The value of citizen science data for passive surveillance of wildlife

RB Dissanayake[1]*, M Stevenson[2], R Allavena[1], J Henning[1]

[1]School of Veterinary Science, University of Queensland, Australia, Gatton, Queensland, 4343, Australia; [2]Faculty of Veterinary and Agricultura[l] Sciences, University of Melbourne, Parkville, Victoria 3010, Australia

*r.dissaayake@uq.eud.au

Abstract

The collection of wildlife distribution data using standard survey sampling methods is expensive. For this reason surveys are often limited to small geographical areas and are usually conducted over short time periods. We used a citizen science dataset on koala sightings reported between 1997 and 2013 across eight local government areas in South East Queensland (SEQLD), Australia (n=11,029) to identify spatial and temporal trends of koala sightings. In addition we used koala sightings data from 2011 (n=353) to model koala distribution. The spatial aggregation of sightings was high in high koala density areas and in areas closer to roads. Temporal trends of sightings mirrored the breeding season of koalas with an increase in the number of sightings per day between September and October. Our results suggest that citizen science data are a useful data source for wildlife population monitoring and are suitable for refining and supporting koala density estimates derived from active survey sampling approaches.

Keywords: *Koala, KoalaBASE, Queensland, surveillance, citizen science*

Introduction

The koala (*Phascolarctos cinereus),* an arboreal folivorous marsupial, is an iconic Australian wildlife species (1). The koala is classified as 'threatened' (2) with its distribution and abundance rapidly declining in SEQLD of Australia over the last few decades (3). The decline in koala numbers is due to the fragmentation and clearing of habitat, urbanisation, hotter and drier weather, vehicular traffic accidents, dog attacks, chlamydiosis and other diseases. Further, low genetic variability has also contributed to the decrease in numbers (4). The Moggill Koala Hospital (MKH) run by the Queensland Government Department of Environment and Heritage Protection (DEHP) maintains a database developed by the University of Queensland, where citizen science data on koalas and clinical information on koalas submitted to koala hospitals is stored. The database, called 'KoalaBASE', is a state-of-the-art multifunctional online SQL database.

Koala sightings, threats or mortalities of koalas are identified by members of the public and this information is passed on to staff of MKH, who respond to these calls and record the information into KoalaBASE. Sick or animals under threat are rescued, and if required, treated at MKH. Animals which recover are released back into the wild, whereas animals with poor prognostic outcomes following a veterinary medical assessment are euthanised.

KoalaBASE represents a passive surveillance system for monitoring koala sightings, threats, rescue details, health information, treatments, and release details. KoalaBASE contains 40,765 geo-referenced records of sightings and submissions of koalas to hospitals between 1997 and 2013. Citizen science, the collection of data (usually relating to the natural environment) by members of the general public, is gaining importance (5) and popularity recently due to the ubiquity of social media, smart phones and web technology, which serve as a cheap and easily accessible methods for collecting data. However citizen science data is not without biases. During an organised one-day citizen science programme for recording of koala sightings in South Australia (6), it was noted that correct sightings of koalas required some experience, although animals on the ground may be identified correctly by inexperienced observers.

To date no research had been conducted to analyse koala sightings in SEQLD. The objective of the study was to identify the spatial-temporal trends in citizen science data and model distribution of koala sightings in SEQLD. Our aim was to determine the usefulness of citizen science data as a passive surveillance tool to inform researchers, wildlife carers and policy makers about wildlife abundance and wildlife diseases.

Materials and methods

We used sightings data from the KoalaBASE collected between 1997 and 2013 across SEQLD.

Our definition for a koala sighting is a koala sighted in a tree or on the ground. Animals that were presented to the hospitals, were reported dead, or needed to be euthanised were excluded from the analysis. Most records were georeferenced, with the longitude and latitude of the sighting recorded by the member of the public. When no GPS location of sighted koalas was recorded, the street address of the location where the koala was sighted was used to identify the location coordinates on Google maps (www.google.com.au/maps). Sightings of more than one animal, or joeys of koalas were recorded as separate incidents at the same location. The same sighting reported more than once in the same location by multiple observers, repeated sightings of the same animal on the same day, and erroneous records were removed from the analysis. Microsoft

Access and Excel was used for data cleaning, coding, sorting, and management of data, the spatstat package was used for modelling in R (16) and ArcMap was used for mapping.

In order to describe the quality and the patterns emerging from sightings data, we analysed spatial-temporal trends, estimated sighting frequencies for data from 1997 to 2013 and fitted a Poisson point process model for 2011 sightings data (*n*=353).

For the 2011 sightings data we selected one koala sighting from each 1 km grid overlayed on the study area to account for spatial bias (15). Then we described the koala density as a function of several explanatory variables known to have impact on koala habitat. Explanatory variables tested in the model were soil clay, soil phosphorous, soil bulk density, soil water, human population, temperature, rainfall, elevation, and distance to different categories of habitat suitability. We quantified the association between each of these explanatory variables and the density of koala sightings using the rhohat procedure in spatstat (16).

We then fitted a Poisson point process model with the density of koala sightings as the outcome and each of the explanatory variables listed above. A Strauss spatial interaction term was used to account for spatial autocorrelation in koala sighting locations. Models were compared by Akaike information criterion (AIC) and lurking variables plots were used to assess the model for misspecification of spatial trend. Predictions were calculated from the final best model and converted to densities per hectare.

Results

The temporal distribution of sightings over the 1997-2013 is shown in Figure 1. Sighting numbers per annum were relatively constant throughout the study period, but in 2011 lower numbers of animals were reported compared to previous and subsequent years.

There was a strong seasonal trend in sightings, with most sightings being observed during the koala breeding period (September-November, Figure 2).

Both night and day time sightings are recorded in KoalaBASE, however only 5% of records contain time information. The highest number of sightings was recorded from 6am to 12 noon (33%) while the lowest number of sightings was recorded during the early morning hours between 12am to 6pm (12%). More sightings were recorded during evening hours (30% from 12pm to 6pm and 25% from 6 am to 12 midnight).

The observed spatial distribution of koala sightings in 2011 is shown in Figure 3. Koala density derived using spatial covariates in the Poisson model and model diagnostic plots are shown in Figure 4.

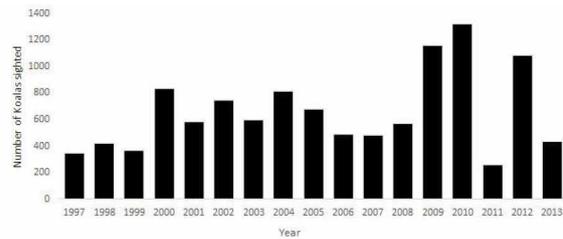**Figure 1.** Number of koalas sighted per year between 1997 and 2013 in SEQLD, Australia.

Figure 2. Number of koalas sighted per month for four-year periods between 1997 and 2013 in SEQLD, Australia.
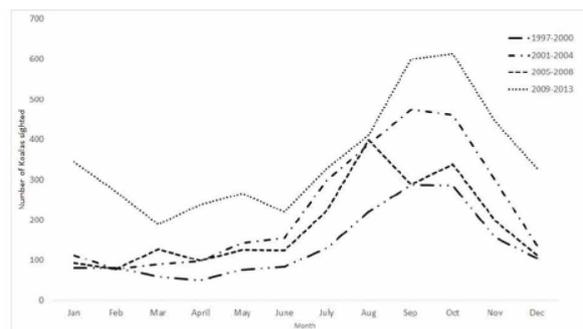

Figure 3. Spatial distribution of sightings in 2011 and study area (square) in SEQLD, Australia.
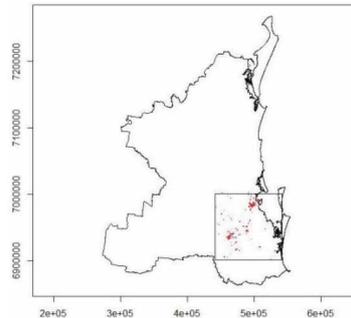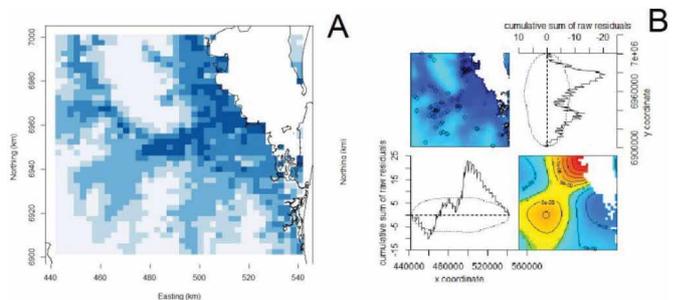

Figure 4. Density estimates from the Poisson model (A) and the diagnostic plot (B)

Discussion

The sighting locations overlayed with Google base maps in ArcMap indicate a higher number of sightings close to roads, where animals have higher visibility. More sightings were

recorded during the breeding seasons in spring and summer, coinciding with increased koala dispersal and movement of animals seeking mating partners.

Reasons for the reduced sightings in 2011 need to be explored, but it may be related to flooding occurring in large parts of SEQLD in this year causing people to prioritise their attention for other important tasks.

Sightings were reported by citizens throughout day and night indicating that citizen scientists are active 24 hours, though higher koala sightings frequency in the mornings and evenings likely also indicates increased animal activity in cooler hours.

The distribution pattern of koala sightings derived from density estimates of the 2011 Poisson point process model (Figure 4A) are quite low which could be due to that in this year a lower number of sightings in a spatially limited area were reported.

Verification of species and location of sightings are useful to ensure that good citizen science data is collected. Many citizen science programmes conducted with the participation of trained volunteers involve species and location validation, as well as recording of search effort as the distance travelled and time spent on the search. Search and reporting effort is probably influenced by a number of variables, such as human population density and accessibility of the area. For example, our density maps (Figure. 4) show high koala density in areas, where human density is also relatively high. Quantifying the search and reporting effort and possibly including it the species distribution models seems paramount.

Sightings information has been previously used in species distribution models using Maxent, boosted regression tree methods (10), generalised linear models (GLM) and generalised additive models (GAM) (11,12). Similar models have been developed using herbarium or museum records for rare plant and animal species (10,13). However, issues of spatial heterogenic of sightings and detection errors have to be addressed.

Citizen science is a useful passive surveillance tool to collect data on animal sightings and disease occurrence. For example, global databases like PROMED (7) and FAO-EMPRES-i (8) have long been used to compile passive surveillance information from various sources, such as news reports and unofficial communications. These databases also provide unlimited access to summarised animal disease events to researchers, policy makers and members of the public alike. Similar databases have been developed for wildlife events in several countries (9).

KoalaBASE is used to record, summarise and to map koala data from SEQLD, but data entry of koala sightings information stopped in 2013. An up-to-date centralised database for koala sightings in SEQLD, including easy-to-

use tools to upload good quality citizen science data would be highly desirable.

Active surveillance such as wildlife data collection using survey sampling methods is expensive, it is usually limited to small geographical areas and is often conducted over short time periods. Incorporating passive surveillance data to improve and confirm estimated wildlife density measures derived from actively collected data seems desirable. Only a few studies have tried to combine different types of datasets. For example, Giorgi *et al.* (2013) developed a geostatistical model accounting for heterogeneity across surveys to correct for spatially structured bias in non-randomised prevalence surveys (14). A similar approach could be applied to combine active wildlife survey sampling data such transect counts (considered as gold-standard data) with sightings information (considered as a convenience sample).

However, limitations and potential biases of citizen data have to be considered. The quality of the citizen science data is influenced by the experience and interest of the citizen in identifying and reporting animals. For example, as the koala is an iconic species, people might show a higher reporting effort compared to other species. People might also monitor a koala's presence if once noticed in an area due to curiosity, and may report several sightings of the same animal.

We conclude that citizen science wildlife data could be a reliable source of information with the potential to monitor population trends and for developing species distribution models. However citizen science data needs to be thoroughly examined for potential biases before it can be fully exploited and used. Wildlife data collected by members of the public might be useful for abundance estimation if the citizen data is combined with actively collected wildlife data that has been obtained using standardised survey methods.

References

1. **Melzer A *et al.** *Conserv Biol.* 14(3), 619-628, 2000
2. **Adams-Hosking C *et al.** *Divers Distrib* 22(3), 249-262, 2016
3. **Preece HJ.** *Modsim.* p1294-1300, 2007
4. **Adams-Hosking C *et al.** *Wildlife Res* 38(2), 122, 2011
5. **van Strien AJ *et al.** *J Appl Ecol* 50(6), 1450-1458, 2013
6. **Sequeira AM *et al.** *Ecol Evol* 4(11), 2103-14, 2014
7. **Woodall J *et al.** *Emerg Infect Dis* 7(3), 563-563, 2001
8. **Martin V *et al.** *J Wildlife Dis* 43(3), S71-S76, 2007
9. **Sullivan BL *et al.** *Biol Conserv* 169, 31-40, 2014
10. **Elith J *et al.** *J Anim Ecol* 77(4), 802-13, 2008
11. **Elith J *et al.** *Divers Distrib* 17(1), 43-57, 2011
12. **Austin M *et al.** *Ecol Model* 200(1-2), 1-19, 2007
13. **Miller-Rushing *et al.** *Front Ecol Environ* 10(6), 285-290, 2012
14. **Giorgi E *et al.** *J Roy Stat Soc A Sta.*178 (2), 445-64, 2015
15. **Sequeira AM.** *Ecol Evol* 4(11), 2103-14, 2014
16. **Baddeley A.** *Journal of Statistical Software* 12(6), 1-42, 2005