey of giant pandas (State Forestry Administration 2015). (Wei et al. 2002, Zhang et al. 2006) are among its critics. It is less reliable in denser population areas or within mating clusters because many individuals may have similar bite sizes. Moreover, some significant variation in bite sizes within individuals could result in an overestimate of the population (Zhang et al. 2006). In addition, this method requires field staff to make very precise measurements to apply the threshold of 2mm (Yin et al. 2005). Measurement errors often preclude this precision (Zhang et al. 2006).

The alternative is using microsatellite analysis with fecal DNA (Zhan et al. 2006). This non-invasive DNA sampling was used with the bite-size estimates method in the fourth national giant panda survey (State Forestry Administration 2015). DNA
estimation is believed to be more accurate (Wei et al. 2015) but it requires the sample to be very fresh to exclude potential degradation and pollution of DNA. Thus, extensive survey effort is required. Challenges in finding sufficient samples have prevented applying this method successfully in large-scale studies. Moreover, the cost of laboratory processing has impeded its use for most practitioners.

There is no apparent sexual dimorphism in the giant panda; it is difficult to identify the sex of giant pandas even in captivity, without a DNA test. Generally, adult males are 10-20% larger than adult females (Smith et al. 2010). However, there is much variation, and it is particularly difficult to identify the sex of a solitary wild animal, outside the breeding season. Because the sexual organs are small and cryptic, it is even harder to sex sub-adults (Yang et al. 1999).

These challenges motivate the development of a robust and flexible technique to balance the accuracy required of a population estimate with the need for a low-cost field tool. The Footprint Identification Technique is a new, promising and cost-effective tool in wildlife conservation (Pimm et al. 2015). This non-invasive technique was first developed for black rhinos (Jewell et al. 2001). More recently it has been successfully adapted and applied for cheetah (Jewell et al. 2016) white rhinos (Alibhai et al. 2008), tigers (Gu et al. 2014), mountain lions (Jewell et al. 2014) and other endangered species.

Footprints have been used as signs of giant panda presence for many years (Fan et al. 2011, Wang et al. 2014, Li et al. 2015). Their footprints are characteristic and easily
found. No efforts have been made in extracting individual identification from footprints of this species until now.

First, thanks to a hugely successful captive breeding program China now has around 375 captive giant pandas (State Forestry Administration 2015). This large captive population provides the opportunity to develop the required FIT training library of footprints that can then be applied to identify unknown individuals in the wild. Secondly, China is now beginning to focus on the reintroduction of giant panda to protected areas. The development of an individual database of footprints for the wild population will be an essential part of monitoring these animals (Personal communication with Sichuan Forestry Administration).

4.2 Methods

4.2.1 Study population

We collected footprint images from captive giant pandas in the China Conservation and Research Centre for the Giant Panda (CCRCGP) in Sichuan, China. The CCRCGP has more than 170 giant pandas, accounting for 60% of the captive population in the world. It has three major bases; Ya’an, Du Jiang Yan, and Wolong National Nature Reserve. The Wolong base was almost destroyed by a magnitude 8.0 earthquake in 2008 and the Ya’an base now holds most of the animals. The Wolong base is in the heart of Wolong National Nature Reserve, which is one of 67 reserves designated by China’s government to protect wild giant pandas (State Forestry
Administration, 2015). Several semi-enclosures are built upon the hills, each with an average area of 0.33 km$^2$. This natural habitat provides perfect conditions for rehabilitating animals who will be reintroduced to free-ranging natural habitat. Of the 41 individuals, we used in the present study, 23 came from the Ya’an and 18 from Wolong. We used two individuals, Zhang Ka and Ye Ye from Wolong in the semi-enclosures with natural habitats as validation.

### 4.2.2 Study period

We collected data from captive animals from March 2014 until April 2016. Most footprints came from a prepared sand substrate since snowfall was infrequent at the lower altitudes where captive panda are held. There are two advantages of using sand. Data collection is independent of snowfall, and secondly sand is convenient to set up and reuse for any enclosure. Fresh sand was used for each animal to avoid any possible disturbance of behaviors from olfactory cues. On occasion, water was sprayed onto the sand when it was very dry. At the same time, we collected footprints on snow from captive animals at Wolong when enough snow had accumulated.

### 4.2.3 Foot anatomy and data collection

In addition to the five digits, each with a claw and digital pad, the giant panda has an unusual feature on the front feet – a ‘sixth finger’ or ‘thumb pad’. Whilst this acts as an opposable digit, it is in fact an enlarged sesamoid bone from the wrist. It enables them to better grab bamboos (Endo et al., 1999). Thus, the front footprint usually shows
six distinct digit pads, the metacarpal and carpal pad. The sesamoid bone imprints are
unique to giant panda prints. They have the advantage, for our purpose, of adding
complexity to the footprint, and affording more details in variables selected for the
development of the FIT algorithm.

Initial trials to investigate the clarity of the prints left by each of the four feet also
indicated that front foot impressions were more distinctive, detailed, and clearly
outlined. This was likely due to a combination of greater weight at the front of the
animal, and less fur on the front feet. We arbitrarily chose the left front foot for the FIT
model development. In common with bear species, pandas tend to over-step. That is,
instead of registering the hind foot impression on top of that made by the front foot, the
hind foot falls in front of the front footprint, leaving a clear front foot impression.
4.2.4 The footprint collection protocol

We define a trail to be an unbroken series of footprints from one animal. Using locally-sourced builder’s sand, we laid sand of approximately 3m length, 1m width and
1cm depth in the enclosure of each individual animal. We packed the sand on the ground using gardening tools and took images of each left front footprint in the trail using the protocol described in Jewell et al (2016). Digital single-lens reflex cameras and smartphone cameras both proved adequate to collect a clear image of the footprint. A carpenter’s scale or two rulers were laid perpendicular along the bottom and left axes of the footprint, with reference to the direction of travel. A paper ID slip which recorded the name, gender, age, date of collection and ID of the footprint was placed adjacent to the scale and included in each image. Footprint ID consisted of the number of the trail (e.g. trail #1) and the number of the footprint within that trail (e.g. 1c for the third footprint). Great care was taken to image from directly above the footprint and perpendicular to the plane of the footprint to avoid parallax error. The dimensions of each footprint vary with the gait of the animal, substrate type, moisture levels, slope of the ground and weather conditions. To account for this variation within the footprints of each individual, we collected multiple footprints from each.

4.2.5 Extracting a geometric profile from footprint images

The whole process of FIT analysis has been integrated into JMP, a software from SAS. After the taking the image, we imported it into JMP, resized and rotated for standardization (Jewell et al 2016). Scale points 1 and 2 are placed on the ruler at an interval of 10 cm. Landmark points are then placed at anatomical positions on the footprint, following software prompts. Landmarks 1 to 7 were selected on the basis of
foot anatomy to include those points which are clearly definable and repeatable across many footprints. These landmarks included five points on the centroids of the toe pads, one on the centroid of thumb pad and one on the distal end of the carpal pad indentation. Points 1 and 5 were used as rotation points along a horizontal axis (Figure 17). Using these landmark points, a further 15 derived points were computed. From these points, we finally extracted a total of 124 metrics consisting of lengths, angles and areas (Appendix F). The full set of points is designed to allow all measurements that one anticipates might prove useful in discriminating between footprints. Apart from the landmark points placed manually, JMP automates the rest of this process (Jewell et al 2016).

![Figure 17: Landmark points and computed derived measurements. At the left, the 7 landmarks that are needed to input manually, 1-5 on centroids of the digital pads, 6 on centroid of the thumb pad and 7 on the distal end of the carpal pad](image)
indentation. At the right, the 7 landmarks with the derived measurements and other variables automatically generated by FIT.

4.2.6 Data analysis

4.2.6.1 Individual identification

The algorithms for classification in FIT are based on a customized model that uses canonical variates of Fisher's approach to discriminant analysis (Jewell et al. 2016) to generate centroid plots. The top explanatory variables are selected using forward stepwise regression using their F-ratio. Then the first two canonical variates are constructed to map the trails in this two dimensional space. The centroid values (multivariate least-square means) and 95% confidence interval ellipses are plotted for each trail. In the FIT model construct, the presence/absence of overlap of the ellipses is used as a classifier. If the ellipses overlap, then these two trails are likely to belong to the same individual (Figure 18).

Figure 18: Two-way canonical plots. The red points are from all the female individuals in the database and blue points are for all the males. Each single point represents a footprint. Each ellipse is for the 95% confidence interval around the mean for a different trail showing trail data. In A, the green and red ellipses are from the
same individual Junzhu, a female, and in B, two different individuals (blue for the male Wu Jun and red for Jun Zhu).

The distance between the centroids is relative, depending on the matrix of within-group variations and the relative-position vector of the centroids. Thus, any changes to a testing set (adding or removing individuals) would alter the positions of the centroid values as well the ellipses. To solve this problem, we applied two modifications (Jewell et al. 2001). First, we applied the centroid plot technique on a pairwise basis, comparing two trails at a time. Second, we constructed a "reference centroid value (RCV)" using the other known individuals in the library as a reference point in the canonical space. The RCV functioned to stabilize the location of any test groups with respect to each other (Alibhai et al 2008, Jewell et al 2016).

When testing the accuracy of a species FIT algorithm it is necessary to optimize the values of three elements within the FIT model construct: the number of variables, the size of the confidence intervals around the ellipse, and the threshold value of the distances between the means. The Appendix F discusses the process of identifying the optimal combination of these three aspects.

We ran the holdback test to test the robustness of our model. We divided our 30 pandas into training set and test set. By varying the number of individuals in the training set, we can test accuracy of the model in predicting the number of individuals in the test set. We increased the training set size from 3 to 27, with test set size varying
accordingly from 27 to 3. For each test set size, we repeat three times using different sections of animals.

A dendrogram was produced to show the model results. It identified the Ward distance between each pair of trails, which is the distance between two clusters is the ANOVA sum of squares between the two clusters summed over all the variables. It is the base for identifying whether each pair of trails are from the same individuals or two different individuals. The same individuals predicted by the algorithm were clustered together and given the same color-code.

We divided our dataset to three parts, one training set and two test sets. The first test set had the individuals from the semi-enclosure where their footprints were collected from the natural panda habitats, to examine whether it applies to the field conditions. The second test set had the individuals with fewer than 6 footprints, and we used it to test limited sample size. We ran three accuracy assessments, one within model accuracy and two validations with the two test sets.

4.2.6.2 Sex discrimination

We applied linear discriminate analysis with stepwise variable selection to develop the sex discrimination algorithm. This process selects the parsimonious set of variables that provides the most discriminating power based on the F ratios to discriminate sex (Gu et al. 2014). It also excludes the highly correlated variables which may bias the estimate. The number of variables used in the model also displays a bell
shape distribution with the accuracy. Thus, by plotting the number of variables against accuracy, we can identify the optimal number for the most effective model. We ran five-fold cross-validation within the test set. This randomly partitions the data into five subsamples and repeat five times. For each time, it holds one subsample as test set and builds the model with the other four. Then we ran a blind validation with the independent test set of nine new individuals.

4.3 Results

In total, we collected 521 usable footprints along 80 trails from 41 individuals (details of individual information in Appendix F). We analyzed the data in three stages. First, of the 41 individuals, 30 had trails with a minimum of 6 footprints per trail. Footprint images for these individuals were collected in a sand substrate and these were used to extract the algorithm in FIT for individual identification. Table 9 summarizes the sex ratio, age, numbers of footprints and trails for the 30 individuals. Second, we collected footprint images for two individuals, with a minimum of 6 footprints per tail (Appendix F) in outdoor enclosures in snow. We analyzed the data for these individuals together with the 30 database individuals separately. Third, as a matter of interest, we subjected the rest of the individuals with <6 footprints per trail for validation.

<table>
<thead>
<tr>
<th># of individuals</th>
<th># of footprint images</th>
<th>Mean # of footprints (range)</th>
<th># of trails</th>
<th>Mean trails/individual (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>16</td>
<td>273</td>
<td>17.1 (8 – 31)</td>
<td>38</td>
</tr>
</tbody>
</table>

Table 9: Summary for 30 pandas
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>14</td>
<td>7.5 (2 – 14)</td>
<td>204</td>
<td>14.6 (6 – 33)</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>8.6 (2 – 15)</td>
<td>477</td>
<td>15.9 (6 – 33)</td>
<td>67</td>
</tr>
</tbody>
</table>

**4.3.1 Individual identification**

**4.3.1.1 Model details**

The FIT customized model for classifying trails is based on pairwise comparison of trails using discriminant analysis. During this process, each pair combination of trails is held back with the rest of the trails utilized for model building. The final output is in the form of a dendrogram giving a predicted number of individuals based on the number of trails used in the analysis. If the trails are identified to be from the same individual, they are clustered together with the same color. Figure 19 shows the output for a small sample of the dataset for six known individuals with a total of 12 trails subjected to FIT analysis. The analysis predicted six individuals with all the trails being clustered correctly.

The program allows the user to change the number of predicted individuals in the interactive window by varying the threshold and to get the likelihood of the varying number of individuals. This dendrogram remains the same. Setting the threshold to return 5 individuals would group Su Xing with Fa Fa, but the likelihood of this would drop to 10% when compared to the best classification. Setting the predicted number to return 7 individuals generates a likelihood of 97%.
Figure 19: Dendrogram of individual identification. Each row in the

dendrogram represents a trail from a known individual. The letter following the

name of an individual shows the ID of different trails. This shows the results

without changing the default setting for giant pandas, which correctly predict 6

individuals. The diamond shows the threshold value. Any trails branch out right to it

are identified as the same individual.

There were five variables that occurred more than 96% of the models for trail
discrimination, the total area of the footprint (minimum polygon constructed from 7
landmarks), Ar1/107 (ratio between total area and the angle between the lines from
centroids of toe pad 1 to carpal pad to the fake thumb), V16 (distance between centroids
of second toe pad and carpal pad), Area 8 (the area between the five toe pads and carpal
pad) and the distance between first and fifth toe pad centroids.

4.3.1.2 Holdback test

The FIT provided a more accurate estimate when the training set was more than
12 individuals and the test size fewer than 18 individuals. Figure 20 shows the varying
test set size plotted against itself, the predicted values and the means of predicted
values. For each test set size, the process was iterated 10 times. With smaller test set sizes
and large training set sizes, the predicted test set size matched the actual test set size very accurately and the range around the mean remained small. However, at a ratio of 15 test/15 training and 18 test/12 training, the range around the mean increased and beyond that, the level of accuracy of prediction declined.

Figure 20: Result of a comprehensive holdback trial for 30 pandas. If the test set size equals 3, then the training set size is 27 to build the model. The process was iterated 10 times for each test set size. The red line represents the 1:1 line of the true values. The grey triangles are the 10 predicted values at each test set size. The black squares are the means of predicted test set size. The trend is shown by linking the points.
4.3.1.3 Test of model for 30 pandas

Using the above algorithm for the giant panda, we tested its efficacy in three stages. First, we ran the FIT analysis for the dataset of 30 individuals with 477 footprints and 67 trails (Table 10). In this analysis, we compared all pairwise combinations for the 67 trails leading to a cluster dendrogram showing a prediction for the number of individuals and the relationship between self and non-self-trails. FIT model predicted 28 individuals (93.3% accuracy) with seven of the 67 trails misclassified (89.6% accuracy). Varying the threshold, the relative estimated likelihood of accuracy for 27 individuals was reduced to 78%. However, the likelihood for 29 or 30 individuals remained high at 91% and dropping off thereafter to 72% for 31 individuals. In other words, the FIT predicted 28 – 30 individuals.

Second, we iterated the analysis with a total of 32 individuals including two trails collected from semi-enclosures in a snow substrate from two known individuals. Figure 21 shows that the model predicted 29 individuals (90.6% accuracy) with the two added trails from Yeye and Zhangka being identified correctly as separate individuals.
Figure 21: Cluster dendrogram for 32 individuals including trails collected in snow from two known individuals Yeye and Zhangka from a semi-enclosure. The model predicted 29 individuals with the two added individuals, identified correctly as separate individuals. The highlighted trails are from these two individuals that are semi-enclosure in the natural habitats. The diamond shows the threshold value. Any trails branch out right to it are identified as the same individual.
Third, we ran the accuracy assessment with the additional trails from the 9 individuals that had fewer footprints than the requisite minimum number (6) for the FIT model. We analyzed these nine trails using the same algorithm in the FIT model. The model predicted eight individuals giving a surprisingly high accuracy (88.9%) (detail dendrogram in the Appendix F).

4.3.2 Sex discrimination analysis

35 variables were selected using stepwise selection for discrimination analysis (Figure 22). There were two key variables in this analysis present in all the models. The first one was Area 1, the total area of the minimum polygon constructed from the seven landmarks. The second one was the distance between two derived points. The first derived point was the intersection between the line that connecting centroid points of first toe pad and the fifth toe pad, and the line connecting centroid point 2 and 6. The second derived point was the intersection between line between centroid point 1 and 5, and the line between centroid point 3 and 7.
Figure 22: Model accuracy with varying numbers of variables used in the sex discriminant analysis. The accuracy increases with the number of variables in the model until it gets to 35 variables.

The subadults tended to confuse the algorithm, so we excluded the individuals that were younger than 3 years old. Footprints from males and females distributed separately in the canonical space in our discriminate analysis (Figure 23). The accuracy for the overall model was 89.8%, the average accuracy of 5-fold cross validation was 85.5±1.2%. The accuracy for the nine-individual test set was 75.8%.
Figure 23: Canonical plot from sex discrimination. Each red point is for female footprint and blue is for male. The separation of 95% confidence interval ellipses shows the success of differentiating sex.

4.3.3 Influence of substrates

We collected footprints from three individuals (Fa Fa, Hua Pu, Hua Rong) on both sand and mud to test the influence of substrate. Because the limited number of footprints, we could only use Fa Fa, which had more than three footprints on each substrate to test. Fa Fa A, which was the trail on mud was classified as the same individual with sand trail Fa Fa B. Zhang Ka, Ye Ye and Caocao that had footprints on snow were correctly classified as well. The locations of these individuals in the
classification dendrogram represented their similarities with other individuals. Instead of on a greatly separated branch, they mixed with other individuals (indicated by the shaded areas in Figure 21), which showed that the substrate did not make a significant difference in the measurements (variables) or the model results.

4.4 Discussion

First, sex discrimination is hard as pandas do not show apparent sexual dimorphism. That said, we could differentiate sex with an accuracy of 85% with trails of more than six footprints and 75% with trails fewer than six footprints.

Second, our method showed high accuracy (around 90%) for individual identification. Although with 12 individuals as training set could already provide robust discrimination, we used 30 individuals to establish a more comprehensive database for the future use. The complexity of individual identification is that it not only requires to predict the right number of individuals but also correctly cluster trails from the same individuals.

Third, our method is robust to various substrates. The success of correctly identifying snow or mud footprints also validate our hypothesis that using centroids could reduce the influence from different substrates. Two individuals Yeye and Zhangka who stayed in the natural habitats were also accurately identified, which indicates the application of this technique is not only limited to artificial conditions but also expands to field survey of the wild population.
Non-invasive sampling to identify individuals is more desirable in ecology and conservation studies and projects (Taberlet and Luikart 1999) than invasive methods. It avoids catching, handling, anesthetizing the animals, which could cause stress, possible injuries, change of behaviors or reduction of fertility (Alibhai et al. 2001). However, the need to understand the structure, dynamics and trend of the population is strong especially for rare and cryptic species.

We have considered the complex landscape features in wild panda habitats, which could cause unclear edges. Thus, we developed this unique technique to use only centroids instead of points on edges for our landmarks, which is different from many other species (Alibhai et al. 2008, Jewell et al. 2014, Jewell et al. 2016). Although we lose much information by doing so, our study still shows very desirable accuracy and are more practical to use in the field.

The good quality of footprints is crucial in determining the accuracy. The minimum requirement for the image is that key features - five toe pads, fake thumb and carpal pad can be recognized from the image. Although our method is robust for different substrates, it is unable to deal with distorted footprints, which may happen when the animals climb up and down the steep slopes and slide the feet. To improve the usability of images, we recommend allocating survey efforts of footprints on ridges and valleys where are flat. This condition can produce complete and clearer images. These areas are also ecologically important for pandas and frequently used as trails, water
sources, and territory marking sites (Schaller 1985, Liu et al. 2005, Hull et al. 2014). In addition, most of the field survey or monitoring program, camera trapping projects use these flat trails most frequently. Thus, using valleys and ridges allows both the feasibility of field surveys and usability of footprint images. Because leaf litter and mosses cover the ground in most of the panda habitats, it is relatively difficult to find footprints on this vegetation. Thus, we recommend carrying out footprint surveys in the snow.

Usually, 6-8 footprints per trail collected in the field can produce a reliable estimate. However, our blind validation dataset had 9 individuals with fewer than six footprints. The results are still desirable. However, we would still encourage future field collection to aim for more than six footprints per trail to produce a more reliable estimate.

The models that we have shown in this paper are already scripted into an add-in for JMP platform. It can be distributed and used easily by any conservationists for free. Any digital cameras or phones can be used to take photos following the standard procedure. The only cost will be the survey cost of footprints. Thus, FIT provides a cost-effective way to identify individual and sex, monitor population dynamic, and to carry out research and conservation plan.

Our technique has its advantages and limitations. The collection of fresh and clear footprints requires a certain amount of fieldwork, especially in snow. Tracking
down a trail of footprints is different from searching for feces or patrolling fixed routes. Field staff need the training to identify the correct foot from pandas and divert from regular route to search for a trail of footprints. Except for working in snow conditions and around water sources, we hope more studies could examine innovative ways to combine this technique with traditional methods such as camera trapping or invent lightweight and environmental friendly substrate to collect footprints year-round.

This cost-effective and non-invasive technique provides a powerful tool to study the population dynamics under different threats such as human disturbance and climate change. It also empowers the local conservation practitioners to monitor their target group and evaluate conservation projects on their own. Moreover, this technique can be used for citizen science. With the tourism development in southwest China, the tourists can get more engaged with panda conservation and participate in the data collection during their visits. Thus, we believe this technique can greatly benefit the future giant panda conservation.
Appendix A: Map of China with geographical features discussed in the text

Figure A1: Map of China with geographical features discussed in the text
Appendix B: Refining the IUCN ranges by elevation range and available habitat

As expected, threatened species have smaller range sizes than non-threatened species for all three taxa (Figure B1 a-c). Refining by first elevational range and then by remaining habitat must inevitably reduce the ranges sizes: the question is whether they do so similarly across all three taxa. On average, 51% for mammals, 59% for birds and 46% for amphibians of the original range, are in the species’ preferred elevational range. That is, original range maps for birds incorporate better elevational range information than for mammals (One-way Anova, F (2, 244) =5.489, p=0.0047; Post-hoc, Tukey-Kramer HSD, p>0.05) and amphibians (Tukey-Kramer HSD, p=0.039).

After further trimming by suitable vegetation types, amphibians and birds are similar to in the extent of reduction (36.2% and 38.9% of the remaining range from the last step respectively), while mammals experience a significant higher ratio (46.4%) compared to amphibians (One-way ANOVA, F (2, 244) =3.860, p=0.0224; Post-hoc, Turkey-Kramer HSD, p=0.016). This reduction means that within the elevational range, more areas are not forests (or forest and shrubland) for mammals. After both refinement processes, the original ranges from IUCN are reduced to 26%, 36%, and 28% for of the published ranges for mammals, birds, and amphibians respectively (Figure B1, a-c). Mammals and amphibians show a significant higher proportion of range reduction than
birds (One-way ANOVA, F (2, 244) = 5.297, p=0.0056; Post-hoc, Turkey-Kramer HSD, p=0.013 between mammals and birds, p=0.015 between amphibians and birds).

These results also show that some species that the IUCN designates as non-threatened have smaller ranges after trimming by elevation and remaining habitat than do threatened species. Furthermore, 18 non-threatened mammal species, 24 bird species and 25 amphibian species have their remaining ranges smaller than 20,000 km$^2$, which is the threshold to consider a species as threatened when applied to original ranges.
Figure B1: Box plots of range changes of forest species of a) mammals, b) birds and c) amphibians. The first panel shows the box plots of range size distribution for each species from IUCN. The second panel displays the remaining range after refined by the elevational range. The third panel shows the remaining range after refined by both elevation and habitat. The whiskers were drawn to the furthest point within 1.5×IQR (3rd quartile minus the 1st quartile).
Appendix C: Sub-regions for future conservation

There are four major gap species concentrated areas, central Sichuan, central Yunnan, Hainan and Nan Mountains that along the borders of Guangxi, Guangdong.
Guizhou and Hunan Province. For Sichuan, the prefecture cities Ya’an Leshan and Meishan have the most GAP species concentration. Some of the areas are protected by provincial nature reserves, which need to be improved for their management power.

The second area is central Yunnan. It includes three prefecture cities: Pu’er, Chuxiong Yizu Autonomous Prefecture, Yuxi Prefecture City. Hainan, which is the island south to mainland, also has a high concentration of gap species. It falls mainly in the south-western island including Baoting Lizu Miaozu Autonomous County, Qiongzhong Li and Miao Autonomous County, Baisha Lizu Autonomous County, Changjiang Lizu Autonomous County, Dongfeng County, Ledong Lizu Autonomous County, Northern part of San Ya prefecture city, northern tip of Lingshui Lizu Autonomous County and western part of Wanning county. The Nan Mountains area mainly includes Qiandongnan Miao and Dong Autonomous Prefecture of Guizhou Province, Guilin of Guangxi Province and Yongzhou of Hunan Province.

A note on data sources:

In our analysis, two mammal species, Lepus yarkandensis (near-threatened) and Proedromys bedfordi (vulnerable), receive almost no protection from the national nature reserves in their remaining ranges, as is the case with one bird species Garrulax courtoisi (critically endangered) and 20 amphibian species. IUCN has reported that Lepus yarkandensis and Proedromys bedfordi occur in some national nature reserves (IUCN, 2014). For Lepus yarkandensis, after refined by elevation and habitat, it occurs only in a
small area outside the national nature reserves. The species Proedromys bedfordi has been recorded in Jiuzhai Gou National Nature Reserve (IUCN, 2014). However, the range map of this species from IUCN is outside this national nature reserve. Thus, we would not infer any conclusion because of the conflicting information. Although no national nature reserve covers the existing range of Garrulax courtoisi, we have noticed that a national wetland park has been established specifically to protect its habitat.
Appendix D: Species by species estimates

In the three tables, we list the species, its IUCN status, whether it occurs only on Hainan or Taiwan, and our estimate of its minimum and maximum elevation.

We then simplify its habitat to whether it occurs in forest or in forest and scrub.

The next three columns are the original area of its range (as extracted from the IUCN maps), how much of this range is within a species’ elevation range, and how much of that area has suitable habitat. These — and all other areas — are in km². The next three columns show the proportional reductions by trimming the original range by elevation, then by habitat given elevation, then by both.

The next column is the area of a species’ range that is protected and the percentage of the range protected.

The next six columns are whether the species falls into various classes described in the text. The penultimate column is the area a species’ range overlaps with the range of the giant panda and the final column the proportion of a species’ range protected by panda National Nature Reserves.

Because of the dataset is huge, please refer to the online data archive:

http://onlinelibrary.wiley.com/store/10.1111/cobi.12618/asset/supinfo/cobi12618-sup-0001-Suppmat.pdf?v=1&s=723925fe83c8bcf31bc5cfac60789e27a5903bc8
Appendix E: Mapping Forest and Evaluation

We identified 1,036,378 km$^2$ of forested area, or 39.1% of the study area. The accuracy assessment for the forest map is below.

**Accuracy Assessment**

Figure E1 (A) shows the fraction of validation sites in natural forests correctly classified for 30, 40, 50, and 60% tree cover thresholds against annual rainfall. For example, at 1400mm of rain, the 60% threshold correctly classified only 65% of the validation points whereas the 30% threshold classified 93%. Above 1800mm, all four thresholds nearly always classified the selected points as forests. This dropped rapidly in drier areas. We added drier validation sites until the fractions of dry and wet sites were approximately proportional to their relative areas. Figure E1 (B) shows the points correctly (solid points) and incorrectly (red points) classified with the 60% threshold against a background where wet areas have >1800mm annual rainfall (blue) and dry areas have <1800mm. Below 1800 mm, the 30% threshold classifies at least 80% of the points correctly. Figure E1 (C) shows the correct and incorrectly classified points where we employ the 60% threshold for areas with >1800mm of rain and the 30% threshold below this.

Table E1 summarizes our accuracy assessment. For natural forests, we correctly classified 88.7%, while the other 11.3% were mainly in dry areas with less than 30% tree cover. While we can further lower the tree cover threshold to reduce the omission error,
we risk having a larger commission error. Commission errors overestimate a species’
range by including rubber, oil palm, and other likely unsuitable habitats within our
estimates of natural forest. We successfully excluded 95% of oil palms and 82% of rubber
from our forest classification.

**Comparison with ESA forest map and differences in the following analysis**

We compared different forest cover maps, our own and by ESA-CCI, and found
that the discrepancies mainly concentrate in the northern part of the study area,
especially in drier areas (Figure E1). ESA-CCI has higher forest cover in Yunnan and
northern Cambodia while our map has more forests in northern Laos, Vietnam and the
western coast of Myanmar. Our forest map tends to show less forest in dry areas where
the tree cover is sometimes lower than 30%. Identification of forests in this kind of
environment is extremely hard as the trees are sparse and definitions of forest may
change according to which threshold of tree cover to use. The northern study area is
among one of the areas with the greatest discrepancies and uncertainties between
different forest products.

Use of the ESA-CCI forest layer would affect the three taxa to varying extents.
Birds are less impacted than mammals and amphibians. After refined by the ESA-CCI
forest layer, 63% of amphibians, 71% of mammals, and 80% of birds have a range size
within 20% of the remaining range refined by our forest layer. For birds, the differences
range from 41% to 192% when using the two forest maps, while for mammals, the range
extends from 31% to 508%, and for amphibians from 10% to 17.9 times the range refined by our forest cover estimates. In most cases, the ESA-CCI overestimates the areas suitable for a species. While this process does not change the conservation status of any mammals or birds as to whether they are classified as species of concern, it changes the status of some amphibians. If we used ESA-CCI, it would upgrade one amphibian (*Hyalrana maosonensis*) to a species of concern while downgrading four species (*Leptolalax heteropus, Limnonectes poilani, Leptobrachium huashen, Xenophrys jingdongensis*) to a non-concerned species. Species information tables can be downloaded from:

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0160566#sec022

**Table E1: Confusion matrix and error estimate**

<table>
<thead>
<tr>
<th>Classification</th>
<th>Natural forest</th>
<th>Rubber</th>
<th>Oil Palm</th>
<th>Total</th>
<th>Error of Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Natural Forest</strong></td>
<td>173</td>
<td>38</td>
<td>10</td>
<td>221</td>
<td>21.7%</td>
</tr>
<tr>
<td><strong>Non-Forest</strong></td>
<td>22</td>
<td>168</td>
<td>190</td>
<td>380</td>
<td>5.8%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>195</td>
<td>206</td>
<td>200</td>
<td>601</td>
<td>5.0%</td>
</tr>
<tr>
<td><strong>Error of Omission</strong></td>
<td>11.3%</td>
<td>18.4%</td>
<td>5.0%</td>
<td></td>
<td>Overall Accuracy</td>
</tr>
<tr>
<td><strong>Overall Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>88.4%</td>
</tr>
</tbody>
</table>
Figure E1: Validation of forest cover. (A) Fraction of correctly classified forest points versus mean annual rainfall. (B) Correctly (black point) and incorrectly (red point) classified points using a 60% forest cover threshold. (C) The same, using the 60% threshold for areas with >1800 mm of rain and 30% for areas with <1800mm. Blue areas have >1800 of mean annual rainfall while beige areas have <1800 mm.
Appendix F: Methodological details for FIT

There are three components that determine the accuracy of FIT.

1. Number of measurements (variables). Too many measurements may lead to increased separation of self-trails and too few may lead to overlap of ellipses of non-self-trails, resulting in an underestimation. Thus, we ran our algorithm with from 8 to 20 measurements, to find the number which produced the highest accuracy.

2. Size of ellipse (confidence interval around the centroid value for each trail set). We used 95% confidence intervals around the centroid values during the trials.

3. Threshold value (Ward distance). Using the above elements, we used the clustering function in JMP software to generate a cluster dendrogram using the Ward distance. The threshold value which gave the highest level of classification accuracy in the holdback trial was used to generate a predicted value for the number of pandas in the dataset of known individuals.

In order to determine the most effective combination of the above elements, we ran series of tests by keeping the ellipse size (95% confidence interval around centroid) constant and varying the other elements. We ran holdback tests by varying the ratio of test/training sets to see how closely the predicted test set size matched actual test set size. Initially, to expedite the analysis, we ran the tests with random partition of the dataset.
into test/training set at ratios of 5/25, 17/13 and 25/5 and iterate the process three times with the number of variables from 8-20. Keeping the ellipse size (95% CI) and threshold value (1.5) constant, we ran a more detailed holdback trial at intervals of 3 randomly selected individuals for the 30 panda data set, i.e. test/training set ratio of 3/27, 6/24 etc. ending with 27/3 (Figure F1).

Figure F1 shows the result from the first holdback tests varying the number of variables used in the analysis from eight to 20, with the ellipse size (95% CI) and threshold value of ward distance (1.5) constant. The most accurate match between predicted and actual numbers was achieved with 12 variables.

Figure F1: Holdback test for different number of variables, plotting predicted test size against the true test size. The holdback test was performed for each number of variables at test/training size of 5/25, 17/13 and 25/5 for 30 pandas using the FIT
model. The line goes through the means of three trials at each test size. The black line represents 1:1 line of the true values.

Figure F2: FIT Image feature extraction window in JMP software showing the landmark points and computed derived points.
Figure F3: Cluster dendrograms for 67 trails from 30 individuals showing classification of trails and predicted values. The FIT model predicted 28 individuals (A), with a relatively lesser likelihood of 27 individuals (B) but a higher likelihood of 30 individuals (C).
Figure F4: Blind validation for individual identification using independent 9-individual test set with $\leq 5$ footprints per trail. The model predicted eight individuals.

Table F1: The number of variables extracted from the footprint images as lengths (L), angles (AN), and areas (AR).

| V1 | L 01-02 | V2 | L 02-03 | V3 | L 03-04 | V4 | L 04-05 | V5 | L 05-06 | V6 | L 06-07 | V7 | L 07-01 | V8 | L 01-05 | V9 | L 01-06 | V10 | L 02-06 | V11 | L 03-06 | V12 | L 04-06 | V13 | L 07-05 | V14 | L 07-04 | V15 | L 07-03 | V16 | L 07-02 | V17 | L 01-15 | V18 | L 14-15 | V19 | L 13-14 | V20 | L 12-13 | V21 | L 06-12 | V22 | L 02-17 | V23 | L 17-19 | V24 | L 19-20 |
|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|----|---------|
| V32 | L 26-27 | V63 | L 07-11 | V94 | AN 2-3-4 |
| V33 | L 06-27 | V64 | L 07-08 | V95 | AN 2-3-7 |
| V34 | L 02-16 | V65 | L 01-14 | V96 | AN 7-3-6 |
| V35 | L 15-16 | V66 | L 06-14 | V97 | AN 6-3-4 |
| V36 | L 07-15 | V67 | L 03-14 | V98 | AN 1-2-3 |
| V37 | L 03-18 | V68 | L 07-14 | V99 | AN 1-2-7 |
| V38 | L 18-19 | V69 | L 01-18 | V100 | AN 7-2-6 |
| V39 | L 19-14 | V70 | L 05-18 | V101 | AN 6-2-3 |
| V40 | L 07-14 | V71 | L 05-15 | V102 | AN 2-1-7 |
| V41 | L 04-25 | V72 | L 07-15 | V103 | AN 2-1-5 |
| V42 | L 23-25 | V73 | L 04-13 | V104 | AN 5-1-6 |
| V43 | L 20-23 | V74 | L 07-13 | V105 | AN 6-1-7 |
| V44 | L 13-20 | V75 | L 02-15 | V106 | AN 5-1-7 |
| V45 | L 07-13 | V76 | L 07-15 | V107 | AN 6-7-1 |
| V46 | L 05-27 | V77 | L 09-10 | V108 | AN 6-7-5 |
| V47 | L 22-27 | V78 | AN 5-6-7 | V109 | AN 5-7-4 |
| V48 | L 21-22 | V79 | AN 5-6-4 | V110 | AN 4-7-3 |
| V49 | L 12-21 | V80 | AN 4-6-3 | V111 | AN 3-7-2 |
| V50 | L 07-12 | V81 | AN 3-6-2 | V112 | AN 2-7-1 |
| V51 | L 01-16 | V82 | AN 2-6-1 | V113 | AN 5-7-1 |
| V52 | L 16-17 | V83 | AN 5-6-1 | V114 | AR 1 |
| V53 | L 17-18 | V84 | AN 1-6-7 | V115 | AR 2 |
| V54 | L 18-24 | V85 | AN 4-5-6 | V116 | AR 3 |
| V55 | L 24-26 | V86 | AN 4-5-1 | V117 | AR 4 |
The numbers 01-07 refer to the landmark points and 08-27 to the derived points (see Figure F1). Variables 08 – 11 were generated at the intersection of two vectors e.g. V08 refers to vector from points 07 horizontal to 01 vertical. The areas were generated as polygons using the most peripheral points in each case. Area 1 = 1,2,3,4,5,6,7,1, Area 2 = 1,2,3,4,5,1, Area 3 = 1,5,6,7,1, Area 4 = 1,5,7,1, Area 5 = 1,2,3,4,5,6,1, Area 6 = 1,6,7,1, Area 7 = 1,5,6,1 and Area 8 = 1,2,3,4,5,7,1.

### Table F2: Detail information for the 41 panda individuals

<table>
<thead>
<tr>
<th>Panda</th>
<th>Sex</th>
<th>age</th>
<th># of footprint</th>
<th># of trails</th>
<th>substrate</th>
<th>location</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cui Cui</td>
<td>F</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Gong Zhu</td>
<td>F</td>
<td>15</td>
<td>20</td>
<td>3</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Huamei</td>
<td>F</td>
<td>15</td>
<td>11</td>
<td>2</td>
<td>sand</td>
<td>Wolong</td>
<td>Training</td>
</tr>
<tr>
<td>Huarong</td>
<td>F</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>sand</td>
<td>Wolong</td>
<td>Training</td>
</tr>
<tr>
<td>Jun Zhu</td>
<td>F</td>
<td>10</td>
<td>13</td>
<td>2</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Le Sheng</td>
<td>F</td>
<td>13</td>
<td>31</td>
<td>4</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Nana</td>
<td>F</td>
<td>10</td>
<td>21</td>
<td>3</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Qian Qian</td>
<td>F</td>
<td>7</td>
<td>10</td>
<td>1</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Qing Qing</td>
<td>F</td>
<td>7</td>
<td>23</td>
<td>3</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>ShenBing</td>
<td>F</td>
<td>5</td>
<td>28</td>
<td>4</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>ShuQin</td>
<td>F</td>
<td>5</td>
<td>26</td>
<td>3</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Sulin</td>
<td>F</td>
<td>15</td>
<td>14</td>
<td>2</td>
<td>sand</td>
<td>Wolong</td>
<td>Training</td>
</tr>
<tr>
<td>Ximei</td>
<td>F</td>
<td>15</td>
<td>13</td>
<td>2</td>
<td>sand</td>
<td>Wolong</td>
<td>Training</td>
</tr>
<tr>
<td>Xin Nier</td>
<td>F</td>
<td>7</td>
<td>8</td>
<td>1</td>
<td>sand</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>You You</td>
<td>F</td>
<td>15</td>
<td>21</td>
<td>3</td>
<td>sand/snow</td>
<td>Yaan</td>
<td>Training</td>
</tr>
<tr>
<td>Name</td>
<td>Gender</td>
<td>Age</td>
<td>ID</td>
<td>Breed</td>
<td>Location</td>
<td>Status</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>-------------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Zhen Zhen</td>
<td>F</td>
<td>6</td>
<td>16</td>
<td>2 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>An An</td>
<td>M</td>
<td>5</td>
<td>14</td>
<td>2 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Ao Ao</td>
<td>M</td>
<td>6</td>
<td>10</td>
<td>1 sand</td>
<td>Wolong</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Fa Fa</td>
<td>M</td>
<td>2</td>
<td>11</td>
<td>1 sand/mud</td>
<td>Wolong</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Hua Rong</td>
<td>M</td>
<td>2</td>
<td>6</td>
<td>1 sand/mud</td>
<td>Wolong</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Hui Hui</td>
<td>M</td>
<td>8</td>
<td>29</td>
<td>4 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Lan Zi</td>
<td>M</td>
<td>12</td>
<td>7</td>
<td>1 sand</td>
<td>Wolong</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>LuLu</td>
<td>M</td>
<td>14</td>
<td>20</td>
<td>3 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Mei Sheng</td>
<td>M</td>
<td>10</td>
<td>14</td>
<td>2 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Su Xing</td>
<td>M</td>
<td>2</td>
<td>8</td>
<td>1 sand</td>
<td>Wolong</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Tong Tong</td>
<td>M</td>
<td>10</td>
<td>33</td>
<td>4 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Wu Jun</td>
<td>M</td>
<td>6</td>
<td>23</td>
<td>3 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Xi Wang</td>
<td>M</td>
<td>8</td>
<td>6</td>
<td>1 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Yang Yang</td>
<td>M</td>
<td>12</td>
<td>7</td>
<td>1 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Yi Bao</td>
<td>M</td>
<td>8</td>
<td>16</td>
<td>2 sand</td>
<td>Yaan</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Yeye</td>
<td>F</td>
<td>17</td>
<td>9</td>
<td>1 snow</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Zhangka</td>
<td>F</td>
<td>16</td>
<td>7</td>
<td>1 snow</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Caocao</td>
<td>F</td>
<td>14</td>
<td>3</td>
<td>1 snow</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Wolong female 1</td>
<td>F</td>
<td>N/A</td>
<td>2</td>
<td>1 snow</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Yue Liang</td>
<td>F</td>
<td>2</td>
<td>5</td>
<td>1 sand</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Hua Pu</td>
<td>M</td>
<td>3</td>
<td>4</td>
<td>1 sand/mud</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Wu Gang</td>
<td>M</td>
<td>14</td>
<td>3</td>
<td>1 sand</td>
<td>Yaan</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Xi Xi</td>
<td>M</td>
<td>2</td>
<td>3</td>
<td>1 sand</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Yuan Yuan</td>
<td>M</td>
<td>14</td>
<td>4</td>
<td>1 sand</td>
<td>Yaan</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Zi Tao</td>
<td>M</td>
<td>6</td>
<td>2</td>
<td>1 sand</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>Zi Yan</td>
<td>M</td>
<td>6</td>
<td>2</td>
<td>1 sand</td>
<td>Wolong</td>
<td>Test</td>
<td></td>
</tr>
</tbody>
</table>
References


Lu, Z., W. Pan, X. Zhu, D. Wang, and H. Wang. 2000. What has the panda taught us? In: Entwistle A. and Dunstone N. (eds), Priorities for the Conservation of


State Forestry Administration of P.R China, 2015. The giant pandas of China: status quo - major findings of the fourth national survey on giant panda.

http://www.forestry.gov.cn/main/69/content-743562.html


van Noordwijk, M., H. L. Tata, J. Xu, S. Dewi, and P. A. Minang. 2012. Segregate or integrate for multifunctionality and sustained change through rubber-based


Biography

Binbin Li was born in China on May 2nd 1988. She grew up in Beijing and earned her B.S. in Life Sciences with a dual degree in Economics from Peking University in 2010. She then came to the United States and earned her M.S in Natural Resources and Environment from University of Michigan. Binbin stated her Ph.D. study with Dr. Stuart Pimm at the Nicholas School of the Environment in 2012. Her PhD was supported by Chinese Scholarship Council and focused on endemic species conservation and management of protected areas in China.

During her study, she has published several papers including: “Effects of feral cats on the evolution of anti-predator behaviors in island reptiles: insights from an ancient introduction” (2014), “China’s endemic vertebrates sheltering under the protective umbrella of the giant panda” (2015), “Remotely sensed data informs Red List evaluations and priorities mountainous areas for endemism conservation in Southeast Asia” (2016), “Fine with heat, problems with water: microclimate alters water loss in a thermally adapted insular lizard” (2016), and “Incorporating explicit geospatial data shows more species at risk of extinction than the current Red List (2016)”. Her research has been funded by Ocean Park Conservation Foundation Hong Kong, China State Forestry Department: Giant Panda International Collaboration Fund, The Explorer Club, with International Dissertation Research Travel Award and Aleane Webb Dissertation
Research Award from Duke University. In 2014, she was awarded NASA-MSU Professional Enhancement Award and won the best student presentation in US-IALE conference. She serves on the board of experts for Giant panda national park program in Sichuan Eco-environmental Protection Institute.

Binbin is devoted to community service and science communication. She has served as Vice-President in Duke University Chinese Students and Scholars Association for two years and on the board for Duke Story for Nature and People for one year. She is a nature photographer and has been served on the advisory board for Disney nature documentary “Born in China”. She started her own social media platform “Young Conservationist” to promote conservation related issues in China.