USING PATTERN ORIENTED MODELING TO DESIGN AND VALIDATE SPATIAL MODELS: A CASE STUDY IN AGENT-BASED MODELING

by

Jerry Patrick Corum

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

December 2014

DEDICATION

I dedicate this document to my parents and siblings for their constant support, and to my grandmother, without whom I would not have received a college education.

ACKNOWLEDGMENTS

I will be forever grateful to my mentor, Professor Kemp and the professors who introduced me to agent-based modeling as an undergrad, Drs. Marco Janssen and Amber Wutich. Thank you also to my family and friends, without whom I could not have made it this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Complexity in spatial simulation models developed without an iterative development process can lead to models that produce inaccurate or nearly random results. This case study examines how real world moving-object data can be used to inform the model development process. Moving-object analysis provides a template for understanding movement behaviors evident in both empirical data and model output. Moving object data generally consists of the GPS points from tracked animals, and is usually acquired as a comma separated values file. Agent-based simulation model development in this case study is informed by pattern oriented modeling, an iterative process used to control a model's complex variables while gradually improving model design. Three simple agent-based models were constructed and a best fit model whose output most closely matches the spatial characteristics of the Galapagos Swallow-tailed Gull moving object data was identified.

.

CHAPTER ONE: INTRODUCTION

This study demonstrates how ecological analysis of moving-object data can be used to empirically validate agent-based models (ABMs) and presents a combination of methods that can be used to analyze moving-object data to inform model development. Integrating ecological analysis in the modeling process proves effective at providing a framework for model development and validation.

Spatial modeling in scientific terms is both a simplified way of representing a spatial system that is being studied and a tool for understanding and predicting processes and behavior (O'Sullivan and Perry 2013). Models are useful to explore changes in the real world as well as help us understand the processes that generate the patterns we observe in the system.

Models are often used for prediction and to assist in data collection, but can also be used as a tool to enable critical thinking about the real world. Examples of this may include predicting weather, identifying critical habitat areas to investigate with sampling techniques, or providing a framework to explore environmental factors and their effect on an ecological system. In this case study I will use a simulation model to explore the process of movement through space and time in an effort to understand and reproduce seabird movement activities.

Spatial modeling comes with a few challenges, especially that of identifying when too much complexity is present for the model to be useful. It is also necessary to decompose the real-world into component parts that are modeled and decide which factors are more likely to affect the system in a meaningful way. In some cases, a lesson learned early on in exploring agent-based models for this study, too much complexity can slow a model down and also produce unpredictable and unreliable results.

This study demonstrates an iterative process for validating a random walk model, a fundamental building block model of agent-based modeling, designed to simulate bird flight tracks. The objective is to integrate Geographic Information Science (GIScience) methodologies with computer science and ecological epistemologies to inform model development through a gradual process beginning with simple movement. The methodology builds on work in agent-based model validation and ecological analysis to provide a multidisciplinary tool for examining agent-based models during the development process using moving-object data.

1.1 Understanding moving-object behavior

Home range estimation, tortuosity or searching intensity, and linear or sequential spatial autocorrelation are tests of range, behavior and independence in movement patterns and provide important insight into the structure of moving-object data. Ecological studies using records of movement tracks of insects, birds and marine and terrestrial mammals commonly include many such tests of the track trajectories to assist in developing an understanding of the moving-object's behavior (Bence 1995; Benhamou 2004; Brillinger et al 2002; Colomb et al 2012; Fauchald and Tveraa 2003; Kareiva and Shigesada 1983; Lichstein et al 2002; Root and Kareiva 1984). As illustrated in this study, modeling is an effective tool for understanding the structure of a moving object's motive force and assists in the exploration of the real-world data.

Knowledge of movement patterns in animal behavior is imperative for understanding the results of simulated behavior in ABMs. Through the use of analysis techniques developed in ecological studies of animal behavior, it is possible to look at the overall shape of the animal's track without regard to its feeding or nesting behavior, making it a suitable approach for analyzing movement patterns simulated by agent-based models. In Figure 1, the paths of four

Galapagos Swallow-tailed Gulls are displayed. These four birds were tracked using GPS collars and are the moving-object data for this study.

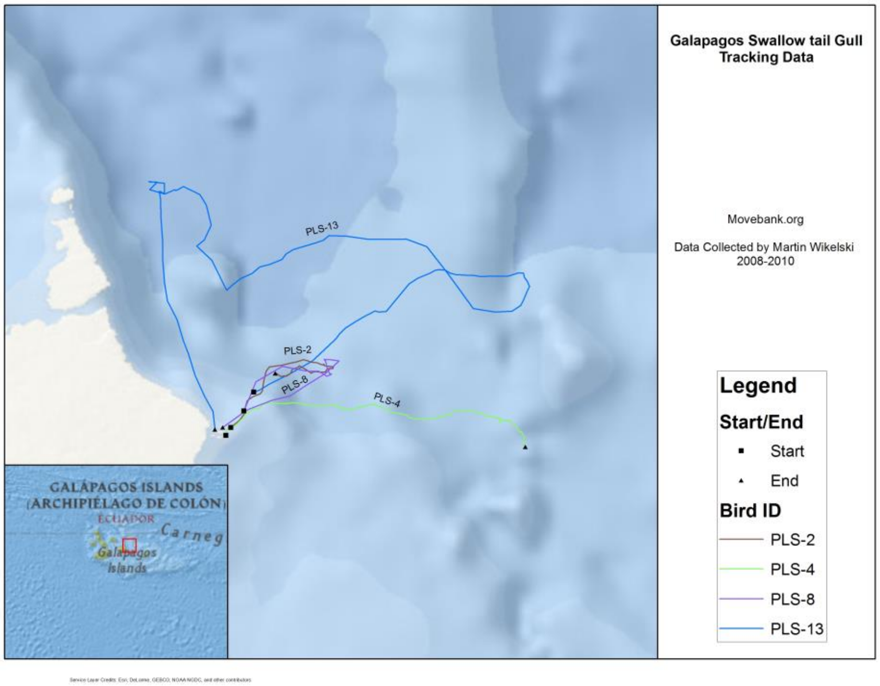

Figure 1 Galapagos Swallow-tailed Gull relocation flight paths constructed from tracking data

It is the goal of this study to decompose the moving-object data's spatial elements to their component parts and use those parts to act as a filter for the agent-based modeling movement processes.

1.2 Overview of the study

The study began with the review of a set of analytical tools in ecological analysis that include statistics for the shape of the movement track, the independence of the movements and the home range area. These were evaluated to design a framework for the analysis of patterns in empirical data available from Movebank.org, a repository of tracking data collected and shared with an open source license. Using these analysis tools, a moving-objects dataset which records the flight paths of Galapagos Swallow-tailed Gulls off the coast of Santa Cruz, Ecuador was analyzed to evaluate movement metrics that describe the patterns observed in a form that is transferrable to ABM. Then, three agent-based models were developed to simulate the birds' movement using increasingly complex forms of a random walk model with identical analytic techniques applied to filter out results that are inconsistent with real world behaviors.

The agent-based models were programmed to simulate three different mathematically modeled movement patterns: a simple random walk, a correlated random walk and a correlated random walk with variations in speed that match the mean and standard deviation of a normally distributed curve modeled from the empirical data. Each model revision was intended to add structure to the simulated movement patterns that better replicated the actual movement tracks. Importantly, the models are not designed to *predict* the movement patterns evident in the seabird dataset, rather they seek to reproduce similar overall movement patterns, and thus provide insight into mechanisms behind the observed patterns. This means it is possible to model reactions to environmental change or habitat incursion without actually changing the real-world environment. The data from Movebank.org acted as a means to assess for and filter out poor program design decisions, eliminating choices that were unable to produce similar trajectories and movement structures.

Agent-based modeling is a bottom up approach to modeling that looks for pattern from process (O'Sullivan and Perry 2013). These models simulate the decisions of multiple individuals simultaneously and allow patterns to emerge from this complexity. One classic example of agent-based modeling is a predator-prey model, which tracks each individual predator and prey's actions and bases the decisions they make on environmental and neighboring attributes. While a predator-prey model is relatively simple, often these models can be incredibly complex. In these cases standards for the ABM empirical validation process are called for but are often ignored (Janssen and Ostrom 2006).

Through agent-based modeling it is possible to examine the processes that create movement patterns in moving-object data. It is possible to use an ABM to explore the effects of habitat extent and resource availability on movement and resource exploitation. The exploration of these effects requires validated simulation output to be meaningful.

The method of empirical validation is specific to each model and the processes and patterns that are modeled. In the analysis of moving-object data in ecology, trajectory and home range analysis can help validate the simulated movements. The ecological techniques may, for example, decompose the animals' track into relative angle changes between each track segment or a habitat area into a cluster of core movements. The 'clusthr' tool identifies three locations with a minimum mean nearest-neighbor joining distance (NNJD), which forms the first cluster. Then, in set steps the search expands and locations (which in this case are relocations, or the sequential locations observed in the dataset) are added based on the next set of NNJD clusters with the minimum mean distance to the first, until 100% of the locations are incorporated. Such decomposed behavior can be simulated with an agent-based model and exported for analysis and validation against real-world data.

Using the seabird data as a case study, the use of empirical data to evaluate the success of interactively developed agent-based models was demonstrated. Applying the same tools and metrics used to evaluate the movement patterns evident in the Movebank.org data to the simulated data resulting from the agent-based models indicates that it is possible to assess whether an agent-based model can accurately reproduce the structural components of moving-object data.

1.3 Research Objective and Process

The objective of this research was to examine the practical aspects of using an analysis of moving-object data to validate an agent-based model during the iterative model development process. The model generates moving-object data for analysis in a geographic information system using the ecological analysis tools in the R-project for statistical computing, including sequential autocorrelation, habitat estimation, and trajectory analysis. Model development began by modeling very basic random walks, incorporating more complex random walk models as analysis progressed. Based on ecological analysis carried out during modeling, it was concluded that correlated random walks and correlated random walks with varying speed produce results that most closely match real-world observations. Further development involving modifications based on animal behavior will be necessary to replicate the empirical data.

The remainder of this document describes the study in detail. Chapter 2 provides the multidisciplinary framework used in this study and establishes the link between agent-based modeling and ecological analysis. Chapter 3 provides detailed methods and results. Chapter 4 discusses the results and Chapter 5 sums up the study, provides conclusions from the results and raises questions for future research.

CHAPTER TWO: THEORETICAL FRAMEWORK

Two scientific fields provide the theory and methods used in this research, agent-based modeling and ecological analysis. Agent-based modeling provides a tool to investigate processes that underlay complex patterns in social science, biology, and other disciplines. The tools of ecological analysis decompose complex patterns and provide quantifiable observations of real-world processes. Ecological analysis laboratory studies of bugs in a maze (Colomb et al. 2012) and tracked animal data (Brillinger et al. 2002) provide a framework for decomposing moving-object data. The building blocks of agent-based modeling (see O'Sullivan and Perry 2013) provide the basis for using decomposed moving-object data as a means of structuring an agent-based model to simulate movement behaviors accurately using an iterative model development process.

Several approaches are used in each discipline, however there are unifying elements such as the assumptions made regarding animal movement, which make this study possible. Simple and correlated random walks are the subject of much discourse in both agent-based modeling and ecological studies (Kareiva and Shigesada 1983, Grimm et al 2005, Haefner 2005). Fundamentally both walking styles form the basis of modeling animal foraging behavior in both agent-based modeling and ecological analysis and even have many applications in economics, psychology, physics, chemistry and biology (Van Kampen 1992, Goel and Richter-Dyn 1974, De Gennes 1979, Weiss 1994).

Several software applications and statistical techniques were used in this study to create the models and analyze the output. Agent-based models were created in the Netlogo modeling environment using the language of the same name. The ecological analysis was done in R using several packages that are tailored to ecological movement analysis, the main package being ‘adehabitatlt’ and its dependencies, ‘sp’, ‘CircStats’ and others installed automatically.

2.1 Ecological Analysis and Moving-object Data

Tracked animal behaviors in ecological analysis form an important part of this study. Ecology is the scientific study of interactions between and among organisms, thus movement takes an important role in an ecological analysis. Moving-object data consists of a set of relocations, either from GPS tracking or sighting and observations. These can be part of an ecological study, lab study, or any set of data with tracked movement. Encoding and storing moving-object data in a useful way is critical not only for analysis but for meaningful sharing of data. Data without a standard, especially without metadata, can be difficult to decipher and use even for the original collector if enough time passes. Storing and using moving-object data is the focus of several associations including the Open Geospatial Consortium, who develop standards for open spatial data storage to extend usability.

2.2 Agent-based modeling

Agent-based (sometimes called individual-based) models (ABM) are composed of collections of individual objects that are unique and autonomous, interacting with each other and their immediate spatial environment (Railsback and Grimm, 2012). While generally based on cellular automata models developed in the 1940’s by John Von Neumann, it was the Game of Life by John Conway that spurred the development of methods of modeling of simple rules to investigate complex patterns (Neumann and Burks 1966; Gardner 1970, 120-123). In many cases agent-based modeling is considered modeling from the bottom up (i.e. from the individual actions that combine into the big picture).

In agent-based modeling an agent is the individual agent that makes decisions based on the patches around it. The Netlogo patches are a raster surface. The patches can be both agents and dynamic variables (resource quantities, environmental conditions, weather, etc.) that change, assess and react during each tick, or time-step, along with agents, also called turtles, which move across the surface. Each agent assesses the world around it during a tick, makes the decisions based on how it is programmed and organizes or prepares any variables it needs for the next tick.

2.2.1 Modeling languages for ABM

Many software platforms have been created for agent-based modeling. Repast, Swarm, Netlogo, and the Multi-Agent Simulator Of Neighborhoods (MASON) are just a few. These languages may have been developed for specific purposes, such as modeling social complexity, as in the case of MASON, or they can be general purpose like Repast and Netlogo. Netlogo was selected for this study because of its supportive community of users, relatively easy learning curve and access to high quality tutorials and textbooks. Netlogo was developed by Uri Wilensky in 1999 at Northwestern University and is derived from the educational programming language Logo, known for Turtle, a robot whose movements across the floor could be easily programmed by school children. Netlogo is likewise an easy to learn programming language that is well documented and supported by a community of active developers including students, professors and professional consultants (www.ccl.northwestern.edu/netlogo). There is a wide variety of open simulations to play with and learn from, and many instructional books. The output is spatial in nature, with x and y coordinates marking location in the Netlogo world, making it intuitive to use the tools of a GIS to analyze the results.

2.2.2 ABM building blocks

While developing and understanding complex, agent-based models about the real-world is usually very difficult, O'Sullivan and Perry (2013) suggest that there are three fundamental building blocks that make up a complete agent-based model's basic structure. These building blocks, grouping, mobility and spread, are discussed below.

Grouping is comprised of segregation and aggregation processes that produce heterogeneity in the landscape. These processes are used as a part of models that explore neighborhood segregation or the evolution of patchy landscapes in ecological studies. Most often these processes use operations like local averaging, in which each tick provides a chance for an agent to move closer to other agents with similar attributes. This behavior is directly observable in the real-world; examples include gentrification and ecosystems models showing clear patches of homogeneity in a landscape.

Mobility is embodied in random walk models. These involve the movement decisions that cause an agent to act on, and react to, the environment around it. Simple random walks involve picking a random direction on a grid and moving one step, then repeating the process. Since there are 360 possible directions to move in the study models, over time the agent is not likely going to be far from its point of origin (O'Sullivan and Perry 2013; Šalamon 2011; Railsback and Grimm 2011). A more complex version of the random walk, called a correlated random walk, provides a more realistic process for movement. In a correlated random walk, the decision on which direction to move is related to the direction previously traveled, meaning if the agent moved east in the previous tick, the agent will have higher odds of choosing a direction that is similar. Restricting an agent's change in direction by limiting the relative angle of movement on subsequent ticks is one way of programming a correlated random walk.

The use of random walks for modeling real-word behavior is supported by biological and ecological studies (Kareiva and Shigesada 1983, 234-238; Benhamou 2004, 209-220). The random walk is considered a stochastic model, but it does not focus on the end product – the combined movements of the animals -- rather it models how each individual contributes to the overall movement of the population (O'Sullivan and Perry 2013). While often used in physics (Berg 1993; Rudnick and Gaspari 2004), pure random walks may seem unrealistic for ecological analysis, since it is likely that almost no living animals move in this way. Nevertheless, variations on random walks are the focus of this particular study, as random walks form a foundation for understanding movement without directly modeling observed movement itself.

Spread, or growth and reproduction form the final building block. This deals with how agents procreate and spread their influence across a landscape. This is different from simply moving through an area as in the random walk because the agent becomes a part of the landscape, acting on it and changing it. It incorporates that landscape into its area of influence, or home range.

2.2.3 Validation of ABM Results

Validation of agent-based models can be a long process and can be very difficult. The entire system must be checked against all the possible variations present in the programs structure. In the simple models in this study, only a few variables exist, such as allowed relative angle changes and speed, however in models that incorporate foraging behaviors, exploration, memory, and links between agents (growth and reproduction) the number of variables that must be tested can be intimidating.

Sensitivity analysis alone represents a significant investment in time (Janssen and Ostrom 2006, 37; Grimm et al. 2006, 115-126; Macal and North 2007, 95-106; Parker and

Meretsky 2004, 233-250; Topping, Høye, and Olesen 2010, 245-255; Valbuena et al. 2010, 185-199). Sensitivity and Uncertainty analysis require repeated model runs with small variations in parameters. Sensitivity analysis often allows a model builder to filter out obviously incorrect model behavior by observing which variables have the most effect on the outcome. Using tools such as Netlogo's BehaviorSpace it is possible to automate the runs and set ranges for variables to vary. This produces a great deal of output, both tabular and spatial that needs stored and analyzed. Both running the model and the analysis can be extremely time consuming. In early models runs using Scott Hekbert's 2013 model, MayaSim, one hundred runs produced over 30 gigabytes of data and took between 24-36 hours to complete on a personal computer.

2.2.4 Pattern Oriented Modeling

Evaluating uncertainty on each building block in the iterative development process will not produce results worth the significant investment of time needed to run the model enough times to perform the analysis. Pattern Oriented Modeling (POM) helps minimize these analysis steps by identifying relevant patterns in a real system that are relevant to the questions being asked, preventing a model from becoming over-parameterized (Grimm and Railsbeck, 2012). Since the model building step begins with identifying patterns, or behaviors that fall beyond random variation that are then reproduced indirectly but purposefully in the model, the analysis step will begin with a series of controlled experiments on the model itself (Grimm and Railsbeck, 2012, Salamon 2011).

POM begins with patterns found in real-world systems and develops a hypothesis to explain the pattern. Predictions based on that pattern are then tested and the model is adjusted accordingly and the process begins again. POM is used to assist in the development of an ABM. This allows model parameters that are not adequately explaining the observed real world patterns to be filtered out. It is an iterative and incremental method to tweak the model and balance the model's complexity with that relevance of what can be learned from it (Railsback and Grimm 2011; Šalamon 2011; O'Sullivan and Perry 2013).

Pattern oriented modeling can be used in any agent-based modeling programming language or environment. The approach presented here should be a part of the iterative model development process (see Salamon 2011; Grimm and Railsback, 2012;), preceding and supporting the sensitivity analysis.

2.3 The R Project and Ecological analysis

The R Project for Statistical Computing is an open source statistical programming language and software environment used for statistical computing and graphics (www.r-project.org). The core functionality of the environment was originally developed at the University of Auckland in New Zealand. R is available for a variety of operating systems. It is community supported, with many online forums for questions and a robust set of tutorials and publications providing tutorials and support. In R, custom sets of tools called *packages* make research using discipline-specific analytical techniques more accessible to non-specialists. There are many packages available through the Comprehensive R Archive Network (CRAN). To make it easy to find packages for specific purposes, CRAN offers several *task views* that organize packages into thematic subsets such as Environmetrics, Spatial or SpatioTemporal groups. Packages are open source and made available in CRAN with a reference manual and optional vignette document with a walkthrough and illustration of the package functions.

2.3.1 R functions and Packages used in this study

ACF and PACF take a time series of recorded values as input with missing values removed. These functions provide a tool for looking at periodicity in the datasets by investigating repeating patterns in the dataset and are also used to investigate sampling error (Venables and Ripley 2002). Used in this study on the changing relative angles of each relocation, both ACF and PACF look for patterns that can show searching behavior of real world tracked seabirds, a movement which may be difficult to reproduce in an agent-based model (Bence 1995, 628-639; Lichstein et al. 2002, 445-463). PACF is primarily used to fit an Autoregressive Integrated Moving Average (ARIMA) model, which does not fall within the scope of this study and is not used. ACF may be used to test sampling error in both the real world dataset and the simulated one (which represents a simulated sample).

R's 'adehabitatLT' package was chosen for this study because of its flexible nature and ability to convert a wide variety of data formats to its 'ltraj' object class, as well as providing compatibility with other packages tailored to ecological analysis and the study of moving-objects. Adehabitat is a collection of tools for analyzing habitat selection by animals and includes tools for modeling error and uncertainty as well as understanding how animal tracks influence habitat range.

A relocation in the 'adehabitatLT' package is considered a sequential or timed observation of an animal's location in space. This data must contain coordinates and a sequence identifier, or time and often includes information related to the observation such as weather, wind speed and direction, elevation, and other miscellaneous data. This data is imported into the 'ltraj' object class in order to begin the analysis. The 'ltraj' object calculates additional measures when the data is converted that enable more advance analysis by the package. Critical to this study is the calculation of relative angle changes, referred to in ‘adehabitatLT’ literature as the shape of an animal’s movement. The ‘ltraj’ object calculates the following measures:

1. Change in x-coordinate (dx)
2. Change in y-coordinate (dy)
3. Distance to previous relocation
4. Absolute angle change
5. Relative angle change

Because the ‘ltraj’ object contains calculated values for the relative angle changes between each relocation, it is possible to begin investigating the overall shape of movement through space and time. Extracting the relative angle changes makes it possible to use time series analysis to investigate periodicity and the overall shape of the relocations. The relative angle values of sequential relocations can then be converted to smoothed cosine values for analysis in the time series. The ‘sliwinltr’ transformation in the ‘adehabitatLT’ package creates a sliding window chart of smoothed cosine values, which is used to investigate tortuosity, or searching behaviors (Benhamou 2004, 209-220). Benhamou states that cosine values near 0.5 are considered tortuous searching behaviors, possibly when the bird is circling an area looking for prey, while values close to 1 or 0 are more linear behaviors, possibly indicating navigating to a location from memory or fleeing a predator.

This study uses several functions found within the core R package as well as the ‘adehabitatLT’ and ‘tseries’ packages. Each of the functions used in this study is summarized in Table 1.

Table 1 Summary of R functions used to investigate ABM

Function Name	Found in package	Brief Description	Example Use
acf	Core R	Autocorrelation function	Examine linear autocorrelation of relative angle changes
wawotest	adehabitatLT	Wald-Wolfowitz test of independence	Finds data in a sequence that doesn't belong.
clusthr	adehabitatHR	Estimates home range by single-linkage cluster analysis and produces a Multiple Convex Hull object to store the data	Identify home range extent from tracked animals relocation data
MCHu2hrsize	adehabitatHR	Calculates home range size from Multiple Convex Hull object with specified percentage levels for the home range	Examines the rate of home range increase – can identify exploration and foraging behaviors
sliwinltr	adehabitatLT	Applies any function to an 'ltraj' object using a sliding window	Used to investigate relative angle changes
ts.plot	stats (core R)	Plots a time series	Used to plot the relative angle changes through time
testang.ltraj	adehabitatLT	Independence test for successive angles (relative or absolute)	Tests for abnormal patterns or periodicity that can result from sampling error

CHAPTER 3: METHODS

In this chapter the basic methods of agent-based modeling and ecological analysis will be reviewed. I have chosen to place the analysis and results of the seabird data in this chapter in order to provide a context for understanding the ecological metrics that are used. Additionally, since the analysis of this real-world data was necessary prior to beginning the model development process, this step is part of the methods.

The statistical software available through the R-Project for Statistical Computing, and several associated packages tailored to GIS and ecological habitat modeling were used to assess the moving-object data. The primary packages used are adehabitatLT and adehabitatHR (Calenge, C., 2006). Using Movebank.org, tracked locations of Galapagos Swallow-tailed Gulls, collected by Martin Wikelski from 2008 to 2010 were downloaded to provide real world, moving-object data to act as a filter for building the agent-based model. These paths are displayed in Figure 2. This data was assessed prior to creating the model to provide statistical measures to be used as a filter for validating the output of the agent-based model (ABM). Both the general method and results are presented below.

Data was imported into R from a comma-separated values (CSV) file downloaded from Movebank or generated by the ABM. Movebank is an online database of animal tracking data stored at the Max Planck Institute for Ornithology. Scientists are free to put their data on the site, sharing it with others, while still enabling them manage the data closely. It contains several datasets of tracked animals, including tortoise, whales and seabirds. The Galapagos Swallow-tailed Gull dataset (Movebank ID 5503590) provided breeding and non-breeding period data for this study. The XY coordinates, time/sequence and a unique identifier for each tracked location are used. In this case the objects are imported as an ‘ltraj’ object of class II, meaning that the exact time of the observations is ignored in favor of placing the relocations in the correct sequence. The use of the class II 'ltraj' object enables comparison between the Netlogo model output and the real-world bird tracking data.

3.1 Investigating moving-object data, the Galapagos Swallow-tailed Gull

In this study, the simulated data captured from the agent-based model contains four columns; x and y coordinates, the unique agent identifier, and the Netlogo tick number. The Galapagos moving-object data contains latitude, longitude, a time stamp, temperature readings, speed, heading, height (above sea level) and the unique identifier.

3.1.1 Seabird track shape and activity

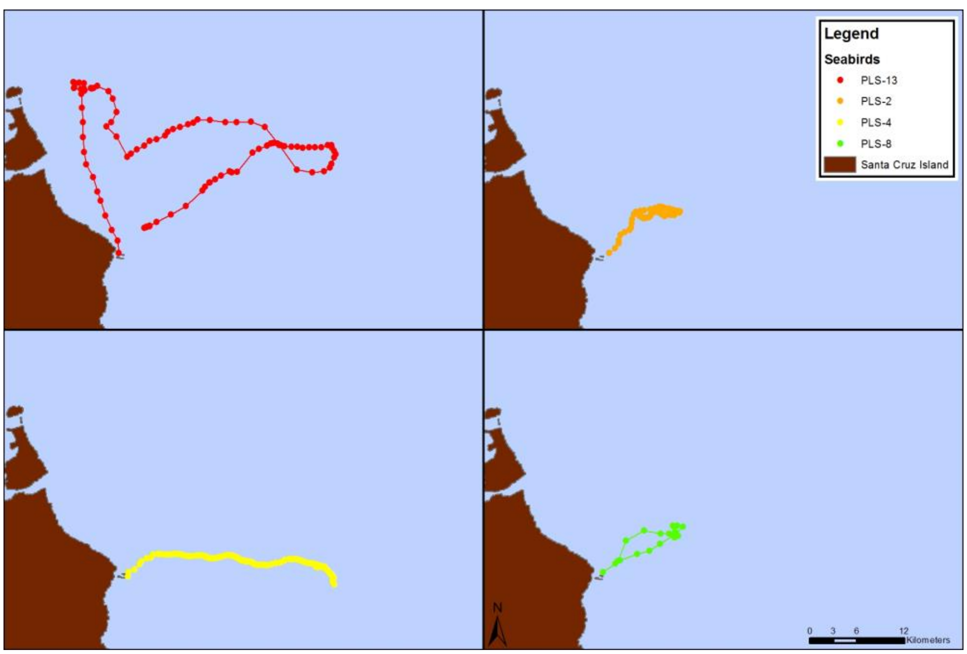

Figure 2 Seabird discrete movement paths sampled at 5-minute intervals

Figure 2 displays the discrete movements of the real-world gulls, sampled every 5 minutes over a time period ranging from approximately 3-8 hours. The starting point is identical for each seabird, right off the coast of the island. Several incidents, including battery consumption and tag damage, can cause the end point. The seabird data imported into the ‘ltraj’ class includes a unique ID for each bird, the number of relocations, or observations taken by the GPS tag, and the number of missing relocations (NA values), or relocations that do not occur every 5 minutes. These numbers are given in Table 2 however in this dataset there are no NA values.

Table 2 Galapagos Swallow-tailed Gull 'ltraj' characteristics

ID	Number of Relocations
PLS-13	99
PLS-2	134
PLS-4	143
PLS-8	22

In order to analyze the overall shape of the seabird relocations, a rediscretizing step is taken, adding or removing point locations to create relocations at regular intervals in space, rather than in time. In mathematics, the process of taking continuous data and breaking it up is called discretizing. Since this has already been done in the seabird data, i.e. the continuous flight path of the gulls is discretized by the GPS sampling their position, the rediscretizing step models a best fit continuous path and discretizes it at the spatial or time intervals chosen for the study. This means that instead of a sequence of relocations 50, 90, 20 and 11 meters distance, all the distances are computed and relocations added or removed to enforce a particular distance as the standard. This effectively fills in the blanks of the relocation data and allows several R tools to assess the changes in relative and absolute angle between each successive relocation (Calenge 2011, Turchin 1998, Benhamou, 2004). The rediscretized movement paths for Galapagos Swallow-tailed Gull observations is given in Figure 3.

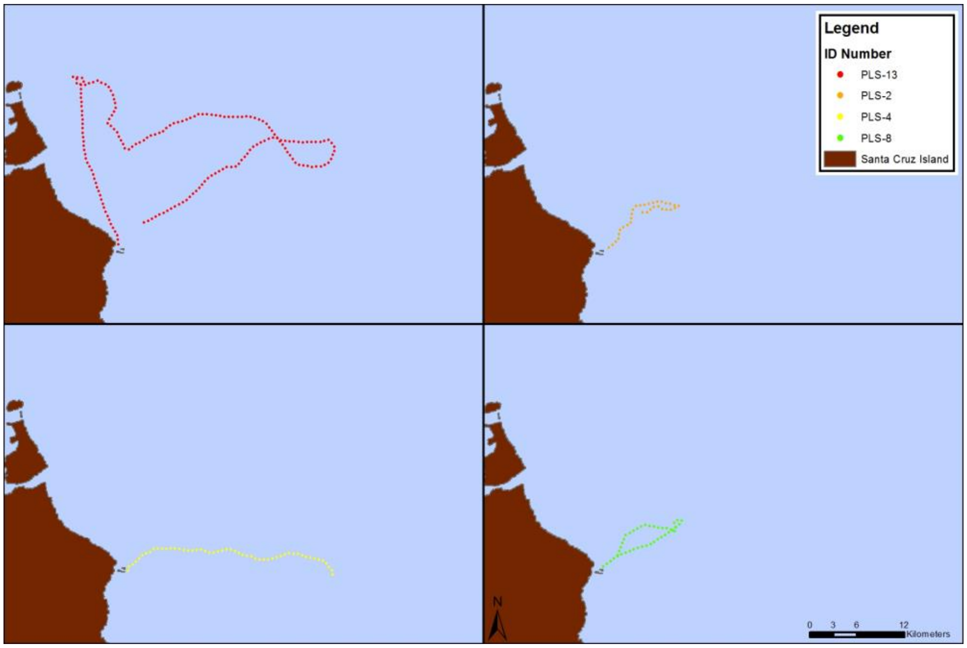

Figure 3 100m Rediscretized movement paths of Galapagos Swallow-tailed Gull observations

Smoothed cosine values of these rediscretized trajectories' relative angles, i.e. the angle changes between each successive movement, provide information on the tortuosity, or intensity of searching behavior in the tracked animal (Benhamou 2004, Colomb 2012). Tortuosity refers to the sharper turn angles of an animal searching for food or a safe location to nest/rest. The use of 'sliwinltr' provides the visual display of these relative angle cosine values in Figure 4. Cosine values near 1 indicate a relatively straight trajectory, while values that approach 0.5 indicate a sharp turn and are commonly considered food or resource searching behaviors (Benhamou, 2004, Calenge, 2011). The seabird data in Figure 4 demonstrates these varied behaviors, linked by Benhamour (2004) to searching behaviors.

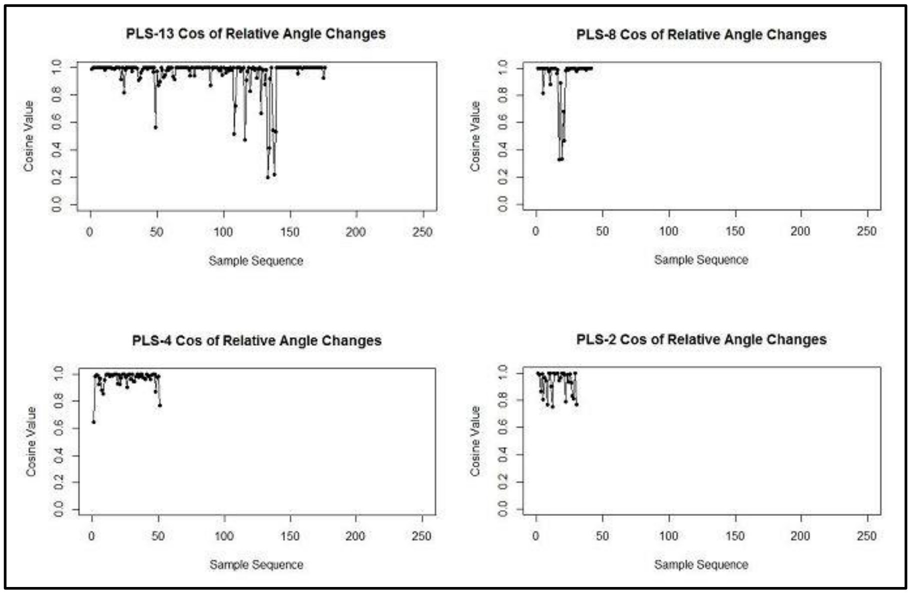

Figure 4 Galapagos Swallow-tailed Gull smoothed cosine values of the relative angles demonstrating searching behaviors

The cosine values of the relative angles in Figure 4 do not show periodic patterns but do demonstrate the same intermittent searching behaviors, with values near 1 and values in the mid-range at times demonstrating tortuous angles, or searching behaviors. Simulating searching behaviors themselves will require a modification of the entire model to include foraging and energetics concepts, additionally predicting the foraging areas will require a model more complex than time will allow for this exploratory study. Thus, foraging locations and behavior are not simulated with these study models. We can test that the underlying movement processes of the foraging behavior are present, specifically that these values vary without any periodic components and that straight line travel and tortuous trajectories are present. The relative cosign values are important for understanding the overall shape of an animal's trajectory. If model development continues beyond the scope of this paper, later iteration of this ABM will necessarily revisit this test as a means of validating foraging behaviors.

Since moving-object data is spatially linear and sequential, the standard tools for investigating error and independence (autocorrelation) in space will not provide useful information. The spatial location of the animal is correlated to its last location because it came from that location; a different approach is needed. Instead, using R and the *adehabitatLT* package 'ltraj' object, it is possible to extract the sequential changes in relative angles to investigate searching behavior. After extracting the data, R's standard autocorrelation ACF function can be used.

The data from the 'ltraj' object is exported to an R dataframe, which is similar to a Microsoft Excel spreadsheet with column names and provides access to the 'ltraj' object's calculated values for the relative angles in each burst, which is the set of relocations for one animal. It is necessary to omit any NA values from the data in order for the ACF function to run, though in this case the real-world seabird data has no NA values.

Figure 5 provides the ACF for individual seabirds in the dataset. Both in the bird dataset and the Netlogo model simulation output, autocorrelation at time equals zero is near 1. The initial angle change has no reference to compare it to, i.e., there is no prior angle for the value to be compared against, causing an edge effect. The ACF values indicate the similarity of observations and identify when abnormal observations occur, whether these are the events of a bird recovering from capture or sampling error from the GPS tracking tags.

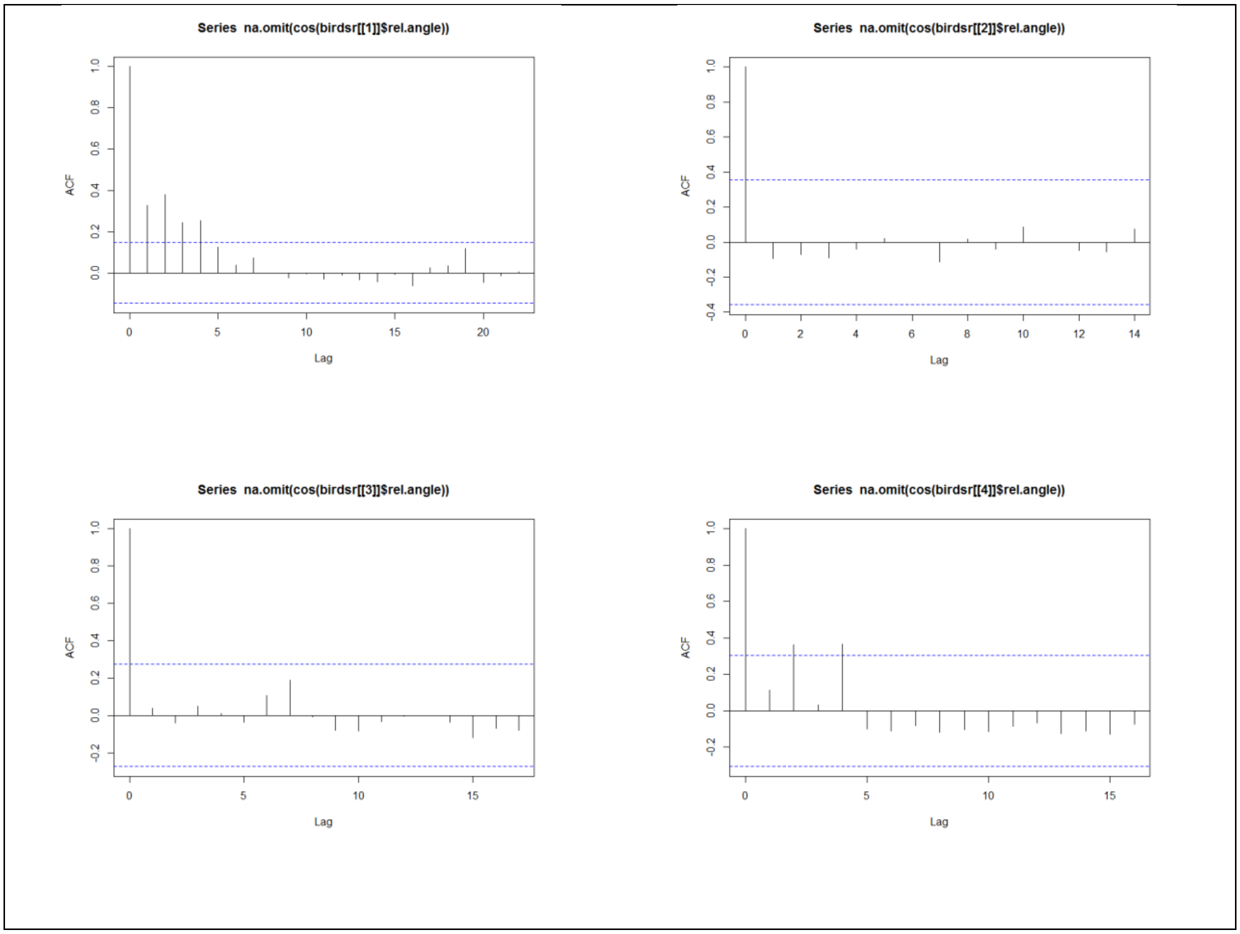

Figure 5 ACF values of seabird relative angles

3.1.2 Seabird relocation independence

Building on the examination of each relocation as a statistically independent event, an overall test of randomness known as the Wald and Wolfowitz Test of Independence (WaWo) is included with the '*adehabitatLT'* package. In the seabird data the p-values are very small for each of the four birds tracked. These values are presented in Table 3. Each is low enough to be considered zero for delta x (dx), delta y (dy), and distance. The WaWo test is a tool for detecting randomness in a sequence of sample values. If the p-values are high, it suggests that there are values present that do not fit in the sequence, i.e. values that are unlikely to occur in a normal distribution.

Table 3 Individual seabird P-values from the Wald-Wolfowitz test of randomness

MPIO Galapagos Swallow Tail Gull WaWo Test P values			
Bird	dx	dy	dist
PLS-13	3.33E-16	1.82E-13	3.87E-09
PLS-2	0	0	0
PLS-4	8.88E-16	0	0
PLS-8	0.000194013	0.04537079	0.000111159

3.1.3 Seabird home range analysis

Home range estimation can be done with several different tools in R and ArcGIS. A simple plot of the birds movement tracks is provided in Figure 2. In this case, the 'ltraj' object is not needed and a staple of R's spatial analysis packages, 'sp', is used instead called a Spatial Points Data Frame. This is basically an Excel sheet but the coordinate values are hidden and accessed through a variety of special calls in the 'sp' package. The 'clusthr' tool provides an intuitive graph of home range size over home-range level, where level is the percentage of points included to calculate the home range (Figure 6).

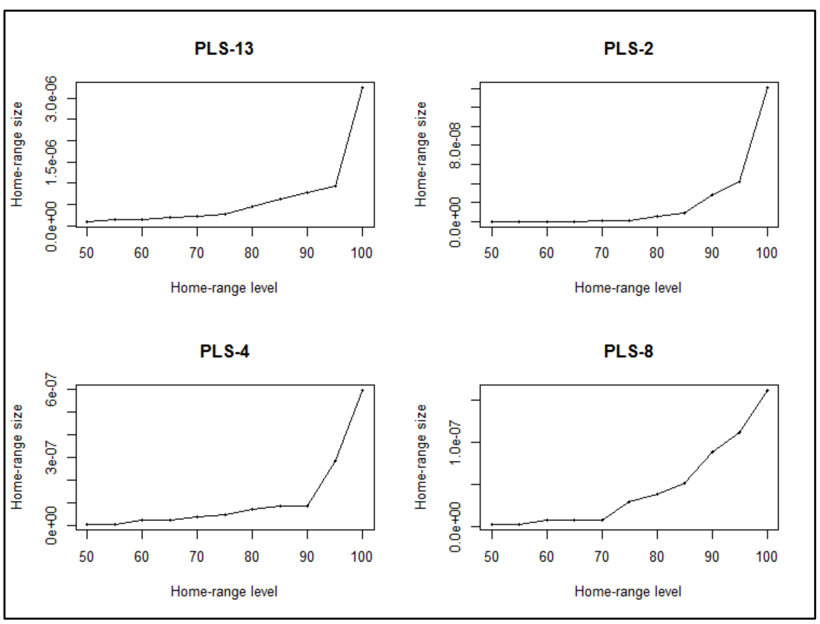

Figure 6 Home range analysis of seabird data using 'clusthr'

It is apparent from these charts that with around 75-80% of points included, the home range begins to increase at an increasing rate. This suggests that a good estimate for the home range will occur with around 75-80% of the points around the first set of clusters identified by the tool. This is important when considering a model of seabird behaviors and suggests that exploration activities may make up around 30% of the relocations for each seabird, though it is possible they are attempting to seek out their home range area after being captured, tagged and released.

3.2 Development of study models through POM

When the real-world data is analyzed, it becomes possible to begin thinking about the creation of the models and how to use the real-world data as a filter on the model output. When comparing a simple random walk model to the moving-object data, there is a large difference in nearly every metric. A truly random walk forms a cluster around the origin point with the probability of movement away from the origin decreasing with distance (O'Sullivan and Perry 2013). The initial form of random walk used in this study is slightly modified to restrict simulated seabird movement over land. After each tick, or time step, the code checks if the seabird is headed toward land or is over land and adjusts the heading out to sea if the movement will take it more than a few hundred meters inland. This location check was developed after examining the tracking data, which revealed rare flights over or near land, and provides a spatial restriction for all three models.

The three simulations model seabird movement off Santa Cruz Island in the Galapagos Islands. The simulation area is modeled to scale after the real-world environment where the tracked data is located. The geography is loaded into the Netlogo environment using Netlogo's

GIS extension. The simulation time is one 8-hour tracking period with observations every 5 minutes to match the 5 minute intervals of the GPS tracking data.

The movement speed for the models is based off the analysis of the seabird data. Each movement in the simple random walk and correlated random walk models is 16 patches per tick. This was calculated based of the average speed of the seabirds, 1217.6 meters per 5 minutes, which when transformed to a Netlogo speed is 16 patches per tick. This is based off the original size of the model area, 53,244.5 m^2, determined using ArcGIS and creating a polygon to cover the extent of the study area, which is used to bound the Netlogo environment. Given that there are two square grids, one 54,244.5 meters in length and the other 701 patches in length, then, the real-world length of one patch in the Netlogo grid is the real world distance divided by the number of patches on one side. This is a ratio based method of transformation and results in approximately 76 meters per Netlogo patch.

These models were constructed specifically for this study, and are very similar. The random walk building block is modified sequentially after investigating the patterns in relation to the real-world data. Code from demonstration models in the Netlogo Modeling Commons, which is included with a Netlogo install, as well as example models from O'Sullivan and Perry (2013) were used to construct each model, with some customization for the environmental restrictions.

3.2.1 Building the basic structure of the models

Three agent-based models were constructed iteratively to investigate the trajectory, movement and home range behaviors of simulated bird movements. I have termed these models the simple random walk (SRW), correlated random walk (CRW), and correlated random walk with variable speed (CRWS) after the elements that were adjusted to get closer to the real-world data patterns. The basic structure of each model is the same. A vector outline of Santa Cruz Island in the Galapagos Islands is used to match the location of the seabird data and can be seen in Figure 7 with airplanes representing the seabird agents in mid-simulation. In this case the black area is the ocean and the brown is the island of Santa Cruz, Ecuador.

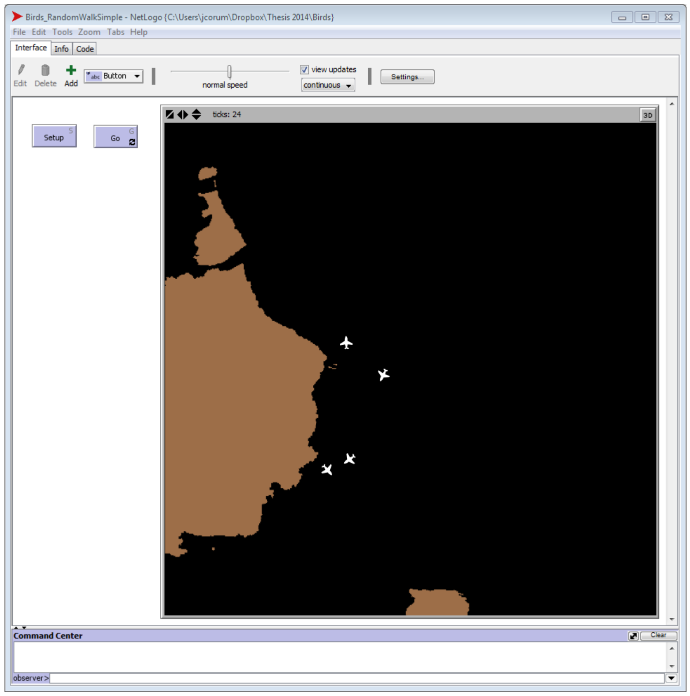

Figure 7 Netlogo modeling environment with vector graphic of Santa Cruz Island

Seabirds return to land only to nest or shelter and spend most of their lives out at sea, thus the agent-based model restricts their movements over land by choosing a random angle that is directed seaward if they end up over the landmass. This behavior is coded into all three models very simply, if the turtle will be over land in one tick, it is directed to pick an angle 180 degrees opposite its current direction, then randomly vary it 90 degrees left or right, enabling the turtle to travel parallel to the coast or go farther out to sea. While this does not prevent a bird from ending up overland entirely, it does allow for some variation in movement when near the coast to approximate seabird behaviors.

CHAPTER 4: ITERATIVE MODEL DEVELOPEMNT AND ANALYSIS

The following sections demonstrate the iterative model development process informed by ecological analysis techniques and presents the results. The simple random walk (SRW) model is the starting point, with the hypothesis that the birds simply have to move through space somehow, ideally in the least complex way possible, avoiding land. The SRW model does not meet the expectations setup by the real-world data and the next step in the iterative process, the correlated random walk (CRW) model is implemented. The final model, the correlated random walk with variable speed (CRWS) model provides the best fit without incorporating behavioral parameters.

4.1 A Starting Point; the Simple Random Walk as a process for bird movement

The simple random walk model is presented as a starting point. This is the simplest way of creating agent movement in Netlogo. Since we need agents to move across the landscape, using the most basic method possible, it was expected the analysis would result in some similarities with the real-world data. Using the SRW, the agent engages these steps:

1. Pick a direction randomly from 360 degrees
2. Check if that will intersect land
    a. If yes, pick a direction 90 degrees plus or minus the angle that is 180 degrees opposite the original direction
3. Move 16 patches in that direction
4. Start over at Step 1

The model will persist with these actions until 96 ticks have been completed, simulating an 8-hour time period of tracking. Other than movement over land, there are no additional restrictions. Model runs where birds reached the edge of the map extent were discarded as this causes an edge effect that is unsuitable for analysis, however if future models need to extend farther from the island, the programming will not need adjustment. The data is exported into a CSV file and the analysis in R can begin.

4.1.1 Shape and the use of space by agents with SRW movement programming

When comparing the SRW output to the seabird data (Figure 8) it is clear through visual inspection that this is not the correct model for movement programming. The simulated agents are tightly clustered and do not exhibit the range or deliberateness found in any of the seabird tracks.

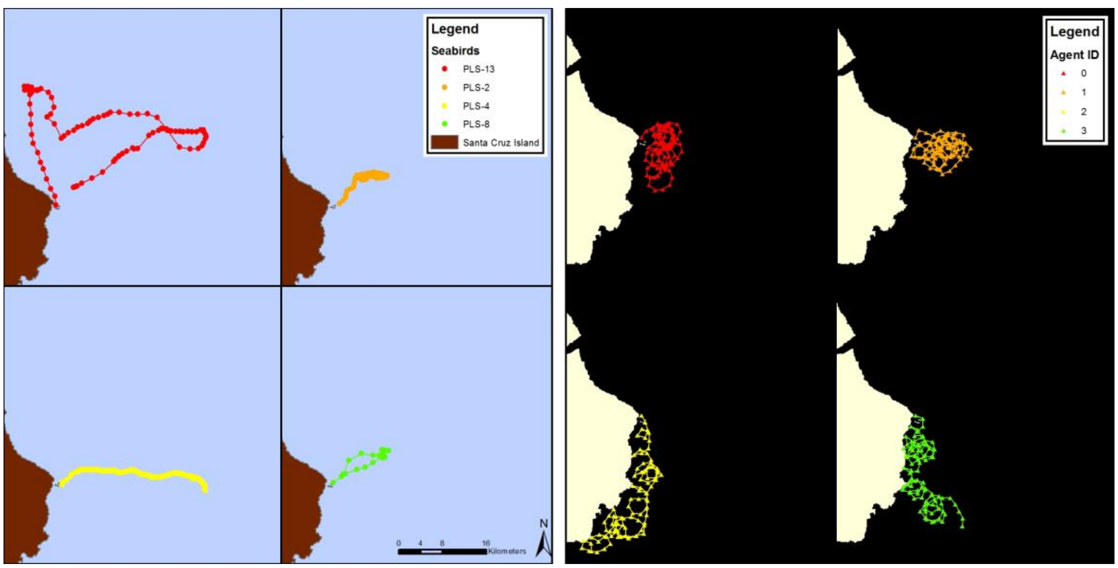

Figure 8 Seabird (left) and SRW Model movement tracks

The movement tracks can be decomposed further by analyzing the spatial components that make up their movement. The 'ltraj' object is used to examine the shape of movements in space and the object itself contains basic characteristics displayed when the object is created which are identical for each model with 96 relocations and 4 simulated birds, referred to as agents. Without device or battery failure there are no missing observations, and the tracks contain the same number of relocations.

Following the steps used for analyzing the seabird data, the SRW data is rediscretized, producing even more tightly clustered areas. In Figure 9, the rediscretization of the SRW paths provides a more continuous path, but the tight clustering and lack of real-world behaviors is apparent.

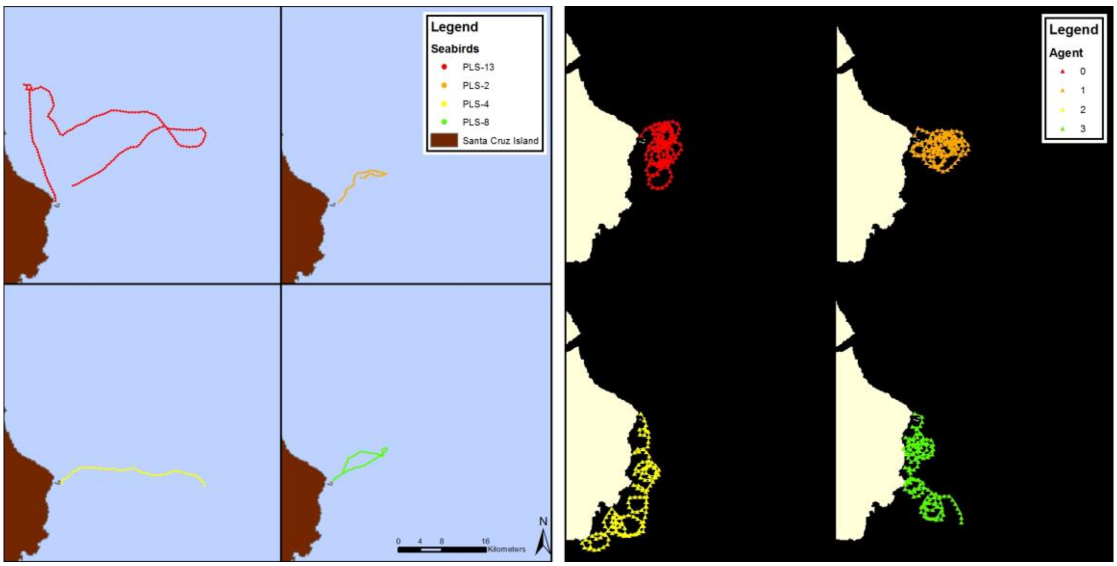

Figure 9 Rediscretized seabird (left) and SRW movement paths

To further decompose the shape of movements into spatial components that can be compared without regard to seabird origin or spatial extent, the cosine values of relative angles changes are used. This method of inspecting the cosine values of relative angle changes also provides an insight into the differences between the seabird data and the simulated movements. To understand these charts it is necessary to know that cosine values near 1 and zero indicate near straight line travel, while values closer to 0.5 indicate what is called a tortuous trajectory, think flying in circles looking for something or a jet fighter pilot avoiding incoming fire. In the SRW output, there is almost no indication of straight line behavior, (Figure 10) instead the cosine values appear to vary widely at every tick indicating movements in completely random directions with no pattern. This implies that the agents are not moving relative to their previous heading, but instead careening around as though in a pinball machine.

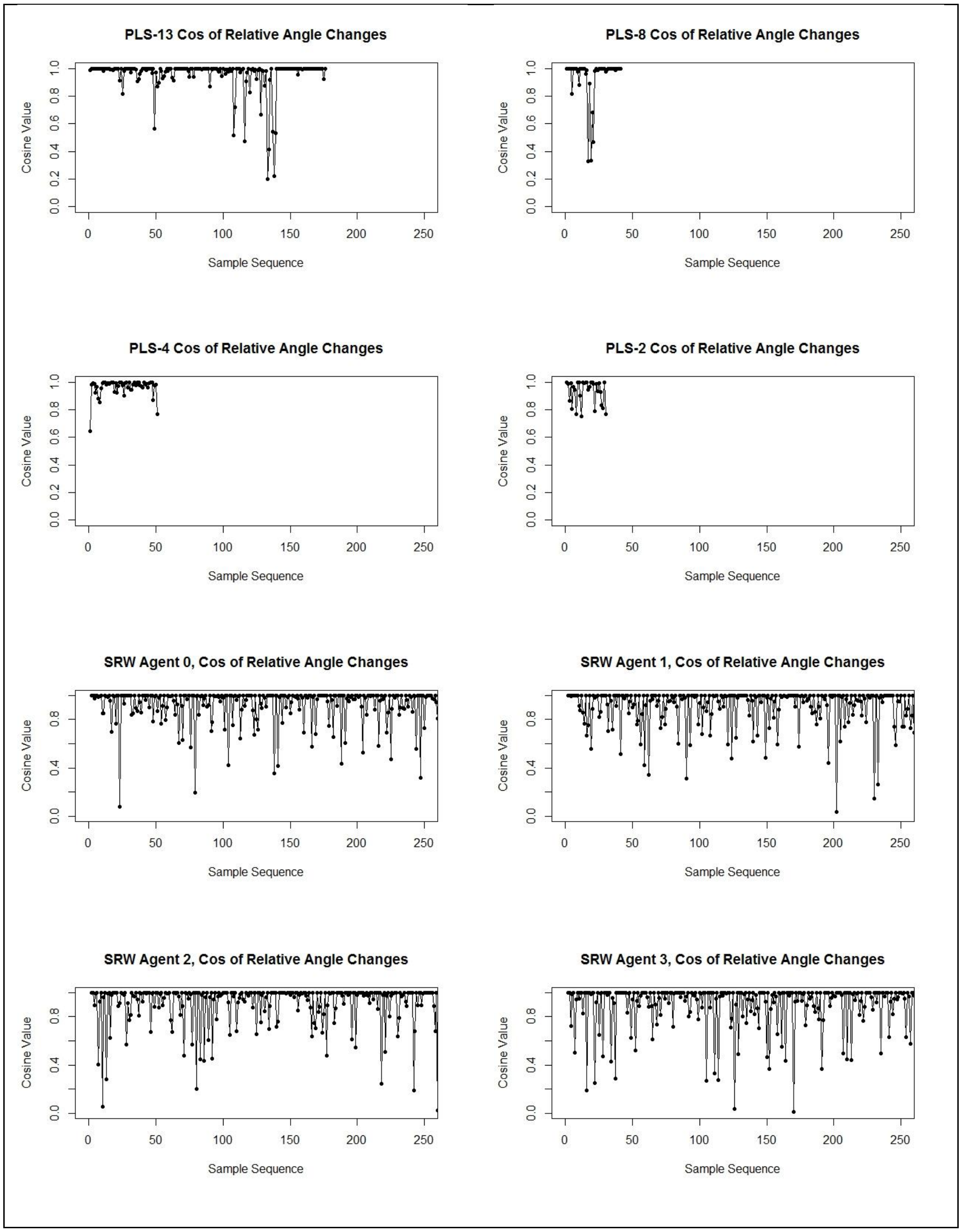

Figure 10 Seabird (left) and SRW smoothed relative angle changes over time

These same relative angle values can be analyzed using R functions that are able to look for patterns in the sequential values. The ACF function tests if the model is selecting truly random numbers or if an underlying pattern is working on the selection of angles. The presence of autocorrelation in these data suggests there is an error in the model and that it is not a random walk. This is true, since we have restricted the agent's movement over land, which forces the system to use a different set of rules and select an angle value from a different distribution. The values above and below the blue lines in Figure 11 indicate autocorrelation that is statistically significant, and that our movement programming needed adjusted for the next iteration of model development.

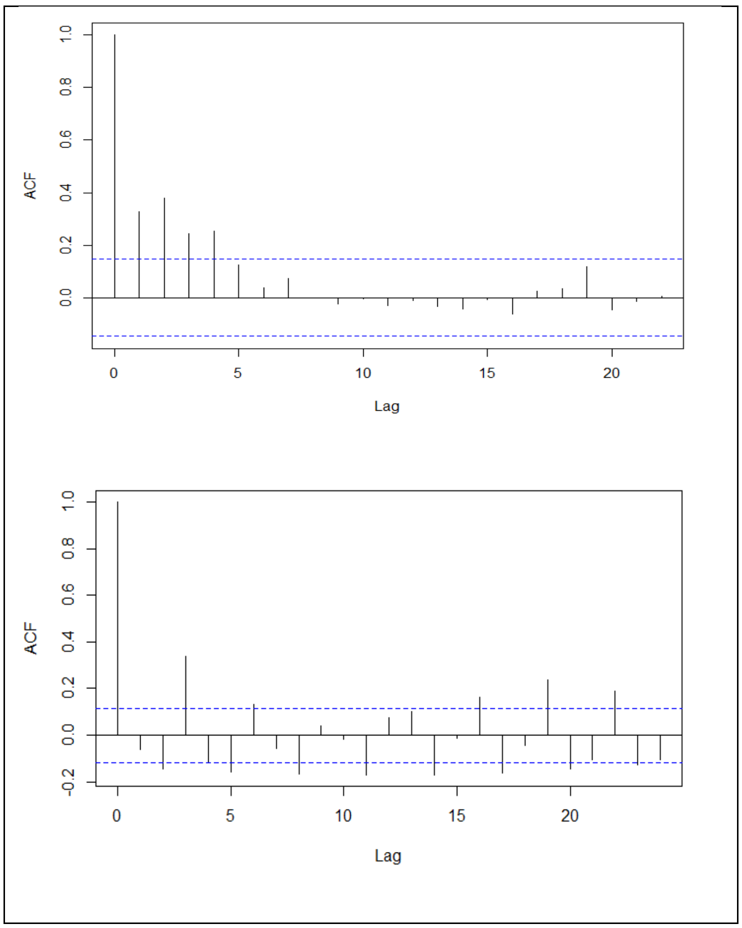

Figure 11 ACF of relative angle changes in seabird (top) and SRW simulation movements

While it would be possible to begin the new iteration here a few remaining tests exists that can help explore and understand the data generated by the ABM and shed light into why a simple random walk is not suitable for modeling moving-object data in the context of animal tracking. The tests above look at the overall shape of the travel paths and begin investigating the underlying values that compose the overall shape. These tests are used to determine the geometric processes acting on the movements, identify searching behaviors and begin testing for independence and sampling error. The tests below will further investigate independence and begin to incorporate the use of space, or home range, into the analysis.

4.1.2 SRW relocation independence, pattern in random values

The WaWo test continues the investigation into the process of movements by examining the distribution of changes in the xy coordinates and distance traveled. WaWo tests examine the sequence of values by looking for values that do not belong and are from a different distribution. Recall that the seabird P-values for dx, dy and dist are very low, meaning that the null hypothesis of the WaWo test holds true, the values are consistent with no unexpected values in the sequence. This test suggests that the movement process of the seabirds has a normal distribution and that they are not moving randomly, but moving with a purpose. While this seems like common sense, obviously birds do not randomly careen about like drunkards, having a mathematical test validate this is useful when examining the simulation data since it allows for the quantitative comparison of the model values with the real-world data and checks the randomness of the movement process in Netlogo.

For the SRW model, the WaWo test results are presented in Table 4. The dx and dy p-values are consistently low enough that the test is confident they are independently drawn from the same distribution. Error here would indicate flaws in the tracking data itself, a bad location fix, or a value in sequence from a different datum. In the simulation, these values are near zero because Netlogo is controlling the coordinate system for all the values. The dist values however are completely different. These values are much higher and the WaWo test indicates that they are not independently drawn from the same distribution. This is true, since the distance traveled is related to the speed of the agent, which we have set to a constant 16 patches per tick.

Table 4 WaWo test results for SRW Model Output

Agent ID	dx	dy	dist
1	1.13E-10	1.84E-10	0.6036293
2	1.43E-09	2.12E-08	0.6631326
3	6.19E-08	2.34E-08	0.45736442
4	5.38E-07	1.53E-07	0.1883668

4.1.3 The use of space: home range

The minimum convex polygon (MCP) that contains 90-95% of all relocations in a set of tracks is a common measure of home range (Kenward, et al 2001, Calenge 2011). The results of the density and linkage estimator ‘clusthr’ are presented in Figure 12. This test identifies a core group of clusters and begins incorporating other clusters into the group until 100% of the relocation points are included. It is often used prior to creating an MCP to identify the percentage of points to include that excludes exploration behaviors. In the seabird data from around 50-75% there is very steady increase in home range area as the percentage of points included increases and demonstrates an exponential increase in area. The simulation results are more parabolic, with increases beginning at or around 50-60% of the home-range level and increasing quickly. This suggests that the SRW model is not utilizing the space in the same way as the seabirds, which is supported by the analysis of shape and space above.

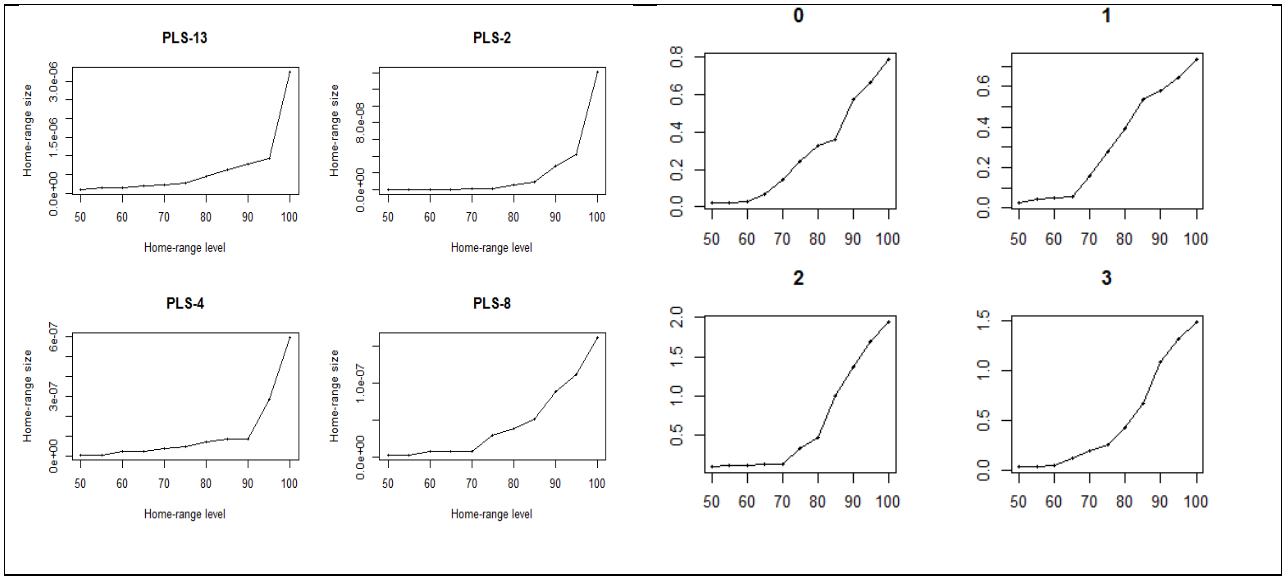

Figure 12 Seabird (left) and SRW home range size from 'clusthr'

In summary, the SRW produces agent movement paths that are tightly clustered, tortuous, and autocorrelated due to programming restrictions. These evidently do not utilize space in the same way as the seabirds. Overall the SRW model clearly does not produce movements accurate enough to model seabird behaviors.

The next stage in the model development was to test a correlated random walk. The correlated random walk is more complex since it changes the process by which the agents pick a direction to move in, making it less random and more directional, it was hypothesized that it would prevent the tightly clustered movement paths and change the overall use of space by the simulation.

4.2 Incorporating a correlated random walk to modify spatial behavior

Adding the correlated random walk changes the behavior of agents at each tick. In this model the agents are still looking out for the landmass, and avoiding it, but the angle changes they are allowed make are restricted. The decision was made to allow the agents to pick a direction that is no more than 45 degrees off their previous heading. This means the agent will not be able to move back toward its origin point without making a more sweeping turn to do so and was expected to bring the model output closer to that of the seabirds since it is unlikely they are making 180 degree turns often.

The correlated random walk model is presented as the second iteration in the basic movement model. When this model is initialized, the agents are given a random heading chosen from the full 360 degrees available for them to move in. After initializing the model, the agents in the correlated random walk model follow these steps:

1. Pick a direction plus or minus 45 degrees off your current heading
2. Check if that will intersect land
    a. If yes, pick a direction 90 degrees plus or minus the angle that is 180 degrees opposite the original direction
    b. If no, continue
3. Move 16 patches in that direction
4. Start over at Step 1

4.2.1 Shape and the use of space by agents with CRW movement programming

In Figure 13 it is apparent that the correlated random walk does bring us closer to the movement model of the seabirds. The agent paths have become much less clustered and tortuous in comparison to the SRW model output. Rediscretizing the movement steps, shown in Figure 14, prior to decomposing the spatial components of the movement tracks, reveals that there are still more clusters in the CRW tracks than are present in the seabird tracks.

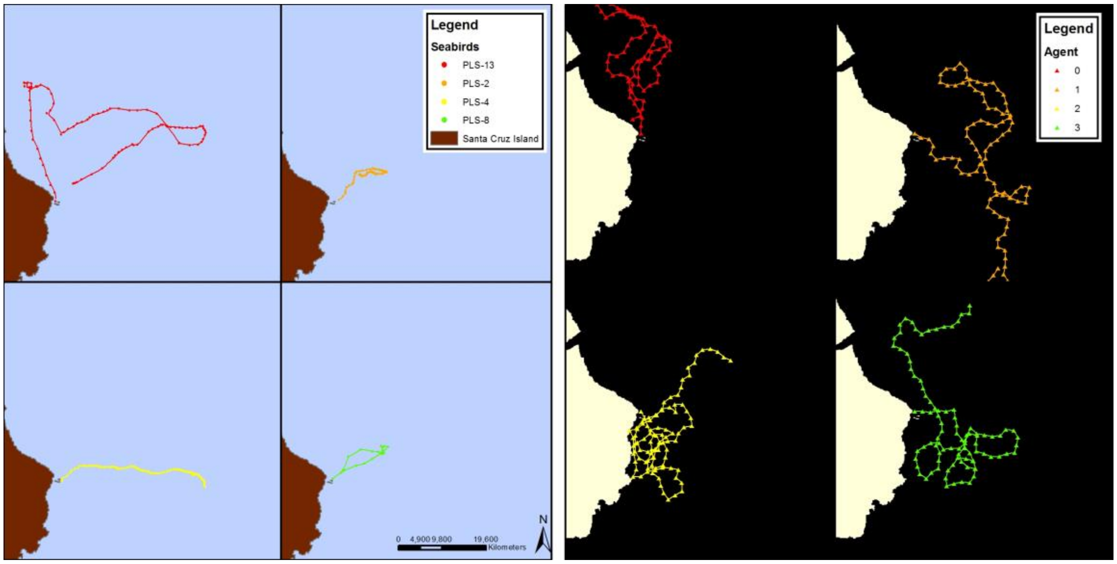

Figure 13 Seabird (left) and CRW movement tracks

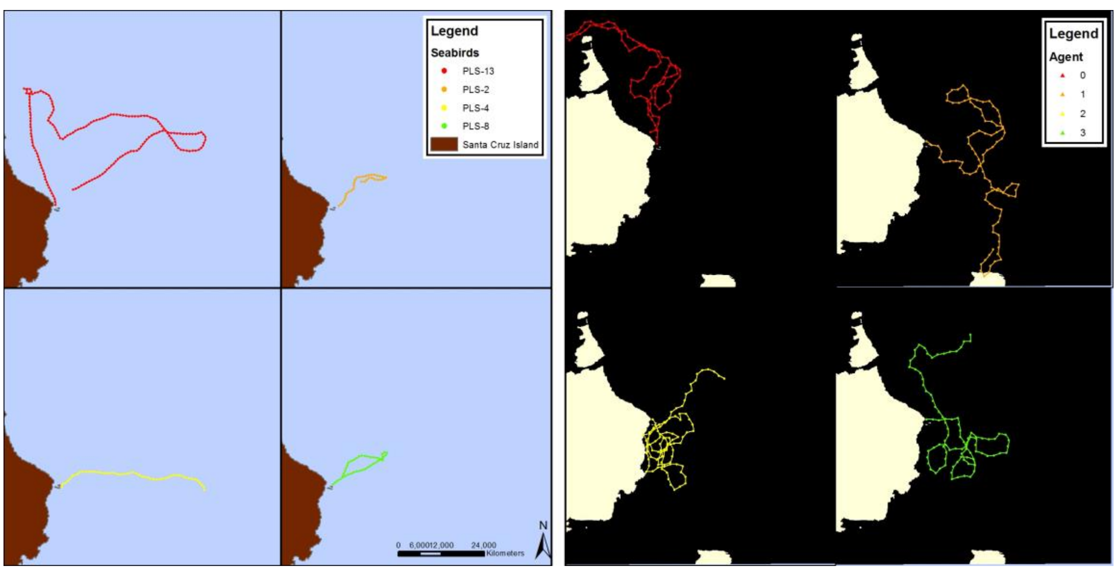

Figure 14 Rediscretized seabird (left) and CRW movement paths

The cosine values are less torturous than the SRW model produces, with some of the wild oscillations seen in the previous model reduced somewhat. While these values still vary a great deal compared to the seabird data, they are showing a reduction in tortuous behaviors overall, with cosine values mostly in the 0.8-0.95 range, indicating relatively straight flight paths with occasionally sharp heading changes.

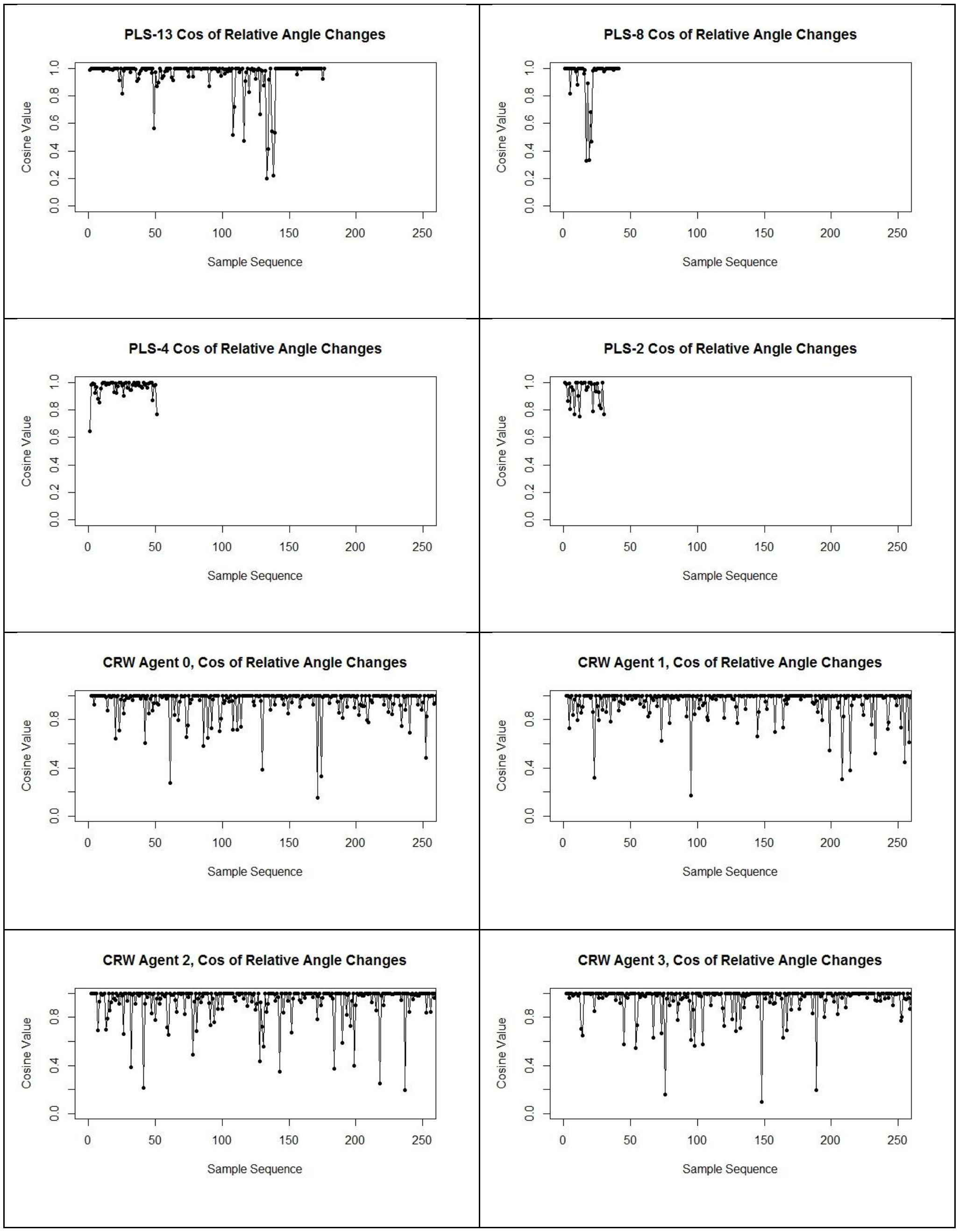

Figure 15 Seabird (left) and CRW smoothed relative angle changes over time

The ACF function for the CRW in Figure 16 reveal autocorrelation very similar to that of the SRW model, with regular spikes of statistically significant autocorrelation. The spikes are nearly identical to that of the SRW. Solla et al (1999) suggest that ecological relationships are often related to the spatial environment and it is common to observe autocorrelation in real-world data, however it is unlikely that it occurs with the regularity observed in the SRW and CRW models.

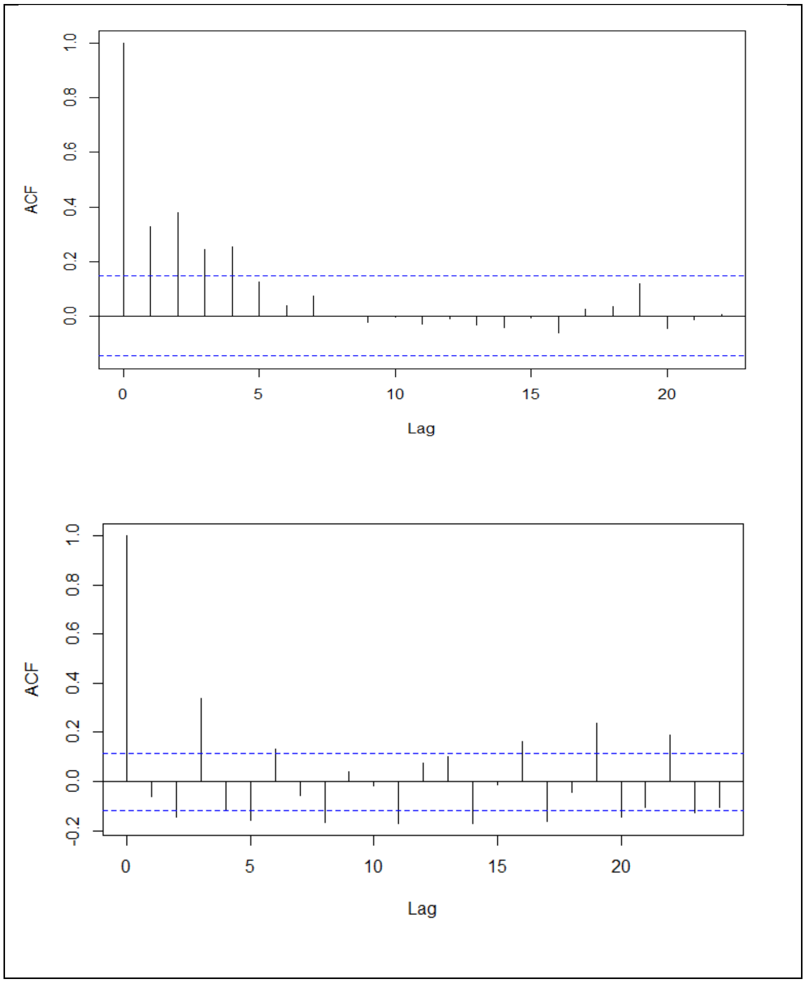

Figure 16 ACF results of relative angle changes in seabird (top) and CRW simulation movements

4.2.2 CRW relocation independence, pattern in random values

The results of the WaWo test on the CRW output are in Table 5. Similar to the SRW the dx and dy p-values remain at zero, but the dist p-value is more interesting. One of the agents achieved a value near zero, however the remaining agents all fail the test of independence. This is interesting because the speed is still set to a constant 16 patches per tick.

Table 5 WaWo test results for CRW Model Output

Agent ID	dx	dy	dist
1	1.82E-14	3.34E-12	2.44E-15
2	1.36E-12	1.16E-13	0.6945514
3	3.53E-09	1.37E-14	0.9069382
4	1.94E-14	8.71E-13	0.9889598

4.2.3 CRW home range analysis

The results of the density and linkage estimator on CRW results are presented in Figure 17. The simulation results are almost linear, with increases beginning at or around 50-60% of the home-range level and increasing at nearly the same rate. This suggests that the CRW model is still not using space in the same manner as the seabirds.

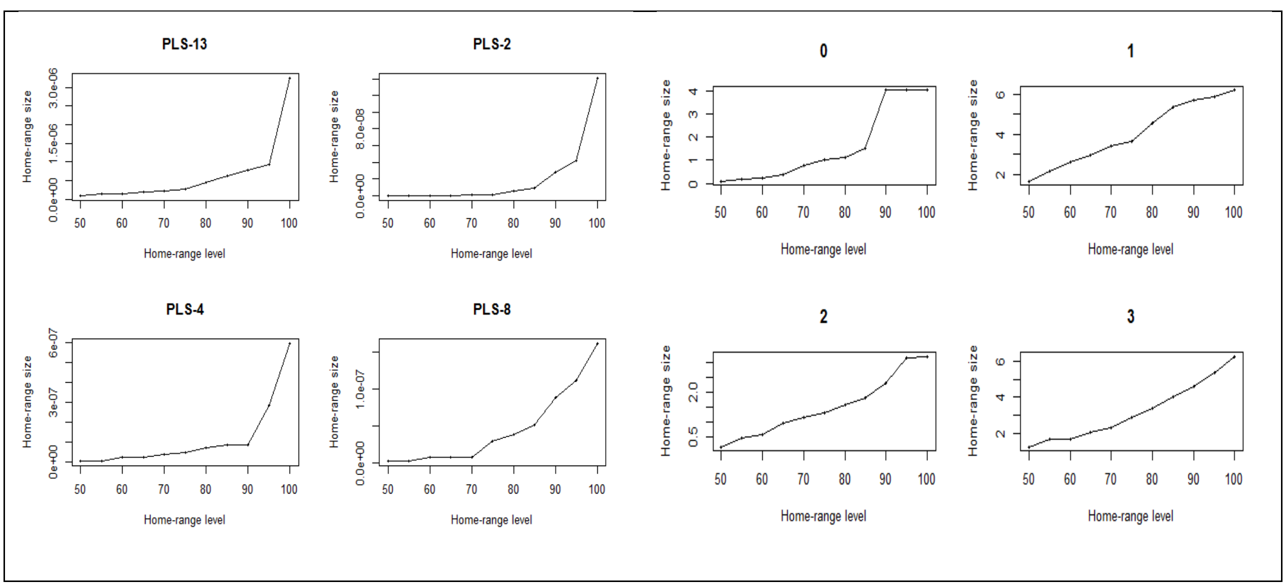

Figure 17 Seabird (left) and CRW home range size from 'clusthr'

To summarize, the movement tracks themselves are closer to that of the seabird data than the SRW model results. There is less clustering, though some still exists in random places. The smoothed relative angle changes are beginning to show less wild oscillations, however something is still causing the WaWo test to indicate that the changes in distance between each relocation are not pulled from the same distribution. Home range levels increase linearly and still do not match the expected gradual increase to 75-80% found in the real world data.

4.3 Incorporating speed variation into the correlated random walk

When looking at the original map of relocations in Figure 8and Figure 13 the ABM output consistently delivered points at regular intervals in both space and time. The seabird data would have many points separated by larger distances and small clusters of points close together; indicating that speed played an important role in the overall shape of the seabird movements when decomposed. Adding the variable speed to the correlated random walk again changes the behavior of agents at each tick. In this model the agents are still looking out for the landmass, and avoiding it, the angle changes they are allowed make are restricted and now their speed will vary with the same mean and standard deviation observed in the aggregated seabird data set.

The CRWS model is presented as the third iteration in the basic movement model. When this model is initialized, the agents are given a random heading chosen from the full 360 degrees available for them to move in. After initializing the model, the agents in the correlated random walk model follow these steps:

1. Pick a direction plus or minus 45 degrees off your current heading
2. Generate a random speed from a normal distribution centered around the mean of 16, with a standard deviation of 16.1, discarding negative values
3. Check if that speed will cause the agent to intersect land

a. If yes, pick a direction 90 degrees plus or minus the angle that is 180 degrees opposite the original direction

b. If no, continue

4. Move at the random speed in that direction

5. Start over at Step 1

4.3.1 Shape and the use of space by agents with CRWS movement programming

In Figure 18 it is apparent that the CRWS model brings us even closer to the movement model of the seabirds. The agent paths have become much less clustered and tortuous in comparison to the SRW model output and the varied speed as made clusters of relocations matching that of the seabirds that should be reflected in the investigations below. Rediscretizing the movement steps, shown in Figure 19, prior to decomposing the spatial components of the movement tracks reveals that there are significantly less clusters than the SRW model and it was expected that the results would match the seabird data more closely.

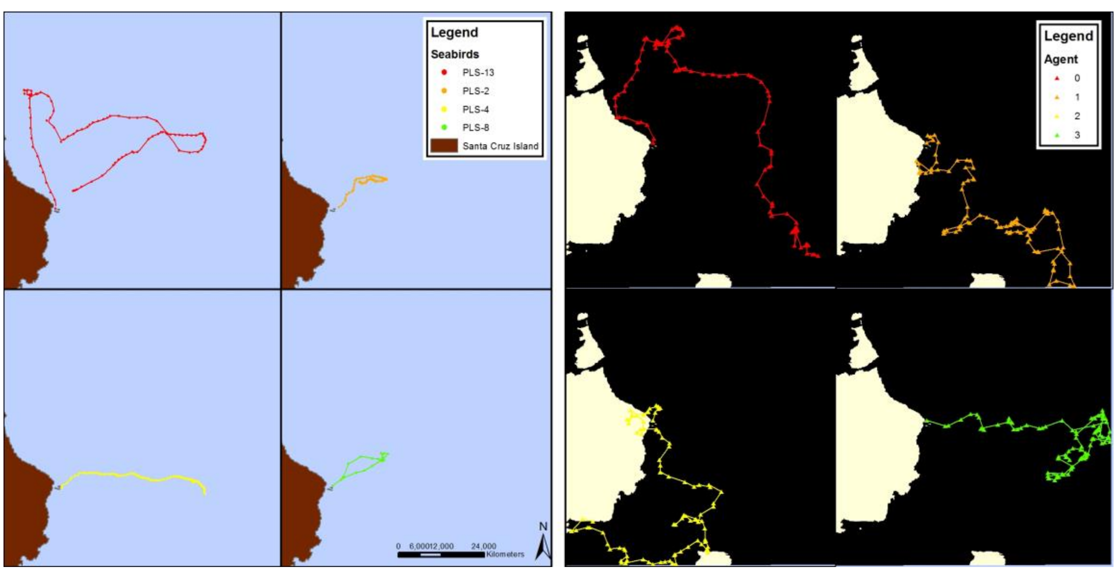

Figure 18 Seabird (left) and CRWS movement tracks

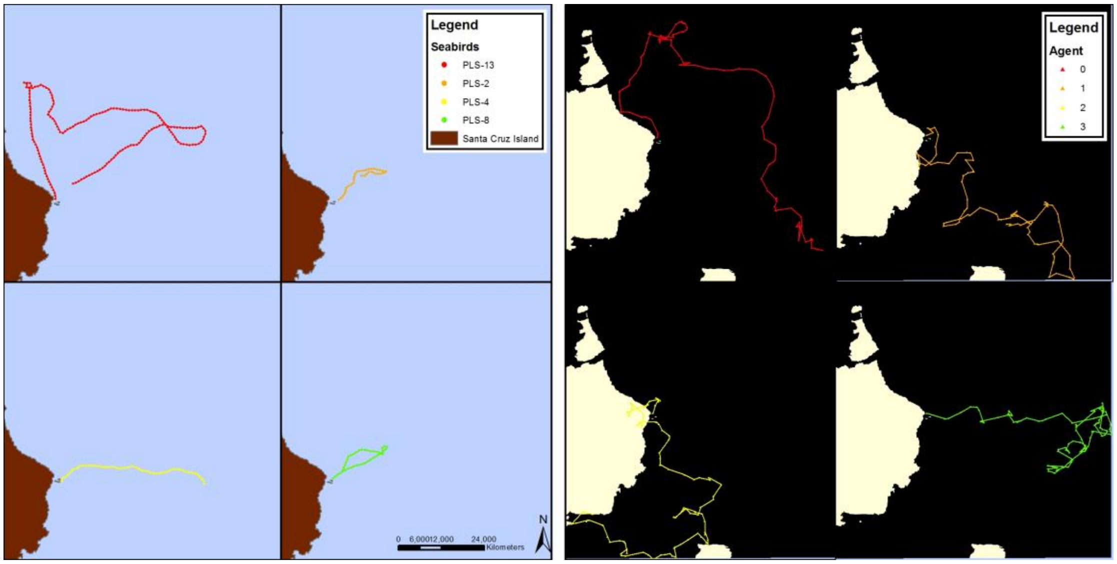

Figure 19 Rediscretized seabird (left) and CRWS movement paths

The cosine values are significantly less torturous than those of the SRW and CRW models, with oscillations that more closely resemble that of the seabird data. Also of note is a more similar range of values to the seabird data. There are still some discrepancies and the values still appear to be influenced by the random number generator control the angle changes. It is possible that GPS accuracy issues may play a role in cloaking the more wild variations observed in the CRWS movements. If the seabird is tightly circling in an area less than 30m, searching for food, the GPS error may not capture these tortuous angle changes. At this point, I am confident that this test is better applied at a point in the iterative process where foraging behaviors can be integrated.

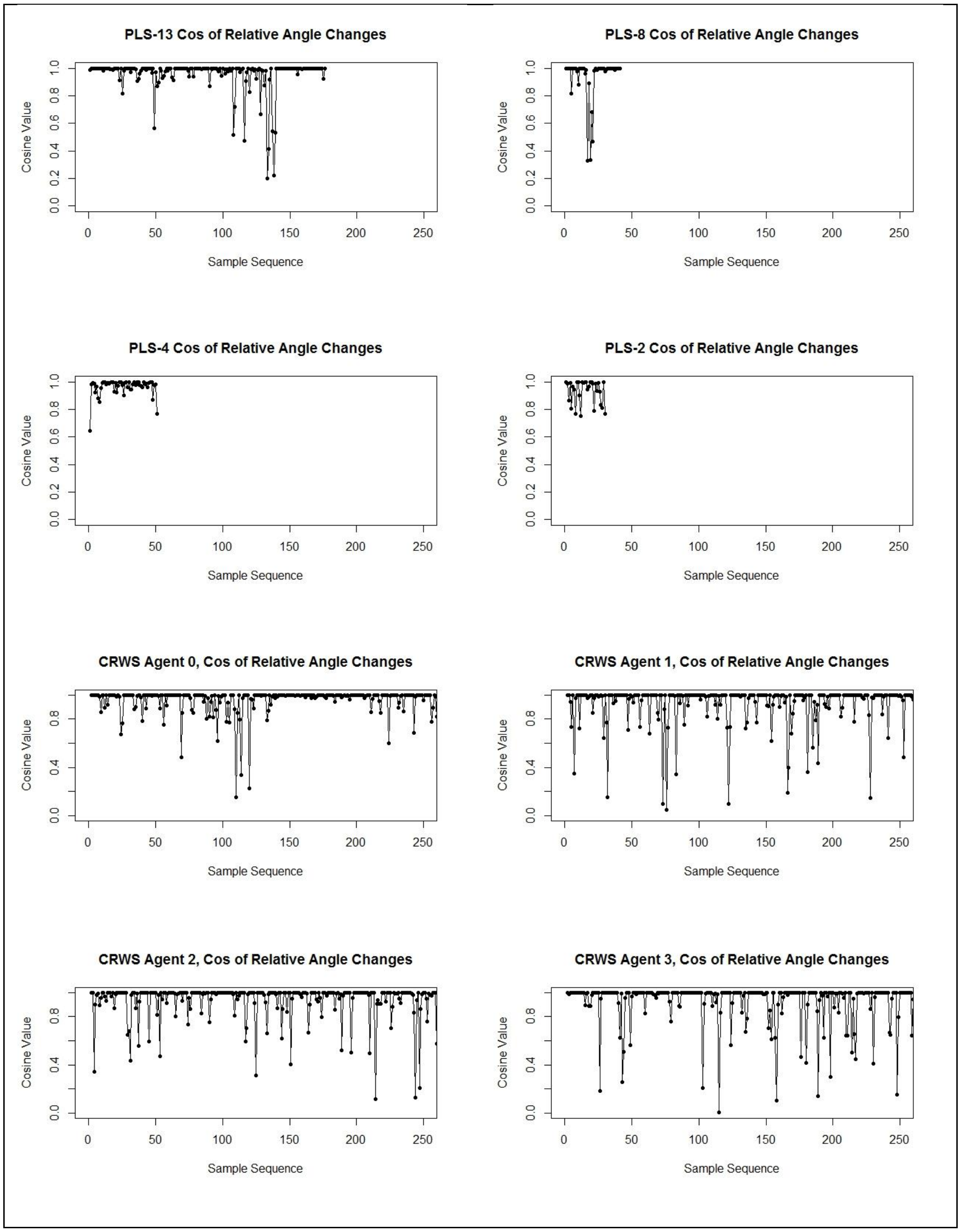

Figure 20 Seabird (left) and CRWS smoothed relative angle changes over time

The ACF function for the CRWS agents in Figure 21 reveal autocorrelation very similar to that of the seabirds, without the regular spikes of autocorrelation found in the CRW and SRW agent paths. Since the only change in the model was varied speed this suggests that speed plays a significant role in the autocorrelation of relative angle changes.

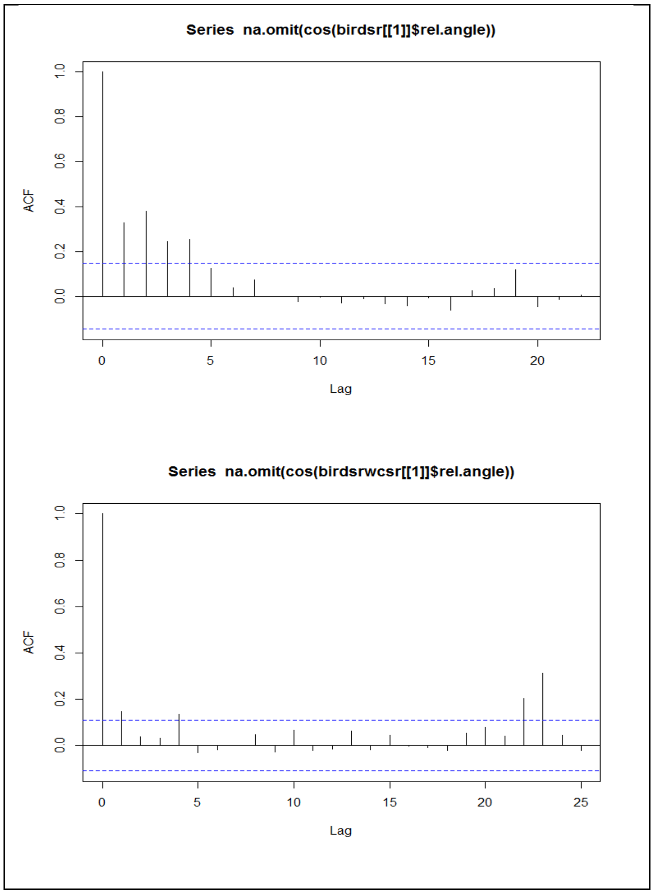

Figure 21 ACF results of relative angle changes in seabirds (top) and CRWS agents

4.3.2 CRWS relocation independence, pattern in less random values

The results of the WaWo test run on the CRWS output are in Table 6. Even with the third model iteration using CRWS agents, the paths are still not a statistical match to the real-world data. Similar to the SRW and CRW agent paths, the dx and dy p-values remain near zero. In this case the dist values become an important indicator. These values are consistently lower than those of the previous models, closer to the seabird values, which suggests that adding a behavioral element to the model, varied speed, made a bigger difference than changing the geometric processes that govern movement. Nevertheless, the CRWS agents are still rejecting WaWo's null hypothesis. This is interesting because the speed value is programmed to come from the same distribution.

Table 6 WaWo test results for CRWS model output

Agent ID	dx	dy	dist
1	0.000476589	0.000592477	0.4073415
2	0.000231885	0.000320847	0.47143726
3	8.91121E-05	0.01447694	0.7902441
4	0.006280511	0.02280497	0.580764

4.3.3 CRWS home range analysis

The results of the density and linkage estimator on CRWS results are presented in Figure 22. The simulation results are very similar to that of the seabird data, with increases beginning at or around 75-80% of the home-range level and increasing rapidly after. This suggests that the CRWS model agents are forming a similar core of movements that the density and linkage estimators are identifying as similar to the seabird movements.

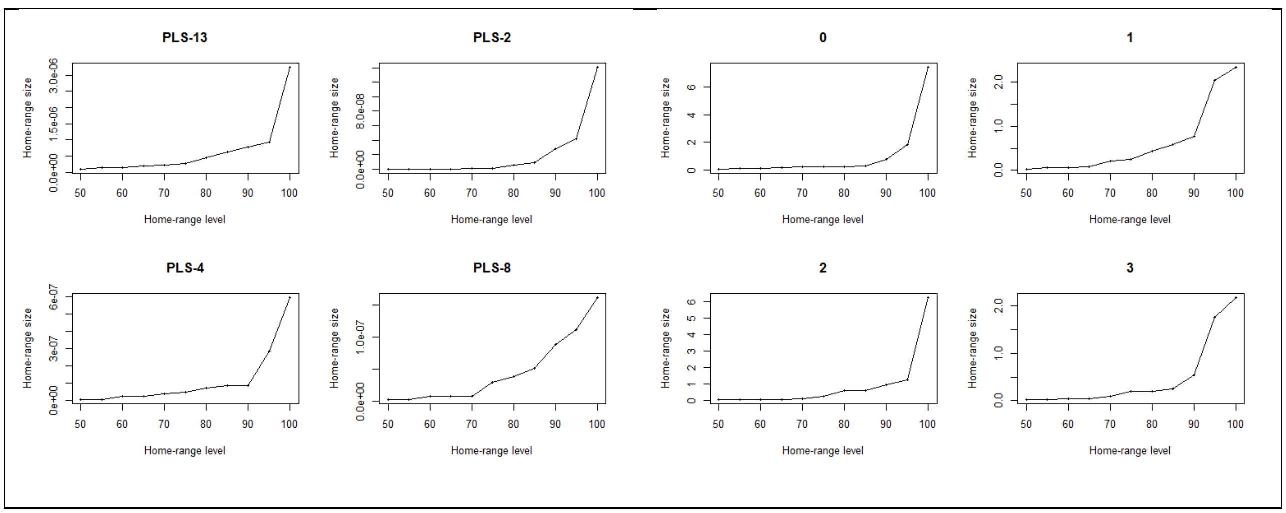

Figure 22 Seabird (left) and CRWS home range size from 'clusthr'

In summary, the movement tracks themselves are closer to that of the seabird data than the any of the previous model agents. There is significantly less clustering in the overall movements and the smoothed relative angle changes show far less wild oscillations. However, something is still causing the WaWo test to indicate that the changes in distance between each relocation are not pulled from the same distribution. Calenge (2011) and the CRWS results suggest that speed is a very significant factor. Home range levels increase linearly and still do not match the expected gradual increase to 75-80% found in the real world data.

Additional testing was done to investigate the programmatic errors that may arise from Netlogo itself. An example of testing the random number generator can be found in Appendix 7. This appendix shows how the speed distribution of the simulated data differs from that of the real-world data.

CHAPTER 5: DISCUSSION

Shape of the sequential relocations in space, independence, and home range are all considered as model outcome key performance indicators. The indicators provide a common language for comparing ecological data with agent-based modeling output. In general, simulations are simplified versions of the real world, implying that some variation between observations and simulations is expected. The problem then becomes selecting a best-fit simulation model that emulates a selection of processes found to be critical in real world data.

5.1 Spatial elements

The overall shape of real-world animal's movement trajectory is likely impossible to predict since an animal's movement track will be unique in an ever changing environment. Variables such as weather and food availability vary a great deal and prevent specific predictions of movement. The mechanisms behind movement can be identified and programmed to produce similar overall behaviors to the real world data, with some caveats. The CRWS model produced results, shown by both the trajectory analysis and home range clustering (Kareive and Shigesada 1983) that mimic the same characteristics as the seabird data. The overall changes in relative angle and percent home range estimates most closely match the output of the CRWS Netlogo model. Other models fail at producing acceptable results on all counts, either displaying no patterns matching the real-world data, as in the case of the simple random walk data or by revealing a completely new pattern. One step in the analysis, that of deconstructing the cosine values of the relative angles and performing a time series analysis is likely better suited at a later point in the iterative development of the model.

The cosine values of relative angles in the real world data reveal signs of tortuous trajectories or searching behaviors (Benhamou 2004). This implies that in order for the agent-based model to accurately mimic the processes that create such trajectories, a great deal of time must be spent informing the model about bird predatory habits and behaviors, as well as incorporating memory and inter-agent communication. Restricting the turn angle to a range around the current direction the simulated agent is facing provides movement paths that more closely mimic that of the real world relocation patterns, which becomes quickly evident when compared, even with modifications to the models to avert seabird movement over land. The smoothed cosine values then inform an iterative step in the model development relating to resource harvesting behavior rather than agent movement. This enables the structure of an ABM to more closely match the processes that develop aggregate behaviors in the real world.

Agent movement, largely controlled by very simple programming, is put to the test using a sequential analysis test that looks for values that are unexpected. Wald and Wolfowitz test of independence (WaWo) results for dx and dy are consistent with the seabird data. The distance variable however seems very dependent on speed, as the results of the CRWS model indicate

WaWo test p-values in dx and dy remain consistently low enough to accept the null hypothesis that the values are independent and normal, dist values however, are not independent with p-values as high as 0.98 in the SRW model. Looking at the structure of the model itself, this is related to the speed value chosen during the analysis steps on the original real-world moving-object data. The gull data was used to find the mean movement speed of the Galapagos Swallow Tail Gull during the model movement, as well as the standard deviation of that speed. In the simple random walk and correlated random walk models the speed was set to that average, while for the correlated walk with variable speed the model was instructed to choose a speed at random during each step from a normal distribution created with the mean and average of the real-world data. The values for the CRWS model are the closest to the real-world data, indicating that variable speed is an important element.

5.2 Conclusions

Several tests with quantifiable or visual comparisons enable a model builder to assess the movement component of an agent-based model. Nearly any model in which independent agents move and have expected behaviors in the real world such as movement restrictions (the birds in the model presented here rarely fly over land), would benefit from quantitative analysis to ensure the model behavior is consistent with real world observations, the goal of Pattern Oriented Modeling.

As demonstrated here, each small step in the iterative model building approach, referred to by Railsbeck and Grimm (2012) as Pattern Oriented Modeling, can take time and careful research. If patterns are identified in the target structure, i.e., the real-world moving-object data, they should be used as filters for the model not only at the end of the programming and development process, but during and before. This is especially so if moving-object data is available for comparison.

Continuing this iterative process through to the development of a complete comprehensive model of the bird behavior would take a considerable amount of time. Incorporating feeding and reproductive behaviors could take up to a year to complete if an appropriate multidisciplinary team could be polled for parameter data. Parameters such as energy consumption, fatigue and resource degradation could be tested in the complete model. Nevertheless, such a complete model would be useful for many reasons. In a real-world system, collapsing a component of the ecosystem, such as a fishery, to see how an animal responds would be impossible to justify. With an ABM, it is possible to simulate the collapse of a resource, or any change in the environment, and observe the patterned outcome. This would allow the ABM to become a valuable tool for policy makers which could be used to inform their decisions on environmental management and business policies as well as to help direct conservation efforts.

5.3 Lessons learned

This study demonstrates the need for iterative model development and testing throughout the programming phase. If movement rules had been chosen that are not a best fit for the real-world object modeling it may not appear until sensitivity analysis is performed several steps further into the model building effort, meaning a return to the foundational programming of a model. It suggests that model complexity must be carefully balanced against the needs of the researcher. Complex models that are not validated may provide data that is unusable, something that can only be discovered after hours of investigation.

5.4 Questions for the future

As more complex agent-based models emerge from the iterative process it becomes necessary to investigate them thoroughly. Initial study designs incorporated models whose complexity was not easily understood and could not produce consistent results. Further research is needed into identifying when a model's complexity reaches a point where it no longer is able to provide useful information and speak meaningfully about the system it is modeling.

Additionally, other methods could be incorporated into the iterative validation process that is specific to a scientific discipline interested in agent-based modeling. It may also be possible to use sets of data that do not include tracked moving-objects but rather single-event observations, such as marine mammal or seabird observations from on board a survey ship. The use of such data would restrict the tools available within the ecological framework but may still provide important insights into the results of an agent-based model.

REFERENCES

Bence, J. 1995. Analysis of short time series: Correcting for autocorrelation. *Ecology*: 628-39.

Benhamou, S. 2004. How to reliably estimate the tortuosity of an animal's path:: Straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology* 229 (2): 209-20.

Gardner, M. 1970. Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American* 223 (4): 120-3.

Grimm, V, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. Heinz, and G. Huse. 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* 198 (1): 115-26.

Janssen, M. and E. Ostrom. 2006. Empirically based, agent-based models. *Ecology and Society* 11 (2): 37.

Lichstein, J., T. Simons, S. Shriner, and K. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72 (3): 445-63.

Macal, C., and M. North. 2007. Agent-based modeling and simulation: Desktop ABMS. Paper presented at Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come, .

Neumann, J. and A. Burks. 1966. Theory of self-reproducing automata.

O'Sullivan, D, and G. Perry. 2013. *Spatial simulation: Exploring pattern and process*John Wiley & Sons.

Parker, D., and V. Meretsky. 2004. Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems & Environment* 101 (2): 233-50.

Railsback, S., and V. Grimm. 2011. *Agent-based and individual-based modeling: A practical introduction* Princeton University Press.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Šalamon, T. 2011. *Design of agent-based models: Developing computer simulations for a better understanding of social processes*Tomáš Bruckner.

Solla, D., R. Shane, R. Bonduriansky, and R. Brooks. 1999. Eliminating autocorrelation reduces biological relevance of home range estimates. *Journal of Animal Ecology* 68 (2): 221-34.

Topping, C., T. Høye, and C. Olesen. 2010. Opening the black box—Development, testing and documentation of a mechanistically rich agent-based model. *Ecological Modelling* 221 (2): 245-55.

Valbuena, D., P. Verburg, A. Bregt, and A. Ligtenberg. 2010. An agent-based approach to model land-use change at a regional scale. *Landscape Ecology* 25 (2): 185-99.

Venables, W., and B. Ripley. 2002. *Modern applied statistics with S*Springer.

Wikelski, M., and Kays, R. 2014. Movebank: archive, analysis and sharing of animal movement data. World Wide Web electronic publication. http://www.movebank.org (last accessed on 5/13/2014).

APPENDICES

APPENDIX 1: TRAJECTORIES

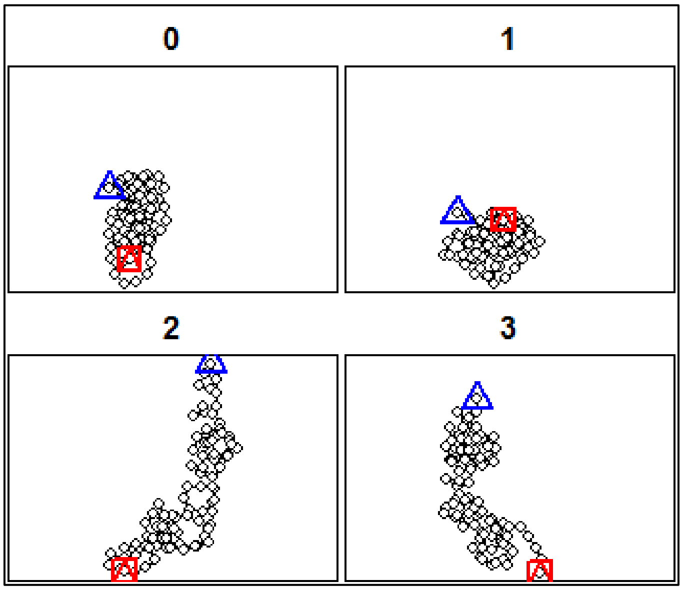

Figure 23 Movement of Netlogo gulls with a simple random walk

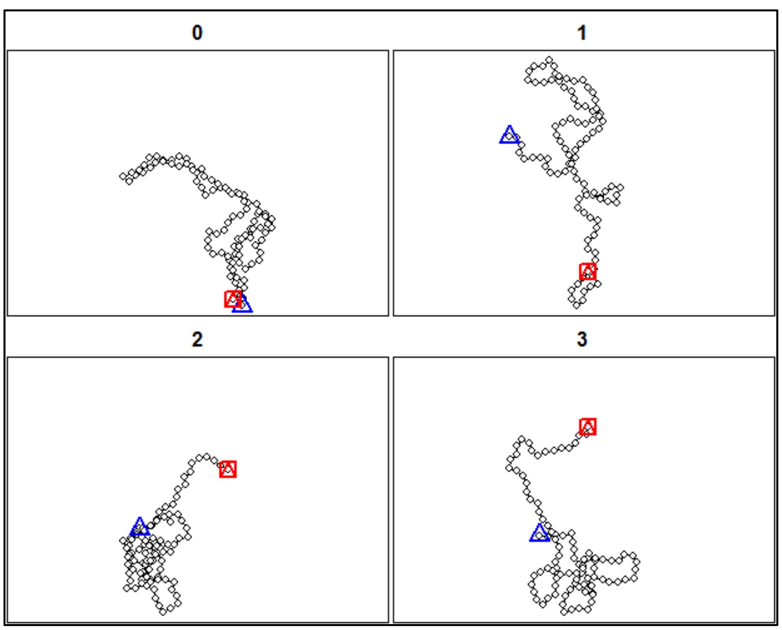

Figure 24 Movement of Netlogo gulls with a correlated random walk

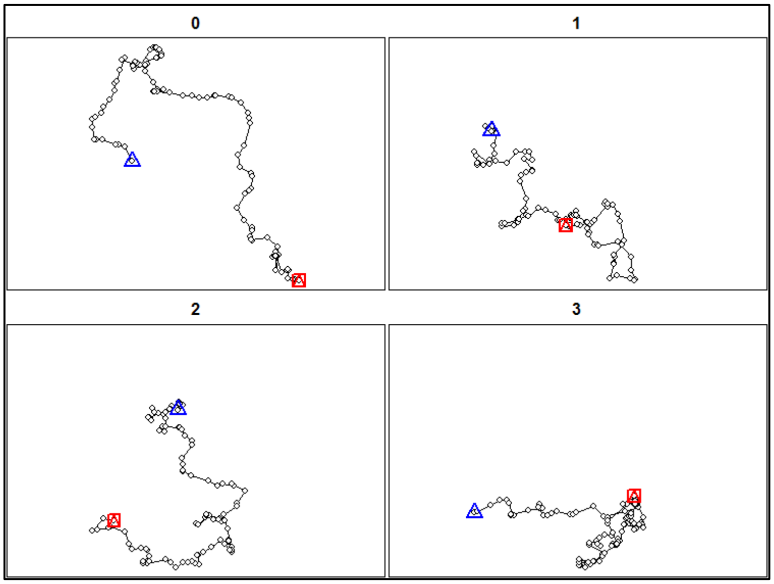

Figure 25 Movement of Netlgo gulls with a correlated random walk and variable speed

APPENDIX 2: REDISCRETIZED TRAJECTORIES

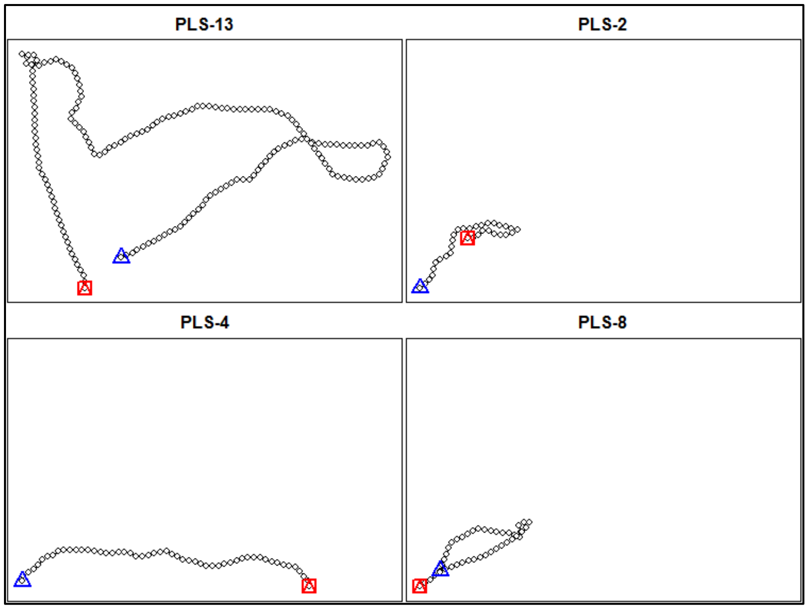

Figure 26 Movebank.org Gull relocations

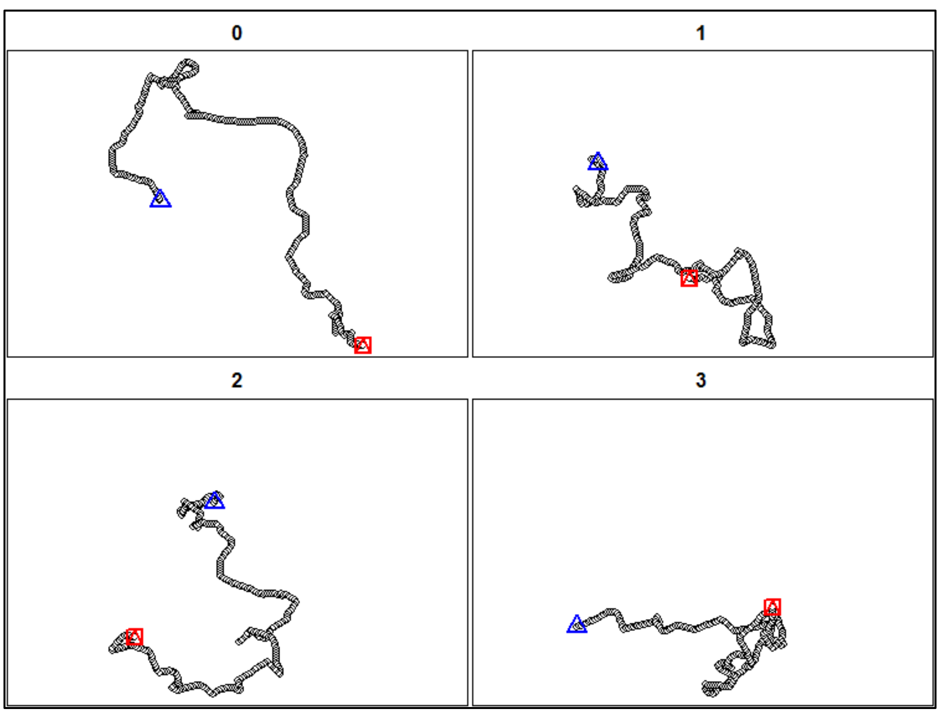

Figure 27 Correlated random walk with variable speed

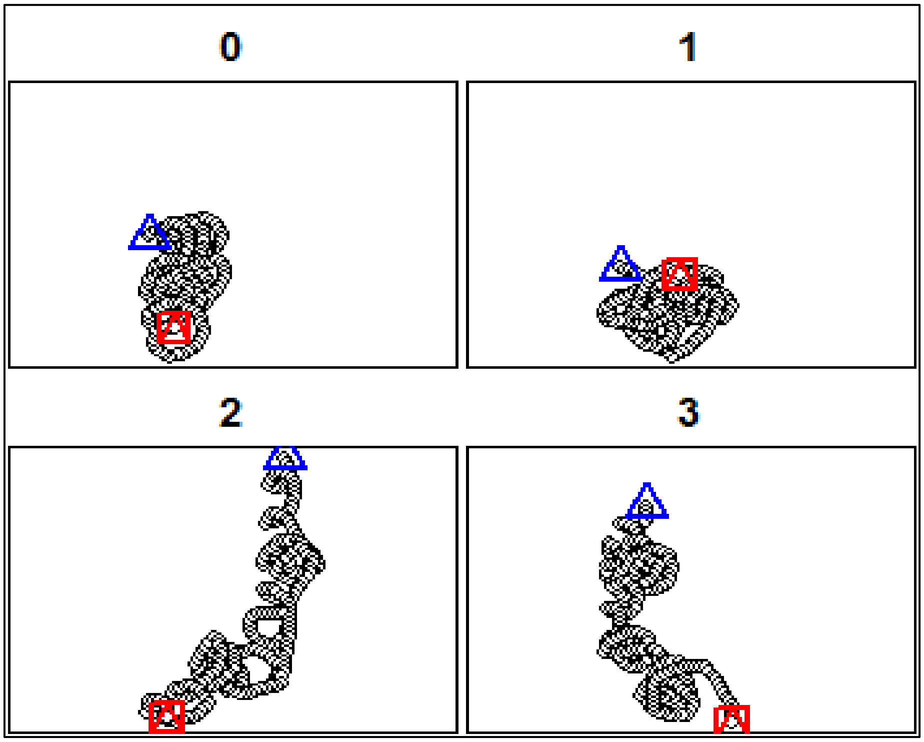

Figure 28 Simple Random Walk

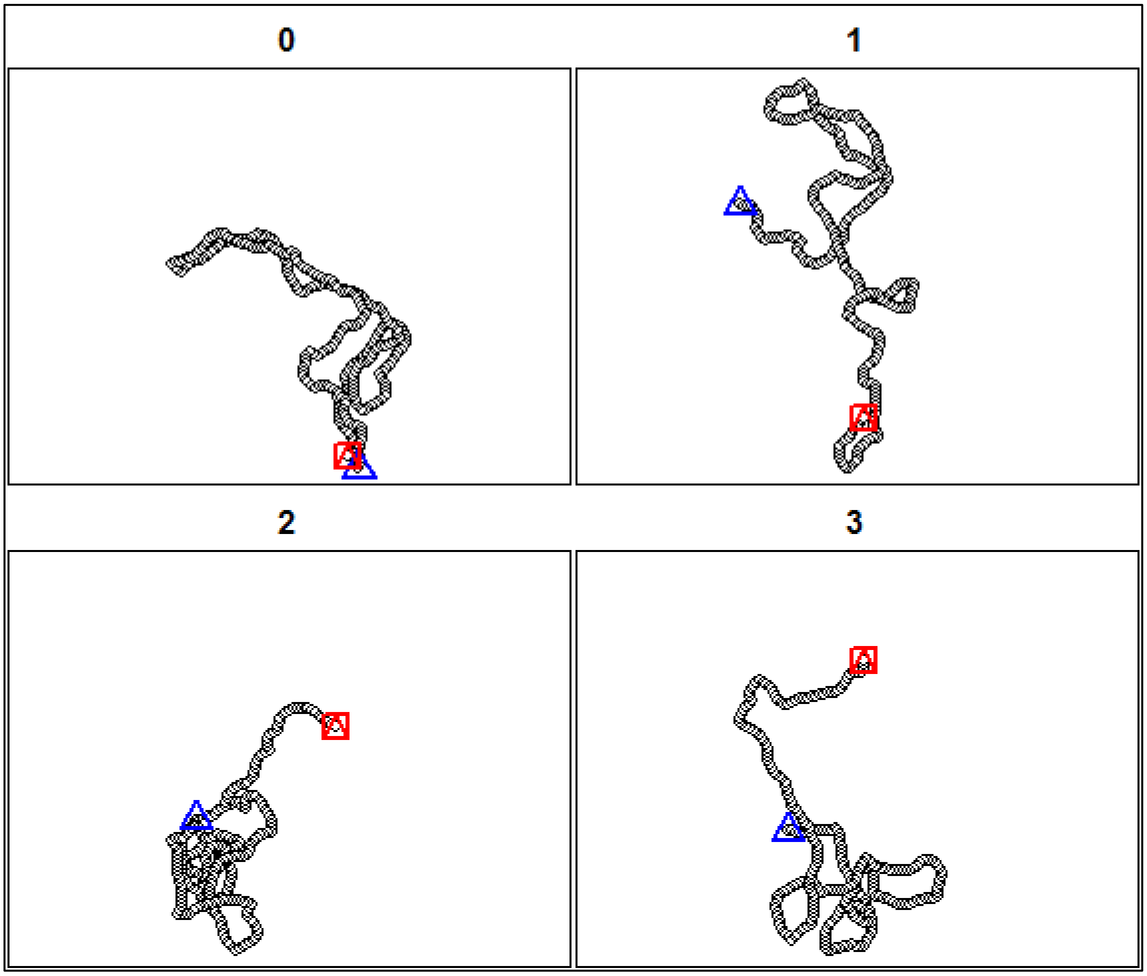

Figure 29 Correlated Random Walk

APPENDIX 3 SMOOTHED COSINE VALUES

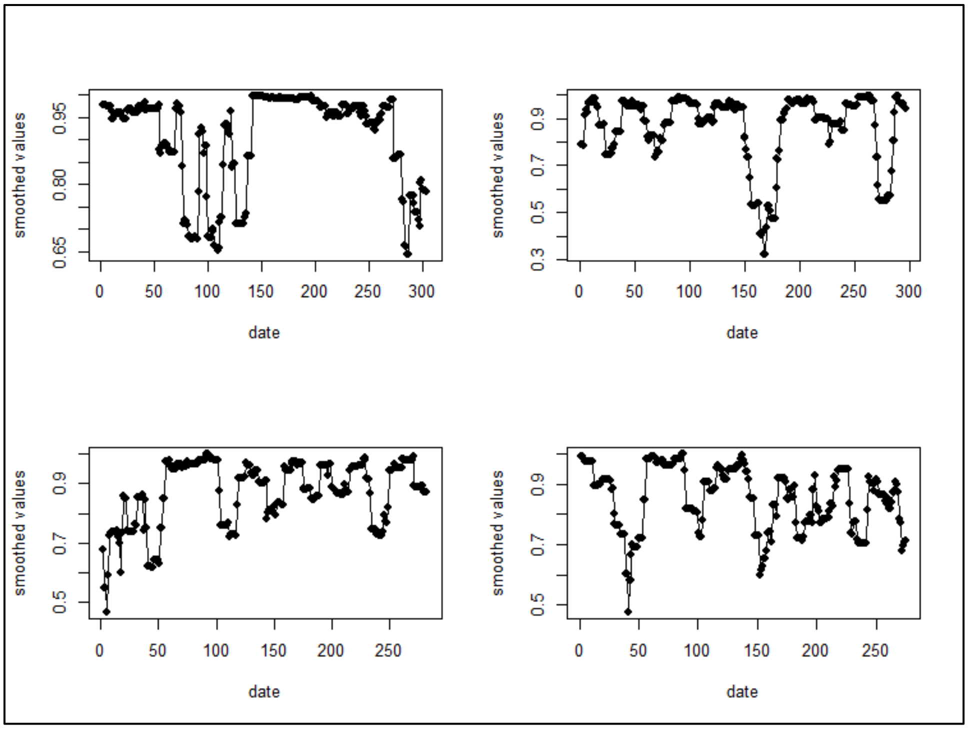

Figure 30 Correlated random walk with variable speed

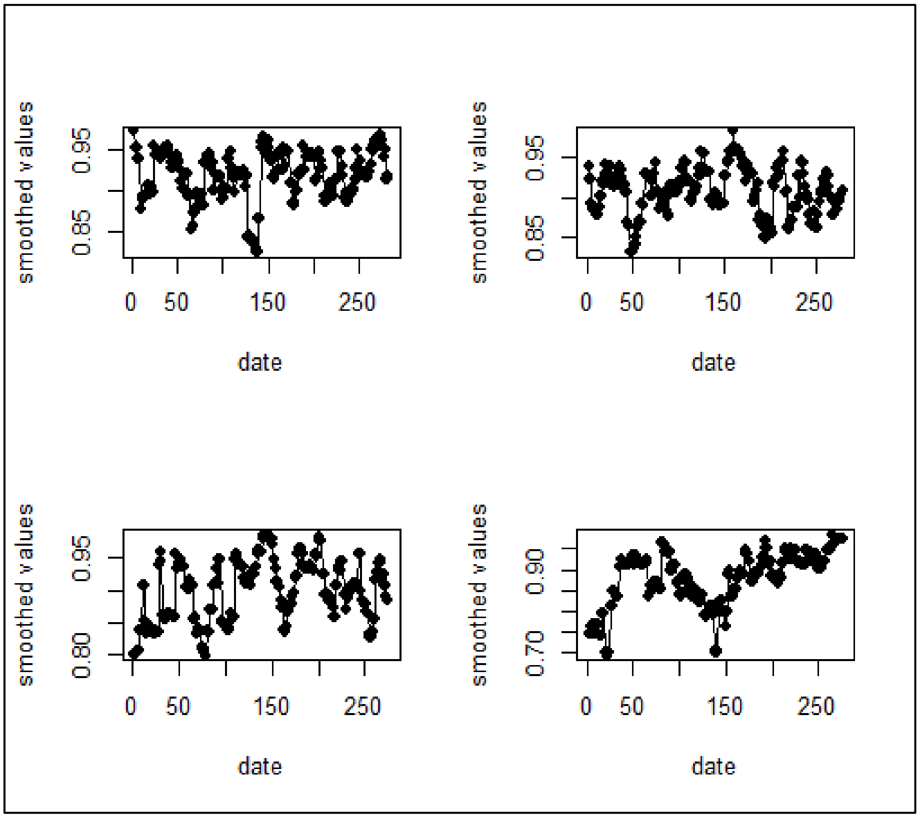

Figure 31 Simple random walk

APPENDIX 4: ACF VALUES FOR INDIVUDAL BIRD AND NETLOGO TURTLE RELATIVE ANGLE CHANGES

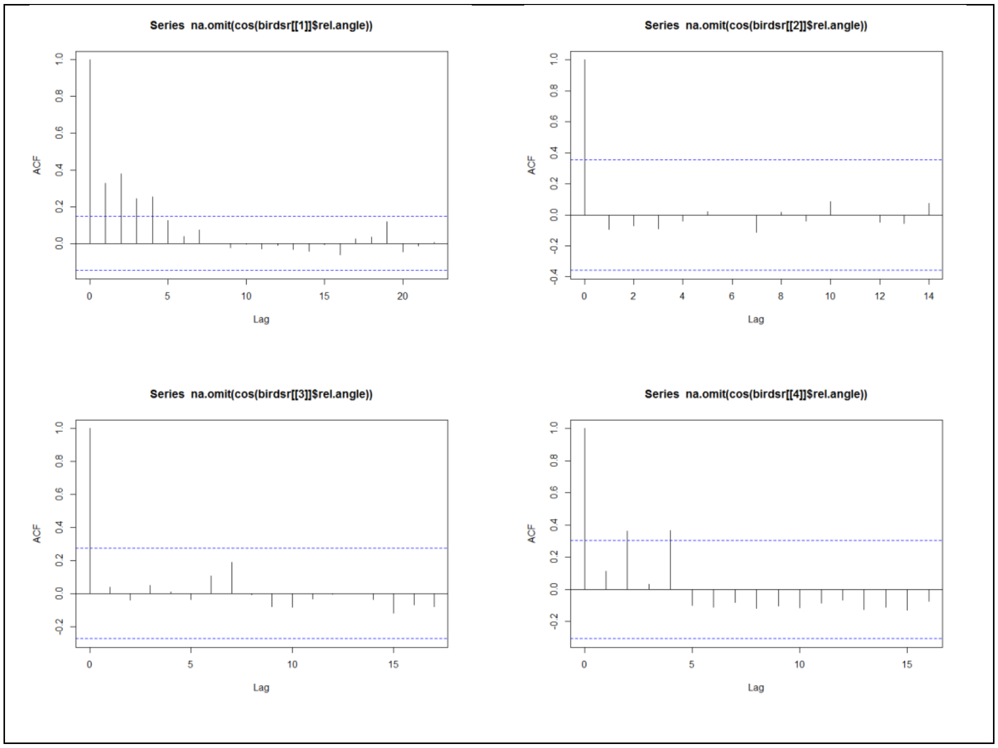

Figure 32 Seabird relative angle ACF, clockwise from upper left, PLS-13, PLS-2, PLS-4, PLS-8

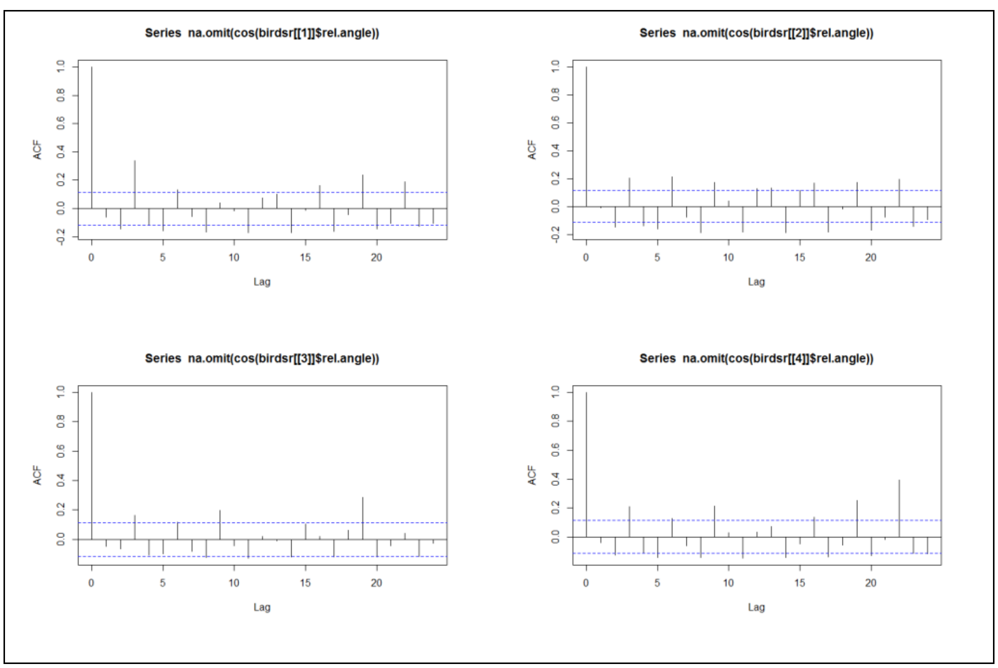

Figure 33 SRW relative angle ACF, clockwise from upper left, bird 0-3

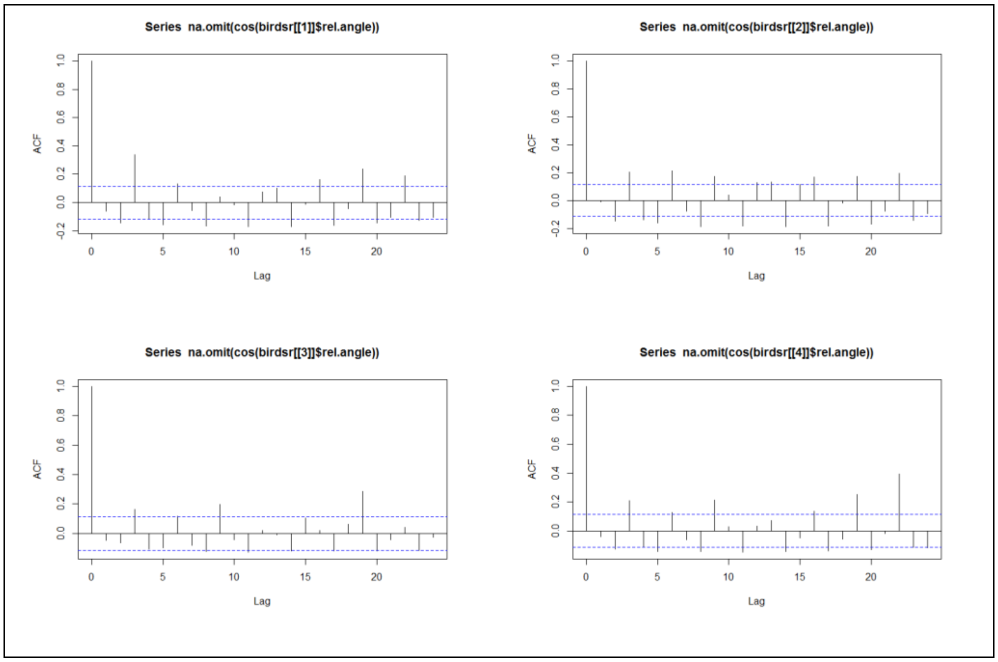

Figure 34 CRW relative angle ACF, clockwise from upper left, bird 0-3

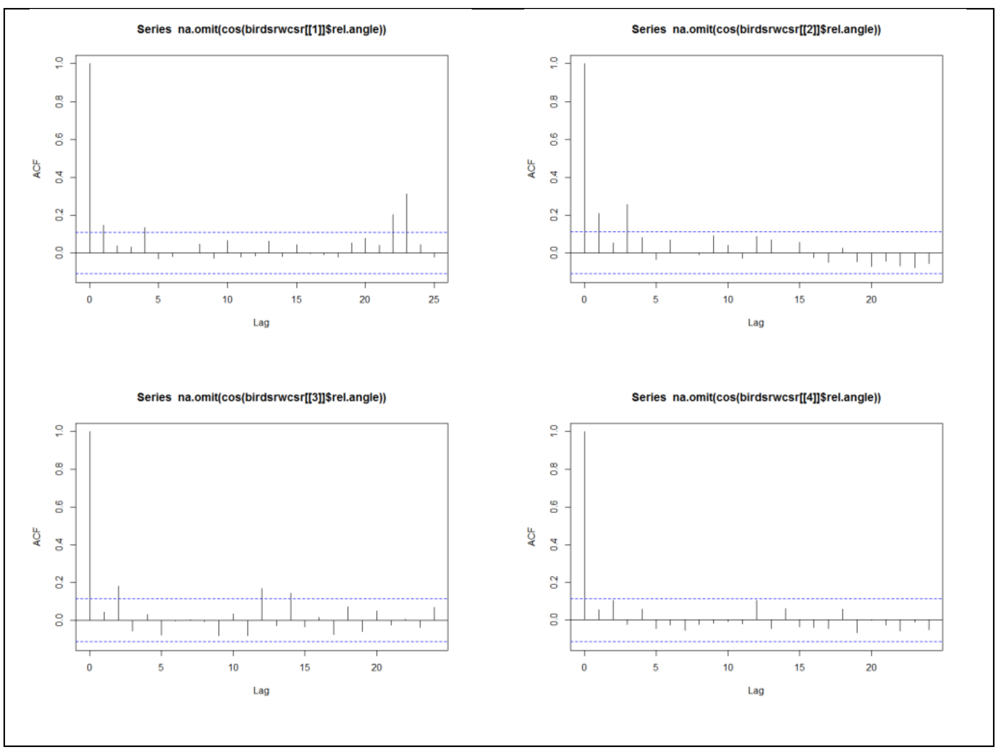

Figure 35 CRWS relative angle ACF, clockwise from upper left, bird 0-3

APPENDIX 5: NEAREST-NEIGHBOR CLUSTERING ANALYSIS OF HOME RANGE ESTIMATES

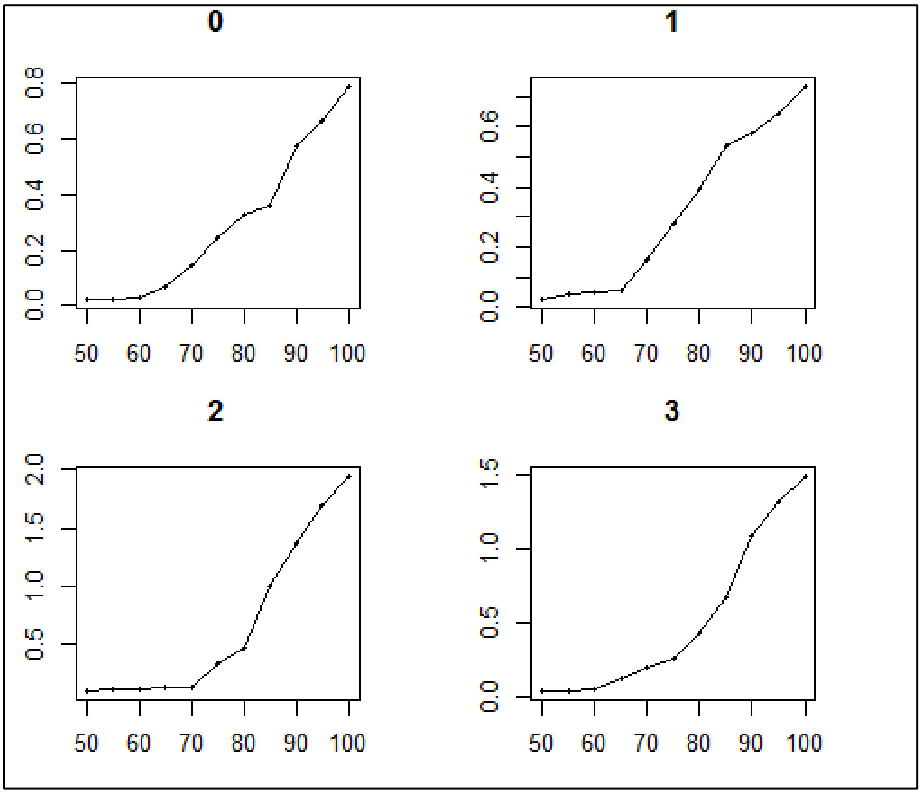

Figure 36 Simple Random Walk 'clusthr' home range results.

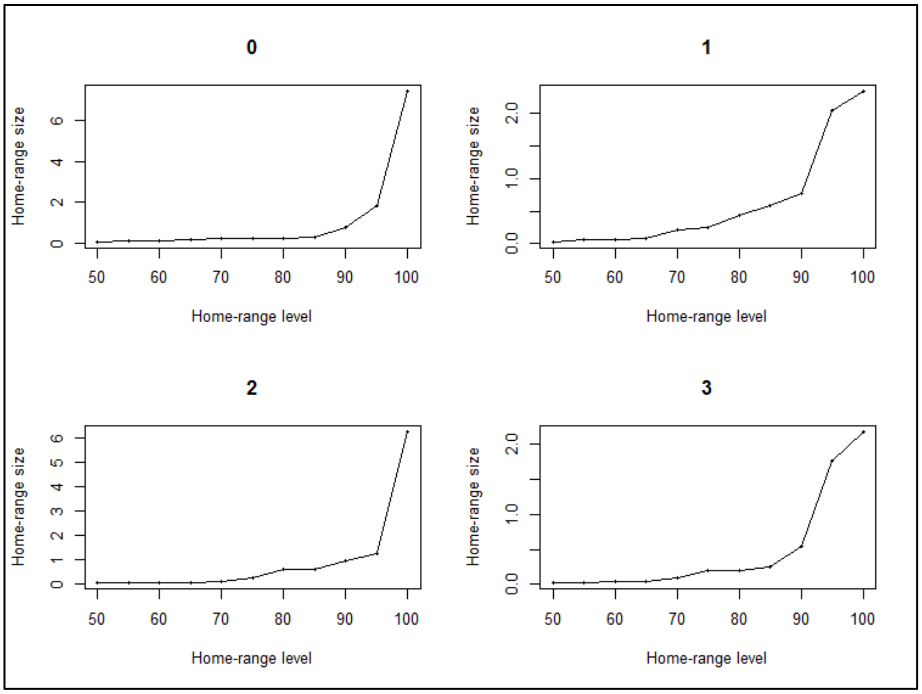

Figure 37 ‘Clustr’ home range results from the Correlated Random Walk with Variable Speed

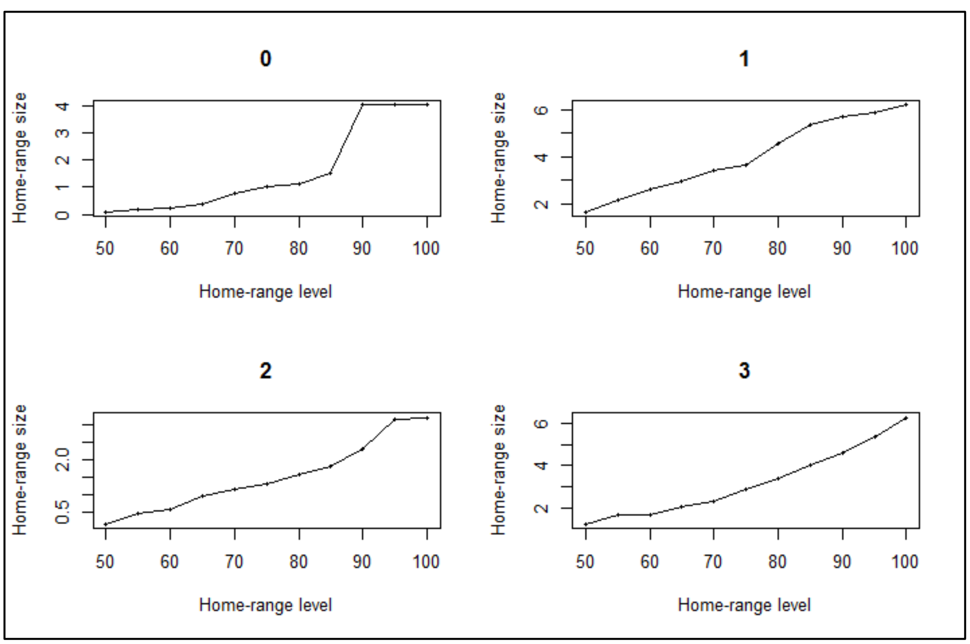

Figure 38 ‘Clustr’ home range results from the Correlated Random Walk

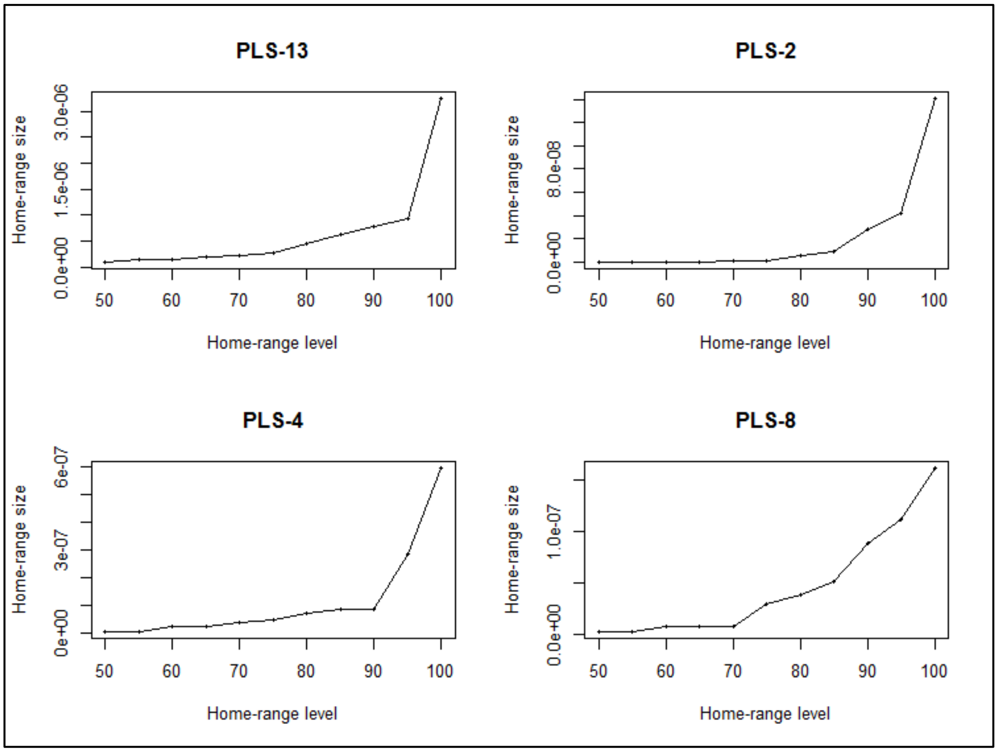

Figure 39 'Clustr' home range results from Moving-Object Bird data set

APPENDIX 6: SPEED, FROM RANDOM TO NORMAL

Speed obviously played a more significant role in two of the tests than I had anticipated when investing ecological analysis methods. After the third and final model iteration, digging deeper into the variation between seabird speed distributions and CRWS speed distributions seemed relevant. The CRWS agents speed distribution, in **Error! Reference source not found.** approximates the distributions in the seabird data in **Error! Reference source not found.**. Small variation is present in the probability distribution of the correlated random walk with variable speed model. It does however follow a very similar pattern to the real world data. The seabird data has several characteristically different curves, however the overall shape of the distribution is identical to the model values.

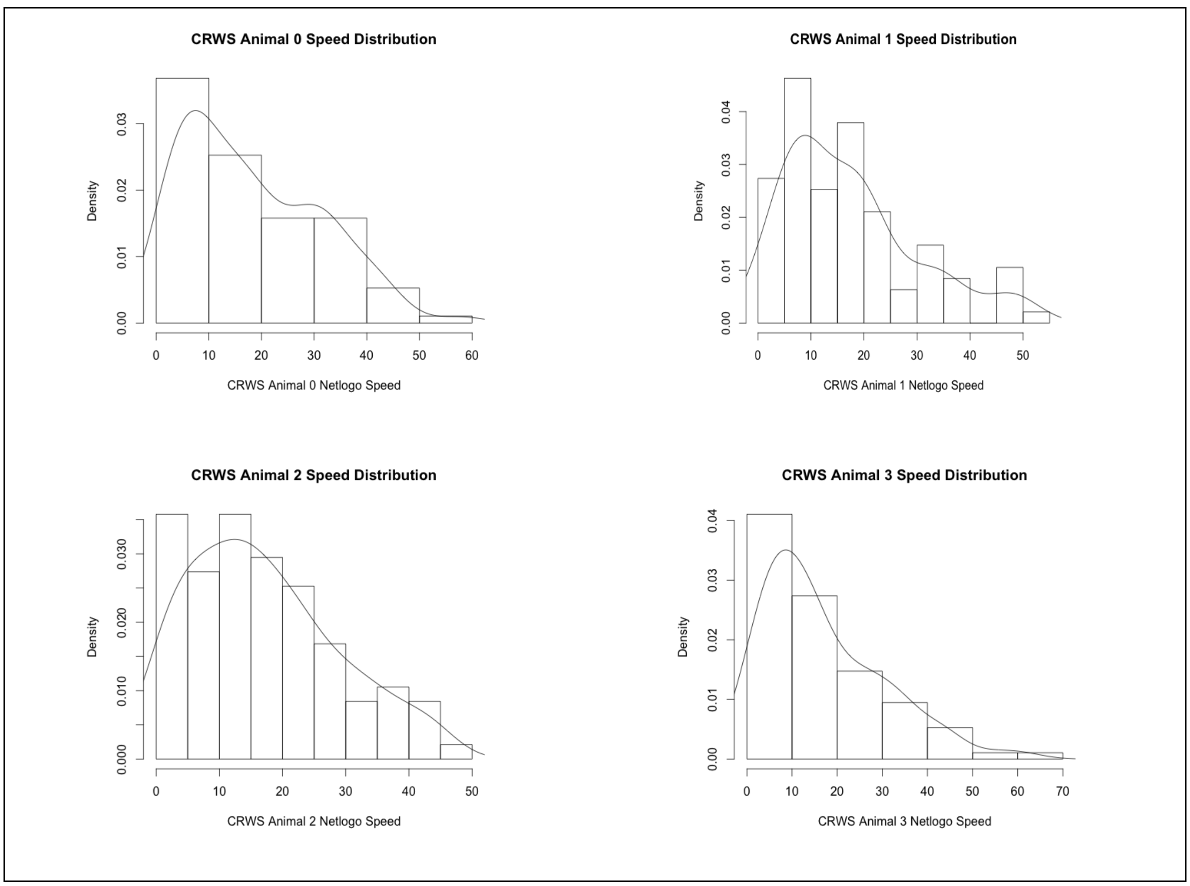

Figure 40 CRWS agents speed distributions

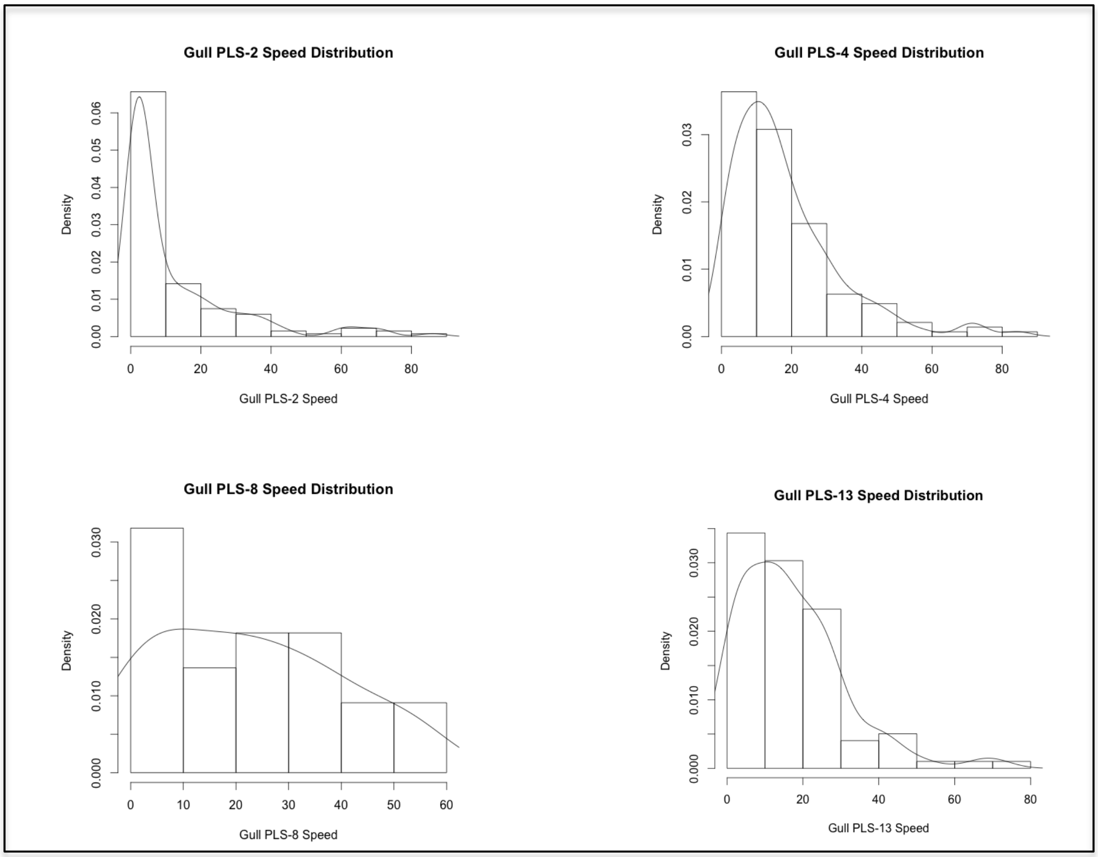

Figure 41 Galapagos Swallow-tailed Gull speed distributions

: