

DEVELOPING A WILDLIFE TRACKING EXTENSION FOR ArcGIS

Cai Chen, B.A.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

May 2009

APPROVED:

Pinliang Dong, Major Professor

Yan Huang, Co-Major Professor

Steven Wolverton, Committee Member

Paul Hudak, Chair of the Department of
Geography

Michael Monticino, Interim Dean of the Robert
B. Toulouse School of Graduate
Studies

Chen, Cai. Developing a wildlife tracking extension for ArcGIS. Master of Science (Applied Geography), May 2009, 51 pp., 2 tables, 28 illustrations, reference list, 76 titles.

Wildlife tracking is an essential task to gain better understanding of the migration pattern and use of space of the wildlife. Advances in computer technology and global positioning systems (GPS) have lowered costs, reduced processing time, and improved accuracy for tracking wild animals. In this thesis, a wildlife tracking extension is developed for ArcGIS 9.x, which allows biologists and ecologists to effectively track, visualize and analyze the movement patterns of wild animals. The extension has four major components: (1) data import; (2) tracking; (3) spatial and temporal analysis; and (4) data export. Compared with existing software tools for wildlife tracking, the major features of the extension include: (1) wildlife tracking capabilities using a dynamic data layer supported by a file geodatabase with 1 TB storage limit; (2) spatial clustering of wildlife locations; (3) lacunarity analysis of one-dimensional individual animal trajectories and two-dimensional animal locations for better understanding of animal movement patterns; and (4) herds evolution modeling and graphic representation. The application of the extension is demonstrated using simulated data, test data collected by a GPS collar, and a real dataset collected by ARGOS satellite telemetry for albatrosses in the Pacific Ocean.

Copyright 2009

by

Cai Chen

ACKNOWLEDGEMENTS

I would like to thank Dr. Dave Anderson for his permission to use the albatross data downloaded from <http://seamap.env.duke.edu/datasets/detail/313> (Dataset credit: National Science Foundation).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
CHAPTER 1 - INTRODUCTION.....	1
1.1 Why Tracking Wildlife.....	1
1.2 Methods for Wildlife Tracking.....	1
1.3 Review of Case Studies	5
1.4 Review of Data Analysis Methods	6
1.5 Review of Software Tools and Systems	7
1.6 Research Objectives.....	8
1.7 Thesis Structure	9
CHAPTER 2 - FRAMEWORK DESIGNS AND IMPLEMENTATIONS	10
2.1 Utilities.....	11
2.1.1 Data Import	11
2.1.2 Points to Polyline	15
2.1.3 Random Points Generation	16
2.2 Wildlife Tracking.....	19
2.2.1 Wildlife Tracking Overview	19
2.2.2 Dynamic Layer.....	20
2.3 Spatial and Temporal Analysis.....	21
2.3.1 Home Range Analysis.....	21
2.3.2 Spatial Clustering Analysis.....	24
2.3.4 Measuring Representatives of Clustering	27
2.3.5 Lacunarity Analysis	28
2.3.6 Modeling Dynamic Membership in Herd	29
2.4 Data Export.....	32
2.4.1 Save as KML Files.....	32
2.4.2 Save as ESRI Shapefiles	33
CHAPTER 3 - SIMULATIONS AND CASE STUDIES.....	34
3.1 The Overview of SEAMAP dataset.....	34
3.2 Migration Paths of Albatrosses.....	36

3.3 Home Range of Albatrosses	37
3.4 Morishita Index of Albatrosses	38
3.5 Mean Crowding of Albatrosses	39
3.6 Lacunarity analysis in Albatrosses	39
3.7 Spatial Clustering---DBSCAN	40
3.8 Simulating Herds and Their Interactions	42
CHAPTER 4 - DISCUSSION AND CONCLUSION	44
REFERENCES	47

LIST OF FIGURES

Figure 1-1. ARGOS system (adapted from BBC).....	4
Figure 2-1. Overview of wildlife tracking extension.....	10
Figure 2-2. Working IDE of tracking extension.....	11
Figure 2-3. Screenshot of wildlife tracking extension.....	11
Figure 2-4. ARGOS file (synthetic data).....	12
Figure 2-5. Quantum GPS collar (telemetrysolutions.com).....	13
Figure 2-6. Plot of quantum GPS data in ArcGIS.....	13
Figure 2-7. Quantum GPS visualization in Google Maps.....	13
Figure 2-8 Overview of file geodatabase design.....	15
Figure 2-9. ARGOS points to polyline.....	15
Figure 2-10. Random points with polygon constraint.....	17
Figure 2-11. Random points generation on a polyline.....	18
Figure 2-12. Random points generation in multiple scales.....	19
Figure 2-13. Overview of wildlife tracking.....	20
Figure 2-14. Architecture of tracking extension.....	21
Figure 2-15. Home range for two different wild animals.....	22
Figure 2-16 Herd interaction graph (synthetic data).....	32
Figure 2-17. Displaying KML data on Google maps.....	33
Figure 3-1. Overview of OBIS-SEAMAP dataset.....	34
Figure 3-2. Sampled points for <i>P. nigripes</i> and <i>P. immutabilis</i>	35
Figure 3-3. Migration paths for <i>P. nigripes</i> and <i>P. immutabilis</i>	36
Figure 3-4. Individual migration paths.....	36
Figure 3-5. Home range of <i>P. immutabilis</i> and <i>P. nigripes</i>	37
Figure 3-6 Morishita index of <i>P. immutabilis</i> species.....	38
Figure 3-7 Mean crowding of <i>P. immutabilis</i> species.....	39
Figure 3-8:Lacunarities for <i>P. nigripes</i> and <i>immutabilis</i>	40
Figure 3-9: DBSCAN clustering result (colors represent different clusters).....	41
Figure 3-10: Generation of the herd interaction graph.....	42

CHAPTER 1 INTRODUCTION

1.1 Why Tracking Wildlife

Migration of wild animals is a process of movements from one place to another. Wildlife tracking is a practice that allows biologists, ecologists, naturalists, and conservation agencies to remotely observe relatively fine-scale movements and discover migratory patterns in free-ranging wild animals using modern-powered technologies, such as electrical radio, remote satellite and Geographic Information Science (GIS) technologies. Wildlife tracking has several significant *purposes and motivations* in the varied applications of wildlife research and monitoring, ecotourism, anti-poaching, and hunting. In order to better understand the interactions of wildlife world, there is a need to develop wildlife tracking and management systems, so that users are able to keep wildlife populations, including endangered animals, at desirable levels determined by human intentions.

Wildlife tracking has given scientists a complete, accurate picture and insight into migration patterns. For example, when scientists used radio transmitters to track one herd of caribou, they might be able to accomplish two crucial things. First, they can learn whether the herd moves as they previously assumed, although those assumptions are simply based upon human's observations. Second, they can prove if the herds will return to about the same location each year to give birth to their next generations, or just to start over the entire migration processes. This understanding and interpretation of their activities might be difficult to obtain purely through visual observations or by human stalking.

1.2 Methods for Wildlife Tracking

Back to 19th century, American naturalist John James Audubon (Burroughs, 1902) began to wonder whether migrating birds would return to the same place together each year after they moved to the south. So he tied a string as a marker around the leg of a bird before it flew south. The following spring, Audubon (Burroughs, 1902) surprisingly discovered that the bird had indeed come back. Although scientists today are still using tags, such as using metal bands, to track movement of animals, metal bands are not always useful tags. That is because the tagged animals have to be caught or seen again by the scientists to get any data from them. Unfortunately, most tagged animals are hardly found again.

To address the above problems, recent technologies have come into play. For example, electronic tags give repeating signals constantly. Typically, in order to ensure the correctness of collected data from wild animals, those signals are chosen by radio devices or satellites. Different frequencies of signal correspond to different devices or satellites. Therefore, scientists now are able to track the locations and movement of the tagged animals without recapturing them or stalking them, both of which require enormous efforts. In addition, these electronic tags can provide a great deal of data. However, there are still several obvious limitations to wildlife tracking. One of them is that electronic tags are more expensive than those low-tech tags that are NOT electronic. Also, because of their weight, electronic tags may harm some animals by slowing them down, or affecting their daily activities. Thus, to overcome those shortcomings, more advanced devices and high-tech approaches are coming into play in wildlife tracking tasks.

In terms of radio tracking, tracking an animal by radio usually required two different devices: one transmitter and one receiver, which are set up in different places. A transmitter attached to the animals sends out a signal in the form of radio waves with different frequencies, just as a common radio station does. This allows a scientist to place the transmitter around an

animal's ankle, neck, wing or any part of its body if applicable, which usually depends on what animal the scientist is studying. A receiver then picks up the signal sent by transmitter attached to wild animals, just like radio machines at home pick up signal from a station nearest to them. In practice, the receiver is usually in a truck or an airplane. In order to keep track of the signal, the scientists have to follow the animal in the trucks or planes due to the range of signal. However, for some endangered species or other scarce wildlife, it can be difficult to track via automobile and planes.

More recently, receivers have been placed in earth-orbiting satellites as well as in airplanes or trucks, which are usually referred as satellite tracking in academic and industrial researches. Usually, groups of satellites are working together to track wild animals. Each satellite in a network picks up electronic signals from a transmitter on an animal. Together, the signals from all satellites are able to provide the precise location of the animals. In addition, the satellites also can track the animal's path as it moves. Since the satellites are consistently moving on the orbits, satellite tracking is especially useful because the scientists do not have to follow after the animals. In nowadays, the satellite networks have successfully tracked the migration of caribou, sea turtles, whales, seals, elephants, bald eagles, and ospreys (Taiga, 1995). It is now the core means to provide global data about wildlife-related data.

Note that radio tracking and satellite tracking sometimes are used together to get more comprehensive tracking accuracies in different scales. One of these examples is Porcupine Caribou Herd Satellite Collar project (Taiga, 1995). It is a cooperative project between a number of wildlife agencies and boards. They used satellite radio collars to document seasonal range use and migration patterns of the Porcupine caribou herd in northern Yukon, Alaska and NWT. Another example is the Wildlife Enforcement Monitoring System (Chandran, 2004), or WEMS,

a GIS-based model designed to monitor wildlife specifically in Asia. It was started as a joint initiative by the United Nations University and the Asian Conservation Alliance.

ARGOS is a unique worldwide location and data collection system dedicated to the study of the environment (ARGOS, 1978). It was first established in 1978 through collaboration between the French space agency (CNES), the US National Aeronautics and Space Administration (NASA) and the US National Oceanic and Atmospheric Administration (NOAA). It helps the scientific community to better monitor and understand our environment, and to comply with environmental protection regulations issued by various governments. As shown in Figure 1-1, ARGOS system consists of five components in general: (1) transmitters on animals or objects send their pulses of data; (2) corresponding satellite collects data and measures signal frequencies; (3) satellites relay data to terrestrial receiving stations after they receive data from transmitters; (4) data processing centers begin to interpret data and determine animal positions; and (5) researchers are able to view and analyze data and provide richer information and deeper understanding of the animals.

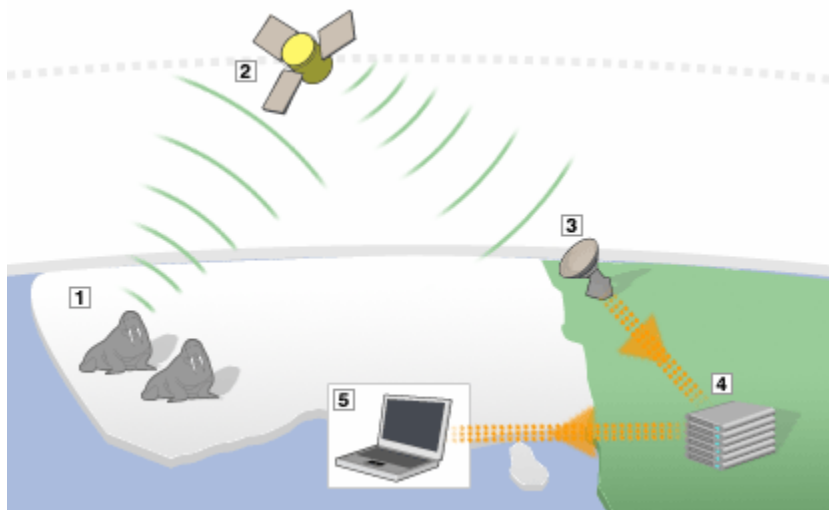


Figure 1-1. ARGOS system (adapted from BBC).

1.3 Review of Case Studies

Historically, tracking animals in the field is difficult and labor intensive because of factors related to terrain and season. For decades, scientists from different domains have been tracking wildlife and the ways they migrate and interact using various techniques. The development of very high-frequency (VHF) radio telemetry in the 1950s greatly improved the research of wildlife migration (LeMunyan et al., 1959; Cochran and Lord, 1963; Ozoga and Verme, 1975). Numerous case studies of wildlife tracking have been reported since 1980s, mainly because of two reasons: (1) advances in technology have improved the accuracy and reduced the cost associated with wildlife tracking; and (2) the growing environmental awareness among the public and policymakers has led to increased research activities on the impacts of human activities on wildlife and their habitats. Examples of radio telemetry applications include animal behavior study (Ozoga et al., 1982), activity (Ozoga and Verme 1975, Vercauteren and Hygnstrom 1998), habitat use (Mooty et al., 1987), and survival (Etter et al., 2002). Gilsdorf et al. (2008) designed and developed a vehicle-mounted VHF-based telemetry system that integrated an on-board antenna, receiver, electronic compass, GPS, and GIS for tracking highly mobile species. The ARGOS system has been used to study movements of caribow (Craighead, 1986, 1987), camels (Grigg, 1987), peregrine falcon (Howey, 1994), emperor penguin (Le Maho and Gendner, 1994), albatrosses (Fernandez et al., 2001), sharks (Gifford et al., 2007), and whale sharks (Hsu et al., 2007). Compared with VHF-based systems and the ARGOS system, the applications of GPS collars are relatively new, including studies on the activity of manged wolves (Young and Coelho, 2006), wolves movements (Meier et al., 2006), wolf predation on deer fawns (Demma et al., 2007), distribution and interaction of white-tailed deer and cattle

(Cooper et al., 2008), and white-tailed deer distribution in response to patch burning on rangeland (Meek et al., 2008).

1.4 Review of Data Analysis Methods

Methods for wildlife data analysis and models of animal movement have experienced many changes in the past decades. As one of the important concepts for wildlife tracking, home range has been studied by many researchers (Burnett, 1943; Jennrich and Turner, 1969; Dixon and Chapman, 1980; Anderson, 1982; Reynolds and Laundré, 1990; Aebischer et al., 1993; Person et al., 2007). A review of the models of home range can be found in Worton (1987). An evaluation of the accuracy of kernel density estimators for home range analysis can be found in Seaman and Powell (1996). In addition to home range analysis, examples of other studies include space use patterns (Ford and Krumme, 1979), activity areas (Dixon and Chapman, 1980), grouping behaviors as they relate to variation in food quality and quantity (Winnie et al., 2008), and independence of observations (Swihart and Slade, 1985). Hebblewhite et al. (2006) provided a statistical approach to correct GPS habitat bias for habitat selection studies using GPS collars. Hins et al. (2008) revealed the evidence of a landscape configuration effect of habitat selection by forest-dwelling caribou in managed boreal forest of eastern Canada. Schwager et al. (2007) discussed robust classification of animal tracking data; Madin et al. (2007) proposed a formal ontology for describing and synthesizing ecological observation data. More discussions on wildlife tracking and data analysis can be found in White and Garrott (1990), and Kays and Wikelski (2007). An important progress in wildlife tracking research in recent years is the development of animal movement models by multidisciplinary researchers in biology, ecology, computer science, mathematics, and statistics. For example, Bian (2000) proposed methods of

component modeling for the spatial representation of wildlife movements; Brillinger et al. (2004) presented an exploratory data analysis of the trajectories of moving animals; Benkert et al. (2008) developed new models for reporting flock patterns; Huang et al. (2008) developed new models for analyzing herds evolution patterns; Patterson et al. (2008) reviewed space-state models (SSM) of individual animal movement; Calenge et al. (2009) discussed computational aspects of trajectories analysis and proposed an object class for trajectories.

In recent years, there has been increased interest in analyzing spatial-temporal patterns and moving paths of wild animals (Ankerst et al., 1999, Cao et al., 2005, Benkert 2006, Huang et al., 2008). Geographic data mining approaches have been proposed to detect those generic spatial-temporal patterns such as flock, leadership, convergence, and encounter in geospatial data (Gudmundsson and Kreveld, 2006; Laube et al., 2002, 2005). Such research attempts to study a group of wild animals together to gain better interpretation of interactions within groups of study animal or outside the group of study animal. As an example, flock pattern is a large subset of moving animals that are constantly moving along paths that are spatially close to each other in a certain pre-defined time. Thus, to discover flock pattern from a large dataset depicting wildlife movement, one needs to consider two sources of information: time duration and space extent.

1.5 Review of Software Tools and Systems

A number of software tools and systems have been developed for wild animal data analysis and visualization in the last decade. Although there were some earlier computer programs for wildlife movement analysis, those were single-function programs that lacked a full suite of analysis options and the ability to handle a wide variety of data (Hooge et al., 1999). With the ability to manage and process spatial and non-spatial data from multiple sources, GIS

becomes an ideal platform for wildlife tracking. Hooge et al. (1999) developed an animal movement program based on ESRI ArcView 3.x, which provided several functions for animal movement analysis, such as home range analysis, location statistics, and some other generic tools for data import and analysis. Wong et al. (2007) developed a WILDSPACE decision support system for species data and knowledge management. Zhang et al. (2007) developed a GBD-Explorer using open source Java GIS for exploring ecoregion-based biodiversity data.

Advances in Internet mapping capabilities have helped development of Web services for serving wildlife data. Best et al. (2007) developed geospatial Web services for serving marine animal biogeographic data from OBIS-SEAMAP (Halpin et al., 2006) and predicting marine mammal habitats in a dynamic environment. They created custom workflows using ESRI Model Builder, ArcGIS geoprocessor, R statistical package, Python scripting language, PostGIS geodatabase, and UMN MapServer. Cagnacci and Urbano (2008) developed ISAMUD, an integrated and modular software platform to manage GPS collar data for wildlife management. ISAMUD is based on an open source spatial database (PostgreSQL and PostGIS) and has functionality for open source data management, geo-statistical analysis, Web services modules (R, QGIS, GRASS, MapServer, Ka-Map) and a front-end database (MS Access).

1.6 Research Objectives

The objective of this research project was to develop a set of GIS tools as an extension of ESRI's ArcGIS 9.x to help biologists, ecologists, and other practitioners to better navigate, visualize, and analyze wildlife data collected by satellite telemetry and GPS collars. Specifically, the research focused on developing the following capabilities in a GIS environment: (1) wildlife tracking capabilities using a dynamic data layer supported by a file geodatabase with 1 TB

storage limit; (2) spatial clustering of wildlife locations; (3) lacunarity analysis of one-dimensional individual animal trajectories and two-dimensional animal locations for better understanding of animal movement patterns; and (4) herds evolution modeling and graphic representation. The application of the extension was demonstrated using simulated data, test data collected by a GPS collar, and a real dataset collected by ARGOS satellite telemetry for albatrosses in the Pacific Ocean (Fernandez et al., 2001).

1.7 Thesis Structure

This thesis contains four chapters. Chapter 1 includes introduction and literature review; Chapter 2 focuses on framework design and implementation; Chapter 3 presents simulations and case studies using simulated data, GPS collar data, and ARGOS data for albatrosses in the Pacific Ocean obtained from OBIS-SEAMAP; Chapter 4 concludes this thesis and discusses future directions of wildlife tracking.

CHAPTER 2

FRAMEWORK DESIGNS AND IMPLEMENTATIONS

The framework of the extension is shown in Figure 2-1 and is described in the following sections.

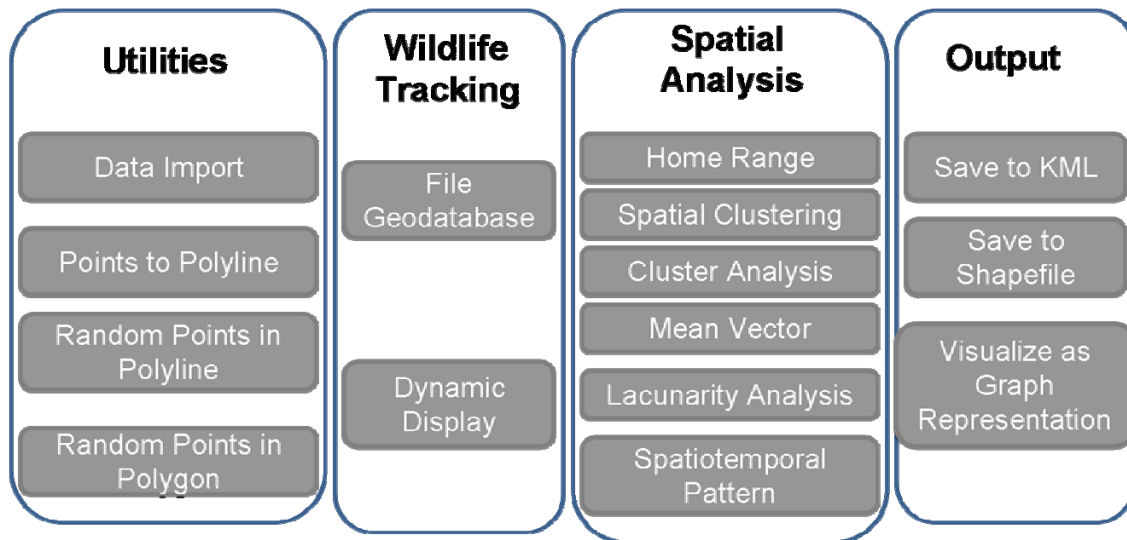


Figure 2-1. Overview of wildlife tracking extension.

This chapter introduces the implementations of major components of this tracking extension. The entire tracking extension is developed using ArcObjects (ESRI, 2007b) and VB.net. The integrated development environment (IDE) of this extension is shown in Figure 2-2. To further extend the ArcGIS functionalities in .NET, ESRI's ArcObjects allows developers to build customized GIS solutions. The screenshot of the wildlife tracking extension in ArcGIS is shown in Figure 2-3.

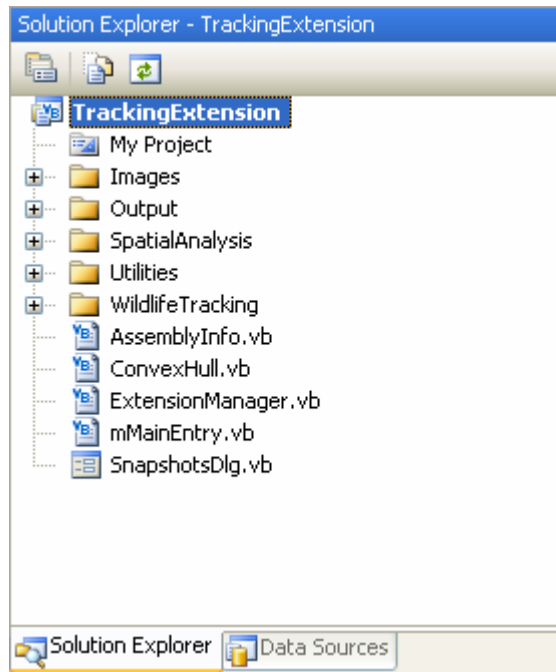


Figure 2-2. Working IDE of tracking extension.

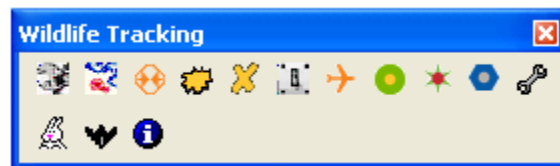


Figure 2-3. Screenshot of wildlife tracking extension.

2.1 Utilities

In this extension, a set of basic utilities for users to import, convert, simulate and analyze wildlife-related data have been implemented. Built with ArcObjects for ArcGIS, the utilities consist of Data Import, Points to Polyline Conversion, Location Data Simulation, and Random Points Generation both on polylines and in polygons.

2.1.1 Data Import

The data import utility provides a practical bridge between raw data and the GIS

environment. It supports data formats from ARGOS and GPS collars as input to GIS.

ARGOS is a satellite telemetry system that has been in use since 1978 (ARGOS, 1978). In the past decades, GPS tracking devices have been linked to the ARGOS Platform Transmitter Terminal (PTT) for data transmission. Users can download their data directly from ARGOS via telnet protocol and process the raw data to extract their transmitted information. The ARGOS data contains:

1. Program number
2. Platform number
3. Platform latitude in degrees and thousandths of a degree
4. Platform longitude in degrees and thousandths of a degree
5. Location class
6. Calendar day and UTC time (in hours and minutes) of most recent data collection
7. Calendar day and UTC time (in hours and minutes) of most recent location
8. Compression index: number of identical messages received during satellite pass

Figure 2-4 shows sample of the raw data gathered by ARGOS System.

```
Prog00003& 00111& -29.99953846 & 22.43700000000 3&061/1503Z&-061/1457&25&01
Prog00003& 00111& -29.94876923 & 22.82623076920 3&061/1503Z&-061/1457&25&02
Prog00003& 00111& -29.86415385 & 23.13084615380 3&061/1503Z&-061/1457&25&03
Prog00003& 00111& -29.47492308 & 24.26469230770 3&061/1503Z&-061/1457&25&04
Prog00003& 00111& -29.17030769 & 25.43238461540 3&061/1503Z&-061/1457&25&05
Prog00003& 00111& -28.91646154 & 25.90623076920 3&061/1503Z&-061/1457&25&06
Prog00003& 00111& -28.56107692 & 26.31238461540 3&061/1503Z&-061/1457&25&07
Prog00003& 00111& -28.358 & 27.02315384620 3&061/1503Z&-061/1457&25&08
Prog00003& 00111& -28.30723077 & 27.02315384620 3&061/1503Z&-061/1457&25&09
Prog00003& 00111& -28.22261538 & 28.90161538460 3&061/1503Z&-061/1457&25&10
Prog00003& 00111& -27.79953846 & 28.98623076920 3&061/1503Z&-061/1457&25&11
Prog00003& 00111& -27.59646154 & 28.64776923080 3&061/1503Z&-061/1457&25&12
Prog00003& 00111& -27.59646154 & 27.88623076920 3&061/1503Z&-061/1457&25&13
Prog00003& 00111& -27.46107692 & 26.93853846150 3&061/1503Z&-061/1457&25&14
```

Figure 2-4. ARGOS file (synthetic data).

The extension can also support various GPS collars, one of them is Quantum GPS collars shown in Figure 2-5 as an example.

1	Date	Time	TTF	Latt	Long	Altitude	Pdop	Hdop	Vdop	Satt	Fix	V1	V2	T1	
2	2006/01/01	15:27:38	38	38:01:05.06	-122:01:57.35			-4	4.2	3.3	0.0	6	3D	3.67V	28.50
3	2008/06/26	16:35:00	0	38:01:05.02	-122:01:57.67			-13	4.2	3.2	0.0	6	3D	3.67V	31.25
4	2008/06/26	16:40:00	0	38:01:04.93	-122:01:56.81			-4	4.0	3.0	0.0	6	3D	3.67V	28.50
5	2008/06/26	16:45:05	5	38:01:04.87	-122:01:55.79			29	3.8	2.8	0.0	6	3D	3.67V	27.62
6	2008/06/26	16:50:03	3	38:01:04.43	-122:01:55.59			12	4.0	3.2	0.0	5	3D	3.67V	27.56
7	2008/06/26	16:55:02	2	38:01:04.99	-122:01:56.38			-4	3.7	1.5	0.0	8	3D	3.67V	27.31
8	2008/06/26	17:00:01	1	38:01:05.90	-122:01:57.34			-37	3.8	1.5	0.0	8	3D	3.67V	26.68
9	2008/06/26	17:05:00	0	38:01:04.80	-122:01:56.20			9	3.8	1.5	0.0	7	3D	3.67V	26.18
10	2008/06/26	17:10:00	0	38:01:05.80	-122:01:57.30			-35	3.8	1.5	0.0	8	3D	3.67V	26.18
11	2008/06/26	17:15:02	2	38:01:06.19	-122:01:57.56			-59	4.0	1.5	0.0	8	3D	3.67V	26.43
12	2008/06/26	17:20:00	0	38:01:05.61	-122:01:57.13			-21	4.0	1.5	0.0	8	3D	3.67V	26.62
13	2008/06/26	17:25:00	0	38:01:06.57	-122:01:58.17			-72	4.2	1.6	0.0	7	3D	3.67V	26.68
14	2008/06/26	17:30:02	2	38:01:04.69	-122:01:56.60			7	4.2	1.6	0.0	6	3D	3.67V	26.75
15	2008/06/26	17:35:03	3	38:01:07.20	-122:01:59.21			-82	4.1	4.9	0.0	6	3D	3.67V	26.81
16	2008/06/26	17:40:03	3	38:01:06.51	-122:01:57.52			-40	4.1	4.7	0.0	6	3D	3.67V	26.62
17	2008/06/26	17:45:02	2	38:01:05.36	-122:01:56.84			-7	4.4	4.5	0.0	6	3D	3.67V	26.00
18	2008/06/26	17:50:02	2	38:01:05.14	-122:01:57.51			-19	4.1	4.2	0.0	6	3D	3.67V	25.50
19	2008/06/26	17:55:02	2	38:01:04.78	-122:01:55.82			33	4.1	4.0	0.0	6	3D	3.67V	25.37
20	2008/06/26	18:00:00	0	38:01:04.75	-122:01:56.68			-21	4.0	1.6	0.0	6	3D	3.67V	25.62
21	2008/06/27	05:01:10	70	38:01:04.79	-122:01:55.53			19	4.0	2.4	0.0	5	3D	3.67V	21.43
22	2008/06/27	05:05:04	4	38:01:03.95	-122:01:57.58			0	3.8	3.0	0.0	5	3D	3.69V	22.56
23	2008/06/27	05:10:05	5	38:01:04.85	-122:01:56.56			8	3.9	1.7	0.0	6	3D	3.67V	22.43
24	2008/06/27	05:15:06	6	38:01:04.51	-122:01:56.62			3	3.5	1.8	0.0	5	3D	3.69V	22.43
25	2008/06/27	05:20:04	4	38:01:04.97	-122:01:56.81			-8	3.7	1.7	0.0	6	3D	3.67V	22.50

Figure 2-5. Quantum GPS collar (telemetrysolutions.com).

With the aid of visualization platforms (e.g., ArcMap and Google Maps), the Quantum GPS dataset can be parsed and plotted for visualizations as shown in Figure 2-6 and Figure 2-7.

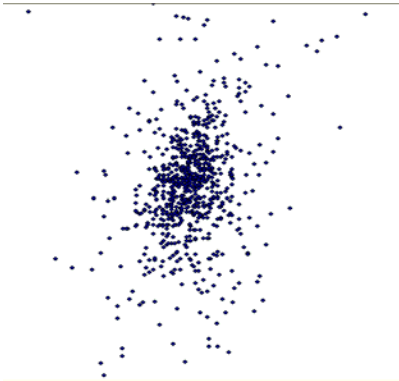


Figure 2-6. Plot of Quantum GPS data in ArcGIS.

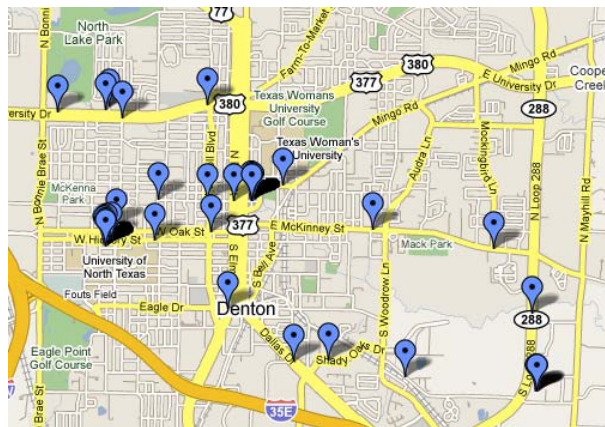


Figure 2-7. Quantum GPS visualization in Google Maps.

Regardless of the GPS data vendors, this extension allows users to parse the GPS raw data into a persistent file geodatabase. File geodatabase is stored as folders in a file system. Each dataset is held as a file that can scale up to 1 TB in size. This option is recommended over personal geodatabases with scalability benefits (ESRI, 2007). This characteristic makes tracking extension fairly extendable to support any GPS collar by simply developing GPS data parsers, although the data formats vary with GPS collars manufacturers. Table 2-1 shows the overview of the file geodatabase design in the wildlife tracking extension. This extension contains seven basic fields in its file geodatabase table.

Schema in File Geodatabase

Field	Description
<i>OBJECTID</i>	The unique indicator of samples points in study animals. This index starts from 1.
<i>EntityID</i>	The unique ID for each study animal. This index starts from 1.
<i>SHAPE</i>	The geometry type of spatial feature in this File Geodatabase table. This tracking extension uses point data type since the GPS sampled data are of discrete forms.
<i>X</i>	X coordinate value of the animal at a specific time.
<i>Y</i>	Y coordinate value of the animal at a specific time.
<i>Z</i>	Elevation of the animal at a specific time.
<i>T</i>	Data collection time for GPS collars or ARGOS satellite system.
<i>Tag</i>	An arbitrary text string for any additional information.

Table 2-1

Note that for future playback and data analysis, this extension does not overwrite (update) each record in the file geodatabase. Instead, it allows users to specify EntityID while importing GPS data files into file geodatabase.

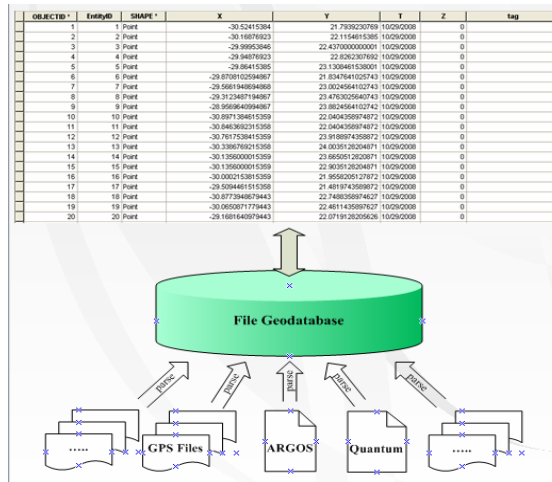


Figure 2-8 Overview of file geodatabase design.

2.1.2 Points to Polyline

The extension provides a way to pre-process data from discrete data points to linear features. GPS devices collect data at different time intervals, whereas wildlife continuously moves around the landscape between time intervals. Thus, the Points to Polyline utility provides capabilities for users to convert discrete data points into linear features. Such conversions are also necessary for analyzing the movement patterns of animals along paths (see examples in Chapter 3). Figure 2-9 is an example of a polyline converted from ARGOS points.

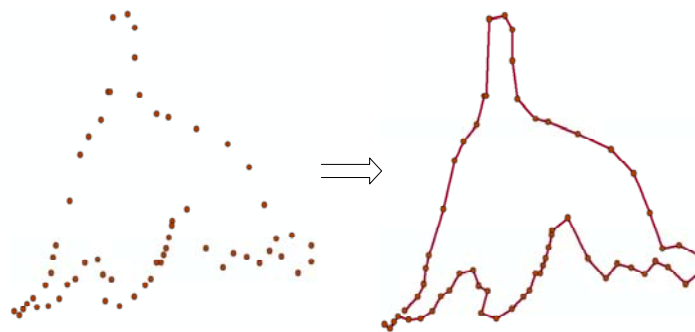


Figure 2-9. ARGOS points to polyline.

To transform discrete points to linear features (polyline), this extension collects points ordered by the time stamp. In its implementations, since IPointCollection interface has provided an access to members that manipulate the points of a Multipoint, Path, Ring, Polyline, Polygon, TriangleFan, TriangleStrip, or MultiPatch, the extension uses the AddPoint method of IPointCollection to feed data from the point layer. The following is the pseudo-code for Points to Polyline conversion:

1. Get an IPointCollection interface,
2. Initiate it is a Polyline Class instance
3. Get an IFeatureCursor interface and an IQueryFilter interface
4. Assign IQueryFilter's select and where clauses
5. Search point layer using IQueryFilter to get its featureclass
6. Do while loop to add points on IPointCollection
7. Create a new polyline feature on polyline layer
8. Assign new created polyline feature with IPointCollection

2.1.3 Random Points Generation

Species presence and absence surveys are commonly used in monitoring environments, examining wildlife populations, and modeling habitat behaviors. However, it is difficult to confirm that a species is absent from a location without enough information about individuals and the entire population (Mackenzie, 2005). Further, it is not possible to collect data in entire population of wildlife with long life spans. As an example, if the migrations of thousands of seasonal birds are to be studied, expensive equipment and intensive labor investment would be required. In practical approaches, representative subsets of the entire population can be selected for further investigation. The Random Points Generation utility provides a capability for users to generate random entities both in simulation studies and in real-world scenarios. For instance, if a

home range of a wild animal has been determined by tracking extension, those home ranges (polygons) can be further used as constraints to generate the random instances (points) for further study. Given certain mobility scenarios, it is possible to extrapolate new home ranges for those animals, which helps us to dynamically analyze the changes in home ranges over time.

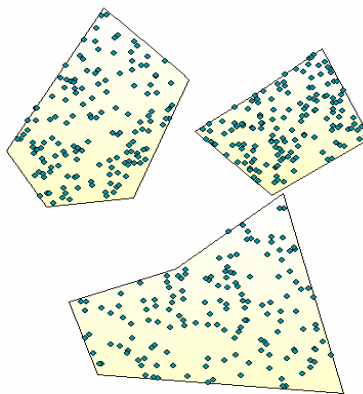


Figure 2-10. Random points with polygon constraint.

To generate random points with spatial constraints, this extension first obtains the boundary of the constraint and utilizes VB.net's built-in function `Rnd()` to generate random numbers in the range of 0 and 1. Further, the points located outside of the polygon constraints should be removed from generated points. Here the method `Contains` in `IRelationalOperator` interface is used to indicate if one geometry contains another geometry. The following is the pseudo-code code for the Random Points Generation utility (with polygon constraint. See Figure 2-10):

1. Get the envelope of polygon constraint (can be either converted from graphics or imported from polygon feature layer)
2. Generate random points with this envelope
3. Use `IRelationalOperator`'s `Contains` method to determine if the `pRandom Pt` is qualified
4. If yes, reserve it as one of generated random points; if no, remove it and repeat step 2.

Similarly, this extension is also able to generate random points on linear features (a.k.a polylines). Figure 2-11 demonstrates the process of generating random points on a polyline. This process involves the QueryPoint method of IPolyline interface to generate a point on the polyline at a specified distance from the beginning of the polyline. It then randomly specifies distances by using built-in VB function Rnd(). The pseudo-code is listed below.

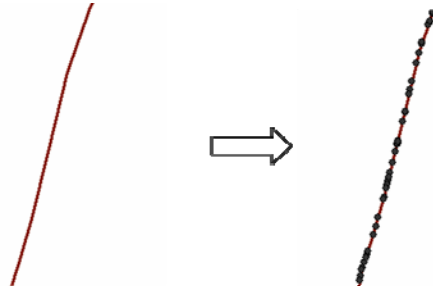


Figure 2-11. Random points generation on a polyline.

1. Get an interface of IPointCollection
2. Assign Polyline class to IPointCollection
3. Generate a point according to the specified distance along this polyline using QueryPoint method in Polyline feature
4. Create a new empty point feature in point layer
5. Assign the point generated in step 3 to this new empty point.

In addition to random point generation, this extension can also generate multi-scale clustered points for simulation purpose. Figure 2-12 shows clustered points with two scales of clusters.

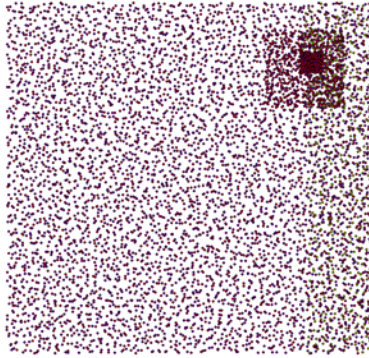


Figure 2-12. Random points generation in multiple scales.

2.2 Wildlife Tracking

In this section, the concept of tracking layer and its architecture in ArcGIS development environment is discussed.

2.2.1 Wildlife Tracking Overview

The satellite constellation continuously transmits signal to receiving stations and GPS collars on wildlife animals. This tracking extension parses satellite telemetry data and stores them in a File Geodatabase. This tracking extension persistently updates data about the queried wildlife in real-time, which allows for dynamic visualization of wildlife movement patterns.

Figure 2-13 demonstrates the process in wildlife tracking.

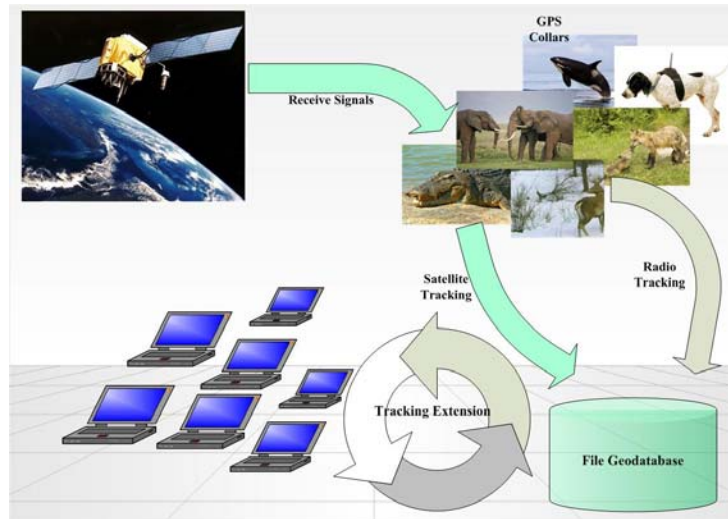


Figure 2-13. Overview of wildlife tracking.

2.2.2 Dynamic Layer

To better display and visualize time-varying data, this tracking extension uses the ArcGIS Dynamic Display to implement a dynamic layer. As shown in Figure 2-14, a dynamic layer is a custom layer that, in addition to the standard interfaces, implements the `IDynamicLayer` interface. `IDynamicLayer` interface allows this layer to draw its moving items into the mode of Dynamic Display and further take advantage of graphics card hardware acceleration. This extension also allows the tracking layer to use simulated data for testing purposes. In case of a real dynamic layer, the dynamic data will be streamed to the layer by a communication port, an Extensible Markup Language (XML) stream, an ArcGIS tracking server, or any other live feed. The tracking extension utilizes a base class named `DynamicLayerBase`, which simplifies the creation of dynamic layers. This extension is fairly extensible. In order to use the `DynamicLayerBase` class, this extension must override the method `DrawDynamicLayer()`. The `DynamicLayerBase` offers default implementation for all other methods and properties. As can be seen in Figure 2-14 legend, the diamond symbol indicates `WildlifeDynamicLayerClass` is a type inherited from `DynamicLayerBase` who has two outbound interfaces denoted by the dot symbols. Further, read

only property and read & write property are denoted with single square arrow and double square arrow, respectively.

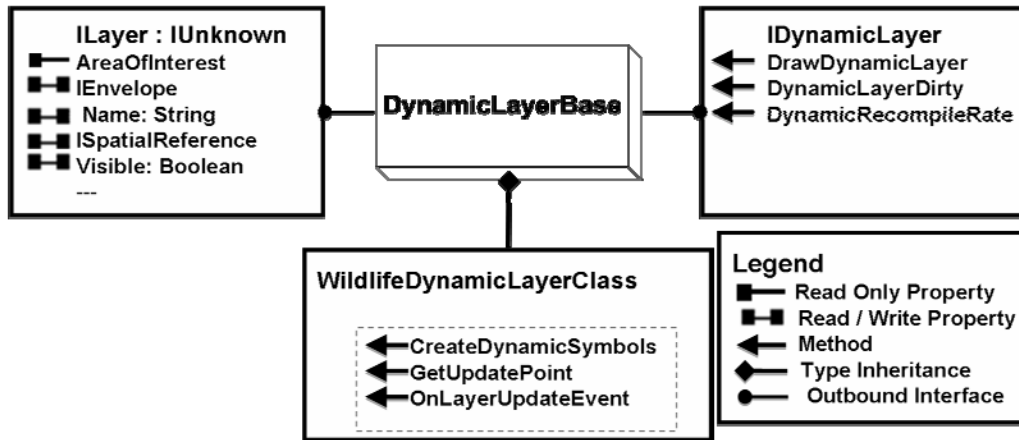


Figure 2-14. Architecture of tracking extension.

2.3 Spatial and Temporal Analysis

In this section, several spatial-temporal analysis implemented in tracking extension will be introduced and further explained.

2.3.1 Home Range Analysis

Home range is a concept which was first appeared in 1943 in the publication written by W. H. Burt (Burt, 1943). Burt delineated the spatial extent or outside boundary of an animal's movement during the course of its everyday activities by the means similar to the way we construct a map in modern GIS systems. There are various methods to analyze home range of an individual or a group of individuals.

2.3.1.1 Minimum Convex Hull

The home range of an individual wild animal is usually generated from a set of

successive points documenting its course of everyday activities, which have been collected over a period of time by satellite GPS devices. As mentioned above, such data are now collected automatically using GPS collars placed on individuals that transmit through satellites. In practice, the simplest way to analyze the home ranges of wild animals is to construct the smallest possible convex polygon around the incident points. This approach is usually referred as the minimum convex polygon (MCP) method, which is still widely employed in many tracking systems and tools, although it has many drawbacks and limitations (Baker, 2001, Creel, 2002). One of them is that it often over estimates the size of home ranges in terms of real extent of the study animal's activities.

This extension supports the implementations of minimum convex polygon (MCP) method and local convex hull (LoCoh) method. Note that calculating the home range for a group of wildlife animals at different times is mathematically equivalent to compute the convex hull for a set of points X in a real vector space V . Figure 2-15 shows two home ranges for two different wild animals at the same observation time.

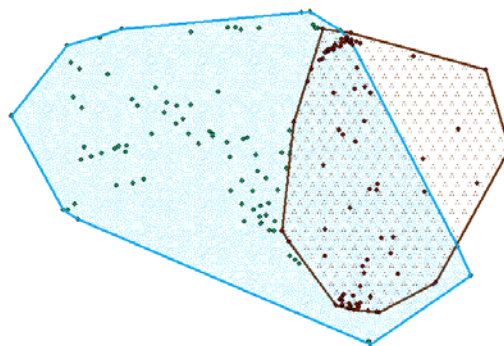


Figure 2-15. Home range for two different wild animals.

In this extension, the ConvexHull method of the ITopologicalOperator interface is used to calculate home range of an animal.

1. Get an ITopologicalOperator interface

2. Set it as the instance of Polygon class
3. Calculate home range by using ITopologicalOperator's ConvexHull()
4. Covert home range(polygon feature) to graphics for display

2.3.1.2 Kernel Approach

Associated with the concept of home range, there is another concept called utilization distribution. Person et al. (2007) have present 15 years of baseline data on the distribution and movements of 72 satellite-collared and 10 GPS-collared caribou from the Teshekpuk caribou herd that have had little to no exposure to oil and gas activities. Utilization distribution is a probability distribution constructed from a set of spatial data regarding the location of a moving animal around the studied area. Without taking time into account, the utilization distribution is a two dimensional probability density function, which indicates the probability of finding an animal in a pre-defined area within its home range. To unveil the real extent of studied wild animals, one of oldest known methods to construct utilization distributions is so-called normal distribution kernel method (Silverman, 1986). The first kernel methods for non-parametric estimation of the utilization distribution from a sample of point-based observations made on an animal in its home range are described by Worton (1989).

Recently, the kernel approach has been further investigated in academic and industrial researches. To construct the utilization distributions, it was extended to include a number of nonparametric methods such as alpha-hull method (Burgman, 2003) and local convex hull, or LoCoh method (Getz and Wilmers, 2004). More recently, the LoCoh method has also been extended from a purely fixed-point LoCoH method to fixed radius and adaptive radius LoCoH methods. In terms of implementation and practicality, more software and tools implement

parametric approaches than nonparametric methods, and it is widely accepted that LoCoH methods generally provide more accurate estimations of home range sizes and have better convergence properties (Getz and Wilmers, 2004).

2.3.2 Spatial Clustering Analysis

Clustering is a process to classify the objects into different disjointed groups based upon certain grouping criteria, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset share some cohesive traits. These traits are usually referred as proximity of studied objects according to some defined distance measure, including Euclidean distance, Manhattan distance, Mahalanobis distance, and Edit distance. The spatial clustering methods can be further classified into four categories: partitioning method, hierarchical method, density based method, and spectral clustering.

2.3.2.1 Partitioning *Clustering*

Partitioning methods such as k-means, k-centroids and Expectation-maximization (EM) clustering are methods that make use of a technique called iterative reallocation to improve the clustering quality from an initial solution. These methods tend to find clusters that are of spherical shape and they are made for minimizing the distance from the data objects to their distance centers.

2.3.2.2 Hierarchical Clustering

Hierarchical algorithms find successive clusters from previously established clusters. Hierarchical algorithms can be agglomerative, either bottom-up or top-down (divisive). In the initial setting, agglomerative algorithms begin with each element as a separate cluster and merge

them into successively larger clusters in the following steps. Divisive algorithms begin with the whole set (consider it as a biggest cluster) and start to divide it into successively smaller clusters. The examples of hierarchical algorithm includes: BIRCH, CURE and CHAMELEON. All of them use complex criteria for compressing and relocating data before merging clusters.

2.3.2.3 Spatial Density-based Clustering

A third classification of clustering algorithms is based on density of data points within a studied region. To discover clusters in a large spatial dataset, one of the best known algorithms, Density-Based Spatial Clustering of Applications with Noise, or DBSCAN (Ester et al., 1996) is proposed. DBSCAN is a density-based clustering algorithm because it finds a number of arbitrary clusters starting from the estimated density distribution of corresponding nodes. Since DBSCAN is one of the most common clustering algorithms and is also the most cited in scientific literature, this thesis has implemented it with R-tree index. Therefore, the implementation of DBSCAN has linear complexity of computing. It has computational cost of $O(n \log n)$.

Basically, DBSCAN requires two parameters: epsilon (eps) and minimum points (minPts). It starts with an arbitrary starting point that has not been visited. It then finds all the neighbor points within distance eps of the starting point: If the number of neighbors is greater than or equal to minPts, a cluster is formed. The starting point and its neighbors are added to this cluster and the starting point is marked as visited. The algorithm then recursively repeats the evaluation process for all the neighbors. If the number of neighbors is less than minPts, the point is marked as noise. If all points in a formed cluster are marked as visited points, the algorithm proceeds to iterate through the remaining unvisited points in the dataset.

The Pseudocode of DBSCAN is also presented. Initially, each point is assigned as cluster 0 and unvisited; then DBSCAN searches the neighboring points by criteria, which are commonly set by users. Finally, DBSCAN determines whether an ongoing point is a noise or belongs to a cluster. Note that DBSCAN is an iterative process that repeats until all of the points in dataset are assigned either as clusters or noise.

1. Initiate cluster index, $C = 0$
2. For each unvisited point P in dataset D
3. Get its neighbors within epsilon, $N = \text{getNeighbors}(P, \text{epsilon})$
4. If numbers of neighbor is smaller than minPts
5. Mark P as a noise
6. Otherwise, increase cluster index C
7. Mark P as visited
8. Add P to cluster C
9. Recurse (N)

2.3.2.4 Spectral Clustering

Given a set of data points A , the similarity matrix may be defined as a matrix S where S_{ij} represents a measure of the similarity between two data points. Spectral clustering techniques then make use of the spectrum of the similarity matrix of the data to perform dimensionality reduction for clustering in fewer dimensions. Typically, spectral clustering techniques are graph-theoretical approaches.

Note that clustering techniques are critical tools for analyzing wildlife grouping behaviors.

A case study of spatial clustering in wildlife management system can be found in Fisher and Trulio (2007).

2.3.4 Measuring Representatives of Clustering

To determine the quality of clusters created by clustering algorithms, this thesis also discusses how to measure quality of clustering:

2.3.4.1 Morishita Index

Morishita Index (Morishita, 1959) is used to measure the dispersion and analysis of distribution patterns in spatial dataset, the index is defined as:

$$I = Q \frac{\sum_{i=0}^Q n_i (n_i - 1)}{N(N - 1)} \quad (1)$$

Where N is the total number of studied points in the sampling network, n_i is the number of samples found within the i^{th} cell (grid-based partitions), and Q is the total number of cells. By displaying the values of the index against the size of the cells, one can investigate the degree of contagion of the sampling network, a probability of two points from the network of falling within the same cell. In a regular distribution of studied population of wild animal, there are several characteristics regarding Morishita Index, which correspond to different circumstances in spatial dataset.

1. Morishita Index (I) increases as the size of the cells is increasing to a value of 1.
2. If the spatial distribution of the samples is random but homogeneous, Morishita Index (I) is independent of the size of cells but vary around a mean value of 1.
3. If clusters are formulated, Morishita Index (I) becomes greater than 1.

2.3.4.2 Mean Crowding

Lloyd (1967) proposed the use of mean crowding (denoted as m : the mean number for each individual of other individuals in the same quadrat) to describe the clustering of animals. Mean crowding can be calculated using the equation:

$$m = \frac{\sum_{i=1}^Q x_i (x_i - 1)}{N} \quad (2)$$

Where Q is the numbers of quadrat and N is the total populations of studied wildlife, x_i represents the number of individuals in quadrat i .

2.3.5 Lacunarity Analysis

Lacunarity was initially introduced as a means of further classifying fractals and textures that have the same fractal dimension and a very different visual appearance (Mandelbrot, 1982). Lacunarity is a metric to determine how the fractal fills in a space. If the fractal is dense, lacunarity will be small. In other words, the lacunarity increases with coarseness of studied objects. Note that lacunarity analysis is a multiscale method for describing patterns of spatial dispersion. Although it was initially developed for fractal objects, the method is more general and can be easily used to describe nonfractal and multifractal patterns (Plotnick et al., 1996 ; Dong, 2000). It can be used both in binary and quantitative data. Lacunarity analysis is broadly applicable to many data sets used in the natural sciences since it can be used in 1-dimensional networks, 2-dimensional cell space, and three-dimensional objects. In this extension, both one-dimension and two-dimension of lacunarity analysis are implemented.

Calculation of lacunarity for binary patterns is introduced by Plotnick et al. (1996). The studied area is first partitioned by cells. An $r * r$ box (e.g., $r = 2$) is then placed over the upper left

corner of the map initially. The number of occupied sites can be considered as the box mass. The box is now moved one column to the right and the box mass is again counted. This process is repeated over all rows and columns producing a frequency distribution of the box masses. The number of boxes of size r containing S occupied sites is designated by $n(S, r)$ and the total number of boxes of size r by $N(r)$. If the map is of size M , then:

$$N(r) = (M - r + 1)^2 \quad (3)$$

This frequency distribution is then converted into a probability distribution $Q(S, r)$ by dividing by the total number of boxes:

$$Q(S, r) = \frac{n(S, r)}{N(r)} \quad (4)$$

The first and second moments of this distribution are now determined:

$$Z^{(1)} = \sum S Q(S, r) \quad (5)$$

$$Z^{(2)} = \sum S^2 Q(S, r) \quad (6)$$

The lacunarity for this box size is now defined as:

$$\Lambda(r) = \frac{Z^{(2)}}{(Z^{(1)})^2} \quad (7)$$

2.3.6 Modeling Dynamic Membership in Herd

The detection and description of spatiotemporal patterns are essential for better understanding of the behaviors of moving objects (animals, vehicles, and people). For example, models of movements can be used to study the ecology of animal behaviors, habitat preferences, and the dynamics of population densities. To model the dynamics of wildlife, several

spatiotemporal patterns have been proposed in literatures. One of them is named Herd (Huang et. al., 2008), which has been implemented in this extension. It captures the dynamic membership of a group of wild animals that is changing over time. In general, those changes in membership can be further categorized as *expand*, *split*, *merge*, and *leave*.

2.3.6.1 Derivations of Herd Algorithm

The way of modeling herd's dynamics in time-varying data is similar to the way of evaluating the performance of information retrieval systems. Traditionally, the Precision (P), Recall (R) and F-Score (F) measurements have been used for evaluating how good an information retrieval system is (Huang et. al., 2008).

$$Recall(R) = \frac{|relevantDocuments| \cap |retrievedDocuments|}{|retrievedDocuments|} \quad (8)$$

$$Precision(P) = \frac{|relevantDocuments| \cap |retrievedDocuments|}{|relevantDocuments|} \quad (9)$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

If moving entities in trajectory data is considered as documents in retrieval system, the above measurements can be applied to the areas of spatio-temporal analysis. Thus, let $H(t)$ be a herd snapshot at time t , i.e. the members of herd H at time t , and let $H'(t+i)$ be another herd snapshot at time $t+i$, the precision (P), recall (R), and F-score (F) measurements are applied to model the relationship between $H(t)$ and $H'(t+i)$ as follows:

$$R(H(t), H'(t+i)) = \frac{|H(t) \cap H'(t+i)|}{|H(t)|} \quad (11)$$

Intuitively, $H(t)$ and $H'(t+i)$ measures the percentage of $H(t)$ that continue to exist

in $H'(t+i)$. Thus, the more the entities left between t and $t+i$ from $H(t)$, the lower the value of $R(H(t), H'(t+i))$ is.

$$P(H(t), H'(t+i)) = \frac{|H(t) \cap H'(t+i)|}{|H'(t+i)|} \quad (12)$$

Intuitively, $H(t), H'(t+i)$ measures the percentage of $H'(t+i)$ that come from $H(t)$.

Thus, the more the new entities joined in $H(t)$ between t and $t+i$, the lower the value of $P(H(t), H'(t+i))$ is:

$$F(H(t), H'(t+i)) = \frac{2 \times P(H(t), H'(t+i)) \times R(H(t), H'(t+i))}{P(H(t), H'(t+i)) + R(H(t), H'(t+i))} \quad (13)$$

$F(H(t), H'(t+i))$ represents the combined results of members left and new members joined and it ranges from 0 to 1, where 0 indicates $H(t)$ is completely different from $H'(t+i)$ and 1 indicates $H(t)$ and $H'(t+i)$ consist of exactly the same members.

2.3.6.2 Herd Interaction Graph

To better visualize the dynamics of moving animal, this extension also implements a graph-based representation, Herd Interaction Graph, which nodes represent the groups (clusters) of wild animal formed at each snapshot; edges represent their interactions between two successive snapshots; edge types represent their changes in membership, including merge, split, leave, join in general.

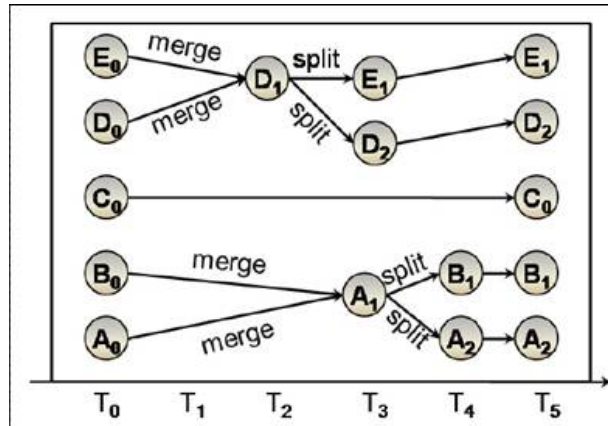


Figure 2-16 Herd interaction graph (synthetic data).

Figure 2-16 illustrates herd interaction graph by using synthetic dataset. The synthetic data is simulated by using our herd simulator and the expected Herd Interaction Graph is generated by our Herd algorithm. There are 6 snapshots (i.e., T_0 to T_5) in this simulated dataset. In the first snapshot T_0 , the DBSCAN clustering method has found 6 disjointed clusters (i.e., A_0 , B_0 , C_0 , D_0 , E_0), they keep the same (go through the quantitative changes) until T_2 since D_1 appears due to the interactions between D_0 and E_0 (E_0 merged into D_0). The subscript denotes the temporal status for each herd. In the same vein, B_0 merged into A_0 as A_1 in T_3 . A_1 and D_1 then split into two Herds. Since the resulting herds are quantitatively identical with Herds appeared on T_0 , Herding Graph labeled them as the same letters but with different subscripts, indicating they are formed at different snapshots.

2.4 Data Export

2.4.1 Save as KML Files

Originally, Keyhole Markup Language (KML) was developed for use with Google Maps and Google Earth, which is an XML-based language schema for expressing geographic annotation and visualization on web-based, two-dimensional maps and three-dimensional Earth browsers.

Later, the KML 2.2 specification was submitted to the Open Geospatial Consortium (OGC) to obtain its status as an open standard for all geo-browsers, making it well-supported by other applications, such as ArcGIS Explorer, Microsoft Virtual Earth. This extension provides users the capability to export results as KML files.

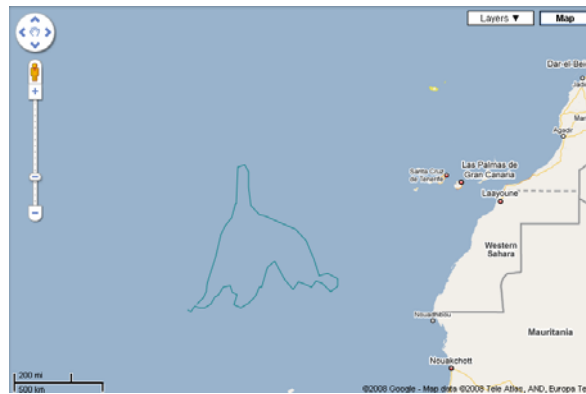


Figure 2-17. Displaying KML data on Google maps.

2.4.2 Save as ESRI Shapefiles

The extension also allows users to save results as shapefiles for further navigation in ArcGIS. In Home Range Analysis, for example, the extension allows users to save home range result as a polygon feature layer so that users can further navigate, compare different home ranges as shapefiles.

CHAPTER 3

SIMULATIONS AND CASE STUDIES

3.1 The Overview of SEAMAP Dataset

OBIS-SEAMAP (Halpin et al., 2006) is an online spatial database which aims to provide geospatial web services for public. It aggregates wildlife-related data, such as marine mammal, seabird and sea turtle from across the globe. The collection of spatial data can be searched and visualized through a set of advanced online mapping applications. The observation data held by OBIS-SEAMAP are collected from various data providers worldwide. They are registered into the SEAMAP database and presented on the web site upon data providers' permission. In this chapter, an OBIS-SEAMAP dataset is used (Figure 3-1, with Google online satellite image background).

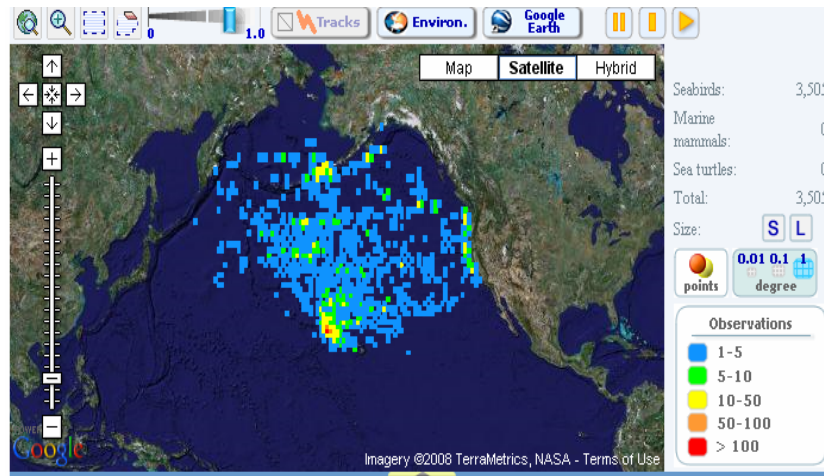


Figure 3-1. Overview of OBIS-SEAMAP dataset,

The SEAMAP dataset used in this thesis was supplied by Ocean Biogeographic Information System (OBIS), a group of investigators, led by Andrew Read and Patrick Halpin of Duke University. The dataset consists of two types of wildlife: *Phoebastria nigripes* and *Phoebastria immutabilis*, with 1488 observations points in *P. nigripes* species and 2017

observations points in *P. immutabilis* species from January 23, 1998 to August 3, 1998 (Fernandez et al., 2001). There are 15 species in *P. nigripes* and 14 species in *P. immutabilis*. Here, their migration paths over season are delineated by this extension. The overview of sampled points for *P. nigripes* and *immutabilis* is shown in Figure 3-1.

The original dataset is pre-processed as follows. The data was supposed to have at least one observation in each day. However, from the original dataset, there are several missing observations throughout the times of observation. Thus, the points are interoperated by calculating mean points between prior and next observations. Therefore, the straight dotted points may be newly generated due to the cause of approximated interpretations. Meanwhile, if more than two observations appeared in a single day, a sampled point is generated by calculating the centroids of those observations (here it is assumed that GPS collars have the same sampling rates). Further, to avoid the error caused from missing data, the data points with two or more consecutive missing days will not be used for analysis. Finally, their migration paths can be produced in the continuous form via the “point to polyline” utilities supported by this extension.



Figure 3-2. Sampled points for *P. nigripes* and *P. immutabilis*.

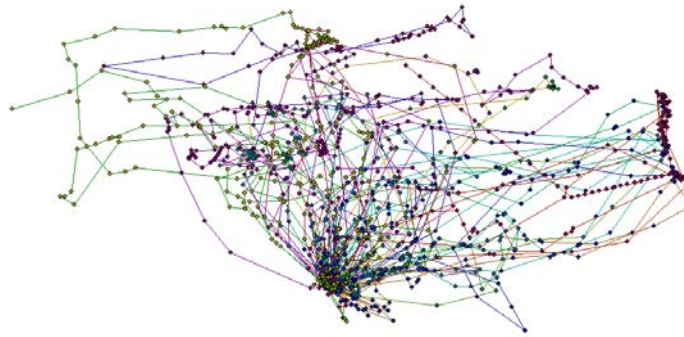


Figure 3-3. Migration paths for *P. nigripes* and *P. immutabilis*.

3.2 Migration Paths of Albatrosses

The extension allows users to generate migration paths for albatrosses by using Point to Polyline utility that serializes discrete point data into continuous paths. As shown in Figure 3-3, the overall migrations paths for the entire populations are displayed in ArcMap. More specifically, Figure 3-4 shows the migration paths for two *P. nigripes* species.

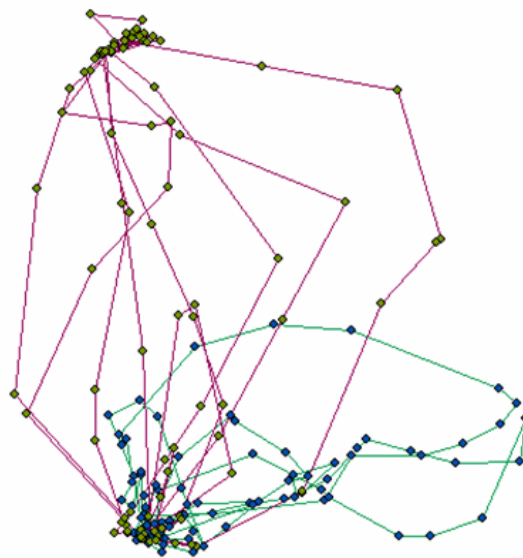


Figure 3-4. Individual migration paths.

3.3 Home Range of Albatrosses

As mentioned earlier, home range is the spatial extent or outside boundary of movement for an animal or a group of animals during the course of its everyday activities. Home Range Analysis enables us to fully understand habitat, preferences for studied animals. Further, two or more home ranges at different times are compared to determine the changes of their everyday activities. In this case study, each albatross trajectory is cut into two segments based on time: before and after February 15, 1998. The reason to use February 15, 1998 as the border point is that the species exhibit two different activities before and after this period of time (Whittow, 1993). Figure 3-5 shows the results in home range analysis for *P. nigripes* (denoted as region A and region B) and *P. immutabilis* (denoted as region C and region D) in two different periods of time.

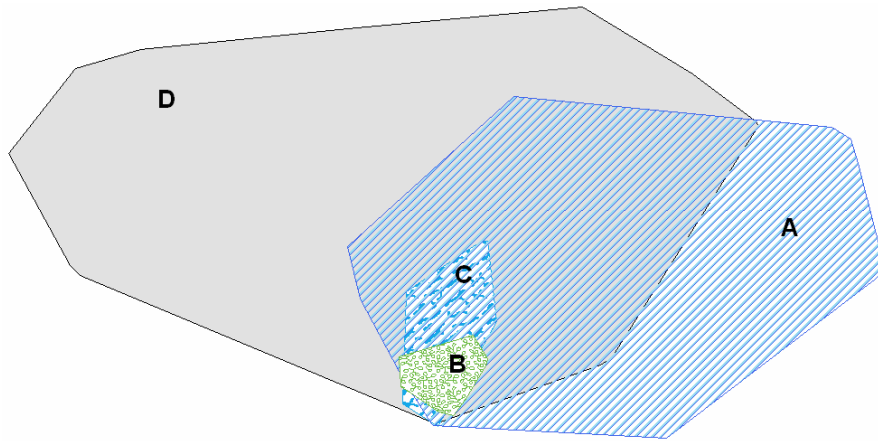


Figure 3-5. Home range of *P. immutabilis* and *P. nigripes*.

As can be seen in Figure 3-5, *P. immutabilis* have smaller uses of space (i.e., region C) during the October to next February, compared to the uses of space (i.e., region D) in post-February period of time because *P. immutabilis* breed mostly on the northwestern Hawaiian Archipelago and off Baja California (Fernandez et al., 2001). Therefore, the entire population of *P. immutabilis* slowed down in migration somehow due to breeding season. And they arrive at

their nesting sites by the end of October, and then incubate and brood chicks from November until February (Fernandez et al., 2001). After February, *P. immutabilis* began moving to North Pacific Ocean and spend the non-breeding season in the North Pacific Ocean and Bering Sea by expanding the use of space (expanding the home range). Figure 3-5 also reveals the same phenomenon happened to *P. nigripes*, rather they began moving to West Coast of North America after the breeding season.

3.4 Morishita Index of Albatrosses

To measure the dispersion and clustering distribution of albatrosses, the entire population of *P. immutabilis* species is selected and different numbers of cell for space partition are used.

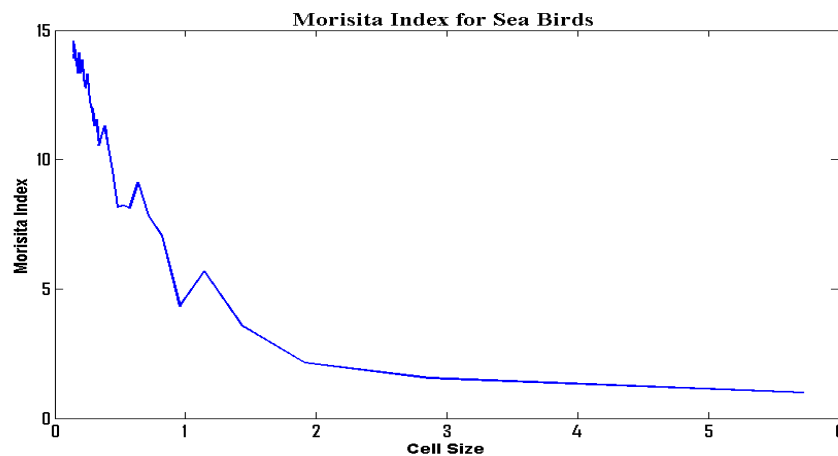


Figure 3-6 Morishita index of *P. immutabilis* species.

As shown in Figure 3-6, the cell size is in the 100 miles per unit in X-axis. Morishita index generally declines and approaches to 1 as the cell sizes are increasing at each computation. At the extreme case, if there is only 1 cell for the entire studied area, the Morishita index will equal to 1: no clustering, no separation. However, Morishita index values vibrate and vary within 200 miles since there are many clusters formed within these distances. This evidence is also supported by DBSCAN clustering, which generated 8 to 10 clusters within the radius of 100 to

200 miles. As the cell size is greater than 600 miles, morishita index is approaching to 1 since if DBSCAN clustering is carried out with radius greater than 600 mile, only 1 to 2 clusters are generated for the entire observations of albatrosses.

3.5 Mean Crowding of Albatrosses

To describe how albatrosses are clustering together during their migrations course, *P. immutabilis* species is used to calculate the mean number for each individual of other individuals in the same quadrat. As shown in Figure 3-7, mean crowding values are increasing as number of cells increase.

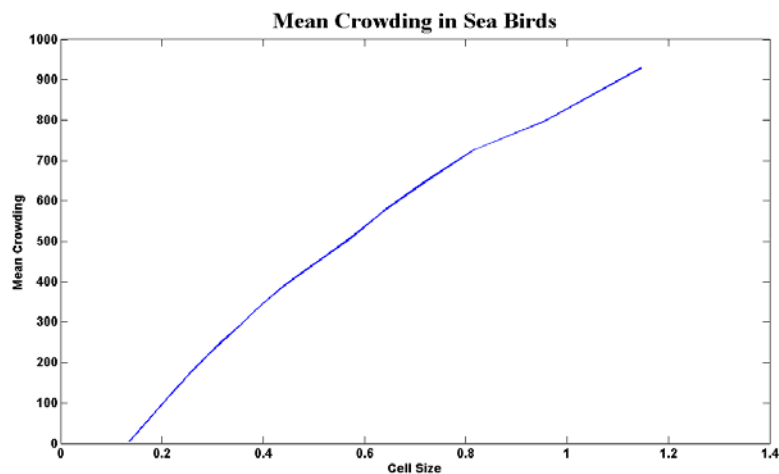


Figure 3-7 Mean crowding of *P. immutabilis* species.

3.6 Lacunarity Analysis in Albatrosses

Followed by home ranges analysis, 1D lacunarity for each entity is examined. In this thesis work, species 292, 312, 513, 522 and 552 are randomly chosen for case study after the data pre-processing. Three of them are *P. immutabilis* and the other three are *P. nigripes*. The purpose of this case study is to determine if two different types of species have different

migration pattern to one another. As shown in Figure 3-8, the lacunarity curves show that each lacunarity value is varying slight different with each other at different scales during their migrations. However, the testing species have similar lacunarity patterns (i.e., curves) with one another, no matter what types of species there are. This case study tells us that although their clustering results are different, they will exhibit and remain similar migration patterns (lacunarity values are closer to each other over scales). The result further suggests that albatrosses have intensive grouping behaviors throughout the entire observation. They tend to join and form the groups inherently while they are migrating to west of North America. The phenomenon also could be found in *P. immutabilis*, testing *P. immutabilis* species tend to form the inherent groups in the course of their migrations to North Pacific Ocean. Note that two dimension lacunarity is also supported in tracking extension. Users are allowed to select which lacunarity to be performed.

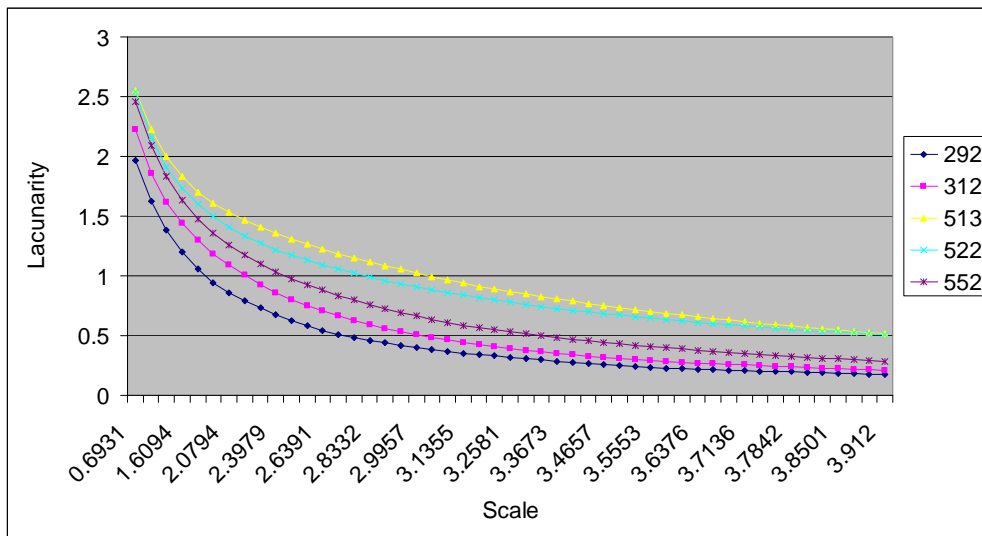


Figure 3-8: Lacunarity for *P. nigripes* and *immutabilis*.

3.7 Spatial Clustering---DBSCAN

The second case study in this thesis is random points generation and spatial clustering on

synthetic data. Random points by several arbitrary polygon constraints with different densities are first generated and clustered by the implemented clustering algorithm: DBSCAN. This case study has found that the DBSCAN algorithm has yielded good quality clusters and also allows noise to exist in testing dataset. There are several variants of DBSCAN addressing various domains. Readers are referred to previous chapter in this thesis to learn how DBSCAN works.

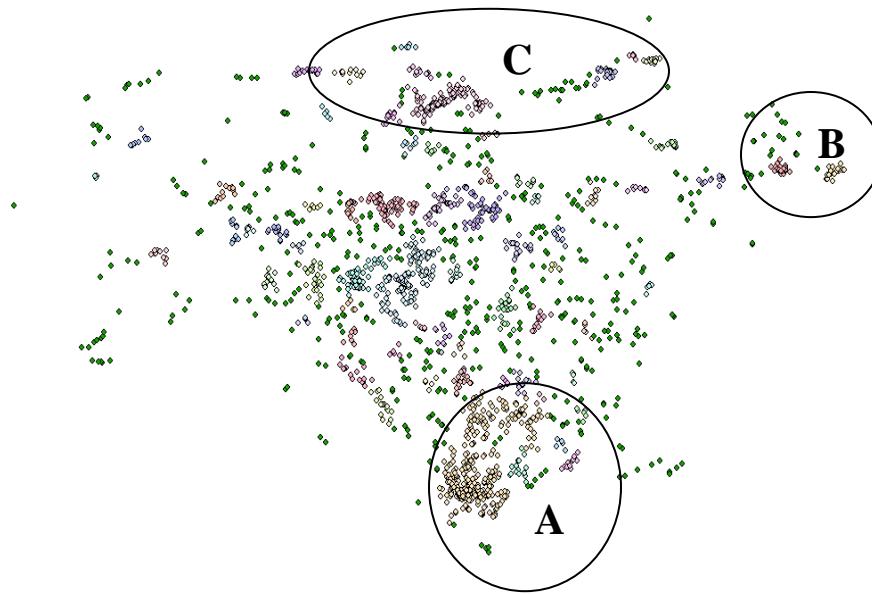


Figure 3-9: DBSCAN clustering result (colors represent different clusters).

Figure 3-9 has shown the result of DBSCAN for *P. immutabilis* species in this tracking extension. In this case study, the radius to search points is set to 146 miles and the minimum points to form cluster is 8 points. As a result, DBSCAN has found 8 clusters in *P. immutabilis* species that are represented by different colors in Figure 3-9. They form clusters differently throughout the entire observations. Region A, for example, is more likely to form clusters since *P. immutabilis* species are in breeding season and migrate together with slower speeds. Region B, on the other hand, was forming clusters mainly because they have arrived in west of North America around July to October. Another possibility to form clusters is in region C is resting and

mating on islands. As shown in region C, since there are many small islands between Bering Sea and Gulf of Alaska.

3.8 Simulating Herds and Their Interactions

Herd simulator allows its users to pre-define the moving pattern for each herd (a group of wild animals) via a computer mouse (Huang et. al., 2008). Herd simulator allows users to select one herd and then specify their spreading speeds and angles for all of the members in that herd. Therefore, an expected interaction between herds can be simulated by creating such paths. Herd simulator and herd interaction graph are evaluated and demonstrated in this chapter. Figure 3-10 shows how to use the herd simulator to model wild animals' dynamics in grouping behaviors and how to generate its interaction pattern by producing herd interaction graph in ArcMap.

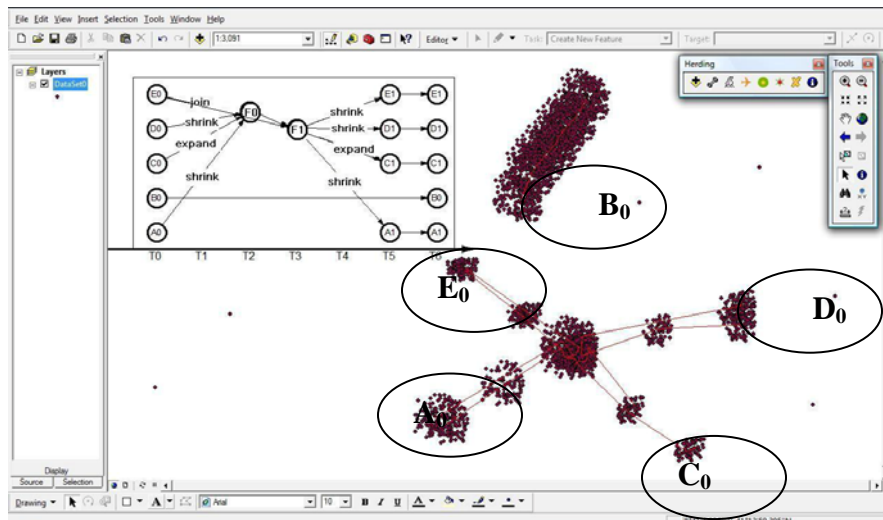


Figure 3-10: Generation of the herd interaction graph.

As can be seen in Figure 3-10, five herds are initiated at the first snapshot (i.e., A_0 , B_0 , C_0 , D_0 , E_0), an intended interaction is specified by herd simulator, which happened in third snapshot, forming a bigger herd (i.e., T_2). Further, shrinking happened in sixth snapshot, resulting smaller herds, which also appeared in the very first snapshot. Herd B_0 , however, does NOT involve any

interaction with other herds, resulting a direct link from the beginning to the end in herd interaction graph.

CHAPTER 4

DISCUSSION AND CONCLUSION

The wildlife tracking extension can be deployed on computers with Microsoft Windows operation system and ESRI's ArcGIS 9.x. The extension uses File Geodatabase as data storage for data from GPS collars and the ARGOS system. It allows users to parse any form of raw data into structured database. Although only data from ARGOS files and Quantum GPS collars are supported as input to the extension, the custom parsers for various GPS collars can be easily developed by extending the data import functionality. Compared with ESRI's Tracking Analyst for ArcGIS, this wildlife tracking extension intends to provide more advanced spatiotemporal analysis tools. Further, case studies conducted by this thesis have shown the basic utilities, such as point to polyline conversion and random point generation, are useful and necessary.

An interesting topic for better understanding wild animals is to determine the spatial distributions in various circumstances. DBSCAN clustering results suggest that clusters are more likely to form in the following circumstances (Whittow, 1993): 1) female animals entered the seasons of breeding. Albatrosses, for example, female birds were carrying their off-springs during the breeding season and began to slow their migrations down; 2) they have been arrived the destinations and are ready to migrate back. As suggested by *P. nigripes*, numbers of *P. nigripes* have been arrived at West Coast of North America around June to August. They were either waiting others to be arrived or settling themselves down for hunting and resting, which are more likely to form higher density of *P. nigripes*; 3) wild animals were mating or resting in islands in their ways of migrations. DBSCAN clustering result has generated several clusters nearby the islands between Bering Sea and Gulf of Alaska.

In this study, lacunarity results show that the migration behaviors of albatrosses are more uniform, rather than individual since the lacunarity values are closer to one another at multiple scales. It is expected that lacunarity curves will be different if the studied animals have very different migration patterns (more individual and independent). Although lacunarity in this case study is for 1D trajectories, this extension also supports 2D lacunarity analysis.

SEAMAP dataset, however, is not suitable for demonstrating how herds evolve and interact with each other due to the lack of interaction. To demonstrate the interaction between wild animals over time and determine how animals form clusters and evolve their membership over time, this extension allows users to generate a simulated data by specifying the moving path for each group of animals through computer mouse. Each click in moving path represents their snapshot. The expected interactions between different groups of wild animals will be presented in a herding interaction graph in the extension, showing how herds would go through quantitative and qualitative changes in membership (variation in quality and quantity) over time.

In summary, this thesis described the development of a wildlife tracking extension for ArcGIS. The extension enables biologists, ecologists, and other researchers to perform complex analysis and visualizations using their own data either in real time mode or offline mode. More specifically, dynamic tracking and several spatial-temporal analysis methods are developed, including (1) Home range analysis for determining the use of spatial extent or outside boundary of wildlife movements during the course of their everyday activities; (2) Spatial clustering using density-based spatial clustering of applications with noise (DBSCAN), for grouping homogeneous objects in the sense of geographic spaces; (3) Lacunarity analysis for describing spatial heterogeneity of 1D animal trajectories and 2D animal locations; and (4) Herd evolution analysis for capturing dynamic variations of quantity and quality in herd membership. The

application of the extension is demonstrated using simulated data, test data from a GPS collar, and real ARGOS data from albatrosses in the Pacific Ocean. The results suggest that the wildlife tracking extension for ArcGIS has great potential for wildlife tracking and relevant applications.

REFERENCES

- Aebischer, N., Robertson, P., and Kenward, R.E., 1993. Compositional analysis of habitat use from animal radio-tracking data. *Ecology*. 74:1313-1225.
- Ankerst, M., Breunig, M., Sander, H., 1999. Optics: ordering points to identify the clustering structure. *SIGMOD Rec.*, 28, 49-60.
- ARGOS, 1978. Worldwide tracking and environmental monitoring by satellite.
<http://www.argos-system.org/>.
- Benkert, M., 2006. Reporting flock patterns. In Proceedings of the 14th conference on Annual European Symposium, pp. 660-671.
- Best, B.D., Patrick, N. Halpin, Andrew, J. Read, Song, S. Qian, Hazen, J., Schick, S., 2007. Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecological Informatics* 210-223.
- Bian, L., 2000. Component modeling for the spatial representation of wildlife movements. *Journal of Environmental Management* 59, 235–245.
- Brillinger, R. David, Preisler, K. Haiganoush, Ager, A., Kie, G., 2004, An exploratory data analysis (EDA) of the paths of moving animals. *Journal of Statistical Planning and Inference* 122 43-63.
- Burgman, C., 2003. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Animal Conservation* 6, pp. 19-28.
- Burroughs, J., 1902. John James Audubon. Boston: Small, Maynard & Company.
- Burt, H., 1943. Territoriality and home range concepts as applied to mammals. In *Journal of Mammalogy* 24, p. 346352.
- Cagnacci, F., Urbano, F., 2008. Managing wildlife: A spatial information system for GPS collars data. *Environmental Modelling & Software* 23: 957-959.
- Calenge, C., Draya, S., 2009. The concept of animals' trajectories from a data analysis perspective. *Ecological Informatics* 34-41.
- Cao, C., Nikos, M., and David, W. Cheung, 2005. Mining frequent spatio-temporal sequential patterns. In Proceedings of the Fifth IEEE International Conference on Data Mining, pages 82–89.
- Chandran, R., 2004. Wildlife enforcement monitoring system. In *Journal of Statistical Planning and Inference* 122, pp. 43-63.

- Cochran, W. W., Lord, R. D., 1963. A radio-tracking system for wild animals. *Journal of Wildlife Management* 27:9–24.
- Craighead, D., 1986. Movements of caribou in western arctic herd. *Argos Newsl.* Nr. 26, 1-3.
- Creel, S., Creel, N. M., 2002. *The african wild dog: behavior, ecology, and conservation.* Princeton University Press, Princeton, New Jersey, p341.
- Dixon, K.R., Chapman, J.A., and Dominic, J. Demma, 1980. Harmonic mean measure of animal activity areas. *Ecology.* 61:1040-1044.
- Dong, P., 2000. Lacunarity for spatial heterogeneity measurement in GIS. *Geographic Information Sciences* 6, 20–26.
- ESRI, 2007a. ArcGIS geostatistical analyst. In *Sixth International Symposium on Wearable Computers.*
- ESRI, 2007b. ArcGIS Desktop Help 9.2. <http://webhelp.esri.com/arcgisdesktop/9.2/>.
- Ester, M., Hans-Peter, K., Jörg, S., Xing, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of 2nd International Conference on KDD, AAAI Press.*
- Fernandez, P., Anderson, D.J., Sievert, P.R., and Huyvaert, K., 2001. Foraging destinations of three low-latitude albatross (*Phoebastria*) species. *Journal of Zoology*, 254: 391-404.
- Ford, R. G, Krumme. D. W, 1979. The analysis of space use patterns. *Journal of theoretical biology.*
- Getz, M., Wilmers, C., 2004. A local nearest neighbor convex-hull construction of home ranges and utilization distributions. *Ecography* 27, pp. 489-505.
- Gifford, A., Leonard J.V. Compagnob, Marie Levine, Alex Antoniou, 2007. *Fisheries Research*, 84: 17–24.
- Giltsdorf, J.M., Vercauteren, K.C., Hygnstrom, S.E., Walter, W.D., Boner, J.R., and Clements, G.M., 2008. An integrated vehicle-mounted telemetry system for VHF telemetry applications. *The Journal of Wildlife Management* 72, 1241-1246.
- Grigg, G., 1987. Tracking camels. *Argos Newsl.* 29, 1-4.
- Gudmundsson, J. and Kreveld, M., 2006. Computing longest duration flocks in trajectory data. In *GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 35–42.

- Halpin, P.N, Read, A.J, Best, B.D., Hyrenbach, K.D, Fujioka, E., Coyne, M.S., Crowder, L.B., Freeman, S.A., Spoerri, C., 2006. OBIS-SEAMAP: developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles. *Marine Ecology Progress Series*. P. 239-246.
- Hebblewhite, M., Percy, M., Merrill, E. H., 2006. Are all global positioning system collars created equal? Correcting habitat-induced bias using three brands in the central Canadian Rockies. *Journal of Wildlife Management* 71(6), 2026-2033.
- Hins, C., Ouellet, J., Dussault, C., and St-Laurent, M., 2008. Habitat selection by forest-dwelling caribou in managed boreal forest of eastern Canada: Evidence of a landscape configuration effect. *Forest Ecology and Management*. P. 636-643.
- Hooge, P.N., Eichenlaub, W., and Solomon, E., 1999. The animal movement program. USGS, Alaska Biological Science Center.
- Howey, W. P., 1994. Tracking the migration of the peregrine falcon. *Argos Newsl.* 48, 19.
- Hsu, H., Joung, S., Liao, Y., Liu, K., Satellite tracking of juvenile whale sharks, *Rhincodon typus*, in the Northwestern Pacific.
- Huang, Y., Chen, C., Dong, P., 2008. Modeling herds and their evolvments from trajectory data. *GIScience 2008*, USA.
- Jennrich, R.I., F.B. Turner. 1969. Measurement of non-circular home range. *J. Theory. Biology*, 22, 227-237.
- Kays, R.W. and Wikelski, M., 2007. Review of the NSF sponsored animal tracking and physiological monitoring workshop. Report published online at www.movebank.org
- Laube, P., Kreveld, M., and Imfeld, S., 2005. Finding REMO: detecting relative motion patterns in geospatial lifelines. In *Developments in Spatial Data Handling*, pp. 201-215.
- Laube, P., Imfeld, S., 2002. Analyzing relative motion within groups of trackable moving point objects. In *2nd International Conference on Geographic Information Science*, pp. 132-144.
- LeMunyan, C. D., White, W., Nyberg, E., and Christian, J. J., 1959. Design of a miniature radiotransmitter for use in animal studies. *Journal of Wildlife Management* 23:107-110.
- Le Maho, Y., Gendner, J.P., 1994. Tracking the emperor penguin across the pack ice. *Argos Newsl.* 48, 13-15.
- Levine, N., 2007. Crimestat: Spatial statistics program for the analysis of crime incident locations. In *Sixth International Symposium on Wearable Computers*.
- Lloyd, M., 1967. Mean Crowding. *J. Anim. Ecol.* 36:1-30. Bureau of Animal Population, Department of Zoological Field Studies, Oxford University, England.

- Mackenzie, D., 2005. Was it there? Dealing with imperfect detection for species presence/absence data. *Australian & New Zealand Journal of Statistics* 47:65-74.
- Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Deana, P., Ferdinando, V., 2007. An ontology for describing and synthesizing ecological observation data.
- Mandelbrot, B., 1982. *The fractal geometry of nature*. Macmillan, San Francisco.
- Meek, M.G., Cooper, S.M., Owens, M.K., Cooper, R.M., Wappel, A.L., 2008. White-tailed deer distribution in response to patch burning on rangeland. *Journal of Arid Environments* 72: 2026– 2033.
- Meier, T.J., Burch, J., and Adams, L., 2006. Tracking the movements of Denali's solves. *Alaska Park Science* 5(1): 30-35.
- Mooty, J. Jack, Patrick, D. Karns., and Fuller, K. Todd, 1987. *Habitat use and seasonal range size of white-tailed deer in northcentral minnesota*, Allen Press.
- Morishita, M., 1959. *Memories of the faculty of science series E, Biology*, vol. 2. Fukuoka, Japan: Kyushu University; 215.
- Ozoga, J. J., Verme, L. J., 1975. Activity patterns of white-tailed deer during estrus. *Journal of Wildlife Management* 39:679-683.
- Ozoga, J. J., Verme, L. J., and Bienz, C. S., 1982. Parturition behavior and territoriality in white-tailed deer: impact on neonatal mortality. *Journal of Wildlife Management* 46:1-11.
- Person, T. Brian, Prichard, K. Alex, Carroll, M. Geoffry, Yokel, A. David, Suydam, S. Robert, George, C. John, 2007. Distribution and movements of the Teshekpuk caribou herd 1990-2005: prior to oil and gas development. *Arctic Articles*, P. 238-250.
- Plotnick, R.E., Gardner, R.H., Hargrove, R.H., Prestegard, K., Perlmutter, M., 1996. Lacunarity analysis: A general technique for the analysis of spatial patterns. *Physics Review*, V53 P5.
- Reynolds, T.D., Laundré, J.W., 1990. Time intervals for estimating pronghorn and coyote home ranges and daily movements. *J. Wildl. Manage.* 54:316-322.
- Saunders, S. C.; Chen, J.; Drummer, T. D.; Gustafson, E. J.; Brosofske, K. D, 2005. Identifying scales of pattern in ecological data: a comparison of lacunarity, spectral and wavelet analyses. *Ecological Complexity*, Vol. 2: 87-105.
- Schwager, M., Anderson, M. D., Butler, Z., Daniela, R., 2007. Robust classification of animal tracking data. *Computers and Electronics in Agriculture* 56 (2007) 46-59.
- Seaman, D.E., and R.A. Powell. 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*. 77:2075-2085.

- Shaffer, S.A., Daniel, P.Costa, 2006. A database for the study of marine mammal behavior: Gap analysis, data standardization, and future directions. *IEEE Journal of Oceanic Engineering*. 31 (1), pp. 82-86.
- Silverman, B. W., 1986. *Density estimation for statistics and data analysis*. Chapman and Hall, London, UK.
- Spencer, S.R., G.N. Cameron, and R.K. Swihart. 1990. Operationally defining home range: temporal dependence exhibited by hispid cotton rats. *Ecology*. 71:1817-1822.
- Taiga, 1995. Porcupine caribou herd satellite collar project. <http://www.taiga.net/satellite/>.
- Terraseer, 2007. Spacestat: International standard for spatial econometric modeling. In *Sixth International Symposium on Wearable Computers*.
- Toby A. Patterson, Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J., 2009. State-space models of individual animal movement. *Trends in Ecology and Evolution* Vol.23 No.2, 87-94.
- Warner, R.E., Etter, S. L., 1983: Reproduction and survival of radio-marked hen ring-necked pheasants in Illinois. *J. Wildl. Managmt.* 47: 369-375.
- White, G. C., Garrott, R. A., 1990. *Analysis of wildlife radiotracking data*. Academic Press, San Diego, California, USA.
- Whittow G.C., 1993. The Laysan albatross (*Diomedea immutabilis*). *The Birds of North America*. The American Ornithologists' Union, Washington, D.C. No. 66. P. 1-20.
- Winnie, J.A., Cross, P., Getz, W., 2008. Habitat quality and heterogeneity influence distribution and behavior in African buffalo (*Syncerus caffer*). *Ecology*, 89(5), pp. 1457-1468.
- Wong, I.W, Fong, P., Russell, R., and Chen, X., 2007. Species at risk: Data and knowledge management within the WILDSPACE. *Decision Support System Environmental Modelling & Software* 22: 423e430.
- Worton, J., 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70:164–168.
- Young, R. J., Coelho, C.,M., 2006. A note on the effect of the full moon on the activity of wild maned wolves. *Behavioural Processes* 73: 228–230.
- Zhang, J., Liu, D.D., 2007. GBD-Explorer: Extending open source Java GIS for exploring ecoregion-based biodiversity data. *Ecological Informatics*. 94–102.