Estimating Chacma baboon population characteristics

in the Garden Route National Park (South Africa)

from opportunistic camera-trap data

Presented by Ethan Cohen
2nd year Master's degree, Biodiversity, Ecology, Evolution, Sorbonne University
2019 - 2020

Supervised by Dr.Chloé Guerbois, Sustainability Research Unit & IRL REHABS, Nelson Mandela University
&
Dr. Roger Pradel, CNRS-CEFE , Université de Montpellier

Sustainability Research Unit (SRU)
Nelson Mandela University (NMU)
George campus Madiba drive
6529 George, South Africa

17/02/2020 – 30/06/2020

Préface

Il me semble pertinent d'expliquer dans une préface les ajustements de stage résultant de la pandémie de Covid-19 pour comprendre la nature de mon sujet de mémoire.

Dans la municipalité de George, comme dans une grande partie de la province du Cap occidental en Afrique du Sud, les conflits hommes-babouins sont quotidiens. Afin de comprendre l'impact des perturbations anthropiques sur cette population de babouins et de proposer des solutions de gestion adaptées, il est nécessaire de mettre en place un suivi de population à long terme. Le premier objectif de cette étude était donc de développer une méthodologie non invasive pour le suivi à long terme des troupes de babouins via l'utilisation de pièges-photographiques sur une troupe connue, couplés à des méthodes d'estimation type Capture-Marquage-Recapture. Le stage a débuté par une période d'étude bibliographique sur l'utilisation de pièges-photographique et le comportement de l'espèce dans la province du Cap. Elle a continué avec un suivi sur le terrain de la troupe, à George, avec une période d'habituation et une période d'identification, de sexage, de classification par âge, de suivi GPS de la troupe et d'identification de méthodes d'échantillonnage les plus pertinentes en testant différents emplacements (sites de repos, sites avec appâts, routes les plus souvent empruntées ou sites pris aléatoirement dans le domaine vital de la troupe). S'en serait suivit, la collecte et l'analyse des jeu de photos et la comparaison des résultats avec les observations de terrain afin de comparer l'efficacité de différentes méthodes d'échantillonnage.

La crise sanitaire ayant donné lieu à un arrêt de toutes les activités de recherches à l'Université Nelson Mandela ainsi qu'à une demande de rapatriement de Sorbonne Université, le travail d'observation n'a pu durer qu'une semaine et ce qui suivait n'a pas pu être effectué. L'étude s'est donc rabattue sur un jeu de données récolté par Lizette Moolman, écologiste aux Parcs Nationaux Sud-Africains (SANParks) et cet étude a débuté à partir de mi-Mars. Le protocole de collecte de photos était destiné à évaluer l'abondance d'une population d'éléphants relictuelle dans le parc national de la Garden Route, une zone d'occurrence d'une population de babouins Chacma. Le placement des caméras, adapté aux éléphants d'Afrique, mais non sans défauts pour l'étude des babouins Chacma, ne nous permettait pas, entre autres, d'identifier chaque individu. Nous avons donc revu nos méthodes et proposons une approche d'estimation du nombre de troupes et un modèle d'estimation d'abondance d'une population non marquée dans une nouvelle zone d'étude avec l'utilisation de données « opportunistes ».

Acknowledgments

My master 2 internship was an incredible experience full of surprise. I would return in South Africa again without hesitation, but, I hope, for more than only ten day this time.

First, I would sincerely thanks Chloé Guerbois for giving me the chance to work on the Baboon project and giving my first real field experience. Even if the dominant male did not seem to like me, my seven days with the troop was awesome. Mainly for that I should like to express all my gratitude. But also, for welcoming me, teaching me about primates and supported me throughout this internship.

I am particularly grateful to Roger Pradel for his investment in my internship, for teach me basics of Bayesian modelisation and as Chloé, for always taking time to answer my questions.

I would also like to thank Hervé Fritz, for giving me a warm welcome in South Africa and for his participation to my supervising.

I would also thank Nelson Mandela university for welcoming me and allowing me to study in the campus.

To my family, thank you for helping me to be able to go in South Africa and for your support. To Juliette, thank you for always be there, I love you.

Abstract

During the recent decades, an increasingly popular monitoring system for biodiversity has been the utilization of camera traps, considered as an effective tool to ensure continuous population sampling. Hence, amount of dataset is increasing, then our two objectives will be to evaluate if we can recycle this dataset and develop an adapted model to assess an unmarked population characteristic and their variations over time. Our dataset contains pictures of each event that occurred in a study area, at North of Knysna, Western Cap, South Africa. Our focal specie Chacma Baboon lived in social group. Then, we used our dataset to combine three different approaches (distance-time ratio between events, individual identification and k-mean clustering), aims to assess the number of troops. Using results of our three approaches with a Bayesian inference model based on capture mark recapture stories, we assess relative age class abundance of each troop. However, survey design of the opportunistic dataset has many defaults that impact our results. Hence, we were not able to correctly assess a number of troop and their characteristics. However, using an opportunistic dataset have advantages and we discuss about some method improvements that could make reliable this approach.

Au cours des dernières décennies, la popularité des systèmes de surveillance de la biodiversité, basés sur l'utilisation de pièges photographiques n'a cessé d'augmenter. Cet outil, considéré comme efficace pour les suivis continue de populations, fait que le nombre de données récoltées par les pièges photographiques ne cessent d'augmenter. Donc l'objectif de cette étude sera de développer un modèle, destiné à évaluer les caractéristiques démographiques d'une population non marqué et ses variations au cours du temps, en s'appuyant sur un de ces jeux de donnée que l'on peut qualifier d' « opportuniste ». Notre jeu de données comprend les photographies de tous les événements montrant des babouins chacma (Papio *ursinus*) sur une zone d'étude au Nord de Knysna au Cap Occidental, en Afrique du Sud. Les babouins chacma étant des primates sociaux, on cherche dans un premier temps à évaluer le nombre de troupe dans la zone, en combinant trois approches permettant de regrouper les événements présentant les mêmes troupes. On utilise ensuite ces résultats avec les histoires de capture marquage recapture des individus dans un modèle d'inférence bayésienne pour évaluer l'abondance relative d'une troupe et des différentes classes d'âges. Néanmoins, le jeu de données présente de nombreux défauts ayant un impact sur nos résultats. Nous ne sommes donc pas en mesure de dire que nos estimations sont fiables. En revanche, l'utilisation de ce type de données a des avantages et nos méthodes peuvent être améliorées.

Summary

Introduction

The urgent need to provide food to an increasing human population induced a rapid intensification of land use during the 20th century at the expenses of losing natural habitats (Ellis *et al.,* 2010). Human encroachment on natural habitats increases human-animal proximity, generating increased competition between humans and wildlife which may lead to conservation conflicts (Heydon *et al.,* 2010; Redpath *et al.,* 2013). In Southern Africa, the Chacma baboon (*Papio ursinus*) is one of the species that can benefit from this proximity to humans (Fehlmann *et al.,* 2017a), thanks to their remarkable behavioural flexibility (Hoffman & O'Riain, 2012). For this reason, baboons are increasingly seen in urban and peri-urban areas (Fehlmann *et al.,* 2017a). Their main diet includes vegetables and fruits, as well as insects and small vertebrates to maximize protein intake (Codron, 2006; Johnson, 2013) but in anthropogenic areas, baboons can also forage on rubbish, and other anthropogenic food resources by breaking inside cars and houses or stealing food directly from humans. This may result in substantial damages, economic losses and contribute to negative attitudes towards baboons, impacting on overall conservation efforts (Fehlmann *et al.,* 2017b; Hill *et al.,* 2018; Kellert *et al.,* 1996). These conflicts between humans and Chacma baboons are affecting most of the urban and peri-urban areas of the Western Cape of South Africa, however very little is known about the status and trends of baboon populations in this region. To understand the relationships between anthropogenic activities and baboon ecology and design sustainable conservation strategy, there is an urgent need to develop long term monitoring of baboon populations and dynamics.

During the recent decades, an increasingly popular monitoring system for biodiversity has been the utilization of camera traps, considered as an effective tool to ensure continuous population sampling (Trolliet *et al.,* 2014). Depending on the focal species and survey design, camera traps can be used in different ways to study population trends. For non-invasive surveys, one way is to use natural marks. For example, Kalle *et al.,* (2011) used a spatially explicit capture–recapture (SECR) model and Bayesian inference to estimate the density of leopard and tiger in the Mudumalai Tiger Reserve (southern India) by identifying single individuals on photos based on their unique strips and rosette pattern. The same model was used by Després-Einspenneret *et al.,* (2017) with a population of western chimpanzees in the Taï National Park (Ivory Coast). If SECR models seem to deliver reliable results when the conditions of the model, i.e., independence and random distribution of movement in the sampling area, are applicable (Efford & Fewster, 2013), a recent study also suggests that identification errors in camera-trap studies

often result in systematic population overestimation (Johansson *et al.,* 2020). Another version of SECR model for unmarked individuals (USCR) by Chandler & Royle (2010) estimated animal density instead of population size within an unknown area. For each sample occasion, generated via a collection of closely spaced "traps", the number of individuals was counted. The model is based on the decreasing probability to catch individuals that expand with Euclidean distance between individuals and their activity centres (Chandler & Royle, 2010; Gilbert *et al.,* 2020). Nevertheless, USCR produces highly imprecise density estimates and is sensitive to assumption violation (Gilbert *et al.,* 2020). Other models exist for unmarked individuals like the space-to-event model where cameras are programmed to take photos at predefined times, or the random encounter model which estimates a density from encounter rates, animal movement speed and camera's detection viewshed (Gilbert *et al.,* 2020). This last model assumes that cameras are randomly deployed in the study area.

The objective of this study is to assess the characteristics (number of troops, troop size, age- and sex-ratios) of a baboon population from an opportunistic data set generated by a survey that does not fit the different assumptions that apply to the models described in the bibliographic review. Due to increasing utilization of camera-traps (Trolliet *et al.,* 2014), and hence an increasing amount of data, it seems interesting to assess the possibility of recycling this data for different research. This opportunistic data set, provided by South African National Parks (SANParks), originates from an elephant monitoring survey in the Knysna and Wilderness sections of the Garden Route National Park (Western Cape, South Africa). The survey used 38 camera trap stations of which 33 stations captured baboons. We used a sub-set of 10 camera trap stations, which contained about 6000 images of chacma baboon to test our method.

We also want to highlight the opportunities and challenges provided by an imposed survey design and an unmarked animal population (Trolliet *et al.,* 2014).

We made different hypotheses about the possibility to calculate an accurate number of troops in the study area:
   1) Troops have a limited travel speed, then with the calculation of distance and time interval between events, we will be able to determine if co-occurring events could be triggered by a same troop or not.

2) Pictures quality allows most times to sex and determine the age-class of each visible individual, and thus to approximate the structure of each troop who triggered the cameras.

3) Even if individuals are unmarked, the quality of the picture would allow to identify a few individuals with conspicuous physical or behavioural characteristics, hence we should be able to use positive individual identification for troop identification.

4) Assuming that the probabilities of detecting an individual in a troop is equal and assuming the troop structure would be stable in our timeframe, counting results could be useful to compare the event's troops structure.

We made further assumptions to determine the troops size:

1) We assume that in our timeframe, baboon troops would be cohesive, and fission would be negligible (Noser and Byrne, 2007).

2) For each event involving several individuals, the entire troop is moving, but some individuals are out of the camera's view range.

3) The low time interval between photos allows us to follow each individual all over the event. Then, for each individual of the event, we can create a CMR history over the series of photos constitutive of the event (i.e., whether it is visible or not on each photo).

4) A Bayesian inference model of the CMR histories serves to derive a troop size estimate for each event using a Markov chain Monte Carlo approach.

5) Events are a troop sample and should be combined to derive consensus estimates for all event recognized to belong to the same troop.

Materials and methods

Study area and data collection

The camera-trap pictures used in this study were produced by South African National Parks (SANParks) as part of a camera trap survey aiming at estimating the size of a relict free ranging elephant population (Western Cape, South Africa). The study area includes the Knysna section of Garden Route National Park covered mostly with Afro-montane forest, as well as some timber plantations and privately-owned forested lands. The elephant camera trap survey started the 15th May 2016 and ended the 10th November 2017, covering an area of 97,32 km^2 through 38 stations stratified along pathways (Moolman *et al*., 2019) (Figure 1). Cameras were fastened to trees about 1 meter away from the paths and at 1-1,15 m height and disposed in

opposition with theirs infrared beams pointing the road (Moolman *et al*., 2019). They were set to take 8 megapixels pictures with a 1 second delay between photographs and 2 second delay between videos, until infrared sensor no longer detects an individual. The flash intensity was set on 'high' when taking photos and on 'normal' when taking videos and sensor sensitivity level was set on "normal".

Baboon population dynamics and ecology

Our focal sub species, *Papio ursinus ursinus* also named Southern Chacma (Duane *et al.,* 2014), is an ubiquitous species living in woodlands, savannahs, grasslands and rocky hills. Baboons are primarily diurnal and spend the night in trees if available or on the faces of cliffs (Davidge, 2010).

They live in stable social troops usually between 15 and 80 individuals (Davidge, 2010) but some studies have reported bigger troops. The adult sex ratio is always female-biased, with 0.1-1 male per 1 female. Troops comprise adult males, adult females, subadults and juveniles (Wallace & Hill, 2012), the age class ratio is variable (Davidge, 2010). The female average gestation is 187 days (173–193) and inter-birth interval ranging between 20 and 38 months (Duane *et al.,* 2014). Males tend to disperse to other troops while females are philopatric (Hamilton, 1990) and troop home-range may vary widely in size. Using global positioning system (GPS) collars in similar environment, Pebsworth *et al.,* (2012) estimate a home-range no less than 15,4 km^2 for a troop of 115 baboons occurring in Wildcliff Nature Reserve located in the Western Cape, South Africa. In the Cape of Good Hope section of the Table Mountain National Park (also Western Cape), Lewis & O'Riain (2017) using GPS and field observations, estimate a home range of 45,3 km^2 for a natural-foraging troop of 56 individuals. Baboons usually travel as a cohesive group with a maximum rate of travel of 1.26 km/h (Lewis *et al.,* 2017). Time allocated to each activity may vary widely between seasons. Travelling time activity can account for 25 % (winter) to 40 % (autumn) of daylight hours (Lewis *et al.,* 2017). However, these home ranges and travel speed were estimated for troops who exploit natural resources, those who exploit anthropogenic foods spend less time foraging, occupy smaller home ranges, and travel shorter distances (Noser & Byrne, 2007). The percentage of home range spatial overlap between different troops varied from 0.7-17.2 % with a mean overlap of 7.3 % according to Hoffman (2011).
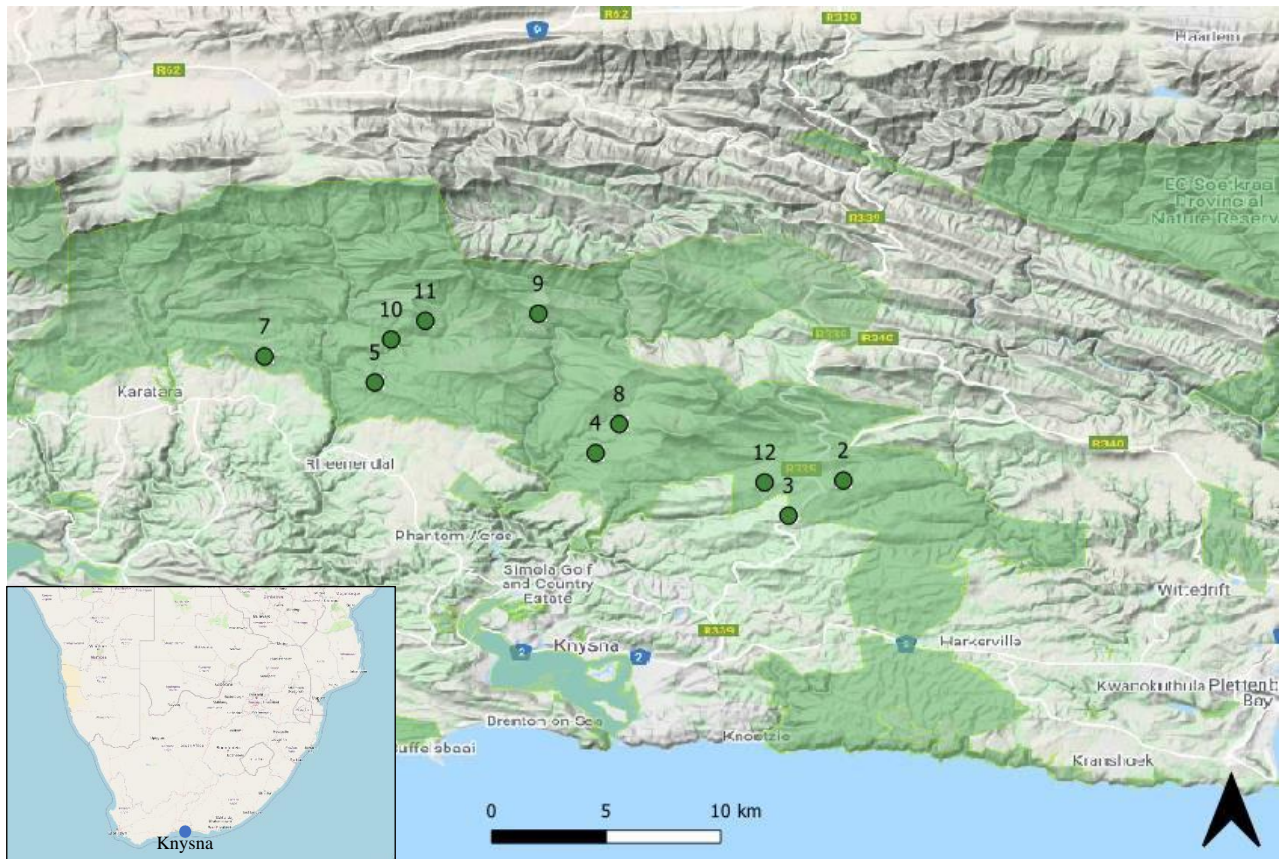
Figure 1: Study site localisation, Knysna, Western Cape, South Africa
(Source: CapeFarmMapper ver 2.3.2.7). Green points are station localizations and green
zones are protected areas.

Data processing

We only used the data from ten stations as the time frame did not allow to process more data. These 10 stations account for a total of 201 baboon capture events and 6008 pictures (Figure 1). Amongst these 10 stations, 4 stations captured 86% of the total events (Figure 2). Due to time constraints, we subsampled 111 baboon-capture events (2732 photos / videos) from those, taken between the 15th May 2016 and the 07th May 2017, to estimate the population base on a full year. An event is defined by consecutive images of baboons, with a maximum interval of ten minutes between two photos or two videos (Moolman *et al*., 2019). Pictures were processed from the most ancient to the most recent event.

For each event, we tried to determine the sex and age of each individual using a combination of physical characteristics: males are often larger than females and have a longer muzzle,

nipples are not visible (Duane, 2014; Davidge, 2010), males butt is characterized by a grey stripe whereas females are easily identifiable with a pink butt and a big pink swell during oestrus, highly visible on photographs (Figure 3).
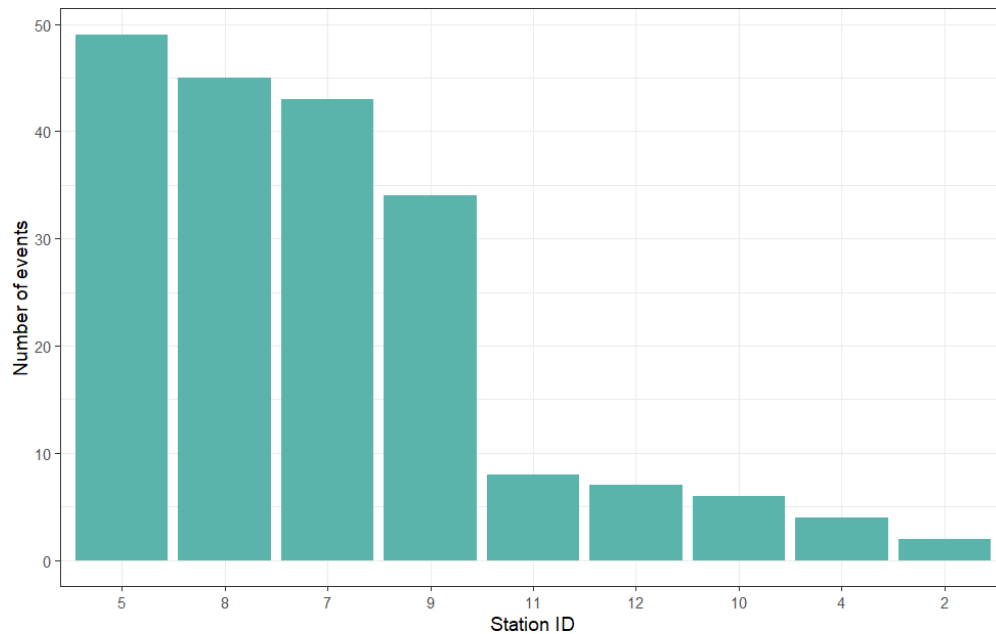

Figure 2: Frequency repartition of the 111 events analysed among stations that observed baboons between the 15th *May 2016 and the* 07th *May 2017*

Figure 3: Difference between male and female chacma baboon. The male (left) presents a gray stripe and is much larger than the female (right) who present a starting pink swell. (Photos from SANParks).

Age class was determined by corpulence, muzzle size, and if individuals were carried or stayed close to another baboon (Wallace & Hill, 2012; Duane *et al.,* 2014). Subadults have not fully grown and frequently exhibit independent behaviour (Wallace & Hill, 2012), while juveniles were carried by their mum or seating close to her. When the sex of an adult was uncertain, it is classified as "undetermined adult". The result of the counts was captured in an Excel datasheet including event ID, station ID, number of photos of the event, date, event start time, end time, event duration, number of individuals count, number of adults, number of adult males, number of adult females, number of subadults, number of juveniles and number of undetermined adults. As we know, age ratio may vary widely between troops while sex ratio is female-biased (Davidge, 2010). In order to differentiate the structures of each troop, our analyses are based on the number and the proportion of individuals in each age-class (adults, subadults and juveniles). To reduce the risk of error when counting, events that displayed technical issues (*i.e. many pictures missing, too dark or too light photos, a baboon who block camera's view*) or showed a too low number of individuals (<10) were discarded. Hence, among the 111 subsampled events, only 47 were used for analysis. Events discarded contains all events in station 2, 4 and 10. Hence, we work on 6 stations instead of 10.

Determining the number of baboon troops in the study area

Three complementary approaches were used to determine the number of troops based on the information gathered for each event:

Distance-time ratio

We first identified pairs of successive events that could not be triggered by the same troop, based on the maximum daily speed of a baboon troop (Lewis *et al.,* 2017). A pair of successive events is defined by one event and its closest event in time. We first calculated spatial and temporal distance matrices using the lubridate, dplyr and sf R packages for each pair of temporally successive events. It allows us to produce an exclusion criterion usable to define if two successive events could belong to different troops based on the spatio-temporal distances (if the spatial (d) - temporal (t) distance between a pair of event > Vmax, two consecutive events can not involve the same troop).

Individual identification

For each pair of successive events not classified as "different troops", we analysed photos to find identifiable and clearly visible individuals. Then, we assessed if two consecutive events shared individuals by comparing pictures to see if one or more identified individuals of the first event matched with identified individuals of the second. When this was the case, both events were indicated as involving the same troop. The combination of physical characteristics used to compare individuals were corpulence, face shape, muzzle size, tail size, hair colour, amputation and / or malformation. Software GIMP 2.10.18 have been used to visualize pictures.

The process of exclusion and inclusion of events in a troop follow the pattern explained Figure 4.
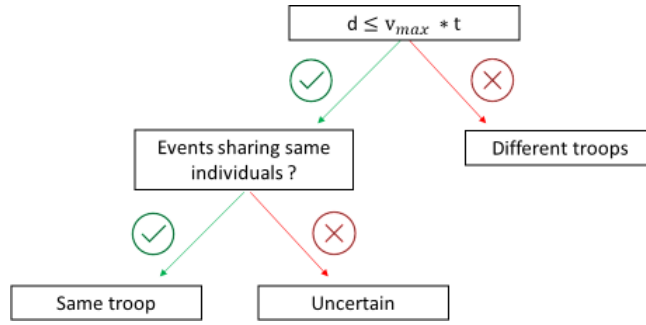

Figure 4: Decision tree to determine if pairs of successive events involved the same troop or not. d = spatial distance between pairs of events, t = temporal distance between pairs of events. vmax = 1.26 km/h maximum travel speed rate, by excluding night-time and taking into account the time allocated to travel activity (Lewis et al., 2017).

K-means clustering method

The K-means clustering aims at partitioning the dataset into k different clusters, each of them represented by a centroid (center of the cluster) (Likas *et al.,* 2003). We used this method to regroup the most similar events by using the number of individuals per age-class. The following variables are retained for clustering : proportion of adults, proportion of subadults, proportion of juveniles, and spatial distances between events calculated with geographic coordinates. The proportion of males and the proportion of females variables are excluded for two reasons: first, the sex ratio variations between two troops are lower than age ratio variation

(Duane *et al.,* 2014; Davidge, 2010), second, the necessity of creating a variable "number of undetermined adults" during counts, introduces uncertainty in the males and females proportions. Moreover, we used spatial distances between stations because we know there is low overlap between different troops, then the probability to have the same troop between two close stations is higher than two distant stations. Also, we expect that clusters with a high density of events to occur at the same station, and to contain fewer events at other stations, with a highly similar troop structure.

To compare variables, data were normalized (Likas *et al.,* 2003). The algorithm chooses k (number of cluster) random events that become the first centroids, then assigns each event to the closest centroid, based on the Euclidean distance between the event and the centroid. For each update of a cluster, the new mean value of all data points in the cluster is calculated and the centroid moved to minimize the total intra-cluster variation (W):

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

where x_i is a data point that belongs to the cluster C_k and μ_k is the mean value of the events associated to C_k (Hartigan & Wong, 1979).

 If after the recalculation of the centroids, none of the cluster centroids has changed, the algorithm stops. The algorithm is used for 25 runs which differ by the initial centroid positions, to minimize the variation due to arbitrary placements (Ding & He, 2004; Likas *et al.,* 2003). The sum of the total intra-cluster variation $\sum_{k=1}^{k} W(C_k)$ must be as small as possible.

A fundamental step for the k-means algorithm is to determine the optimal number of clusters. The Elbow Method is one of the most popular methods to determine this optimal value of k (Syakur *et al.,* 2018). For many runs with different k (2 to 10), we calculate the sum of the total intra-clusters variation (ΣW). The higher is k, and the lowest is ΣW, however ΣW vary non-linearly with k. Consequently, there is an optimum k value for which adding another cluster does not significantly improve ΣW (Syakur *et al.,* 2018; Kaufman & Rousseeuw, 1990).

In addition to estimating the number of troops, the cluster analysis informs which events refer to which troop. This information is then compared with the two other approaches (Spatial / temporal analysis and individual identification) for cross-validation and used in a bayesian inference model to estimate each troop size.

K-means clustering were conducted with Rstudio software 1.2.5033 and the following packages: readr, dplyr, tidyverse, lubridate and ggplot2.

Capture Mark Recapture

One objective of our study was to test the CMR approach with unknown number of individuals to estimate baboon troop size for each event. With some exceptions, individual baboons cannot be recognized on unrelated camera trap photos. However, given the short time interval between successive photos making up one event, it becomes possible to follow one individual as it moves over the background and among the rest of the group. Only for the duration of the event, individuals are marked by an ID. A capture history can be built in this way for each individual, photos represent occasions (T). For each photo, an individual is noted 1 if it is visible and 0 if it is not. The approach aimed to estimate troops size by correcting our counts using the detection probability estimation ($\hat{p}$). Concerning events collected by videos, the too low number of videos for these events do not allow us to create a correct CMR history. In addition, individuals often alternate between visible and non-visible during a video. So, events recorded by videos are not used for Capture Mark Recapture. We created CMR stories for all individuals of 25 events among the 47 analysed, with Microsoft Excel.

Trap dependence

When we analyse the set of capture histories corresponding to one event, we know that the troop is present. We assume that the entire troop is moving as a group each time, but some individuals are out of the camera range. We assume that there is some chance (possibly different between age and sex classes) that a particular individual enters the camera range at any time (detection parameter p). Once visible, it has a high chance to remain visible for the next photo (detection parameter c >> p). When it goes out, we assume again that it reverts to the initial probability to enter the camera field (parameter p).

Hence, we use two different detection probabilities (p and c) depending on whether an individual was caught during the precedent occasion or not.

Data augmentation

From repeated captures of a given individual within an event, our model estimates detection probabilities $\hat{p}$ and $\hat{c}$ for each age class (adults, subadults, and juveniles) based on the number of photos where individuals are visible and on their potential round trip within the camera's field of view. In our case, the utilization of a Bayesian inference model is relevant mainly because we use data augmentation. In most circumstances, some individuals in the study area will be missed, then the counts (C) will be an underestimation of true troops size (N). So, in the

capture history, we added potential unobserved individuals with zero-only encounter histories. The number of individuals in the augmented data set must be much higher than the unknown troop size (N), M >> N. As baboon troops rarely exceed one hundred individuals (Hoffman & O'riain, 2012), we chose to add 150 unobserved individuals. We add a binary variable z, an indicator to determine if an empty capture history, added by data augmentation, correspond to a « real » individual. This indicator z has a Bernouilli prior distribution and Ω, the parameter of this distribution called inclusion probability, is estimable from the data.

<p style="text-align:center">Bayesian inference model</p>

As a prior: $\Omega \sim$ uniform $(0,1)$; $p_i \sim$ uniform $(0,1)$; $c_i \sim$ uniform $(0,1)$ and $z \sim$ bernouilli (Ω), where i represent one age class, 1: adults, 2: subadults, 3: juvenils.

We runned our model using Markov Chain Monte Carlo Method thanks to Winbugs software 1.4.3 and the Rstudio package R2Winbugs, with the following parameters : number of iterations in the sampling phase (ni) = 5000, thinning rate (nt) = 2 (we take in account one iteration on two), burnin lengh (nb) = 500 (we exclude the first 500 iterations) and number of chains (nc) = 3 (number of starting points) (Kéry & Schaub, 2012) for 31 events. By taking into account observations contain in CMR history, the model calculates for each iteration, an estimation of all parameters based on the previous estimations.

For each event, we obtain an estimation of the number of individuals, number of adults, number of subadults and number of juveniles at each iteration of the MCMC algorithm. We end up with a distribution of these quantities for each event. The mean of this distribution is the number of individuals n̂ estimated for the event. To regroup events that belong to the same troop and have a sample of the entire troop, we randomly collect one value in the estimation distribution of each event. We do it one thousand times (with replacement) to make one thousand sub-samples, representative of the troop. Then we calculate the means of each subsample to obtain one distribution with one thousand values and with a mean who will be the estimation N̂.

Results

Number of troops

<p style="text-align:center">Distance/Time between pair of successive events</p>

Based on the maximum travel speed, we identified a total of 3 pairs of successive events that involved different troops (Figure 5).

Concerning these pairs:

1) The first one concern events 5-2044 and 7-658a. It highlights the presence of two different troops, one occurring at station 5 (S5) and the other at S7, the same day at 1 hour 3 minutes and 53 seconds interval.

2) The second pair, 9-1567 and 7-674, shows two different troops, one at S9 and the second at S7, the same day, at 2 hours 53 minutes and 36 seconds interval.

3) The third pair, 7-676 and 12-491, indicates the presence of two different troops, one at S7 and the other at S12.
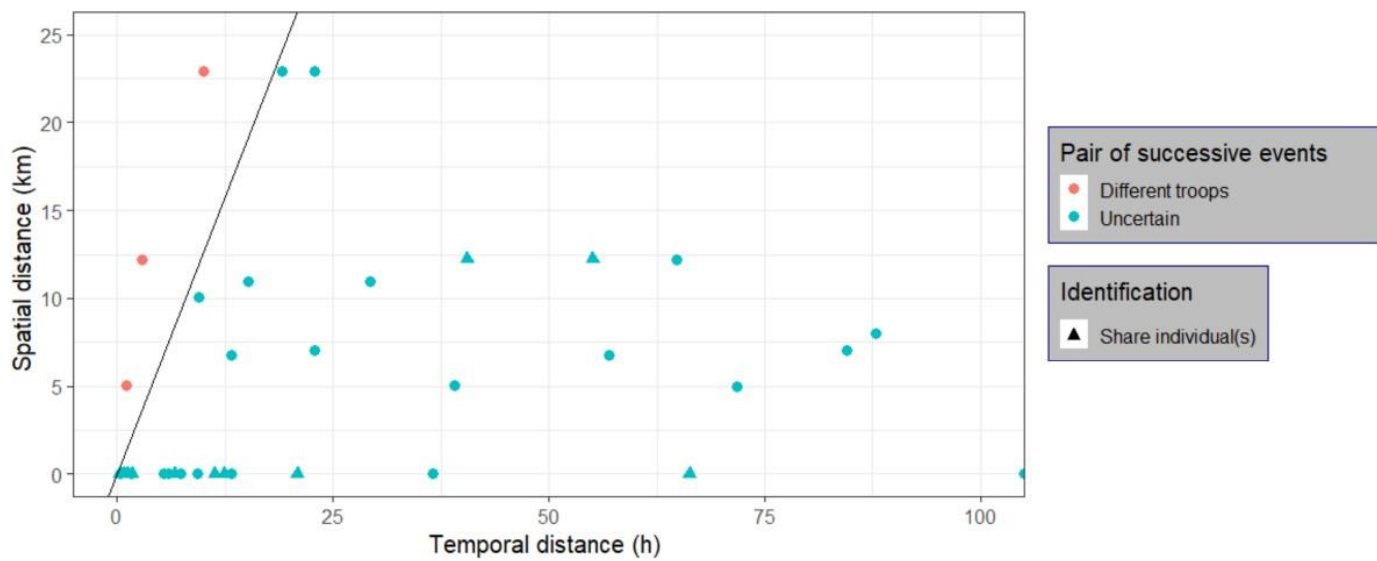

Figure 5: Spatial distance per temporal distance of each pair of successive events. Pairs that belong to two different troops are represented by red dots above the travel speed equation line. Blue dots represent pairs of successive events potentially involving the same troop and blue triangles pairs of successive events that shared individuals (hence captured the same troop). Vmax = 1,26 km/h (Lewis and O'Riain, 2017).

Individual identification

Among events that needed individual identification (blue points in Figure 5) we identified 14 pairs that share the same individuals. Individuals identified in events occurring at S5 are exclusively identified at S5. We obtained the same results for individuals captured at S8. However, it was noted that many events occurring at S7 share individuals with events also occurring at S9.

Overall, our age-class count results indicate 59.18 % of adults, 28.62 % of subadults and 12.2 % of juveniles. Among adults, 21.64 % were males, 33.76 % were females and the sex could not be determined for 44,6 % of them.

The results presented here resulted from clustering including the proportions of each age class for each event and the spatial distance however some clustering was also run using individual counts in each age group (Appendix 7). Based on the Elbow method we find an optimal number of 3 clusters (beyond this value, variations of ΣW are too low to justify the addition of another cluster). Variations of ΣW are presented in figure 6.
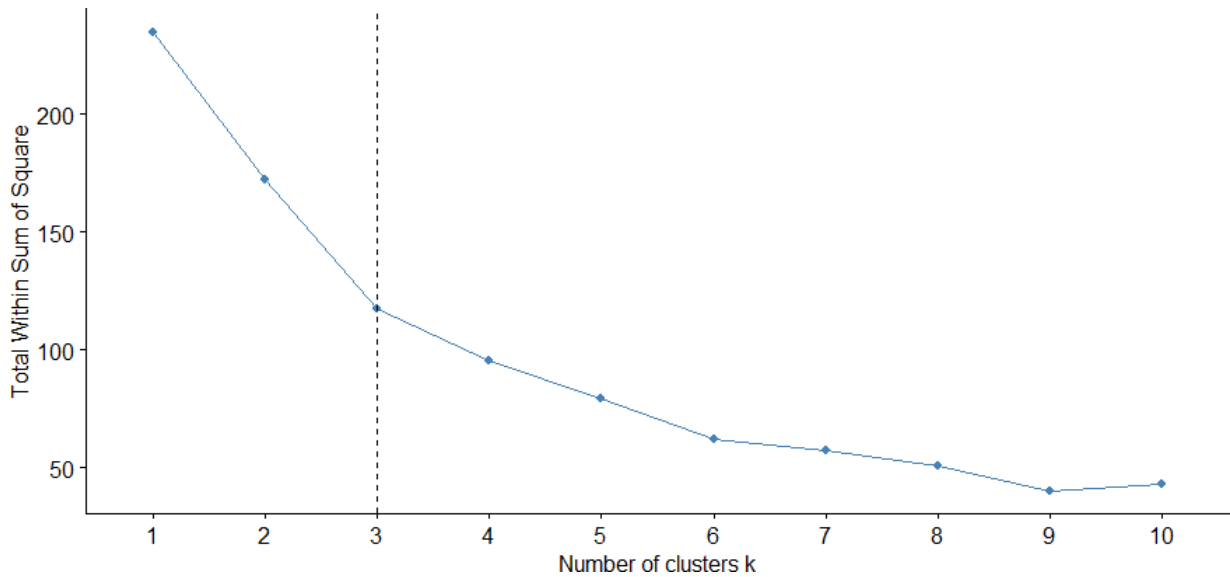

Figure 6: Results of the elbow method of the clustering on proportion data. The total within sum of square decreases with the k value. The "elbow" defines the optimal number of clusters, here k = 3.

The result of the k-means algorithm for k=3 is represented in Figure 7 (clusters for other k values are provided in Appendix 1). The first 2 dimensions capture 73.8 % of the variation. Proportion of subadults and spatial distances positively contribute to 28.8 % and 47.6% of component 1 and proportion of adults negatively contributes to 23.4 %. Proportion of adults negatively contributes to 19.7 % of component 2 and proportion of juveniles positively contributes to 68.0 %. In comparison to the other centroids, the centroid 1 is defined by the

highest proportion of adults and juveniles and the lowest proportion of subadults. The centroid 2 shows the lowest proportion of juveniles and the centroid 3 the highest proportion of subadults and spatial distance and the lowest proportion of adults. Moreover, the total intra-cluster variation (W) is higher in cluster 2 (42.83) and cluster 3 (41.76) than cluster 1 (32.70).
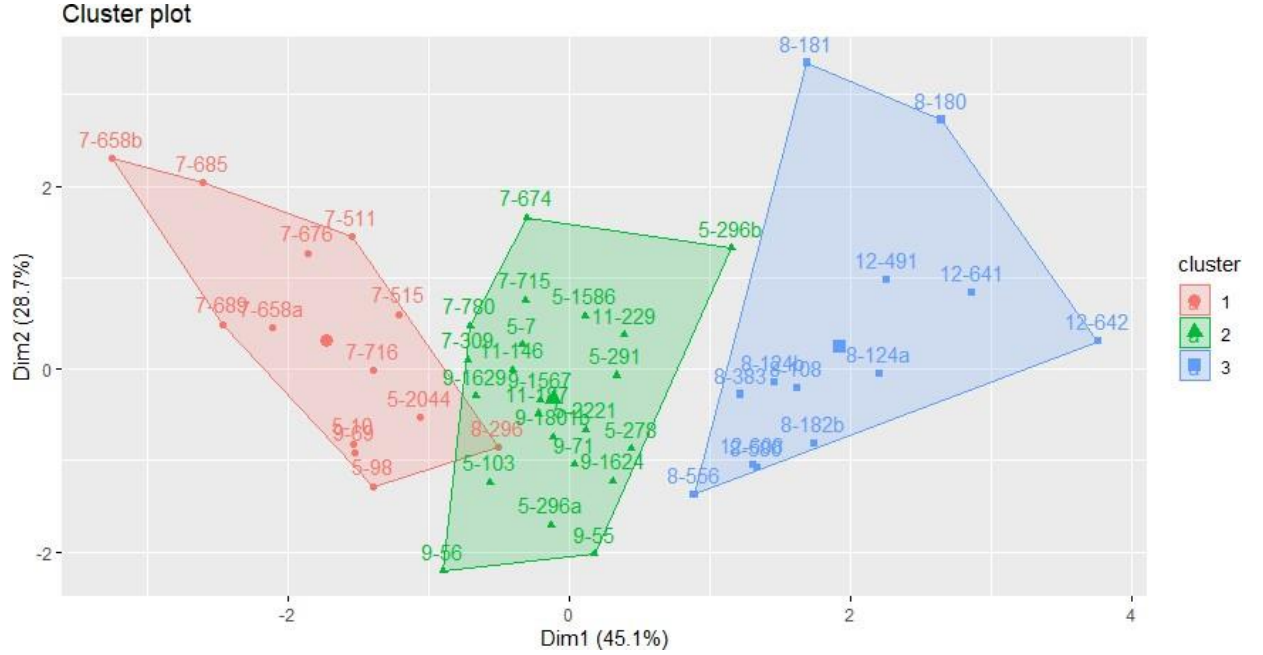

Figure 7: Dataset k-means clustering for k = 3. Each of the 48 events is assigned to the nearest cluster, according to Euclidean distance with centroid.

Clustering validation

Using the results from the spatial/temporal analysis and individuals' identification, we canidentify few mismatches with the clustering (Table 1):

1) Events 5-2044 and 7-658a events are two different troops but are both in cluster 3.
2) Same results for events 9-1567 and 7-674 in cluster 2
3) Events 5-7 and 5-10 share individuals but are in different clusters
4) 5-10 and 5-98 share individuals, hence they must be in the same cluster as 5-7
5) Events 7-674 and 7-676 share individuals but are in different clusters
6) Events 7-689 and 9-1624 share individuals but are in different clusters
7) Events 7-715 and 7-716 share individuals but are in different clusters

We use the combination of all results described in Table 1 to rectify the clusters and choose events that potentially belong to the same troop for the Bayesian inference. Events sharing

individuals in common are moved in same cluster, and pairs identified as different troops by spatial and temporal analysis are excluded if they are in the same cluster. Each of these rectified clusters represent one hypothetic troop. CMR histories of events that belong to the same "new clusters" / troops are used to estimate the parameters $\hat{N}$ a, $\hat{N}$ s and $\hat{N}$ j based on the Bayesian inference. Details regarding rectified clustering results are in appendix 3.

Table 1: Summary of our three approaches. Combination of these results are used to rectify the clustering.

Event 1	Event 2	Spatial/Temporal analysis	Identification	Clustering proportion
5-7	5-10	Uncertain	Individuals in common	Different cluster
5-10	5-98	Uncertain	Individuals in common	Same cluster
5-278	5-291	Uncertain	Individuals in common	Same
5-291	5-296a	Uncertain	Individuals in common	Same
5-296a	5-296b	Uncertain	Individuals in common	Same
8-108	8-124a	Uncertain	Individuals in common	Same
8-124a	8-124b	Uncertain	Individuals in common	Same
7-658a	7-658b	Uncertain	Individuals in common	Same
7-674	7-676	Uncertain	Individuals in common	Different
7-685	7-689	Uncertain	Individuals in common	Same
7-689	9-1624	Uncertain	Individuals in common	Different
9-1624	9-1629	Uncertain	Individuals in common	Same
9-1629	7-715	Uncertain	Individuals in common	Same
7-715	7-716	Uncertain	Individuals in common	Different
5-2044	7-658a	Different troops	Uncertain	Same
9-1567	7-674	Different troops	Uncertain	Same
7-676	12-491	Different troops	Uncertain	Different

Capture Mark Recapture

Based on the rectified clusters (which are variations of clusters showed in Figure 7) we can regroup some events and theirs CMR histories. We can assume that rectified cluster 1 correspond to troop A, cluster 2 to troop B and cluster 3 to troop C. Rectified clusters and CMR histories are presented in the table in Appendix 3.

The Bayesian models results suggest that the number of individuals in troop C is higher than troop B and troop A (Figure 8). Troop A show the highest proportions of adults and the lowest proportion of subadults. However, this troop also has the lowest proportion of juveniles. Moreover, as expected by clustering, troop C have the highest proportion of subadults and the lowest proportion of adults (Table 2).
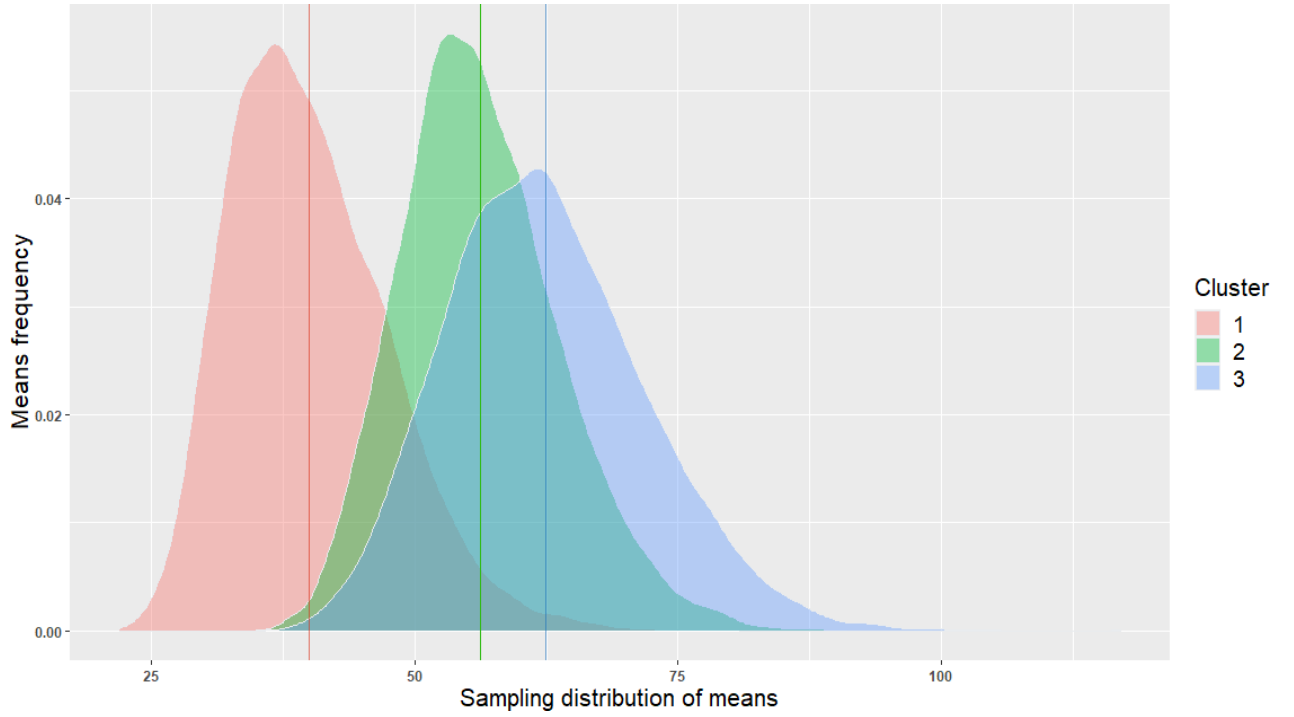
Figure 8: Distribution of the mean number of individuals in each troop. Vertical lines represent mean of the distributions $\hat{N}$ of each troop. Cluster colors are the same as rectified clusters Appendix 3, based on 8 events for cluster 1 (Troop A), 6 events for cluster 2 (Troop B) and 9 events for cluster 3 (Troop C).

Table 2: Count results (N: mean number of individuals, Na: mean number of adults, Ns: mean number of subadults, Nj: mean number of juveniles). $\hat{N}$ parameters are means of the sampling distribution representative of the entire troops and estimated thanks to Bayesian inference. Brackets contain distributions of each parameter. Percentages are the number of each age class estimation / number of individuals estimation (proportions of each age class).

Parameters \ Troops	Troop A	Troop B	Troop C
N	16.0 [10: 26]	21.2 [10: 38]	20.5 [11: 30]
Na	9.7 [6: 16]	12.2 [4: 21]	10.6 [4: 20]
Ns	3.3 [0: 7]	6.9 [3: 15]	7.3 [3: 14]
Nj	3.0 [1: 6]	2.0 [0: 5]	2.5 [1: 4]
$\hat{N}$	39.8 [22.7: 77.4]	56.2 [35.6: 96.6]	62.2 [36.6: 104.7]
$\hat{N}$ a	24.0 [13.1: 46.0]; 60.3 %	34.3 [20.6: 60.2]; 61.0 %	36.9 [20.1: 62.1]; 59.3 %
$\hat{N}$ s	11.1 [4.3: 23.4]; 27.9 %	17.4 [10.2: 28.7]; 31.0 %	20.2 [10.9: 33.4]; 32.5 %
$\hat{N}$ j	4.8 [2.9: 9.6]; 11.8 %	4.7 [2.2: 8.6]; 8.0 %	5.4 [2.3: 10.0]; 8.2 %

Discussion

This study has explored potential methods for using opportunistic data to estimate baboon population characteristics, i.e., the number of troops, troop size and composition.

During photos analysis, we observed that troops often travel in a line, on the road or a path. With this troop configuration, counting and following individuals was relatively easy. In addition to this behaviour, the low time interval between pictures allowed us to easily follow most individuals of a given event. However, for events showing a troop resting or foraging, only a few individuals would trigger the camera trap, hence giving a biased number of individuals. In these instances, counting individuals was a lot more challenging, and could induce more errors. Hence for counting individuals in troops, using frequently travelled path from Bird-eye view seems to be a preferable option.

Pictures quality was good enough to determine individual age-class. However, adult sex-determination was a lot more challenging. Same applied to individual identification, and as a result, our data may contain some misidentification. Physical characteristics reliable to identify individuals like face and ears (M. Bachman unpublished data; C. Guerbois pers.com) were not visible or unclear due to the Bird-eye view (Appendix 2). Identification is a lot easier with adults but because they tend to disperse to other troops (Hamilton, 1990), we avoided to base our assessment on male identification to determine troop similarity. However, individual identification could provide meaningful information in addition to other approaches as exposed here, in order to refine the number of troop from camera trap data. As a consequence a combination of baboon height (70cm) and Bird-eye view camera traps could be envisaged. Also, one issue that affects individuals' visibility is about the number of cameras. Moolman *et al.,* (2019) set two opposing cameras pointing trails at each station, which is usually enough (Tobler, 2008). But we had only access to one camera for each event. It means that sometimes we had the backside of the troop (which is perfect for sexing, but not for identification) and sometimes the front side (which is ideal for identification but not for sexing).

Correct cameras height and access to all cameras pictures could considerably improve counting and identification results, hence may reduce uncertainties about undetermined adults.

We assumed that assessing troop similarities from proportion of each age class would be more robust than using the actual number per class, which is more sensitive to the condition of the event. However, both analyses were conducted, and we found the same optimal number of troops (k = 3). Clustering and Bayesian inference results on counts per age class are presented in the Appendices and expose some similarities and dissimilarities with proportions results. Our two different clustering approaches share three mismatch and proportion results show one more mismatch than counts per age per class.

As explained in materials and methods, the rate of travel used for our analysis is based on the maximum rate of travel register for a troop observed by Lewis & O'Rian (2017), during Autumn (period with the highest travel activity time). Hence, this rate travel speed is higher than usually for a troop and it could be reassessed on the field to be more accurate and specific. This was only run of our sub-sample of our data, but this approach could be very useful to determine co-occurring troops in an area. Working from the pair of events that cannot co-occur could be an effective way for estimating the number of troops in a specific area.

As Table 1 highlights, count results are quite low compared to our model results. Due to picture analyses done by only one person, we cannot estimate an identification error and we can assume for sure that count results have errors. Moreover, recent study suggests that identification errors in camera-trap studies often result in systematic population overestimation (Johansson *et al.,* 2020). Hence it seems highly probable than some individuals are out of the camera's capture range during an event and encourage the utilization of data augmentation in our model.

The utilization of an opportunistic dataset has some costs and benefits. The survey design was not specifically designed for the focal species, consequently some events were unusable due to low individuals' visibility. In addition, pictures must have to be analyzed before being classified as unusable or not. This adds time to the analyses and explains why on the 201 events provided by the dataset, only 111 events had been processed and finally 47 are used on our analyses. Hence, we miss lots of information and the low accuracy of our results may due to our low number of events.

Additionally, to find variation in troop structure and dynamics (in order to assess the possibility to have long-term monitoring), it is necessary to compare two equals periods. However, our

data set only covers eleven-month that overlap between 2016 and 2017, and the number of usable events was too low to cut in half our study period.

Moreover, a large section of our study area, between S5 and S8, is not covered by our data. Taking into account the whole set of stations would help us to refine the number of troops and potential overlaps.

Baboon population characteristics

Our spatial/temporal analysis, identification as well as clustering results suggest the presence of a minimum of three troops occurring in the study area.

The troop A (Cluster 1) occurring at the north-west between S7 and S5 according to the k-means clustering. However, one pair of successive events indicate different troops between S7 and S5, which raises 4 hypotheses:

1) The travel speed of the troop is highly superior to the maximum rate of travel registered by Lewis & O'Rian (2017). Knowing that spatial distance between stations equals 5 kilometers and temporal distance equals 1 hours and 3 minutes, the troop must travel at 4.76 km/h during a period where travel activity equals 100 %, which is highly improbable.

2) The troop A split before the first event, then part of the troop was closer to the second event and able to trigger the camera.

3) Another troop with similar proportions of age-class occurs in this area.

4) Troop B possibly triggered one of the events. Moreover, based on clustering results, many events who occurred in S7 show a similar troop structure than events occurring in S5. Then, we cannot exclude the possibility that troop B travels to S7, even knowing that the range spatial overlap percentage between troops is low (Hoffman, 2011). This could be explained by the fact that lots of anthropogenic food sources are available at the vicinity of S5 (crop fields, camping sites and lodges) which could attract different troops.

The Cluster 2 suggest the existence of a troop roaming between S5, S7, S11 and S9. However, we exclusively recognized troop B at S5. Hence, it is in favour of hypothesis 4 with troop A occurring in S7 and troop B essentially in S5 and sometimes in S7. Then, we know that troop A and troop B cannot be the same. Because of both cluster 1 and cluster 2 contains some S5 events and there is an anthropogenic food source availability close to S5, it is highly probable

that troops overlap in S5. Also, few events occurring at S11 show similar troops structures with troops occurring at S5, then it is in favour of the presence of Troop B at north-west area.

Troop C is exclusively captured at S8. Thanks to spatial and temporal analysis, we know that troop occurring at S8 cannot be troop A or troop B. However, even if events that occurred S12 and S8 are in the same cluster, the absence of identification does not allow us to conclude about the presence of troop C at S12, and we must consider the possibility of another unknown troop.

We acquired complementary information from two sources monitoring baboons in the same area at the same period (Appendix 6). The Goudveld baboon project reports the presence of 4 troops in this area (M. Bachman unpublished data.) The first troop called "Pumba troop" of 69 individuals has been regularly observed close to S5 and a second called "West troop" of more than 80 individuals has been observed at the south and at west of S5. Two other troops, Arnie and Far East, are also found more towards the East. Another source confirms the co-occurrence of 2 troops south of S5 (Appendix 6).

Thanks to Goudveld baboon project we know that troops home range may overlap at S5. Moreover, the project mentioned the tendency of West troop to regularly subdivide in two troops, one of 25 individuals and one of approximately 60 individuals.

The Goudveld baboon project described a troop called "Arnie's Troop" of 62 individuals at south of S8 and S4 and who regularly travel close to S5 for anthropogenic ressources.

Also, a troop called "East Troop" (unknown troop size) occur even further to the east. Their home-range may correspond to two troops who respectively occur at S8 and S12. We do not have informations about home range overlap between these two troops.

Our hypotheses about our model cannot be assessed without more accurate clustering methods. On the four troops described by the Goudveld baboon project, our analysis only discriminates three, but we have a lack of information at far east of the study area. Moreover, given the results, it is highly probable than cluster 1 and cluster 2 are a combination of events triggered by troop A and troop B (West and Pumba's troop).

Mean count results (based on all events) showed an overall population structure of 0.64 male for 1 female and 0.8 adult for 1 immature (subadults and juveniles) the sex-ratio for our population is comparable to that observed by Duane *et al*., (2014) (0.15 – 0.80 male for 1

female) and Davidge (2010) (0.49 male of 1 female, means based on all troops mentioned by Davidge), but with 44,6 % of undetermined adults, our sex-ratio assessment is uncertain.

Our Bayesian inference model allowed us to estimate troop size and number of individuals of each age class, for each troop. Each troop size estimation is comparable to that observed by Davidge (2010) (20 – 85 individuals) and Lewis & O'Rian (2017) (9-109 individuals). However, the model shows a large distribution of parameters for each troop. Even by regrouping events (current state of the distribution indicate that results of each separated event are meaningless), the model shows a low accuracy. Also, considering the mean number of individuals counted of each troop (troop A: 16, troop B: 21.2, troop C: 20.5), the model more than doubles this number, which would mean that we miss lots of individuals during counts.

If troop A correspond to West troop, our model underestimates the number of individuals. We used two events of S9 (9-1624 and 9-1629) in the Bayesian inference model. These two events share individuals with events that occurred in S7. However, results can be explained by the tendency of West troop to subdivided in one large troop and one smaller troop. Then we possibly combine events that show the large troop with events that show the smaller troop and hence estimate a troop size between the two subtroops of smaller troop size. Nevertheless, troop overlap is usually low and we have no field information about the presence of West troop at S9. S5 seems to be an exception due to the high presence of anthropogenic resources that attract all troops (Fehlmann *et al.,* 2017a). Then, we cannot exclude the possibility that we did mistake during identification. Additionally, these two events have closer troop structure to events in S5 than events in S7.

Our troop B has an estimated size of 56,3 individuals while the Pumba's troop size is 69 individuals. Moreover, Bayesian inference is done on a majority of events occurring at S5 who share individuals and two events at S9 where Pumba's troop may occur according to field observations. S5 is a key station that overlap with the three troops. Hence, we cannot exclude the possibility that we did mistake during identification and clustering and then, combine events triggered by Pumba's troop with events triggered by another troop. Then our model underestimates the number of individuals for Troop B.

The troop C estimation of 62,2 individuals highly correspond to Arnie's Troop size with 62 individuals. Moreover, except one event which occur at S12, all events used for Bayesian

inference occur at S8. Due to Goudveld baboon project observations, it is highly probable that events occur at S8 correspond to Arnie's troop. Then, the estimation seems to indicate that with a correct clustering, our Bayesian model is reliable.

Perspectives:

Constrained K-means method

Our clustering method is an unsupervised method. It means that the clustering algorithms discover hidden patterns in data and study the intrinsic structure of the data without other information (Wagstaff *et al.,* 2001). However, spatial/temporal analysis and identification allowed us to attribute with some events to the right cluster after correction. But this method, regarding results and field observation of the the Goudveld baboon project have defaults and used an alternative method like Constrained k-means, in order to combine our three approaches, could improve our rectified clusters.

Constrained k-means is also based on the notion of similarity in the data set, but the algorithm integrates background information. The algorithm is an alternative version of k-means clustering used for this study, but it considers two types of pairs:

1) Must-link who constraints that two events must be in the same cluster
2) Cannot-link who constraints that two events must not be placed in the same cluster.

Must link pairs should be used for events that share individuals and Cannot link for the three events which show with certainty two different troops. Additionally, clustering results with the constrained method seem to be more accurate than the unsupervised k-means method (Wagstaff *et al.,* 2001). Hence, this new approach should be considered for future analysis, aims to have more accurate rectified clusters.

Scale change

The main challenge in our study, it is the combination of k-means clustering and opportunistic dataset. The utilisation of opportunistic dataset induces a loss of data and more time-consuming analysis, while k-means clustering requires a lot of data to be more accurate. Hence, it could be considered to not use this method. According Goudveld baboon project field observations, there is high overlap between troops in the study area, while our method is ideally suited for low or inexistent overlap.

There are many advantages to use camera trap method, especially if the purpose is to long term monitor a population. Also, it has been proved that wild animals (and especially primates who tend to fly when an observer approach) react strongly in response to observer presence, and one solution is the utilization of camera traps (Crofoot *et al.,* 2010). This is a non-invasive method and photographic sampling eliminates interactions between observers and baboons, possibly reducing stress and behavioural changes (Gerber *et al.,* 2014). Another huge advantage of camera traps compared to other survey techniques is the potential long-term cost saving opportunities. Rovero and Marshall (2009) estimated that excluding the cost of purchasing cameras, camera trapping was cheaper than standard trapping who need more human resources. Additionally, even if camera-trapping does not generate information as accurate as most field observations, they have promising potential to detect variation (McCarthy *et al*., 2018). Moreover, recycling data by using opportunistic dataset limit the financial cost of buying and maintain camera-traps.

Conclusion

Using an opportunistic dataset kindly provided by the South African National Parks*,* we tested different approaches to assess the number of troops of chacma baboons in a section of the Garden Route National Park and their size from over a 1-year period dataset. Our results highlight the presence of minimum three troops, in the study area.

These troops show a number of individuals and a structure comparable of what researchers usually observed for this species. However, field observations by the Goudveld baboon project and residents (Appendix 6) show the presence of fours troops and we assess the presence of three troops, thanks to fields observation. We can conclude than our three approaches are complementary and perfectible but seem reliable as show by the case of troop C. However, using an opportunistic data seems to become a loss of data and time. Utilization of a dataset with a more baboons specific survey design, would have enabled us to have more useful data that positively impact our clustering and Bayesian inference model results.

A great work remains to be done to confirm our results, and some method improvement like constrained clustering could be tested. However, we succeed in extracting information from a non-adapted dataset taken among the large number of data collected by camera traps and our method may be a basis for future improvement, aiming at estimating unmarked population characteristics.

Bibliography

Codron, D., Lee-Thorp, J.A., Sponheimer, M., de Ruiter, D., and Codron, J. (2006). Inter- and intrahabitat dietary variability of chacma baboons (Papio ursinus) in South African savannas based on fecal δ13C, δ15N, and %N. *American Journal of Physical Anthropology.* 129, 204–214. https://doi.org/10.1002/ajpa.20253

Crofoot, M.C., Lambert, T.D., Kays, R., and Wikelski, M.C. (2010). Does watching a monkey change its behaviour? Quantifying observer effects in habituated wild primates using automated radiotelemetry. *Animal Behaviour* 80, 475–480. https://doi.org/10.1016/j.anbehav.2010.06.006

Davidge, C. (2010). Ecology of baboons (*Papio ursinus*) at Cape point. *Zoologica Africana,* 13, 22. https://doi.org/10.1080/00445096.1978.11447633

Després-Einspenner, M.-L., Howe, E.J., Drapeau, P., and Kühl, H.S. (2017). An empirical evaluation of camera trapping and spatially explicit capture-recapture models for estimating chimpanzee density. *American Journal of Primatology* 79, e22647. https://doi.org/10.1002/ajp.22647.

Ding, C., and He, X. (2004). K -means clustering via principal component analysis. Association for Computing Machinary Press. *In the Proceedings of the Twenty-First International Conference on Machine Learning*, Banf, Alberta, Canada, p. 29. https://doi.org/10.1145/1015330.1015408

Duane A. Schlitter, Adam W. Ferguson, and Molly M. McDonough (2014). Mammals of Africa (Vol. I-VI.), *Journal of Mammalogy* 95(6), 1299-1303. https://doi.org/10.1644/14-MAMM-R-122.

Efford, M.G., and Fewster, R.M. (2013). Estimating population size by spatially explicit capture-recapture. *Oikos* 122, 918–928. https://doi.org/10.1111/j.1600-0706.2012.20440.x

Ellis, E.C., Klein Goldewijk, K., Siebert, S., Lightman, D., and Ramankutty, N. (2010). Anthropogenic transformation of the biomes, 1700 to 2000: Anthropogenic transformation of

the biomes. *Global Ecology and Biogeography* 19, 589–606. https://doi. org/10.1111/j.1466-8238.2010.00540.x

Fehlmann, G., O'Riain, M.J., Kerr-Smith, C., Hailes, S., Luckman, A., Shepard, E.L.C., and King, A.J. (2017a). Extreme behavioural shifts by baboons exploiting risky, resource-rich, human-modified environments. *Scientific Reports*, 7, 15057. https://doi.org/10.1038/s41598-017-14871-2

Fehlmann, G., O'Riain, M.J., Kerr-Smith, C., and King, A.J. (2017b). Adaptive space use by baboons (*Papio ursinus*) in response to management interventions in a human-changed landscape. *Animal Conservation*, *20*, 101–109. https://doi.org/10.1111/acv.12293.

Foley, J.A. (2005). Global Consequences of Land Use. *Science,* 309, 570–574. https://doi.org/10.1126/science.1111772

Gerber, B.D., Williams, P.J., and Bailey, L.L. (2014). Primates and Cameras: Noninvasive Sampling to Make Population-Level Inferences While Accounting for Imperfect Detection. *International Journal of Primatology* 35, 841–858. https://doi.org/10.1007/s10764-014-9761-9

Gilbert, N.A., Clare, J.D.J., Stenglein, J.L., and Zuckerberg, B. (2020). Abundance estimation of unmarked animals based on camera-trap data. *Conservation biology: the journal of the Society for Conservation Biology*. https://doi.org/10.1111/cobi.13517

Hamilton, W.J., and Bulger, J.B. (1990). Natal male baboon rank rises and successful challenges to resident alpha males. *Behavioral Ecology and Sociobiology*, 26, 357–362. https://doi.org/10.1007/BF00171102

Hartigan, J.A., and Wong, M.A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28, 100. https://doi.org/10.2307/2346830

Heydon, M.J., Wilson, C.J., and Tew, T. (2010). Wildlife conflict resolution: a review of problems, solutions and regulation in England. *Wildlife research*. 37, 731. https://doi.org/10.1071/WR10006

Hoffman, T. S., & O'Riain, M. J. (2012). Troop Size and Human-Modified Habitat Affect the Ranging Patterns of a Chacma Baboon Population in the Cape Peninsula, South Africa. *American Journal of Primatology*, 74 (9), 853–863. https://doi.org/10.1002/ajp.22040

Hoffman, T.S. (2011). The spatial ecology of chacma baboons (*Papio ursinus*) in the Cape Peninsula, South Africa. PhD diss., University of Cape Town.

Johnson, C.A., Raubenheimer, D., Rothman, J.M., Clarke, D., and Swedell, L. (2013). 30 Days in the Life: Daily Nutrient Balancing in a Wild Chacma Baboon. *PLoS ONE* 8, e70383. https://doi.org/10.1371/journal.pone.0070383

Kaufman, L and Rousseeuw, P.J ,1990. Clustering Large Applications (Program CLARA). eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.), *Wiley Series in Probability and Statistics*, pp. 126–163. https://doi.org/10.1002/9780470316801.ch3

Kalle, R., Ramesh, T., Qureshi, Q., and Sankar, K. (2011). Density of tiger and leopard in a tropical deciduous forest of Mudumalai Tiger Reserve, southern India, as estimated using photographic capture–recapture sampling. *Acta Theriologica,* 56, 335–342. https://doi.org/10.1007/s13364-011-0038-9.

Kays, R., Kranstauber, B., Jansen, P., Carbone, C., Rowcliffe, M., Fountain, T., and Tilak, S. (2009). Camera traps as sensor networks for monitoring animal communities. *Institute of Electrical and Electronics Engineers 34th Conference on Local Computer Networks*, Zurich, Switzerland, pp. 811–818. https://doi.org/10.1109/LCN.2009.5355007

Kellert, S.R., Black, M., Rush, C.R., and Bath, A.J. (1996). Human Culture and Large Carnivore Conservation in North America. *Conservation Biology 10*, 977–990. https://doi.org/10.1046/j.1523-1739.1996.10040977.x

Kery M, Schaub M. 2012 Bayesian population analysis using WinBUGS: a hierarchical perspective. New York, NY: Academic Press.

Kummer, H. (1984). From laboratory to desert and back : A social system of hamadryas baboons. *Animal Behaviour*, 32(4), 965–971. https://doi.org/10.1016/S0003-3472(84)80208-0

Lewis, M.C., and O'Riain, M.J. (2017). Foraging Profile, Activity Budget and Spatial Ecology of Exclusively Natural-Foraging Chacma Baboons (Papio ursinus) on the Cape Peninsula, South Africa. *International Journal of Primatology*, 38, 751–779. https://doi.org/10.1007/s10764-017-9978-5

Likas, A., Vlassis, N., and J. Verbeek, J. (2003). The global k-means clustering algorithm. Pattern Recognition *36*, 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2

Pebsworth, P.A., Morgan, H.R., and Huffman, M.A. (2012). Evaluating home range techniques: use of Global Positioning System (GPS) collar data from chacma baboons. *Primates*, 53, 345–355. https://doi.org/10.1007/s10329-012-0307-5

Pebsworth, P.A., and LaFleur, M. (2014). Advancing Primate Research and Conservation Through the Use of Camera Traps: Introduction to the Special Issue. *International Journal of Primatology,* 35, 825–840. https://doi.org/10.1007/s10764-014-9802-4

McCarthy, M.S., Després-Einspenner, M.-L., Samuni, L., Mundry, R., Lemoine, S., Preis, A., Wittig, R.M., Boesch, C., and Kühl, H.S. (2018). An assessment of the efficacy of camera traps for studying demographic composition and variation in chimpanzees (*Pan troglodytes*). Am J Primatol *80*, e22904. https://doi.org/10.1002/ajp.22904

Moolman, L., Morney, M.A. de, Ferreira, S.M., Ganswindt, A., Poole, J.H., and Kerley, G.I.H. (2019). And Then There was One: A Camera Trap Survey of the Declining Population of African Elephants in Knysna, South Africa. *African Journal of Wildlife Research* ,49, 16 - 26 https://doi.org/10.3957/056.049.0016

Noser, R., and Byrne, R.W. (2007). Travel routes and planning of visits to out-of-sight resources in wild chacma baboons, Papio ursinus. *Animal Behaviour* 73, 257–266. https://doi.org/10.1016/j.anbehav.2006.04.012.

Redpath, S.M., Young, J., Evely, A., Adams, W.M., Sutherland, W.J., Whitehouse, A., Amar, A., Lambert, R.A., Linnell, J.D.C., Watt, A., et al. (2013). Understanding and managing conservation conflicts. *Trends in Ecology & Evolution*, 28, 100–109. https://doi.org/10.1016/j.tree.2012.08.021

Rovero, F., and Marshall, A.R. (2009). Camera trapping photographic rate as an index of density in forest ungulates. *Journal of Applied Ecology,* 46, 1011–1017. https://doi.org/10.1111/j.1365-2664.2009.01705.x

Syakur, M.A., Khotimah, B.K., Rochman, E.M.S., and Satoto, B.D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering,* 336, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Tobler, M.W., Carrillo-Percastegui, S.E., Leite Pitman, R., Mares, R., and Powell, G. (2008). An evaluation of camera traps for inventorying large- and medium-sized terrestrial rainforest mammals. *Animal Conservation* 11, 169–178. https://doi.org/10.1111/j.1469 1795.2008.00169.x

Trolliet, F., Huynen, M.-C., Vermeulen, C., and Hambuckers, A. (2014). Use of camera traps for wildlife studies. A review, Biotechnology, Agronomy, Society and Environment. 18 (3), 446-454. http://hdl.handle.net/2268/165854

Wagstaff, K., Cardie, C., Rogers, S., and Schroedl, S. (2001) Constrained K-means Clustering with Background Knowledge. *Proceedings of the Eighteenth International Conference on Machine Learning*, p. 577–584.

Wallace GE, Hill CM (2012) Crop Damage by Primates: Quantifying the Key Parameters of Crop-Raiding Events. *PLoS ONE 7*(10): e46636. https://doi.org/10.1371/journal.pone.0046636

Appendix 1: Clustering on the dataset for k = 2 to 5.
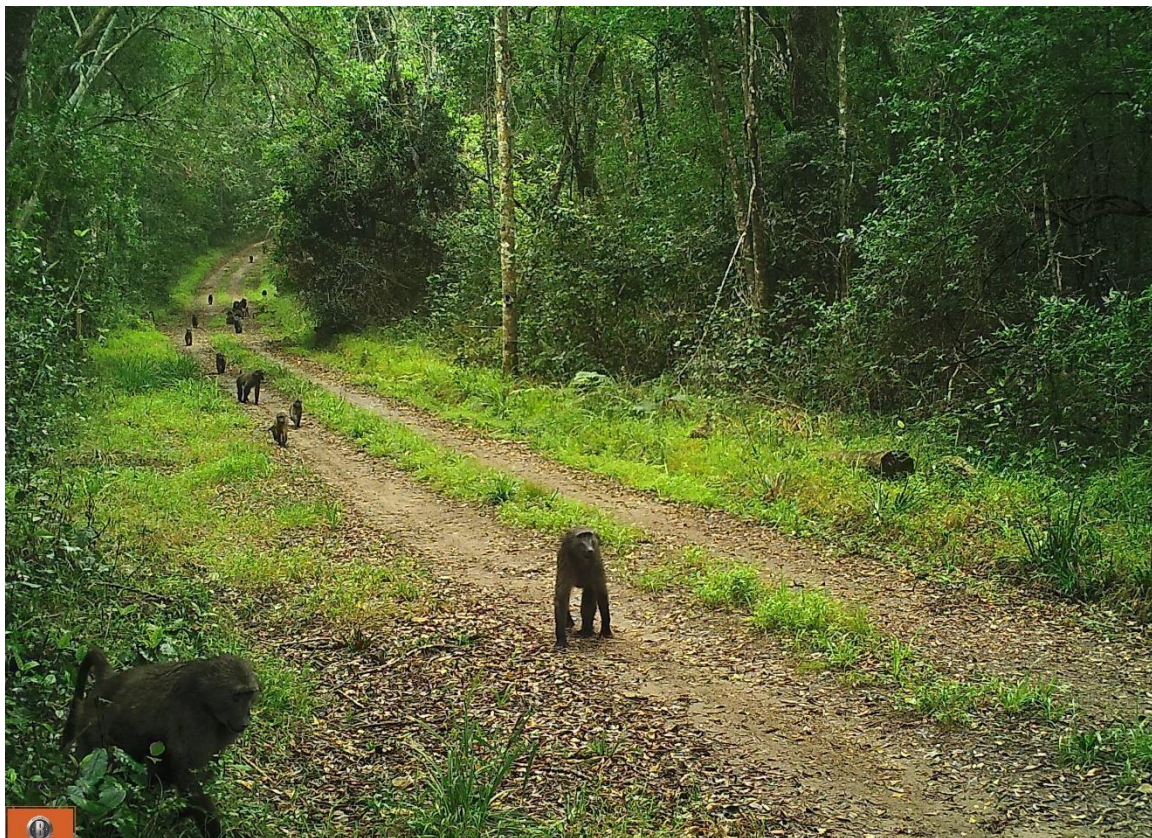
*Appendix 2: Pictures of two different events that occurred station 7 (up) and station 5 (down).*

Appendix 3: Rectified clusters. Red events are excluded. Orange events contains CMR histories used to estimate troop A parameters with our Bayesian inference model. Green events correspond to troop B and blue events to troop C.

Event	Rectified cluster		
5-7	2	7-511	1
5-10	2	7-515	1
5-98	2	8-383	3
5-103	2	5-2044	Excluded
5-278	2	7-658a	Excluded
5-291	2	7-658b	1
5-296a	2	9-1567	2
5-296b	2	7-674	1
9-55	2	7-676	1
9-56	2	12-491	3
9-69	1	7-685	1
9-71	2	7-689	1
11-146	2	9-1624	1
8-108	3	9-1629	1
8-124a	3	7-715	1
8-124b	3	7-716	1
8-180	3	5-2221	2
8-181	3	12-608	3
8-182b	3	8-556	3
11-187	2	9-1801b	2
7-309	2	8-580	3
11-229	2	12-641	3
8-296	3	12-642	3
5-1586	2	7-780	2
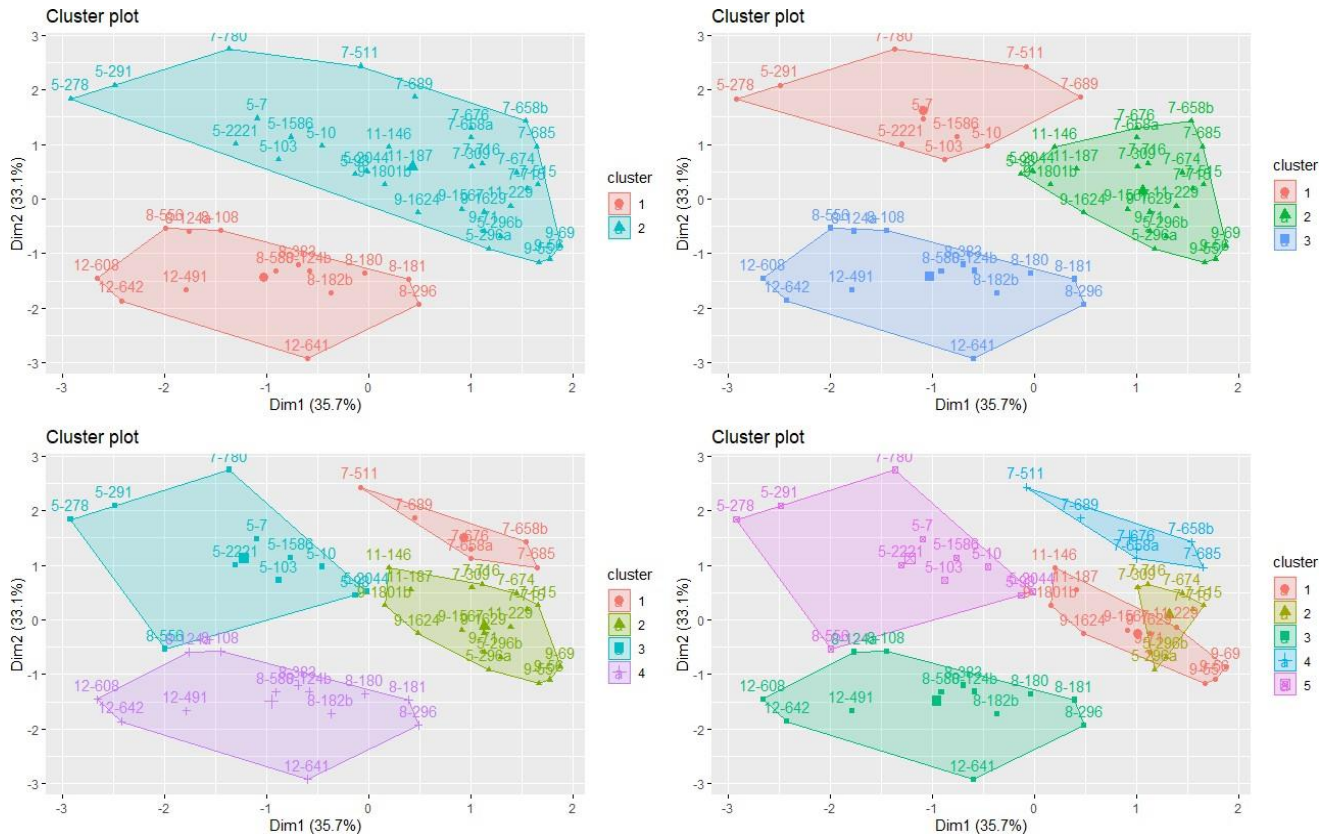
Appendix 4: Clustering on the dataset for k = 2 to 5 using number of individuals instead proportions.

Appendix 5: Counting results (N: mean number of individuals, Na: mean number of adults, Ns: mean number of subadults, Nj: mean number of juveniles).

$\hat{N}$ *parameters are means of the sampling distribution representative of the entire troops and estimated thanks to Bayesian inference. Brackets contain distributions of each parameter. We obtain these results with rectified clusters by using number of individuals instead of proportions like main report.*

Parameters \ Troops	Troop A	Troop B	Troop C
N	16.3 [12: 23]	28.4 [22: 38]	17.3 [11: 24]
Na	9.8 [6: 16]	17.2 [16: 21]	9 [4: 12]
Ns	3.8 [1: 7]	9.2 [5: 15]	5.8 [1: 10]
Nj	2.7 [1: 4]	2 [1: 4]	2.5 [1: 4]
$\hat{N}$	45.0 [20.3: 119.0]	77.0 [47.7: 118.9]	54.8 [28.5: 108.0]
$\hat{N}$ a	26.8 [14.0: 69.2]	44.2 [25.3: 69.2]	31.0 [15.3: 57.5]
$\hat{N}$ s	12.0 [4.7: 39.8]	26.2 [15.3: 39.8]	17.7 [9.0: 33.0]
$\hat{N}$ j	6.2 [2.2: 12.5]	6.7 [3.3: 12.2]	5.9 [2.2: 12.3]

A

B

Appendix 6: Ground-truthing information provided by (A) Madison Bachman from the Gouveld project and (B) a local resident, suggesting the existence of at least 4 troops in our area. The circles in B seem to correspond to the 2 groups (Pumba and Arnie) monitored by the Gouveld project.

Appendix 7: Summary of our three approaches. Combination of this results are used to rectified clustering by using number of each age class individuals instead of proportions.

Event 1	Event 2	Spatial/Temporal analysis	Clustering	Identification
5-291	5-296a	Uncertain	Different	Individuals in common
5-7	5-10	Uncertain	Same	Individuals in common
5-10	5-98	Uncertain	Different	Individuals in common
5-278	5-291	Uncertain	Same	Individuals in common
5-296a	5-296b	Uncertain	Same	Individuals in common
8-108	8-124a	Uncertain	Same	Individuals in common
8-124a	8-124b	Uncertain	Same	Individuals in common
7-658a	7-658b	Uncertain	Same	Individuals in common
7-674	7-676	Uncertain	Same	Individuals in common
7-685	7-689	Uncertain	Different	Individuals in common
7-689	9-1624	Uncertain	Different	Individuals in common
9-1624	9-1629	Uncertain	Same	Individuals in common
9-1629	7-715	Uncertain	Same	Individuals in common
7-715	7-716	Uncertain	Same	Individuals in common
5-2044	7-658a	Different troops	Different	Uncertain
9-1567	7-674	Different troops	Same	Uncertain
7-676	12-491	Different troops	Different	Uncertain