# Developing a Kalman filter approach to home range estimation: Applied to the Atlantic bluefin tuna (Thunnus thynnus)

Daniel Badger
*University of New Hampshire, Durham*

DEVELOPING A KALMAN FILTER APPROACH TO HOME RANGE ESTIMATION:

APPLIED TO THE ATLANTIC BLUEFIN TUNA (*THUNNUS THYNNUS*)

BY

DANIEL BADGER

BS, University of Washington, 2004

THESIS

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Master of Science
in
Natural Resources

December, 2007

UMI Number: 1449575





ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis has been examined and approved.

Thesis Director, Andrew B. Cooper
Research Assistant Professor of Earth, Oceans and Space
Affiliate Assistant Professor of Natural Resources

Maryellen M. Lutcavage
Research Associate Professor of Zoology

Andrew A. Rosenberg
Professor of Natural Resources Policy and Management and
Earth, Oceans and Space

Date

## DEDICATION

This degree and all that I accomplish will always be dedicated to, and be a direct result of the love and support of my father, Richard Badger, and my mother, Dian Badger's enduring memory of love and pride.

# ACKNOWLEDGEMENTS

This project was generously supported through a fellowship from the UNH Large Pelagics Research Center, and I am extremely grateful for the opportunity the support has given me. The completion of this thesis is due to the assistance of many people. A standout is my advisor, Dr. Andy Cooper, who has been an incredible asset, providing amazing insight, substantial commitment of his time, and notable patience as I journeyed through this Master's project. This thesis would certainly not have been possible without him. My other committee members, Drs. Molly Lutcavage and Andy Rosenberg, have also shown me great patience along the way, as well as remarkable support in advice and resources. Thank you to all three of you for everything you have done.

My friends (including all of my labmates) and family have also been invaluable. Special thanks go to Ben, Francois, Nuno, Wally, Hanna, Sarah, Eric and Rich for reviewing my thesis and for all of your support every step of the way. Francois and Ben, thanks so much for all the insight into ADMB and the Kalman filter! Finally, a special thanks goes to Taylor Luke and Julia Hammer. Taylor really made this all possible by keeping me going through much of the process, such as by feeding me throughout the late nights, and by being my pillar and friend to lean on. Julia, thanks for giving me the last pillar I needed to finish! To everyone I mentioned, and many more...thank you.

## TABLE OF CONTENTS

## LIST OF FIGURES

## ABSTRACT

DEVELOPING A KALMAN FILTER APPROACH TO
HOME RANGE ESTIMATION:
APPLIED TO THE ATLANTIC BLUEFIN TUNA (*THUNNUS THYNNUS*)

by

Daniel Badger

University of New Hampshire, December, 2007

Accurate estimation of an animal's home range, or utilization distribution, is of great importance to understanding the animal's role in the ecosystem, and for effective population management. Current methods for home range estimation often do not incorporate uncertainty in the observations of monitored animals. Given days without observations, they also have the potential to omit migration corridors when describing important habitat. Here the Extended Kalman filter is modified to return daily predicted geolocations, creating a most probable estimation of the true path the observed animal followed. Markov Chain Monte Carlo methods were used to map the uncertainty in this path to create a probability of use distribution, representing the animal's utilization distribution. The modified method was applied to Atlantic bluefin tuna (*Thunnus thynnus*) observed using pop-off satellite archival tags with light-based geolocation. The home range estimation technique developed can be used for any animal with a time-series of locations.

# CHAPTER 1

# INTRODUCTION

Throughout the development of fishery science, managers and scientists have been faced with an increasing need for detailed knowledge of the ocean's resources (Smith 1994). Current terminology refers to the description of where an animal is, or is likely to be, during a given time as its utilization distribution (UD), alternatively referred to by the more general concept of "home range" (Seaman and Powell 1996). Specifically, the home range is defined by Burt (1943) as "...that area traversed by the individual in its normal activities of food gathering, mating, and caring for young. Occasional sallies outside the area, perhaps exploratory in nature, should not be considered as in part of the home range." Therefore the rare excursion should not be included in the home range definition, and the home range is not the area in which you could potentially find an animal at all times. Rather, a home range is where there is a given probability (often set at 0.95; Anderson 1982, Worton 1989) that an animal will utilize the area during the time in question (Jennrich and Turner 1969). The 0.95 that is often used for definition of the home range refers to the contour inside which the animal spends 95% of its time, hereafter referred to as the '95% home range.' This can then be expanded to ask how the animal distributes its time throughout the home range, by calculating the UD. The utilization distribution is the continuous, probabilistic depiction of an animal's spatial use of an area, visualized by contour lines encompassing various levels of use.

Animal movements have been the focus of many population studies, with several techniques developed in an attempt to best describe an animal's spatial use of its environment in probabilistic terms (Burt 1943, Jennrich and Turner 1969, Anderson 1982, Worton 1989). In the terrestrial environment, the description of an animal's UD is often defined using stationary features of the landscape that can be consistently located from year to year (*e.g.* landmarks; Bethke *et al.* 1996). In the oceanic environment, direct and completely accurate observation of animals can be difficult. The fluctuating structure in which pelagic fishes reside and this difficulty in observing movements of fishes make a strictly landmark/coordinate-based definition of UD inappropriate. Those fluctuations in the environment (*e.g.* temperature) can greatly influence where the animal spends its time (Kitagawa *et al.* 2000, Itoh *et al.* 2003). With the corresponding fluctuation in UD through time, it is particularly important to have a realistic and accurate estimate of the UD in order to relate it to the movements of prey, temperature fields, and more. Current approaches for estimating UD are insufficient as they are derived from periodically locating an individual and estimating its UD without accounting for the temporal correlation of the monitoring structure, the biological capabilities of the animal (*e.g.* how far it can travel in a day), or the accuracy of the observations. An accurate approach for defining home range, sensitive to the specific organism being studied, is necessary in order to understand the behavior and adequately manage aquatic species. Such an approach will also serve to illuminate the influences that environmental characteristics have on a fish's spatial use (*e.g.* sea surface temperature fronts and prey distributions) (Brill *et al.* 2002, Gutenkunst *et al.* 2007).

Many benefits can be gained from knowing the probable location of a species, and predictability in the fish's movements would prove invaluable to scientific study and management plans. Adaptive management of a species can be quite difficult without knowing the environmental characteristics that can be used to locate that species, or how it interacts spatially with other species. Defining the UD for commercially important species like the Atlantic bluefin tuna (*Thunnus thynnus*) can aid in the identification of the relationships with environmental and biotic factors that influence the structure of the UD, eventually enabling the prediction of tuna locations.

This thesis provides a proof of concept for an approach to UD estimation that is more biologically appropriate and potentially accurate than current approaches. The state-space Extended Kalman Filter model was modified and applied to geolocation records of tagged Atlantic bluefin tuna, *T. thynnus*, in the Gulf of Maine. This new technique lays the building blocks for improved correlation of the tuna's UD with that of their prey, competitors, and predators for which time-series of location data also exists. Sea surface temperatures and other physical features of the ocean can also be correlated to tuna locations (Humston *et al.* 2000, Schick *et al.* 2004). The validity of such correlations, however, depends on using as realistic an estimate of home range as possible.

The developed technique will be useful for UD, and thus home range, estimation of any species for which time-series of location data exists, both in the marine and the terrestrial environments.

## 1.1 The Importance of Utilization Distribution Estimation

An animal's UD may depend on the time of year, the location of other animals with which it may interact, and highly variable environmental conditions, which may shift from day to day and year to year. If a correlation of these environmental factors with an animal's UD can be identified and analyzed (McLoughlin and Ferguson 2000, Adams 2001), it could lead to improved predictability of where the animal is likely to be under particular conditions (Hinton and Nakano 1996, Schaefer and Fuller 2002). Realistic and accurate estimates of the UD are a necessary step towards application of environmental factors to predict the tuna's spatial use of an area (Brill and Lutcavage 2001, Macdonald and Rushton 2003, Newlands et al. 2004) as well as estimating resource use patterns (Millspaugh *et al.* 2006).

## 1.2 Gap in knowledge of *Thunnus thynnus*

There is a pressing need for a model capable of predicting the location and identifying the distribution of Atlantic bluefin tuna, which appears to have high inter-annual and geographic variability (Powers and Porch 2003, Fromentin and Kell 2007). Estimates of the spawning biomass for Atlantic bluefin tuna remain controversial (Butterworth and Punt 1993, Restrepo *et al.* 1994, Restrepo 1996, Fromentin and Powers 2005), though conservative estimates predict that the current population size is one-eighth of that needed to produce maximum sustainable yields (Sissenwine *et al.* 1998). Current catches have been depressed compared to previous years, inhibiting both industry and management (Lutcavage 2004) and it remains unknown if the fishery has overexploited the stock (as suspected by: ICCAT 2003), which could have led to a seemingly sudden collapse, or if the fish are simply residing somewhere not yet known.

## 1.3 Importance of an Improved Utilization Distribution

Describing the utilization distribution of *T. thynnus* and relating it to environmental factors could greatly improve stock assessment and management. Unfortunately, techniques for the estimation of utilization distributions have some major dilemmas to their effective use. The need for an effective and applicable method extends beyond this tuna species. Indeed, it extends into all of fisheries management, and further, for the study of any animal or population often needs information on the spatial distribution and movement patterns of the animal in question.

Recently, fishery management theory has sought to expand beyond managing single-species and move away from the assumption that these species are independent from other factors. Instead, efforts are now underway to conduct ecosystem-based management (as called for by, for example: Pew Oceans Commission 2003, U.S. Commission on Ocean Policy 2004). This has posed substantial challenges for managers attempting to shift their methods appropriately (Hanna 1999). The shift in approach may require substantial improvements in the understanding of how various members in the ecosystem interact (Molsa *et al.* 1999, Hill *et al.* 2006). An effective method that accurately defines an animal's UD could improve identification of environmental and ecosystem dynamics that influence the animal's probable location. If such a method is developed, it would provide enhanced analyses of how decisions made for one species will impact the entire ecosystem.

One such advantage that is gained by an understanding of where an animal like tuna spends its time is the ability to correlate the tuna's UD with that of prey species like herring (*Clupea Harengus*), a known bluefin tuna prey species (Crane 1936, Bigelow and Schroeder 1953, Chase 2002, Estrada *et al.* 2005, Golet *et al.* 2007). This can provide insight into the role that tuna play within the ecosystem, as well as some detail regarding the factors that influence where the tuna travel. Previously developed UD estimation techniques do not always operate under completely realistic assumptions, potentially hindering these kinds of efforts. While each technique has distinct advantages and disadvantages, they typically all share at least one of a few shortfalls to be discussed in Section 3.0. A UD estimation approach is needed that utilizes information regarding the temporal proximity of observations, is sensitive to the distance the animal in question could have traveled during the time elapsed between observations, and accounts for the uncertainty in the observations.

# CHAPTER 2

# HOME RANGE AND UD ESTIMATION TECHNIQUES

The best method for estimating home range remains disputed (Adams 2001, Jonsen *et al.* 2003, Horne and Garton 2006), but most involve defining the home range, or the UD, with a frequency distribution. Most methods currently used for home range estimation tend not to account for serial correlation between observations of an animal or temporal gaps in observations, and assume the observed locations are exact. Typically, home range estimation falls into the category of non-parametric techniques (meaning that there are no assumptions that the distribution pattern follows a prescribed pattern, such as a circle) and are based on density estimation. Although examples of parametric-like home ranges that are circular, elliptical, and, occasionally, linear parametric shaped home ranges have been observed (Fitch 1958, Calhoun and Casby 1958, Stumpf and Mohr 1962), these situations are not common. A brief review of some prominent home range and UD estimation techniques follows.

## 2.1 Minimum Convex Polygon Method (Mohr 1947)

The minimum convex polygon (MCP) method estimates the minimum area that all observed locations of an animal represent by enclosing the locations by imaginary lines connecting the outermost positions of the animal in the smallest possible convex polygon. For example, suppose a hypothetical tuna is located in the positions displayed in Figure 1A. The MCP would define the area of the home range as that shown in Figure 1B, encompassed by the gray lines. For this example where the locations are densely and uniformly packed, this method may be useful. However, it encounters problems when faced with locations that do not have as compact of a structure, as in Figure 2 where a land mass makes the true use pattern concave.

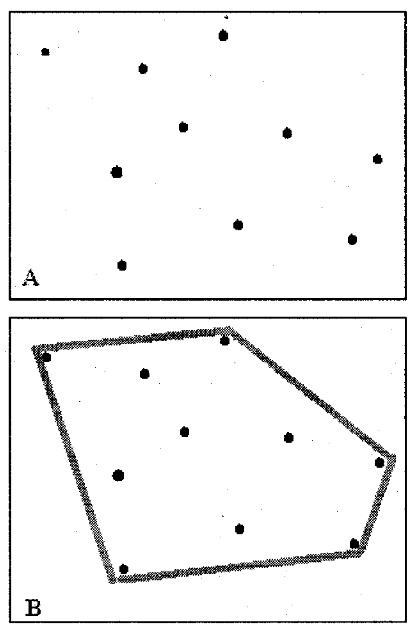


**Figure 1 – Minimum Convex Polygon Method. Positions of a hypothetical tuna. The gray lines in B show the area that the minimum convex method would estimate to be the 100% home range of the tuna.**

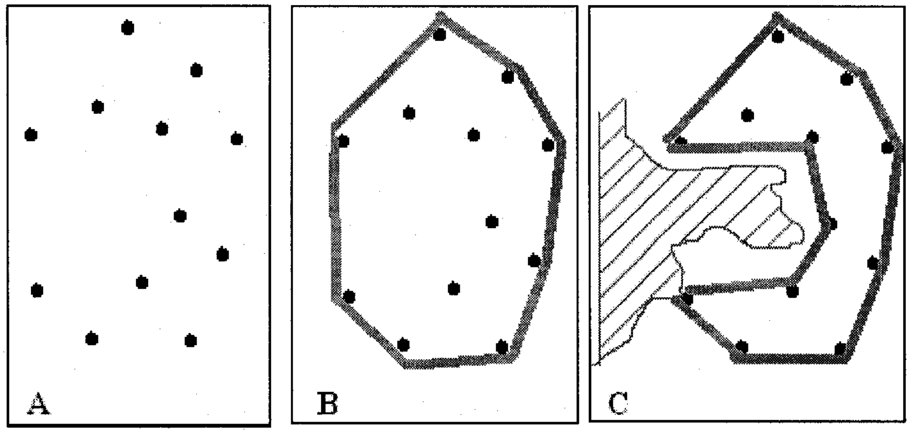


**Figure 2 – Overestimation by Minimum Convex Polygon Method. A pitfall of the minimum convex polygon method. (A) Hypothetical locations of a tuna. (B) The area that the MCP method would classify as the home range (encompassed by the gray line). (C) Striped area represents a landmass that the tuna cannot traverse, and thus its home range could not include this area as is assumed in B. The displayed gray lines outlining the home range in C is not consistent with the MCP method, as it creates a concave polygon.**

It may be argued that in such a case, one could dictate that the area of the land mass would be omitted from the home range estimate. However, features that restrict or dramatically influence animal movements are not always so easily identified. A further drawback to the minimum convex polygon method is that the home range area estimated is dependent on the sample size, with an increasing sample size resulting in an increased estimated home range area without a decrease in estimated variance (Jennrich and Turner 1969, Anderson 1982). In addition, if the home range is not convex, the MCP tends to overestimate the home range.

Finally, this approach estimates the home range that encompasses all the observations (the 100% home range). It does not produce a UD, which makes correlation between environmental factors and spatial use difficult. The MCP assumes that observations are known without error, are independent, and ignores temporal distance between observations.

## 2.2 Bivariate Normal (Jennrich and Turner 1969)

The bivariate normal approach operates with the assumption that the home range of an animal can be described by probability ellipses surrounding a center of activity, otherwise described as a bivariate normal distribution. This method can be quite effective in encompassing the entire range of the observed animal, but often includes areas that are not at all utilized by the animal, therefore inflating the estimated area utilized. The approach can be effective for many animals that may sleep in a consistent location (*i.e.* a den) and radiate their activity out from that location. If the animal is free-ranging, however, and does not distribute its activity in an easily describable geometric shape, parametric approaches to home range such as this are not appropriate.

Like the MCP, the bivariate normal approach assumes all the observed locations are known without error. This may not be an issue for some animals, but for animals such as tuna where accurate observations are unusual, the bivariate normal home range estimate would not be appropriate. However, the bivariate normal can produce utilization distributions, which is a distinct advantage over the MCP. It cannot incorporate available information such as the temporal sequence, distance between the observations, or the movement rates of the animal.

## 2.3 Harmonic Mean (Dixon and Chapman 1980)

Dixon and Chapman (1980) commented on the inadequacy of using the arithmetic mean of locations around which to center an estimated home range, as was done in the bivariate normal approach. They argued that the arithmetic mean of locations could fall in an area that was never used by the animal if, for example, the animal spent its time in multiple, spatially separated areas of high utilization. In Figure 3 it is shown that the approach of Jennrich and Turner (1969) produces an ellipse that effectively captures the area of use, but also includes large areas where there was little or no observed utilization by the animal. Meanwhile, the harmonic mean approach produces an area describing the home range that encompasses 95% of the observations, and with further distinction describing the areas of intensive use.

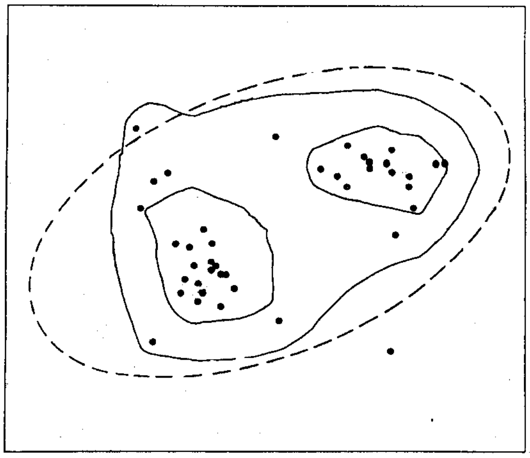

**Figure 3 – Harmonic Mean Method. Adapted from Dixon and Chapman (1980), observed locations of an animal are plotted as the points. The dashed line represents the probability ellipse as what would be estimated by the bivariate normal approach. The outer solid line indicates the area defined by the harmonic mean approach as the area containing 95% of the loci, while the inner solid lines surround the "area of greater activity intensity."**

This approach is based on calculating the areal moment such that the $n^{th}$ moment at a location *j*, is

$$M'_n \text{ at } j = \frac{\sum_{x=1}^{P} r_{jx}^n}{P} \qquad (1)$$

where P is the number of observed locations in the data, and $r_{jx}$ is the distance between *j* and the observed location *x*. The harmonic mean center is the minimum value of $\sqrt[-1]{M'_{-1}}$ calculated as the minimum of

$$\frac{1}{\frac{1}{P}\sum_{x=1}^{P}\frac{1}{r_{jx}}}. \qquad (2)$$

Essentially, this finds the point at which the summed inverse distance from all the observed locations is minimized. Thus, if there are the occasional points outside the area of heavy use, it will not drastically impact the location of the harmonic mean. Dixon and Chapman (1980) applied this approach to time series data by a moving 11-observation portion of the data so that the harmonic mean was updated with changing hubs of activity, allowing for the distinction between multiple areas of high use throughout the observation's time frame. The areas between the hubs of activity are given some importance as being used by the animal but with a lower frequency not identified by the high intensity contour line. This ability is important because managers often need to know the area an animal traverses in order to go from one area of high activity to another (Eggleston and Dahlgren 2001, Mumby 2006, Rouget *et al.* 2006). However, like many other UD estimators, the accuracy depends somewhat on the size of grid used to define the *j* points. The harmonic mean approach, like the MCP and bivariate normal, assumes that the observations are known without error. This assumption is often violated when dealing with oceanic animals of which direct observations are difficult.

## 2.4 Fourier Transform Method (Anderson 1982)

The Fourier method is similar to estimating a probability density function by using a histogram of observation frequency within a given square of a grid placed over the observation area. However, such a histogram has a severe limitation due to its dependency on the resolution of the grid used (Anderson 1982). The Fourier transform method avoids this issue by placing infinitesimally thin but tall columns upon each observation location. The sharp change in height from these tall cylinders to adjacent areas with no observations of the animal can then be smoothed to create a probability of observation density surface similar to the kernel density approach (Section 2.5). From this, contour lines of the probability of use are created, delineating the UD. While this method works generally well for modeling where the animal was observed, like the other home range estimators it does not account for where the animal was between observations in a biologically meaningful sense (*i.e.* the dispersal of the animal from the observation point is not modeled, and therefore is not sensitive to the study animal's biology). It assumes observations are independent and known without error. The UD may then be underestimated. The density estimates produced can also be negative values (Worton 1989), which is not a realistic situation.

## 2.5 Kernel Density

At present, the most accepted utilization distribution estimation techniques are the kernel density estimators (Seaman and Powell 1996), first introduced to ecologists as a home range estimator by Worton (1989). In this approach, a kernel consisting of a probability density, visualized by concentric rings that represent lower probability as the distance from the observation point grows, is placed over each observation point in the sample. A grid is then superimposed on the data, and within each cell the densities from all kernels that overlap that cell are summed. Observations nearer to a point of evaluation will contribute more to the summed density than an observation far from it, thus the density estimate will be relatively higher in areas with many observations. An example of such a kernel is the bivariate normal density kernel of *f(x)*, which can be defined as

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{3}$$

where the kernel *K*( ) is a unimodal symmetrical bivariate probability density function and *h* is a smoothing parameter that can be varied by the modeler. *f(x)* is the probability density function of an unknown utilization distribution where $X_i$ is a random sample of *n* independent points. The form of the kernel *K* can be defined by various shapes. For instance, the biweight kernel $K_2$ (Silverman 1986:76) defined as

$$K_2(x) = \begin{cases} \left(3/\pi\right)\left(1 - x'x\right)^2 & \textbf{for} \quad x'x < 1 \qquad 2.1 \\ 0 & \phantom{\textbf{for}} \quad x'x \geq 1 \qquad 2.2 \end{cases} \tag{4}$$

where *x'x* is the distance from the evaluation point to the observation point divided by the smoothing parameter, *h*. This particular member of the kernel family would place a probability density upon each observation like that shown in Figure 4.

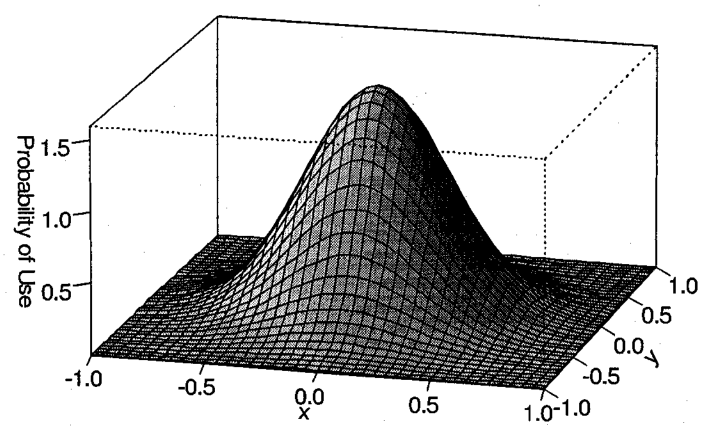


**Figure 4 – Biweight kernel $K_2$**

This method utilizes information on animal locations and analyzes the densities of those locations over space, creating a surface plot of the probability of finding the animal in any given region. The kernel density estimator has many positive characteristics such as being uninfluenced by effects of grid size and placement (Seaman and Powell 1996). Particularly, it is nonparametric, allowing it to estimate densities of any shape, as is appropriate when dealing with animal home ranges. However, there are drawbacks of the kernel density approach. As can be noted in equations 3 and 4, there is no variable in the kernel density approach that accounts for the elapsed time between observations, ignoring the serial correlation of observations. Indeed, in order to remove autocorrelation, previous studies using this home range estimator intentionally gather less data

(Kernohan *et al.*1998), loosing potentially valuable information. There is also little consideration of the movement rates that the animal in question typically undertakes. Movement rates can be somewhat accounted for by selecting among the various kernel types and smoothing parameter, but not in a direct, biologically meaningful way. Because of this omission, should the kernel density approach be fed identical locations and timings of observations for a fast-swimming animal like a tuna and for a slower-swimming animal like a loggerhead turtle (*Caretta caretta*), the resultant home range estimations would be identical for both species. However, the home range of a tuna has the potential to be larger than that of the turtle, as the tuna tends to travel farther on a daily basis (Papi *et al.* 1997, Wilson *et al.* 2005). Furthermore, if the distance between observations is large enough, then without making the kernels un-informatively massive, the modeled area that the animal used may have gaps between kernels in areas that the animal must have traversed to move from the location of one observation to another. Finally, as with the other techniques discussed, observations are assumed to be known without error. Because of these assumptions made by the currently used UD estimators that are often violated, a new approach appears warranted.

## 2.6 Kalman Filter Home Range Estimator: A New Approach

To address the need for an improved, more realistic method of defining an animal's utilization distribution, a modification to the Kalman filter (Harvey 1989) application in the software AD-Model Builder (Otter Research Ltd) is proposed. The Kalman filter is a set of recursive state-space equations that Sibert and Fournier (2001) were among the first to apply to tracking data. The filter analyzes position estimates to estimate a ‘most probable track,’ tag geolocation errors, and relevant parameters of a biased random walk model that simulates an individual's likely movements. Sibert *et al.* (2003) later applied the Kalman filter to position estimates of bigeye tuna, *Thunnus obesus*, derived from archival tags to estimate the horizontal movements of the individual tagged tuna. These methods were applied by Wilson *et al.* (2005) to pop-up satellite archival tag data (PSAT) of Atlantic bluefin tuna (*Thunnus thynnus*) in the Gulf of Maine to produce most probable tracks of tuna in that region. In these publications, utilization distributions were estimated using the fixed kernel approach upon the geolocations of the most probable track.

Using the data gathered by Wilson *et al.* (2005), all possible tracks of a single fish with associated likelihoods are estimated with the Kalman filter, and from these estimates a utilization distribution is created which accounts for the data's biological and temporal specifics. The available data has locations derived on a daily time-scale. Traditional applications of the Kalman filter estimate the underlying, true movements of the animal on a time scale identical to that of the observations, mirroring any gaps in those observations. Instead, by adjusting the modeling program to update the process model's estimated location to regular and adjustable time intervals, locations can be estimated for time steps without observations. A grid will be placed over these possible tracks, and a histogram of use for a given location will be produced. The resulting density field will represent the probability of use for a given location by the individual, creating a surface plot of the area where the tuna was likely to have been during the time period in question.

A modified approach to estimating a tracked animal's UD with this application will be developed that addresses the assumptions the previously discussed methods often violate. The result will be an improved approach for utilization distribution estimation that produces more realistic results than current approaches, which can lead to improved correlation with environmental variables.

# CHAPTER 3

# METHODS

## 3.1 Data Collection

This thesis uses the data gathered by Wilson *et al.* (2005) on bluefin tuna in the southern Gulf of Maine (N=66 in 2002, N=61 in 2003). Bluefin tuna were tagged with pop-up satellite archival tags (PSAT, model PTT-100, Microwave Telemetry, Inc., Columbia, MD) as part of a study to ascertain bluefin tuna movement, behavior and their interaction with the environment in the north Atlantic.

PSATs were attached to a fish's dorsal musculature using metal or plastic darts (Lutcavage *et al.* 1999, Graves *et al.* 2002, Wilson *et al.* 2005). After 500 days of collecting data while attached to the fish, the tags were programmed to send an electrical charge to the nose-cone to initiate a reaction with the saltwater that corrodes the attachment wire. The tag then floated to the surface and transmitted its data through the Argos satellite system. If the tag remained at a constant depth for more than four days, the tag was programmed to initiate a release under the assumption that such behavior indicated the animal had died or that the tag had been shed (Sibert *et al.* 2003, Wilson *et al.* 2005).

The tags carried an internal clock, sensors to measure light level, ambient temperature and pressure; and a battery voltage meter. The position of the tag upon the Earth, termed its geolocation, was calculated using the readings of the light levels and time of day. Geolocation of the tag was derived via a proprietary algorithm of Microwave Telemetry, although the theory underlying this algorithm is well understood (Hill 1994, Hill and Braun 2001, Ekstrom 2004). Longitude was estimated by measuring the time of local noon for a given day (when the sun reaches its zenith), relative to Greenwich Mean Time. The times of sunrise and sunset, characterized by the maximum rates of change in light levels for the day, and the associated day length yielded an estimate for latitude. The accuracy of the latitude estimate is highly affected by time of year, along with other sources of error, and this error should be corrected before use in movement studies.

## 3.2 Geolocation correction

Light-based geolocation estimates are often characterized by substantial error (Metcalfe 2001, Shaffer *et al.* 2005, Nielsen *et al.* 2006). As described in section 3.1, the PSATs used in this study carry a sensor to detect the time of local noon and the times at which the ambient light level is changing at its maximum rate indicating dawn or dusk. Accuracy of geolocation estimates can be affected by several factors including drift of the tag's internal clock, algal biofouling of the light-sensor housing, movement of the fish between dawn and dusk, or variability in the attenuation of light at depth (Welch and Eveson 1999).

The diving behavior of the tagged animal can also affect the accuracy of geolocation estimates. As the location is based on documenting the time of dawn and dusk, the tuna would need to be near the surface at those times in order to record the changing daylight. Bluefin tuna are an ideal fish for these tags because they have been known to spend the majority of their time in the uppermost 30m of the water column (Lutcavage *et al.* 2000, Wilson *et al.* 2005), although Lutcavage *et al.* (2000) did suggest that bluefin tuna may dive at times of light transition to feed on sandlance, a shallow-water species, rising off the bottom. If the tagged fish dive deep for too great a time at the beginning or end of the day, a reading of day length can not be attained for that day, creating gaps of observation data. This is not a likely problem with bluefin tuna in general, because the study that showed diving behavior at dawn and dusk was specific to an inshore location in the Gulf of Maine.

The estimate of latitude is particularly sensitive to errors in light measurements. Latitude errors exhibit fluctuating patterns of magnitude correlated with the time of year and its proximity to an equinox. Whereas error in longitude does not change throughout the year (Hill 1994, Hill and Braun 2001, Musyl *et al.* 2001), day lengths are essentially 12 hours at all latitudes during the summer and winter equinoxes. This makes day-length associated latitudes indistinguishable and their accurate estimation more difficult. For example, at a latitude of 50°N and five days away from the equinox, if the day length estimate was off by just one minute, there will an error of about 1.5° (over 150 km) in estimated latitude (Welch and Eveson 1999). Under optimal conditions, Hill and Braun (2001) suggest it is possible to estimate longitude and latitude with standard errors of 0.32° and 0.7°, respectively. These errors can be substantially inflated under field conditions near the tropics with upwards of 5.5° error in latitude (Musyl *et al.* 2001). These erroneous geolocation estimates can result in the tags reporting that a fish traveled hundreds of miles on land. Thus, it is important to filter out this error when basing a UD estimate on these geolocations.

Kalman (1960) developed a method useful for minimizing these uncertainties. Originally described for the engineering field, the Kalman filter has recently been adapted to apply to tracked animal positions to produce a 'most probable path' that the animal traveled while tracked (Anderson-Sprecher and Ledolter 1991, Sibert and Fournier 2001, Sibert *et al.* 2003, Nielsen 2004).

### 3.2.1 The Kalman Filter

The Kalman filter is a state-space statistical model comprised of: recursive equations describing the transition of a system from one state to the next (*i.e.* progression of time-steps); equations describing the observation model with errors in measurement of the state of the system; and a set of recursive relationships that, based on the observation model, updates the estimated state of the system and the components of variance at each step (the state model) (Sibert *et al.* 2003). The Kalman filter can also include equations for adjusting errors to the geolocation estimates derived from the archival tags based on proximity to the solstice. To represent the observation model, let $y_i$ be a two dimensional vector in terms of latitude and longitude, of the estimated sequential position of the tagged fish at observation moment $i$, and let

$$y_i = Z_i \alpha_i + d_i + \epsilon_i, \quad i = 1, \ldots, T \tag{5}$$

where $\alpha_i$ is its true position (described later in equation (8)), $Z_i$ transforms from coordinates of the plane expressed in nautical miles (nm) to coordinates on the sphere expressed in degrees of longitude and latitude, and is defined in equation 13, $d_i$ is a two dimensional vector of the bias in observing the position, $T$ is the number of measurements in the time series, and $\epsilon_i$ is a serially uncorrelated 2-dimensional random vector with mean 0 and 2 X 2 covariance matrix, $H_i$:

$$d_i = \begin{pmatrix} b_x \\ b_y \end{pmatrix} \quad \text{and} \quad H_i = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_{yi}^2 \end{pmatrix}, \tag{6}$$

where $\sigma_x^2$ and $\sigma_{yi}^2$ are the mean-squared errors in estimating longitude and latitude, respectively. $b_x$ and $b_y$ can be interpreted as the mean 'raw' error bias, while $\sigma_x$ and $\sigma_{yi}$ are the standard deviation of the 'raw' error (Musyl *et al.*(2001). Longitude estimation is determined by observing the time at local noon. Latitude estimation is determined by the length of day, and so accuracy diminishes greatly during the equinoxes when the day length is constant for all latitudes. Errors in latitude can be modeled in several ways (Sibert *et al.* 2003), but for the purposes of this study a location error variance model will be adopted in which observations near an equinox are considered highly uncertain while the uncertainty is lowest around the solstice (Nielsen 2004).

The latitude error structure is modeled as:

$$\sigma_{y_i}^2 = \frac{\sigma_{y_0}^2}{\left(\cos^2\left(2\pi(J_i + b_0)/365.25\right) + a_0\right)} \tag{7}$$

where $\sigma^2_{y_0}$ is the average latitude geolocation error, $J_i$ is the number of days since the first solstice prior to the start of the track, $b_0$ is a parameter to be estimated expressing the number of days before the equinox at which the latitude error variance is at its maximum and $a_0$ is a non-negative, dimensionless model parameter inversely affecting the general magnitude of the variance fluctuation.

Variability in day length has less impact on longitude estimation. Thus $\sigma_x$ is assumed to be constant over time, as are the biases, $b_x$ and $b_y$.

Using the observed locations, the movement of a fish along a time series is assumed to be a biased random walk on a plane, described by the transition equation,

$$\alpha_i = \alpha_{i-1} + c_i + \eta_i, \quad i = 1,...,T \tag{8}$$

where $\alpha_i$ is a two-dimensional vector describing the position of the fish at time $i$, $c_i$ is a 2-dimensional vector representing the bias of the random walk, and $\eta_i$ is a 2-dimensional vector of serially uncorrelated random variables with mean 0 and 2 X 2 covariance matrix, $Q_i$.

This application of the Kalman filter assumes that a fish's movements can be modeled as a biased random walk. The advection-diffusion equation is the continuous case of a biased random walk (Okubo 1980), and therefore the parameters of the biased random walk can be described by the advection-diffusion equation. Animals dispersing according to an advection-diffusion process distribute in such a way that the probability of observing an animal at point $x$ at time $t$ has been shown by Feller (1966, 1968) to be a normal probability density function,

$$p(t,x) = \frac{1}{\sqrt{4\pi D\Delta t}} e^{\left(-0.5 * \frac{(x-u\Delta t)^2}{2D\Delta t}\right)}. \tag{9}$$

After an amount of time at large, $t$, the mean position of the animal will be $ut$ and the variance $2Dt$, where $u$ is the mean rate of movement and $D$ is the rate at which the uncertainty of the position increases over time. It is the incorporation of these latter two variables that allows the Kalman filter to be sensitive to the biological tendencies of the animal in question (*i.e.* at what rate can it move) and to the temporal characteristics of the sampling design (*i.e.* increasing uncertainty of the position as the time between observations increases).

The parameters of the transition equation (8) are as follows

$$c_i = \begin{pmatrix} u\Delta t \\ v\Delta t \end{pmatrix} \quad \text{and} \quad Q_i = \begin{pmatrix} 2D\Delta t & 0 \\ 0 & 2D\Delta t \end{pmatrix} \tag{10}$$

where the change in time ($\Delta t$) is $t_i - t_{i-1}$ and $u$ and $v$ are mean longitudinal and latitudinal biases of the fish's random movements, respectively. $D$, $u$ and $v$ are from the advection-diffusion model (Equation 9). Note that $u$ and $v$ are to be interchangeable in Equation 9.

The Kalman filter is then comprised of a set of recursive relations that update the estimated position and the components of its variance at each time step (Harvey 1989, Sibert *et al.* 2003), such that for $i = 1, 2, \ldots, T$ the true position of the tagged fish is estimated from the random walk,

$$a_{i|i-1} = a_{i-1} + c_i \tag{11}$$

where $a_{i|i-1}$ is an estimate of the 'true' position of the tagged fish.

$$P_{i|i-1} = P_{i-1} + Q_i \tag{12}$$

updates the variance of that position. The total variance is then computed by combining the variance from the random walk, $Q_i$, with the variance of the observation, $H_i$, via:

$$F_i = Z_i P_{i|i-1} Z_i^{'} + H_i . \tag{13}$$

The position is updated by the tag following equation (5),

$$\tilde{y}_i = Z_i \alpha_i + d_i \tag{14}$$

where $Z_i$ is a 2 X 2 matrix which converts between coordinates on the plane expressed in nautical miles (nm) and coordinates on the sphere expressed in degrees of longitude and latitude. Let $a_{i|i-1}$ denote the optimal estimator of $\alpha_i$ conditioned on all observations up to and including $y_{i-1}$ such that

$$Z_i^{-1} = \begin{pmatrix} 60\cos(a_{i|i-1,2}/60) & 0 \\ 0 & 60 \end{pmatrix} \tag{15}$$

where $a_{i|i-1,2}$ is the estimated latitude position of the tag north of the equator in nm at the beginning of time step *i*, and 60 is the number of nm per degree of latitude and per degree of longitude at the equator.

The residual between the location estimated by the tag and that by the biased random walk is calculated as,

$$w_i = y_i - \tilde{y}_i . \tag{16}$$

The most probable position's parameters for the time step is then calculated as

$$a_i = a_{i|i-1} + P_{i|i-1} Z_i' F_i^{-1} w_i \tag{17}$$

and

$$P_i = P_{i|i-1} - P_{i|i-1} Z_i' F_i^{-1} Z_i P_{i|i-1} . \tag{18}$$

Equation (17) computes the most probable position, and equation (18) estimates its variance, as a tradeoff between the random walk position and the position estimated by the tag based on the relative variance of the two estimates, creating the 'most probable' track through the sequence of points $a_i = 1,2,\ldots,T$ with an estimated level of uncertainty around that track. The parameters describing the fish's movements to be estimated are the values of *u*, *v*, *D*, $b_x$, $b_y$, $\sigma_x^2$, $\sigma_{y0}^2$, $b_0$ and $a_0$, that maximize the log-likelihood function,

$$\ln L = -T \ln 2\pi - 0.5 \sum_{i=1}^{T} \ln |F_i| - 0.5 \sum_{i=1}^{T} w_i' F_i^{-1} w_i . \tag{19}$$

The Kalman filter takes noisy geolocations and returns the most likely position of the animal at the time of observation, given the observed geolocations and the estimated uncertainty in those observations. Sequenced together, they describe the 'most probable track' that the animal traveled during the time it was tagged. The '*kftrack*' package (Sibert *et al.* 2003, Sibert and Nielsen 2004) written for the statistical language program R (Ihaka and Gentleman 1996) provides easy application of the Kalman filter.

The program's example is for a bigeye tuna (*Thunnus obesus*) near Hawaii. The program produces a predicted track based solely on the previously observed positions, and a most probable track that is based on all observations. As shown in Figure 5, both the predicted track (solid grey line) and the most probable track (black line) are far less variable than the raw observations.

As written, however, *kftrack* predicts a location only on days for which an observation exists. Unfortunately, PSATs often do not report geolocations for 100% of the days. Once a tag comes off the fish and floats to the surface, it can take several days for the data to upload through the Argos satellites during which the upload can be interrupted by factors such as heavy seas sending the antennae below the water (Lutcavage *et al.* 1999, Galuardi 2006). On any given day, the tag may be unable to make a geolocation observation for a variety of reasons. For instance, the tag may have sustained damage from interactions with other animals (NMFS 2004) or other causes discussed in Section 3.2. These limitations can lead to as many as 79% of the days not having observations (De Metrio *et al.* 2003). Because the Kalman filter predicts locations based on the information from all the observations, locations between observations can easily be interpolated with slight modifications to the Kalman filter approach, as recommended by Galuardi (2006: 52).

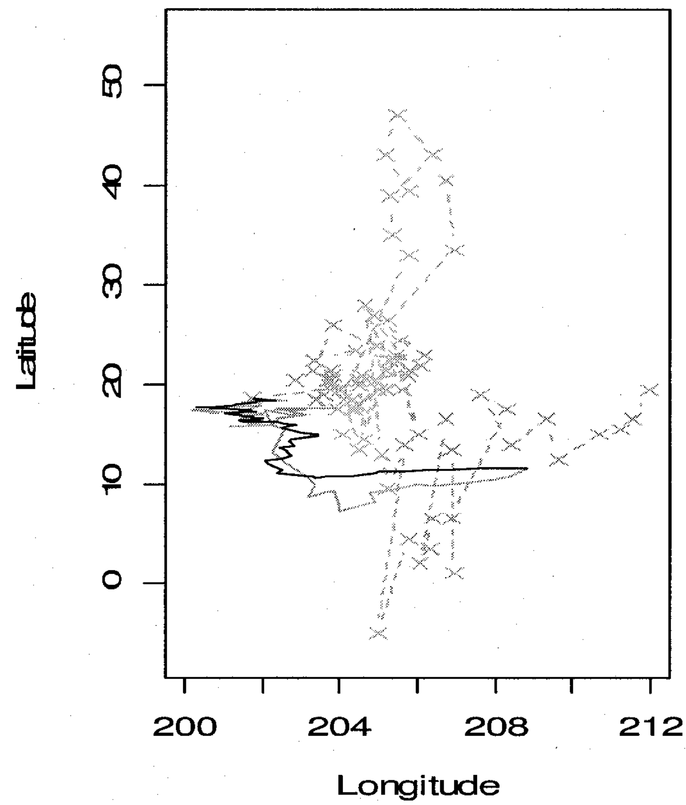


**Figure 5 – Example *kftrack* Most Probable Track. X's indicate observed locations. The dashed grey line indicates the temporal progression from point to point. The solid grey line indicates the predicted track, with each position estimate based solely on the previous locations, while the black line is the 'most probable track' that is smoothed using all observations.**

### 3.2.2 Why the random walk?

The approach presented in this thesis models bluefin tuna's movements as a random walk. Any movement model that describes an animal's movement patterns may be applied to the Kalman filter home range approach. An unbiased random walk was used for this thesis in part because it is general, easily incorporated into the Kalman filter, and has previously been incorporated into the

Kalman filter (Sibert *et al.* 2003, Nielsen 2004, Galuardi 2006). Most of the previous applications of the Kalman filter used the biased random walk for its movement model. However, combining the bias of the random walk with the MCMC procedures to be described in Section 5.4 resulted in instability in the model results. More complex movement models can be used and should be explored in the future (*i.e.* Moorcroft *et al.* 1999), as should implementation of a bias to the random walk to improve the biological validity of the movement structure.

## 3.3 Analyzing Effects of Observation and Prediction Frequencies

### 3.3.1 Predicting all days

The code for *kftrack* (Sibert and Nielsen 2004) was written using AD Model Builder (ADMB - Otter Research Ltd), utilizing R as a user-friendly interface. For this thesis the code was modified within ADMB (Appendix A) to predict geolocations of the tagged fish at times for which no observation exists. This modified code was written for the purposes of the Kalman filter utilization distribution estimation method, and will be referred to as *kfud. Kfud* forms a matrix with as many rows as time steps, where time steps are equal increments (*i.e.* one day) between the release date and end date of the track. The rows of this matrix were then populated with the observed positions. Rows of time steps for which observations do not exist were filled with zeros. The Kalman filter is an application to smooth observations. The parameters of the observation model were therefore not updated for time steps without observations. For these time steps, only those parameters associated with the process model were updated. Similarly, the likelihood function is only affected by the deviation of the observed location at a given time step, and the corresponding predicted location.

The interpolated positions predicted for time steps between observations serve to delineate where the animal may have gone while we could not observe it, as well as provide information about the uncertainty surrounding the most probable track. The *kftrack* package reports the uncertainty of its most probable track via a covariance matrix around each position. With no observations on unobserved days, *kftrack* does not address the uncertainty about the track between observations. Regular temporal spacing of estimated positions via *kfud* provides a more complete picture of the area that a tagged animal likely traveled, with the associated uncertainty.

### 3.3.2 Thinning Data

Observation rates can vary drastically with the animal being tracked based on the habitat, animal's behavior, and tracking technique (NMFS 2004, Nielsen *et al.* 2006), as can the accuracy of the observations. To assess how a decreased observation rate may affect the estimate of home range, the original observations were sub-sampled. In one scenario, every other day was assumed to have no observation, thus cutting the reporting rate in half. Another track was made with the same reporting rate by shifting the omitted observations. The first scenario had every other observation omitted beginning with the 2$^{nd}$ observation. The next scenario started the omissions with the 3$^{rd}$ observation. The 1$^{st}$ observation was kept either way, as that was the release point, and is assumed known without error. To more drastically thin the data, another scenario was explored in which 5 of every 6 observations were assumed void, cutting the reporting rate to 17% of the original observations.

### 3.3.3 Multiple Predictions a Day

*kfud* was designed to allow predictions of locations as frequently as once an hour, if desired. To review the effect that an increased prediction to observation rate might have on the resulting most probable track and UD, estimates were performed with 2 and 6 predictions a day, spaced 12 and 4 hours apart, respectively. Geolocation observations were assumed to be at noon (1200h). This is because the longitude estimated using light-based geolocation is based on time of local noon. To estimate latitude, an entire day's duration of sunlight is needed, and is therefore not associated with a particular time of day.

### 3.3.4 Altering the Animal's Mobility: Changing D

The home range of an animal whose mobility is more restricted than a tuna is expected to utilize a smaller home range. Similarly, by increasing the distance we assume the animal tends to have traveled each day, we would expect the uncertainty around the most probable path to increase, and create a larger home range estimate. To observe how the approach could be applied to animals with varying degrees of mobility, 95% home range estimates were produced with *D* declared as 50 and as 500 or 1000 $nm^2$/day.

## 3.4 Home Range

Part of the *kftrack* results is a covariance matrix for each predicted geolocation, which can be used to create 95% confidence ellipses around those positions using the *ellipse* package (Murdoch and Chow 1996) in R (Figure 6). While the ellipses can give a general idea of where the bluefin tuna was at the time of observation with 95% certainty, it does not produce either a utilization distribution or a representation of the home range as a single polygon. Furthermore, the ellipses only represent the uncertainty around the locations, rather than the uncertainty around the track.

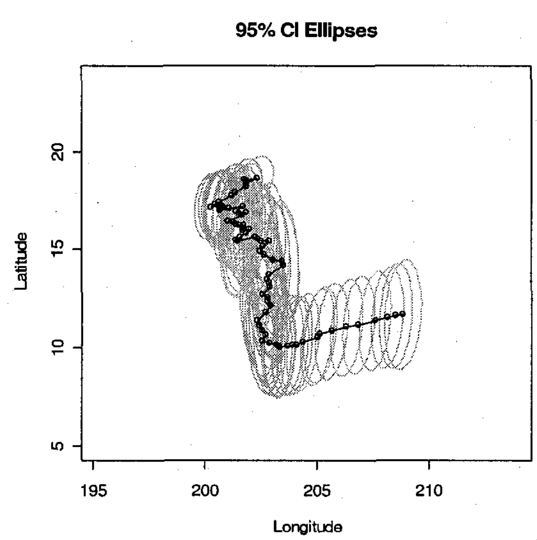


**Figure 6 – *kftrack* Confidence Interval EllipsesMost probable track (black line) of *kftrack*'s example bigeye tuna, with predicted geolocations (black circles) everyday. The grey ellipses are the 95% confidence intervals. Note that there is substantially more uncertainty in the latitudinal direction than longitudinally.**

### 3.4.1 UD Estimation and Representation

One challenge to using utilization distributions is in establishing the proper method of visually representing the animal's probabilistic use of an area. A number of methods exist within R to display utilization distributions through binning and smoothing functions, such as the function *bkde2D* (referred to here as the *2D Kernel Density* method) of the *KernSmooth* package (Wand 1994, Wand and Jones 1995) and the *kernelUD* function of the *adehabitat* package (Silverman 1986, Worton 1989, Bullard 1991, Worton 1995, Seaman and Powell 1998). This study presents 95% home range estimates using both of these methods showing the similarity of the pictures they produce of the home range. These methods place bivariate distributions (typically Gaussian) upon each position as discussed in Section 2.5 to create a smoothed picture of the home range. They then place a grid over the area and bin overlapping probabilities of the animal's presence in each grid cell. Another package in R, *hist2d*, written by Gregory Warnes in the *gplots* package creates a simple 2-dimensional histogram, using a grid over the study area and binning up the frequency of occurrence in each cell with no smoothing (simple binning). To utilize this method, high densities of positions are necessary, and cannot be effectively performed on the single, most probable track.

The key to estimating a realistic UD is to directly illustrate the probability of all possible paths, identifying areas that the tuna likely utilized. The three UD estimators mentioned above cannot provide a representative assessment of space use if only given the locations of the most probable path, as that would ignore the many other possible paths. Uncertainty was thus estimated through application of a Markov Chain Monte Carlo (MCMC) approach. A 100,000 run MCMC was applied to each scenario with a burn-in period of 2,000. Every $50^{th}$ run was extracted, producing 2,000 possible tracks based on the variance of the model's parameter estimates and the uncertainty in the observations. This produced geolocations of all possible tracks, the density of which represented the probabilistic use of a given area. The kernel density estimators and the simple binning approach introduced above were then applied to the MCMC results, illustrating a UD based on all possible tracks derived from the model's uncertainty. The *2D Kernel Density* approach was found to be substantially faster and robust, and produced a UD image similar to the other visualization methods. It therefore was used for analysis of most of the scenarios explored in this thesis.

## 3.5 Selection of Tracks

The Kalman filter home range approach was applied to tag tracks with patterns typical of bluefin tuna. Tag 37008 (Figure 7) exhibits a typical track for a tagged bluefin tuna which had several, separate areas of localized movement connected by periods of apparent directed travel. Tag 37011 (Figure 8) provides an example where the bluefin tuna doesn't appear to spend large amounts of time in any one area, but also doesn't seem to have a general trend in movement bias. Tag 3817 (Figure 9) provides a final test track in which the observations were made with very little uncertainty. This track is the observed track of the tag *after* it released from the tuna, which was determined by Doppler positioning, and can be accurately located within 150 meters (Argos 1996). The positions are found at least once a day. For this analysis, the observation nearest to noon of a given day was used as the daily position, and no days were missing. This track has small measurement error, a distinct bias in travel direction, and a daily observation record absent of gaps. It is therefore indicative of a tagging study for an 'animal' who's tracking does not have the typical complications that tuna tracking experiences.

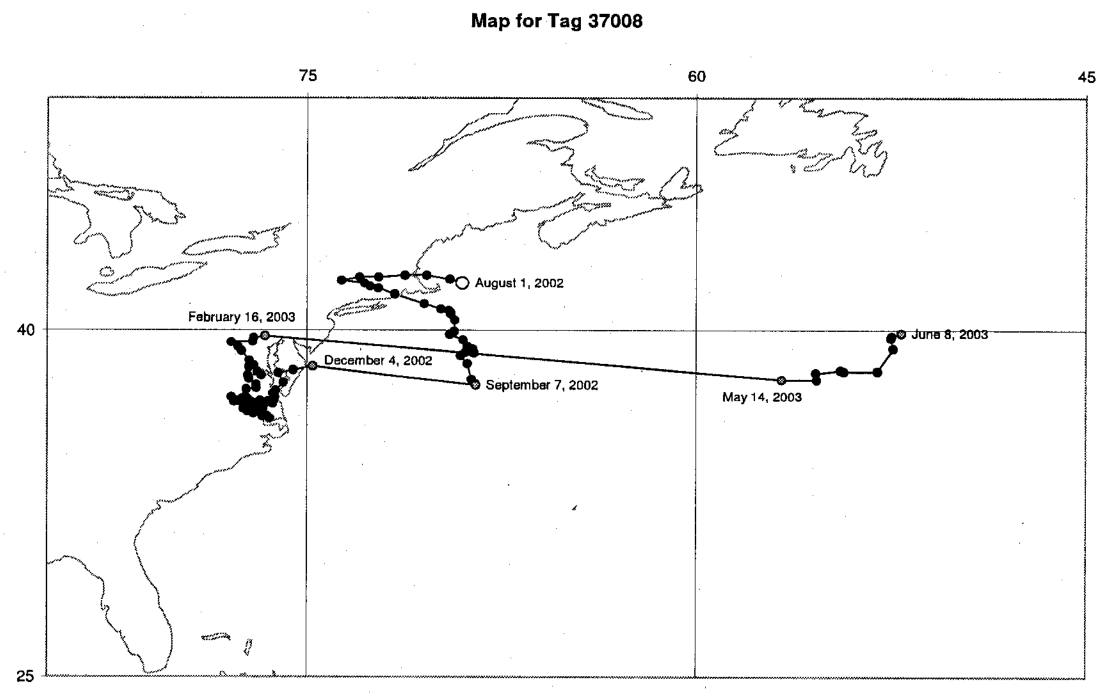


**Figure 7 – *Kftrack* corrected observed positions (black points) and most probable track (black line) of a bluefin tuna with tag # 37008, released August 1, 2002 at the open black circle.**

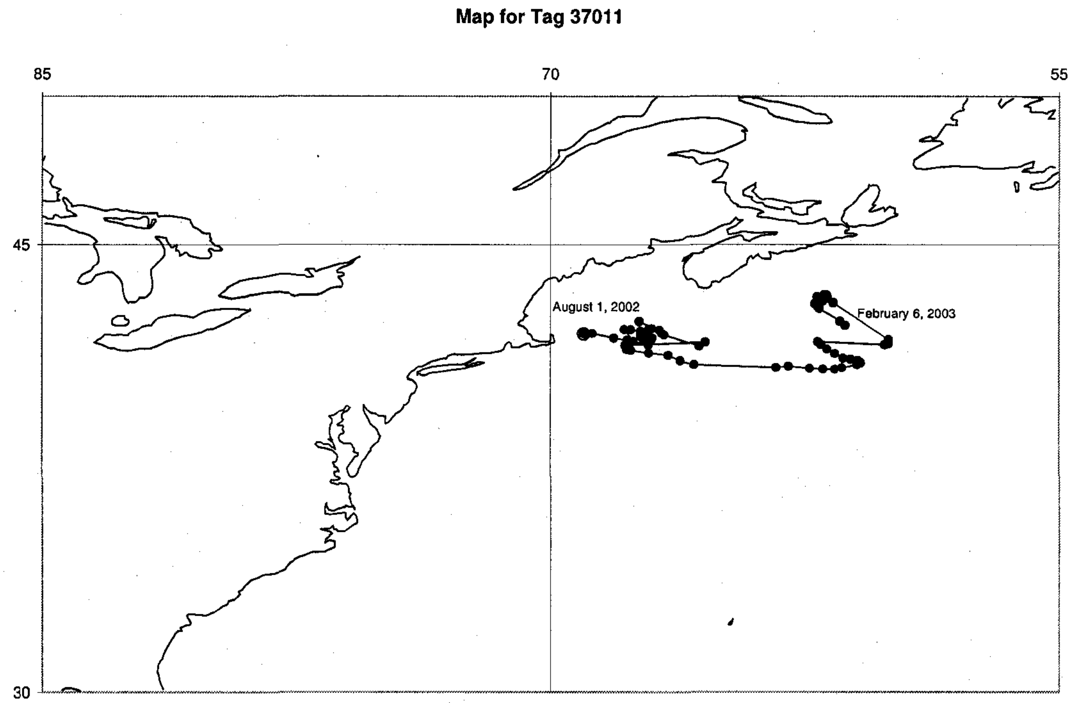


**Figure 8 – Kftrack of bluefin tuna with tag # 37011, released August 1, 2002.**

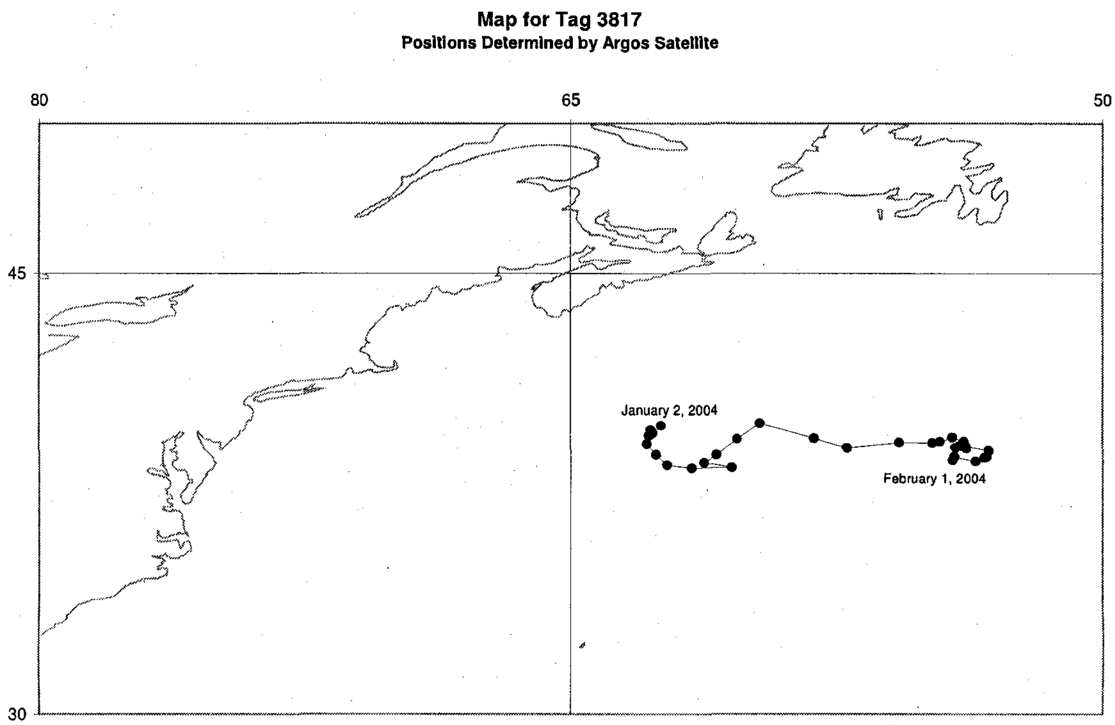


**Figure 9 – Observed geolocations, obtained through a link with the Argos Satellite. The track is ideal for estimating a home range for a straight track measured with very little error.**

# CHAPTER 4

# RESULTS

## 4.1 Extension to "*kftrack*"

Modifications made to the *kftrack* code to predict locations for days without observations were found to produce identical results to the *kftrack* estimates of bigeye tuna Track 241 when no unobserved days' positions were predicted (Figure 10). This indicates that the new code effectively performs the function of the previous without altering the estimates.

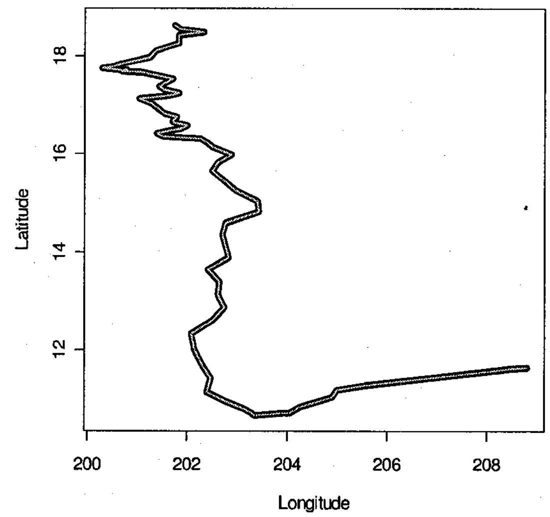


**Figure 10 – Comparing most probable track of *kftrack* (thick black line) and *kfud* (grey line).**

## 4.2 Application to bluefin tuna data – Tag 37008

### 4.2.1 Original Observations

Figure 11 shows the Kalman filter-corrected points of the bluefin tuna tagged with Tag 37008 as produced by *kftrack* (with predicted locations only for days with observations). Ellipses based on the covariance around each location represent the uncertainty in the track's individual geolocations. Note that the covariance ellipses suggest there is little likelihood that the tuna traversed the area between the clusters of observed locations. Applying the *kernelUD* function in the statistical program R to the track to estimate the 95% home range has a similar result (Figure 12). Though the estimated home range encompasses more area around areas densely-packed, the area in the middle is still considered unlikely to be an area in which to find the tuna. This gap in the estimated home range still exists when extracting the tracks from the MCMC of *kftrack* (Figure 13).

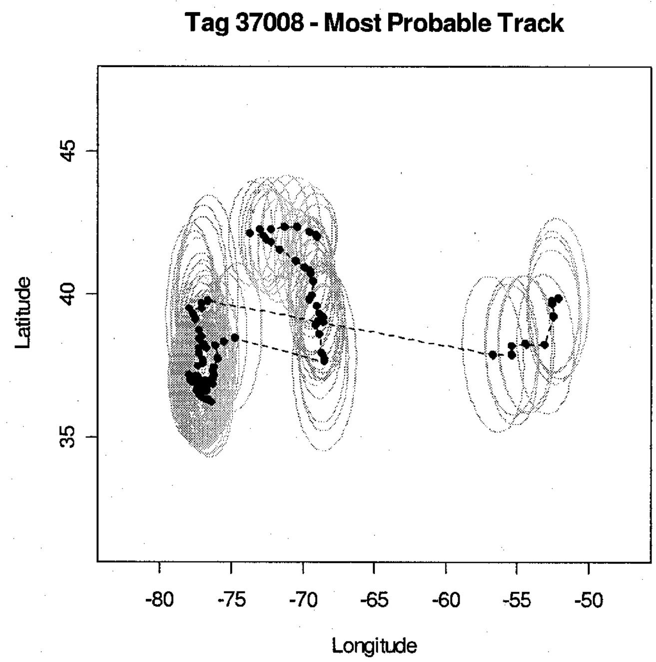


**Figure 11 – Tag 37008 *kftrack* confidence interval ellipses of the most probable track (black dotted line) of bluefin tuna tag 37008. Grey ellipses show the 95% confidence intervals around each observed geolocation.**

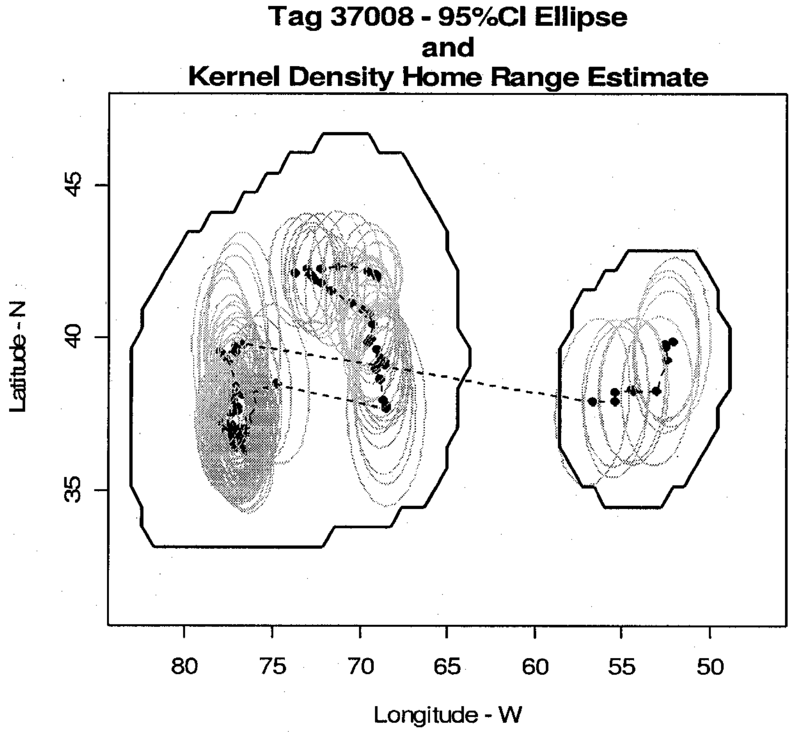


**Figure 12 – 95% confidence intervals (grey ellipses) around each location and 95% Home Range (black outline) estimated by *kernelUD,* based only on the observed geolocations.**

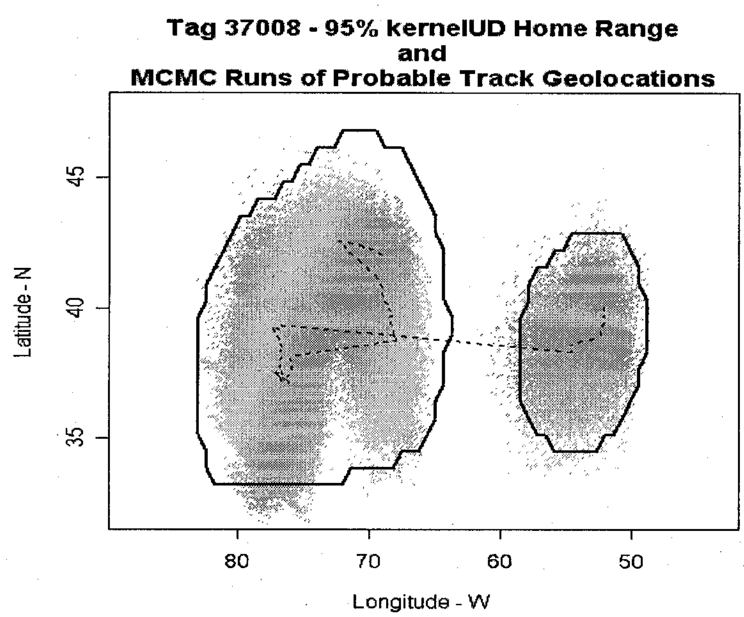


**Figure 13 – *kernelUD* estimate of 95% home range (black outlines) and geolocations from MCMC runs of kftrack (grey points). Dashed line is the most probable track.**

4.2.2 Prediction Every Day

Predicting a location every day for Tag 37008, as expected, placed estimated geolocations between observations for days without observations – one geolocation for every missing day (Figure 14). Note that compared to *kftrack*, the *kfud* track is altered slightly because of the increased smoothing effect of making predictions everyday. The effect of predicting every day's geolocation is evident in Figure 15 where the home range estimate by *kfud* bridges the gap between the hubs that the *kftrack* does not include.

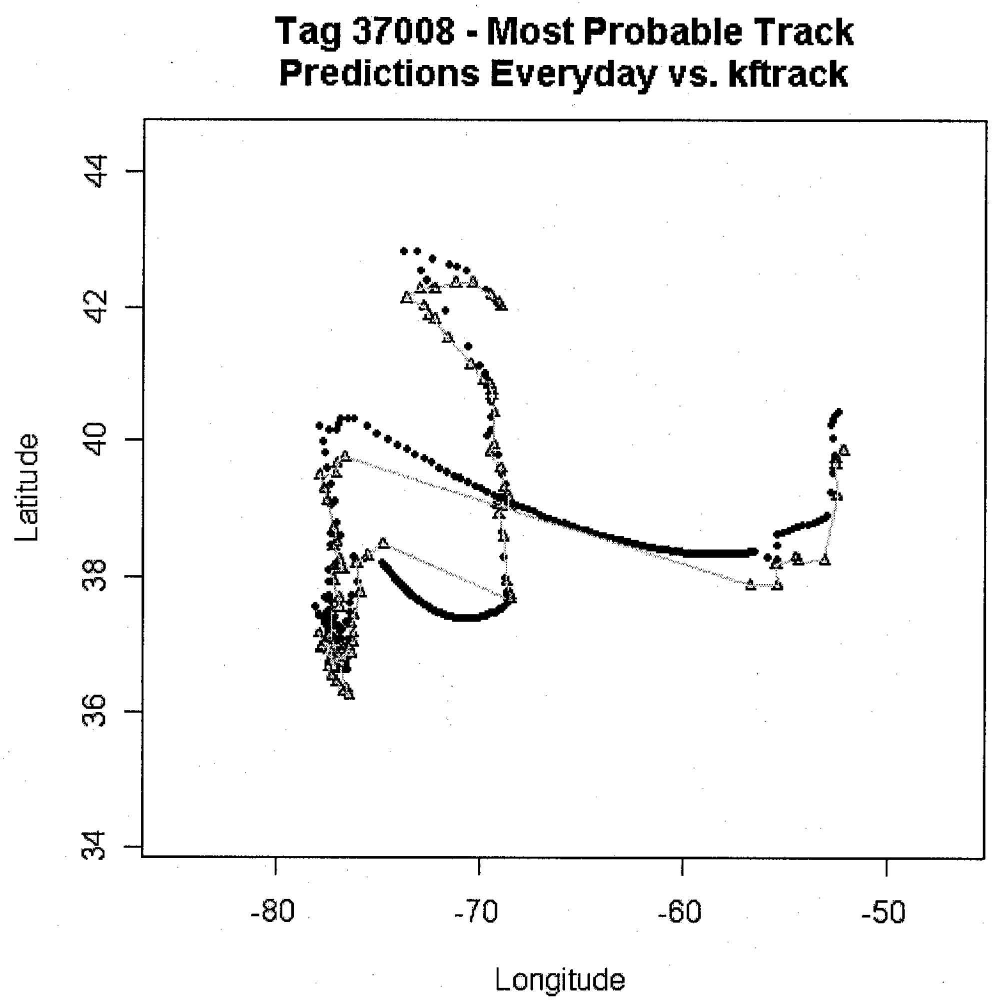


**Figure 14 – Most probable track for *kftrack* (grey line, black triangles) and for *kfud* with predictions every day (black points).**

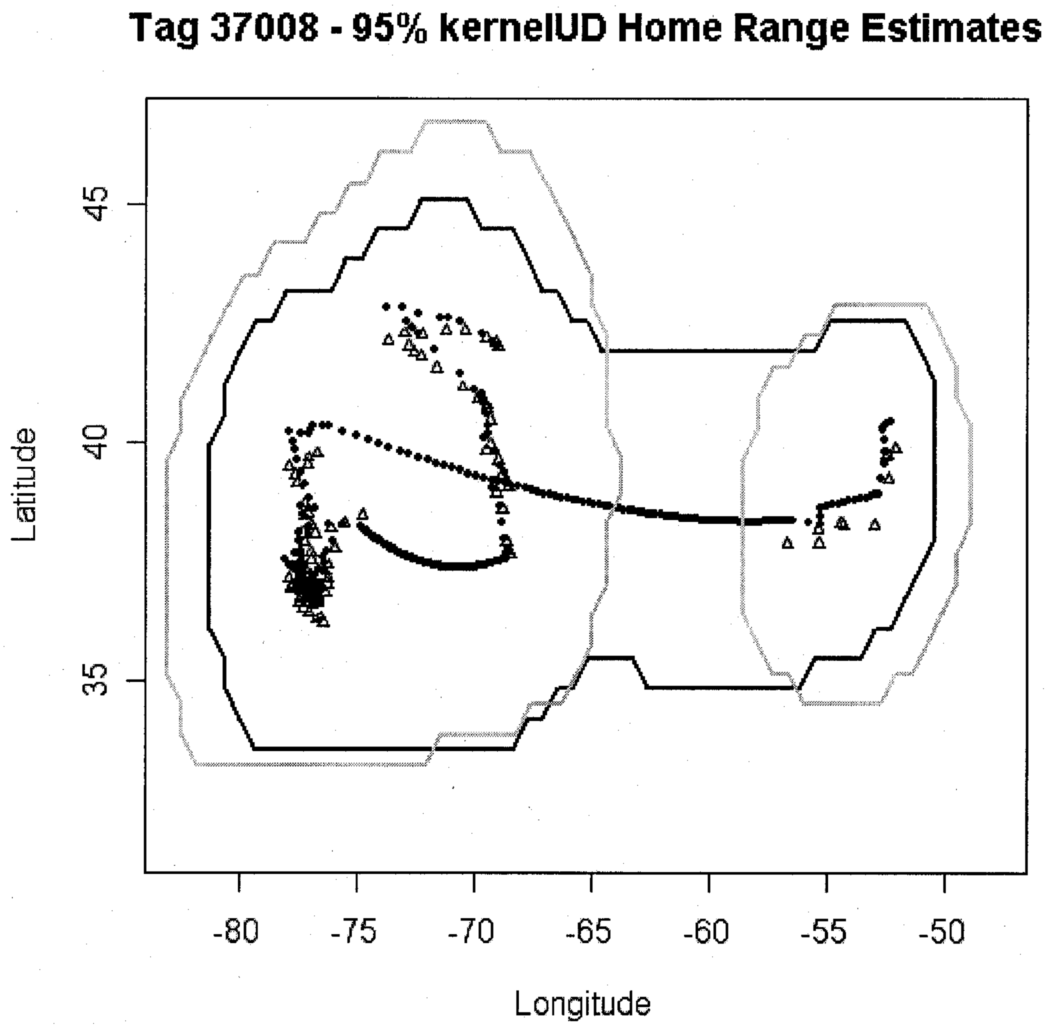


**Figure 15 – *kernelUD* estimate of 95% home range for *kftrack* (grey outlines) and for *kfud* (black outline) surrounding the most probable track (dashed line)**

Note that the *kernelUD* application was used to visualize this home range because the other two methods for binning and smoothing the data require a greater density of points (*i.e.* from an MCMC) in order to effective. Though the gap is bridged with *kfud*, the home range estimate is still assuming observed locations are known exactly, and it does not represent the tuna's potential area of use due to its species-specific ability to move particular distances in a day. To make this home range estimate more realistic, uncertainty can be mapped using an MCMC.

### 4.2.3 Home Range from MCMC Methods – Daily Predictions

Applying the home range estimation techniques to the estimated tracks produced by an MCMC yielded similar estimates for home range (Figure 16 and Figure 17) with no gaps in the 95% home range. The estimated 95% home range based solely on the geolocations of the most probable track did produce a noticeable, qualitative difference compared to the produced home range estimate based on the MCMC positions (Figure 17). For this track, traditional kernel density UD estimators applied only to the observed positions would have overestimated the tuna's home range. This is probably because the tuna's maximum daily movements it is capable of are not incorporated when not using the MCMC positions.

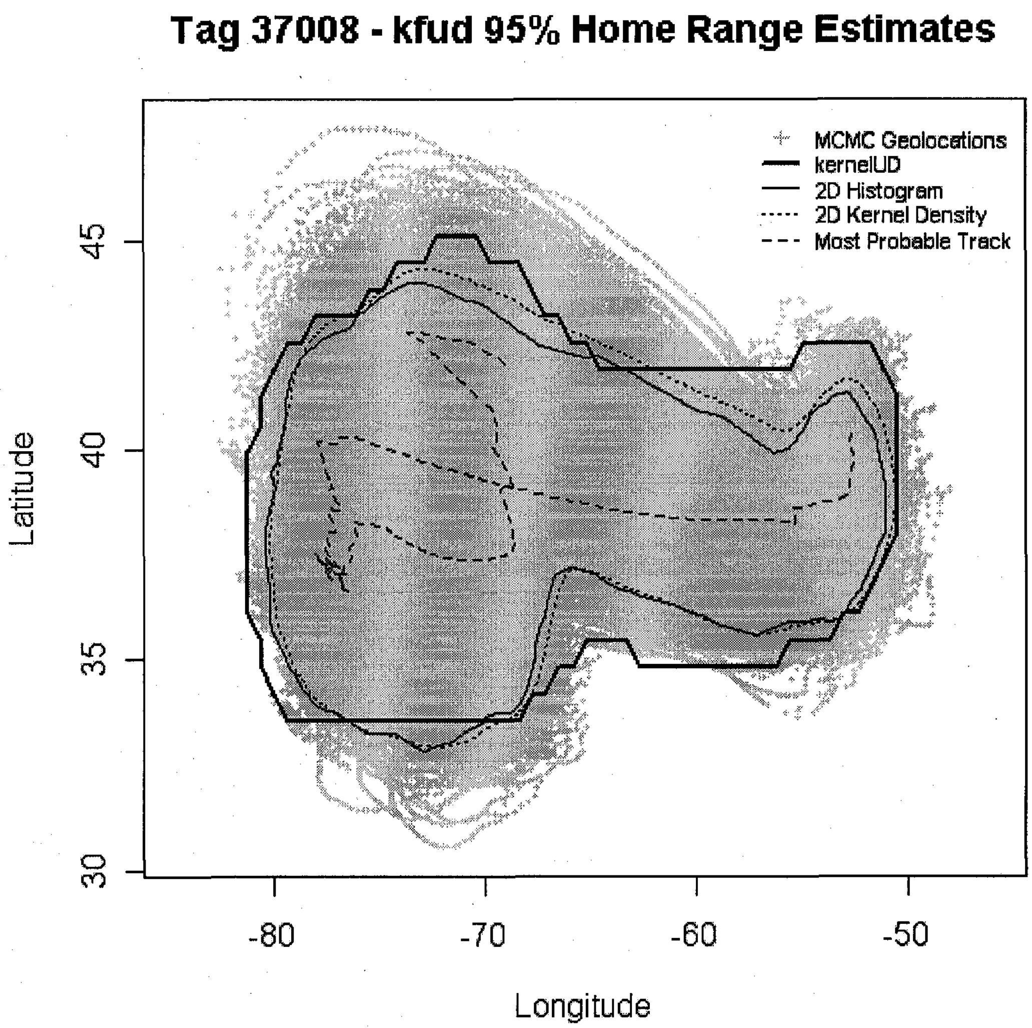


**Figure 16 – 95% Home range estimates from three methods available in the R program. Here, the *kernelUD* method is based solely on the most probable track while the *2D Histogram* and the *2D Kernel Density* methods are based on the MCMC geolocations.**

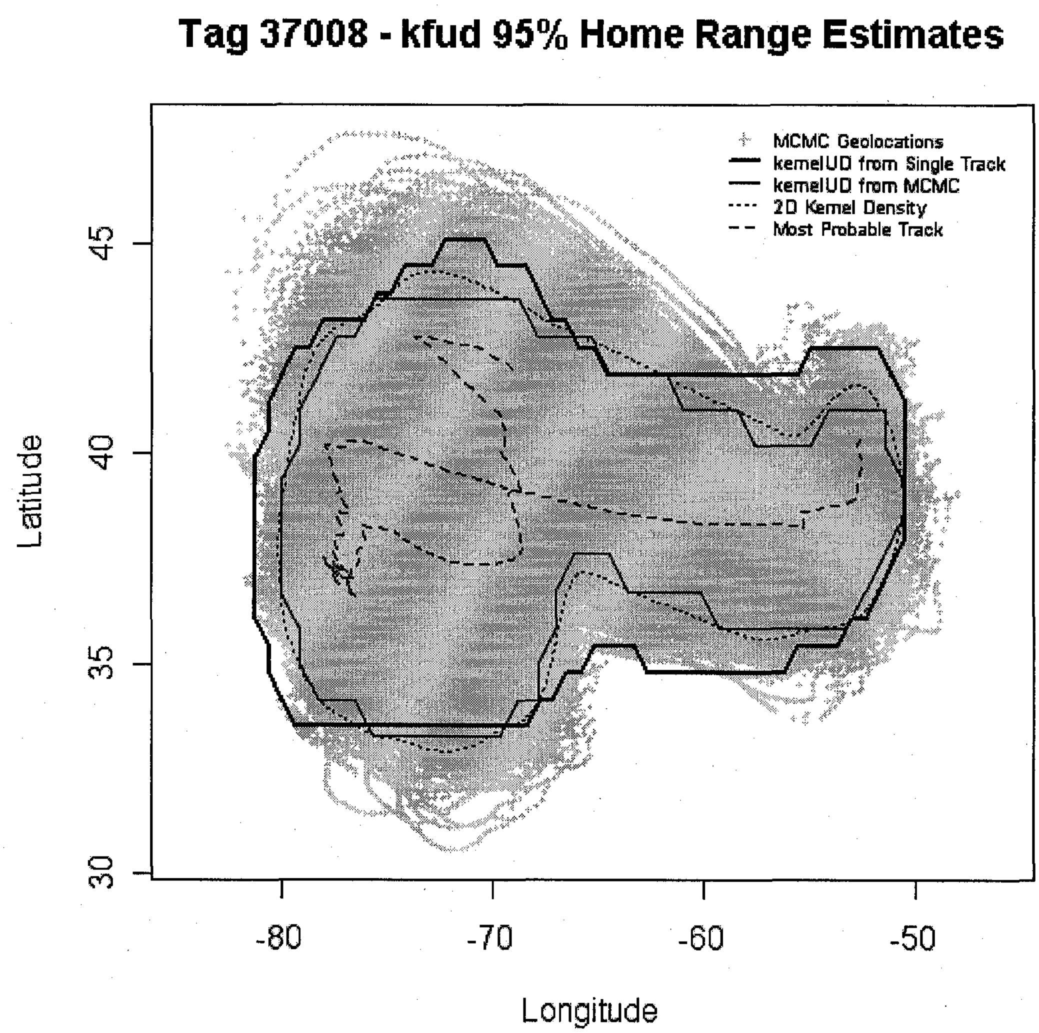


**Figure 17 – *kfud* 95% Home range estimates based on MCMC. The thick line is the *kernelUD* estimate when based solely on the most probable path, while the thin line is the *kernelUD* estimate based on all the MCMC geolocations. The *2D Kernel Density* estimate is also based on all MCMC geolocations.**

4.2.4 Thinned Data

Thinning the original data by removing every other observation provides a demonstration of the sensitivity that the home range estimate has to infrequent data. It was found that lower reporting rates results in substantial shifts in home range estimate (Figure 18), and the nature of that impact can change depending on which observations were removed. The changes, however, do appear to keep the different estimated home ranges within the same general region. This thinning of the data increases the influence that each particular observation has over the estimated track and thus outliers can cause the substantial shifts in the home range estimates observed. Parameter estimation is also susceptible to thinned data, as evidenced by the substantial change in the *D* parameter estimates for the different scenarios of Figure 18. Further thinning of the data, by removing five of every six original days of observations, expectantly inflates the potential for a shifted home range (Figure 19).

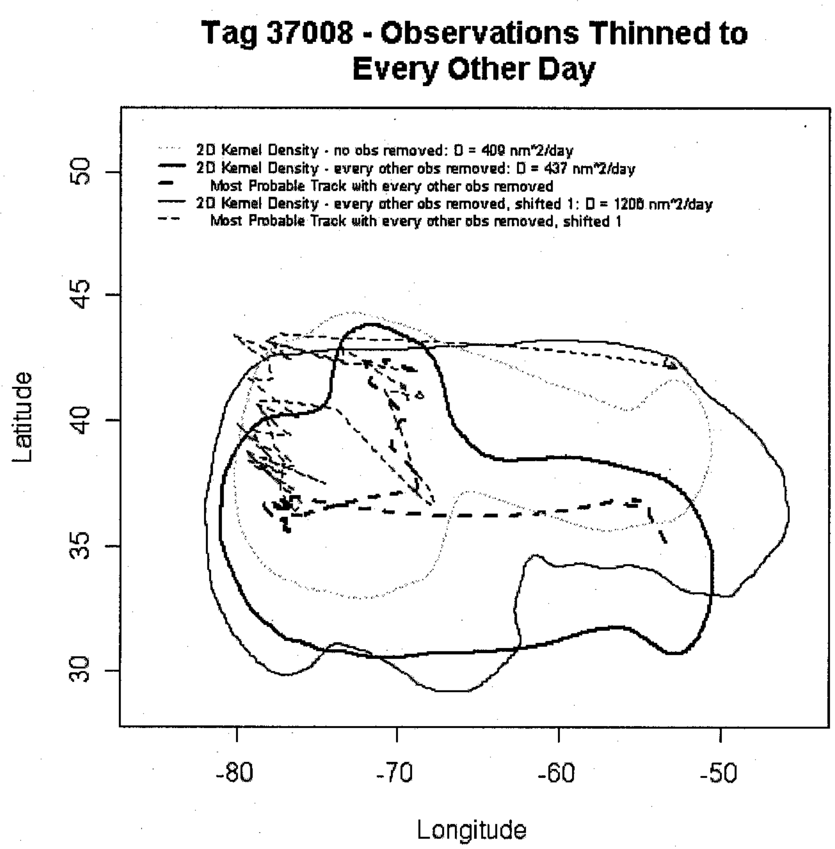


**Figure 18 – 95% Home range estimates for tuna tag 37008. Every other observation was omitted, to see the effect of lower reporting rate.**

## Tag 37008 - 5 of Every 6th Observations Removed

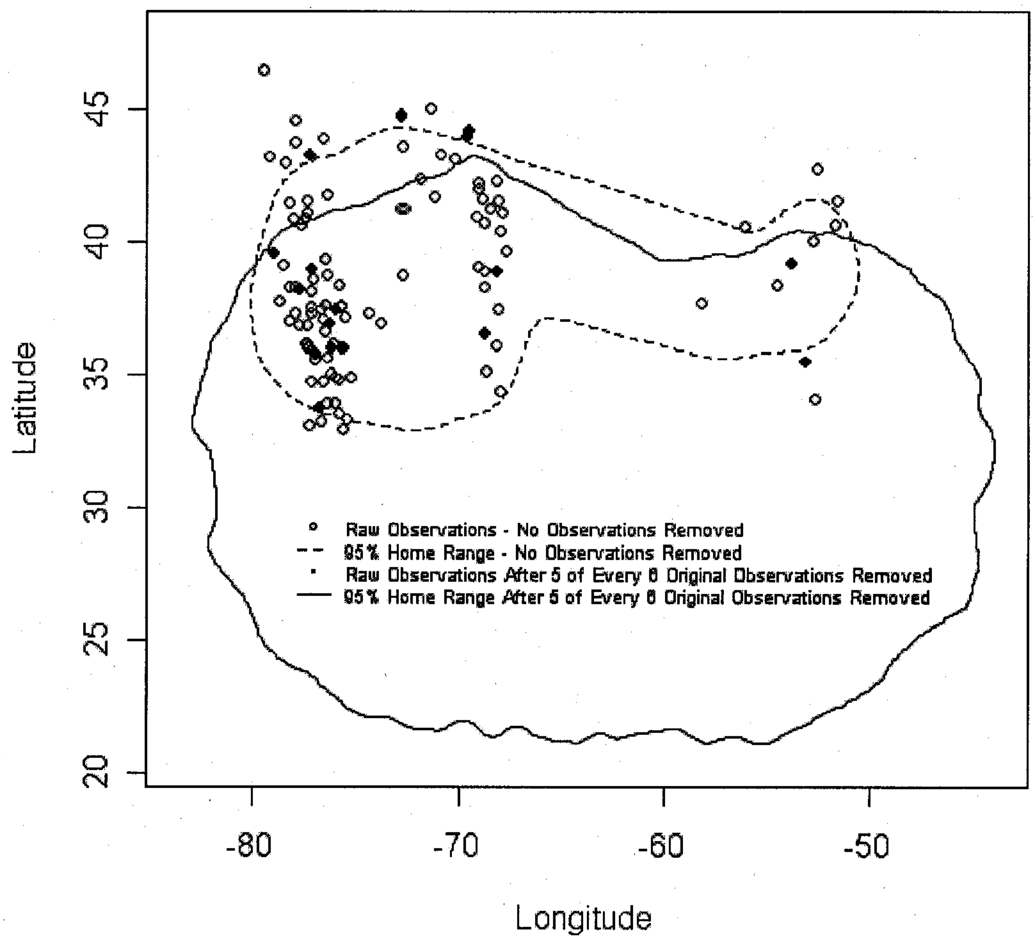


**Figure 19 – 95% home range estimate comparison *between* when all observations are used (black dashed line) and when 5 of every 6 observations are removed (black solid line).**

### 4.2.5 Predicting Multiple Times a Day

Increasing the frequency of geolocation prediction did affect the home range estimate. The 95% home range estimated from two predictions a day (Figure 20) is slightly smaller than the home range estimate based on the same data with 6 geolocation predictions made a day (Figure 21).

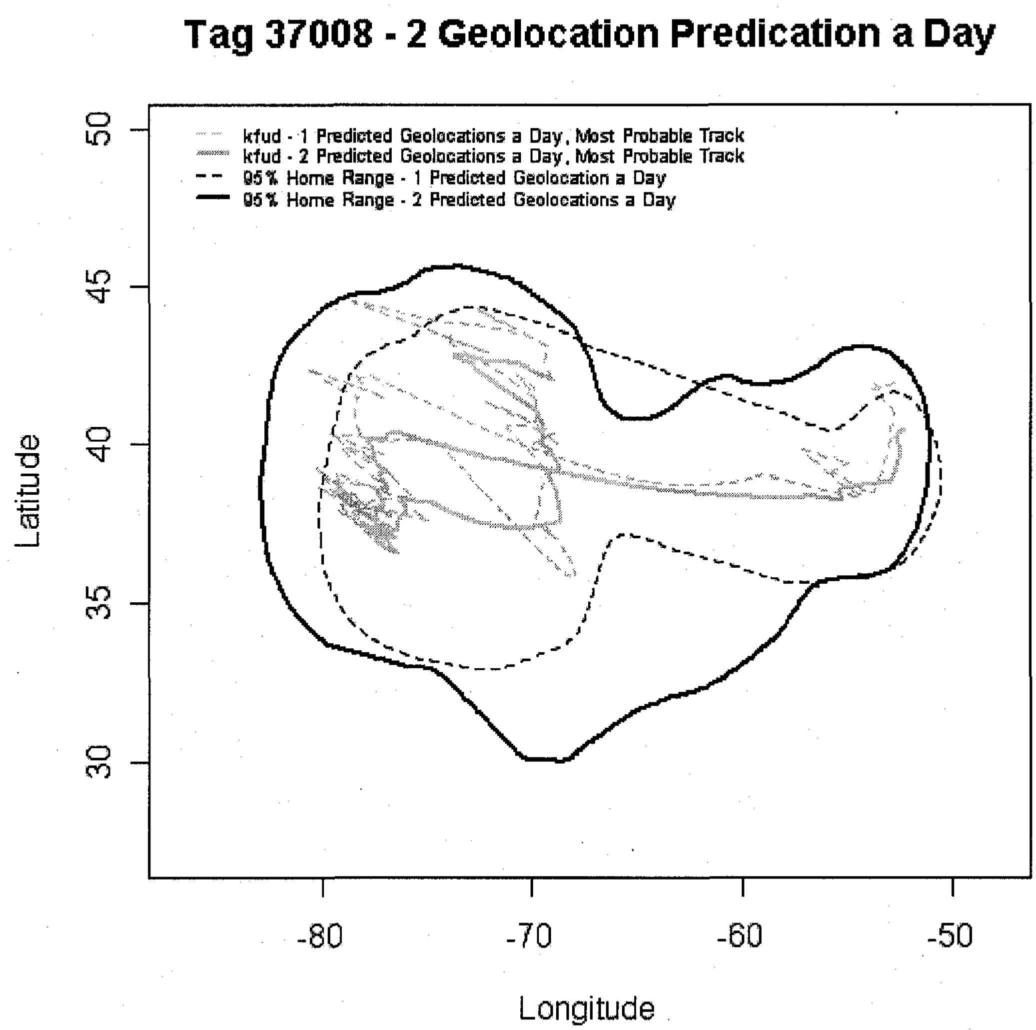


**Figure 20 – Home range estimate comparison between when the tuna's position is predicted 2 times a day (solid black line) rather than once a day (dashed black line).**

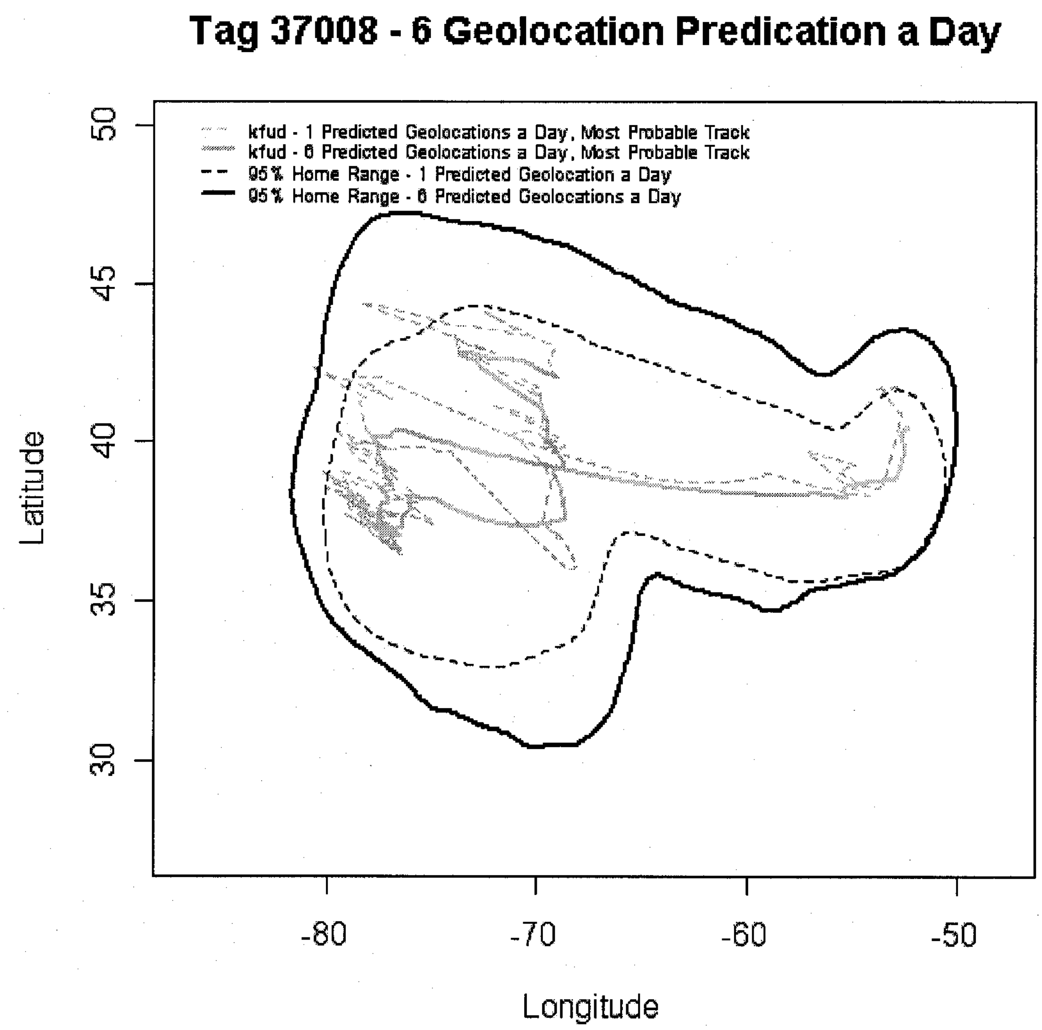


**Figure 21 - Home range estimate comparison between when the tuna's position is predicted 6 times a day (solid black line) rather than once a day (dashed black line).**

Making more predictions of geolocation per day, decreasing the observation to prediction rate, appears to allow the random walk more time to expand the estimated home range. For data that is rarified to 17% of its original number of observations, predicting geolocations four times a day does not appear to result in any appreciable change in the estimate (Figure 22).

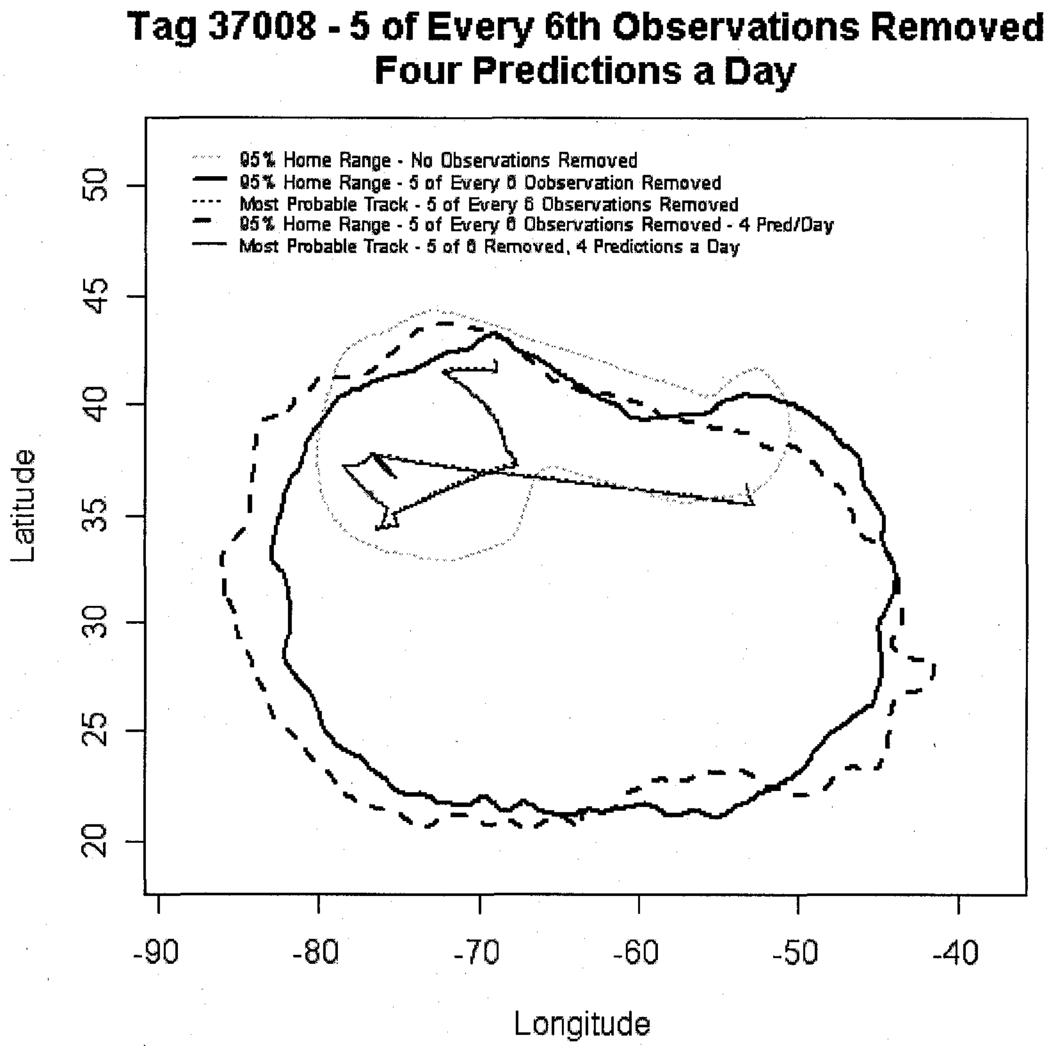


**Figure 22 – 95% Home range estimates for thinned data when using daily and quarter-daily predictions.**

4.2.6 Fixing D

To assess the affect of how this technique might apply if it were being applied to animals of different capacity for travel, *D* was fixed at a high and a low value, relative to the most likely *D* estimated from *kfud*. As expected, if the model is given an absolute prior for *D* that is smaller than what it found to be most likely given the observations, the resulting home range was smaller (Figure 23). Likewise, a larger *D* resulted in a larger home range.

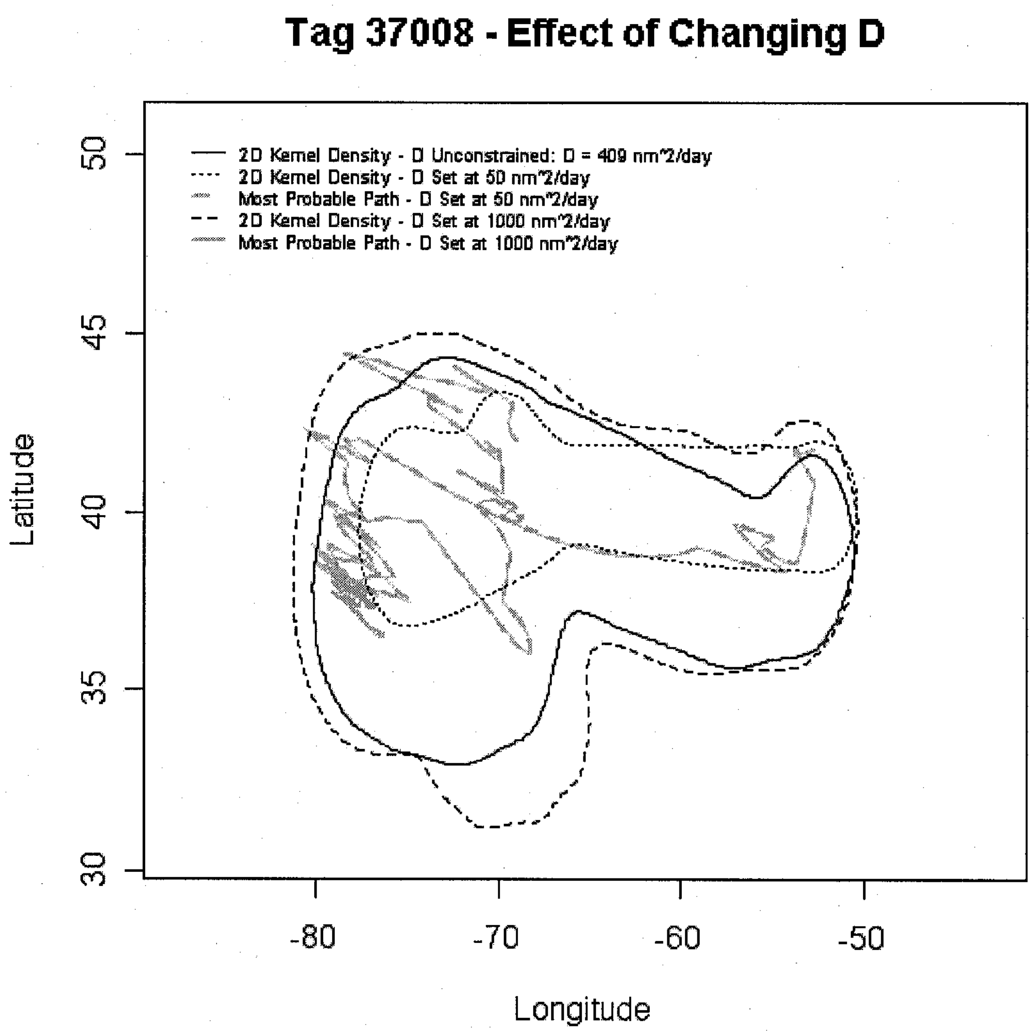


**Figure 23 – Effect of *D* upon 95% home range estimates. 'D unconstrained' indicates that for that track the model estimated the *D* most appropriate for the data. *D* was arbitrarily set for the other two scenarios.**

## 4.3 Application to Bluefin Tuna Data – Tag 37011

### 4.3.1 Prediction Every Day

The observed locations of Tag 37011 provide an opportunity to observe *kfud*'s performance on a track that is relatively centralized, meaning that while tagged, the tuna did not travel great distances between areas of high use. As with Tag 37008, producing geolocation estimates everyday for Tag 37011 through the *kfud* code does alter the overall path from the results of *kftrack* (Figure 24). However, because the observed locations are all within a relatively tight cluster, there are no gaps between hubs of observations. This results in home range estimates that are similar whether the 2D KernelUD smoothing function was applied to the most probable track of *kfud* or *kftrack* (Figure 25).

It is noticeable that again with Tag 37011, as with the previously analyzed track, the specific technique for binning and smoothing the home range based on the MCMC locations does not greatly influence the resulting home range estimate. Figure 27 shows that each of the methods used showed home range estimates that are roughly the same.

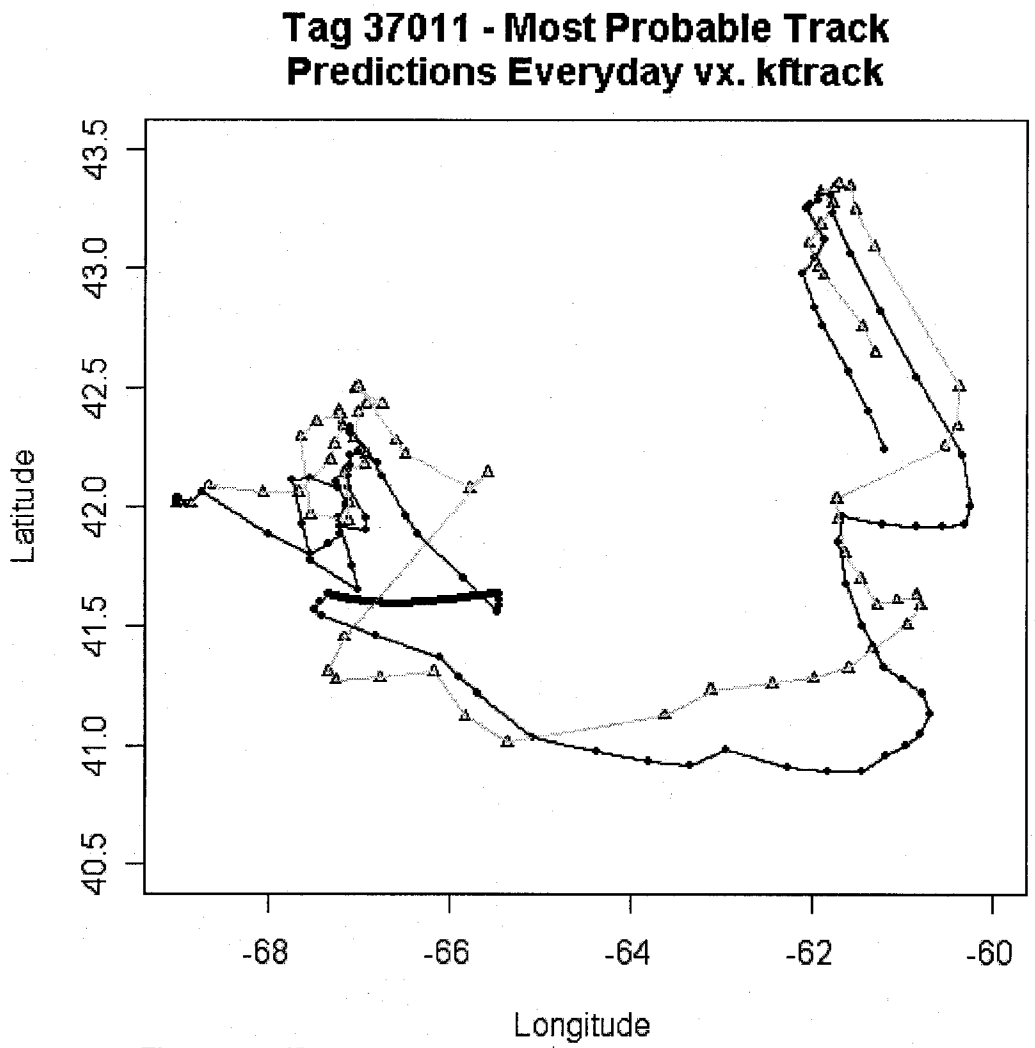

**Figure 24 – Tag 37011 *kftrack* (grey line) vs *kfud* (black line) predicted most probable tracks.**

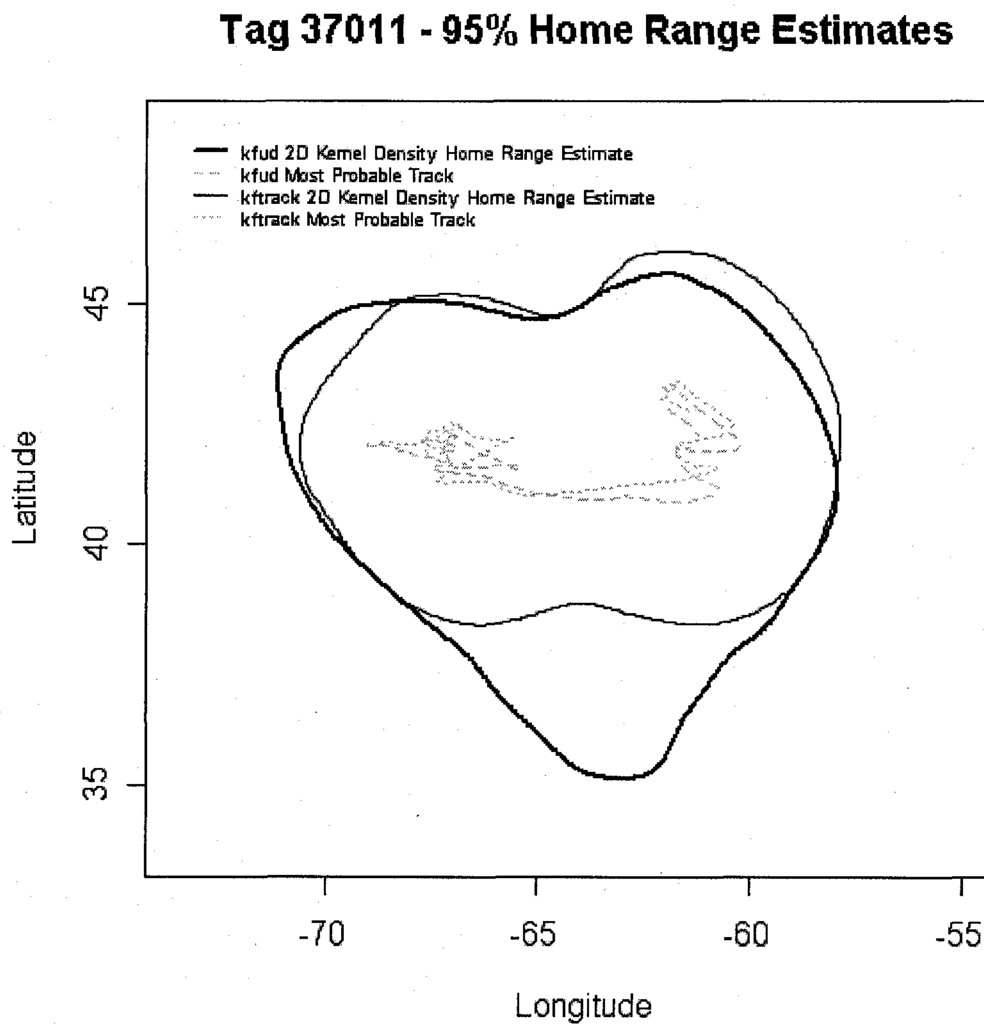


**Figure 25 – 95% Home range estimates based on MCMC runs of the most probable track positions of *kftrack* and *kfud*.**

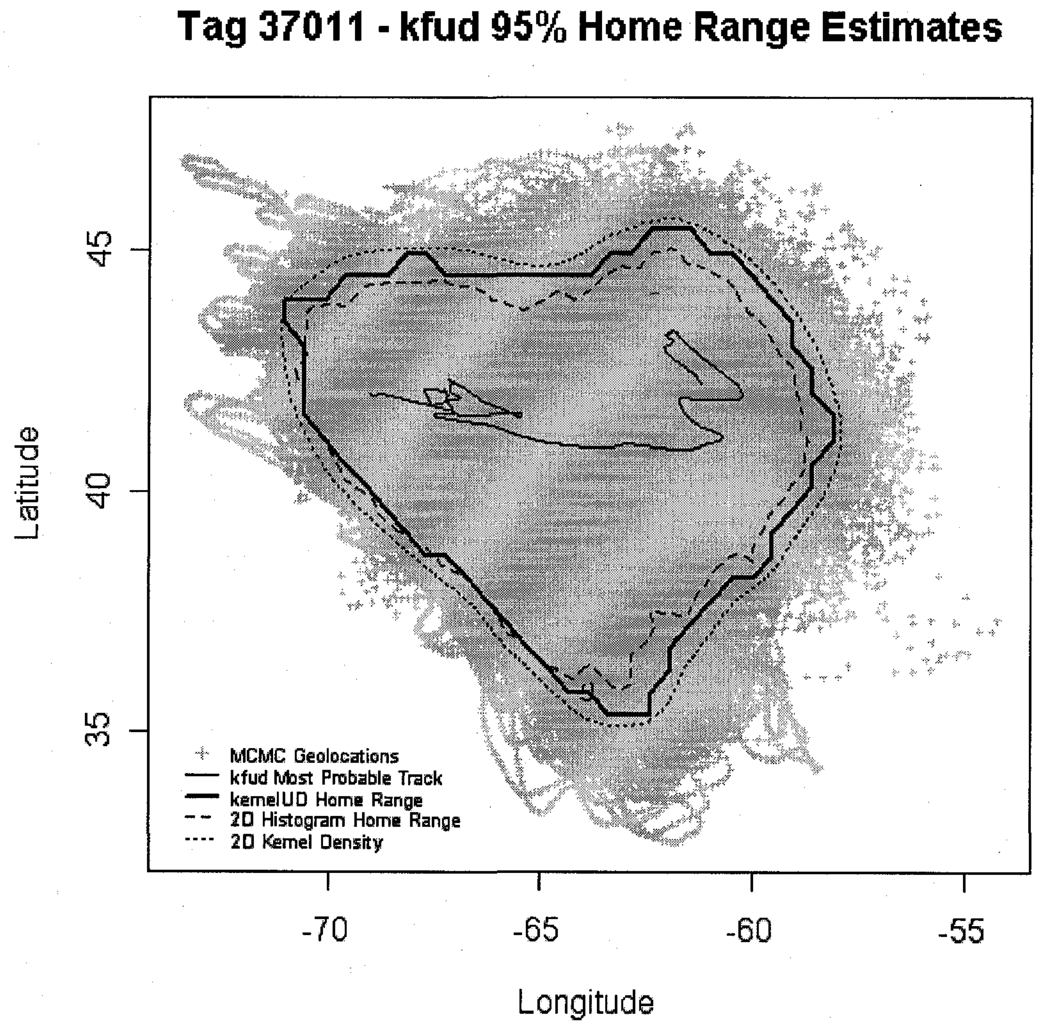


**Figure 26 – 95% home range estimates from the three R functions. Note that all represent very similar home ranges.**

### 4.3.2 Thinned Data

Removing every other observation from the track of Tag 37011 (Figure 27) produced a change to the home range compared to the estimates from a complete data set, similar to what was seen for the previously analyzed track (Tag 37008, Figure 18). For this track, the home ranges estimated in both scenarios were altered in slightly different ways, extending the home range in different directions.

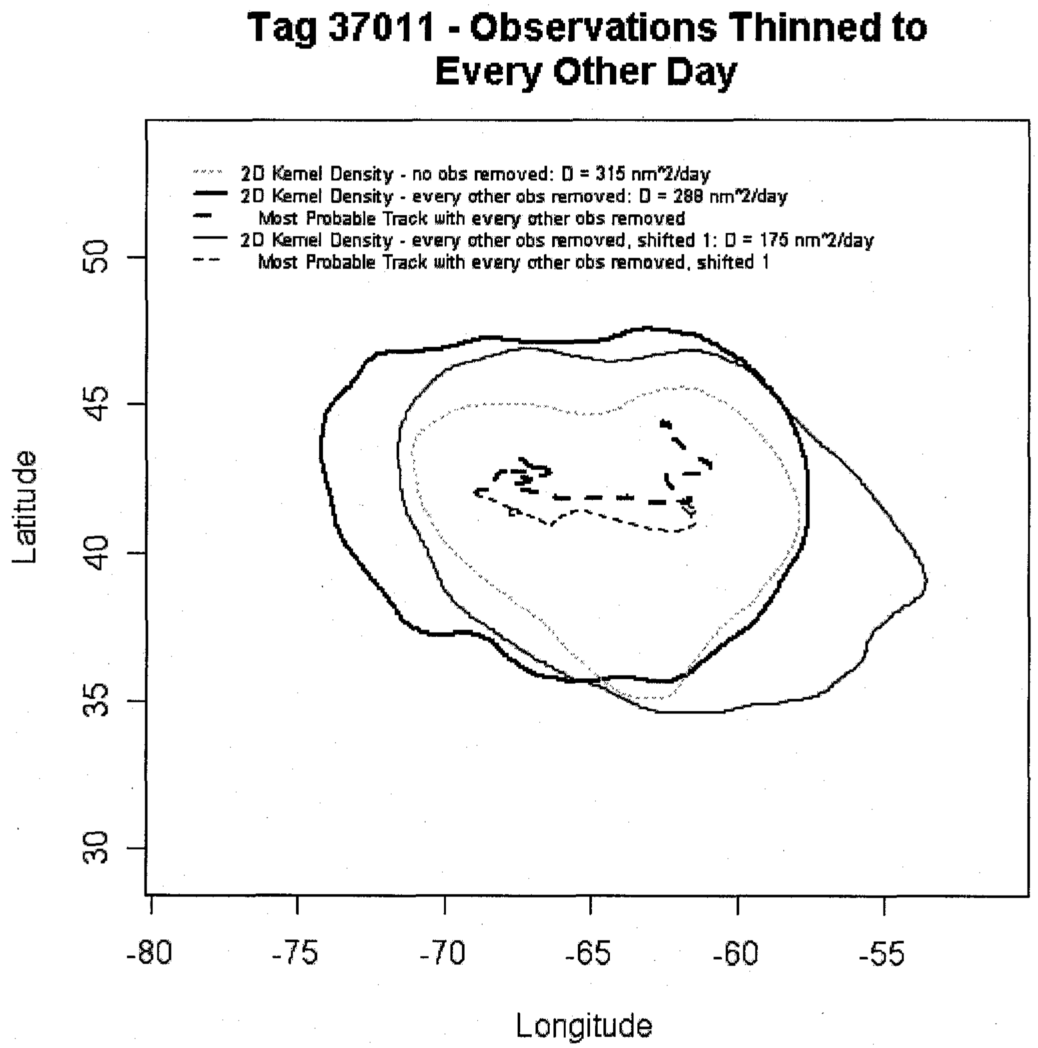


**Figure 27 – 95% Home range estimates when every day's observed location is removed. "Shifted 1" refers to using the observations for the days that the other scenario had removed, thus shifting the observed days by one.**

4.3.3 Predicting Multiple Times a Day

Estimating the location multiple times did have an effect on the estimated home range for Tag 37011 (Figure 28). As more interpolation each day was done (2 times a day vs. 6 times a day), the estimated home range increased in size. This may have something to do with the random walk having more occasions with which to randomly wander away from the most probable path.

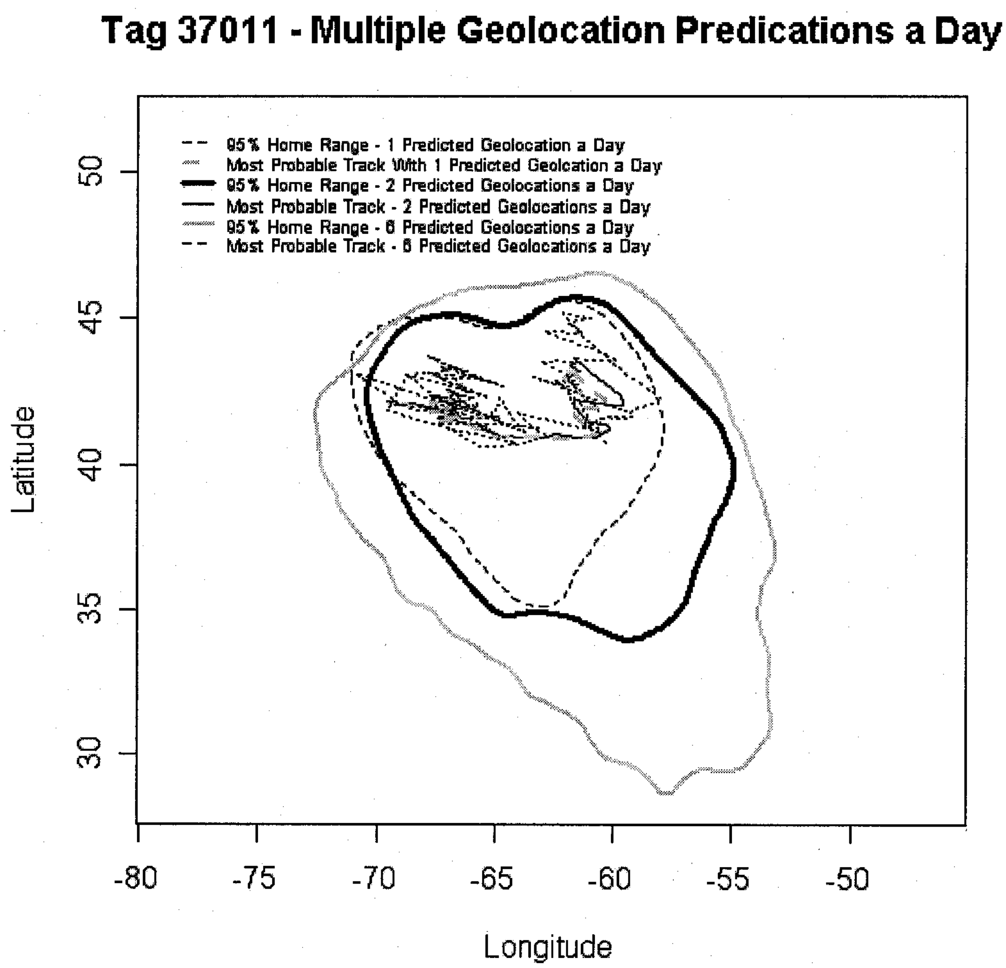


**Figure 28 – 95% Home range estimates for Tag 37011 with multiple positions estimated each day.**

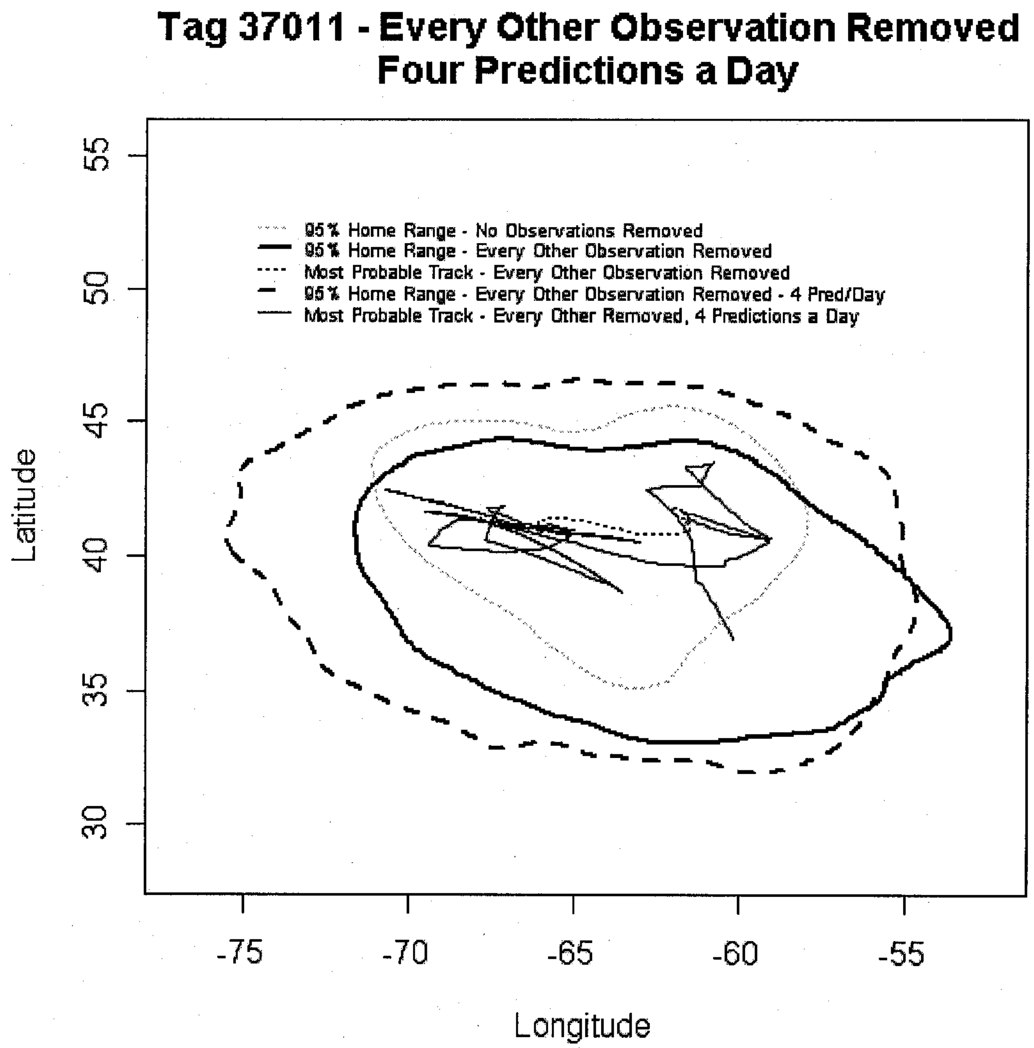


**Figure 29 – 95% Home range estimates for Tag 37011 with the data thinned and geolocation estimates made four times a day.**

## 4.3.4 Fixing D

Setting *D* at less than a sixth of what was estimated to be appropriate for *D* shrunk the home range estimate substantially, as is what would be expected (Figure 30). Like Tag 37008, a larger *D* did indeed translate into a larger home range. With an enlarging *D*, the model assumes the tuna can go farther than it estimated, and therefore the tuna can cover more area in a day. This ends up enlarging the home range estimate. With a small *D*, the model conditions more on D than the observations, operating under the assumption that the animal is so slow it would not be able to reach the next day's observation as the observed geolocations were reported. This in effect shrinks the distance the animal can travel, and in turn shrinks the home range estimate. With a large *D*, the random walk is allowed to go a greater distance than it otherwise may have, but is not conditioned on doing so. Therefore the home range is only slightly larger than it is when *D* is not declared to be a certain value.

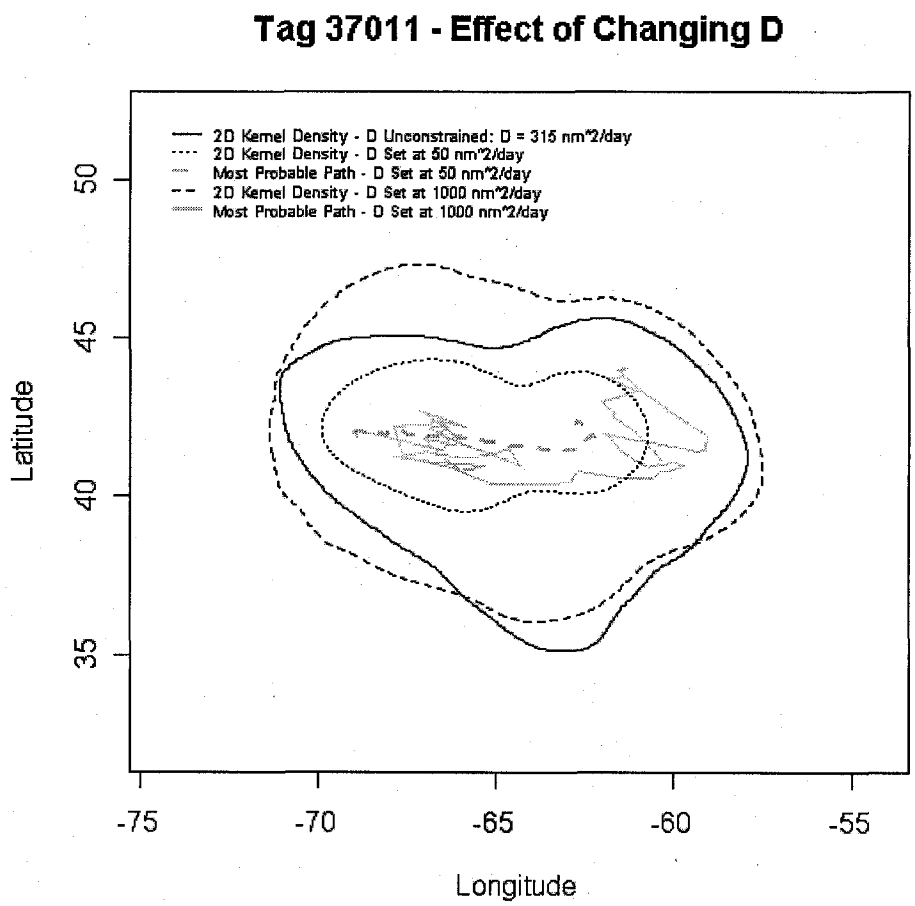


**Figure 30 – 95% Home range estimates for Tag 37011 when D is constrained to relatively large and small values.**

## 4.4 Application to a Track with Low Measurement Error – Tag 3817

### 4.4.1 Prediction Every Day

The Doppler derived geolocations (Tag 3817), with data successfully collected every day, have very low measurement error. This provides a relatively precise track on which to perform the developed UD analysis. The mean squared errors in estimating longitude and latitude were estimated to be $1.72 \times 10^{-5}$ and $6.87 \times 10^{-6}$ degrees, respectively. For comparison, the estimated mean squared errors of longitude and latitude of Tag 37008 were 1.14 and 2.22 degrees, respectively. This low amount of uncertainty in the positions leads to a relatively small home range (Figure 31).

### 4.4.2 Thinned Data

Due to the precision of the positions of Tag 3817, and the relatively straight path it traveled during its observation, removing alternate portions of the data did not appear to substantially change the home range estimates (Figure 32).

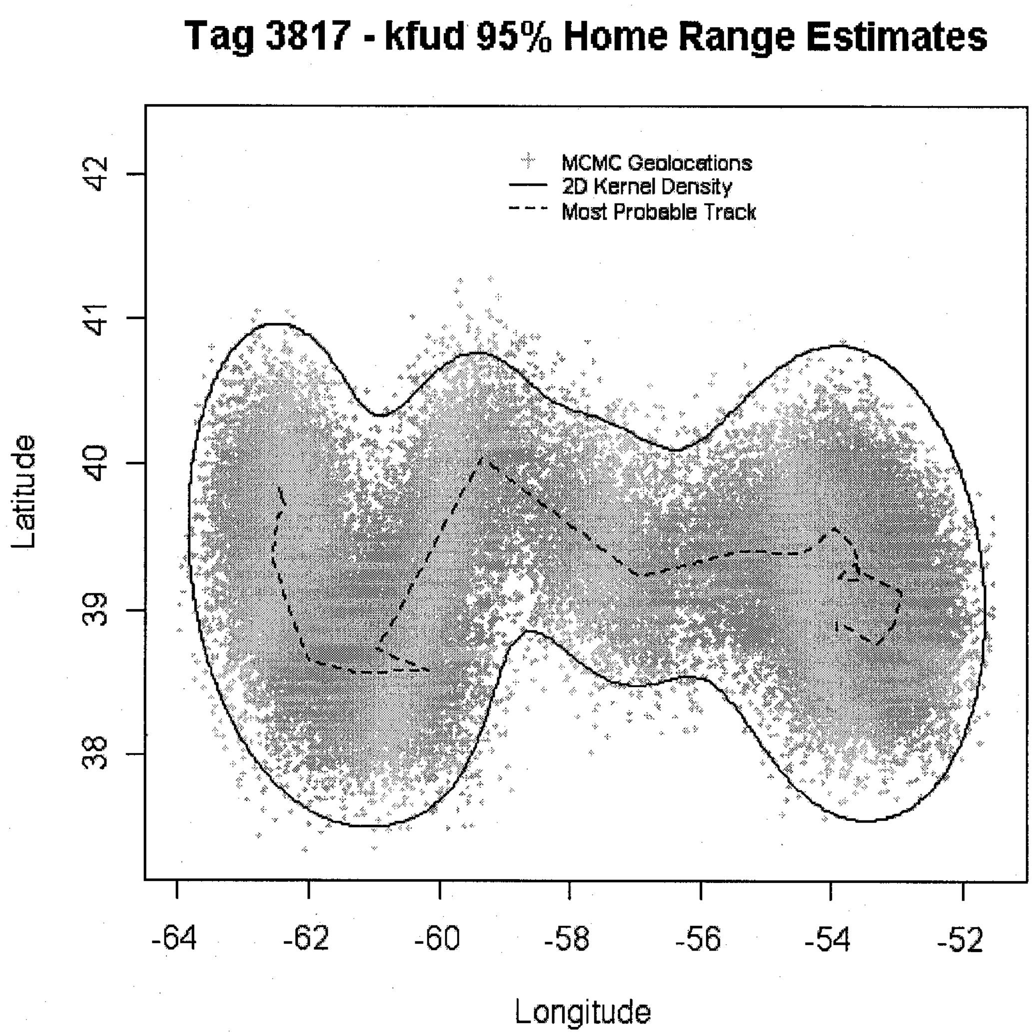


**Figure 31 – 95% home range estimates of Tag 3817 *after* releasing from a tuna. Locations were observed via Doppler positioning, and have very little error.**

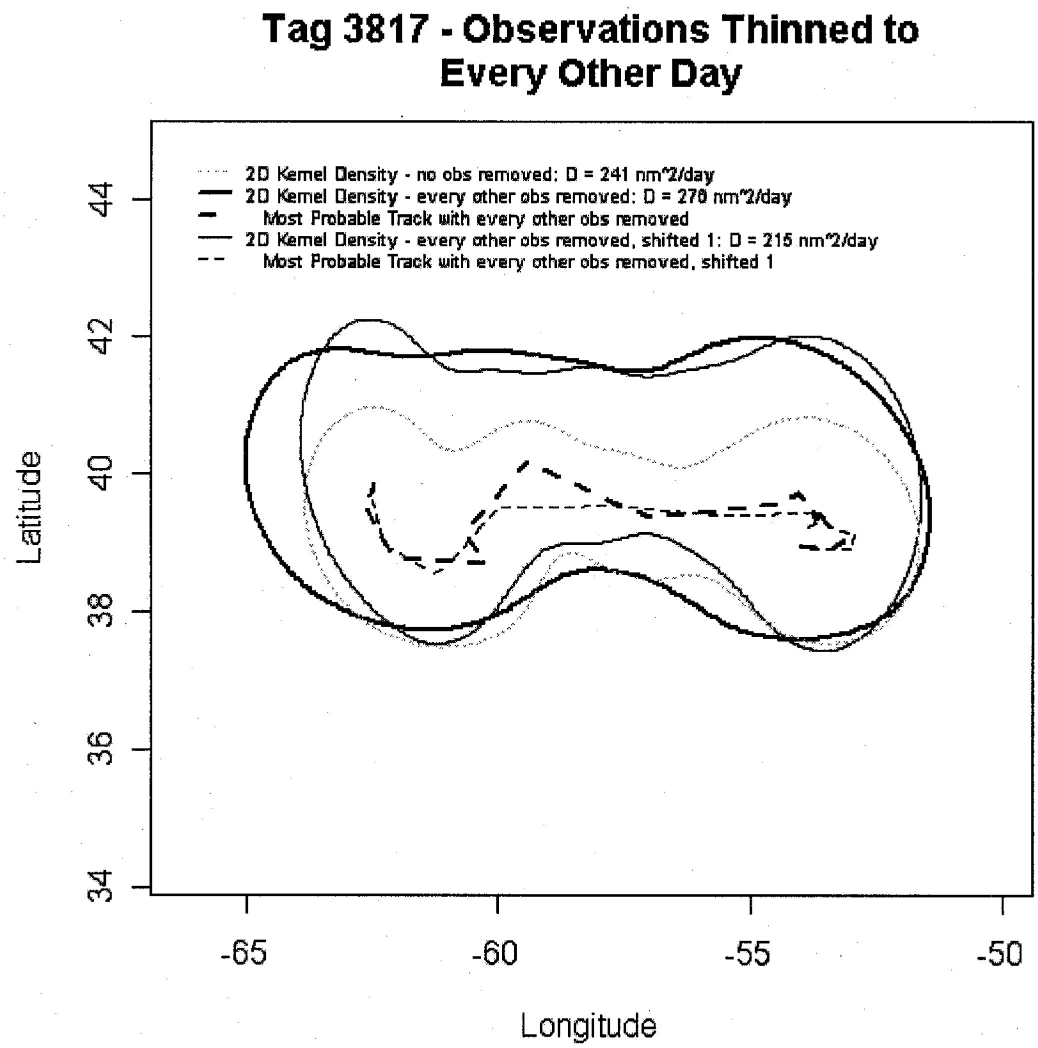


**Figure 32 – 95% home range estimates for Tag 3817 with every other observation omitted.**

### 4.4.3 Fixing D

With the straight path and accurate measurements of Tag 3817, changing the *D* parameter also did not cause a great deal of change in the resulting home range estimates (Figure 33). However, some differences can be noted. The smaller *D* doesn't allow the tag to be modeled as traveling as far as the observations indicated. The associated home range is thus forced to shorten longitudinally. This shows how prior knowledge about the mobility capabilities of the 'animal' can be used to influence the estimated home range.

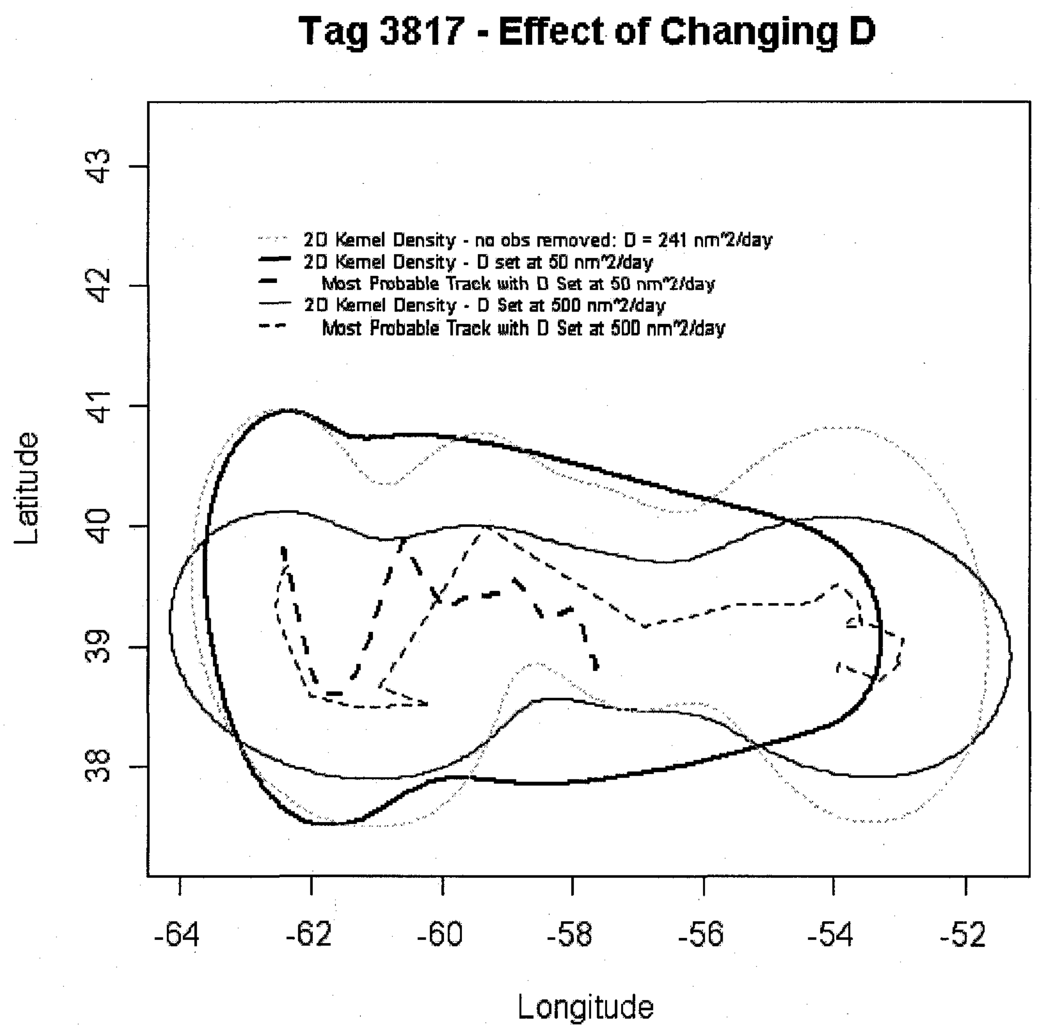


**Figure 33 – 95% home range estimates with changing the *D* parameter.**

## 4.5 Utilization Distribution

It should be remembered that the 95% home range estimates used in these comparisons are only a portion of the power provided by a method that produces a UD. The contour plot of the UD for Tag 37008 and 37011 (Figure 34 and Figure 35) clearly indicates areas that the tuna used frequently, and the floating Tag 3817 (Figure 36) shows increased precision in the estimate due to the increased precision in observations. Other areas that were likely only used for travel by the tuna can be noted in the less restricted contours of the UD such as the 95% and 99% contours, but are omitted from the more exclusive features of the plot. This detail of how the tuna uses its home range is necessary to insightfully correlate probabilistic space use to environmental factors.

## Tag 37008 - Utilization Distribution

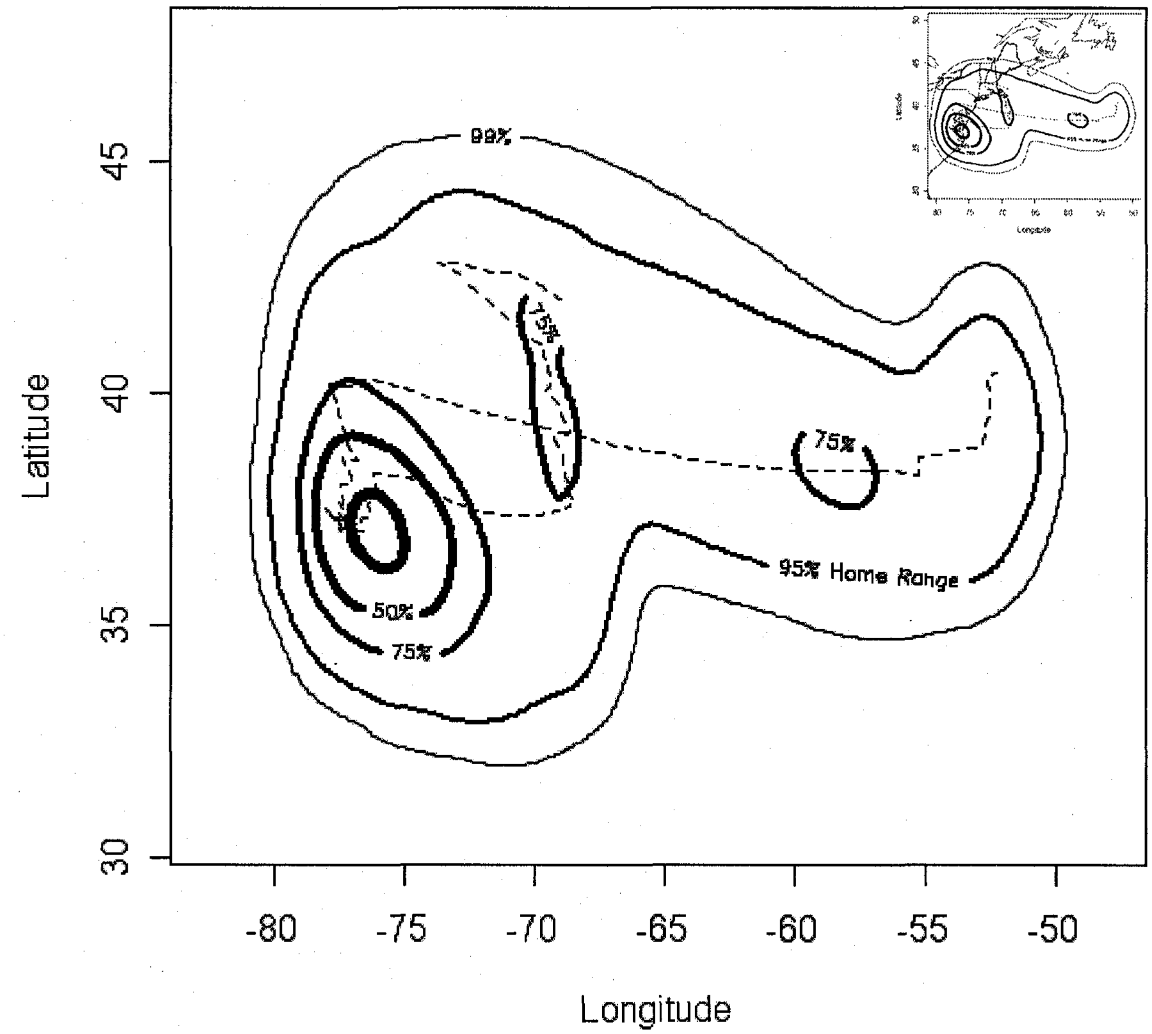


**Figure 34 – Utilization distribution (UD) of tuna 37008. Each contour label corresponds to the percent of the MCMC positions that reside inside the contour.**

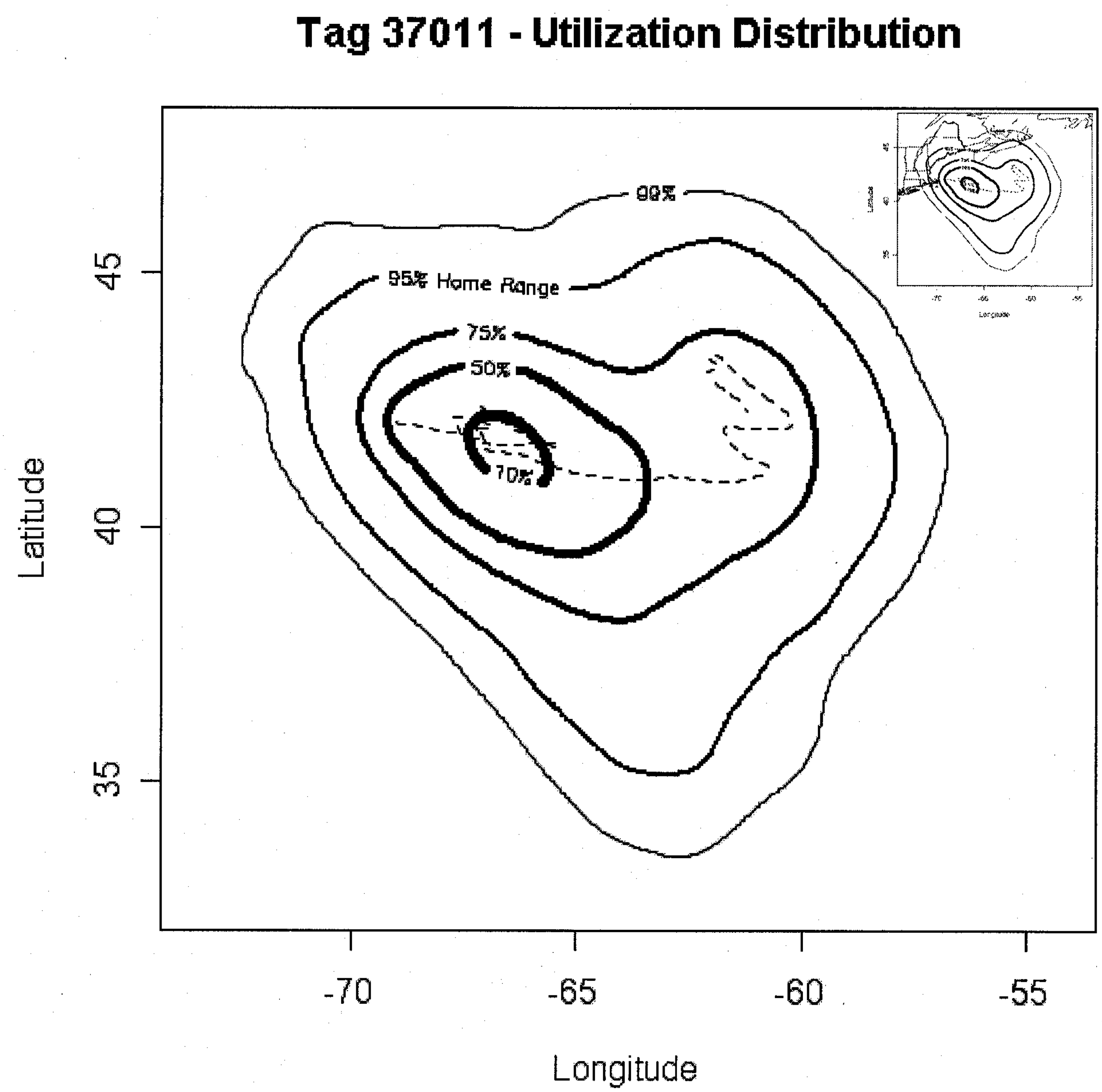


**Figure 35 – Utilization distribution (UD) of tuna 37011. Each contour label corresponds to the percent of the MCMC positions that reside inside the contour.**

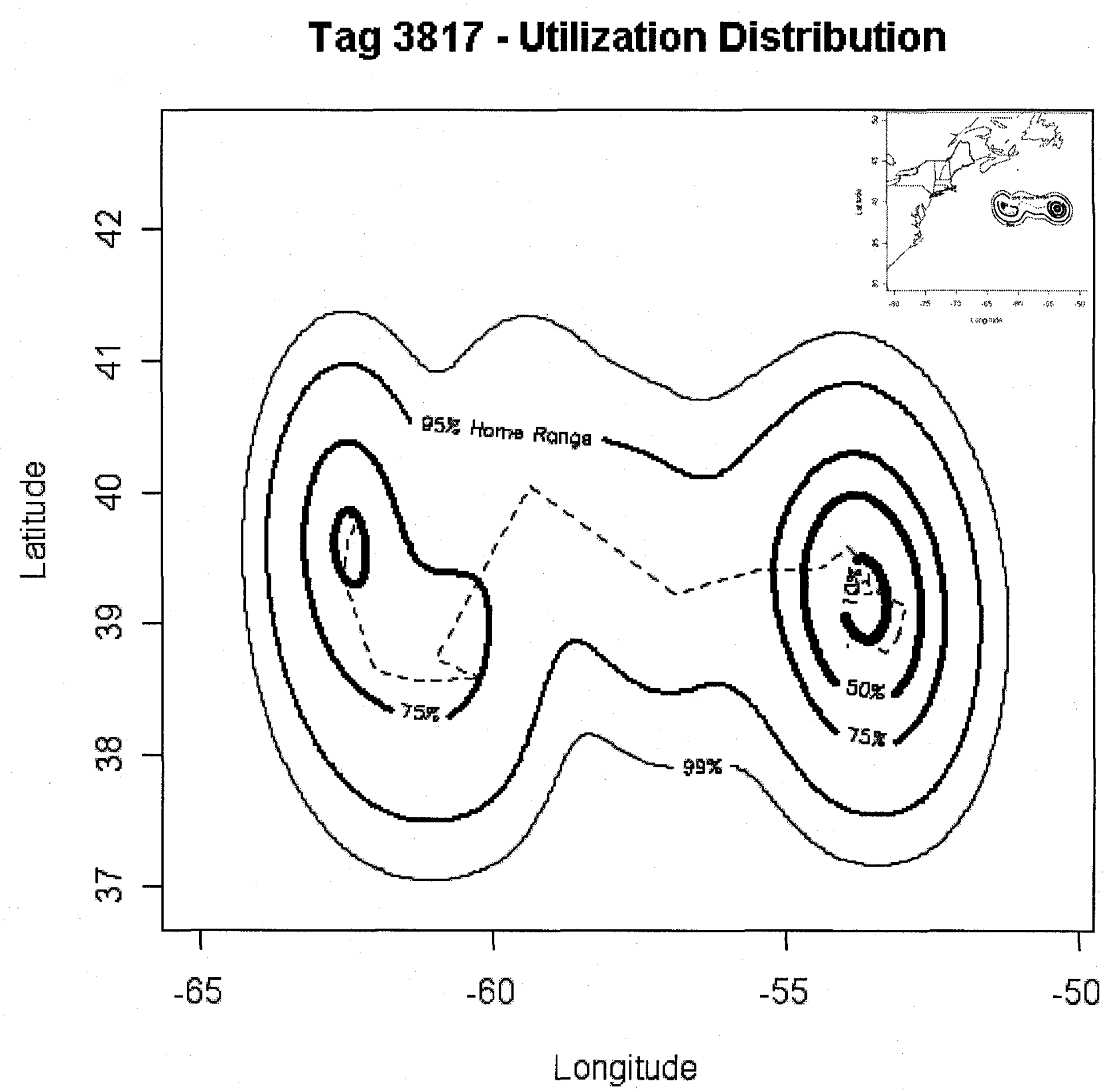


**Figure 36 – Utilization distribution (UD) of tuna 3817. Each contour label corresponds to the percent of the MCMC positions that reside inside the contour.**

# CHAPTER 5

# DISCUSSION

The developed UD estimator provides a marked advantage over current methods. With other methods, if there are spatial gaps in the observations, there can be gaps in the UD estimate (Figure 13). Gaps in the UD can be misleading, as they give the impression that the areas of these gaps, through which the animal must have traveled at some point, are unimportant habitat for the animal. The importance of corridors through which animals travel from one area of localized activity to another is well understood (Eggleston and Dahlgren 2001, Mumby 2006, Rouget *et al.* 2006) and should not be overlooked when analyzing habitat use. Furthermore, estimations of home ranges as those represented in Figure 13 can lead to the assumption that the area between the defined home ranges was only omitted from the estimate because the tuna traveled through it quickly, utilizing the area sparsely in both space and time. Yet when the large temporal gaps exist between subsequent observations such as those for Tag 37008 (Figure 7), it is not adequate to view the gap in observations as indicative of a gap in actual use. The tuna may have traveled between the two areas with dense observations directly and rapidly, or it may have meandered using much of the space not included in the home range estimate. Also the temporal sequence of the observations is an important consideration. If a series of sequential observations alternate between the hubs of observations, then it can be assumed that the gap must have been traversed more frequently than if the gap was traversed only once. The adjustments made to *kftrack* to create *kfud,* which made estimates of geolocation for all days regardless of the presence of an observation, allow for the potential use of the connecting areas by the tuna to be included in the home range estimate (Figure 15).

## 5.1 Effects of Observation Frequency

The effect of missing data is made apparent in Figure 19 where a sub-sampling of one of every six observations originally made, resulted in a home range that was both phased southward and is less precise. Another example of the effect missing observations can have is shown in Figure 18. In this case, every other observation was ignored. Note that the home ranges estimated are out of phase, but the direction of the phase shift depends on if every other observation was omitted beginning with the $2^{nd}$ or the $3^{rd}$ observation. The missing values can also substantially alter estimates of individual parameters. The parameter *D* in this case was dramatically affected, influencing the daily movement rates we would have derived from the analysis depending on which set of observations were actually made. When observations that were not ignored happened to be close to each other (average distance 9.83 nm/day), the daily displacement of the tuna was estimated to be low, and in turn, *D* was small (437 $nm^2$/day), leading to a more confined home range estimate. When the retained observations were far apart given the amount of time between them (average distance 21.03 nm/day), *D* became larger (1206 $nm^2$/day) and the home range was slightly broader. Generally, if a UD estimate is inflated, perhaps due to uncertainty in observed geolocations, the resulting estimate may not be precise, but it would likely be less accurate if the data was rarified. Therefore, it seems tracking an animal via frequent observations is of great importance when estimating a UD.

## 5.2 Increasing Geolocation Prediction Rate

For this study, an effect on the most probable track and the UD was observed for a prediction frequency of both 2 and 6 times a day. The 95% home range was not affected drastically by increased frequency of prediction beyond once a day (Figure 20 and Figure 21). The home range did expand slightly upon each increase of frequency beyond once a day. It is not clear if increasing prediction frequency to a time scale more minute than the observations is of benefit. With each interpolated prediction, the model has greater opportunity for the model to wander. This thesis employed an unbiased random walk in *kfud* as the theoretical movement model of the tuna. If a biased random walk was instead used, it might mitigate this increased home range size so increasing interpolation frequency will not continually expand the home range estimate.

## 5.3 Effect of No Measurement Error

Many forms of tracking an animal are currently available for a variety of habitats, each with its own temporal observation structure, and with different degrees of measurement error. Indeed, some tags such as Argos or GPS have very little measurement error relative to PSAT tags that use light-based geolocation (Galuardi 2006). To simulate how the home range estimation method behaves operating on geolocations with very little measurement error, the *kfud* approach was applied to Argos positions. Lower measurement error substantially reduced the uncertainty in the home range (Figure 31), and indeed the effects of rarefied data (Figure 32). When observations are known with little error, removing portions of the original data does affect the UD estimates, but the magnitude of the shifts seen for Tags 37008 and 37011 were not seen in 3817. It may then be suggested that while frequent observations is beneficial to this UD estimation technique, directionally biased estimates can generally be avoided if using positions known with very little error. The priority, therefore, should be to gather position data in as accurately a way as possible, even if for some reason it requires a sacrifice in observation frequency. With accurate observations, the Kalman filter based UD estimate may be robust to sample size and frequency.

## 5.4 Model Performance

The *kfud* model used an unbiased random walk to model movement, and as mentioned earlier, it would have been preferable to have used a biased random walk. Unfortunately, unless the biases of the movement models (parameters *u* and *v*) are set to zero, the model proves unstable during MCMC analysis, driving *D* exponentially larger with each subsequent run of the MCMC. The *–mcdiag* option was also used in the MCMC command of ADMB, without which *D* also would inflate during the MCMC far beyond the value estimated for the most probable track. This command option produces a diagonal covariance matrix, removing any covariance between the parameters. With these details addressed, the MCMC behaved as expected with parameters not drifting during the sequence of runs.

# CHAPTER 6

## CONCLUSIONS AND THE NEXT STEP

Estimating an animal's UD through application of the Kalman filter and MCMC is successful in producing realistic estimates. The estimates are sensitive to the serial correlation of the observations, the daily migratory abilities of the species being monitored, and the uncertainty of the geolocation estimates. However, the large amount of uncertainty associated with light-based geolocation demands a frequent observation structure, with very few missing days of observations. As data becomes more rarefied, the likelihood of imprecise and inaccurate UD estimates increases. Geolocations with greater accuracy diminishes the UD estimator's sensitivity to rarefied data.

One of the problems encountered from the inaccurate measurements is that the geolocation and associated UD are often estimated to be on land. To make the UD estimate truly realistic, it would need to not include land in its estimate for tuna, or any other strictly marine or aquatic animal. The UD estimate presented here does not resolve this issue, however one promising method has been developed to change the filter to exclude the possibility of the tuna traveling on land. Work underway at the UNH Center for Large Pelagics and by John Sibert's Pelagic Fisheries Research Program at the University of Hawaii at Manoa incorporates the extra data gathered by the PSATs such as temperature and dive depth to improve the accuracy of the tag's locations and tracks (Royer *et al.* 2005, Galuardi 2006, Nielsen *et al.* 2006, Royer *et al.* 2006). This can help restrict tag location estimates to the water. The tags record temperature on an hourly basis. This temperature data can then be used to correct for the inherently inflated latitudinal geolocation errors through comparison with sea surface temperatures (SST) gradients in the area. This approach would exclude land from being estimated as a possible area of use because there are no SSTs on land to match the temperature read by the tag.

Many features of the environment that may restrict movement of an animal are not so readily identified. Accurate and frequent observations are the best way to restrict home range estimates from extending into areas that are not potential habitats for the animal. For tuna, methods such as using SST or bathymetry to improve accuracy of the geolocation estimates and to block the model from going on land, will lead to a more realistic and informative utilization distribution estimate.

This UD estimation approach can be modified to fit a large number of situations. The code developed for this study, *kfud*, can be applied to any animal for which a time-series of locations has been observed. *Kfud* produces UD estimates that are more connected, informative, and unique to the animal's capabilities. Further modification of the Kalman filter can be in the application of a third dimension, as has been done with SST (Royer *et al.* 2005, Galuardi 2006, Nielsen *et al.* 2006, Royer *et al.* 2006), or into a more realistic movement model such as the biased random walk. The filter itself could be altered as well. This code utilizes the Extended Kalman Filter. However, there are several other versions of the Kalman filter that have been developed such as the Ensemble (Evensen 2003), Unscented (Julier and Uhlmann 1997), and Particle (Bolviken and Storvik 2001) that can, and have been applied to correct geolocation estimates (Royer *et al.* 2006). The filers may aid in the further reduction of geolocation error, and thus cause an increase in the robust quality of a resulting UD. Finally, with the more realistic nature of the UD estimation developed in this thesis, more accurate and illuminating correlations between utilization distributions and the environmental factors influencing the animal's movements can be pursued.

# LIST OF REFERENCES

Adams, E.S. 2001. Approaches to the study of territory size and shape. *Annual Review of Ecology and Systematics* 32: 277-303.

Anderson, D.J. 1982. The home range: a new nonparametric estimation technique. *Ecology* 63: 103-112.

Anderson-Sprecher, R. and J. Ledolter. 1991. State-Space Analysis of Wildlife Telemetry Data. *Journal of the American Statistical Association* 86(415): 596-602.

ARGOS. 1996. Argos User's Manual. CLS (Service Argos), Toulouse, France.

Bethke, R., M. Taylor, S. Amstrup and F. Messier. 1996. Population delineation of polar bears using satellite collar data. *Ecological Applications* 6(1): 311-317.

Bigelow, H.B. and W.C. Schroeder. 1953. Fishes of the Gulf of Maine. *Fishery Bulletin (Washington DC)* 53: 1-577.

Bolviken, E. and G. Storvik. 2001. Deterministic and stochastic Particle filters in State-Space model. *In*: Doucet, A. and N. de Freitas (eds) Sequential Monte Carlo Methods in Practice. Springer-Verlag, NY.

Brill, R.W., M. Lutcavage, G. Metzger, P. Bushnell, M. Arendt, J. Lucy, C. Watson and D. Foley. 2002. Horizontal and vertical movements of juvenile bluefin tuna (*Thunnus thynnus*), in relation to oceanographic conditions of the western North Atlantic, determined with ultrasonic telemetry. *Fishery Bulletin* 100(2): 155-167.

Brill, R.W. and M.E. Lutcavage. 2001. Understanding environmental influences on movements and depth distributions of tunas and billfishes can significantly improve population assessments. *American Fisheries Society Symposium* 25: 179-198.

Bullard, R. 1991. Estimating the home range of an animal: a Brownian bridge approach. Master of Science, University of North Carolina, Chapel Hill.

Burt, W.H. 1943. Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy* 24: 346-352.

Butterworth, D.S. and A.E. Punt. 1993. The robustness of estimates of stock status for the western north Atlantic bluefin tuna population to violations of the assumptions underlying the associated assessments models. *ICCAT Collective Volume of Scientific Papers* 43(1): 192-210.

Calhoun, J.B. and J.U. Casby. 1958. Calculation of home range and density of small mammals. *Public Health Monograph* No. 55. Washington, D.C.: US Government Printing Office.

Chase, B.C. 2002. Differences in diet of Atlantic bluefin tuna (*Thunnus thynnus*) at five seasonal feeding grounds on the New England continental shelf. *Fishery Bulletin* 100(2): 168-180.

Crane, J. 1936. Notes on the biology and ecology of giant tuna, *Thunnus thynnus* Linnaeus, observed at Portland, Maine. *Zoologica* 21: 207-211.

De Metrio, G., G.P. Arnold, J.M. de la Serna, P. Megalofonou, G. Sylos Labini, M. Deflorio, A. Buckley, J.L. Cort, C. Yannopoulos and M. Pappalepore. 2003. Movements and migrations of North Atlantic Bluefin tuna tagged with pop-up satellite tags. In: Spedicato, M.T., G. Lembo, G. Marmulla (eds). *Aquatic telemetry: advances and applications*. Proceedings of the Fifth Conference on Fish Telemetry held in Europe, Ustica, Italy, 9-13 June 2003: 161-168.

Dixon, K.R. and J.A. Chapman. 1980. Harmoic Mean Measure of Animal Activity Areas. *Ecology* 61(5): 1040-1044.

Eggleston, D.B. and C.P. Dahlgren. 2001. Distribution and abundance of Caribbean spiny lobsters in the Key West National Wildlife Refuge: relationship to habitat features and impact of an intensive recreational fishery. *Marine and Freshwater Research* 52(8): 1567-1576.

Ekstrom, P. 2004. An advance in geolocation by light. *Memoirs of the National Institute of Polar Research* 58: 210-226.

Estrada, J.A., M. Lutcavage and S.R. Thorrold. 2005. Diet and trophic position of Atlantic bluefin tuna (*Thunnus thynnus*) inferred from stable carbon and nitrogen isotope analysis. *Marine Biology* 147(1): 37-45.

Evensen, G. 2003. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics* 53: 343-367.

Feller, W. 1966. An Introduction to Probability Theory and its Applications, Vol. II. New York: John Wiley & Sons.

Feller, W. 1968. An Introduction to Probability Theory and its Applications, Vol. I. (3rd edn). New York: John Wiley & Sons: 509pp.

Fitch, H.S. 1958. Home ranges, territories, and seasonal movements of vertebrates of the Natural History Reservation. *University of Kansas Publications, Museum of Natural History* 11(3): 63-326.

Fromentin, J. and L.T. Kell. 2007. Consequences of variations in carrying capacity or migration for the perception of Atlantic bluefin tuna (*Thunnus thynnus*) population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences* 64(5): 827-836.

Fromentin, J. and J. Powers. 2005. Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. *Fish and Fisheries* 6(4): 281-306.

Galuardi, B. 2006. Geolocation of pop-up satellite archival tags: Optimal smoothing of sea surface temperature for Kalman filter estimation. Master of Science Thesis. Zoology Department, University of New Hampshire, Durham, New Hampshire, USA: 84 pp.

Golet, W.J., A.B. Cooper, R. Campbell and M. Lutcavage. 2007. Decline in condition of northern bluefin tuna (*Thunnus thynnui*) in the Gulf of Maine. *Fishery Bulletin* 105(3): 390-395.

Graves, J.E., B. Luckhurst and E.D. Prince. 2002. An evaluation of pop-up satellite tags for estimating postrelease survival of blue marlin (*Makaira nigricans*) from a recreational fishery. *Fishery Bulletin* 100(1): 134-142.

Gutenkunst, R., N. Newlands, M. Lutcavage and L. Edelstein-Keshet. 1999. Inferring resource distributions from Atlantic bluefin tuna movements: An analysis based on net displacement and length of track. *Journal of Theoretical Biology* 245(2): 243-257.

Hanna, S.S. 1999. From single-species to biodiversity – making the transition in fisheries management. *Biodiversity and Conservation* 8: 45-54.

Harvey, A.C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge, U.K: Cambridge University Press: 554pp.

Hill, R. 1994. Theory of geolocation by light levels. In: Le Boeuf, B. and R. Laws (eds) *Elephant seals: population ecology, behavior, and physiology.* University of California Press, Berkeley, CA: 227-236.

Hill, R. and M. Braun. 2001. Geolocation by light-level, the next step: latitude. In: Sibert, J. and J.L. Nielsen (eds). *Electronic tagging and tracking in marine fisheries: Methods and Technologies in Fish Biology and Fisheries.* Kluwer Academic Publishers, Dordrecht: 443-456.

Hill, S.L., E.J. Murphy, K. Reid, P.N. Trathan and A.J. Constable. 2006. Modelling Southern Ocean Ecosystems: Krill, The Food-Web, And The Impacts Of Harvesting. *Biological Reviews of the Cambridge Philosophical Society* 81(4): 581-608.

Hinton, M.G. and H. Nakano. 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (*Makaira nigrans*) catch and effort data from Japanese longline fisheries in the Pacific. *Inter-American Tropical Tuna Commision Bulletin* 21(4): 69-200.

Horne, J.S. and E.O. Garton. 2006. Selecting the best home range model: An information-theoretic approach. *Ecology* 87(5): 1146-1152.

Humston, R., S.J. Ault, M. Lutcavage and B.D. Olson. 2000. Schooling and migration of large pelagic fishes relative to environmental cues. *Fisheries Oceanography* 9(2): 136-146.

ICCAT. 2003. Report of the 2002 Atlantic bluefin tuna stock assessment session. *ICCAT Collective Volume of Scientific Papers* 55(3): 710-937.

Ihaka, R. and R. Gentleman. 1996. R: A Launguage for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* 5: 299-314.

Itoh, T., S. Tsuji and A. Nitta. 2003. Simming Depth, Ambient Water Temperature Preference, and Feeding Frequency of Young Pacific Bluefin Tuna (*Thunnus orientalis*) Determined With Archival Tags. *Fishery Bulletin* 101(3): 535-544.

Jennrich, R.I. and F.B. Turner. 1969. Measurement of non-circular home range. *Journal of Theoretical Biology* 22(2): 227-237.

Jonsen, I.D., R.A. Myers and J. Mills Flemming. 2003. Meta-analysis of animal movement using state-space models. *Ecology* 84(11): 3055-3063.

Julier, S.J. and J.K. Uhlmann. 1997. New extension of the Kalman filter to nonlinear systems. *In*: Kadar, I. (ed) Signal Processing, Sensor Fusion, and Target Recognition VI: 182-193.

Kalman, R. 1960. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Jounral of Basic Engineering* 82(Series D): 35-45.

Kamler, J.F., W.B. Ballard, R.L. Gilliland and K. Mote. 2003. Spatial relationships between swift foxes and coyotes in northwestern Texas. *Canadian Journal of Zoology* 81: 168-172.

Kernohan, B.J., J.J. Millspaugh, J.A. Jenks and D.E. Naugle. 1998. Use of an adaptive kernel home-range estimator in a GIS environment to calculate habitat use. *Journal of Environmental Management* 53(1): 83-89.

Kitagawa, T., H. Nakata, S. Kimura, T. Itoh, S. Tsuji and A. Nitta. 2000. Effect of ambient temperature on the vertical distribution and movement of Pacific bluefin tuna *Thunnus thynnus orientalis*. *Marine Ecology Progress Series* 206: 251-260.

Lutcavage, M.E. 2004. Personal Communication. Department of Zoology, University of New Hampshire, Durham, NH 03824, U.S.A.

Lutcavage, M.E., R.W. Brill, G.B. Skomal, B.C. Chase and P.W. Howey. 1999. Results of pop-up satellite tagging of spawning size class fish in the Gulf of Maine: do North Atlantic bluefin tuna spawn in the mid-Atlantic? *Canadian Journal of Fisheries and Aquatic Sciences* 56: 173-177.

Lutcavage, M.E., R.W. Brill, G.B. Skomal, B.C. Chase, J.L. Goldstein and J. Tutein. 2000. Tracking adult North Atlantic bluefin tuna (*Thunnus thynnus*) in the northwestern Atlantic using ultrasonic telemetry. *Marine Biology* 137: 347-358.

Macdonald, D.W. and S. Rushton. 2003. Modelling space use and dispersal of mammals in real landscapes: a tool for conservation. *Journal of Biogeography* 30: 607-620.

Millspaugh, J.J., R.M. Nielson, L. McDonald, J.M. Marzluff, R.A. Gitzen, C.D. Rittenhouse, M.W. Hubbard, S.L. Sheriff. 2006. Analysis of resource selection using utilization distributions. *Journal of Wildlife Management* 70(2): 384-395.

Mohr, C.O. 1947. Table of equivalent populations of North American small mammals. *American Midland Naturalist* 37: 223-249.

Molsa, H., J.E. Reynolds, E.J. Coenen and O.V. Lindqvist. 1999. Fisheries research towards resource management on Lake Tanganyika. *Hydrobiologia* 40: 1-24.

Mumby, P.J. 2006. Connectivity of reef fish between mangroves and coral reefs: Algorithms for the design of marine reserves at seascape scales. *Biological Conservation* 128(2): 215-222.

McLoughlin, P.D. and S.H. Ferguson. 2000. A hierarchical pattern of limiting factors helps explain variation in home range size. *Ecoscience* 7: 123-130.

Metcalfe, J.D. 2001. Summary report of a workshop on daylight measurements for geolocation in animal telemetry. In: Sibert J. and J.L. Nielsen (eds) *Electronic Tagging and Trackin in Marine Fisheries Reveiws: Methods and Technologies in Fish Biology and Fisheries.* Kluwer Academic Press, Dordrecht: 443-456.

Moorcroft, P.R., M.A. Lewis and R.L. Crabtree. 1999. Home range analysis using a mechanistic home range model. *Ecology* 80(5): 1656-1665.

Murdoch, D.J. and E.D. Chow. 1996. A graphical display of large correlation matrices. *The American Statistician* 50: 178-180.

Musyl, M.K., R. Brill, D.S. Curran, J.S. Gunn, J.R. Hartog, R.D. Hill, D.W. Welch, J.P. Eveson, C.H. Boggs and R.E. Brainard. 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. In:

Sibert, J.R. and J. Nielsen (eds) *Electronic Tagging and Tracking in Marine Fisheries*. Kluwer Academic Publishers, Dordrecht: 343-367.

Newlands, N.K., M.E. Lutcavage and T.J. Pitcher. 2004. Analysis of foraging movements of Atlantic bluefin tuna (*Thunnus thynnus*): individuals switch between two modes of search behaviour. *Population Ecology* 46: 39-53.

Nielsen, A. 2004. Estimating Fish Movement. Ph.D. Dissertation. Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Frederiksberg, Denmark: 119 pp.

Nielsen, A., K.A. Bigelow, M.K. Musyl and J.R. Sibert. 2006. Improving light-based geolocation by including sea surface temperature. *Fisheries Oceanography* 15(4): 314-325.

NMFS. 2004. 2004 report to Congress pursuant to the Shark Finning Prohibition Act of 2000 (Public Law 106-557). National Marine Fisheries Service: 37.

Okubo, A. 1980. Diffusion and Ecological Problems: Mathematical Models. New York: Springer: 254pp.

Papi, F., P. Luschi, E. Crosio and G.R. Hughes. 1997. Satellite tracking experiments on the navigational ability and migratory behaviour of the loggerhead turtle *Caretta caretta*. *Marine Biology* 129: 215-220.

Pew Oceans Commission. 2003. America's Living Oceans: Charting a Course for Sea Change. *A Report to the Nation*. May 2003. Pew Oceans Commission, Arlington, Virginia: 164pp.

Powers, J.E. and C.E. Porch. 2003. Approaches to incorporating mixing and movement of Atlantic bluefin tuna into management and assessment. *Theme Session V. Mixed and Multi-Stock Fisheries-Challenges and Tools for Assessments, Prediction, and Management. ICES Council Meeting documents*

Restrepo, V. 1996. An implementation of an age-structured production model with application to western Atlantic bluefin tuna fisheries. *ICCAT Collective Volume of Scientific Papers* 66(2): 348-353.

Restrepo, V.R., C.E. Porch, S.C. Turner, G.P. Scott and A.A. Rosenberg. 1994. Combination of spawner-recruit, spawning biomass-per-recruit and yield-per-recruit computations for the estimation of the long term potential for west Atlantic bluefin tuna. *ICCAT Collective Volume of Scientific Papers* 42(1): 214-222.

Rouget, M., R.M. Cowling, A.T. Lombard, A.T. Knight and I.H.K. Graham. 2006. Designing large-scale conservation corridors for pattern and process. *Conservation Biology* 20(2): 549-561.

Royer, F., J.M. Fromentin and P. Gaspar. 2005. A state space model to derive bluefin tuna movement and habitat from archival tags. *Oikos* 109: 473-484.

Royer, F., B. Galuardi, P. Gaspar and M. Lutcavage. 2006. Movement Patterns of Atlantic Bluefin Tuna in the Gulf Stream Area: an Improved Multi-Year Analysis using Pop-up Satellite Archival Tags and Oceanographic Fields. *EOS, Transactions, American Geophysical Union* 87(36) suppl.

Royer, F., S. Wilson and M. Lutcavage. 2006. Migratory and depth behaviour of bluefin tuna in the Gulf Stream Area: how Mercator can help in revealing habitat use by pelagic species. *Mercator Quarterly Newsletter* 20: 7-11.

Schaefer, K.M. and D.W. Fuller. 2002. Movements, behavior, and habitat selection of bigeye tuna (*Thunnus obesus*) in the eastern equatorial Pacific, ascertained through archival tags. *Fishery Bulletin* 100: 765-788.

Schick, R.S., J. Goldstein and M.E. Lutcavage. 2004. Bluefin tuna (*Thunnus thynnus*) distribution in relation to sea surface temperature fronts in the Gulf of Maine (1994-96). *Fisheries Oceanography* 13(4): 225-238.

Seaman, D.E. and R.A. Powell. 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology* 77(7): 2075-2085.

Seaman, D.E. and R.A. Powell. 1998. Kernel home range estimation program (kernelhr). Documentation of the program. <URL: ftp://ftp.im.nbs.gov/pub/software/CSE/wsb2695/KERNELHR.ZIP>.

Shaffer, S.A., Y. Tremblay, J.A. Awkerman, W. Henry, S. Teo, D.J. Anderson, D.A. Croll, B.A. Block, D.P. Costa. 2005. Comparison of light- and SST-based geolocation with satellite telemetry in free-ranging albatrosses. *Marine Biology* 147(4): 833-843.

Sibert, J. and Fournier, D.A. 2001. Possible models for combining tracking data with conventional tagging data. In: *Electronic Tagging and Tracking in Marine Fisheries Reviews: Methods and Technologies in Fish Biology and Fisheries*. J. Sibert and J. Nielsen (eds) Dordrecht: Kluwer Academic Press: 443-456.

Sibert, J.R., M.K. Musyl and R.W. Brill. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. *Fisheries Ocenaography* 12(3): 141-151.

Sibert, J.R. and A. Nielsen. 2004. *kftrack*. https://www.soest.hawaii.edu/tag-data/tracking/*kftrack*/.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. London, UK: Chapman and Hall.

Sissenwine, M.P., P.M. Mace, J.E. Powers and G.P. Scott. 1998. A commentary on western Atlantic bluefin tuna assessments. *Transactions of the American Fisheries Society* 127: 838-855.

Smith, T. 1994. Scaling fisheries: the science of measuring the effects of fishing, 1855-1955. New York: Cambridge University Press: 392pp.

Stumpf, W.A. and C.O. Mohr. 1962. Linearity of home ranges of California mice and other animals. *Journal of Wildlife Management* 26(2): 149-154.

U.S. Commission on Ocean Policy. 2004. An Ocean Blueprint for the 21st Century - Final Report – Pre-Publication Copy. Washington, D.C. ISBN#0-9759462-0-X: 610pp.

Wand, M.P. 1994. Fast Computation of Multivariate Kernel Estimators. *Journal of Computational and Graphical Statistics* 3: 433-445.

Wand, M.P. and M.C. Jones. 1995. Kernel Smoothing. Chapman and Hall, London: 224pp.

Welch, D.W. and J.P. Eveson. 1999. An assessment of light-based geoposition estimates from archival tags. *Canadian Journal of Fisheries and Aquatic Sciences* 56(7): 1317-1327.

Wilson, S.G., M.E. Lutcavage, R.W. Brill, M.P. Genovese, A.B. Cooper and A.W. Everly. 2005. Movements of bluefin tuna (*Thunnus thynnus*) in the northwestern Atlantic Ocean recorded by pop-up satellite archival tags. *Marine Biology* 146: 409-423.

Worton, B.J. 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70(1): 164-168.

Worton, B.J. 1995. Using Monte Carlo simulation to evaluate kernel-based home range estimators. *Journal of Wildlife Management* 59: 794-800.

# APPENDICES

## APPENDIX A – *KFUD* CODE

```
// KFtrack program by John Sibert <jsibert@soest.hawaii.edu> (2001)
// Minor modifications by Anders Nielsen <anielsen@dina.kvl.dk> (2002+3)
// Modifications by Daniel Badger <daniel.badger@unh.edu> (2006) to produce
geolocation estimates everyday
//
// This version includes:
//
// Extended Kalman filter
// Smoothing
// minor modification of likelihood function
// Known recapture position option (but weight is not used anymore)
// first need not to be known
// error estimates on most probable track
// commands to extrand likelihood profiles of the parameters


GLOBALS_SECTION
 #include <fstream.h>
 #include <math.h>
 #include <fvar.hpp>
 #include <azimuth.cpp>
 #include <adstring.hpp>
 #include "yrmonday.h"
 #include "trace.h"
 #include <strstream>
 #include <iostream>
 using std::ostrstream;
 #undef REPORT
 #define REPORT(object) report << #object " = " << object << endl;
 #define MREPORT(object) report << #object ": \n" << object << endl;

 // function prototypes
//  adstring make_banner();
//  double azimuth(const double& y, const double& x);
//  dvariable azimuth(const dvariable& y, const dvariable& x);
//  dvariable gc_dist(const dvector& y1, const dvar_vector y2);
//  int previous_solstice(const int y0, const int m0, const int d0);

 // global variables
```

```
 const double mpg = 60.0; // Nautical miles per degree
 const double rmpg = 1.0/mpg;
 const double mpi180 = M_PI/180.0;
 const double two_pi = 2.0*M_PI;
 const double epss = 1e-8;  //small number to avoid divide by 0
 ofstream clogf("kftrack.log");
 int u_phase = -1;
 int v_phase = -1;
 int D_phase = -1;
 int bx_phase = -1;
 int by_phase = -1;
 int vx_phase = -1;
 int vy_phase = -1;
 int cos_phase = -1;
 int a0_phase = -1;
 int b0_phase = -1;
 int dev_phase = -1;
 int t;

 adstring copyright("\n  (c) 2001 John Sibert\n"\
      "  Pelagic Fisheries Research Program, University of Hawaii\n");

DATA_SECTION
 init_int npoint;
 !!TRACE(npoint)
 init_int N;
 init_int m;
 init_int col;
 init_int dal;
 init_int no_data_days;
 init_int step;
 init_int half_step;
 init_int release_point;
 !!TRACE(release_point)
 init_int recap_point;
 !!TRACE(recap_point)
 !!TTRACE(N,m)
 init_int u_active;
 init_int v_active;
 init_int D_active;
 init_int bx_active;
 init_int by_active;
 init_int vx_active;
 init_int vy_active;
 init_int a0_active;
 init_int b0_active;
```

```
 init_int D_wt_phase;
 init_int rnwalk_wt_phase;

// Read in initial values for the following parameters
 init_number init_u; // longitudinal bias
 init_number init_v; // latidunial bias
 init_number init_D;// rate at which the uncertainty of the position increases over
time
 init_number init_bx;        // inherent longitudinal bias of the tag, in degrees
 init_number init_by;        // inherent latitudinal bias of the tag, in degrees
 init_number init_vx;        // mean squared error in estimating longitude
 init_number init_vy;        // mean squared error in estimating latitude
 init_number init_a0;        // affects the variability of latitude measurement
// error throughout year
 init_number init_b0;        // number of days prior to the equinox
 init_number D_prior_variance;
 init_number init_D_wt;
 init_number init_rnwalk_wt;
 init_number avg_long;
 init_number avg_lat;

 init_int cos_errors;
 init_int dev_errors;
 init_number vy_dev_penalty_wt;

 number point;
 number length;
 number time_length;
 !! length = (npoint)*step-half_step+no_data_days*step;
 !! time_length = (npoint)*step-half_step+no_data_days*step;
 number interp;
 !! interp =  length-npoint;
 number step_2;
 !! step_2 = step;

 int nphase;
 matrix Y(1,length,1,N)
 matrix y_deg(1,length,1,N)
 matrix P0(1,m,1,m);
 vector vy_t(1,length);
 vector time(1,length);

 init_matrix dat_mat(1,npoint+no_data_days,1,col);
 !! P0.initialize();
 number start_long;
```

```
LOCAL_CALCS
  nphase = -1;
  if (u_active || v_active || D_active)
  {
   nphase ++;
   if (u_active)
     u_phase = nphase;
   if (v_active)
     v_phase = nphase;
   if (D_active)
     D_phase = nphase;
  }
  if (vx_active || vy_active || bx_active || by_active)
  {
   nphase ++;
   if (bx_active)
     bx_phase = nphase;
   if (by_active)
     by_phase = nphase;
   if (vx_active)
     vx_phase = nphase;
   if (vy_active)
     vy_phase = nphase;
  }
  if (cos_errors)
  {
   nphase ++;
   cos_phase = nphase;
   if(a0_active)a0_phase=cos_phase;
   if(b0_active)b0_phase=cos_phase;
  }
  if (dev_errors)
  {
   nphase ++;
   dev_phase = nphase;
  }
  TTRACE(u_active,u_phase)
  TTRACE(v_active,v_phase)
  TTRACE(D_active,D_phase)
  TTRACE(bx_active,bx_phase)
  TTRACE(by_active,by_phase)
  TTRACE(vx_active,vx_phase)
  TTRACE(vy_active,vy_phase)
  TTRACE(cos_errors,cos_phase)
  TTRACE(dev_errors,dev_phase)
```

```
PARAMETER_SECTION
 matrix a(1,length,1,m);
 matrix a1(1,length,1,m);
 matrix aSmooth(1,length,1,m);
 matrix T(1,m,1,m)
 matrix TT(1,m,1,m)
 vector d(1,N)
 matrix c(1,length,1,m)
 matrix Q(1,m,1,m)
 matrix H(1,N,1,N)
 matrix v(1,length,1,N);
 3darray P(1,length,1,m,1,m);
 3darray P1(1,length,1,m,1,m);
 3darray PSmooth(1,length,1,m,1,m);
 3darray PSmoothTrans(1,length,1,m,1,m);
 3darray PStar(1,length,1,m,1,m);
 matrix ySmooth(1,length,1,N);
 vector next_y(1,N);
 vector blk(1,length);
 vector observed(1,length);
 init_vector rnpar_lat(1,interp);
 init_vector rnpar_long(1,interp);
 matrix rnwalk(1,length,1,m);

 init_bounded_number uu(-50.0,50.0,u_phase);
 init_bounded_number vv(-50.0,50.0,v_phase);
 init_bounded_number D(0.0,5000.0,D_phase);
 init_bounded_number vx(0.0,50.0,vx_phase);
 init_bounded_number vy(0.0,50.0,vy_phase);
 init_bounded_number bx(-50.0,50.0,bx_phase);
 init_bounded_number by(-50.0,50.0,by_phase);
 init_bounded_number a0(0.0,50.0,a0_phase);
 init_bounded_number b0(-80.0,80.0,b0_phase);
 init_bounded_vector vy_dev(2,length,-500.0,500.0,dev_phase);
 init_bounded_number D_wt(0.00,1.00,D_wt_phase);
 init_bounded_number rnwalk_wt(0.00,1.00,rnwalk_wt_phase);
 matrix Z(1,N,1,m)
 objective_function_value f;
 matrix expanded(1,length,1,col+1);
 number kalman_like;
 number recap_err;
 number gc_recap_err;
 number e1;
 number e2;
 number e3;
 number e4;
```

```
number dt;
likeprof_number u_prof;
likeprof_number v_prof;
likeprof_number D_prof;
likeprof_number vx_prof;
likeprof_number vy_prof;
likeprof_number bx_prof;
likeprof_number by_prof;
likeprof_number a0_prof;
likeprof_number b0_prof;
likeprof_number last_long;
likeprof_number last_lat;
likeprof_number test_lat;
likeprof_number test_lon;

sdreport_number sduu;
sdreport_number sdvv;
sdreport_number sdD;
sdreport_number sdbx;
sdreport_number sdby;
sdreport_number sdvx;
sdreport_number sdvy;
sdreport_number vxy;
sdreport_number hdg;
sdreport_number spd;
LOCAL_CALCS
 cout << "D_phase = " << D_phase << endl;
 cout << "step_2 = " << step_2 << endl;
 uu = init_u;
 vv = init_v;
 D  = init_D;
 bx = init_bx;
 by = init_by;
 vx = init_vx;
 vy = init_vy;
 D_wt = init_D_wt;
 rnwalk_wt = init_rnwalk_wt;
 if (D < epss)
   D = epss;
 TTRACE(init_D,D);
 if (vx < epss)
   vx = epss;
 TTRACE(init_vx,vx)
 if (vy < epss)
   vy = epss;
 TTRACE(init_vy,vy)
```

```
if (cos_phase < 0){
  a0 = epss;
  b0 = 0.0;
}else{
  a0 = init_a0;
  b0 = init_b0;
}
d.initialize();
Z.initialize();
T.initialize();
T(1,1)=1; T(2,2)=1;
TT=trans(T);

dvector yLong = column(dat_mat,4);
TTRACE(min(yLong),max(yLong))
start_long = min(yLong)+0.5*(max(yLong)-min(yLong));
TRACE(start_long)
clogf << "\nInput data:" << endl;
clogf << "index      date sday    long      lat       x        y"
     << endl;
cout << "length = " << length << endl;
cout << "time_length = " << time_length << endl;
if(step == 1){
  for(int i = 1; i <= length; i++){
    time(i) = 12;
  }
}else{
  time(1) = 12;
  for (int i = 2; i <= length; i++){
    if(time(i-1)+24/step < 24){
      time(i) = time(i-1) + 24/step;
    }else{
      time(i) = -24 + time(i-1) + 24/step;
    }
  }
}
cout << "interp = " << interp << endl;
cout << "time" << time << endl;

      // to create a matrix that has room for mult values per day
for (int w = 1; w <= col; w++) expanded(1,w) = dat_mat(1,w);
expanded(1,col+1) = time(1);
for (int i = 2; i <= time_length; i++){
  expanded(i,col+1) = time(i);
  if((i-1)%step==0){                    //if step is divisible by i, then we do the
following
```

```
      for(int j=1; j<=col; j++){
        expanded(i,j)=dat_mat((i-1)/step+1,j);
      }
    }else{
      if(time(i) > 12){
        expanded(i,1) = expanded(i-fmod((i-1),step),1);
        expanded(i,2) = expanded(i-fmod((i-1),step),2);
        expanded(i,3) = expanded(i-fmod((i-1),step),3);
      }
    }
  }
  for (int i = 2; i <= time_length; i++){
    if(time(i) < 12){
      expanded(i,1) = expanded(i+step-fmod((i-1),step),1);
      expanded(i,2) = expanded(i+step-fmod((i-1),step),2);
      expanded(i,3) = expanded(i+step-fmod((i-1),step),3);
    }
    if(expanded(i,4)==0){   // To allow unobserved day's initial position values to
be set
                            // to appropriate scale
      expanded(i,4) = avg_long;
      expanded(i,5) = avg_lat;
      expanded(i,8) = expanded(i-1,8)+(1/step_2);
    }
  }

  for(int i =2;i<=time_length;i++){
    expanded(i,6) = expanded(i-1,6)+1;
  }


  blk(2) = 1/step;
  for (int t = step+1; t <= length; t++){
    if(value(expanded(t-step,7)) == 0){
      blk(t) = blk(t-1) + 1;    // allows the number of days since an
                                // observation to increase based on a non-observation
day
    }else{
      blk(t) = 1;
    }
  }

  for (int g = 1; g <= length; g++){ // tells me the point at which it is midway
                                     // between observations on a given day and
month
```

```
    if((value(expanded(g,9)) == 0)&&(value(expanded(g,1)) ==
9)&&(value(expanded(g,2)) == 2)){
      point = g;
    }
  }

  cout << "dat_mat =" << dat_mat << endl;
  cout << "expanded = " << expanded << endl;


  for (int i = 1; i <= length; i++)
  {
    // set up estimated geographic posision from tag data
    y_deg(i,1) = value(expanded(i,4));
    y_deg(i,2) = value(expanded(i,5));

    // shift origin longitude
    Y(i,1) = y_deg(i,1) - start_long*value(expanded(i,7));
    Y(i,2) = y_deg(i,2);

    dvar_vector Yi=zInv(Y(i));
    clogf << setw(5) << expanded(i,8)
         << setw(10) << Y(i,1) << setw(10) << Y(i,2)
         << setw(10) << Yi(1) << setw(10) << Yi(2)
         << endl;
  }
  clogf << "\nFinished LOCAL_CALCS in PARAMETER_SECTION.\n" << endl;
  cout << "  Finished LOCAL_CALCS in PARAMETER_SECTION.\n" << endl;

PROCEDURE_SECTION
 setup_d();
 setup_H();
 int counter = 1;
 for(int h = 1; h<=length;h++){
   if(expanded(h,7)==0){
     rnwalk(h,1)=rnpar_long(counter);
     rnwalk(h,2)=rnpar_lat(counter);
     counter+=1;
   }
   if(expanded(h,7)==1){
     rnwalk(h,1)=0;
     rnwalk(h,2)=0;
   }
 }
 f+=kalman_filter();
```

```
  vxy = sqrt(vx*vx+vy*vy);
  sduu = uu;
  sdvv = vv;
  sdD = D;
  sdbx = bx;
  sdby = by;
  sdvx = vx;
  sdvy = vy;
  spd = sqrt(uu*uu+vv*vv+epss);
  hdg = azimuth(vv,uu);

FUNCTION setup_d
  d(1) = bx;
  d(2) = by;

FUNCTION setup_Q
  Q.initialize();
  for (int i=1;i<=m;i++)
    Q(i,i) = 2.0*D*dt;

FUNCTION setup_H
  H.initialize();
  H(1,1) = vx*vx;
  H(2,2) = vy*vy;

FUNCTION dvar_vector varA(dvar_matrix Y)
  dvar_vector A1(1,length), A2(1,length);
  dvar_vector tmp(1,2);
  for(int i=1; i<=length; ++i){
    tmp=zInv(Y(i));
    A1(i)=tmp(1); A2(i)=tmp(2);
  }
  tmp(1)=pow(std_dev(A1),2); tmp(2)=pow(std_dev(A2),2);
  return(tmp);

FUNCTION dvar_vector z(dvar_vector alpha)
  dvar_vector tmp(1,N);
  tmp(1)=alpha(1)/(mpg*cos(alpha(2)/mpg*mpi180));
  tmp(2)=alpha(2)/mpg;
  return(tmp);

FUNCTION dvar_vector zInv(dvar_vector y)
  dvar_vector tmp(1,m);
  tmp(1)=y(1)*mpg*cos(y(2)*mpi180);
  tmp(2)=y(2)*mpg;
  return(tmp);
```

```
FUNCTION dvar_matrix ZHatFun(dvar_vector a)
 dvar_matrix tmp(1,N,1,N);
 dvariable ex1=mpg*cos(a(2)/mpg*mpi180);
 tmp(1,1)=1.0/ex1;
 tmp(1,2)=a(1)*sin(a(2)/mpg*mpi180)/(ex1*ex1)*mpi180;
 tmp(2,1)=0;
 tmp(2,2)=1.0/mpg;
 return(tmp);

FUNCTION dvariable kalman_filter(void)
 dvar3_array F(1,length,1,N,1,N);
 dvar3_array Finv(1,length,1,N,1,N);
 dvar_matrix Ptemp(1,m,1,m);

 a(1,1) = mpg*(Y(1,1))*cos(mpi180*Y(1,2));
 a(1,2) = mpg*Y(1,2);

 if(!release_point){
  dvar_vector tmp(1,2); tmp=varA(Y);
  P0(1,1)=value(tmp(1)); P0(2,2)=value(tmp(2));
 }

 P(1)=P0;
 rnwalk(1,1)=0;
 rnwalk(1,2)=0;
 // This is the Kalman filter recursion. The objects tmp1
 // and tmp2 hold common calculations to optimize a bit
 for (t=2;t<=length;t++)
 {
  dt = 1/static_cast<double>(step);

  setup_Q();
  c(t,1)=uu*dt;
  c(t,2)=vv*dt;

  a1(t)=T*a(t-1)+c(t);
  P1(t)=T*P(t-1)*TT+Q;
  if(expanded(t,7)==0){
   P1(t,1,1)=P1(t,1,1)+rnwalk(t,1);
   P1(t,2,2)=P1(t,2,2)+rnwalk(t,2);
  }

  Z=ZHatFun(a1(t));
```

```
  if((recap_point)&&(t==length)){
   next_y=z(a1(t));
  }else{
   next_y=z(a1(t))+d*expanded(t,7);     // Makes 'd' be zero in the matrixfor any
time step
                                          // where there was not an observation
  }
  v(t)=Y(t)-next_y*expanded(t,7);
  dvar_matrix tmp1=P1(t)*trans(Z);

  if (cos_errors){
   int sdx = (int)fmod(value(expanded(t,8)),365.25);
   int bdx = (int)(sdx/182.625) + 1;
   e1 = cos(two_pi*(pow(-1.0,bdx)*b0+value(expanded(t,8)))/365.25);
   e3 = vy*1.0/sqrt(e1*e1+a0);
   H(2,2) = e3*e3;
  }
  if (active(vy_dev))
  {
   e4 = vy*(exp(vy_dev(t)));
   H(2,2) = e4*e4;
  }
  if((recap_point)&&(t==length)){
   H(1,1)=0; H(1,2)=0; H(2,1)=0; H(2,2)=0;
  }
  vy_t(t) = sqrt(value(H(2,2)));
  F(t)=Z*tmp1+H;
  Finv(t)=inv(F(t));
  dvar_matrix tmp2= tmp1*Finv(t);
  P(t)=P1(t)-tmp2*Z*P1(t)*expanded(t,7);
  a(t)=a1(t)+tmp2*v(t)*expanded(t,7)+rnwalk(t)*(-1*(expanded(t,7)-1));
 }
 int sgn=0;
 kalman_like = (npoint-1)*log(two_pi);

 u_prof = uu;
 v_prof = vv;
 D_prof = D;
 vx_prof = vx;
 vy_prof = vy;
 bx_prof = bx;
 by_prof = by;
 a0_prof = a0;
 b0_prof = b0;
 kalman_like +=D_wt*(log(sqrt(D_prior_variance))+.5*log(two_pi)+((log(D)-
log(init_D))*(log(D)-log(init_D)))/(2*D_prior_variance))+log(sqrt(2*D*dt)) +
```

```
.5*log(two_pi); //the last two terms are connected with the likelihood equation
below, but do not need to be repeated for every t
 for (t=2;t<=length;t++)
 {
  dvariable tkl = (v(t)*Finv(t)*v(t))*value(expanded(t,7));
  kalman_like
+=(0.5*ln_det(F(t),sgn)+0.5*v(t)*Finv(t)*v(t))*value(expanded(t,7))+rnwalk_wt*(1-
expanded(t,7))*(((rnwalk(t,1)-0)*(rnwalk(t,1)-0))/(4*D*dt)+((rnwalk(t,2)-
0)*(rnwalk(t,2)-0))/(4*D*dt)); //
 }
 dvariable f = kalman_like;

 if (active(vy_dev))
 {
   f += vy_dev_penalty_wt*norm2(vy_dev);
 }

 //Smoothing loop
 if(recap_point){
  PSmooth(length,1,1)=0; PSmooth(length,1,2)=0;
  PSmooth(length,2,1)=0; PSmooth(length,2,2)=0;
 }else{
  PSmooth(length)=P(length);
 }

 if(recap_point){
  aSmooth(length)=zInv(Y(length));
  ySmooth(length)=Y(length);
 }else{
  aSmooth(length)=a(length);
  ySmooth(length)=z(aSmooth(length)); //notice without bias term
 }

 for(int i=(length-1); i>=1; --i){
  PStar(i)=P(i)*inv(P1(i+1));
  aSmooth(i)=a(i)+PStar(i)*(aSmooth(i+1)-a(i)-c(i+1));
  ySmooth(i)=z(aSmooth(i)); //notice without bias term
  PSmooth(i)=P(i)+PStar(i)*(PSmooth(i+1)-P1(i+1))*trans(PStar(i));
 }

 for(int i=1; i<=length; ++i){
PSmoothTrans(i)=ZHatFun(aSmooth(i))*PSmooth(i)*trans(ZHatFun(aSmooth(i)));
 }
 last_long = ySmooth(length,1)+start_long;
 last_lat = ySmooth(length,2);
 test_lon = rnwalk(10,1);
```

```
 test_lat = rnwalk(10,2);

 dvar_vector lon = column(ySmooth,1);
 dvar_vector lat = column(ySmooth,2);

 if (mceval_phase()){
   cout << "lat: " << lat << " " << "long: " << lon+start_long << endl;
// Comment in the below statement and comment out the above statement to
have the
// -mceval command return the parameter values throughout the MCMC runs
//   cout << "D: " << D << " " << "uu: " << uu << " " << "vv: " << vv << " " << "bx: "
<< bx << " " << "by: " << by << " " << "vx: " << vx << " " << "vy: " << vy << " " <<
"a0: " << a0 << " " << "b0: " << b0 << endl;
 }
 return f;

REPORT_SECTION
 REPORT(current_phase())
 char flags[80];
 ostrstream ss(flags,80);
 ss <<active(uu)<<active(vv)<<active(D)<<active(bx)<<active(by)
   <<active(vx)<<active(vy)
   <<cos_errors<<active(vy_dev) << ends;
 REPORT(flags)
 int days_at_liberty = dal;
 REPORT(days_at_liberty)
 REPORT(npoint)
 double reporting_rate = (double)npoint/(double)days_at_liberty;
 REPORT(reporting_rate)
 REPORT(npoint)
 int Number_of_parameters =  initial_params::nvarcalc();
 REPORT(Number_of_parameters)
 REPORT(f)
 REPORT(kalman_like)
 REPORT(recap_point)
 if (recap_point)
 {
  REPORT(recap_err)
  REPORT(gc_recap_err)
 }
 REPORT(uu)
 REPORT(vv)
 REPORT(D)
 REPORT(D_wt)
 REPORT(rnwalk_wt)
 REPORT(bx)
```

```
  REPORT(by)
  REPORT(vx)
  REPORT(vy)
  REPORT(a0)
  REPORT(b0)
  REPORT(spd)
  REPORT(hdg)
  REPORT(vxy)
  REPORT(vy_dev_penalty_wt)
  REPORT(norm2(vy_dev))
  REPORT(c)
  MREPORT(Q)
  REPORT(d)
  MREPORT(H)
  MREPORT(rnwalk)

  if (last_phase())
  {
   adstring gmt_name("gmt_");
   gmt_name += adstring(flags);
   gmt_name += adstring(".dat");
   REPORT(gmt_name);
   adstring rstuff_name("rstuff_");
   rstuff_name += adstring(flags);
   rstuff_name += adstring(".dat");
   REPORT(rstuff_name);
   adstring mpt_name("mpt_");
   mpt_name += adstring(flags);
   mpt_name += adstring(".dat");
   REPORT(mpt_name);
   ofstream rstuff(rstuff_name);
   rstuff << "i date time dt j vy ax ay ox oy px py smoothX smoothY
Psmooth11 Psmooth12 Psmooth21 Psmooth22 observed?" << endl;
   dvector PY(1,N);
   for (int i = 1; i <= length; i++)
   {
    if(i%i==0){
     dt = 0;
     double vyt = 0.0;
     PY = y_deg(i);
     if (i > 1)
     {
      dt = 1/static_cast<double>(step);
      if (no_data_days == 0) dt = value(expanded(i,6)) - value(expanded(i-
step,6));
      vyt = vy_t(i);
```

```
        PY = value(z(a(i)));
        PY(1) += start_long;
      }
      if(value(expanded(i,7))==1){
        rstuff    << setw(5) << expanded(i,6) << " " << value(expanded(i,3)) << "/"
<< value(expanded(i,2)) << "/" << value(expanded(i,1)) << setw(4) << time(i) <<
":00" << " " << setw(6) << dt << " "
              << setw(5) << expanded(i,8) << " "                          //
setw(5) says to make 4 spaces before performing next thing
              << setw(10) << setprecision(4) << vyt <<" "
              << setw(10) << setprecision(5) << a(i,1) <<" "
              << setw(8) << setprecision(4) << a(i,2) <<" "
              << setw(11) << setprecision(6) << y_deg(i,1) <<" "
              << setw(8) << setprecision(5) << y_deg(i,2) <<" "
              << setw(11) << setprecision(6) << PY(1) <<" "      // Predicted Track
              << setw(9) << setprecision(5) << PY(2) <<" "       // Predicted Track
              << setw(10) << setprecision(6) << ySmooth(i,1)+start_long      <<" " //
Smoothed most probable track
              << setw(9) << setprecision(5) << ySmooth(i,2) <<" "                 //
Smoothed most probable track
              << setw(11) << setprecision(5) << PSmoothTrans(i,1,1) <<" "
              << setw(11) << setprecision(5) << PSmoothTrans(i,1,2) <<" "
              << setw(11) << setprecision(5) << PSmoothTrans(i,2,1) <<" "
              << setw(11) << setprecision(5) << PSmoothTrans(i,2,2) <<" "
              << setw(31) << "locations_observed"
              << endl;
      }else{
        rstuff    << setw(5) << expanded(i,6) << " " << value(expanded(i,3)) << "/"
<< value(expanded(i,2)) << "/" << value(expanded(i,1)) << setw(4) << time(i) <<
":00" << " " << setw(6) << dt << " "
                << setw(5) << expanded(i,8) << " "
                << setw(10) << setprecision(4) << vyt <<" "
                << setw(10) << setprecision(5) << a(i,1) <<" "
                << setw(8) << setprecision(4) << a(i,2) <<" "
                << setw(11) << setprecision(6) << y_deg(i,1) <<" "
                << setw(8) << setprecision(5) << y_deg(i,2) <<" "
                << setw(11) << setprecision(6) << PY(1) <<" "
                << setw(9) << setprecision(5) << PY(2) <<" "
                << setw(10) << setprecision(6) << ySmooth(i,1)+start_long     <<" "
                << setw(9) << setprecision(5) << ySmooth(i,2) <<" "
                << setw(11) << setprecision(5) << PSmoothTrans(i,1,1) <<" "
              << setw(11) << setprecision(5) << PSmoothTrans(i,1,2) <<" "
                << setw(11) << setprecision(5) << PSmoothTrans(i,2,1) <<" "
                << setw(11) << setprecision(5) << PSmoothTrans(i,2,2) <<" "
                << setw(31) << "locations_NOT_observed"
              << endl;
```

```
      }
    }
  }
  ofstream mpt(mpt_name);
  mpt << "# npoint" << endl;
  mpt << setw(5) << npoint << endl;
  mpt << "# i     date  time    dt   j      vy      ax      ay      ox      oy
px      py   smoothX   smoothY  Psmooth11  Psmooth12  Psmooth21
Psmooth22          observed?" << endl;
  for (int i = 1; i <= length; i++)
  {
    dt = 0;
    double vyt = 0.0;
    PY = y_deg(i);
    if (i > 1)
    {
      dt = 1/static_cast<double>(step);
      if (no_data_days == 0) dt = value(expanded(i,6)) - value(expanded(i-
step,6));
      vyt = vy_t(i);
      PY = value(z(a(i)));
      PY(1) += start_long;
    }
    if(value(expanded(i,7))==1){
      mpt    << setw(5) << expanded(i,6) << " " << value(expanded(i,3)) << "/" <<
value(expanded(i,2)) << "/" << value(expanded(i,1)) << setw(4) << time(i) <<
":00" << " " << setw(6) << dt << " "
          << setw(5) << expanded(i,8) << " "
          << setw(10) << setprecision(4) << vyt <<" "
          << setw(10) << setprecision(5) << a(i,1) <<" "
          << setw(8) << setprecision(4) << a(i,2) <<" "
          << setw(11) << setprecision(6) << y_deg(i,1) <<" "
          << setw(8) << setprecision(5) << y_deg(i,2) <<" "
          << setw(11) << setprecision(6) << PY(1) <<" "
          << setw(9) << setprecision(5) << PY(2) <<" "
          << setw(10) << setprecision(6) << ySmooth(i,1)+start_long        <<" "
          << setw(9) << setprecision(5) << ySmooth(i,2) <<" "
          << setw(11) << setprecision(5) << PSmoothTrans(i,1,1) <<" "
          << setw(11) << setprecision(5) << PSmoothTrans(i,1,2) <<" "
          << setw(11) << setprecision(5) << PSmoothTrans(i,2,1) <<" "
          << setw(11) << setprecision(5) << PSmoothTrans(i,2,2) <<" "
          << setw(31) << "locations_observed"
          << endl;
    }else{
```

```
    mpt    << setw(5) << expanded(i,6) << " " << value(expanded(i,3)) << "/" <<
value(expanded(i,2)) << "/" << value(expanded(i,1)) << setw(4) << time(i) <<
":00" << " " << setw(6) << dt << " "
             << setw(5) << expanded(i,8) << " "
             << setw(10) << setprecision(4) << vyt <<" "
             << setw(10) << setprecision(5) << a(i,1) <<" "
             << setw(8) << setprecision(4) << a(i,2) <<" "
             << setw(11) << setprecision(6) << y_deg(i,1) <<" "
             << setw(8) << setprecision(5) << y_deg(i,2) <<" "
             << setw(11) << setprecision(6) << PY(1) <<" "
             << setw(9) << setprecision(5) << PY(2) <<" "
             << setw(10) << setprecision(6) << ySmooth(i,1)+start_long       <<" "
             << setw(9) << setprecision(5) << ySmooth(i,2) <<" "
             << setw(11) << setprecision(5) << PSmoothTrans(i,1,1) <<" "
           << setw(11) << setprecision(5) << PSmoothTrans(i,1,2) <<" "
             << setw(11) << setprecision(5) << PSmoothTrans(i,2,1) <<" "
             << setw(11) << setprecision(5) << PSmoothTrans(i,2,2) <<" "
             << setw(31) << "locations_NOT_observed"
           << endl;
   }
  }
  const int npma = 5;
  const double rmaden = 1.0/(double)npma;
  const int n2 = npma/2  + 1;
  double sumx = 0.0;
  double sumy = 0.0;
  dmatrix ZP(1,m,1,m);
  ZP.initialize();
  ofstream gmt(gmt_name);
  // these labels will cause GMT to complain, but shouldn't cause an error
  // they work with R
  gmt << "ox  oy  px  py  mx  my  ex  ey  smoothX  smoothY" << endl;
  for (int i = 1; i <= length; i++)
  {
   if ((i > 1)&&(step==1)&&(no_data_days==0))
   {
    if (blk(i) > 1)
      gmt << ">  >  >  >  >  >  >  >  >  >" << blk(i) <<endl;
   }
   sumx = 0.0;
   sumy = 0.0;
   if (i < n2)
   {
    sumx = y_deg(i,1);
    sumy = y_deg(i,2);
   }
```

```
      else  if (i > (length-n2) )
      {
        sumx = y_deg(i,1);
        sumy = y_deg(i,2);
      }
      else
      {
        int n1 = i - n2 + 1;
        int n2 = n1 + npma - 1;
        for (int nn = n1; nn <= n2; nn++)
        {
          sumx += y_deg(nn,1);
          sumy += y_deg(nn,2);
        }
        sumx *= rmaden;
        sumy *= rmaden;
      }

      gmt << y_deg(i,1) << "  " << y_deg(i,2);
      if (i==1)
        gmt << "  " << y_deg(i,1) << "  " << y_deg(i,2);
      else
      {
        dvector ta = value(z(a(i)));
        gmt << "  " << (ta(1)+start_long) << "  " << ta(2);
        dmatrix PP = value(P(i));
        ZP = ((value(Z)*PP+epss)/i);
      }
      gmt << "  " << sumx << "  " << sumy;
      gmt << "  " << ZP(1,1) << "  "  << ZP(2,2);
      gmt << "  " << ySmooth(i,1)+start_long << "  "  << ySmooth(i,2);
      gmt << endl;
    }
    report << "\nPhase " << current_phase() << " tracks written to files "
         << mpt_name << " and "
         << gmt_name << endl;
    clogf << "\nPhase " << current_phase() << " tracks written to files "
         << mpt_name << " and "
         << gmt_name << endl;
    cout << "\nPhase " << current_phase() << " tracks written to files "
         << mpt_name << " and "
         << gmt_name << endl;
  }
TOP_OF_MAIN_SECTION
  arrmblsize=20000000;
  gradient_structure::set_MAX_NVAR_OFFSET(800315);
```

```
gradient_structure::set_CMPDIF_BUFFER_SIZE(3000000);
gradient_structure::set_GRADSTACK_BUFFER_SIZE(1000000);
gradient_structure::set_NUM_DEPENDENT_VARIABLES(1000);
```

## APPENDIX B - THE COOKBOOK:

## STEP-BY-STEP PROCESS FOR HOME RANGE ESTIMATION USING *KFUD*

- The first step is preparing the data file. Follow the instructions in the DATA section of the tpl file (Section 0) to ensure you have needed information. Be sure to set the active parameter for *u* and *v* as '0' to keep the filter using an unbiased random walk.
    - I found it easiest to prepare in Excel, and then save it as a .txt file with a name like kfud.dat. Be sure that the first portion ('kfud') be the same as a the .tpl file's name.
- Compile the tpl file. This can be done through a dos window. I like to do it in the program Textpad. If using Textpad, go to Tools -> Run. Command is Command for compiling is 'admb'. Parameters: 'kfud'.
- To get most probable track:
    - Command: 'kfud.exe'
    - Parameters: 'kfud'
    - You can view the various parameter estimate results in the kfud.rep file.
    - Three files will be produced that can be used to view the geolocation estimations: mpt_111111110.dat, gmt_111111110.dat, rstuff_111111110.dat
    - To plot track in R:
        - track = read.table("c:/.../kfud/rstuff_001111110.dat",head=TRUE)
        - pts = matrix(c(track$smoothX,track$smoothY),,2)
        - plot(pts,type='b')
- To get likelihood profiles of parameters and geolocations of interest
    - Ensure you have defined the parameters as described in Section 0
    - Run kfud:
        - Command: 'kfud.exe'
        - Parameters '-lprof kf'
    - A file will be produced titled 'variablename.plt' which returns the likelihood profile values
- To perform MCMC
    - Run kfud:
        - Command: 'kfud.exe'
        - Parameters: '-mcmc 100000 –mcscale –mcsave 50 kfud –mcdiag'
            - 100000 is the number of runs, and it will draw the results from every 50th run.
    - Run kfud again:
        - Command 'kfud.exe'
        - Parameters: '-mcmc 100000 –mcscale –mcsave 50 kf –mcdiag –mceval'

- On the screen, the results of the MCMC for the parameters you noted in the "mceval" function in *kfud* will be displayed. Once displayed, remove all text before and after the strings of lats and longs; save as a text file titled "Command Results."
- Distributions can also be viewed in the kfud.hst file

➢ To view MCMC lat/long results plotted in R
  - o tabl<-read.table("C:/.../kfud/Command Results.txt")
  - o perform the command dim(tabl)
    - Note the number of columns. The way this is set up, the lat and longs are on the same rows, so we need to read particular columns to extract the appropriate numbers in the proper order.
  - o Perform the following commands
    - long<-tabl[,y:z] #where z is the total number of rows, learned from the dim() function. y is z/2+2
    - lat<-tabl[,2:x] # where x is z/2
    - lat<-as.matrix(lat)
    - long<-as.matrix(long)
    - lat_t <- as.vector(t(lat))
    - long_t <- as.vector(t(long))
    - plot(long_t,lat_t,cex=.2,pch=3)

➢ To make the 95% home range contour using the *2D Kernel Density* method:
    - library(KernSmooth)
    - est_kfud <- bkde2D(cbind(long_t,lat_t), gridsize=c(201, 201), bandwidth=c(.5,.5))
    - kfud_z <- est_kfud$fhat/max(est_kfud$fhat)
    - contour(est_kfud$x1, est_kfud$x2, kfud_z, zlim=c(5*max(est_kfud$fhat)/100,max(est_kfud$fhat)),nlev=20,label="95% Home Range",levels=.05)