Imperial College London

Assessing the validity of Hainan's 400-year gazetteer record for establishing historical baselines

Connor Walsh

August 2016

A thesis submitted for the partial fulfilment of the requirements for the degree of Master of Science at Imperial College London

Submitted for the MSc in Conservation Science

Declaration of Own Work

I declare that this thesis:

Assessing the validity of Hainan's 400-year gazetteer record for establishing historical baselines

is entirely my own work, and that where material could be construed as the work of others, it is fully cited and referenced, and/or with appropriate acknowledgement given.

Signed.

Name of student: Connor Walsh

Name of Supervisors: Dr Samuel Turvey; Clare Duncan

Table of Contents

List of acronyms

CE: Current Era

OLE : Optimal linear estimation

Assessing the validity of Hainan's 400-year gazetteer record for establishing historical baselines

Abstract

Conservationists have sampled the enormous collection of local gazetteers in China to find species information dating from as much as 1,500 years ago until the 20th century. However, the inherent statistical completeness and biases of the gazetteers, and their ability to provide a robust historical baseline of past biological conditions and faunal changes has yet to be tested. This paper collates gazetteer references to the world's rarest ape, the Hainan gibbon, and other mammal species on Hainan island, China's southernmost province. The records cover 1521—1935, with 114 gibbon records. The density of records increases through time towards the present. After qualitatively scoring these records they were analyzed with optimal linear estimation (OLE), a technique to estimate extinction date based on limited historical sightings. The findings show the gazetteer record is robust because it predicts survival for species that are still extant today, past the end of the gazetteer record. This suggests gazetteers may provide useful information on past environmental conditions and change. The gazetteers also record wolf and dhole, species that are not otherwise confirmed to have occurred on the island. Further OLE on those 2 species suggests they may have been present on Hainan well into the 20th century.

Keywords: Dhole, Hainan gibbon, optimal linear estimation, taxonomic boundaries

Word count: 5967

Acknowledgements

Introduction

Historical baselines are used in conservation to help make sense of current environmental patterns and processes, inform management decisions, and place conservation into wider cultural and intellectual contexts (Szabó 2010; Rick & Lockwood 2013; Davies 2014). Shifting baseline syndrome, where the perspective of what is 'normal' is degraded over time, can mislead not just on population size (Pauly 1995) but even the possible former regional existence of large mammal species. However, historical sources seldom match the requirements of a contemporary biodiversity survey (Rick & Lockwood 2013): records may be incomplete, organisms may be described according to folk classifications which ignore or pre-date standard scientific nomenclature (Fleck et al 1999); and over centuries the individual observer will change many times. Therefore while historical baselines are valuable, the sources must be critically evaluated before they can confidently be applied to inform conservation science and management.

With its ancient written record (Wilkinson 2015), China appears to hold strong potential as a place to identify historical baselines. The need is urgent: anthropogenic pressures have led to extreme environmental destruction in the 20th and 21st centuries. In the modern era China has seen repeated waves of environmental damage: since the founding of the People's Republic of China in 1949 the human population has doubled, post-war economic prioritization led to land conversion, which was followed by the Great Leap Forward from 1958 to 1961, another period of major environmental damage (Edmonds 1999); and subsequent economic growth spurts in the 1980s, 1990s, and 21st century were not matched by conservation action (Liu & Diamond 2005). The Environmental Protection Law came into effect only in 1989. With the likely loss of the Yangtze river dolphin or baiji (*Lipotes vexillifer*), China is the scene of one of only 2 extinctions of an entire mammal family in the past 2 centuries (Turvey 2009). In this context native biodiversity baselines are urgently needed.

Previous investigations into the past status of Chinese biodiversity using historical sources have been conducted by Wen (2009, 2013), Chatterjee et al (2012), and Turvey et al (2015b). Their approaches began in the scholarly tradition of sinology (the study of pre-modern China) and advanced to conservation biology with GIS and statistics, potentially identifying the most appropriate habitat types to restore, and predicting

responses to climate change (Chatterjee et al 2012; Turvey et al 2015b). However before historical baselines can be constructed using historical texts, the original records still need to be critically examined to see just how robust a source they are.

Of all the biodiversity challenges in China, arguably the most urgent is that of the Hainan gibbon (*Nomascus hainanus*), the rarest ape on Earth, with fewer than 30 individuals living, limited to just one nature reserve, on the island of Hainan (Turvey et al 2015a). In this situation any new information on the species' ecology, distribution, and dynamics are valuable. In this paper I analyze historical gazetteer records from Hainan Province dating from 1521 to 1935 to assess mentions of Hainan gibbons and other mammals over time, with the broader goal of examining the validity of the sources that previous studies have taken when using gazetteers to inform Hainan gibbon conservation (Chatterjee et al 2012; Turvey et al 2015b). There are no known Hainan gibbon population estimates to compare with before the late 20th century (Turvey et al 2015a), so the analysis is based on whether the animals were likely extant in 1935 (by virtue of being extant today). This can simply indicate whether the gazetteer record is robust.

In this study the biases and errors to contend with include inexpert observations; editorial characteristics of the gazetteer style; irregular publication dates; and different ideas of what constitutes a species. A qualitative process involving historical, linguistic, statistical and taxonomic steps can aspire to see past these biases.

Identifying species in these historical sources is resource-intensive. While the gazetteers list a range of fauna and flora, I have restricted this study to mammals. The various authors' understandings of plants and animals will differ greatly, both between kingdoms and temporally, therefore comparing records from different taxa over time may introduce additional biases. Only comparing mammals aims to limit this.

The spatial focus of this study, Hainan, is a tropical island province of about 34,000 km^2. It lies south of the Chinese mainland at about 20° north and as close as 240 km from Vietnam. The region it sits in, Southeast Asia, has (along with South Asia) the highest density of threatened land mammals (Schipper et al 2008). In the 20th century Hainan faced huge environmental pressures, with 35% of potential gibbon habitat lost between 1991 and 2008 (Zhang et al 2010). Potentially confirming the gazetteers as a valid tool for researching

local historical baselines may open up biological and social conservation strategies, meeting an urgent need.

To test the Hainan gazetteers I apply optimal linear estimation (OLE) (Roberts & Solow 2003; Solow 2005) to the qualitatively scored sightings of gibbons and other mammals. Comparing the OLE extinction estimates, based on gazetteer data, to the otherwise known presence (or not) of species can indicate the internal coherence of the gazetteer sightings record.

In this paper I evaluate the Hainan gazetteers as a potential source for establishing species historical baselines up to 1935. To do this I synthesize and analyze data on past species records in the Hainan gazetteers, to then evaluate the validity of this historical archive at predicting species persistence. I then use that assessment to assess records of otherwise unconfirmed mammals on Hainan.

Methods

The local gazetteer is a historical genre in China of which 8,584 editions exist up to 1912, covering 1,500 areas of China and some neighboring countries (Wilkinson 2015). They often served as hand-over documents for civil servants assigned from elsewhere in China. As a need-to-know guide to the taxable capacity of a county, prefecture, or province, they include relevant local biographies, culture, topography, and produce (Bol 2001; Wilkinson 2015). The scattered books and fragments of Hainan's historical gazetteers were located, edited, and published in 2003 as a collection of 68 hardback volumes (Hong 2003). After a digitization process (see Supporting Information for the full process) I had 39 machine-readable volumes dating from 1521 to 1938 which form the reference text for this study. I discarded 23 corrupted volumes and 6 missing volumes as bad data. The final volume, from 1938, was also discarded because it deals exclusively with other islands under the administration of Hainan, but not Hainan island itself. This leaves 1935 as the final date in the study for which potential environmental data are available from the gazetteers. The 38 volumes formed a corpus which I searched with AntConc version 3.4.4 (Anthony 2014). My initial keyword searches were based on terms for apes or gibbons, two variants of *yuán* (see Supporting Information for the Chinese characters of the search terms). I collected these entries in a spreadsheet along with the gazetteer name, publication year, administrative unit, date, specific geographical information, page number, translation, and the other mammals listed in the surrounding paragraphs (see Supporting

Information for a reduced version of the spreadsheet). I checked that places named in the records were indeed on Hainan—it was at times administered as a single unit with the nearby mainland provinces, necessitating confirmation of which entries were spatially relevant. The spatial scale of the gazetteers is variously county, prefecture, whole-island, or province. The quality of maps is in the gazetteers is also variable, with internal boundaries not appearing until 1930 (Fig. 1). Thus the spatial unit in this study is the island of Hainan, regardless of the individual gazetteer's unit.

Figure 1: Historical maps of Hainan created in 1689 (Hong 2003) and 2007 (CHGIS 2007) indicate the spatial difficulties inherent in using the gazetteers for identifying historical land mammal range.

I scored the entries for reliability according to the following criteria:

EX: Record exists but it is excluded from analysis, because it describes an animal which is called an ape but is definitely not a gibbon (e.g. it has a tail; it is the wrong color); or is identifiably referring to another part of China.

0: no gibbon record of any kind available for a given gazetteer. Counted towards gibbon absence data.

1: Record is mythical or fantastical, e.g. inkpot gibbon, fantastical poetry, or extremely limited, such as place names incorporating 'ape'. E.g.: *In addition, Qiongzhou* [Hainan] *has stone apes, the small ones the size of a fist, they grow when fed water from a well. Also* [known as] *black*

1689

2007

apes, they can rub ink [from an inkstone], *jumping into the ink pot when finished.*

2: Record contains a brief mention of a gibbon, but with no or very limited detail; record consists of 'naturalistic poetry'; or record is a standard text quoted from an earlier gazetteer. E.g.: *Gibbons: there are three types: Golden ones are yellow, jade-faced are black, and there are also pure black ones. The golden and jade-faced ones are all rare.*

3: Record is original, and contains relatively detailed description of either gibbon characteristics/biology and/or geographical location of gibbons. E.g.: *Gibbons: The male is black and the female pale. They are like monkeys but their arms are very long, climbing and grabbing branches in the canopy with great agility. If they ever fall to the ground they stiffen up like a tree. As they are good at climbing trees but cannot walk, they need to stay in the trees to be raised, as if they get near the earth vapors they fall ill and die.*

These scored gibbon data can be categorized to provide time-series of sightings records with either greater resolution (any score) or reliability (mean and top scores). I separated it into tables of top score per year; mean score per year; and summed score per year, to see what insights into the data this might reveal when OLE is applied.

I identified the book-sections which usually had gibbon references, searched for sections with those titles across the corpus, and identified those that did not mention gibbons. I then further recorded every mammal in the 'beasts' (*shòu*) sections across the corpus. I identified all of the mammal species possible by cross-referencing historical Chinese dictionaries (Cihai 1948; Guo 1981; Hanyu dacidian 1986; Guhanyu changyongzi zidian bianxiezu 1998; Duan 1997), 19th century British accounts (Swinhoe 1870) and the most comprehensive modern Chinese mammal guide (Smith & Xie 2008), and then conducted new full-corpus searches with them. This was a more rapid system than that used with gibbons: I counted only entries which scored equivalent to 2 or 3 in the above criteria, and saved only the number of entries, gazetteer name, and year. I took one or more records of a given species in one gazetteer edition to equal one sighting. In a year with multiple publications, each edition with a record was counted as an individual sighting.

To run OLE analysis on the animal records, I used R version 3.3.0 (R Core Development Team 2016) with the package sExtinct (Clements 2013). Confidence intervals can be determined during the calculations. OLE accuracy improves with

the number of records, requiring at least 5 and preferably more than 10 (Collen & Turvey 2009; Boakes et al 2015).

To look for a trend in how the records were distributed over time, I applied Mann-Kendall trend analysis to the variations on the gibbon data in R version 3.3.0 with the trend package (Thorsten 2016). For gazetteer publication, gibbon records scored 2, scored 3, and scored 2 or 3 and other species this is a time series of 1 for any occurrence in a given year, 0 for no occurrence. I also ran Mann-Kendall trend analysis on time series of the annual top scores, mean scores, and summed scores.

Results

My searches revealed 114 total references to the search terms *yuán* in 29 individual gazetteers, of which 28 records were excluded for clearly not referring to gibbons, 35 were scored 1, 47 were scored 2, and 4 were scored 3. The archaic words for ape *xīngxing* and *fèifei* (van Gulik 1967) are not found in the corpus. Of the 114 gibbon references, 21 were found in sections on 'beasts' (*shòu*). A search of the other editions revealed only 3 which had a 'beasts' section that did not mention gibbons. Four years had multiple publications, where gazetteers were produced on a county scale, referring to different parts of Hainan.

The initial count of other named types of mammal in the Hainan gazetteers was 128 but only 16 of these types could be confidently identified as corresponding to recognized species concepts, of which 13 are mammals, and 8 both robust and plentiful enough to analyze. Most of the excluded animals have pre-modern names and little or no description. By way of example there are 16 variations on *lí*, usually translated as civet, and 15 variations on *shǔ*, a broad term for rodents; these two categories together could cover a range of civets, mustelids, small felids, rodents, and perhaps binturong, and I could only reasonably identify 5 of

these mammal types to modern species concepts.

Four types of bear are listed (*rénxiong*, *mǎxióng*, *gǒuxióng*, *zhūxióng*) as well as the generic bear (*xióng*). They are described as having identical bodies but different faces, probably ruling out sun bear (*Helarctos malayanus*) which is found in the region. I collapsed all of these into one and assumed it was Asiatic black bear (*Ursus thibetanus*), the species of bear known to exist on Hainan today, as safe reading, given that e.g. the giant panda (*Ailuropoda melanoleuca*) was traditionally classified as a type of leopard (Li 1957). Variations on wild pig (*yězhū*) and mountain pig (*shānzhū*) likewise are counted as referring to wild boar (*Sus scrofa*). Two terms (*chuānshānjiǎ*, *línglǐ*) are presented as synonyms for pangolin (*Manis pentadactyla*). Both the generic monkey (*hóu*) and more specific macaque (*mí*) are taken as rhesus macaque (*Macaca mulatta*), the only monkey known to have occurred historically on Hainan. Some leopard accounts distinguish between three types (*bào*, *àibao*, *jīnqiánbào*) which I have taken as clouded leopard (*Neofelis nebulosa*), because this is known to exist on Hainan (Lau et al 2010) and all three names contain the generic 'leopard' (*bào*). In the Chinese naming, as in English but not in zoology, all three are 'leopards'. The

patterning of leopard (*Panthera pardus*) is described explicitly in *jīnqiánbào*, which appears 7 times, however each entry is an identical description quoting a 16th century pharmacopeia (Li 1957). The lack of any original information or language in the entries suggests this is defaulting to standard texts rather than reporting based on local knowledge, and likely a shorthand interpretation for clouded leopard.

The final two species that could be identified relatively confidently to modern mammal species concepts appear to be dhole (*Cuon alpinus*) and wolf (*Canis lupus*) (*chái, láng* respectively). Neither is confirmed on Hainan. Supporting evidence suggests that the use of these names may be valid—both species are or were historically present in Chinese and Vietnamese regions adjacent to Hainan, and would likely have been known to the officials compiling the gazetteer (Smith & Xie 2008; Lau et al 2010). The 19th century naturalist Robert Swinhoe (1870) took '*the first* [dhole] *to imply a species of* Lupus, *and the other* [wolf] Nyctereutes procyonides, *Gray; but it is very questionable whether either occurs in Hainan.*' However, illustrated 20th century dictionaries suggest the raccoon dog is more likely to be described as a type of *lí* (Hanyu dacidian 1986), and the raccoon dog is not known from Hainan, making it no more likely as a possible species identification based on independent evidence (Smith & Xie 2008). The gazetteers record 11 types of *lí* to which I could not ascribe a species, so it may be more likely that the raccoon dog, if present at all, would fit in there, leaving the *láng* more likely to be, as it is commonly understood, wolf. Where described in the gazetteers, a standard text from about 1,500 earlier is re-used: '*like a hound, with a pointed head and pale cheeks, tall at the front and wide at the rear*' (Duan 1997). Dhole is recorded 13 times, from 1618 to 1935. Seven of the records share 2 standard descriptions, again standard texts pre-dating the gazetteer collection by more than 1,000 years.

The gibbon records become more frequent in later years (Fig. 2).

Fig. 2: Spread of gibbon records in the Hainan gazetteers, according to (a) mean score per record, (b) top score per year, (c) sum of scores per year, and (d) total number of mentions per year.
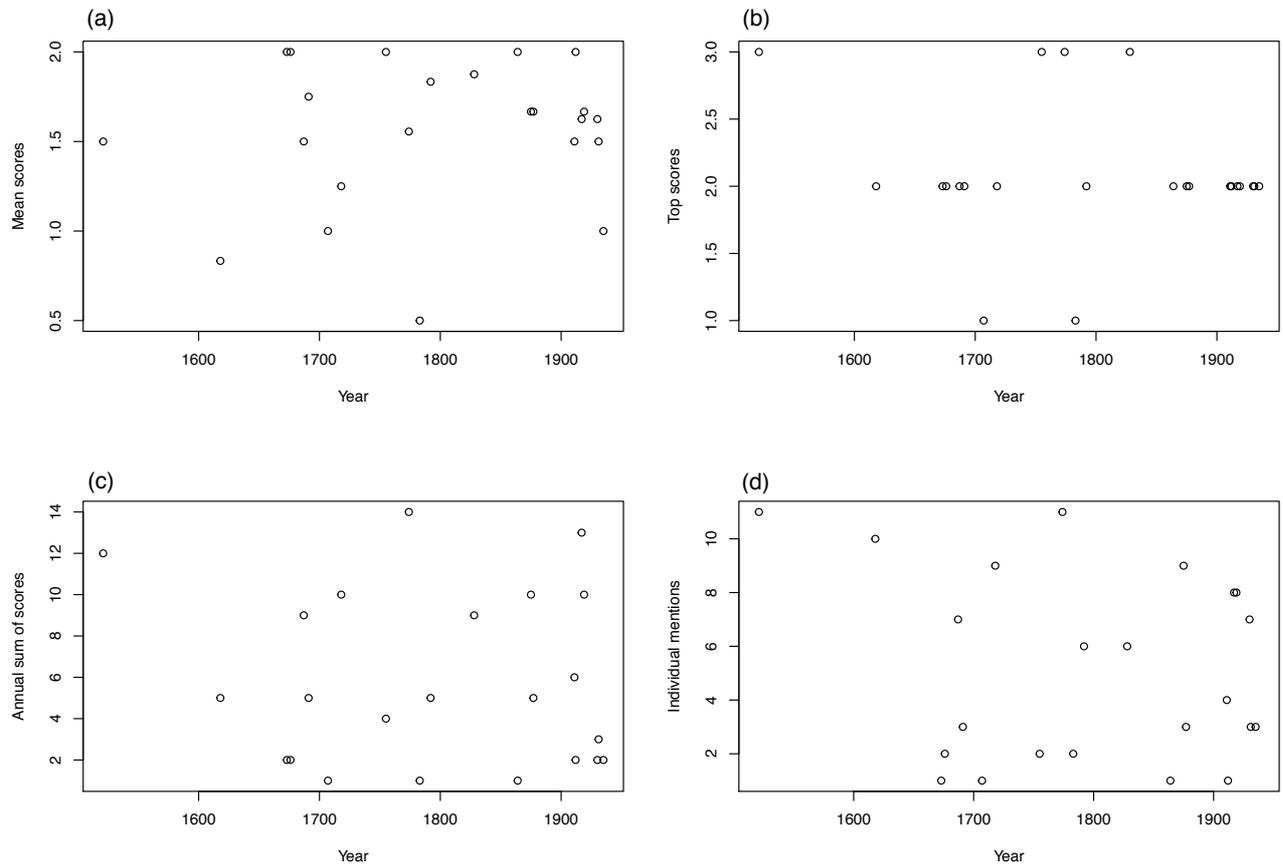
This is further confirmed by Mann-Kendall trend analysis which indicates all the samples show a positive trend (Table 1) for both gazetteer publication dates and sightings not just of gibbons but of all the study-species.

Table 1: Mann-Kendall trend analysis showing Kendall's tau, test statistic S and p-values rejecting the null hypothesis of no trend in the records over time.

	p-value	Kendall's tau	S
Gibbons: mean annual scores	0.00198	0.039	3407
Gibbons: count of annual entries scored 2 or 3	0.00348	0.035	3082
Gazetteers publication	< 2.22e-16	0.521	90832
Bear	0.00102	0.039	3372
Boar	0.000148	0.045	3896
Macaque	0.00124	0.039	3399
Pangolin	0.00336	0.036	3158
Clouded leopard	0.0146	0.033	2856
Dhole	0.0182	0.023	2022
Wolf	0.0153	0.02	1736

In each case the data show a positive trend, with more records in more recent times.

The OLE with 95% confidence intervals (CI) for the gibbons according to: mean score; scored 2 or 3; and summed score; and with the same CI for other species all scored equivalent to 2 and 3 (Table 2), shows the predicted survival of all species past the final gazetteer record, i.e. 1935.

Table 2: The Optimal Linear Estimation (OLE) of the year a species went extinct, with 95 % CI, and all species other than gibbons scored equivalent to 2 and 3.

Species	Estimated extinction	Lower CI	Upper CI	Number of sightings
Gibbons, by mean annual score:	1938	1935	1961	23
Gibbons, scored 3[*]	1898	1829	5260	4
Gibbons, scored 2	1938	1935	1961	16
Gibbons, scored 2 or 3	1938	1935	1957	20
Gibbons, sum of scores	1938	1935	1958	23
Gibbons, scored 1, 2 or 3	1937	1935	1950	26
Bear	1938	1935	1961	20
Boar	1936	1935	1948	23
Pangolin	1938	1935	1959	22
Macaque	1937	1935	1956	22
Clouded leopard	1938	1935	1959	27
Dhole	1942	1935	1999	14
Wolf	1953	1932	2062	9

---

[*] Only 4 records were scored 3, making it inadequate for reliable OLE, which requires at least 5 entries. It is included here just for completeness.

Discussion

This study suggests the Hainan gazetteer can be used as a robust source towards developing historical baselines of the mammals found on Hainan. Thus validated, the gazetteers also suggest the presence of two previously unrecorded large carnivores, wolf and dhole, into the 20th century. This may have implications on how the ecosystem of the island is understood, which in turn may be relevant to Hainan gibbon conservation action.

A statistical approach to validating such historical sightings often has to work with very limited data, given, in the case of Hainan gibbons, no reliable population studies for decades after the last gazetteer record (Turvey at al 2015a). A reference point may come from non-statistical data. Observations today can help infer historical presence: if a species is not extinct on an island today, it was likely not yet extinct in 1935. I compared the gazetteer dataset with tabulated strengths and weaknesses of models to estimate extinction (Boakes et al 2015). OLE is a useful tool to assess extinctions, be they global or local, from just time-series sightings data (Collen et al 2010; Boakes et al 2015). It is a probabilistic method, which estimates a year of extinction as a weighted-sum of the sighting times

(Solow 2005) based on a Weibull distribution, a two-parameter model from weakest-link analysis (Crawley 2005). Accuracy improves with the number of records, requiring at least 5 and preferably more than 10 (Collen & Turvey 2009; Boakes et al 2015). In this dataset there were 387 years between 1521 and 1935 with no publications, which could be seen as times when the observation effort was zero. Larger gaps between sightings lead to more conservative extinction estimates, however the method likewise does not assume that observation effort is constant, just remaining above zero (Collen et al 2010; Boakes et al 2015). Boakes et al (2015) say this is largely unimportant provided the sightings are comparatively rare. Given the information available, OLE appears to be an appropriate tool to assess the gazetteers.

The OLE analysis indicates persistence for species beyond the end of the gazetteer record, with extinction estimates of 1936 to 1953, and upper CIs of at least 12 years after the last sighting. This includes the still extant gibbon, bear, and pangolin (Lau et al 2010; Turvey 2015a; Nash et al 2016), suggesting the gazetteers are a reliable source within their own timeframe. Species that appear in the records but are known not to have existed on historical Hainan, such as tiger (*hǔ*), rhino (*sì*), and water monster (*yù*)

(Schafer 1970, Lau et al 2010) are filtered out by the qualitative scoring process, thus species that I put through OLE are all more plausibly present at last sighting than not. This adds validation to the findings of previous studies which have been based on the gazetteer record (Chatterjee et al 2012; Turvey et al 2015b).

The formulaic nature of the gazetteer as a genre provides little scope for agency on the part of the individual observer (Pooley 2013): the official who writes the account completes the same 'produce' section of the 'geography' chapter referencing the same standard works as his predecessors did. It might be hoped that widening the search to include other chapters on literature, biography, and trade, would escape this and perhaps include fresh detail. In the Hainan gazetteers, this is only partially the case, because while they did enlarge the dataset, none of my records from outside the geography or territory chapters scored above 2.

The Mann-Kendall trend analysis shows more records in more recent years. I sought to overcome this by the qualitative method of scoring entries for reliability. Further qualitative approaches may bring greater data resolution. It would be possible to temporally map the broad trends

influencing the officials who wrote the gazetteers: political and educational norms change and may affect the scientific awareness of the authors, or their likelihood of quoting older sources, and thus influence the quality of individual records. Such a system may further overcome the effects of higher record densities in later years, and would contribute to the GIS and statistics-based Chinese primate extinction early-warning system called for by Pan et al (2016) by increasing the resolution of historical baselines. However such an inter-disciplinary process would be highly resource-intensive, and studies made in China may be subject to contemporary political concerns.

Improved spatial resolution may also be possible. The gazetteers are published for administrative units ranging from county, to island-wide, to province. My attempts to identify the historical boundaries met contradictory reports between different gazetteers and historical mapping services (CHGIS 2007). However there can be consistency within a given dynastic reign. This may provide some potential for further defining geographical spread, but over shorter temporal scales.

The greatest challenge to using the gazetteers to establish historical baselines is identifying species. In this study that

excludes 50 types of animal recorded but difficult to identify, such as the 16 rodent-like (*shǔ*). The deer of Hainan are a illustrative example: The gazetteers indicate 12 names for types of deer, plus variations for sex. Accounts will claim two terms as being alternative names for the same type, but not consistently. Some clarity comes in the form of western scientific studies from the 19th and 20th centuries (Swinhoe 1870; Ohtaishi & Gao 1990), but even so this requires more work with additional sources before drawing taxonomic lines.

The apparent robustness of the gazetteers for identifiable species has implications. Wolf and dhole are recorded and apparently identifiable, despite neither being scientifically confirmed on Hainan (Swinhoe 1870; Lau et al 2010). On a global scale, that a large mammal carnivore should be extirpated from an area in the 20th century is on-trend: large mammals are twice as likely to be threatened as smaller, while in China the range of tigers has shrunk 95% (Macdonald et al 2013). With its large scale land-use change (Zhang et al 2010), Hainan can conceivably follow the same pattern.

The gazetteers record wolf and dhole 9 and 14 times respectively. The OLE results suggest that wolf and dhole may have persisted on Hainan into the 20th century.

The more questionable is wolf, first recorded in 1707, continuing until 1931. However only a standard description ever appears, and this is from a dictionary complied in 121 CE (Duan 1997; Wilkinson 2015). No mention is made of the typical wolf behavior of hunting in packs (Macdonald 2009), while some of the deer references do describe large herds, so this behavior is among the characteristics that the authors could have noted. A 1774 entry in the gazetteers specifically states that there are no wolves on Hainan—but the very same edition also contains a record of wolf presence, as do 8 other editions. This presents an opportunity for further qualitative research looking at the overall and zoological reliability of these 9 editions, an exercise which could further test OLE as a method of assessing gazetteer robustness. Including Hainan in the historical wolf distribution maps would mean extending the range south by about 800 km (Mech & Boitani 2010).

Dhole is recorded 14 times, across almost the entire time-span of the gazetteer collection. More than half the records share 2 descriptions: 5 say '*Like a dog but with a longer tail*' while 2 go into more detail: '*A type of wolf, deep fur and dog-like feet, a long tail, pale cheeks, colored yellow/brown*'. Published in 1875 and 1931 these come after Swinhoe (1870) dismissed the likelihood of dhole on the island. The

historical range maps of dhole just miss out Hainan, reaching the nearby landmass of Vietnam (Kamler et al 2015), and the presence of dhole on Hainan into the 20th century seems likely.

The ecosystem may provide clues as to whether dhole, wolf, or both were on Hainan within the past 100 years. Both would have had prey in the form of the island's deer and small mammals (Hayward et al 2014). This addition of two large carnivores to the confirmed bears and clouded leopard would likely have implications for the ecosystems that were in place on Hainan in the early 20th century (Ripple et al 2014). There may be need for a radical re-think of habitat restoration targets on Hainan, including possible trophic knock-on effects for Hainan gibbon food trees, due to carnivores controlling the herbivore population and dispersal (Marris 2014; Ripple 2014; Turvey et al 2015a). There are nearly as many written sources of Chinese history from 1911 to 1949 as there are in the previous 3 millennia, including archives and newspapers, in both the People's Republic of China and Taiwan (Wilkinson 2015). Possible 20th century sources to examine include local newspapers and the accounts of communist fighters, who were based in the mountains before 1949 and may have

encountered cryptic species (Murray 2011). The gazetteers may again inform ecology research because as well as the mammals in this study, they also list plants, trees, birds and fish over the centuries, and at least one account of the wild animal pelts found at markets run by the indigenous *Lí* people. Hunting by humans is reported to exert an influence in ecosystem cascades involving large carnivores such as wolves and bears (Ripple et al 2014), and ungulates.

Taken directly, the Hainan gazetteers provide little more than sightings data. However, this study shows that they can be a robust record of the historical presence of animals, and can be used to inform conservation management planning. At least 2 of the secondary study species here, wolf and dhole, are not reported on Hainan today. The main study species, the Hainan gibbon, is the rarest ape in the world. With fewer than 30 remaining it is an extreme case but with anthropogenic pressures, including climate change, increasing, it is unlikely to be the last example of a species facing extinction in Hainan or China. As such including the gazetteer record to help establish a historical baseline is a meaningful process. In the case of Hainan it is valid, and throughout China both likely to be needed and worth repeating for other species.

Supporting Information description

Chinese characters and transliteration (Appendix S1), Digitization workflow (Appendix S2) Raw data: gibbons (Appendix S3), and Raw data: other species (Appendix S4), are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature cited

Anthony L. 2014. AntConc. Waseda University, Tokyo, Japan.

Boakes EH, Rout TM, and Collen B. 2015. Inferring species extinction: the use of sighting records. Methods in Ecology and Evolution **6**(6)**:**678-687.

Bol PK. 2001. The Rise of Local History: History, Geography, and Culture in Southern Song and Yuan Wuzhou. Harvard Journal of Asiatic Studies **61**(1)**:**37-76.

Chatterjee HJ, Tse JSY, and Turvey ST. 2012. Using Ecological Niche Modelling to Predict Spatial and Temporal Distribution Patterns in Chinese Gibbons: Lessons from the Present and the Past. Folia Primatologica **83**(2)**:**85-99.

CHGIS January 2007. China Historical GIS. Harvard Yenching Institute, Cambridge, MA. Available from http://www.fas.harvard.edu/~chgis/data/chgis/downloads/v4/ 2016.

Cihai: hedingben. 1948. 2nd edition. Zhonghua shuju, Shanghai.

Clements C. 2013. sExtinct. 1.1st edition.

Collen B, and Turvey ST. 2009. Probabilistic methods for determining extinction chronologies. Pages 181-191 in ST Turvey, editor. Holocene Extinctions. Oxford University Press, Oxford.

Collen B, Purvis A, and Mace GM. 2010. BIODIVERSITY RESEARCH: When is a species really extinct? Testing extinction inference from a sighting record to inform conservation assessment. Diversity and Distributions **16**(5)**:**755-764.

Crawley MJ. 2005. Statistics: an introduction using R. John Wiley, Chichester.

Davies AL, Colombo S, and Hanley N. 2014. Improving the application of long-term ecology in conservation and land management. Journal of Applied Ecology **51**(1)**:**63-70.

Duan Y. 1997. Shuowen jiezi zhu=Annotated Shuowen jiezi. 1st edition. Jiangsu guangling gujikeyinshe, Yangzhou.

Edmonds RL. 1999. The Environment in the People's Republic of China 50 Years On. The China Quarterly **159:**640-649.

Fleck DW, Voss RS, and Patton JL. 1999. Biological Basis of Saki (Pithecia) Folk Species Recognized by the Matses Indians of Amazonian Perú. International Journal of Primatology **20**(6)**:**1005-1028.

Guhanyu changyongzi zidian bianxiezu. 1998. Guhanyu changyongzi zidian: 1998 nian ban. Shangwu yinshuguan, Beijing.

Gulik RHv, 1910-1967. 1967. The gibbon in China: an essay in Chinese animal lore.

Guo P. 1981. Erya yintu. 1st edition. Guangwen shuju, Taibei.

Hanyu dacidian. 1986. Di 1 ban. edition. Shanghai cishu chubanshe, Shanghai.

Hayward MW, Lyngdoh S, and Habib B. 2014. Diet and prey preferences of dholes (Cuon alpinus): dietary competition within Asia's apex predator guild. Journal of zoology **294**(4)**:**255-266.

Hong S., and Hainan difang wenxian congshu bianweihui. 2003. Hainan difangzhi congkan = Collected Hainan Gazetteers. Hainan chubanshe, Haikou.

Kamler JF, Songsasen N, Jenks K, Srivathsa A, and Sheng, L. & Kunkel, K. 2015. Cuon alpinus. The IUCN Red List of Threatened Species **2015** (e.T5953A72477893)**:**15 August 2016.

Lau MW, Fellowes JR, and Chan BPL. 2010. Carnivores (Mammalia: Carnivora) in South China: a status review with notes on the commercial trade. Mammal Review **40**(4)**:**247-292.

Li S. 1957. Bencao gangmu. 1st edition. Renmin weisheng chubanshe, Beijing.

Liu J, and Diamond J. 2005. China's environment in a globalizing world. Nature **435**(7046)**:**1179-1186.

Macdonald DW, Boitani L, Dinerstein E, Fritz H, and Wrangham R. 2013. Conserving large mammals: are they a special case? Pages 276-312 in D. W. Macdonald and K. J. Willis, editors. Key Topics in Conservation Biology 2. John Wiley and Sons Ltd, Chicester.

Macdonald DW. 2009. The encyclopedia of mammals. New edition. Oxford University Press, Oxford.

Marris E. 2014. Rethinking predators: Legend of the wolf. Nature **507**(7491)**:**158-60.

Mech LD, and Boitani L. 2010. Canis lupus. The IUCN Red List of Threatened Species 2010 **2010**(e.T3746A10049204)**:**Downloaded on 15 August 2016. http://dx.doi.org/10.2305/IUCN.UK.2010-4.RLTS.T3746A10049204.en.1.

Murray JA. 2011. Culturing revolution: the local Communists of China's Hainan Island. UC San Diego.

Nash HC, Wong MHG, and Turvey ST. 2016. Using local ecological knowledge to determine status and threats of the Critically Endangered Chinese pangolin (Manis pentadactyla) in Hainan, China. Biological Conservation **196:**189-195.

Ohtaishi N, and Gao Y. 1990. A review of the distribution of all species of deer (Tragulidae, Moschidae and Cervidae) in China. Mammal Review **20**(2-3)**:**125-144.

Pan R, Oxnard C, Grueter CC, Li B, Qi X, He G, Guo S, and Garber PA. 2016. A new conservation strategy for China—A model starting with primates. American Journal of Primatology.

Pauly D. 1995. Anecdotes and the shifting baseline syndrome of fisheries. Trends in Ecology & Evolution **10**(10)**:**430 http://dx.doi.org/10.1016/S0169-5347(00)89171-51.

Pooley S. 2013. Historians are from Venus, Ecologists are from Mars. Conservation Biology **27**(6)**:**1481-1483.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rick TC, and Lockwood R. 2013. Integrating Paleobiology, Archeology, and History to Inform Biological Conservation. Conservation Biology **27**(1)**:**45-54.

Ripple WJ, Estes JA, Beschta RL, Wilmers CC, Ritchie EG, Hebblewhite M, et al. 2014. Status and Ecological Effects of the World's Largest Carnivores. Science **343**(6167)**:** 10.1126/science.12414841.

Roberts DL, and Solow AR. 2003. Flightless birds: When did the dodo become extinct? Nature **426**(6964)**:**245-245.

Schafer EH. 1970. Shore of pearls. Univ. of California Press, Berkeley and Los Angeles.

Schipper J, Chanson JS, Chiozza F, Cox NA, Hoffmann M, Katariya V, et al. 2008. The Status of the World's Land and Marine Mammals: Diversity, Threat, and Knowledge. Science **322**(5899)**:**225-230 10.1126/science.11651151.

Smith AT, and Xie Y, editors. 2008. A guide to the mammals of China. Princeton University Press, Princeton, N.J.

Solow AR. 2005. Inferring extinction from a sighting record. Mathematical Biosciences (195)**:**47-55.

Swinhoe R. 1870. The natural history of Hainan. Horace Cox, London.

Szabó P. 2010. Why history matters in ecology: an interdisciplinary perspective. Environmental Conservation **37**(04)**:**380-387.

Thorsten P. 2016. trend: Non-Parametric Trend Tests and Change-Point Detection.

Turvey, ST, Traylor-Holzer, K, Wong, MHG., Bryant, JV, Zeng, XY, Hong, XJ, Long, YC. 2015a. International Conservation Planning Workshop for the Hainan Gibbon: Final Report. Zoological Society of London/IUCN SSC Conservation Breeding Specialist Group, London & Apple Valley, MN.

Turvey S. 2009. Witness to extinction: how we failed to save the Yangtze River dolphin. Oxford University Press, Oxford.

Turvey ST, Crees JJ, and Di Fonzo MMI. 2015b. Historical data as a baseline for conservation: reconstructing long-term faunal extinction dynamics in Late Imperial–modern China. Proceedings of the Royal Society of London B: Biological Sciences **282**(1813)**:** 10.1098/rspb.2015.12991.

Wen R. 2013. Zhongguo gudai yeshengdongwu dili fenbu = Geographical distribution of wild animals in ancient China.

Wen R. 2009. Zhongguo zhenxi yeshengdongwu fenbubianqian = Changes in distribution of China's endangered wild animals.

Wilkinson EP. 2015. Chinese history: a new manual. Harvard University Asia Center, Harvard University Press, Cambridge, Mass.

Zhang M, Fellowes JR, Jiang X, Wang W, Chan BPL, Ren G, and Zhu J. 2010. Degradation of tropical forest in Hainan, China, 1991–2008: Conservation implications for Hainan Gibbon (Nomascus hainanus). Biological Conservation **143**(6)**:**1397-1404.

Appendix S1. Chinese characters and transliteration

This paper shows simplified Chinese characters, as these were used in the received version of the base text. Transliterations into the Roman alphabet are in Hanyu Pinyin. Both these are also the ISO standards.

Glossary:

Hainan 海南 *Hǎinán*

Yangtze river dolphin or baiji 白暨 *báijì*

Local gazetteer 地方志 *dìfāngzhì*

Ape, gibbon 猿 *yuán* and 猨 *yuán*

Ape (archaic) 猩猩 *xīngxing*; 狒狒 *fèifei*

Beast, approximating mammal: 兽 *shòu*

Civets etc. 狸 *lí*

Rodents etc.鼠 *shǔ*

Bears:人熊 *rénxiong*; 马熊 *mǎxióng*; 狗熊 *gǒuxióng*; 猪熊 *zhūxióng;* generic 熊 *xióng*

Boar 山猪 *shānzhū*，野猪 *yězhū*

Pangolin 穿山甲 *chuānshānjiǎ*, 鲮鲤 *línglǐ*

Monkey 猴 *hóu*

Macaque: 猕 *mí*

Leopard: 豹 *bào*

Clouded leopard: 金钱豹; *jīnqiánbào* 艾豹 *àibào*

Dhole: 豺 *chái*

Wolf 狼 *láng*

Geography 地理志 *dìlǐzhì*

Territory 舆地志 *yúdìzhì*

Produce 物产 *wùchǎn*

Yu water monster 蜮 *yù*

Li (ethnic group) 黎 Lí

Tiger 虎 *hǔ*

Rhino 兕 *sì*

Quoted passages:

In addition, Qiongzhou [Hainan] has stone apes, the small ones the size of a fist, they grow when fed water from a well. Also [known as] black apes, they can rub ink [from an inkstone], jumping into the ink pot when finished. 琼州又有石猨，小者拳许，饮以井水即长。又墨猨能磨墨，磨毕跳入笔筒中。

Gibbons: there are three types: Golden ones are yellow, jade-faced are black, and there are also pure black ones. The golden and jade-faced ones are all rare. 猿 有三种。金丝者黄，玉面者黑，又有纯黑者。金丝、玉面皆难得。

Gibbons: the male is black and the female pale. They are like monkeys but their arms are very long, climbing and grabbing branches in the canopy with great agility. If they ever fall to the ground they stiffen up like a tree. As they are good at climbing trees but cannot walk, they need to stay in the trees to be raised, as if they get near the earth vapors they fall ill and die. 猿雄黑雌白，似猴而两臂甚长，攀树援岭往来甚捷，一堕平地，则木强矣。以善援木而不能走也，畜之亦须置树间，近土气即病而死

Like a dog but with a longer tail 豺似狗而长尾

A type of wolf, deep fur and dog-like feet, a long tail, pale cheeks, colored yellow/brown 豺狼属，深毛而狗足，长尾，白颊，色黄

Chinese literature cited:

洪寿祥 Hong Shouxiang. 2003. 海南地方志丛刊 *Hainan difangzhi congkan*. 海南地方志丛刊 Hainan difangzhi congkan. 1st ed. Haikou: 海南出版社 Hainan Chubanshe.

辞海 *Cihai: Hedingben*.1948. ed. Zhonghua shuju. 2nd ed. ed. Shanghai: Zhonghua shuju.

古汉语常用字字典编写组 Guhanyu changyongzi zidian bianxiezu., 1998. 古汉语常用字字典 *Guhanyu changyongzi zidian : 1998 nianban*. 北京 Beijing: 商务印书馆 Shangwu yinshuguan.

汉语大词典 *Hanyu dacidian* 1986. eds. 汉语大词典编辑委员会；主编罗竹风 Zhufeng Luo, Hanyu dacidian bianji weiyuanhui. and 汉语大词典编纂处 Hanyu dacidian bianzuanchu. Di 1 ban. ed. 上海 Shanghai: 上海辞书出版社 Shanghai cishu chubanshe.

文榕生. 2009. 中国珍稀野生动物分布变迁. 山东科学技术出版社, Shandong.

Appendix S2. Digitization workflow

In 2003 the Hainan Publishing house compiled all extant Hainan gazetteers, editing them into a published collection. However they did so only in a 68 volume printed edition, not making it available digitally. An unofficial ebook version of many Chinese historical gazetteers is available from an antiquarian book seller on China's leading shopping website, Taobao.com. These are delivered on demand as DVDs containing scans of the paper book, supplied as individual images of each page, organized in folders equivalent to each volume. I had to make these searchable by running the 1.9 GB of image data through optical character recognition software FineReader Pro (ABBYY 2016): at this stage I found 23 volumes were corrupted and 6 missing, and discarded them as bad data. I found one missing volume online as a lower-resolution PDF, so it was included in the corpus. In this volume the OCR error rate is likely to be higher than for the others, which have higher image-quality. This left 39 volumes to work with.

The western or Gregorian calendar was introduced to China after publications in this dataset. I converted each volume's traditional dates to calendar years, based on the year of printing or the most recent revision, as listed in the Forward of each gazetteer. The Forwards may variously state the year the book was commissioned, revised, or printed. There is no way of telling for sure when between those years the animal sightings were actually made–or, indeed, whether they were seen at all.

Appendix S3. Raw data: gibbons

The spreadsheet recording gibbon mentions had the following column headings:
Id; Gazetteer name; hanzi county; county; hanzi date; date; Species accounts; Specific region; Area today; Original text (yuan); Translation; Referring to; txt file; PNG; Other species listed in adjacent text (or animal list if no yuan in volume); Rank; Other species text; yuanYN.

In the attached file some columns have been removed to accommodate the single PDF file submission requirement. Complete Excel spreadsheets are available from the author on request.

Appendix S4. Raw data: other species

These were collected in the paragraphs surrounding gibbon mentions, and in the 'beasts' sections. In rare cases these sections include non-mammal species.