DETERMINING THE OPTIMAL REVISIT TIME FOR PRIMATE SPATIAL DATA COLLECTION

MARCUS JACK DOSTIE
Bachelor of Science, University of Lethbridge, 2014

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTERS OF SCIENCE

Department of Geography
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

DETERMINING THE OPTIMAL REVISIT TIME FOR PRIMATE SPATIAL DATA COLLECTION

MARCUS J. DOSTIE

Approved:

Signature	Rank	Highest Degree	Date
Louise Barrett Co-Supervisor			
Stefan Kienzle Co-Supervisor			
Peter Henzi Committee Member			
Shawn Bubel Committee Member			
_____ Chair, Thesis Examination Committee			

ABSTRACT

A major limiting factor in the study of animal movement is the ability to collect affordable high resolution spatial data. Tracking data often requires interpolation; this research will focus on determining the amount of estimation error that can be expected at different sampling frequencies. This will allow researchers to find a sampling frequency which best captures an animal's movement, while also allowing for the application of margins of error to previously collected data.

High resolution tracking data for individual adult vervet monkeys and chacma baboons from two sites in South Africa was used for this analysis. Their paths were resampled at increasing intervals from one minute up to sixty minutes. The increased temporal simplification of animal tracks caused by growing revisit times was anticipated to cause multiple changes to the overall path geometry which allowed for the quantification of the resulting changes in the resampled paths.

ACKNOWLEDGEMENTS

I will start my thanks where I first stepped onto this path, to my comrade Dave Horness. We started our careers in the military together and fifteen years later he inspired me with a couple of books and his own example to pursue a university degree. Once on this path, I would have never succeeded without the patience and support of my family. Tina, Jeff, Amanda, and Cheyanne, you all endured countless hours without my presence when duty called, then when I was at home and I asked you to give me solitude to study you did so without question. To my parents I thank you for your unconditional support throughout the years no matter how far away my choices took me from you.

There have been several Professors who helped me get to this point. Professor Stefan Kienzle who helped me get a position as a Research Assistant which ended up turning into this research and Master's Degree, I have benefited from your insight into spatial analysis. A special thank-you needs to be extended to Professors Louise Barrett and Peter Henzi who have taken me into their lab, allowing me to conduct this research and providing me with the opportunity to carry out field research at a private game reserve in South Africa; I will always appreciate the opportunities you have both given me. Professor Shawn Bubel has always been willing to give me her insight, advice and time. Her open door policy has saved me many stressful hours, thank you for your time and sound advice over the years. And to Professors Craig Coburn, Kevin McGeough, Tom Johnston, and Ivan Townshend who have all taught me to think critically about my work.

Thanks needs to be given to all the people who provided or collected data for this project: Eric Matlock and Jonathan Jarrett who followed individual vervets for me for ten

to eleven hours in the African sun, and Dr. Tony Phelps who took time out of his own research to follow baboons for me. And finally to Dr. Perry Clarke, Dr. Richard McFarland, Alena Matlock, Brittany Thomas, and Nic Ducheminsky who all collected data for the lab which made me start to think about the questions I ask in this research; I may not have come up with this research without looking at your datasets.

TABLE OF CONTENTS

LIST OF FORMULAS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DMS = Degree, Minute, Seconds
GME = Geospatial Modelling Environment
GPS = Global Positioning System (refers to handheld receiver)
SED = Synchronized Euclidean Distance
UTM = Universal Transverse Mercator
VBA = Visual Basic Application
WGS = World Geographic System

“When you want to know how things really work, study them when they're coming apart.”

— William Gibson, *Zero History* (2010)

CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

Since researchers started tracking animals of all sizes, and long before the advent of Global Position System (GPS) tracking, there has been an intrinsic problem with the data, which is how best to connect the points that comprise the tracking data set. Whether recording animal movement with a pencil, map, and compass, a handheld GPS, or using remote tracking collars, all such movement data are collected as time series, with discrete records of position that vary from seconds to minutes or days. Positional fixes alone have limited utility; it is the connection of multiple points that enables them to represent the spatial behaviour of an animal or a group of animals.

Studies that focus on the relationship between animals and their environment, such as the Crist et al. (1992) study of *eleodes* beetles' movement in shortgrass prairie, typically require only a positional fix for the group or of single individuals being studied. Data at the level of a group of gregarious animals do not allow the analysis of internal mechanisms (individual-level data) that may underpin estimates of group-level behaviour. Recent work into group decision-making and the aggregation patterns of animals, such as the movement of pigeon flocks (Nagy et al. 2010, Freeman et al. 2011), collective motion of fish schools (Herbert-Read et al. 2011), subgrouping and fission-fusion dynamics (Aureli et al. 2008, Aureli et al. 2012), social coordination (Barrett and Henzi 2013), or the democratic/despotic decision making model as seen in primate foraging strategy studies (King et al. 2008), all highlight the need for individual data. In all the studies above, group decisions is thought to evolve from collective decisions or behaviours which are best studied by taking the individual locations for all members of

the group and analysing their interactions with each other across space and time. Collective decision making can be thought of as the inter-individual interactions by which individuals can influence the behavior of other group members (Couzin and Krause 2003).

In order to analyze the interplay between animals living in social groups and determine possible interdependencies that may manifest as group level behaviour, it is not only necessary to have positional data on all the animals involved in a given social interaction, it is often also necessary that their positions be temporally synchronized. High resolution individual paths are difficult to capture without either dedicating one researcher to each individual animal within the social group, or without investing in tracking collars. Tracking collars can only collect high resolution data for a finite period of time, which is limited directly by the size of the animal which in turn limits the size of the tracking collar battery and consequently its life. For example, pigeons equipped with 16g loggers can collect high resolution data (0.2 sec) for only 1h40m (Nagy et al. 2010). Failures of the breakaway unit or the GPS on the collar are an added risk which cannot be mitigated during the study due to the lack of human interaction (Bowman et al. 2000), while the lack of a human dimension to the data collection also excludes the gathering of any observational data. Following each individual with handheld GPS units is limited by manpower and the study species tolerance to multiple observers. Excessive disturbance by human observers could impede the opportunity to record natural behaviour, and would likely invalidate the data (Rasmussen 1991, Williamson and Feistner 2003). Due to the constraints limiting the collection of high resolution positional data for all individuals at the same time, it is often more effective to interpolate the spatiotemporal positions of

animals from discrete data collected by one or a few observers (Heezen and Tester 1967, Harris et al. 1990, Reynolds and Laundre 1990, Breitenmoser et al. 1992, Johnson et al. 2002). Data could be collected in the latter by moving among the subjects in a cyclic pattern capturing points as often as possible while minimizing the time between each new cycle of data collection (Figure 1). This will produce a data cloud of points over the study area that is later parsed out to create individual paths.

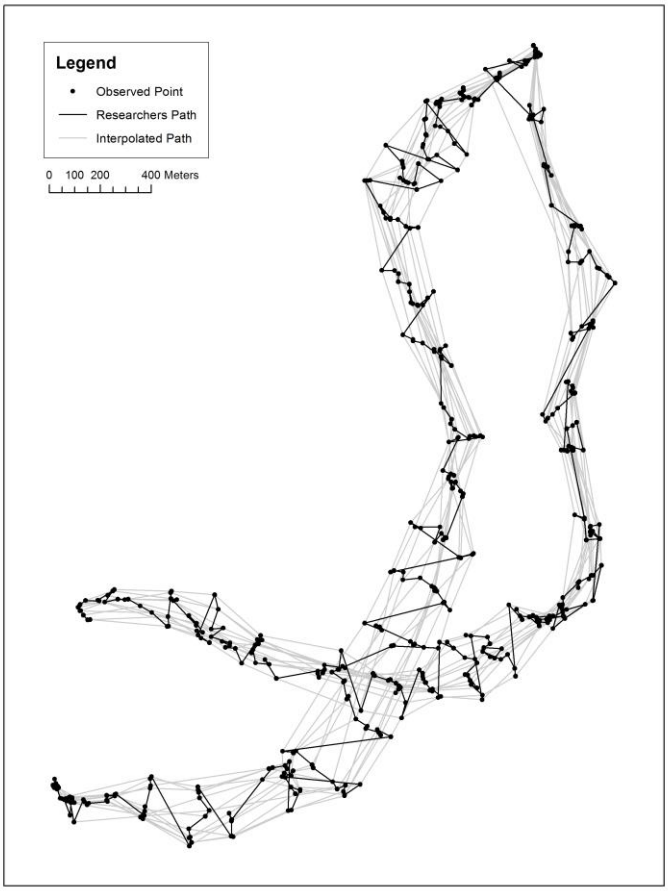

Figure 1: An example of data collection on the sequential spatial position of individual adult members of a baboon group (N = 13) using a handheld GPS data logger (Data collected at De Hoop Nature Reserve by Dr. Parry Clarke on March 27th 2007). The solid black line represents the researcher while a grey line represents baboons by sequentially connecting their observed positions.

Such data collection may well be an efficient compromise between disturbing the animals and ensuring that data are returned for all subjects without the need to re-capture or replace GPS collars, but it suffers from one major flaw, which is that with every set of

points which represent an animal's trajectory there is a coinciding gap in space and time for which there is no record. This gap can vary from seconds and minutes, to hours or even days, and while in many cases one can estimate what is happening during these gaps simply by connecting the points, the question remains: "How accurate are these estimations?" This, in turn, leads to another question: "Why not simply use the actual points which have been collected to analyze the interactions between animals, when interpolation will invariably introduce further error?" As mentioned above it is highly unlikely that a cluster of points which represents a group of animals will be synchronized (Figure 2), which means that even if the original points collected in the field are used to conduct an analysis, there is still temporal error being introduced. Linear interpolation of tracking data is one way to address the problem of uneven sampling (Tremblay et al. 2006).

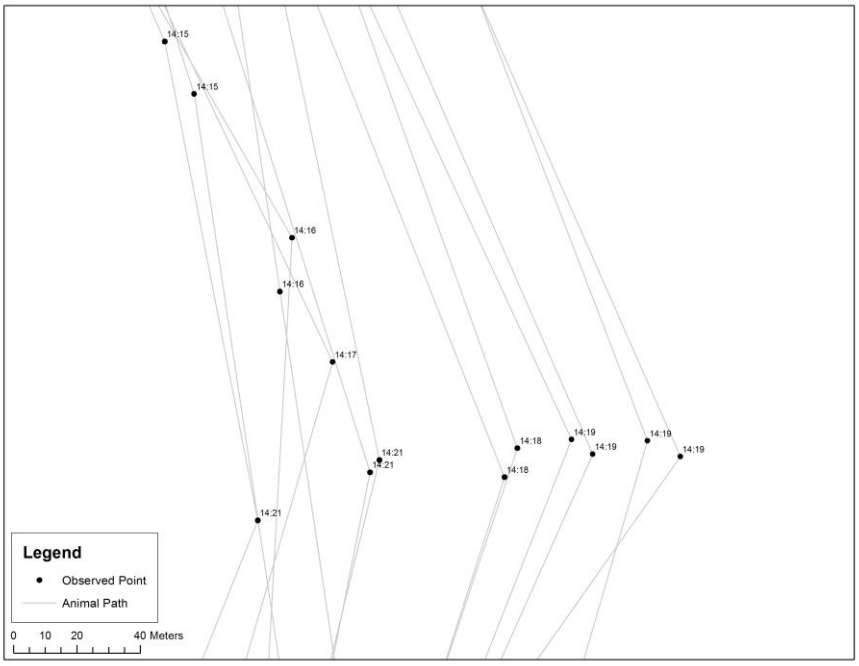

Figure 2: An example of the temporal offset caused by collecting spatiotemporal data on multiple moving targets using one researcher (Data collected at De Hoop Nature Reserve by Dr. Parry Clarke on March 27th 2007). Grey lines represent the interpolated paths, which sequentially connect the time stamped points collected in the field.

Animals are typically dynamic creatures moving about in their environment, which means that temporal and spatial errors are bound together just as distance and time are bound to each other in the formula for velocity. Torsten Hagerstrand went so far as to propose that space and time were inseparable (Hägerstraand 1970). So, if points are used that were not collected at - or interpolated to - the same time, it cannot be assume that there is any more than a tenuous and indeterminate spatiotemporal relationship between them. This research was conducted with the intent to not only quantify how much error exists between two points, but also provide a means to minimize this error during data collection. This is important because the production of complex patterns is a primary characteristic of the biological realm (Katz 1986). These complex patterns often manifest as some form of movement, whether it is at the cellular level (Trinkaus 1984) or the locomotion of animals (Lett and Mirabet 2008). Movement analysis is an essential part of the study of these biological phenomena and often involves the categorization of these patterns (Katz and George 1985), which often brings with it the spatial uncertainty of interpolation. These complex patterns rarely happen in isolation, often the movement patterns of multiple biological entities interact with one another, and this interaction brings with it more complexity and a greater need for accurate estimates of location.

The key to collecting accurate individual movement data at the group level is to find the appropriate balance between high and low resolution data. Turchin points out that, determining the spatiotemporal resolution with which to measure an organism's path is one of the most important analytical decisions and any method should avoid both undersampling and oversampling (Turchin 1998).

1.2 ANALYTICAL APPROACHES

Interpolation is often done using a linear estimation, where a straight line is drawn between two points due to the lack of additional data (Turchin 1998). This can have a great influence on accuracy of location, derived speed, and use of space estimations (Musiani 1998, Wentz, Campbell and Houston 2003, Pépin et al. 2004, Burdett et al. 2007). While much research has been conducted with a focus on the type of interpolation such as, mathematical interpolation (Friedman 1962), inferring network paths from point observations (Matisziw and Demir 2012), Bayesian estimation (Sumner, Wotherspoon and Hindell 2009), curvilinear interpolation (Tremblay et al. 2006), linear weighted distance, and constrained random walks (Wentz et al. 2003), few discuss methods to optimize the temporal interval between records (Reynolds and Laundre 1990). In highly mobile species, such as vervet monkeys (*Chlorocebus pygerythrus*) (Thomas 2013), or spider monkeys (*Ateles geoffroyi*) (Wentz, Campbell and Houston 2003), which live in complex three dimensional environments, it is easy to lose an individual or even the group for extended periods of time during field observations. This causes variability in the revisit times for individuals, which makes interpolation and the associated error a complex and central issue in the spatial analysis of primates. The reduction of temporal intervals also improves the amount of biologically significant information available in a dataset, thereby increasing its viability in many analyses (Reynolds and Laundre 1990). While curvilinear or smoothing techniques may improve the accuracy of interpolation on a finite scale, if at all (Tremblay et al. 2006), it is ultimately a reduction and optimization of revisit or collection times that will increase interpolation accuracy by decreasing the overall uncertainty of a path (Lonergan, Fedak and McConnell 2009).

1.3 RESAMPLING

While it is recognized that an interval in data collection no matter how short would make the data set discrete, for the sake of clarity paths collected at 5 second intervals during this research will be referred to as continuous paths. In order to analyze the error caused by interpolation, the original continuous path must be available for comparison (Figure 3a). This can be achieved using a time-based strategy (Andrienko et al. 2008) following one individual all day, capturing its route automatically by following the individual animal with a handheld GPS set to a very short interval. These continuous paths can then be resampled temporally (Figure 3b) to create a simplified subset of the original points which simulate the revisit times researchers might achieve in the field as they move from individual to individual. Resampling for this thesis is more akin to the downsampling used in signal processing (Strang and Nguyen 1996), as the paths are resampled at a set interval, in this case a set temporal interval. This downsampling technique was used to create uniform intervals which could be quantified and plotted out to show the change in error as revisit times increased, rather than a random resampling, (Figure 3c) which would not provide a quantifiable temporal interval.

Another method that can be used to resample the continuous paths is the Douglas-Peucker line reduction algorithm (Douglas and Peucker 1973) (Figure 3d), which was created to spatially simplify contour lines on maps. Because not all points in a series carry the same weight of information (Potamias, Patroumpas and Sellis 2006), points arranged in a predictable location - such as a straight line - can often be eliminated with minimal change to the line segment. The Douglas-Peucker algorithm works by capturing turns or changes in trajectory also known as trajectory evolutions within a defined

tolerance, while discarding those points that carry little relevant information about the path. The Douglas-Peucker algorithm, while ideal for the spatial simplification of data sets which have been oversampled, does not entirely address the reality of collecting spatiotemporal data in the field. Important trajectory evolutions are difficult to capture in the field unless a researcher is present for every turn the animal makes. Whether using handheld GPS loggers or tracking collars with pre-set data collection intervals, it is unlikely that all individual trajectory evolutions will be captured as they come about in a group setting.

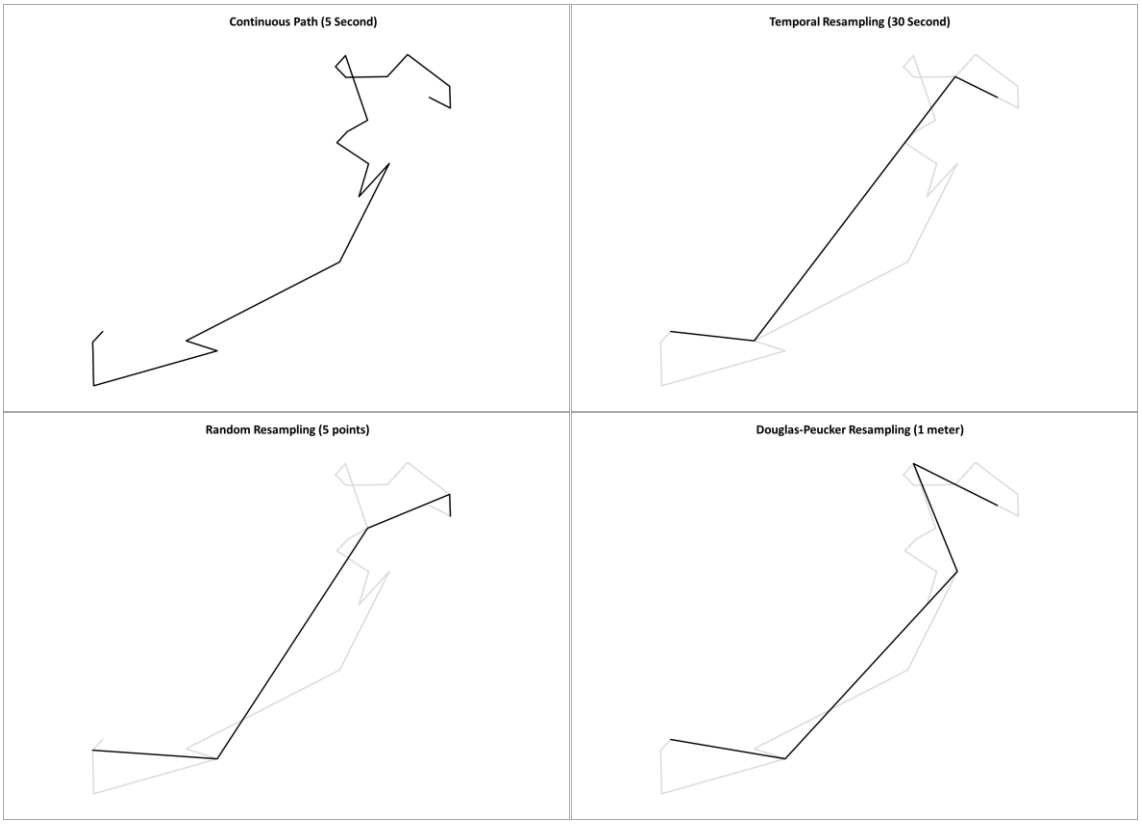

Figure 3: Demonstration of temporally, randomly, and spatially resampled paths overlaid on the unsampled continuous path. (a) Continuous path collected at 5 second intervals. (b) Temporally resampled path at a 30 second interval. (c) Randomly resampled path using 5 random points. (d) Spatially resampled path using the Douglas-Peucker algorithm (Douglas and Peucker 1973) set to a 1 meter tolerance. In (b), (c) and (d) the continuous path is grey overlaid with the black resampled paths.

1.4 AIMS OF THIS STUDY

The impitus of this project is the real-life problem posed by the collection of spatial data on baboons and vervet monkeys, where revisit times to individual animals was necessarily constrained by the demands of moving from one subject to the next to collect GPS points (see Figure 1 and Chapter 2). The purpose of the original data collection was to generate data that could be used to answer questions about how individuals interact spatially. However, before this is possible, it is necessary to both determine the extent of the intrinsic error and to estimate the improvement, if any, offered by this analysis. Consequently, this research focused on determining the amount of estimation error that can be expected at different sampling frequencies, both in relation to these specific data sets and to act as a guide for researchers intending to collect data of this kind. This will allow researchers to find a sampling frequency which best captures an animal's movement, while minimizing unnecessary oversampling of its path. It also allows for the application of margins of error to previously collected data. To do so, fine scale tracking data collected from individual adult vervet monkeys and chacma baboons (*Papio hamadryas ursinus*) at two sites in South Africa will be used. These paths were resampled at increasing intervals from one to sixty minutes in order to determine the extent of error associated with increasing granularity, by measuring length reduction, changes in sinuosity, spatial offset, and comparing the differential symmetry of the polygons formed by the resampled paths.

CHAPTER TWO

STUDY SITE AND METHODS

2.1 STUDY ANIMALS

The movement patterns of two species of primates were studied during this analysis: vervet monkeys and chacma baboons.

2.1.1 VERVETS

Vervet monkeys are semi terrestrial Old World monkeys. Next to baboons, they are the most widely distributed of the nonhuman African primates (Struhsaker 1967a, Wolfheim 1983, Pasternak 2011). Occurring primarily in riparian woodland throughout sub-Saharan Africa, ranging from Senegal to Ethiopia and as far south as South Africa (Struhsaker 1967b, Pasternak 2011). Home ranges can vary in size from 0.06 km^2 to 1.78 km^2 (Harrison 1983, Willems and Hill 2009), with the minimum distance travelled each day ranging from 135 m to 2251 m (Struhsaker 1967a, Wrangham 1981, Pasternak 2011).

2.1.2 BABOONS

Chacma baboons are Old World primates which, unlike vervets, are essentially terrestrial (Melnick and Pearl 1987, Hamilton and Bulger 1992). They are found in wooded highlands, grassland steppe, and savannah habitats throughout sub-Saharan Africa (Napier and Napier 1985, Melnick and Pearl 1987, Jolly 1993). Home ranges vary between 9.1 km^2 and 33.7 km^2, with a daily range of a troop fluctuating between 1.6 km and 14.5 km (Melnick and Pearl 1987).

2.2 STUDY SITES

Data were collected from two study sites: the Samara Game Reserve and the De Hoop Nature Reserve in South Africa. While Samara is only 230 km latitudinally north of De Hoop, they are two very different climatic regions. De Hoop is located on the coast

of the Southern Cape (Barrett et al. 2004) while Samara lies on the southern edge of the Great Karoo highland region (Pasternak et al. 2013).

2.2.1 SAMARA GAME RESERVE, SOUTH AFRICA

The vervet study site is located on the Samara Game Reserve, South Africa, with the center of the site situated at 32°22’06”S and 24°51’41”E (Figure 4). The site is dominated by a narrow riparian strip of Acacia Karroo (Heyne) woodland along a south flowing intermittent river that has an environmental influence on path choice for both vervets and researchers.

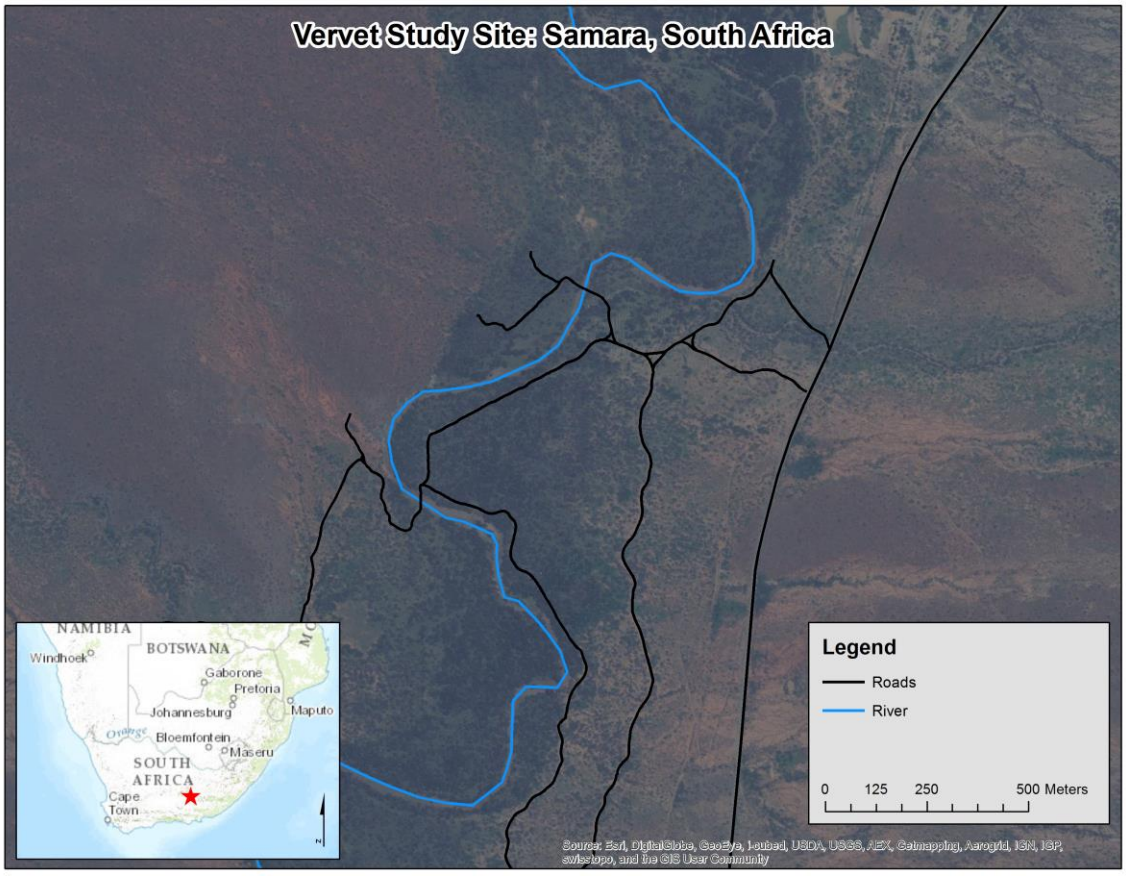

Figure 4: Local map of the vervet study site on the Samara Game Reserve. A red star shows the site’s location within South Africa (inset). A dark green north/south strip represents an acacia dominated riparian zone, which straddles the Milk River (blue line).

2.2.2 DE HOOP NATURE RESERVE, SOUTH AFRICA

The baboon study site is located next to the De Hoop Vlei on the De Hoop Nature Reserve, South Africa, with the center to the site situated at 34°28’10”S and 20°25’00”E (Figure 5). The site is covered by natural shrubland called fynbos, which only grows in a narrow belt along the western Cape of South Africa.

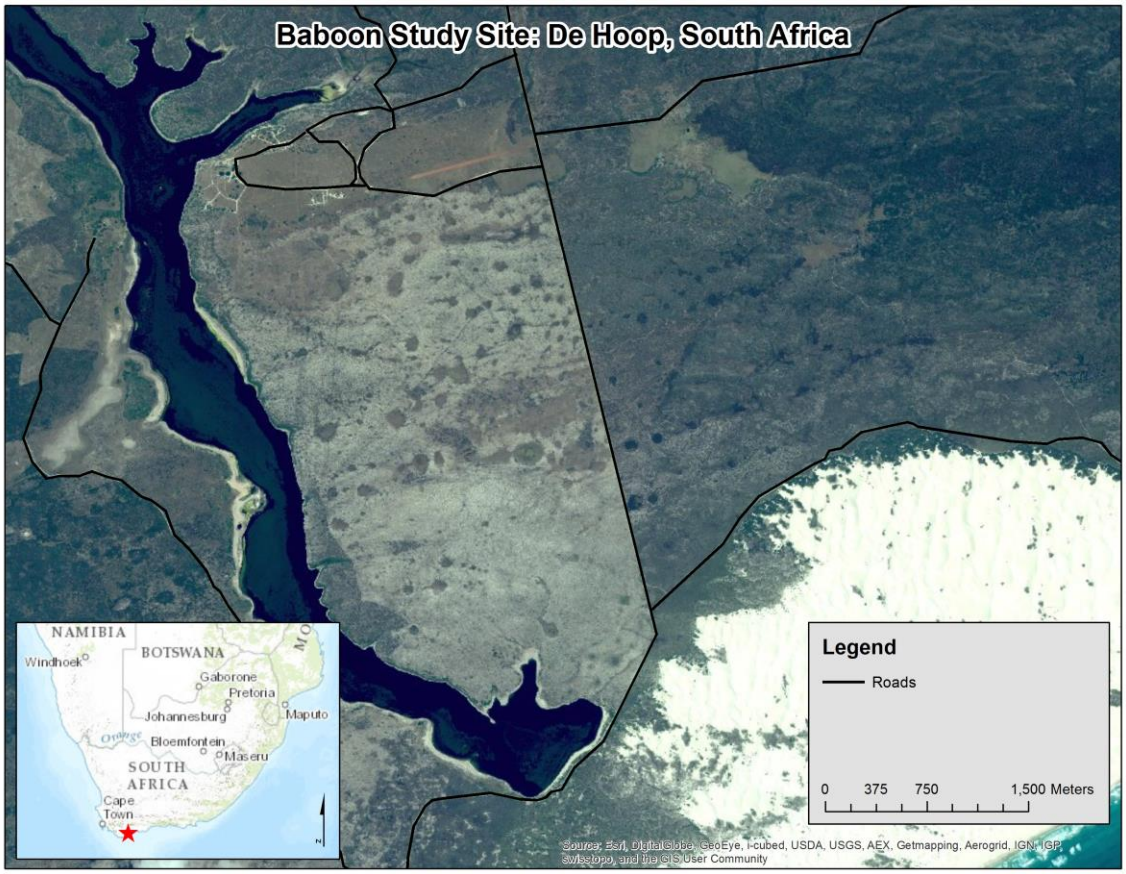

Figure 5: Local map of the baboon study site on the De Hoop Nature Reserve. A red star shows the site’s location within South Africa (inset). A white patch in the southeast corner represents sand dunes on the coast of the Indian Ocean, and the dark blue water body on the west edge of the site is the De Hoop Vlei.

2.3 DATA COLLECTION

Dissimilarities in habitat and daily travel ranges caused the two study species to use movement strategies suited to the navigation of their spatial environment, presenting

two varied spatial data sets. This along with the species tolerance to human observation made them ideal subjects for this study.

The two primary types of data that were collected at these study sites were spatial data, and social behavioural data. While they were often collected simultaneously this research only required the spatial data. Two data sets were used for each species in this study. The original discrete data sets, which were collected as explicitly spatial data sets with behavioural observations, had data points varying in temporal spacing from each other. Continuous spatial data with regular five second spacing between data points collected specifically for this research were also used.

The original Samara data used for comparison on this project were collected by Brittany Thomas and the Barrett-Henzi research team for seventy-nine days between January and July of 2012 (Appendix 1, Table 1). Adults from one of two troops in the area were the primary focus of data collection. Alternating between troops, there would be two to three observers following the focal troop from sleep site to sleep site, with one observer assigned to follow the other troop. While troop size did vary during the study period due to deaths and changes in troop membership, ~30,000 data points were collected on seventy-five adult vervets across two troops (Figure 7). Continuous data for this project was collected on ten individuals by the same team between August 2012 and July 2013 (Appendix 1 Table 5).

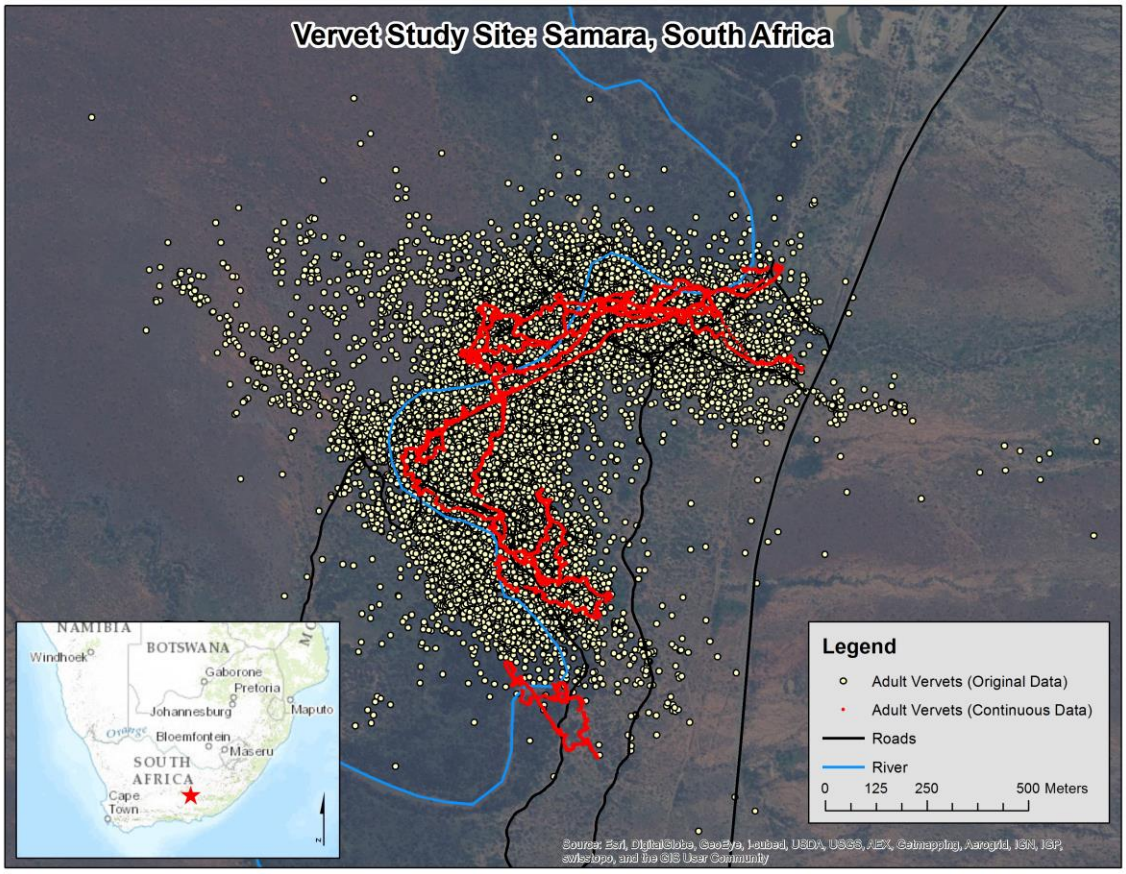

Figure 6: Original and continuous vervet data clouds overlaid on a local map of Samara Game Reserve, South Africa. Original data collected by Brittany Thomas between January and July of 2012, and continuous data collected by Eric Matlock and Jonathan Jarrett from March to April of 2013. Light yellow points represent individual observations from the original dataset, while red lines represent the continuous data sets.

Baboon data was originally collected at De Hoop by Dr Parry Clarke, who followed thirteen adult group members from sleep site to sleep site for 74 days between March 2007 and April 2008. He collected a total of ~60,000 data points (Appendix 1, Table 1) (Figure 6). Continuous data from De Hoop were collected for 21 individuals by Dr. Tony Phelps from March to October of 2013. From this data set 14 days were selected for the primary analysis (Appendix 1, Table 5).

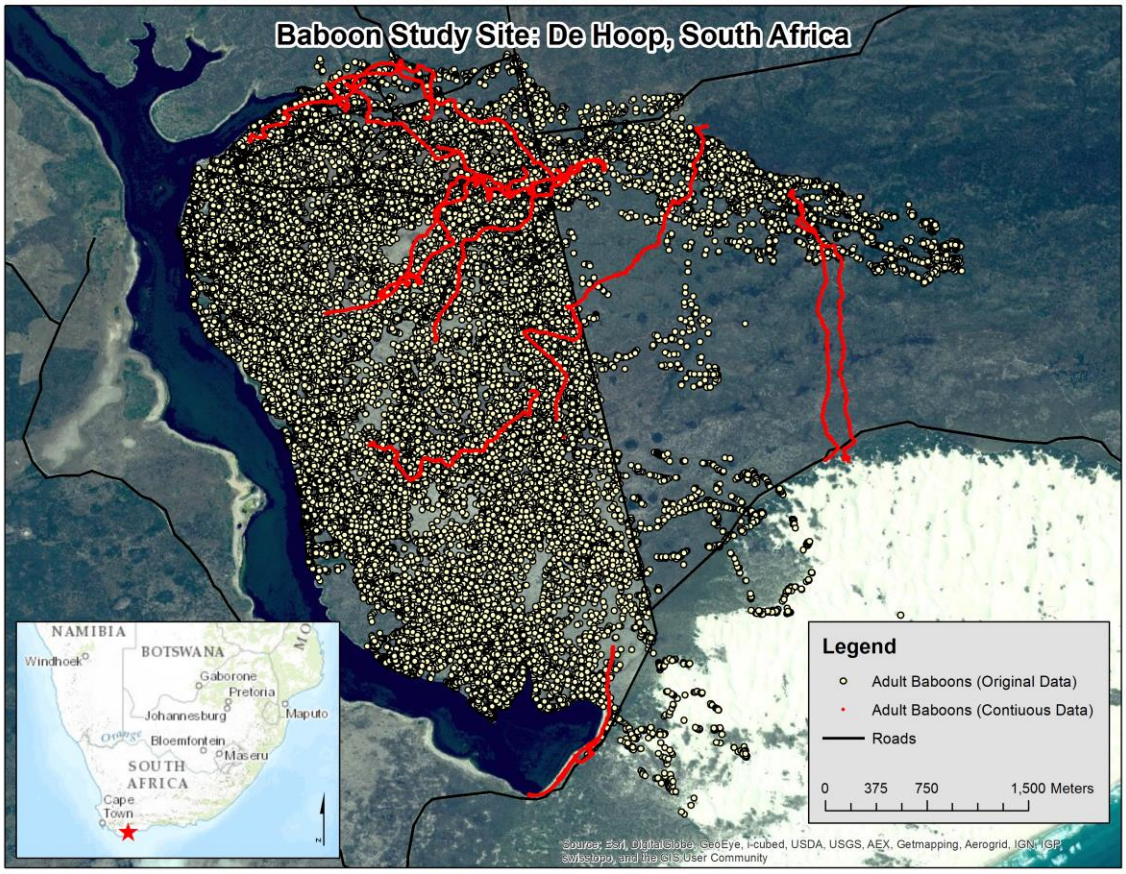

Figure 7: Original and continuous baboon data clouds overlaid on a local map of De Hoop Nature Reserve, South Africa. Original data collected by Dr. Parry Clarke between March 2007 and April 2008, and continuous data collected by Dr. Tony Phelps from March to April of 2013. Light yellow points represent individual observations from the original dataset, while red lines represent the continuous data sets.

To generate appropriate comparative data for the interpolations, continuous data for the vervets were collected using a Garmin Legend CX handheld GPS receiver, which has a reported accuracy of less than ±15 m. Observers closely followed an individual animal belonging to one of two habituated troops from sleep site to sleep site, with a typical follow lasting eleven hours. Initial data collection (six tracks) was done from August 3rd to September 7th, 2012, which is the beginning of summer for this region. Four more tracks were collected during the summer of 2013 to increase the sample to ten continuous tracks.

Continuous data for the baboons were collected using a Garmin eTrex 30 handheld GPS receiver, which has a reported accuracy of ±4 m. An observer followed one individual from a habituated troop at a time for two to three hours. These follows were much shorter due to time constraints and limited availability of local observers. Data collection of an initial eight tracks were conducted during March 2013, with six more samples in April of the same year (n=14).

The differential temporal durations of the two species paths did raise concerns about the validity of a comparison across the two data sets. In order to address these concerns, longer paths were created to test the validity of the shorter paths. Two methods were considered: agent-based modelling and the concatenation of several short paths to create a longer path. Time constraints did not allow for the creation of a sufficiently detailed agent-based model, therefore the concatenation technique was used here. Paths were concatenated by translating the points in one data set so that the start of one path coincided spatially with the end of another (Figure 8). This was done by subtracting the difference in the (x, y) coordinates of the start and end point of the two sets being concatenated from all points in the second set. The concatenated paths were compared to the original paths using Synchronized Euclidean Distance and exhibited similar patterns of change as the original shorter paths.

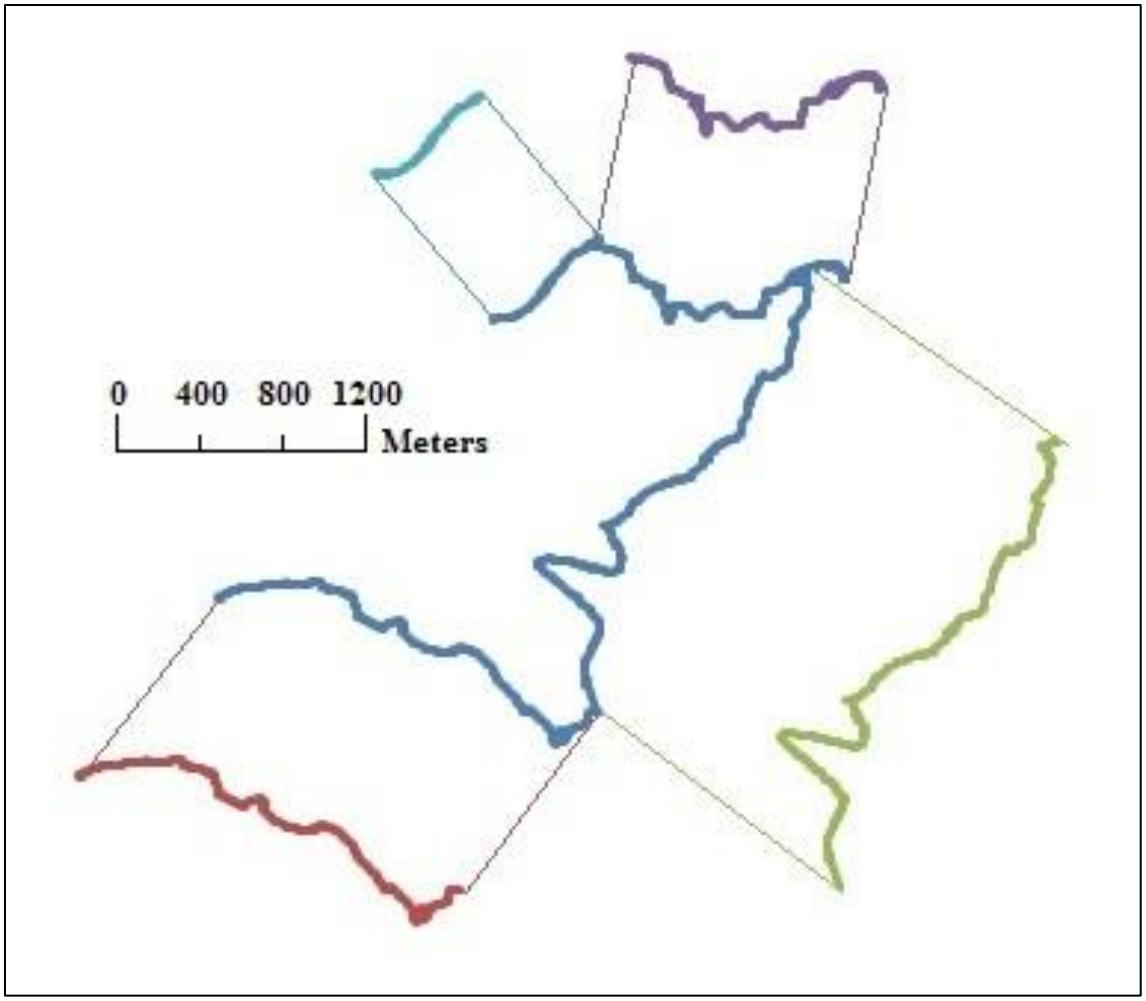

Figure 8: An example showing four baboon paths concatenated together to create one longer path. The red, green, purple and turquoise paths were collected on different days. By spatially transforming their location so the start of each path matches the end of the previous path, the longer blue concatenated path is created.

At both study sites points were collected at five-second intervals in order to achieve as close to a continuous path as possible. while still retaining enough power to last the entire day so battery changes did not interrupt the follow. Animals were followed in their exact track from a distance of 5 m. Detours from the animal's path were minimized. When detours were necessary, the GPS was paused until the researcher was back on track. Raw data were not post-processed or corrected in any way before being subjected to analysis.

2.4 SOURCES OF ERROR DURING INTERPOLATION

In general, interpolation error is the amount of error caused by connecting two points in space and time. Where the original path between the points is uncertain, the interpolated line is a theoretical path which may or may not conform to the original path. Interpolation error can manifest in many different forms which are covered below.

2.4.1 LENGTH LOSS

Length loss is the path length lost when the interpolation of points along an entities route creates a simplified path in comparison to the original path. It is measured by subtracting the length of the interpolated path from the length of the continuous path. Using linear interpolation, this comparison will typically result in a length loss or, in the rare case of a straight original path, no error.

2.4.2 SINUOSITY CHANGE

Sinuosity is a measure of how much a line curves, which is the ratio of total line length along the curve and the shortest (straight line) distance between its end points (Leopold, Wolman and Miller 1964). This ratio increases from the unity of two straight lines as the path becomes more sinuous (Figure 9). Sinuosity is exaggerated by circular movement, such as a day's travel involving the return of the animal to its point of origin. A circular path will result in a very short distance between end points with an exceptional high sinuosity in comparison to any portion of the day. Taken from hydrological research a sinuosity value greater than 1.5 is considered meandering (Leopold et al. 1964).

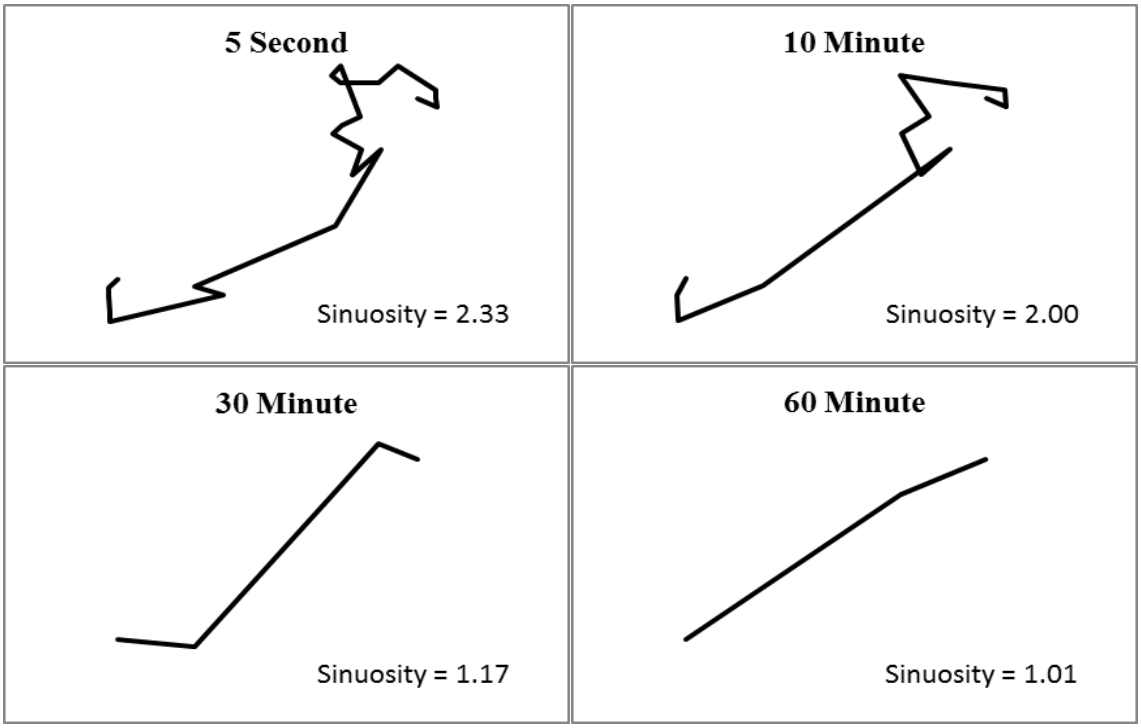

Figure 9: An example of sinuosity change from a temporally simplified 5 second interval path at ten, thirty, and sixty minute revisit intervals. Sinuosity is listed in the bottom right of each panel to highlight the change in sinuosity, as sinuosity approaches 1 the path becomes straighter (A sinuosity of one denotes a straight line).

2.4.3 SHAPE CHANGE

Paths can be looked at as shapes or polygons by simply connecting the start and end points. By looking at a path as a polygon it is possible to measure changes in path area as well as length. Shape change measurements assess how isometric two shapes are, in this case how far a subset of the original shape has moved from isometry with the original shape. Isometric comes from Greek, and means “having equal measurement”. Isometry for this study refers to direct isometry, where no rotation or translation has occurred. In this analysis, the change between the original shape created by the animal’s path was compared to that created by the interpolated path of the same animal using symmetrical differencing (Halmos 1960). Symmetrical differencing essentially ignores

the area of two overlapping shapes where no change has occurred and measures the area where the two shapes do not coincide (Figure 10).

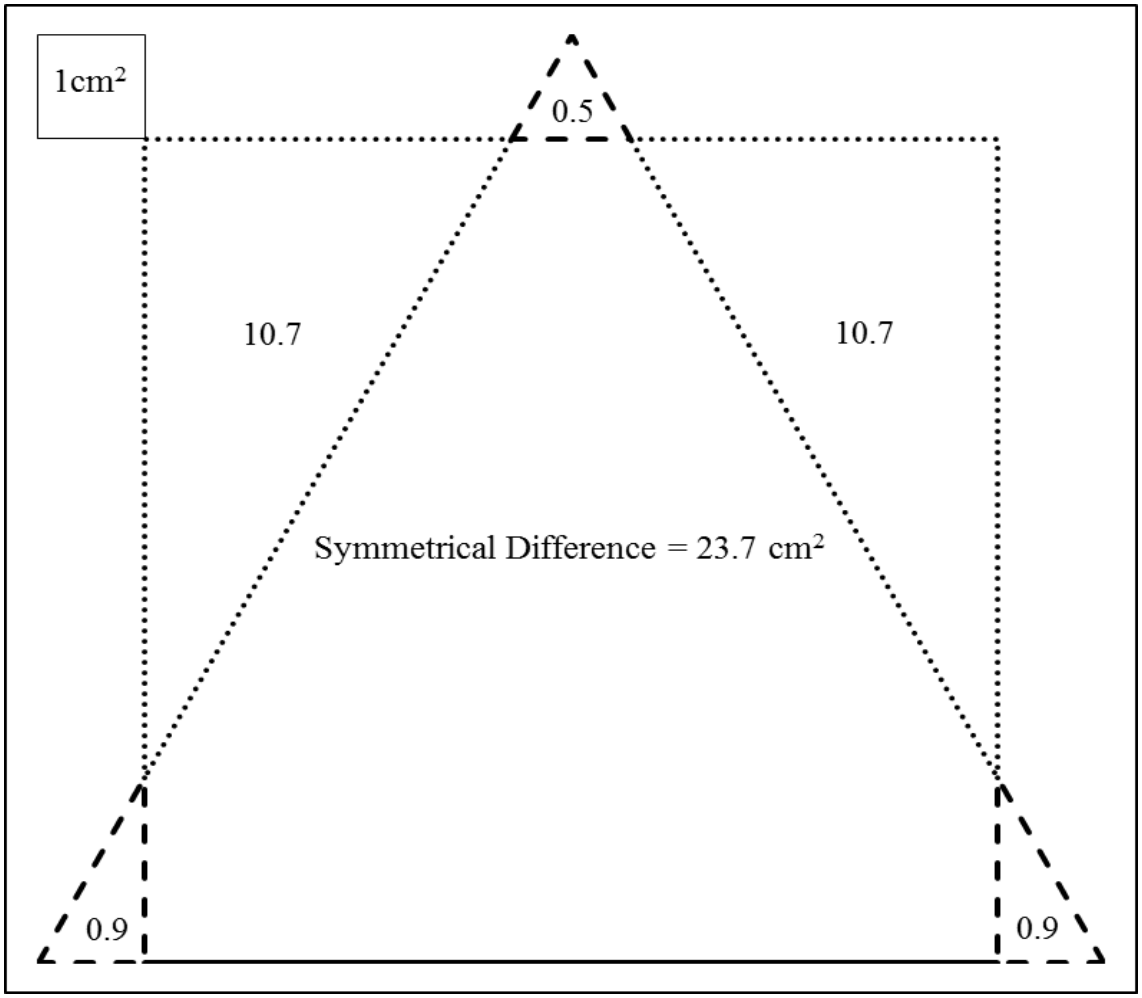

Figure 10: Symmetrical differencing example using a square and isosceles triangle. The symmetrical difference for these two shapes is represented by the sum of the non-congruent areas (cm^2) in the dashed line for the triangle and the dotted line for the square.

2.4.4 SPATIAL OFFSET OF INTERPOLATED POINTS

Synchronized Euclidean Distance (SED) is a measure of the Euclidean distance between two points which coincide temporally. SED takes into account both the geometric and the temporal relationship of the points along a trajectory (Zheng and Zhou 2011). For this study, SED was used to measure the spatial offset caused by interpolation between the original points and their temporal pairs on the interpolated line (Figure 11).

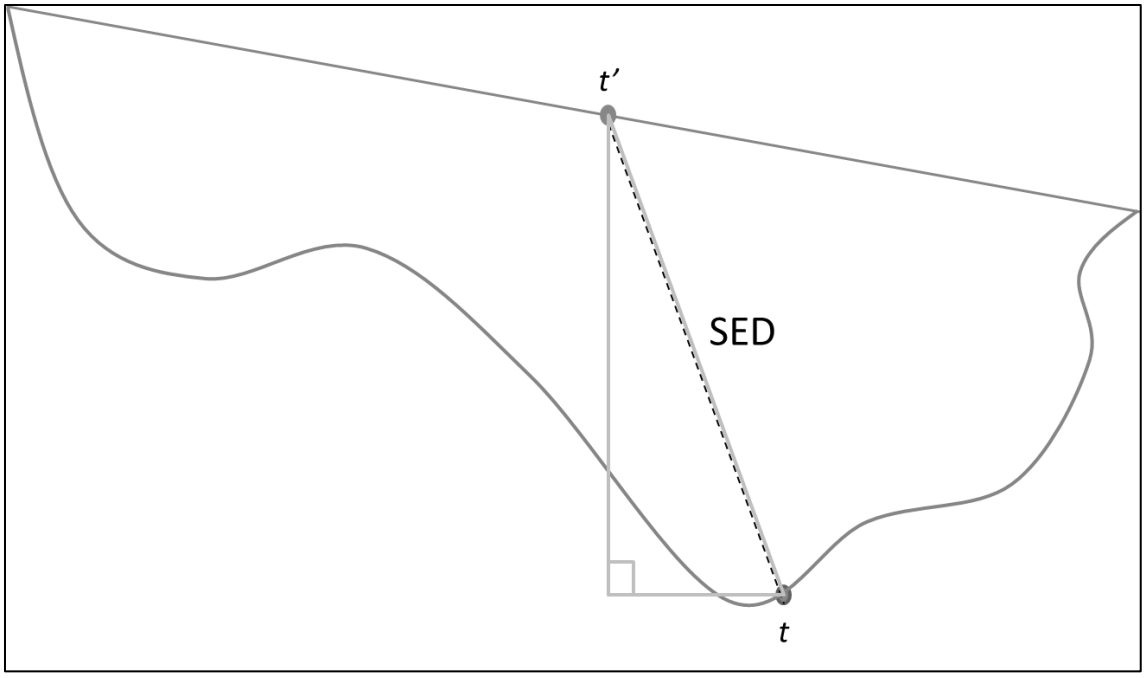

Figure 11: Synchronized Euclidean Distance (SED) measured between temporally synchronized points on both the continuous and interpolated lines. The point on the continuous line is labelled *t*, and the point on the interpolated line is marked *t'*, with a dashed line connecting them visually representing the SED.

2.5 ANALYSIS

2.5.1 PRE-PROCESSING

Raw data from the field were collected in the WGS1984 projection as Degree-Minute-Second (DMS) coordinates and then exported from the handheld units in the GPS Exchange Format (.gpx). These files can be imported into ArcGIS one of two ways: directly using the "*GPX to Feature*" tool, or by adding the data from Excel then using the "*Display XY Data*" function. For the following analyses, the data were imported into ArcGIS 10.1 (ESRI 2012) from Excel 2010 (Microsoft 2010) in order to facilitate the removal of unused fields, the conversion of DMS coordinates to decimal degrees, and to allow for the parsing of dates and time from text/string fields. Once the data were in ArcGIS, they were converted from the Geographic Coordinate System WGS1984 projection to a Projected Coordinate System (Samara: Cape UTM Zone 35S, De Hoop:

Cape UTM Zone 34S) in order to allow for the later use of Euclidian geometry in the analyses. Once projected, the "*Add XY Coordinates*" tool in ArcGIS was used to create UTM, X and Y fields. The data were then imported back to Excel where the date and time could be extracted from the timestamp string using Excel's *left(), right()* and *mid()* commands. The final step was to convert the time stamps from their native Microsoft Windows format to a simple numeric format of "hh.mm", where minutes were expressed as a base 10 decimal value (i.e. 5:30 is expressed as 5.5) in order to simplify later processing (Figure 12).

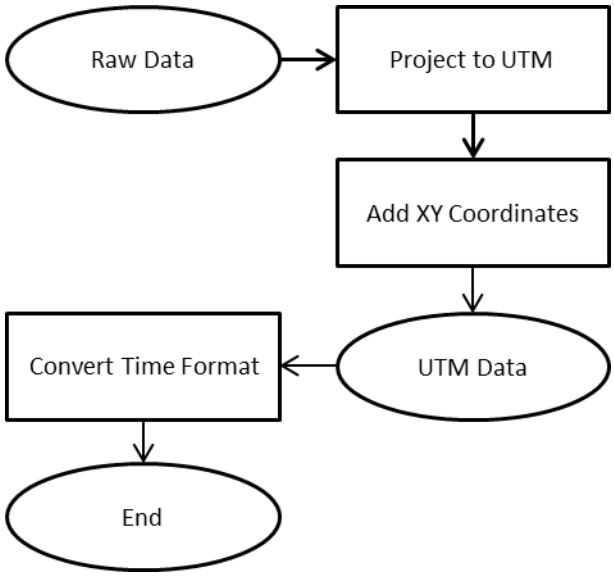

Figure 12: Analysis Preprocessing Flowchart. Ovals represent the inputs and outputs, with squares representing the processes.

2.5.2 LENGTH LOSS

Coordinates from the preprocessed data were resampled at increasing intervals (*i*) from one to sixty minutes using one minute steps. The core of the total length module is the primary routine used to resample the continuous data sets to simulate increased revisit times and is reused in all of the later analyses. Once each new data set was extracted for a

given interval (*i*), the Euclidean distance for every step (*n*) within the new data set was calculated using the Pythagorean formula [1].

$$d = Step_n^i = \sqrt{(x_n - x_{n+1})^2 + (y_n - y_{n+1})^2} \quad [1]$$

Where: *d* = step length (m), *n* = the number of steps in a path, *i* = the interval between points or step length in minutes, and *x* and *y* = coordinates (m). The Euclidean distance (*d*) for each step (*n*) was added to give a new total path length (*D*) for each interval (*i*) [2].

$$D = Path\ Length_i = \sum_n^i Step \quad [2]$$

The Euclidean distance of the original five-second data was also measured in order to establish a baseline distance for each animal's path (Figure 13).

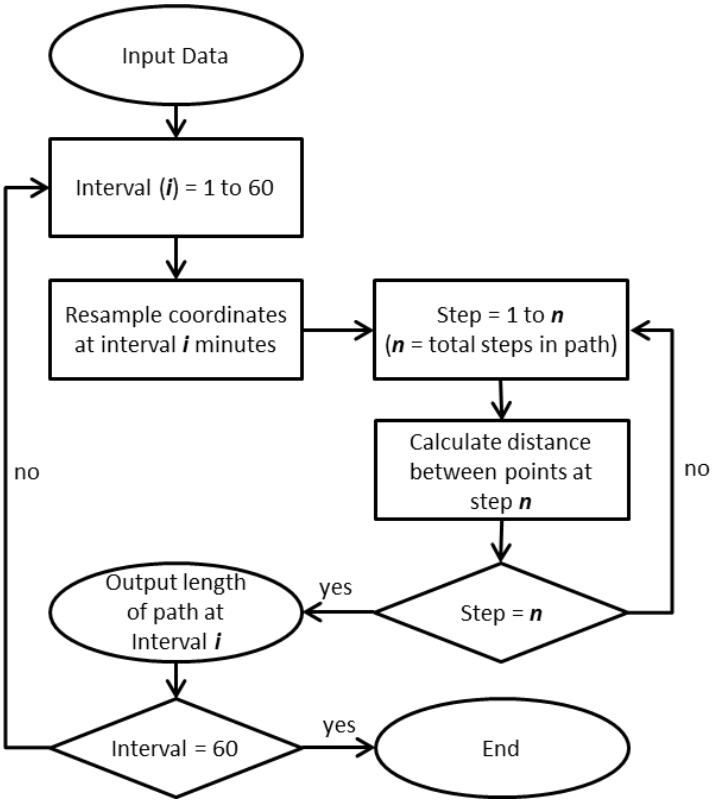

Figure 13: Total Length Loss Flowchart. Ovals represent the inputs and outputs, with squares representing the processes, and diamonds representing decision branches.

2.5.3 SINUOSITY CHANGE

Total path sinuosity, also referred to as tortuosity in the literature (Kelvin and Tait 1885), was calculated in a similar manner as the total path length calculation with one added step (Figure 14). Once the path length at interval *n* was determined, it was divided by the straight-line distance from start to finish, which is termed the shortest path length in the formula below [3]. This metric was applied only to the entire day's path.

$$Sinuosity = \frac{Actual\ Path\ Length}{Shortest\ Path\ Length} \quad [3]$$

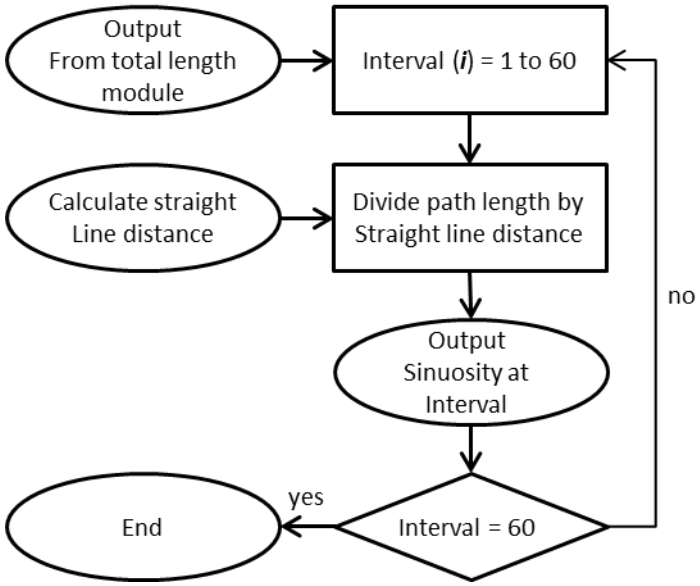

Figure 14: Sinuosity Change Flowchart. Ovals represent the inputs and outputs, with squares representing the processes, and diamonds representing decision branches.

2.5.4 SYMMETRICAL DIFFERENCING

In order to measure the symmetrical difference of two paths it is first necessary to convert all point layers into polygons. This was done using the Geospatial Modelling Environment's (GME) "*Point to Polygon*" tool (Beyer 2012). Symmetrical differencing requires at least two polygons or sets (often referred to as Set A and B in Set Theory) and is defined by the formula:

$$Symmetrical\ Difference = (A - B) \cup (B - A) \quad [4]$$

In this model the primary polygon (Set A) was that of the original continuous path, which provided a baseline for the comparison of change in the path shape relative to the original continuous path. Another method investigated was to use a popular spatial generalization technique, the Douglas-Peucker algorithm, to create an alternate (simplified) baseline, with the logic that by repeating the symmetrical differencing using this new baseline a value approaching zero could be interpreted as a 'best fit'. The difference of the area between the two polygons was measured using the "*Symmetrical Difference*" and "*Calculate Geometry*" tools in ArcGIS 10.1 (Figure 15). This tool does not perform a union of A-B and B-A, but logs them as two separate shapes/areas. The output (area) from ArcGIS was imported into Excel using a VBA macro, where the area of the two sets were added and then graphed. This comparison gave an area measurement that showed the extent by which the shape changed as revisit times increased (Figure 15).

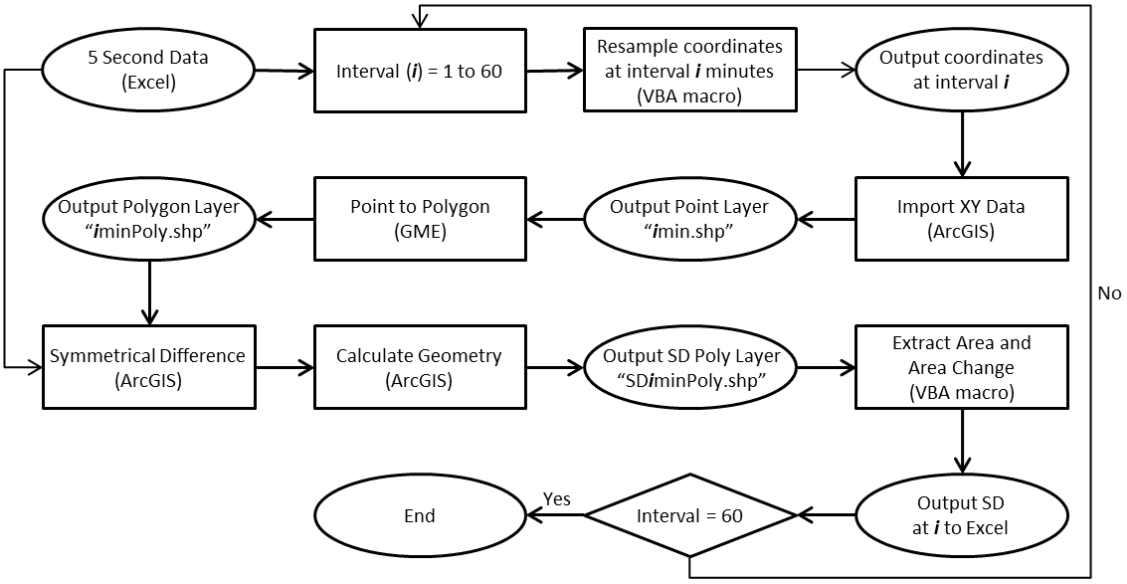

Figure 15: Symmetrical difference flowchart. Ovals represent the inputs and outputs, with squares representing the processes, and diamonds representing decision branches.

2.5.5 SYNCHRONIZED EUCLIDEAN DISTANCE

Synchronized Euclidean Distance represents the spatial offset between two points which represent the same entity at the same moment in time, but have a different spatial location due to interpolation or path simplification. This is measured using the following formula for Synchronized Euclidean Distance (SED):

$$SED_t = \sqrt{(x_t - x_t')^2 + (y_t - y_t')^2} \quad [5]$$

Where: t = time, x and y = coordinates on continuous path, and x' and y' = coordinates on interpolated path.

As Synchronized Euclidean Distance is a measure of what is happening between the resampled points, it is necessary that the resampled paths be interpolated to an interval smaller than the smallest resample interval. For this study, resampled paths were interpolated at one minute spacing (Figure 16).

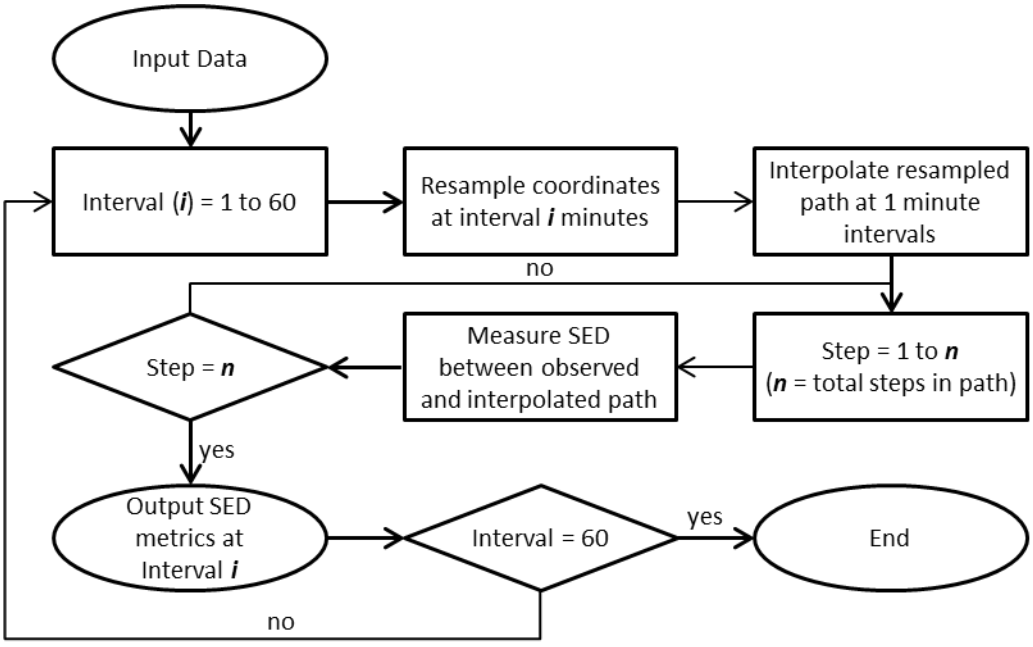

Figure 16: Synchronized Euclidean Distance Flowchart. Ovals represent the inputs and outputs, with squares representing the processes, and diamonds representing decision branches.

CHAPTER THREE

RESULTS

3.1 LENGTH LOSS

As the path-point resolution becomes coarser, length loss due to interpolation highlights two distinct modes of change: the initial and very rapid smoothing of the path followed by unpredictable corner cutting as critical trajectory evolutions are missed. Smoothing of the continuous path as it was resampled to increasing revisit times typically occurred in the initial one to ten minute resamples. In the baboon study group, ~35% of the overall length of the continuous path was lost in this initial path smoothing phase (Figure 17). A similar pattern with an increase in amplitude emerged from the vervet study group with ~50% loss (Figure 17).

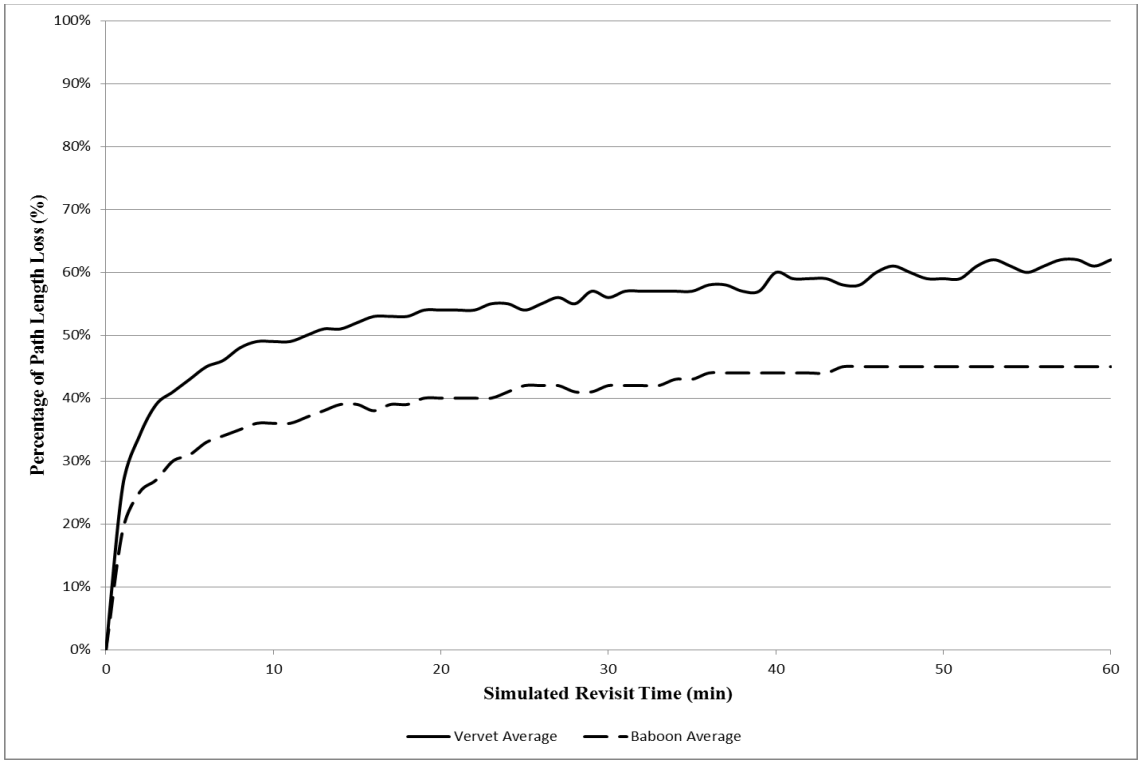

Figure 17: Average Length Loss due to interpolation for vervets and baboons. Length Loss is measured as the percentage of loss compared to the continuous path as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the continuous path. The solid line represents vervet length loss, while the dashed line symbolizes baboon length loss.

Length loss is highly correlated to revisit times for both species (Baboons: R = 0.75, N=60, R^2 = 0.56, P < 0.01, Vervets: R = 0.76, N=60, R^2 = 0.57, P < 0.01). Interspecies similarities in the length loss graphs were confirmed by their very high correlation (R = 0.99, R^2 = 0.99, N=60, P < 0.01).

3.1.1 PATH SMOOTHING

The initial rapid length loss established in the results above (Figure 17) were believed to be the result of linear interpolation smoothing out small noisy movements which did not constitute course changes (Tourtellot, Collins and Bell 1991). In order to confirm the processes underlying this notion, a circular path was created in NetLogo 5.0.5 (Wilensky 1999) (Appendix 2). Step length was set to 1 meter with 3600 steps (1 step = 5 seconds) changing course in 0.1 degree increments to create a full circle. The smooth circle demonstrated a fairly linear degradation in length due to interpolation with no initial rapid data loss (Figure 20). Noise was then introduced into this smooth circular path by inducing a regular 45º wiggle into the movement (Figure 18).

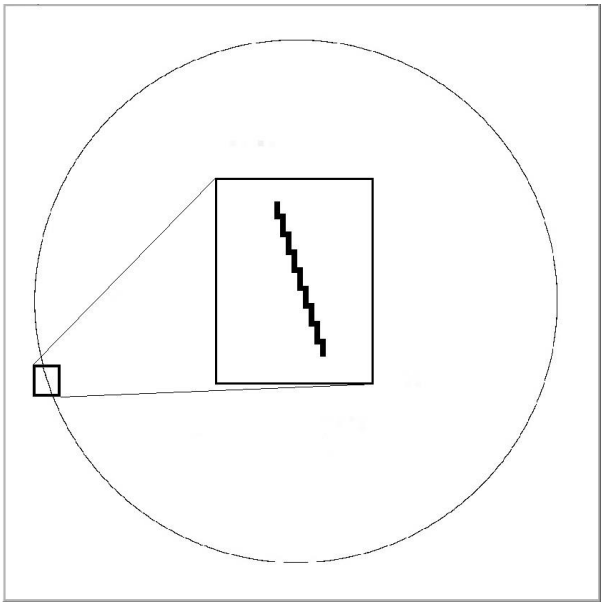

Figure 18: Circular path created in NetLogo agent based modelling environment to simulate a noisy path by inducing a small wiggle in the movement. The inset is a magnification that illustrates this wiggle.

The wiggle created a predictable but noisy path which exhibited a pattern of rapid initial length loss due to path smoothing (Figure 19) analogous to the length loss graphs above (Figure 17). The steep initial drop is due to the regular amplitude of the wiggle, which caused an immediate change in the graph rather than a smooth transitioning curve. Had the wiggle varied in length and time as would be expected in a natural path, a smoother initial curve would emerge.

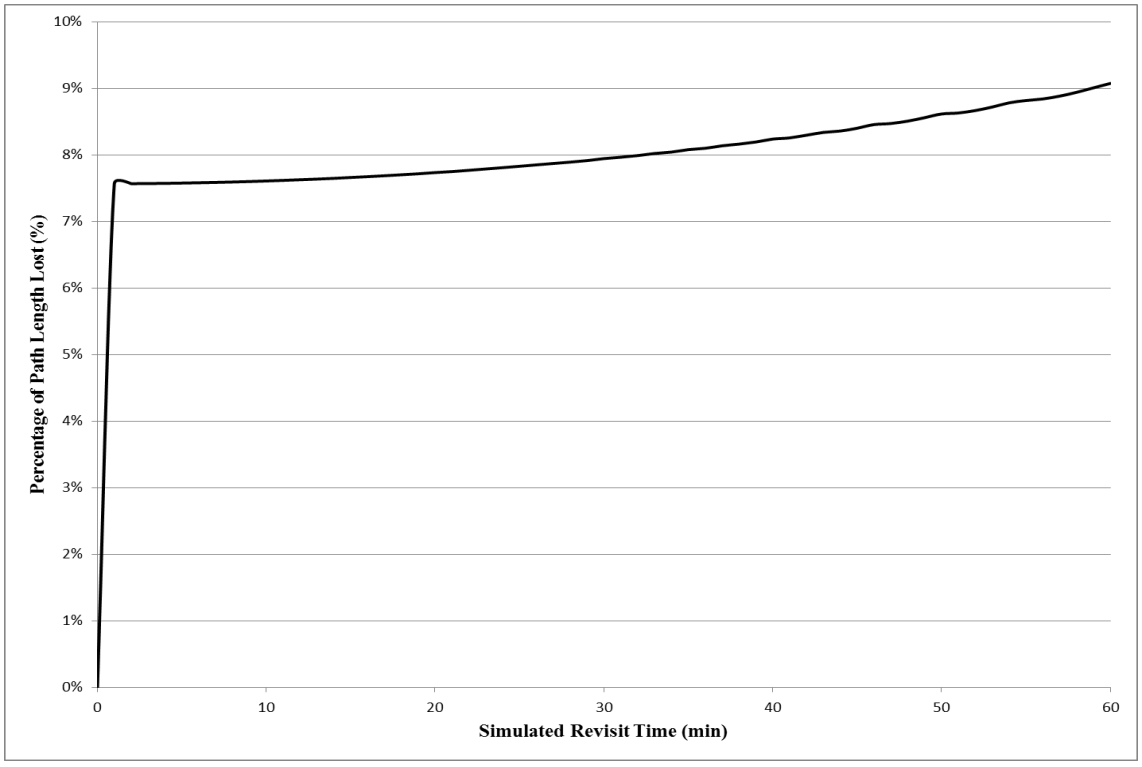

Figure 19: This graph demonstrates the rapid initial length loss caused by linear interpolation, as the noisy movement in the model of a circular path (Figure 19) is flattened. Length Loss is measured as the percentage of loss compared to the continuous path as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the continuous path.

3.1.2 CORNER CUTTING

Oscillations in the length loss graph as it reaches asymptote were thought to be the result of corner cutting. This was caused by a cyclic progression in the plotting of corners when increasing revisit times were plotted relative to the actual turns in the continuous path. As such the path length varied relative to corner-point proximity, which

directly influenced overall length loss. To demonstrate the cyclic progression of an interpolated corner, three successive revisit times were plotted onto the circular path below (Figure 20). Since the starting point for all interpolated paths were the same as the original continuous path regardless of the revisit interval (*i*), any revisit time of *i* minutes resulted in the individual points representing the corners (*n*) at a given time (*t*) being moved forward relative to the previous revisit time (*i-1*) at a growth rate of *n* minutes from the point of origin.

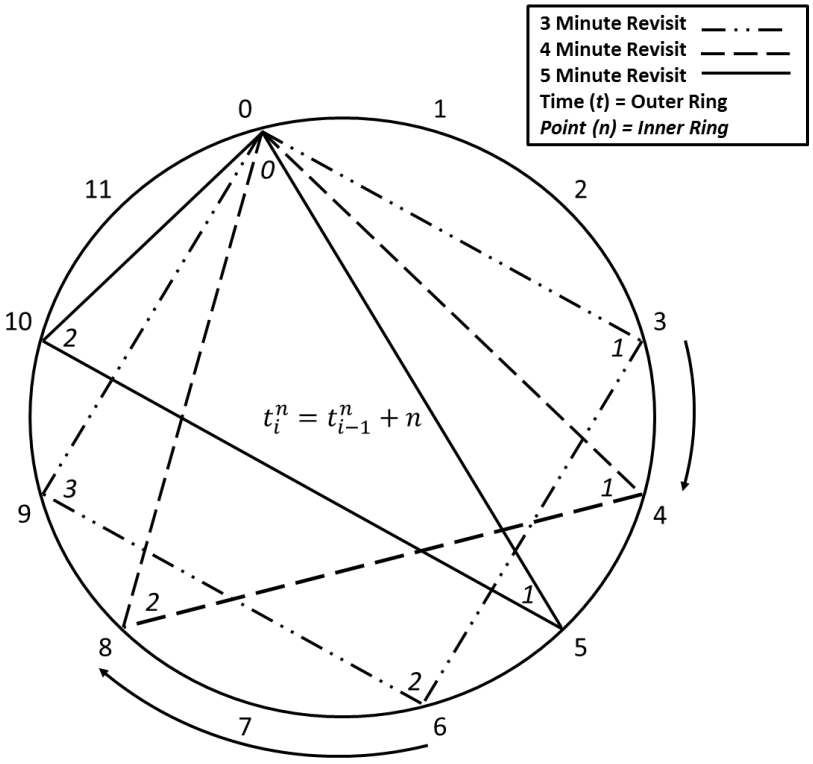

Figure 20: An example demonstrating the cyclic progression of corners expected in a movement path as revisit times increase using a circular path. Time (t) = numbers outside of circle, point (n) = numbers inside of circle. Note how the corners move clockwise as the revisit time increases.

The corner cutting concept was tested using the same NetLogo model as above (Section 3.1.1), but with the wiggle removed to provide a perfectly smooth circular path. It was hypothesized that the lack of corners in a circle would produce a perfectly smooth length loss graph if corner cutting is the only process causing oscillations. When this was

first tested oscillations in the length loss graph were minimal, but not completely removed as demonstrated by the dashed line in the graph below (Figure 21). It became apparent that revisit times which did not divide evenly into the total track time had uneven final revisit times (i.e. a one hour path would have its first uneven revisit at a 7 minute interval, with 8 points collected 7 minutes apart and the last point collected at 4 minutes). These temporally shortened final revisits caused small dips in the graph at the respective revisit times which did not divide equally into the total track time and, as such, did not exhibit linear growth. These uneven revisits were normalized in the analysis to provide consistent results for a given revisit time by dividing the length of the last leg by its actual revisit time, then multiplying the result by the expected revisit time. Normalized results are shown with a solid line.

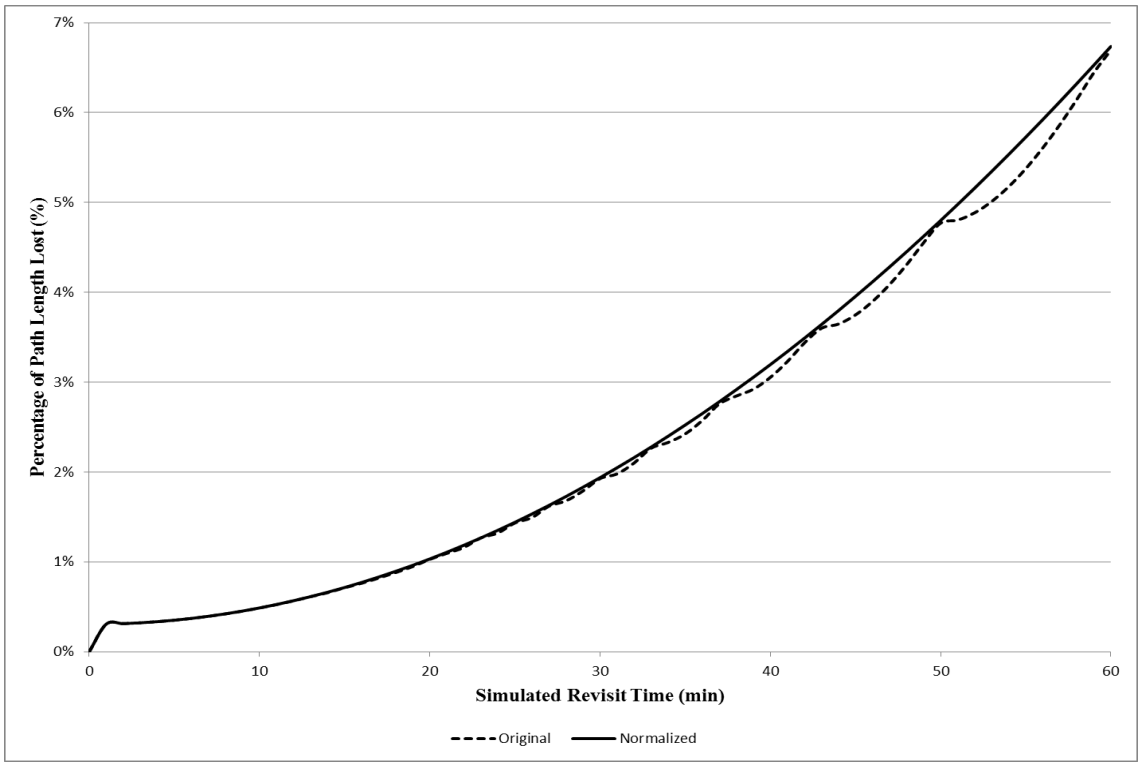

Figure 21: Length loss results for a smooth circle demonstrating the expected curve for a path lacking defined corners. Length Loss is measured as the percentage of loss compared to the continuous path as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the continuous path.

3.1.3 PATH COMPACTNESS AND CORNER CUTTING

Corner cutting is clearly shape dependent and is expected to be less pronounced in smooth predictable curves as demonstrated by the circle example above (Figure 19). However, it is not the only factor affecting corner cutting; path compactness is thought to control the magnitude of corner cutting. This is largely due to the fact that compact paths are likely to have shorter segments between corners, which in turn create shorter corners. By having shorter corners the length loss caused by corner cutting will reduce the magnitude of oscillations in the graph below (Figure 22), which shows the length loss curves of two individual vervets.

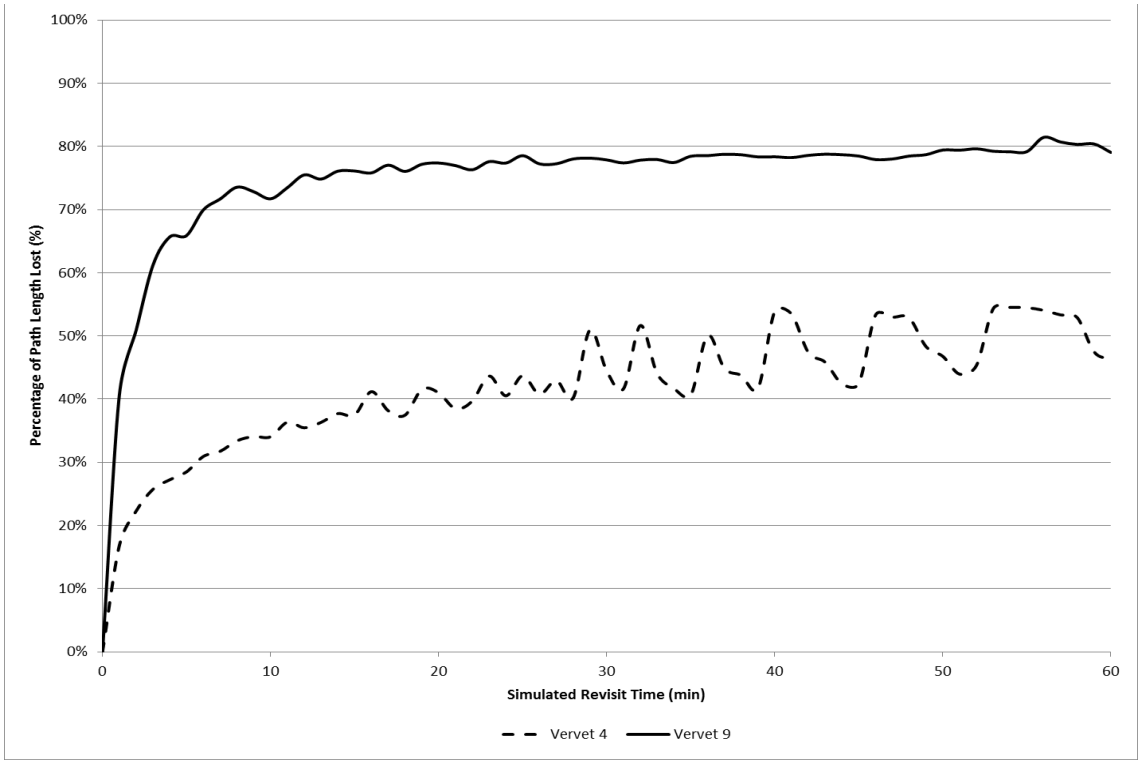

Figure 22: Vervets four and nine length loss graphs exhibit different magnitudes of oscillation due to corner cutting. Length Loss is measured as the percentage of loss compared to the continuous path as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the continuous path. The solid line represents Vervet 9 length loss, while the dashed line symbolizes Vervet 4 length loss.

The examples below illustrate how a similar pattern of corner cutting at different scales (Figure 23) can result in a different magnitude of corner cutting error (Figures 22). Vervet 9 travelled 3 km in 5.5 hours, but its path was visibly more compact in comparison to Vervet 4 who travelled more slowly, covering only slightly more ground (4.2 km) in twice as much time (11.3 hrs). As predicted, the highly compact path of Vervet 9 (Figure 23) exhibits low amplitude oscillations during the corner cutting phase (Figure 22) than the less compact path of Vervet 4.

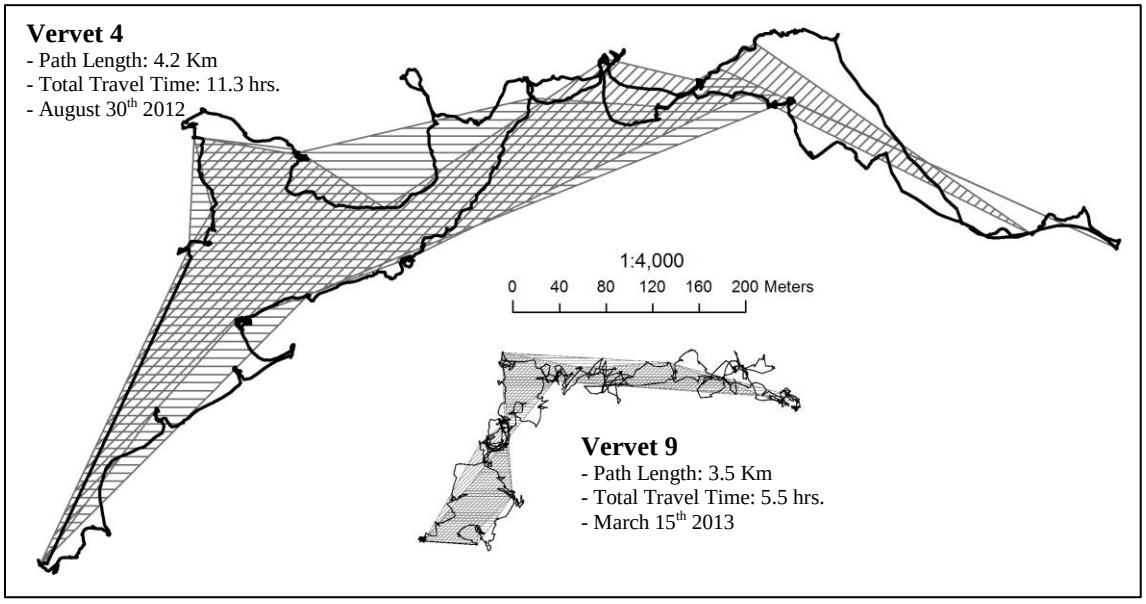

Figure 23: Comparison of path compactness for vervets four and nine, both paths use the same scale. The thick black line represents the continuous path for both vervets. The crosshatched polygons represent two consecutive revisit times interpolated paths, demonstrating how different corners and magnitudes of corner cutting occur at different revisit times.

3.1.4 ISOLATION OF PATH SMOOTHING AND CORNER CUTTING

While the initial path smoothing phase was visually distinguishable from the later corner cutting phase, these two phases were also isolated using a piecewise regression (Figure 24).

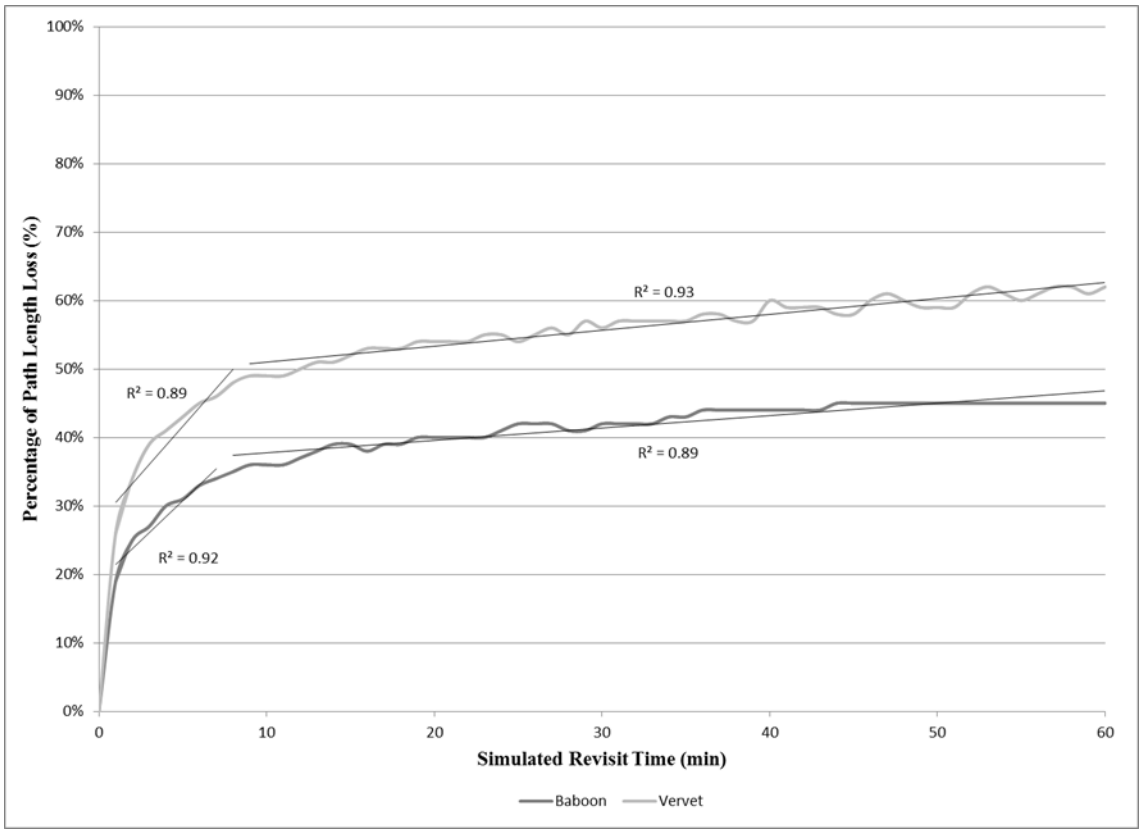

Figure 24: Identification of path smoothing and corner cutting phases using piecewise regression. Length Loss is measured as the percentage of loss compared to the continuous path as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the continuous path. The dark grey line represents baboon length loss, while the light grey line symbolizes vervet length loss. In both cases piecewise regression shows an obvious break just before the 10 minute revisit.

3.2 SINUOSITY CHANGE

Path sinuosity was originally intended as a metric to apply to all the paths collected. It became apparent that this would not be required, changes in path sinuosity due to interpolation, occurred at precisely the same rate as changes in path length (Figures 25). This is ascribed to the fact that path sinuosity is a ratio of path length compared to the shortest distance from the point of origin to the end of the path. Since the point of origin and the end point are constant throughout the analysis, the only variable that changes is path length. Therefore, a comparison of changes in sinuosity and length over different simulated revisit times will yield the same results. The only difference is

the metric being used to express the result, where the path length was measured in meters lost, while sinuosity was expressed as a change in the unitless sinuosity value. When both are expressed as a percentage of change from their initial continuous state, the graphs align seamlessly.

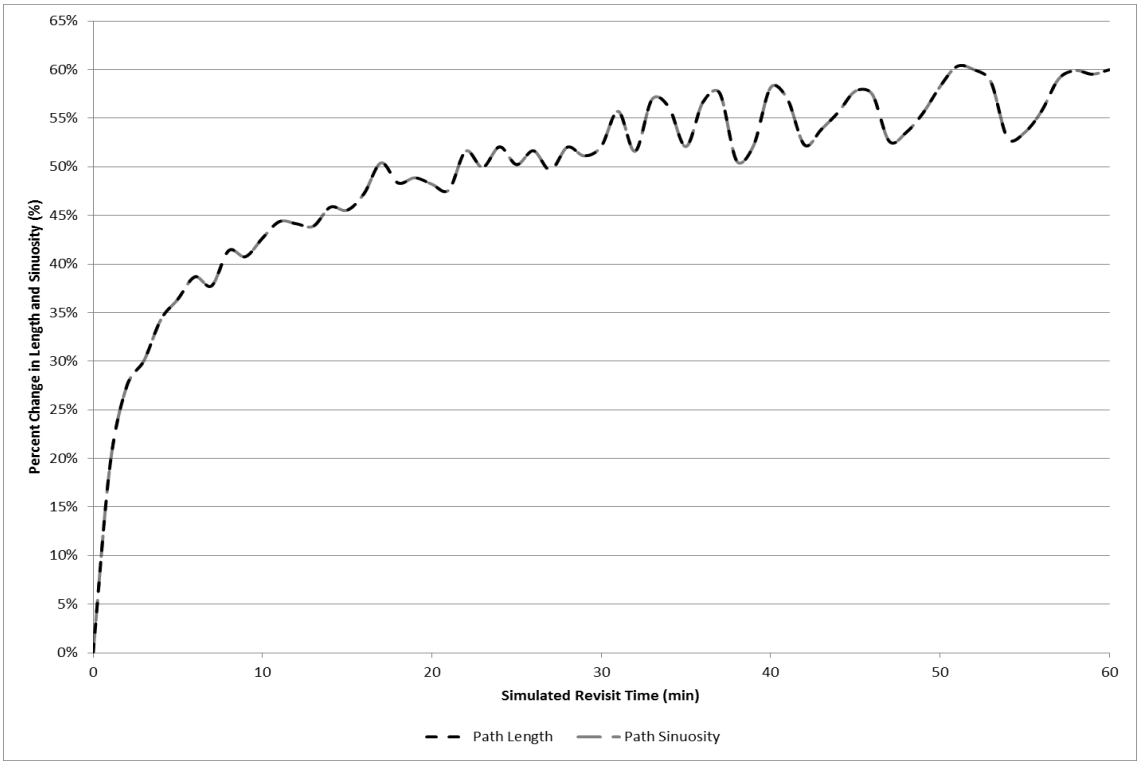

Figure 25: Example showing the uniformity between path length and path sinuosity graphs. Length and sinuosity change are both measured as the percentage of loss compared to the continuous path as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the continuous path. The dark grey line represents length loss, while the light grey line symbolizes sinuosity change.

3.3 SYMMETRICAL DIFFERENCING

Symmetrical differencing for the vervets shows a similar pattern of data loss as the length loss analysis (Figure 26), with a steep initial slope prior to ten minute revisits followed by an oscillating slope which approaches a maximum length loss. These oscillations are caused by the same cyclic motion of the points relative to the initial continuous path as described with the circle example above (Figure 20). As with length loss the oscillations also represent corner cutting, as is illustrated in Figure 23 by the

polygons overlaid on the continuous paths. These polygons represent higher revisit times which have failed to capture many of the corners of the continuous path.

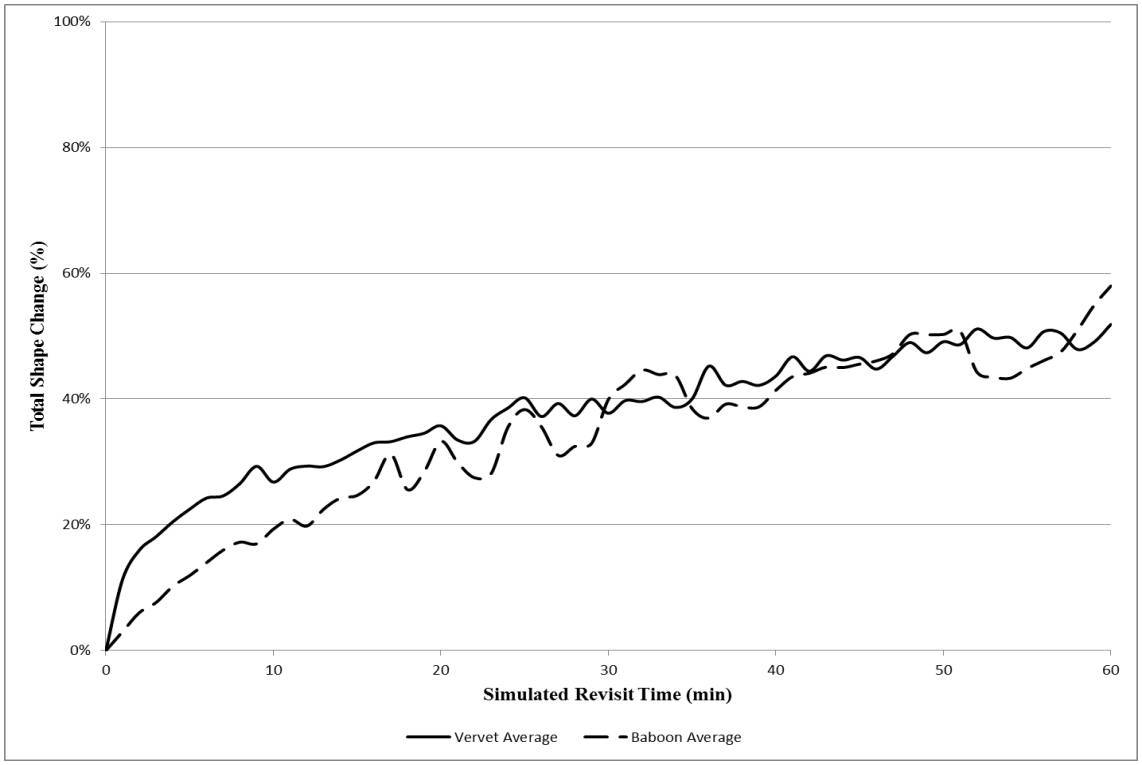

Figure 26: Change shape due to interpolation measured using average symmetrical difference for vervets and baboons. Shape change is measured as the percentage of change in set B (interpolated path) compared to set A (continuous path) as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the set A path. The solid line represents vervet shape change, while the dashed line symbolizes baboon shape change.

Like length loss, shape change also exhibited a very high correlation between species ($R = 0.96$, $N = 60$, $R^2 = 0.91$, $P < 0.01$). The data also indicate that shape change is very highly correlated to length loss in both species (Baboons: $R = 0.84$, $N = 60$, $R^2 = 0.71$, $P < 0.01$, Vervets: $R = 0.93$, $N = 60$, $R^2 = 0.86$, $P < 0.01$).

3.4 DOUGLAS-PEUCKER ALGORITHM

Symmetrical differencing with the Douglas-Peucker algorithm showed the same pattern of initial loss as the previous analyses (Figure 26). When compared to the original symmetrical differencing results, which used the original continuous path as a baseline,

the Douglas-Peucker symmetrical differencing results (Figure 27) were nearly identical ($R = 0.99$, $R^2 = 0.99$, $N = 14$, $P < 0.01$). The resolution of the simulated revisit was set to five minutes for the initial run of this comparison. When the similarity in the results was established, further higher-resolution runs were deemed redundant.

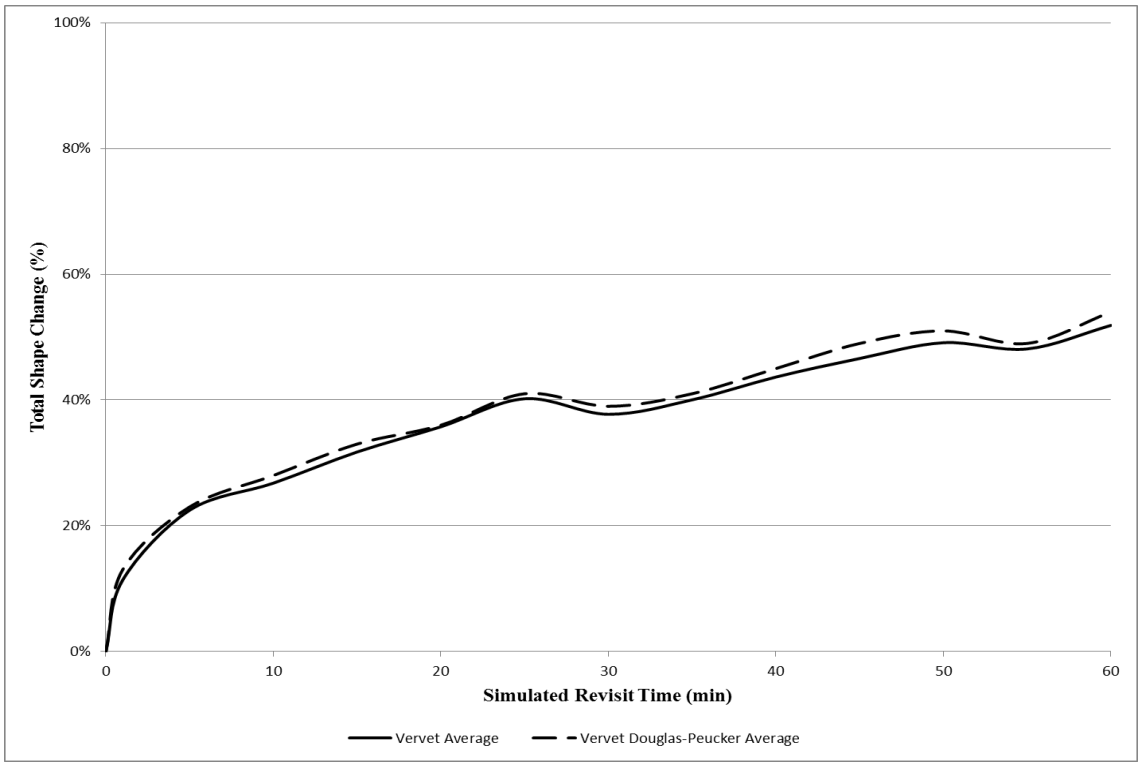

Figure 27: Symmetrical difference results for vervets using the Douglas-Peucker paths set to a 5 meter tolerance as Set A, compared to the original results which use the continuous path as Set A. Shape change is measured as the percentage of change in Set B (interpolated path) compared to Set A as revisit time's increased from 0 to 60 minutes, with the 0 minute revisit time representing the Set A path. The solid line represents the original shape change results, while the dashed line symbolizes the Douglas-Peucker results.

This similarity is not surprising under the circumstances. The Douglas-Peucker algorithm is capable of preserving critical trajectory evolutions when set to a reasonably low tolerance, which was set to five meters for this analysis. Therefore, the general shape of the original path, which was used in the initial symmetrical differencing, was preserved. This also allowed for the extraction of turns as a by-product of the Douglas-Peucker analysis, which were not always easily detected on continuous paths due to many

turns forming gentle sweeping arcs. Turn frequency in the following figure (Figure 28) has been binned according to the nearest minute interval between two turns. This gives an indication of how frequently (temporally) actual turns occurred on the continuous path. When the Douglas-Peucker algorithm set to a 5 m tolerance is used, 99% of the turns for the baboon data occurred within 24 minutes of each other, with more than half (56%) occurring within a minute of each other (Figure 28). Vervet turns were more temporally spread, with 99% occurring as far apart as 31 minutes, but, as with the baboon turns, approximately half were within one minute of each other (Figure 28).

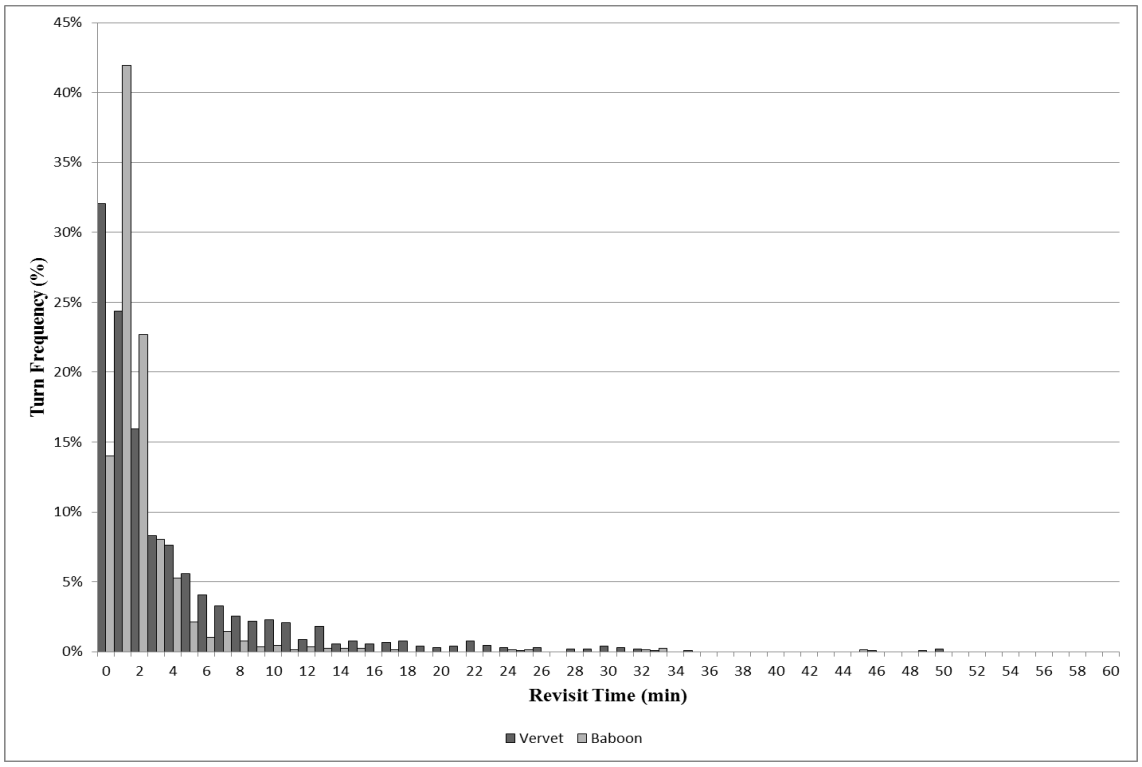

Figure 28: Turn frequency for vervets and baboons extracted to the nearest minute using the Douglas-Peucker line-smoothing algorithm. Dark grey bars represent the vervet turn frequencies, and light grey bars represent baboon turn frequency.

3.5 SYNCHRONIZED EUCLIDEAN DISTANCE

Synchronized Euclidean Distance (SED) results are displayed in Figures 29a and 29b as a combination of (a) the maximum point displacement from the interpolated line,

(b) the percentage of points which fall within the given proximity of the line, and (c) the revisit times from 5 to 60 minutes. This allows for the estimation of the spatial proximity that an interpolated point should be from its temporally coincidental point on the actual path, as well as the corresponding probability based on the revisit time of the interpolated line. For example, if an interpolated point for a vervet falls between two actual points with a revisit time of 20 minutes, based on the results displayed in Figure 29a the interpolated point has a 90% probability of being within ~27 meters of the animals actual location, or a 50% probability of being within ~8 meters.

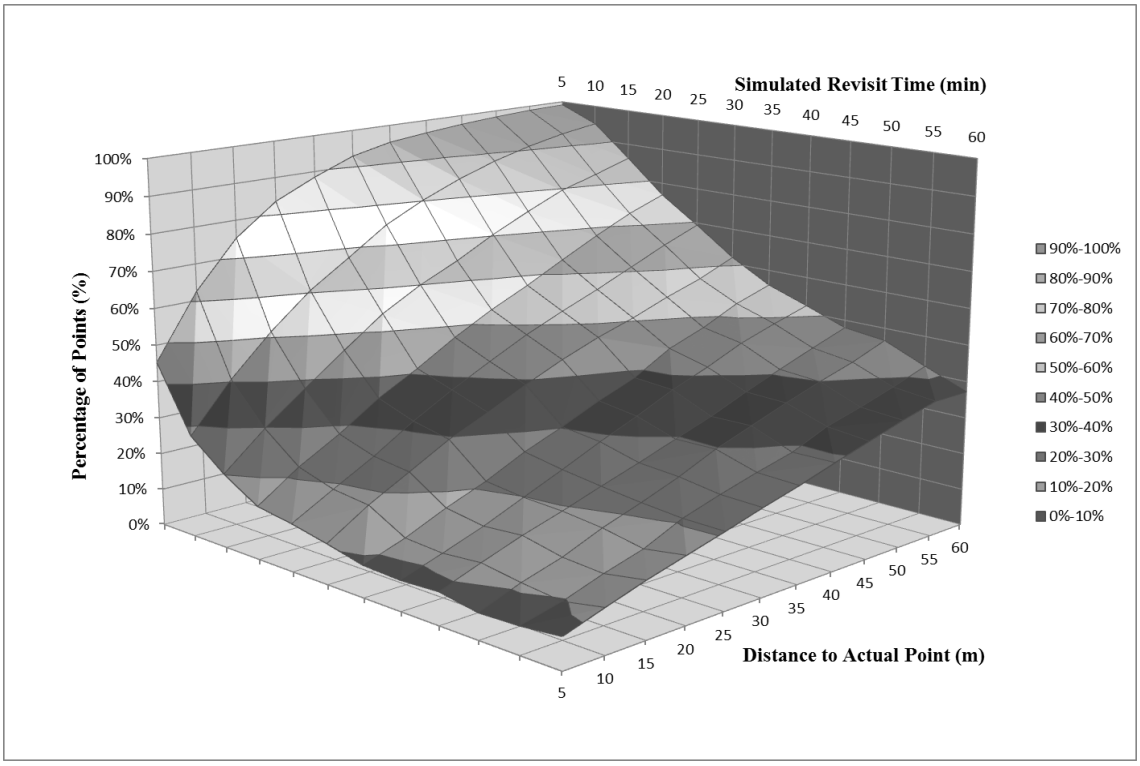

Figure 29a: Synchronized Euclidean Distance probability surface for baboons. The surface represents a probability surface, showing how probable it is for a point to lie a given distance from the interpolated line at any revisit time from 5 to 60 in 5 minute increments.

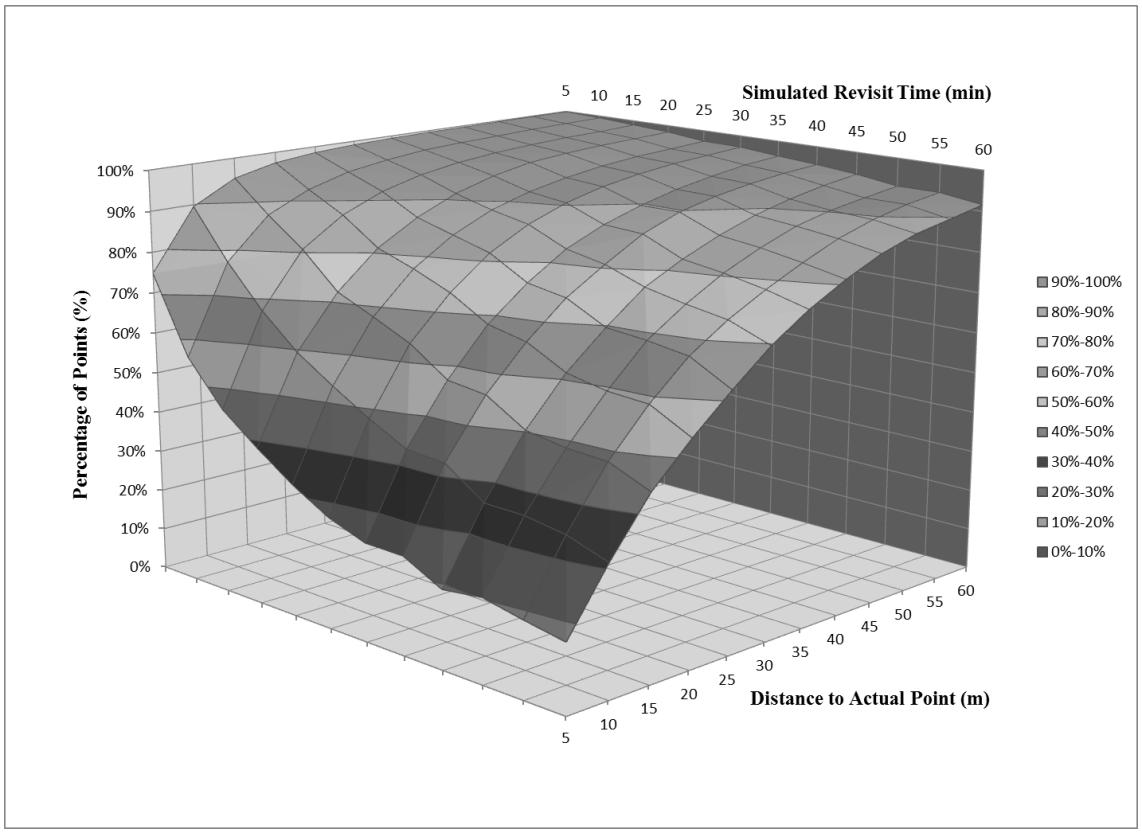

Figure 29b: Synchronized Euclidean Distance probability surface for vervets. The surface represents a probability surface, showing how probable it is for a point to lie a given distance from the interpolated line at any revisit time from 5 to 60 in 5 minute increments.

The two SED surfaces are different in shape, with the baboon surface (Figure 29a) maintaining a concave shape along the simulated revisit time axis as it progresses up the distance to actual point axis, while the vervet graph (Figure 29b) reaches a plateau at around 60 m displacement on the distance to actual point axis. This variance is due to differences in the scale of movement of the two species, with the baboons moving ~4.7 times farther than the vervets. If the maximum distance to the actual point axis for the baboon graph is extended from 60 m out to 120 m, a similar plateau begins to form as more points fall within the increased point displacements (Figure 30). This plateau is not as flat as it is in the vervets, with only 70% of the 60-minute revisit points explained by the graph at 120 m proximity from the interpolated line, compared to the vervet 60-

minute points which level off at 90% within 60 m of the line. While the distance to the actual point axis could be extended out until all points are explained at all revisit times, it is unlikely that a ±60 m tolerance will be useful to many research questions let alone one of ±120 m. If higher tolerances are acceptable the graph is easily extended.

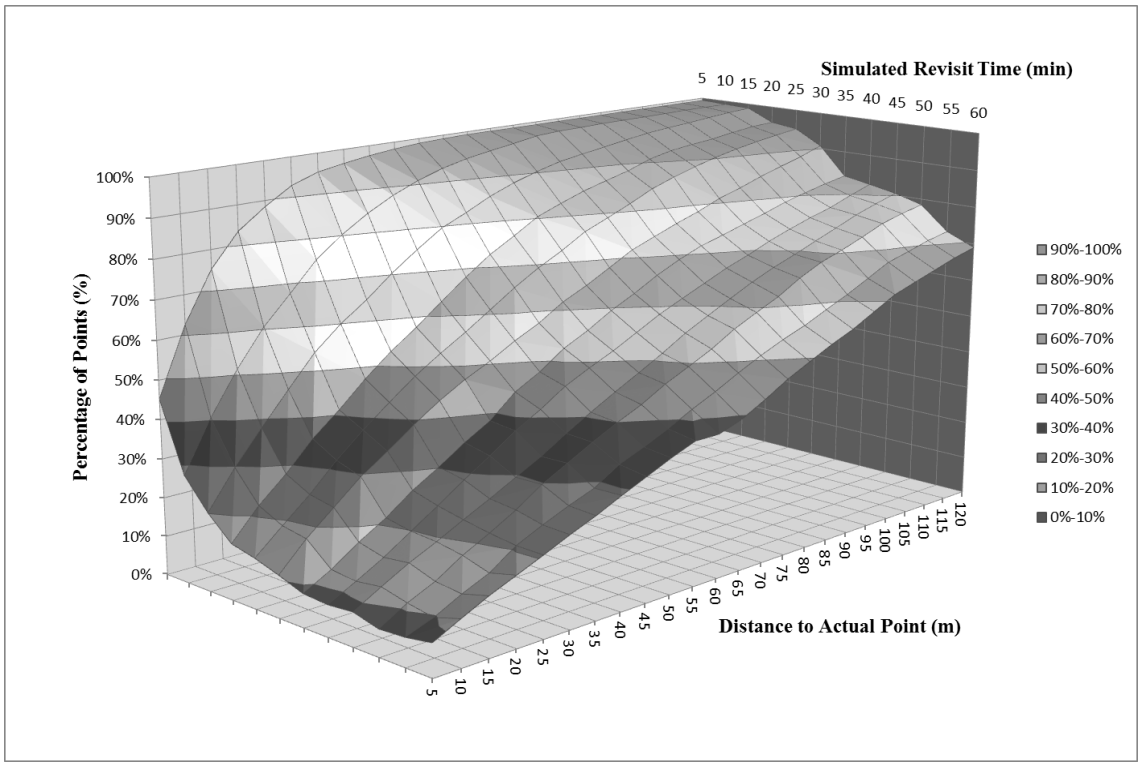

Figure 30: Synchronized Euclidean Distance probability surface for baboons extended out to 120 Meters. The surface represents a probability surface, showing how probable it is for a point to lie a given distance from the interpolated line at any revisit time from 5 to 60 in 5 minute increments. Decreasing revisit times shows a greater influence on the surface than increasing distance from the line.

Baboon paths that have been concatenated to create longer duration paths (Figure 31a) exhibited the same surface shape as shorter baboon follows (Figure 29a). The oscillations along the simulated revisit time axis of both of the scaled baboon graphs (Figure 31) are due to only one path being displayed rather than an averaging of fourteen paths, as in the previous probability graphs (Figures 29 and 30).

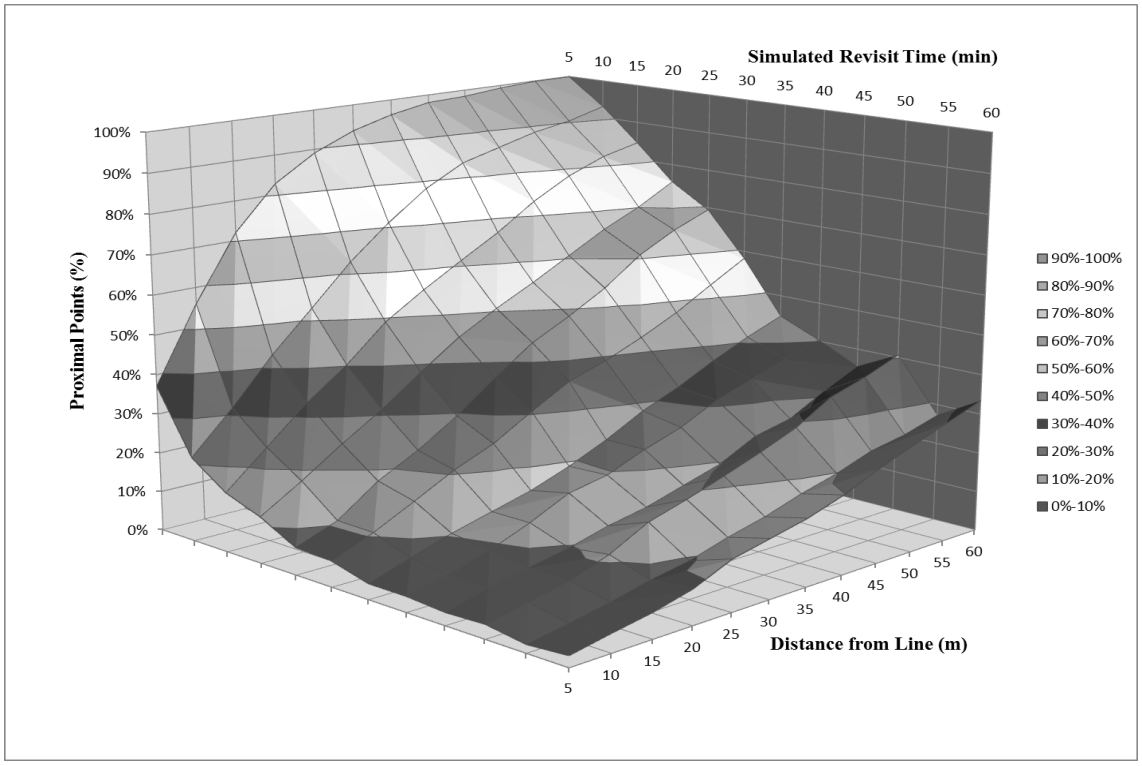

Figure 31a: Synchronized Euclidean Distance probability surface for baboons, scaled up temporally using path duration. The surface represents a probability surface, showing how probable it is for a point to lie a given distance from the interpolated line at any revisit time from 5 to 60 in 5 minute increments. Temporal scaling shows little surface change, while spatial scaling produces a surface similar to the vervet SED surface (Figure 30).

When the path/step length for a baboon is reduced by a factor of 4.7, its SED surface (Figure 31b) approximates the same shape as the vervet SED surface with its telltale plateau (Figure 29b). This demonstrates that scale was not a path duration issue caused by collecting shorter follows on the baboons, but a path/step length scale difference between species due to increased daily travel distances.

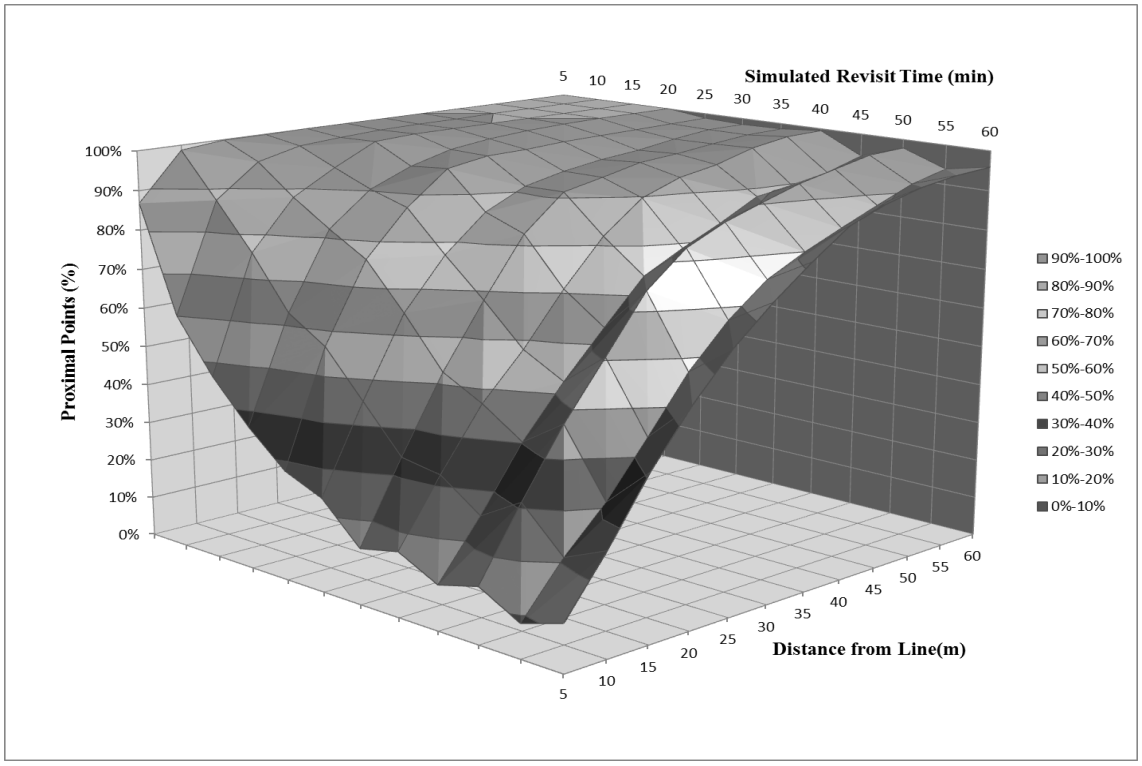

Figure 31b: Synchronized Euclidean Distance probability surface for baboons, scaled down spatially using path/step length. The surface represents a probability surface, showing how probable it is for a point to lie a given distance from the interpolated line at any revisit time from 5 to 60 in 5 minute increments. Temporal scaling shows little surface change, while spatial scaling produces a surface similar to the vervet SED surface (Figure 30).

3.6 ORIGINAL DATA SETS

3.6.1 ORIGINAL DATA TURN FREQUENCY

i. Baboons. Revisit times for the original baboon data displayed a positively skewed leptokurtic distribution (Kurtosis = 10.09, Skew = 3.27), with the peak of the distribution at the 6 minute revisit time. The high kurtosis and skew were caused by outliers as high as 920 minutes. When these outliers were not considered, the distribution approached a positively skewed mesokurtic distribution (Kurtosis = 2.08, Skew = 1.79). Over 90% of the baboon data points were collected within 15 minutes or less of each other, with 76% of these points collected in the 10 minute or less range (Figure 32). The high number of short revisit times was likely due to the relatively flat and simple

environment, enabling researchers to move quickly and easily while maintaining line of sight on the rest of the troop during data collection.

ii. Vervets. The original vervet revisit times were more temporally spread when compared to the baboon data, likely due to the animal's size and the complex three-dimensional nature of their habitat. The original vervet data also displayed a positively skewed leptokurtic distribution (Kurtosis = 15.79, Skew = 3.94), with the peak of the distribution at the five minute revisit. Like the baboon data, there were a number of outliers, when these were removed kurtosis and skew did decrease. Nevertheless, the vervet data did remain leptokurtic with a positive skew (Kurtosis = 7.46, Skew = 2.83). Fewer than 50% of the vervet data points were collected within fifteen minutes of each other, with ~35% of these points occurring in the 10 minute or less range (Figure 32).

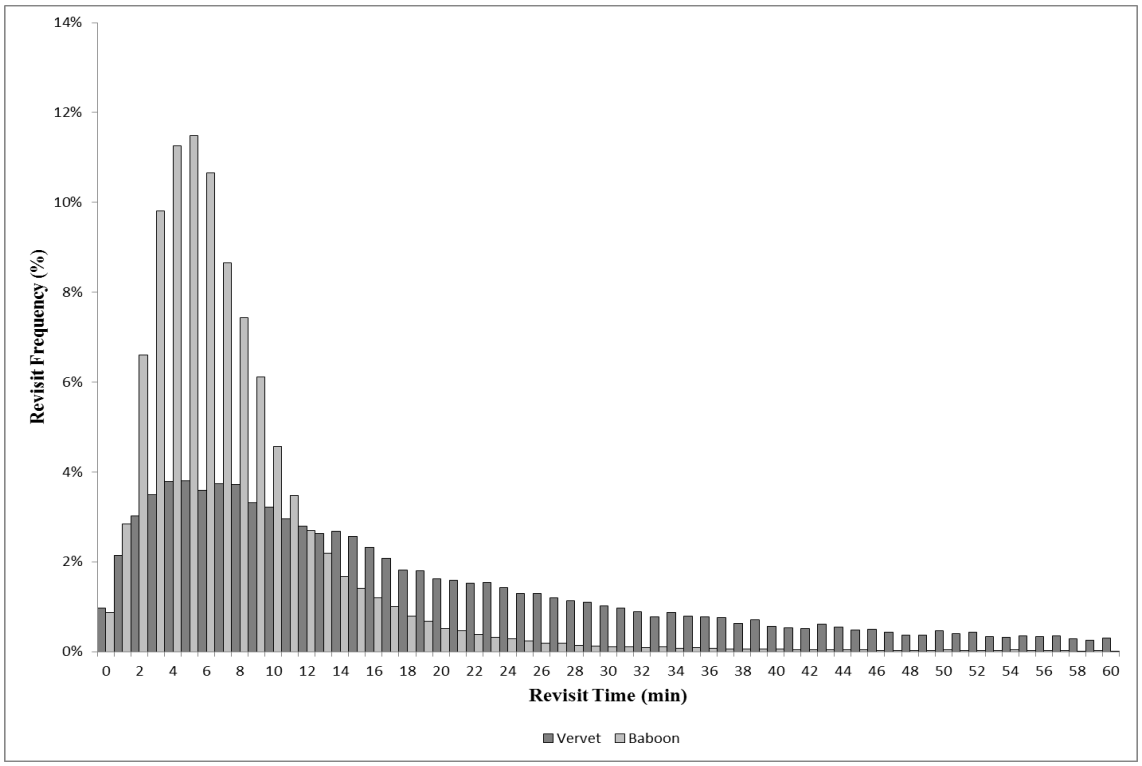

Figure 32: This graph represents the revisit frequency with which the original vervet and baboon datasets were collected. Over 99% of the baboon revisits and 88% of the vervet revisits are represented on this graph, these remaining revisits beyond 60 minutes were excluded.

3.6.2 LENGTH LOSS COMPARISON

i. Baboons. When the original baboon data are compared to the rate of length degradation observed during the length loss analysis described in Section 3.1, it is clear that the bulk of the original data were collected during the initial smoothing period, but well before the post 15 and 30 minute corner cutting oscillations (Figure 33).

ii. Vervets. When the original vervet data are compared to the expected length loss graph, it becomes obvious that, while ~50% of the data were collected during the initial smoothing phase, the remaining data points have revisit times that place them in the minor (post 15 minute) and major (post 30 minute) corner cutting phases (Figure 33).

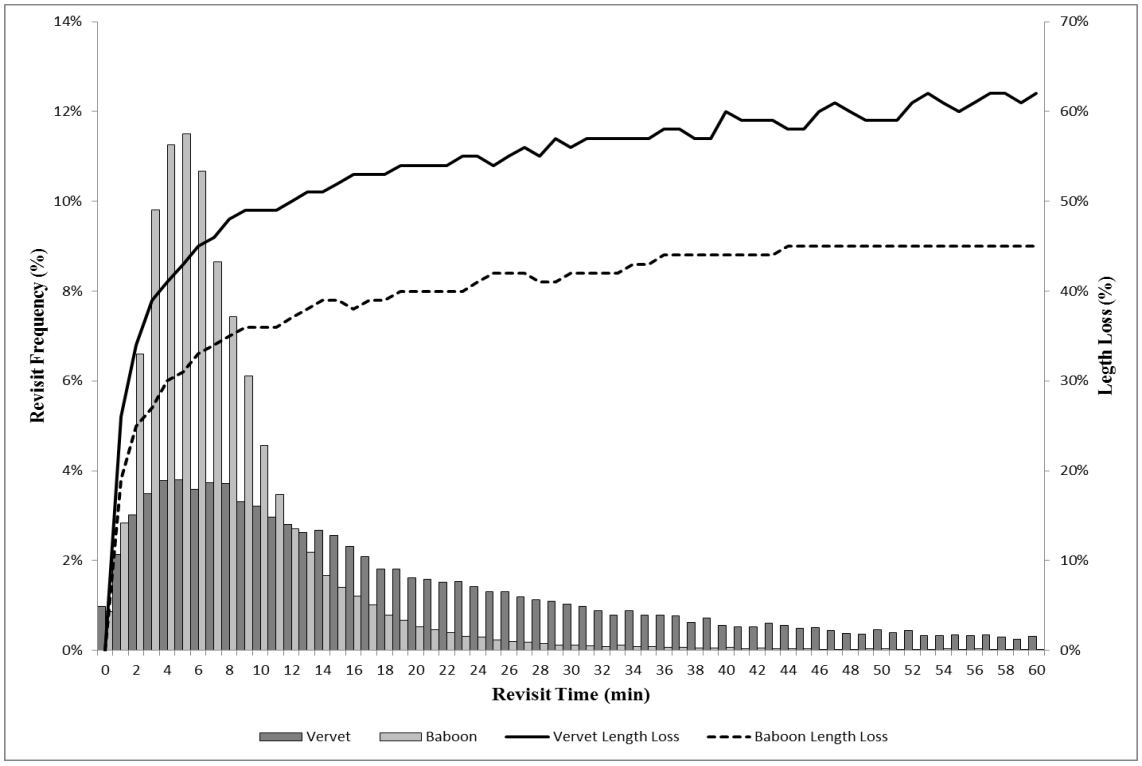

Figure 33: Comparison of length loss results to the original datasets collection frequencies. Revisit frequency is displayed as a bar graph along the primary axis, and length loss is displayed along the secondary axis as a line graph.

3.6.3 SYMMETRICAL DIFFERENCE COMPARISON

i. Baboons. Comparison of symmetrical differencing results to the original baboon data showed that 80% of the data were collected prior to the path shape degrading more than 21% from its original form, and before the oscillations caused by corner cutting occurred (Figure 34). Most of the baboon data (90%) were collected prior to the first major corner cutting oscillation at seventeen minutes.

ii. Vervets. The vervet symmetrical differencing graph exhibits a similar trend: only 35% of the revisits occurred prior to the corner cutting phase, at which point 29% of the shape had been lost (Figure 34). The first major oscillation in the vervet graph occurred at 21 minutes; by this point 60% of the points had been collected.

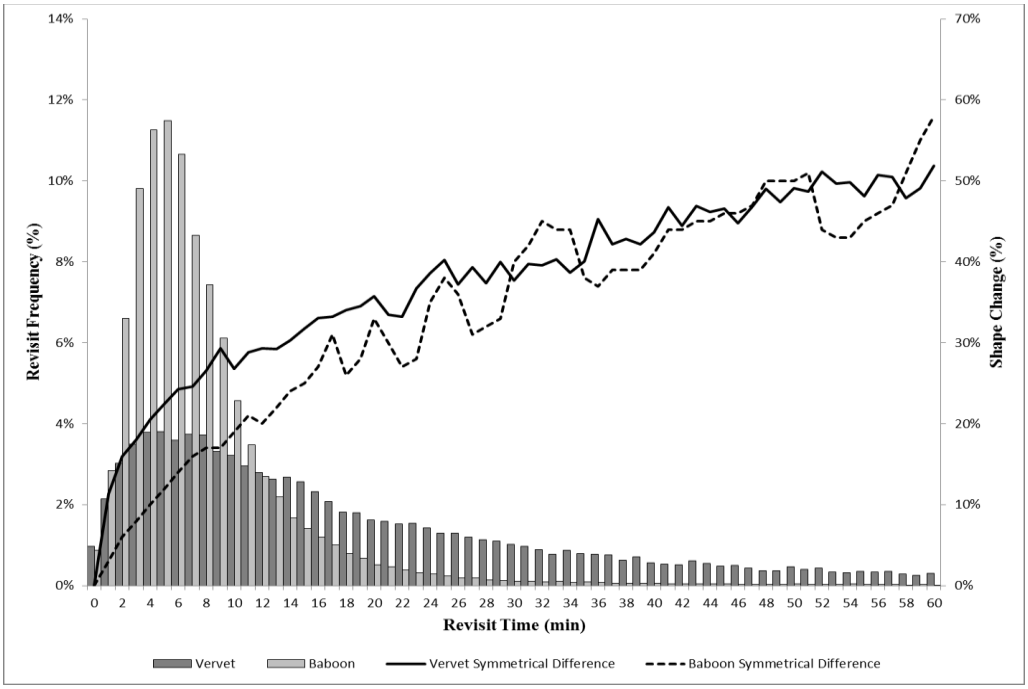

Figure 34: Comparison of symmetrical difference results to the original datasets collection frequencies. Revisit frequency is displayed as a bar graph along the primary axis, and shape change is displayed along the secondary axis as a line graph.

3.6.4 DOUGLAS-PEUCKER TURNS COMPARISON

i. Baboons. Approximately 90% of the turns extracted from the baboon continuous data using the Douglas-Peucker algorithm occurred within five minute of each other. Of those, over 50% occurred in the first minute. When all of the points in the original baboon data set are considered, only 43% were collected within five minutes of each other, with ~4% collected within a minute of each other.

ii. Vervets. Vervet data showed that ~90% of the turns extracted from continuous data occurred within 10 minutes of each other. Of those, approximately 50% occurred in the first minute. Only 35% of the points in the original vervet data set were collected within 10 minutes of each other, with ~3% collected within a minute of each other.

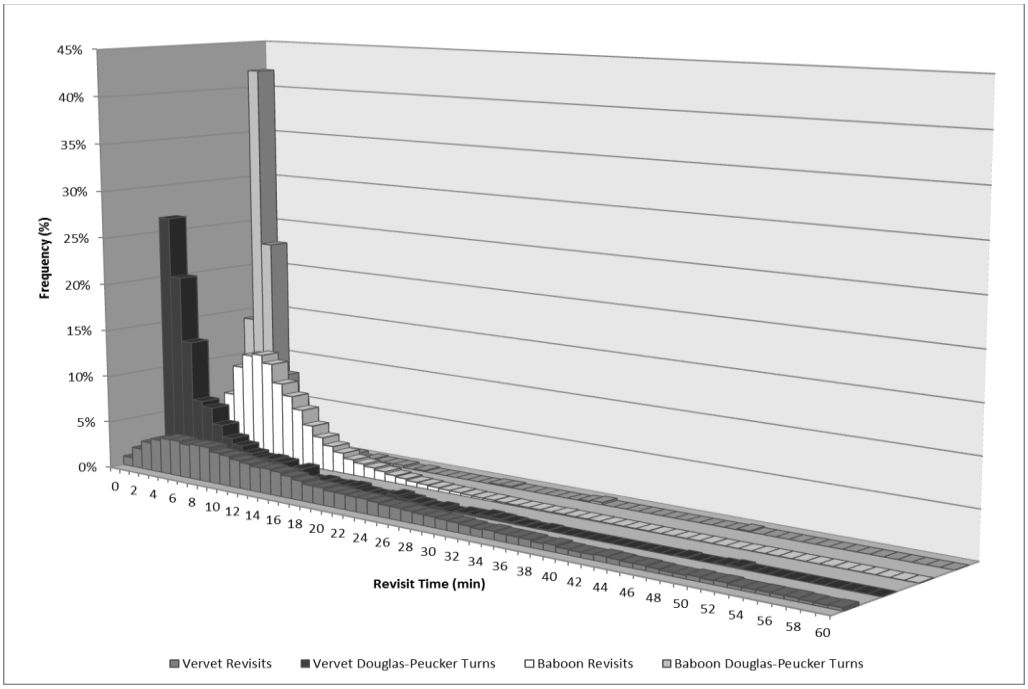

Figure 35: Douglas-Peucker extracted turns from continuous paths compared to original revisit times. Vervet revisits and turns are represented in the foreground by the darker bar graphs, while baboon's turns and revisits sit in the background in lighter tones.

3.6.5 SYNCHRONIZED EUCLIDEAN DISTANCE COMPARISON

i. Baboons. When the original baboon revisit times were compared to the Synchronized Euclidean Distance results, it was possible to infer that interpolated points would land within ~65 m or less of the actual path if they were placed on lines created with sub-15 minute revisit intervals. Of those sub-15 minute revisit points, ~50% would be within ~20 m or less of the actual path (Figure 36).

ii. Vervets. The vervet results showed much lower values when their revisit times were compared to the Synchronized Euclidean Distance results, 90 % of the sub-15 minute points were expected to be within ~25 m or less of actual path, and of those points ~50% could be expected to fall within ~5 m or less of the path (Figure 36).

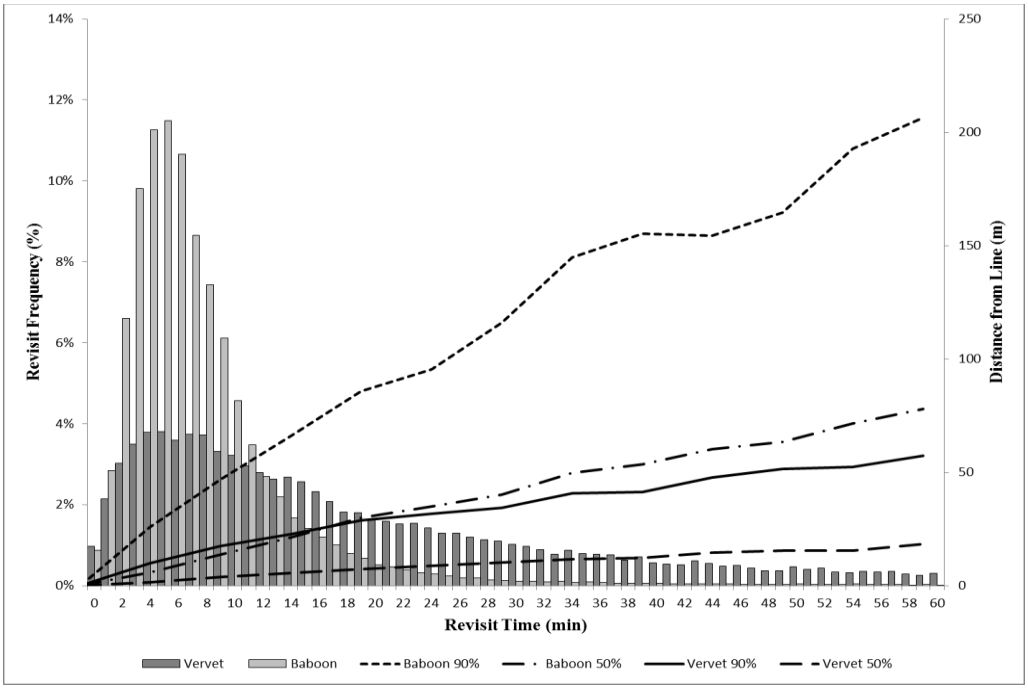

Figure 36: Revisit frequencies of vervets and baboons compared to the Synchronized Euclidean Distance 50% and 90% probability results. Revisit frequency is displayed as a bar graph along the primary axis, and Synchronized Euclidean Distance is displayed along the secondary axis as a line graph.

Like length loss and shape change, SED probability bands also exhibited a very high correlation between species, demonstrated here using the 90% band (R = 0.99, N = 13, R^2 = 0.98, P < 0.01), and the 50% band (R = 0.99, N = 13, R^2 = 0.99, P < 0.01).

3.7 ORIGINAL DATA PATH LENGTH CORRECTION

It is not possible to recreate the shape of a path, or even the original lines based on the percentage of area or length that was lost during increased revisit times. Because of the very strong correlation between revisit times and length loss it is possible to correct the length of the original path to a more accurate approximation using the revisit times in the original data and the results from the length loss analysis (Section 3.1).

$$PLC = \frac{l}{1-c} \quad [6]$$

PLC = Path length correction, l = the length of the interpolated line, c = the correction factor, or the percentage of underestimation for the interpolated line (expressed as a decimal). Path length correction was validated using extra continuous baboon paths that were collected late in the analysis, and as such were never used in the creation of the data being used for path length correction. Each path was randomly resampled to provide variation in the revisit times, 1% of the original 5-second data produced variable revisit times at a frequency similar to that which might be expected in the field (1% of 720 points per hour at 5 second intervals equals ~7 points per hour). Once resampled, the length of each leg of the path was corrected by the expected length loss results (Section 3.1; Appendix 1, Table 4), summed, and then compared to the actual path length. Based on these corrections, a path length reconstructed from 1% of the original points will be, on average, within 16% of the original line length, which provides an average improvement of 26% between the interpolated and corrected length (Appendix 1, Table

2). The more points available, the more accurate the adjustment was, 5% random resamples came within an average of 8% of the original path length, while 10% resamples were within 5% on average. When this is compared to results from an equal number of follows which were used in the length loss analysis (Section 3.1; Appendix 1, Table 3), it becomes clear that length correction for data not used in the original analysis performs well; to within 5% in comparison to data used to create the correction values. When path length correction was applied to the original vervet and baboon data sets, baboon path length increased by 51 to 60%, with an average growth of 56%, while vervet path length increased by 76 to 150%, with an average growth of 114% (Figure 37).

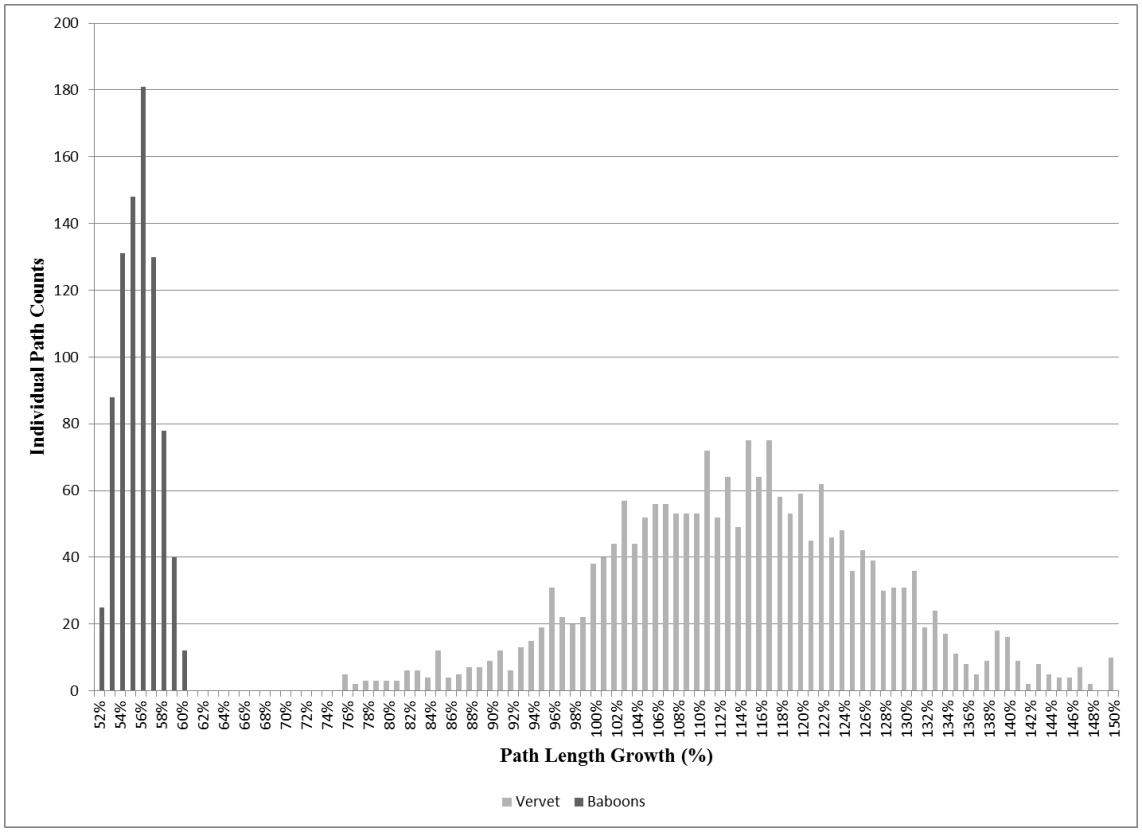

Figure 37: Path growth of original vervet and baboon paths using the length loss results to correct the interpolated lengths. The range of growth found in discrete paths is represented in dark grey for the baboons, and light grey for the vervets.

3.8 SYNCHRONIZED EUCLIDEAN DISTANCE BUFFERS

Just as length loss results cannot recreate the original path, SED measurements cannot pinpoint the original location of a point. However, SED can be used to assign buffers to points which provide an area of probability, which will also have a margin of error assigned to them based on the values from the SED analysis. Synchronized Euclidean Distance exhibits a very strong correlation to revisit times, when measured using the 90% probability data (Baboons: $R = 0.99$, $R^2 = 0.98$, $P < 0.01$, Vervets: $R = 0.98$, $R^2 = 0.98$, $P < 0.01$). Using the 90% and 50% values, one of the interpolated points below has been assigned buffers to represent the area and probability where the point is expected to be positioned (Figure 38). By adding more buffers at regular intervals (every 5 or 10%), a probability surface can be created for any point.

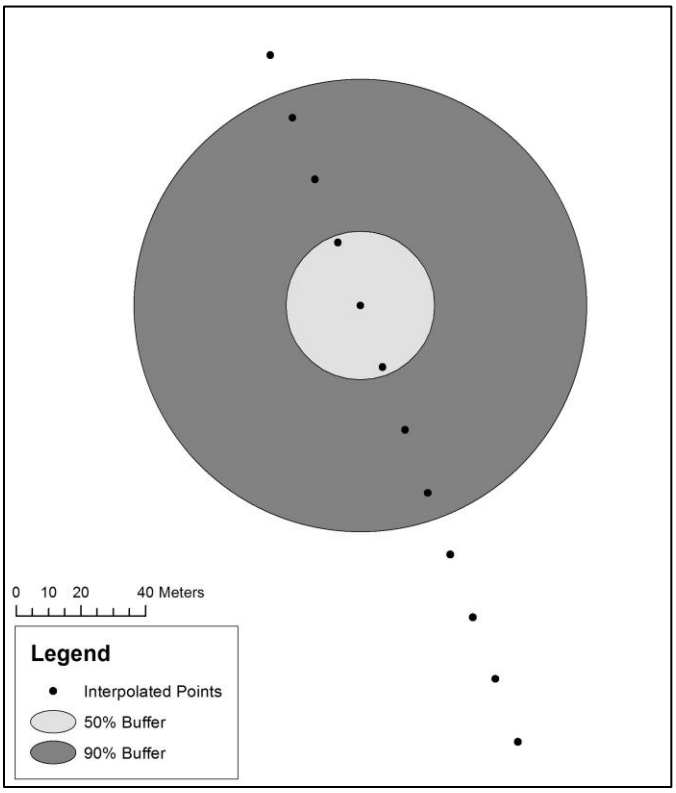

Figure 38: Buffers representing the probable location of the original point using the Synchronized Euclidean Distance results to determine buffer radius. The light grey buffer has a 50% chance of containing the original point, while the dark grey buffer has a 90% of containing it.

This can be applied to topological problems where the relative location of multiple individuals is being measured (Figure 39). By area weighting the buffers it is possible to determine the adjacency of a point. The likelihood that Point 3 will be to the right of points 1 and 2 can be easily measured using this technique; if they occupied the proximal edges of their respective buffers with Point 3.

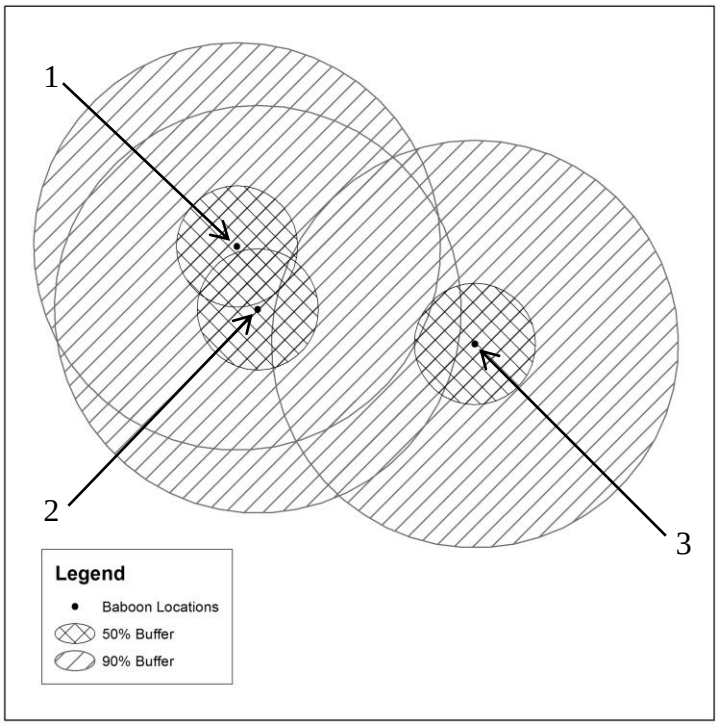

Figure 39: Buffers around interpolated baboon locations can be used to determine the topological adjacency of the original points. Points where buffers overlap can be given a probability of whether they are left, right, above or below another point using area weighting.

CHAPTER FOUR

DISCUSSION

4.1 INTERPRETATION

The length loss data analysis established both visually and through piecewise regression that two phases of path change and, consequently, data loss occurred during path interpolation from points: an initial path smoothing phase, followed by a corner cutting phase. The path smoothing phase occurred prior to ten minute revisits, representing the smoothing of short legs in the subject's movement, while the corner cutting phase became evident at around 15 minute revisit times, increasing in magnitude around the 30 minute revisits. Interspecies similarities in the length loss graphs were found to be highly correlated (Section 3.1), suggesting that the two phases are the product of linear interpolation, and are not limited to any one species, while also establishing that this technique can be used across species.

Similar to length loss, shape change also exhibited a very high correlation between baboons and vervets. The data also indicate that shape change is likewise very highly correlated to length loss in both species. Shape change results revealed path smoothing and corner cutting phases similar to those found in the length loss results, and while the breakpoint could not be distinguished using piecewise regression as it was in the length loss data, oscillations in the corner cutting phase were more pronounced than those observed in the length loss analysis, providing clear indication of when the latter phase started (Figure 26). This indicates with both species that data should be collected using sub 15 minute revisit times if path shape is to be preserved, keeping in mind that the lower the revisit time is, the more accurate the shape will be. The vervet path shape

did degrade more steeply during the sub 15 minute revisits than the baboon path shape, highlighting a possible need to collect data more frequently for vervets.

Turns extracted using the Douglas-Peucker algorithm indicated that most critical trajectory evolutions occurred in the sub 5 minute range for baboons and sub 10 minutes for the vervets (Figure 35). For studies with a focus on recording turns it would be crucial to ensure that revisit times are set to the lowest common denominator. Consequently, this would require for baboons and vervets that points be collected every minute or as continuous data. Even if points were collected at the above rates, it may still prove difficult to extricate every turn, as many turns were not sharp angles followed by long legs, but instead occurred as sweeping curves. As such it may not be practical to attempt to capture turns with anything less than continuous data, which is subsequently run through a line simplification algorithm as mentioned above.

Of all the analyses carried out during this research, Synchronized Euclidean Distance exhibited the strongest correlation to revisit times. SED measurements calculated every minute between the interpolated line and the actual line provided a revealing picture of error across revisit times (Figure 29). Interspecies differences in the SED probability surfaces reflect different scales of movement, with baboons travel distances manifesting their increased travel distances as greater levels of spatial offset during the higher revisit times. However, when the bands of the probability surface are isolated to provide a two-dimensional view, a highly correlated linear relationship between SED and revisit times becomes obvious (Figure 36), suggesting that like length loss and shape change, increases in SED are the direct product of linear interpolation and not isolated to any one species.

In order to minimize data loss it seems best to capture data points during the early stages of path smoothing and preferably prior to the corner cutting phase. While this will hold true in most instances, it will also depend on the analytical questions being explored with the data. If the research question involves using accurate path shapes, then minimizing revisit times will be important as the later stages of corner cutting will have adverse effects on path shape as shown in Section 3.3. However, if path length is the focus, then this research provides a means of correcting path length shortfalls due to linear interpolation (Section 3.5), both original vervet and baboon data sets showed an average growth in path length of over 50% (Figure 37). For questions were the relative position of multiple entities is being studied, SED can be used to assign buffers to points, providing a radius were the original point could have been, and the probability that the original point was in the buffer based on the values from the SED analysis (Figures 38 & 39).

4.2 FUTURE WORK

It is possible to measure sinuosity for shorter periods or to constrain a circular path to a channel, so that the shortest path length is measured along the channel to avoid extremely high values. One possible approach considered was to treat the edges of the group in which the individual was travelling as the sides of a channel, with the assumption that by choosing to travel with a group the individual constrains itself to travel within the group's perimeter. The task of determining the appropriate period or exploring methods of channeling the path was beyond the scope of this research and may be reconsidered at a later date.

In this study SED results were reported as an average result for each revisit interval. In future work it may prove more useful to report the SED results based on their distance from the original points. It is predicted that interpolated points closer to the original point will have a lower SED when compared to the interpolated points found near the mid-point between the original points. This technique is expected to provide a series of expanding and contracting buffers as the spatio-temporal distance between the interpolated and original points increased and decreased, shrinking the majority of the buffers to provide a more realistic estimation of SED.

Interpolated paths for this study were not physically constrained by an established network (streets, sidewalks, trails). As such, the subjects were theoretically able to move about the environment unhindered. This leads to an interesting question about animal movement: Are there underlying, identifiable networks of paths and environmental constraints which consistently influence animal movement, and if so can interpolation be confined to this network in order to reduce error?

4.3 CONCLUSION

As more scientists investigating animal behaviour with a focus on individual spatio-temporal interactions which influence group behaviour (Couzin and Krause 2003), the need for accurate estimations of individual positions will continue to be an important factor in their research. This study has shown a very high positive correlation between revisit times and various aspects of change in the spatial accuracy of an interpolated path, such as length loss, shape change, and synchronized Euclidean distance, confirming the original expectation that it is ultimately a reduction and optimization of the temporal gaps between revisit or collection times that will increase interpolation accuracy, by

decreasing the overall uncertainty of a path. This reduction in revisit times must be balanced out by what is realistically achievable in the field.

Two significant phases of data loss, path smoothing and corner cutting have also been established by this research, representing two different processes which drive the spatial uncertainty in linear interpolation. This research provides a means to assign a level of spatial certainty to previously collected intermittent data (Clarke 2008, Thomas 2013) by measuring the extent of the estimation error caused by linear interpolation in continuous paths and applying these results to the intermittent paths. These methods offer other researchers a means to test their own study species and find the optimum target revisit time for their data collection, while establishing the level of error which they can expect as the temporal gap between data points increases.

In summary, data collected using higher revisit times comes with a cost, increased error during interpolation. Based on the results of this study it is possible to propose that in most cases data should be collected prior to the corner cutting phase, while error is the most predictable, and before increased variability in error is introduced during the corner cutting phase. For the two species used in this study this would provide a target revisit time of less than ten to fifteen minutes. Data collected during the corner cutting phase should be used only after giving careful consideration to the spatial uncertainty being introduced into the analysis.

REFERENCES

Andrienko, N., G. Andrienko, N. Pelekis & S. Spaccapietra. 2008. Basic concepts of movement data. In *Mobility, Data Mining and Privacy*, 15-38. Springer.

Aureli, F., C. M. Schaffner, N. Asensio & D. Lusseau (2012) What is a subgroup? How socioecological factors influence interindividual distance. *Behavioral Ecology,* 23**,** 1308-1315.

Aureli, F., C. M. Schaffner, C. Boesch, S. K. Bearder, J. Call, C. A. Chapman, R. Connor, A. Di Fiore, R. I. Dunbar & S. P. Henzi (2008) Fission‐fusion dynamics. *Current Anthropology,* 49**,** 627-654.

Barrett, L., D. Gaynor, D. Rendall, D. Mitchell & S. P. Henzi (2004) Habitual cave use and thermoregulation in chacma baboons (Papio hamadryas ursinus). *Journal of Human Evolution,* 46**,** 215-222.

Barrett, L. & S. P. Henzi (2013) Social Coordination: Patience Is a Virtue for Vervet Monkeys. *Current Biology,* 23**,** R311-R313.

Geospatial Modelling Environment. 0.7.2.1, Spatial Ecology LLC.

Bowman, J. L., C. O. Kochanny, S. Demarais & B. D. Leopold (2000) Evaluation of a GPS Collar for White-Tailed Deer. *Wildlife Society Bulletin,* 28**,** 141-145.

Breitenmoser, U., C. Breitenmoser-Würsten, G. Zuleta, F. Bernhart & M. O'Donoghue (1992) A method to estimate travel distances of fast moving animals. *Biotelemetry XII (R. Mancini, S. Fioretti, C. Cristalli, and P. Bedini, eds.). Litografia Felici, Ancona, Italy***,** 318-326.

Burdett, C. L., A. M. Ron, G. J. Niemi & L. D. Mech (2007) Defining Space Use and Movements of Canada Lynx with Global Positioning System Telemetry. *Journal of Mammalogy,* 88**,** 457-467.

Clarke, P. 2008. DeHoop Nature Reserve Baboon Data.

Couzin, I. D. & J. Krause (2003) Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior,* 32.

Crist, T. O., D. S. Guertin, J. A. Wiens & B. T. Milne (1992) Animal Movement in Heterogeneous Landscapes: An Experiment with Eleodes Beetles in Shortgrass Prairie. *Functional Ecology,* 6**,** 536-544.

Douglas, D. H. & T. K. Peucker (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization,* 10**,** 112-122.

ArcGIS. 10.1 SP1, ESRI, Redlands, CA.

Freeman, R., R. Mann, T. Guilford & D. Biro (2011) Group decisions and individual differences: route fidelity predicts flight leadership in homing pigeons (Columba livia). *Biology letters,* 7**,** 63-66.

Friedman, M. (1962) The Interpolation of Time Series by Related Series. *Journal of the American Statistical Association,* 57**,** 729-757.

Hägerstraand, T. (1970) What About People in Regional Science? *Papers in Regional Science,* 24**,** 7-24.

Halmos, P. R. 1960. *Naive Set Theory. The University Series in Undergraduate Mathematics*. Van Nostrand Reinhold Company, New York, NY.

Hamilton, W. J. & J. Bulger (1992) Facultative expression of behavioral differences between one‐male and multimale savanna baboon groups. *American Journal of Primatology,* 28**,** 61-71.

Harris, S., W. Cresswell, P. Forde, W. Trewhella, T. Woollard & S. Wray (1990) Home-range analysis using radio-tracking data - a review of problems and techniques particularly as applied to the study of mammals. *Mammel Rev.,* 20**,** 26.

Harrison, M. J. (1983) Territorial behaviour in the green monkey, Cercopithecus sabaeus: seasonal defense of local food supplies. *Behavioral Ecology and Sociobiology,* 12**,** 85-94.

Heezen, K. L. & J. R. Tester (1967) Evaluation of Radio-Tracking by Triangulation with Special Reference to Deer Movements. *The Journal of Wildlife Management,* 31**,** 124-141.

Herbert-Read, J. E., A. Perna, R. P. Mann, T. M. Schaerf, D. J. Sumpter & A. J. Ward (2011) Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences,* 108**,** 18726-18731.

Johnson, C. J., K. L. Parker, D. C. Heard & M. P. Gillingham (2002) Movement Parameters of Ungulates and Scale-Specific Responses to the Environment. *Journal of Animal Ecology,* 71**,** 225-235.

Jolly, C. J. 1993. Species, subspecies, and baboon systematics. In *Species, species concepts and primate evolution*, 67-107. Springer.

Katz, M. & E. George (1985) Fractals and the analysis of growth paths. *Bulletin of Mathematical Biology,* 47**,** 273-286.

Katz, M. J. 1986. *Templets and the explanation of complex patterns*. CUP Archive.

Kelvin, W. T. B. & P. G. Tait. 1885. *Elements of natural philosophy*. University Press.

King, A. J., C. M. S. Douglas, E. Huchard, N. J. B. Isaac & G. Cowlishaw (2008) Dominance and Affiliation Mediate Despotism in a Social Primate. *Current Biology,* 18**,** 1833-1838.

Leopold, L. B., M. G. Wolman & J. P. Miller. 1964. *Fluvial Processes in Geomorphology*. San Francisco: W.H. Freeman and Company.

Lett, C. & V. Mirabet (2008) Modelling the dynamics of animal groups in motion. *South African Journal of Science,* 104**,** 192-198.

Lonergan, M., M. Fedak & B. McConnell (2009) The effects of interpolation error and location quality on animal track reconstruction. *Marine Mammal Science,* 25**,** 275-282.

Matisziw, T. C. & E. Demir (2012) Inferring network paths from point observations. *International Journal of Geographical Information Science,* 26**,** 1979-1996.

Melnick, D. & M. Pearl. 1987. Cercopithecines in multimale groups: genetic diversity and population structure. In *Primate Societies,* eds. B. Smuts, D. Cheney, R. Seyfarth, R. Wrangham & T. Struhsaker, 121-134. Chicago: University of Chicago Press.

Microsoft Excel. 14.0.6023.1000 (32-bit), Microsoft Corporation.

Musiani, M., Okarma, H., & Jędrzejewski, W. (1998) Speed and actual distances travelled by radiocollared wolves in Białowieża Primeval Forest (Poland). *Acta Theriologica,* 43**,** 7.

Nagy, M., Z. Ákos, D. Biro & T. Vicsek (2010) Hierarchical group dynamics in pigeon flocks. *Nature,* 464**,** 890-893.

Napier, J. & P. Napier. 1985. *The Natural History of the Primates*.

Pasternak, G., L. R. Brown, S. Kienzle, A. Fuller, L. Barrett & S. P. Henzi (2013) Population ecology of vervet monkeys in a high latitude, semi-arid riparian woodland. *Koedoe,* 55**,** 01-09.

Pasternak, G. M. 2011. Environmental effects on group structure and vigilance in vervet monkeys. In *Psychology*. Lethbridge, Alberta: University of Lethbridge.

Pépin, D., C. Adrados, C. Mann & G. Janeau (2004) Assessing Real Daily Distance Traveled by Ungulates Using Differential GPS Locations. *Journal of Mammalogy,* 85**,** 774-780.

Potamias, M., K. Patroumpas & T. Sellis. 2006. Sampling trajectory streams with spatiotemporal criteria. In *Scientific and Statistical Database Management, 2006. 18th International Conference on*, 275-284. IEEE.

Rasmussen, D. (1991) Observer influence on range use of Macaca arctoides after 14 years of observation. *Laboratory Primate Newsletter,* 30**,** 6-11.

Reynolds, T. D. & J. W. Laundre (1990) Time Intervals for Estimating Pronghorn and Coyote Home Ranges and Daily Movements. *The Journal of Wildlife Management,* 54**,** 316-322.

Strang, G. & T. Nguyen. 1996. *Wavelets and filter banks*. SIAM.

Struhsaker, T. T. (1967a) Ecology of vervet monkeys (Cercopithecus aethiops) in the Masai-Amboseli game reserve, Kenya. *Ecology***,** 892-904.

--- (1967b) Social structure among vervet monkeys (Cercopithecus aethiops). *Behaviour***,** 83-121.

Sumner, M. D., S. J. Wotherspoon & M. A. Hindell (2009) Bayesian estimation of animal movement from archival and satellite tags. *PLoS One,* 4**,** e7324.

Thomas, B. L. 2013. Population density, spatial dynamics and territoriality in vervet monkeys In *Psychology*. Lethbridge, Albeta: University of Lethbridge.

Tourtellot, M. K., R. D. Collins & W. J. Bell (1991) The problem of movelength and turn definition in analysis of orientation data. *Journal of theoretical biology,* 150**,** 287-297.

Tremblay, Y., S. A. Shaffer, S. L. Fowler, C. E. Kuhn, B. I. McDonald, M. J. Weise, C. A. Bost, H. Weimerskirch, D. E. Crocker, M. E. Goebel & D. P. Costa (2006) Interpolation of animal tracking data in a fluid environment. *J Exp Biol,* 209**,** 128-40.

Trinkaus, J. (1984) Mechanisms of Cell Movement in vitro. *Cells into Organs. The Forces that Shape the Embryo***,** 179-244.

Turchin, P. 1998. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sunderland, Massachusetts: Sinauer Assosiates, Inc. Publishers.

Wentz, E. A., A. F. Campbell & R. Houston (2003) A comparison of two methods to create tracks of moving objects: linear weighted distance and constrained random walk. *International Journal of Geographical Information Science,* 17**,** 623-645.

NetLogo. 5.0.5, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Willems, E. P. & R. A. Hill (2009) Predator-specific landscapes of fear and resource distribution: effects on spatial range use. *Ecology,* 90**,** 546-555.

Williamson, E. A. & A. T. Feistner (2003) Habituating primates: processes, techniques, variables and ethics. *Field and laboratory methods in primatology: A practical guide***,** 25-39.

Wolfheim, J. H. 1983. *Primates of the world: distribution, abundance and conservation*. Psychology Press.

Wrangham, R. (1981) Drinking competition in vervet monkeys. *Animal Behaviour,* 29**,** 904-910.

Zheng, Y. & X. Zhou. 2011. *Computing with spatial trajectories*.

APPENDIX ONE

TABLES

Table 1: Composition of adult vervets (Thomas 2013), and baboons (Clarke 2008) in the study groups.

Vervets	RBM Troop	RST Troop
Males	13	25
Females	21	22
Baboons		
Males	3	
Females	10	

Table 2: Corrected path length using follows not used in length loss analysis.

Path	Sample Size	Actual Length (m)	Resampled Length (m)	Corrected Length (m)	Resampled Loss (%)	Corrected Length (%)	Length Improvement (%)
Bab15	10%	2112	1628	2020	23%	4%	19%
	5%	2112	1388	1865	34%	12%	23%
	1%	2112	994	1547	53%	27%	26%
Bab16	10%	3041	2463	3088	19%	2%	17%
	5%	3041	2317	3110	24%	2%	22%
	1%	3041	1941	3158	36%	4%	32%
Bab17	10%	4961	4057	5056	18%	2%	16%
	5%	4961	3819	5154	23%	4%	19%
	1%	4961	3354	5246	32%	6%	27%
Bab18	10%	1604	1188	1494	26%	7%	19%
	5%	1604	1102	1480	31%	8%	24%
	1%	1604	745	1159	54%	28%	26%
Bab19	10%	4594	3472	4280	24%	7%	18%
	5%	4594	3243	4385	29%	5%	25%
	1%	4594	2685	4242	42%	8%	34%
Bab20	10%	1672	1192	1507	29%	10%	19%
	5%	1672	1075	1461	36%	13%	23%
	1%	1672	821	1306	51%	22%	29%
Bab21	10%	3433	2832	3512	18%	2%	15%
	5%	3433	2746	3774	20%	10%	10%
	1%	3433	2483	4105	28%	20%	8%

Table 3: Corrected path length using follows from length loss analysis.

Path	Sample Size	Actual Length (m)	Resampled Length (m)	Corrected Length (m)	Resampled Loss (%)	Corrected Length (%)	Length Improvement (%)
Bab01	10%	3061	2620	3292	14%	8%	7%
	5%	3061	2493	3386	19%	11%	8%
	1%	3061	2309	3678	25%	20%	4%
Bab03	10%	2876	2235	2797	22%	3%	20%
	5%	2876	2040	2781	29%	3%	26%
	1%	2876	1695	2561	41%	11%	30%
Bab05	10%	1877	1459	1821	22%	3%	19%
	5%	1877	1369	1794	27%	4%	23%
	1%	1877	1219	1927	35%	3%	32%
Bab07	10%	2615	2024	2506	23%	4%	18%
	5%	2615	1794	2356	31%	10%	22%
	1%	2615	1280	2125	51%	19%	32%
Bab09	10%	2555	2138	2670	16%	5%	12%
	5%	2555	1873	2455	27%	4%	23%
	1%	2555	1666	2655	35%	4%	31%
Bab11	10%	2553	1989	2535	22%	1%	21%
	5%	2553	1917	2578	25%	1%	24%
	1%	2553	1622	2782	36%	9%	27%
Bab13	10%	2963	2372	2975	20%	0%	20%
	5%	2963	2281	2996	23%	1%	22%
	1%	2963	2056	3340	31%	13%	18%

Table 4: Length loss results table used during path length correction.

Revisit (min)	Length Loss (%)		Revisit (min)	Length Loss (%)	
	Vervet	Baboon		Vervet	Baboon
1	26%	19%	31	57%	42%
2	34%	25%	32	57%	42%
3	39%	27%	33	57%	42%
4	41%	30%	34	57%	43%
5	43%	31%	35	57%	43%
6	45%	33%	36	58%	44%
7	46%	34%	37	58%	44%
8	48%	35%	38	57%	44%
9	49%	36%	39	57%	44%
10	49%	36%	40	60%	44%
11	49%	36%	41	59%	44%
12	50%	37%	42	59%	44%
13	51%	38%	43	59%	44%
14	51%	39%	44	58%	45%
15	52%	39%	45	58%	45%
16	53%	38%	46	60%	45%
17	53%	39%	47	61%	45%
18	53%	39%	48	60%	45%
19	54%	40%	49	59%	45%
20	54%	40%	50	59%	45%
21	54%	40%	51	59%	45%
22	54%	40%	52	61%	45%
23	55%	40%	53	62%	45%
24	55%	41%	54	61%	45%
25	54%	42%	55	60%	45%
26	55%	42%	56	61%	45%
27	56%	42%	57	62%	45%
28	55%	41%	58	62%	45%
29	57%	41%	59	61%	45%
30	56%	42%	60	62%	45%

Table 5: Metadata for continuous follows of baboons and vervets.

ID	Collection Date	Travel Time	Point Total
Bab01	12-Mar-13	2:16:50	1643
Bab02	20-Mar-13	2:18:24	1574
Bab03	12-Mar-13	2:22:00	1705
Bab04	13-Mar-13	2:07:10	1527
Bab05	08-Mar-13	1:58:35	1424
Bab06	19-Mar-13	2:15:10	1623
Bab07	10-Mar-13	2:18:05	1658
Bab08	08-Mar-13	3:51:50	1557
Bab09	14-Apr-13	2:11:05	1574
Bab10	12-Apr-13	2:05:25	1506
Bab11	19-Apr-13	2:25:10	1738
Bab12	16-Apr-13	2:27:50	1775
Bab13	18-Apr-13	2:02:06	1466
Bab14	21-Apr-13	2:44:20	1973
Bab15	15-Sep-13	2:14:45	1618
Bab16	06-Oct-13	2:37:25	1890
Bab17	07-Oct-13	4:21:15	3136
Bab18	08-Oct-13	1:44:50	1259
Bab19	09-Oct-13	4:08:00	2977
Bab20	09-Oct-13	2:16:20	1637
Bab21	11-Oct-13	2:47:20	2009
Ver01	10-Aug-12	10:23:32	7321
Ver02	07-Sep-12	11:43:50	8270
Ver03	22-Aug-12	10:50:20	7767
Ver04	30-Aug-12	11:17:19	7934
Ver05	06-Sep-12	11:05:34	7413
Ver06	03-Aug-12	10:34:58	6995
Ver07	07-Mar-12	10:24:38	7282
Ver08	14-Mar-13	10:44:00	7146
Ver09	15-Mar-13	5:34:37	3670
Ver10	22-Jul-13	5:56:00	3999

APPENDIX TWO

NET LOGO CODE

```
to setup
  clear-all
  crt 1 [setxy 100 400 set color yellow pen-down]
  ask turtles [set heading 0]
  reset-ticks
end

to go
  if ticks >= 3600 [export-output (word "E:/ <Directory Location> " Name ".csv") stop]
  ask turtles [rt .1 left 45 forward .5]
  ask turtles [output-type xcor output-type ", " output-print ycor]
  ask turtles [rt 45 forward .5]
  ask turtles [output-type xcor output-type ", " output-print ycor]
  tick
end
```