Animal Population Estimation Using Flickr Images

Sreejith Menon
Bloomberg L.P.
New York, New York 10022
smenon59@bloomberg.net

Tanya Berger-Wolf
University of Illinois at Chicago
Chicago, Illinois 60607
tanyabw@uic.edu

Emre Kiciman
Microsoft Research
Redmond, Washington 98052
emrek@microsoft.com

Lucas Joppa
Microsoft Research
Redmond, Washington 98052
lujoppa@microsoft.com

Charles V. Stewart
Rensselaer Polytechnic Institute
Troy, New York 12180
stewart@cs.rpi.edu

Jason Parham
Rensselaer Polytechnic Institute
Troy, New York 12180
parhaj@rpi.edu

Jonathan Crall
Rensselaer Polytechnic Institute
Troy, New York 12180
crallj@rpi.edu

Jason Holmberg
WildMe.org
Portland, Oregon 97217
jason@wildme.org

Jonathan Van Oast
WildMe.org
Portland, Oregon 97217
jon@wildme.org

Abstract

While the technologies of the Information Age have produced staggering amounts of data about people, they are by and large failing the world's wildlife. Even the simplest and most critical piece of information, the number of animals of a species, is either unknown or is uncertain for most species. Here, we propose to use images of wildlife posted on social media platforms, together with animal recognition software and mark-recapture models, to estimate population sizes. We show that population size estimates from social media photographs of animals can produce robust results, yet more work is needed to understand biases inherent in the approach.

ACM Reference format:
Sreejith Menon, Tanya Berger-Wolf, Emre Kiciman, Lucas Joppa, Charles V. Stewart, Jason Parham, Jonathan Crall, Jason Holmberg, and Jonathan Van Oast. 2016. Animal Population Estimation Using Flickr Images. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 4 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 Introduction

Of the 8.7 Million terrestrial species which are estimated to exist on earth, humans have discovered and described a mere 14% [17]. The International Union for Conservation of Nature (IUCN) Red List of Threatened Species™[23] is an internationally recognized official body for tracking conservation status of species across the world [25]. IUCN currently maintains and tracks conservation status of over 79,000 species. Of those, over 23,000 species are threatened with extinction [13, 25]. The Living Planet report, the most comprehensive effort to track the population dynamics of species around the world, includes 10,300 populations of just 3,000 species [10]. Such scarce data make it exceptionally difficult to assess some of the most pressing problems in conservation - how healthy are wildlife populations, and are our conservation management actions having a positive effect? Answering those questions is impossible without leveraging computational tools - the operational and financial burdens of large scale censuses sharply limit their use. Today, one of the most abundant sources of information about wildlife are images, taken by scientists, trail cameras, field assistants, tourists, and opportunistic photographers. The latter two image sources are available only if the photographs are shared, directly with the scientists or publicly with the world. A promising solution to tracking wildlife populations at large spatio-temporal scales is to turn to an opportunistic form of citizen science: mining publicly available social media photos of animals.

The current era is marked by extensive use and influence of social media platforms in our day-to-day lives. The proportion of total world population active on social media has seen a sharp rise in the past five years and is likely to rise up to 2.67 billion by the year 2018 [8]. One of the primary activities on social media is sharing images; over 3 billion images are uploaded and shared on various social networks every day [14]. While most of those images do not pertain to wildlife, the amount that do still dwarfs the information available from standard scientific surveys. However, combining these new sources of data with traditional population and range models is not straight-forward. In this paper, we discuss how accurately can we estimate population of Grevy's zebra *(Equus grevyi)* from images that are scraped from Flickr[1] and demonstrate a proof of concept approach for using social media images for wildlife population estimates.

2 Methods

2.1 Estimating population of wildlife species

A standard method of estimating population of a wildlife species is using the Mark-Recapture approach [6, 22, 24]. Typically, it proceeds in two stages. First, some number of animals are captured and these animals are marked with some kind of unique identifiable marking. At the second stage, considered to be independent of the first, another set of animals from the same population is captured and of those some already have been marked in the first stage. The simplest mark-recapture approach assumes that the ratio of animals with initial markings to the number of animals captured in the second sample has to be proportional to number of animals

[1] https://www.flickr.com

Conference'17, Washington, DC, USA

DOI: 10.1145/nnnnnnn.nnnnnnn

captured in the first sample to the actual population of the species. Thus, let n be the number of individuals captured during the first stage (capture), K be the number of individuals captured during the second stage, and k be the number of individuals with the unique marking from first stage that were captured in the second stage (*i.e.,* the animals that were re-captured). Then the estimated population size, N_{est}, is computed using the simple ratio formula known as the Lincoln Petersen estimator [20].

$$N_{est} = \frac{Kn}{k}.$$

There are several major assumptions that this method makes. First, that every animal has an equal probability of being captured [15], that the sampling strategy is the same in the first and second stages and they are independent of each other, and, finally, that there are no new animals joining or animals leaving the population during the sampling period.

Animal population estimation using photographic data can be done using a very similar approach. The marks and recaptures can be transformed into sights and re-sights across different pictures [3, 12]. For the extent of our studies, we extend the same sight-resight model but in a social media setting. All the pictures that are known to be shared are divided into various time epochs and a population size can be estimated across any (preferably consecutive) pair of epochs. This method can be used not only for population size estimation for a given period but can be used to observe trends and dynamics of population changes over a longer period of time.

Note, however, that using the simple Lincoln Petersen estimator may not be appropriate as its assumptions may be violated by the social media image data. It is not guaranteed that the population is closed and, more importantly, that every animal has the same probability of being photographed (captured). However, more sophisticated models of mark-recapture, such as Jolly-Seber [15], require knowledge of more parameters of the sampling process and the biases herein than we currently possess of the social media image data. Thus, in absence of this information we use the simplest method and leave the estimation of the biases of the social media for future research (see Section 4).

2.2 Identifying individual animals using computer vision

A population estimation method for wildlife species requires an ability to detect and identify individual animals, in our case, in photographic data. HotSpotter [7] is a fast and accurate algorithm for identifying individual animals of species with unique markings on their bodies, from their pictures. HotSpotter algorithm can identify species such as Grevy's zebra, Plains zebra, Reticulated giraffe, Masai giraffe, Humpback Whales (using flukes), Bottlenose dolphins (using dorsal fins), Iberian lynx, Giant sea bass, Geometric tortoises, Hawksbill sea turtles, Giant mantas, Ragged tooth nurse sharks, and many others.

Wildbook™ [29] is an open source (GPL v2) software system that starts with images of animals and connects the image analysis with a data-management layer to enable queries about animals. Wildbook™ provides an interface for citizen scientists to report their sightings of animals and upload animal images. These images are passed through a detection algorithm where species of the animal, a bounding box of where the animal is located in the picture, and other aspects are determined. Every detection is assigned a unique annotation ID. Every annotation in an image is then passed through the HotSpotter algorithm where it is compared to already labelled annotations in the reference database to determine whether the animal was previously sighted or not. Previously unseen animals are assigned a new, unique ID, while previously sighted animals are reassigned their existing IDs. Access to the photographic detection and identification modules is exposed as REST APIs within Wildbook™ [21].

3 Experiment

3.1 Grevy's Zebra

For the purpose of our study we chose the species of the Grevy's zebra. The population of Grevy's zebra in the wild is limited to only some parts of Kenya and Ethiopia [26]. This gives us a unique opportunity to estimate global population of a species and, also, the chance to validate the accuracy of the results since the ground truth population of Grevy's zebra is known and is a closed population. In addition, the stripe pattern on Grevy's zebra provides the unique marking which allows HotSpotter to uniquely identify individuals.

The ground truth population size, as well as the images for the reference database of the existing population of Grevy's zebra, come from the massive citizen science event, The Great Grevy's Rally [1, 21], which took place in central and northern Kenya in January 2016. The Great Grevy's Rally spanned over two days, about 160 people participated in the rally and over 40,000 images were captured in the event. Using HotSpotter 1,942 individuals were uniquely identified and named. The population of Grevy's zebra was then calculated using standard Lincoln-Petersen estimator dividing the animal sights and resights across day 1 and day 2 of the rally. These estimates are also used as the official population size estimates by the IUCN Redlist [26].

We use the population estimation stated on the IUCN's Redlist as the ground truth population against which our estimates are compared. The ground truth population as of year 2016 is 2250 ± 93 [26]. The number is from January 2016 and previous population size estimates indicate higher numbers in years prior to 2014 (albeit with a much lower confidence level).

3.2 Data

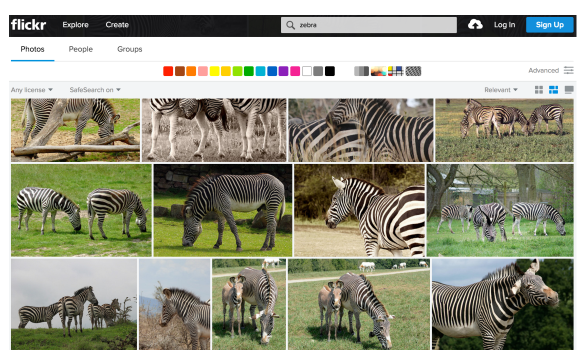

Figure 1. Snapshot of images results returned by searching using the search term "zebra" on www.flickr.com

Using Flickr's image search API's [9], we scraped publicly shared images that were tagged as "zebra". These images were then uploaded to the HotSpotter instance and then detections and identifications were performed on these images. Out of all the images that were obtained, we used the images that had at least one Grevy's zebra in it and then used those images to estimate the population size of the species. Thus, 1,701 images taken between 2004—2015 were scraped from Flickr. Both Plains zebra and Grevy's zebra images were found in this sample. A total of 2,047 annotations were detected and 1,080 unique individuals were identified and named. For the images scraped from Flickr, we also extract spatio-temporal information about every image, *i.e.* when and where each image was taken. Only images of wild Grevy's zebras (not zoo) taken in Kenya were used for analysis. The images were then divided into sets on the basis of the year in which the image was taken. Table 1 summarizes the images that were obtained from scraping the publicly shared albums on Flickr.

Search query	zebra
Number of images	1701
Number of annotations	2047
Species detected	Grevy's zebra, Plains zebra
Individuals identified	1080
Date range	2004-2015

Table 1. Search query, total number of images, annotations, species detected and individuals identified from the images scraped from Flickr.

The population was then estimated using Lincoln-Petersen estimator by assigning the images taken in a particular year as marks and the images taken on the subsequent year as recaptures. This process is then repeated for all the years starting from 2004 to 2015.

3.3 Analysis details

Population estimate of Grevy's zebra does not take into account animals in zoos. This is primarily because of the higher likelihood of zoo animals to be photographed in comparison to the animals in the wild. A higher likelihood of capture means a higher recapture rate. A higher recapture rate in the mark-recapture model indicates lower population size. Thus, inclusion of zoo animals in the corpus of images used for population estimation will lead to an underestimate of the population size. Therefore we removed all images of zoo animals from the corpus.

Let P be the ground truth population, $\hat{P}$ be the estimated population and ϵ be the percent estimation error or simply, error. We assume the value of P to be equal to the upper limit of known ground truth population to simplify our calculations. We then compute the error in population estimate ϵ using the below formula.

$$\epsilon = \frac{\hat{P} - P}{P} \times 100.$$

3.4 Results

We computed population size estimates from Flickr images from years 2004—2015, using sight-resight estimates on pairs of consecutive years. Table 2 summarizes the results and the error *against the Januray 2016* population size estimate. In the years 2014 and 2015, the estimate is very accurate, with 3.8 and 0.9 percent error, respectively. The population of the Grevy's zebra is considered to have been roughly stable over the last 10 years [26].

Sight year	Resight Year	Population Estimate	Error
2004	2005	No resights	N/A
2005	2006	No resights	N/A
2006	2007	2035	-9.5%
2007	2008	1842	-18.13%
2008	2009	1521	-32.4%
2009	2010	2782	23.64%
2010	2011	6580	192.4%
2011	2012	3624	61.1%
2012	2013	4447	97.6%
2013	2014	2336	3.8%
2014	2015	2272	0.9%

Table 2. Population size estimates calculated annually using simple sight-resight model for Grevy's zebra using Flickr images.

4 Discussion and Future Work

The results of our experiment and analysis, as summarized in Table 2, provide evidence that population estimation from social media images is a viable and reasonable approach. While there are errors in some years, we have a good estimate from the pictures that were taken in the years 2014 and 2015 and the estimates for prior years may, too, be accurate when compared to the actual population numbers in those years. The census event for Grevy's zebra took place in the January of year 2016, thus giving us a good benchmark against which we can compare our estimates for the years 2014 and 2015.

The next steps in our research involves repeating the experiments and estimating population of more species. In principle, for every species that can be detected and identified using Wildbook™ image analysis tools, we can estimate the population of that species using images from social media. While Flickr has been an important source for all publicly shared images, we are currently extracting images from Bing and exploring options to scrape images from other social media platforms such as Twitter, Facebook, and others.

While estimating population of a species from social media directly, we completely neglected bias in any form that influences the type of image samples we scraped from the Flickr database. There are multiple biases that influences the final outcome of estimating population of a certain species from images that are obtained from social media. Some of the most prominent biases which influence the data we obtain from social media are outlined in Figure 2. There are several layers of biases, accumulating in the resulting bias of estimating animal population properties from images. First, there are biases in the types of animals that people typically photograph in sufficient numbers in the first place. These may be charismatic or endangered species, or simply the ones easily observed. Second, there are biases in what images people take versus which ones they decide to share publicly on social media. These range from the Hawthorne Effect [2, 19, 27, 28] of changing behavior when knowing to be observed, to biases introduced by the demographics of the person sharing [4, 19] and the choice of the social media platform [5, 11, 18]. There are biases of our notions of beauty and aesthetics and cultural differences. Any mark-recapture model used to estimate the population size makes many assumptions and introduces its own biases. The fundamental question, however, is: *Do any of these actually affect the estimates of the population size and*

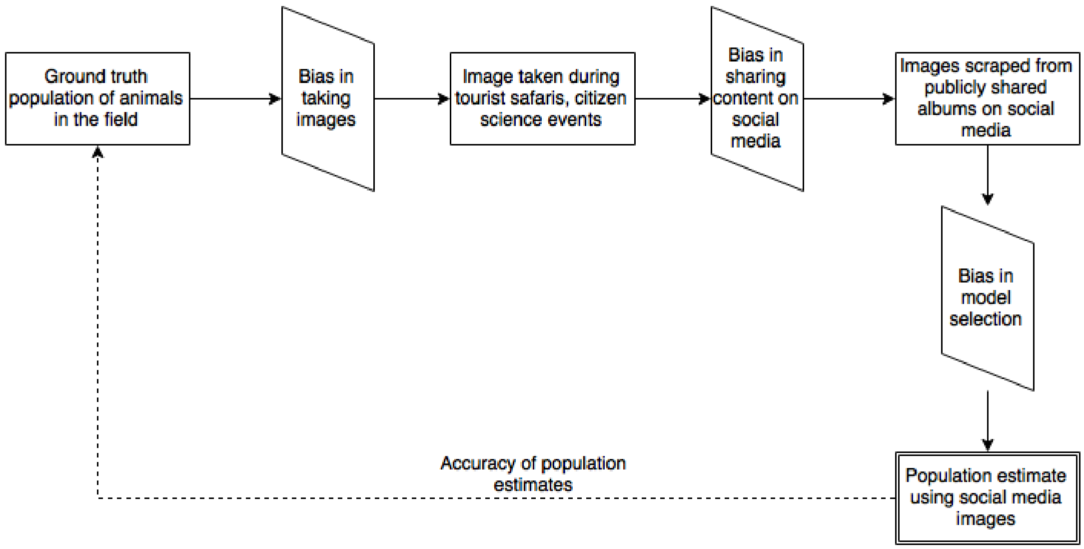

Figure 2. High level schematic representation of the problem of population estimation of wildlife species using images from social media, its challenges and biases in play.

other parameters and if so, how? Menon has begun to answering this question [16] but a lot of work remains to be done.

Once we find out a way to quantify some of these biases and investigate the effects of biases on a particular species of animals, we can build an accurate population estimator and adjust the results of the estimator for the known biases. Moreover, social media data can serve as an augmenting source to more traditional data sources or provide the first baseline where none exists.

References

[1] T. Berger-Wolf, J. Crall, J. Holmberg, J. Parham, C.V. Stewart, B. Low Mackey, P. Kahumbu, and D. Rubenstein. 2016. *The Great Grevy's Rally: The Need, Methods, Findings, Implications and Next Steps.* Technical Report. Technical Report, Grevy's Zebra Trust, Nairobi, Kenya.

[2] M. S. Bernstein, A. Monroy-Hernández, D. Harry, P. André, K. Panovich, and G. G. Vargas. 2011. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community.. In *ICWSM*. 50–57.

[3] D. T. Bolger, T. A. Morrison, B. Vance, D. Lee, and H. Farid. 2012. A computer-assisted system for photographic mark–recapture analysis. *Methods in Ecology and Evolution* 3, 5 (2012), 813–822. DOI: http://dx.doi.org/10.1111/j.2041-210X.2012.00212.x

[4] J. Brenner and M. Duggan. 2013. The demographics of social media users. *Consultado en* (2013).

[5] A. Bruns and S. Stieglitz. 2014. Twitter data: what do they represent? *it-Information Technology* 56, 5 (2014), 240–245.

[6] M. Chap and M. Chap. 1975. Fallow Deer, their history, distribution and biology. (1975).

[7] J. Crall, C. Stewart, T. Berger-Wolf, D. Rubenstein, and S. Sundaresan. 2013. Hotspotter - patterned species instant recognition. *Applications of Computer Vision (WACV)* (2013), 230–237.

[8] eMarketer; Statista. 2017. Number of social media users worldwide from 2010 to 2020 (in billions). (2017). https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ Accessed May 1, 2017.

[9] Flickr. 2005. The App Garden. flickr.photos.search. (2005). https://www.flickr.com/services/api/flickr.photos.search.html Accessed May 14, 2017.

[10] World Wide Fund for Nature. 2016. The Living Planet Report. (2016). http://wwf.panda.org/about_our_earth/all_publications/lpr_2016/ Accessed May 1, 2017.

[11] S. Gonzalez-Bailon, N. Wang, A. Rivero, J. Borge-Holthoefer, and Y. Moreno. 2014. Assessing the bias in samples of large online networks. *Social Networks* 38 (2014), 16–27.

[12] L. Hiby, W. D. Paterson, P. Redman, J. Watkins, S. D. Twiss, and P. Pomeroy. 2013. Analysis of photo-id data allowing for missed matches and individuals identified from opposite sides. *Methods in Ecology and Evolution* 4, 3 (2013), 252–259. DOI: http://dx.doi.org/10.1111/2041-210x.12008

[13] C. Hilton-Taylor. 2000. *2000 IUCN red list of threatened species.* IUCN.

[14] Kleiner Perkins Caufield Byers (KPCB). 2016. Internet Triends 2016 - Code Conference. (2016). http://www.kpcb.com/internet-trends Accessed May 1, 2017.

[15] C.J. Krebs. 1999. *Ecological Methodology.* Benjamin/Cummings. https://books.google.com/books?id=1GwVAQAAIAAJ

[16] S. Menon. 2017. *Animal Wildlife Population Estimation Using Social Media Images.* Master's thesis. University of Illinois at Chicago, Chicago, IL.

[17] C. Mora, D. P. Tittensor, S. Adl, A. G. Simpson, and B. Worm. 2011. How many species are there on Earth and in the ocean? *PLoS Biol* 9, 8 (2011), e1001127.

[18] F. Morstatter, J. Pfeffer, and H. Liu. 2014. When is it biased?: assessing the representativeness of twitter's streaming API. In *Proceedings of the 23rd International Conference on World Wide Web.* ACM, 555–556.

[19] A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman. 2016. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. (2016), 1–44.

[20] S. W. Pacala and J. Roughgarden. 1985. Population experiments with the Anolis lizards of St. Maarten and St. Eustatius. *Ecology* 66, 1 (1985), 129–141.

[21] J. Parham, J. Crall, C. Stewart, T. Berger-Wolf, and D. Rubenstein. 2017. Animal Population Censusing at Scale with Citizen Science and Photographic Identification.. In *The AAAI 2017 Spring Symposium on AI for Social Good (AISOC).*

[22] R. Pradel. 1996. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics* (1996), 703–709.

[23] IUCN Global Species Programme. 2017. The IUCN Red List of Threatened Species website. (2017). http://www.iucnredlist.org/ Accessed May 14, 2017.

[24] D. S. Robson and H. A. Regier. 1964. Sample Size in Petersen Mark–Recapture Experiments. *Transactions of the American Fisheries Society* 93, 3 (1964), 215–226. DOI: http://dx.doi.org/10.1577/1548-8659(1964)93[215:SSIPME]2.0.CO;2 arXiv:http://dx.doi.org/10.1577/1548-8659(1964)93[215:SSIPME]2.0.CO;2

[25] A. S. Rodrigues, J. D. Pilgrim, J. F. Lamoreux, M. Hoffmann, and T. M. Brooks. 2006. The value of the IUCN Red List for conservation. *Trends in ecology & evolution* 21, 2 (2006), 71–76.

[26] D. Rubenstein, B. Low Mackey, Z.D. Davidson, F. Kebede, and S.R.B. King. 2016. *Equus grevyi.* The IUCN Red List of Threatened Species. e.T7950A89624491 (2016). http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T7950A89624491.en

[27] S. Y. Schoenebeck. 2013. The Secret Life of Online Moms: Anonymity and Disinhibition on YouBeMom. com.. In *ICWSM.* Citeseer.

[28] M. Shelton, K. Lo, and B. Nardi. 2015. Online media forums as separate social lives: A qualitative study of disclosure within and beyond reddit. *iConference 2015 Proceedings* (2015).

[29] WildMe. 2016. Wildbook. (2016). http://www.wildbook.org Accessed May 1, 2017.