

Is the visual data collected and processed by citizen scientists scientifically valid for research?

Michelle Medina

Miami University

GFP EE Baja 1: Summer 2016

Introduction

Camera traps have been shown to be invaluable for capturing data about wildlife presence and behavior, especially cryptic, nocturnal, or solitary animals and those that inhabit difficult to reach areas. The image data obtained from camera traps can be utilized to determine distribution, population estimates, activity patterns, intra-community interactions and monitoring (Trolliet, Huynen, Vermeulen, & Hambuckers, 2014). For those animals, which present individualistic and visually identifiable marks, scientists can follow individuals via mark-recapture techniques to study dispersal, home ranges and behaviors over their lifespan.

With the use of camera traps/trail cameras and digital imaging for wildlife studies growing exponentially over the last twenty years, immense amounts of photographic data has been collected. Large-scale projects can easily accumulate millions of photographs and accompanying data, which translates to terabytes of digital data. Much of this data can take months, or even years, for researchers to process. As a result, some researchers are utilizing web-based databases to collect the data, and then process it using the power of citizen scientists.

Snapshot Serengeti is a very popular and successful citizen science web-based project, in which volunteers processed 1.2 million sets of camera trap collected images in three years, by identifying species, recording the presence or absence, and counting of animals (Swanson et al., 2015). The use of citizen scientists to process photographic images can be successful because many find the activity to be rewarding (McShea, Forrester, Costello, He, & Kays, 2016) and share in the excitement of discovery (Sheil, Mugerwa & Fegraus, 2013).

Citizen scientists have also been utilized to gather visual, observational data. Some of the most well-known historic examples of observational data collection from citizen scientists involve bird watchers, the annual Christmas Bird Count began in 1900 and the North American Breeding Bird Survey began in 1966 (Saur, Link, Fallon, Pardieck, & Ziolkowski, 2013). Currently there are several web-based database collections of fish, amphibian and reptile, bird, and mammal images and observations, which have utilized citizen scientists in their creation. These web-based collections of visual data have been utilized to create an atlas of fish in Japan (Miyazaki et al., 2014), and amphibians and reptiles in the Carolinas (Price & Dorcas, 2011). By 2013, the eBird database “collected over 140 million observations submitted by 150,000 different observers, with 10.5 million hours in the field” (Callaghan & Gawlik, 2015, p298).

Utilizing citizen scientists to gather and process visual data not only assists researchers with wildlife studies, their participation may also foster continued interest and education of conservation issues. While the use of citizen science when combined with conservation goals can be powerful (Dickinson et al., 2012), there remains skepticism among some researchers of the scientific validity of their contributions.

Discussion

There are currently different methods that citizen science is utilized to assist with wildlife studies involving visual data. While researchers may set up camera traps to collect photographs, they are increasingly using volunteers to deploy and manage large arrays of cameras. The researchers themselves may process the collected images, but due to the vast amount of data collected, they may again utilize citizen scientists to assist.

Observational data submitted from citizen scientist volunteers is also utilized in wildlife studies. Large web-based databases have been created to collect observations, in some of these projects the observations are verified with a photograph submitted by a citizen scientist.

Camera traps: Use of Citizen Scientists to set-up and manage sites

Armies of volunteers are increasingly being utilized for deploying arrays of camera traps for large-scale studies. Volunteers are trained to manage and maintain camera traps, and given coordinates and instructions where to place the camera, the collected images then uploaded to a database by the volunteers.

Appalachian Trail Corridor Camera traps have been used in large-scale studies, Erb, McShea and Guralnick (2012) studied the anthropogenic effects on the presence of mammals along the Appalachian Trail. Erb, McShea and Guralnick recruited and trained volunteers from trail clubs; the volunteers were trained on camera maintenance and management, camera placement and their sampling protocol (2012). A website was created by the National Park Service for the volunteers to upload images and enter data which were later reviewed by the researchers (Erb, et al., 2012). Erb, McShea and Guralnick (2012) concluded that based on their results, their camera-trapping protocol resulted in obtaining high quality data useful for a landscape scale study.

Jachowski, Katzner, Rodrigue, and Ford (2015) developed the Appalachian Eagle Monitoring Program (AEMP) to study raptor distribution patterns. The AEMP used an array of over 180 baited camera trap/trail cameras placed and managed by volunteers, in sites from Maine to Alabama (Jachowski, et al., 2015). The volunteers were instructed on camera placement, as well as how to bait the stations with a deer carcass, and how often to revisit the site. State coordinators collected the images from the volunteers and the data then forwarded to one of the researchers, Rodrigue, for review (Jachowski, et al., 2015). This one researcher was responsible for the storing, organizing, and analysis of the 2.5 million images that were used in the study. The researchers acknowledged how time-consuming this process is and concluded that the placement of the images into a computer database was needed, and including the involvement of citizen scientists to assist in the identification would be helpful (Jachowski, et al., 2015).

eMammal McShea, Forrester, Costello, He and Kays (2016) developed the eMammal program and database to better survey using camera traps on a large scale. The program was initially used to help with a large-scale camera trap project studying the relationship between domestic cat and coyote presence/detection in protected areas versus urban forests (Kays et al., 2015). The two-year study extended over six states in the eastern United States, with camera traps deployed by 486 trained volunteers (Kays et al., 2015). In addition to the utilization of citizen scientists being useful in large-scale projects, they can also offer access to private properties, which are often habitats of interest to suburban ecologists (Nagy, Weckel, Toomey, Burns, & Peltz, 2012). The photographs obtained from Kays et al. (2015) resulted in 94% of volunteer deployments being correct. The volunteers used eMammal software for initial identification of species in the collected images, entered camera data such as location and then uploaded the images into a database (Kays et al., 2015). The researchers then used the “eMammal Expert Review Tool” to review and process 2.6 million images over a period of 2 years, and found 67-100% of the images accurately tagged by the volunteers (Kays et al., 2015). Kays et al. (2015) state that in a recent project experienced volunteers were almost 100% accurate in camera trap deployment, and were over 90% accurate with identifying 15 of 20 species of wildlife, having less accuracy “distinguishing sympatric species of foxes and squirrels” (p.63). The researchers state further that they are developing algorithms along with crowd-sourcing, similar to Snapshot Serengeti, to validate the most commonly tagged mammals (Kays et al., 2015).

Image processing by citizen scientists: Snapshot Serengeti

While researchers process some collections of camera trap images, there are successful large-scale projects that utilize a large community of citizen scientists to process large amounts of photographic data. Using a model of successful citizen science projects in astrophysics, which use algorithms to produce expert-quality data sets, Snapshot Serengeti uses an algorithm to combine information into a “consensus dataset,” where images circulate in the database to multiple users, and data is entered and processed until a criteria is met (Swanson et al., 2015). Snapshot Serengeti deployed 225 camera traps within the Serengeti Lion Project study area in Serengeti National Park, Tanzania, with the goal of evaluating spatial and temporal inter-species dynamics (Swanson et al., 2015). The cameras have operated continuously since 2011, and by 2013 generated 1.2 million image sets (Swanson et al., 2015).

Snapshot Serengeti is hosted by the world’s most popular online citizen science platform, The Zooniverse, and within three days of launching, “volunteers contributed 1 million species classifications and processed an 18-month backlog of images” (Swanson, Kosmala, Linott, & Packer, 2016). On the Snapshot Serengeti site, citizen scientists “identify species, count the number of individuals, classify behavior, and indicate the presence/absence of young,” and a ‘nothing here’ option is offered to classify those photographs with no animal present (Swanson et al., 2015). Swanson, et al., (2016) found that even though the volunteers participating were neither trained nor asked to exhibit skills in identifying species, the aggregated data submitted to Snapshot Serengeti were 97.9% accurate overall, in comparison, projects that utilize trained volunteers report 85-95% accuracy, and even individual experts are reported at 96.6% accuracy.

The use of and contribution of observational data to web-based databases by citizen scientists

There are several examples of web-based databases that utilize crowd-sourced images and observations from different communities to collect data. Examples include the Carolina Herp Atlas for herpetology enthusiasts (Price & Dorcas, 2011), sport fishing fans are a base for WEB *sakana-zukan* (Japanese Internet atlas of fishes) and the Image Database of Fishes of the Kanagawa Prefectural Museum of Natural History (KPM-NR) which includes a majority of submissions from scuba divers (Miyazaki et al., 2014). One of the most famous crowd-sourced

databases is eBird, which collects observations and data from birders around the globe (Sullivan et al., 2014).

eBird In a little over ten years eBird aggregated more than 140 million observations, recorded by 150,000 unique users (Sullivan et al., 2014), it is one of the most massive collections of biodiversity data providing global access for study (eBird, n.d.). Over 90 peer-reviewed articles have been published utilizing eBird data (Sullivan et al., 2014). The data collected has been used to create species range distribution maps, and monitor range changes, study temporal distribution patterns, and migrations (Sullivan et al., 2009). eBird data is verified using automated filters, records flagged by filters are then reviewed by a network of over “500 regional editors composed of local experts” (Sullivan et al., 2009). Sullivan et al., (2009) do acknowledge potential biases in the data collection; for example, birds that are easily found are reported more often than those species that are difficult to locate and observe, incomplete checklists may be submitted, and locational data may not be accurate as well.

To study the validity of eBird data, researchers have compared eBird data to that collected in traditional surveys. Callaghan and Gawlik, (2015) compared a year’s worth of traditional shorebird survey data collected by trained observers to a year of data collected from eBird for the same location. The researchers found only a minor difference in the estimates of species richness, derived from the higher number of observations from eBird with 35,289 person-hours of observations versus only 2,126 person-hours performed by the standardized surveyors. When this difference in amount of observations was accounted for there was no significant difference between the two data sets and Callaghan and Gawlik concluded that the “use and value of eBird as a tool for land managers and conservationists may be greater than currently realized” (2015).

Conclusion

Camera traps have been shown to be advantageous for collecting data for in situ wildlife studies. As study areas enlarge, especially to landscape size, the appropriate amount of cameras necessitates assistance in setting up the arrays of cameras. Trained volunteers are now more commonly utilized to set up, maintain, manage, collect data and upload images for researchers (Erb, et al., 2012; Jachowski, et al., 2015; Kays et al., 2015; McShea, et al., 2016). Kays et al.

(2015) trained 486 volunteers for their study and achieved 94% accuracy in camera placement. Corrective feedback was given to the volunteers with less than standard camera set ups, resulting in 100% accuracy in a more recent study (Kays et al., 2015). Kays et al. demonstrated that the training and gathering of image data from a large number of volunteers was not only possible, but also completely successful.

With more cameras being utilized, more images are obtained, easily reaching millions of images, translating into terabytes of data. Since many researchers did not trust the validity of citizen scientists to process photographic data, which may include identifying species, counting individuals, and classifying behavior, many researchers process the image data collected. In the AEMP study, one researcher was solely responsible for the storing, organizing and analyzing 2.5 million images (Jachowski, et al., 2015). This seems unrealistic, and when we consider Swanson et al. (2015) states that experts have been reported at 96.6% accuracy, the protocol of one individual responsible for the analysis of such an enormous amount of photographic data may be unwise. In retrospect, Jachowski, et al. (2015) acknowledged that their protocol was time-consuming and that utilizing citizen scientists and a web-based database would be beneficial for such a project.

McShea, et al. (2016) developed eMammal to assist such large studies with processing data utilizing citizen scientists, and they developed an “eMammal Expert Review Tool” to review data processed by citizen scientists for quality control. Kays et al. (2015) used the review tool to process 2.6 million images over a two year period, demonstrating that quality control protocols by researchers reviewing data is hindering the amount of time to process image data. Snapshot Serengeti utilizes algorithms, and circulates images in their web-based database for citizen scientists to process, and when a consensus and specific criteria are met for a specific image, the image is pulled from circulation and processing is deemed complete. Swanson, et al. (2016) state that citizen scientists participating in Snapshot Serengeti not only processed 18-months of images in just 3 days of the website’s launch, but achieved 97.9% accuracy in processing the data, including identifying species, which was a higher accuracy than that reported for experts. Kays et al. (2015) acknowledge the power and validity of the Snapshot Serengeti model for citizen scientists processing photographic data accurately and quickly, and plan to develop algorithms and use crowd-sourcing in the future to improve eMammal. Snapshot

Serengeti, currently processing its ninth season, is so successful that the new bottleneck is not processing but obtaining images.

In addition to the utilization of citizen scientists for collecting and processing photographic data, volunteers also assist researchers by submitting observational data. Web-based databases such as eBird (Sullivan et al., 2014) have demonstrated their value to global avian research. Callaghan and Gawlik (2015) demonstrated the validity of observational data submitted by citizen scientists by comparing data obtained from citizen scientists and that of trained observers for the same site. The caveat remains that there are reporting biases with observations of more easily detected species of birds, or amphibians and reptiles (Price & Dorcas, 2011) being submitted more frequently than those species which may be more difficult to locate. Regardless of the reporting bias, the observational data does have value for the presence of species and behavioral studies.

In conclusion, Snapshot Serengeti and eBird are two great examples of web-based database platforms, powered by citizen scientists, which have resulted in valid contributions to biodiversity research.

References

- Audubon. (n.d.) Christmas Bird Count: History of the Christmas bird count. Retrieved from <http://www.audubon.org/conservation/history-christmas-bird-count>
- Callaghan, C., & Gawlik, D. (2015). Efficacy of eBird data as an aid in conservation planning and monitoring. *Journal Of Field Ornithology*, 86(4), 298-304. doi:10.1111/jof.12121
- Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, (6), 291.
- eBird. (n.d.). About eBird. Retrieved from <http://ebird.org/content/ebird/about/>
- Erb, P., McShea, W., & Guralnick, R. (n.d). Anthropogenic influences on macro-level mammal occupancy in the Appalachian trail corridor. *Plos One*, 7(8)
- Jachowski, D. S., Katzner, T., Rodrigue, J. L., & Ford, W. M. (2015). Monitoring landscape-level distribution and migration phenology of raptors using a volunteer camera-trap network. *Wildlife Society Bulletin*, 39(3), 553-563. doi:10.1002/wsb.571
- Kays, R., Costello, R., Forrester, T., Baker, M., Parsons, A.W., Kalies, E.L., Hess, G., Millspaugh, J. & McShea, W. (2015). Cats are rare where coyotes roam. *Journal of Mammalogy*, 96(5), 981-987. doi:10.1093/jmammal/gyv100
- McShea, W. J., Forrester, T., Costello, R., He, Z., & Kays, R. (2016). Volunteer-run cameras as distributed sensors for macrosystem mammal research. *Landscape Ecology*, 31(1), 55-66.
- Miyazaki, Y., Murase, A., Shiina, M., Naoe, K., Nakashiro, R., Honda, J., Yamaide, J., & Senou, H. (2014). Biological monitoring by citizens using Web-based photographic databases of fishes. *Biodiversity and Conservation*, 23(9), 2383-2391.
- Nagy, C.M., Weckel, M. E., Toomey, A., Burns, C.E., & Peltz, J. (2012). Validation of a citizen science-based model of coyote occupancy with camera traps in suburban and urban New York, USA. *Wildlife Biology In Practice*, 8(1), 23-35. doi:10.2461/wbp.2012.8.3

- Price, S. J., & Dorcas, M. E. (2011). The Carolina Herp Atlas: An online, citizen-science approach to document amphibian and reptile occurrences. *Herpetological Conservation and Biology*, 6(2), 287-296.
- Saur, J.R., Link, W.A., Fallon, J.E., Pardieck, K.L. & Ziolkowski, D.J. (2013). The North American Breeding Bird Survey 1966-2011: Summary analysis and species accounts. *North American Fauna*, 79, 1–32. doi:10.3996/nafa.79.0001
- Sheil, D., Mugerwa, B., & Fegraus, E. (2013). African golden cats, citizen science, and serendipity: tapping the camera trap revolution. *South African Journal of Wildlife Research*, 43(1), 74-78.
- Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., & Kelling, S. (2009). eBird: a citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142, 2282-2292.
- Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., & ... Kelling, S. (2014). Review: The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation*, 169, 31-40. doi:10.1016/j.biocon.2013.11.003
- Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, 2, 150026. <http://doi.org.proxy.lib.miamioh.edu/10.1038/sdata.2015.26>
- Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, 30(3), 520-531.
- Trolliet, F., Huynen, M., Vermeulen, C., & Hambuckers, A. (2014). Use of camera traps for wildlife studies. A review. *Biotechnology, Agronomy and Society and Environment*, 18(3), 446-454.