Transactions in GIS

Research Article

Transactions in GIS, 2015, 19(4): 516–534

Analyzing Animal Movement Characteristics From Location Data

Dipto Sarkar,* Colin A. Chapman,† Larry Griffin‡ and Raja Sengupta†

*Indraprastha Institute of Information Technology, New Delhi
†Department of Anthropology and McGill School of Environment, McGill University
‡WWT Caerlaverock Wetland Centre

Abstract

When individuals of a species utilize an environment, they generate movement patterns at a variety of spatial and temporal scales. Field observations coupled with location technologies (e.g. GPS tags) enable the capture of detailed spatio-temporal data regarding these movement patterns. These patterns contain information about species-specific preferences regarding individual decision-making, locational choices and the characteristics of the habitat in which the animal resides. Spatial Data Mining approaches can be used to extract repeated spatio-temporal patterns and additional habitat preferences hidden within large spatially explicit movement datasets. We describe a method to determine the periodicity and directionality in movement exhibited by a migratory bird species. Results using a High Arctic-nesting Svalbard Barnacle Goose movement data yielded undetected patterns that were secondarily corroborated with expert field knowledge. Individual revisits by the geese to specific locations in the breeding and wintering grounds of Svalbard, Norway and Solway, Scotland, occurred with a periodicity of 334 days . Further, the orientation of this movement was detected to be mostly north-south. During long-range migration the geese use the north-south oriented Norwegian islands as "stepping stones", Short-range movement between mudbank roosts to feeding fields in Solway also retained a north-south orientation.

1 Introduction

Animal movement is a fundamental activity for many species and establishing an understanding of it is a necessary prerequisite to gaining knowledge about their ecology, life history and behavior (Rubenstein and Hobson 2004; Nathan, 2008), or to simulate within agent- or individual-based models (Tang and Bennett 2010). Further, studies of this nature serve a variety of purposes, including understanding disease dynamics (Bonnell et al. 2010) and factors impacting wildlife conservation (Chetkiewicz et al. 2006). The study of animal movement has gained impetus in recent years with improvements in telemetric technologies which enable higher accuracies when gathering location coordinates. Current GPS tags typically have a spatial accuracy of 10 m in most ecological settings and the frequency of fixes can be altered to suit the field site, study duration, and movement speeds of the organism in question (Johnson and Ganskopp 2008). Current movement data (i.e. sequential locations of the subject recorded on relatively short time scales) obtained through GPS tags are very data rich when coupled with information about the landscape in which the animal moves. Analysis of these data provides a means to understand animal preferences (e.g. preferred habitat, avoidance of high risk predator or disease areas, territorial defence, social behavior).

Address for correspondence: Dipto Sarkar, Indraprastha Institute of Information Technology, New Delhi, India. E-mail diptos@iiitd.ac.in
Acknowledgments: We would like to thank Baz Hughe of Caerlaverock Wetland Centre, Wildfowl and Wetlands Trust, for reviewing the manuscript and providing expert field knowledge.

 doi: 10.1111/tgis.12114

GIScience has borrowed from and expanded on Computational Science and Spatial Ecology methods to analyze spatially-explicit movement data (Long and Nelson 2013). In this article, we propose a methodology to extract periodic and directional behavior from spatially explicit, time-stamped data to make inferences about animal movement. This information can be used to quantify the spatio-temporal behavior patterns in relation to the environment in which the movement is occurring. The information obtained becomes particularly important when used to understand the factors that underlie the behavior of an individual of a migratory species; GPS data from a single migratory Svalbard Barnacle Goose is used as a case study. In this article, we are concerned with finding areas that are visited often and with a defined periodicity by the Barnacle Goose. These areas assume special importance in the case of migratory birds such as the Bernacle Goose. The great distance covered by animals such as the Bernacle Goose and the remote locations they visit during migration makes it important to automate the procedure of finding these locations from large datasets. Once the locations have been identified, conservation efforts can be targeted at these locations.

Prior work on Barnacle Geese has shown that the geese migrate over a very narrow time span and usually keep to well defined routes that differ slightly in spring and autumn (Owen and Gullestad 1984; Griffin 2008). The birds do get frequently displaced if the weather during migration is unfavourable. Studies have also indicated that routes can change because of the competitive pressure in staging areas, as for example for the Russian Barnacle Goose population where birds have begun to exhibit a later start to the spring migratory period and a pronounced westward expansion of the breeding distribution (Eichhorn et al. 2006, 2009). It is thought that the geese are skipping the western Baltic staging sites due to increased competition for food, but in doing so also shortening their migratory pathways by up to 700 km. The skipping of spring staging areas in Norway has also been witnessed in the Svalbard Barnacle Goose (Griffin 2008; see Table 10 in Griffin et al. 2011).

2 Related Work

Recent interest in movement data and "movement ecology" (Nathan 2008), particularly those collected via GPS Telemetry, has given rise to a number of useful applications and numerous studies have been undertaken that leverage the in-depth information provided by GPS tags. Rodgers and Anson (1994) concluded that "GPS-based animal-location systems will set a new standard for habitat-resource utilization studies of large animals over the next five to 10 years". These studies range from estimating the home ranges of animals to understanding the space use and movements of animals (Burdett et al. 2007). Furthermore, several researchers have provided the means to analyze such telemetry data in order to understand the unknown rules followed by the study animals (Merrill and Mech 2003; Ropert-Coudert and Wilson 2005). This is achieved by augmenting the inherent locational information provided by the GPS with the spatial information handling capabilities of GIS, enabling more complex data analysis.

From an ecological perspective, several applications have been coupled with spatial analysis methods to provide a better understanding of animal behaviors (Patterson et al. 2009; Bonnell et al. 2013). This has led to the development of software packages built to exploit information from telemetry-based movement data. Though general purpose GIS software can still be used for analyzing these data, special purpose open source software are available that takes the environment in which a movement occurs into consideration. For example, the Geospatial Modelling Environment (Beyer 2012) analyzes animal movements from an ecological perspective, and allows these movements to be decomposed into component localized movements that can be correlated with environmental or habitat information and thus understood in the context in which they are occurring. In recent years, systems like Env-DATA (Dodge et al. 2013) have made it possible to augment entire GPS tracks with environmental data. A myriad of environmental information is usually available from disparate sources and in different formats. Env-DATA provides a convenient way to accumulate information from numerous sources, convert them into a standard format and annotate the GPS track. The richly annotated tracks opens up a new possibilities for researchers to explore and understand how the surrounding environment affects animal movement.

It is important to note that there are some technological issues related to GPS that affects their accuracy and thus utility. Kolodzinski and Resources (2010) studied the effects of home range measurements due to different sampling rates and concluded that errors of between 30–50% are possible due to the inappropriate selection of sampling frequencies. A more generalized study that quantified the effect of sampling rates on estimates of distance travelled, proximity and resource use was conducted by Johnson and Ganskopp (2008). Laube and Purves (2011) point out that temporal error in the GPS signal may affect locational analysis simply because of the continuous nature of the data collected. Hence, two temporally consequent data points collected by GPS in the exact same location differ in their coordinate values only because of this error. These studies are particularly important in the context of analyzing the periodicity and direction of movement data. A misrepresentation in the data might result in weak or inappropriate hypotheses being generated.

Further, analyzing the relative periodicity and directionality of movement is an important and integral (albeit understudied) part of a broader framework of movement-related studies in GIScience (Long and Nelson 2013). Specifically, pattern and cluster methods help identify similarity of movement behavior or locate places of repeat interaction or use (Long et al. 2010). For example, the use of “episodal movements” is important to capture repeat patterns in the behavior of a moving point object (Laube et al. 2007; Long et al. 2013). When integrated with distance (e.g. as in the case of the Dynamic Interaction methodology), it provides information about similarity of movement patterns for a pair of moving objects.

The huge amounts of data collected by GPS receivers are also a challenge to analyze manually. As a result, researchers have come up with automated data mining techniques to efficiently find interesting and useful patterns from the large datasets (Miller and Han 2009). Large scale deployment of radio collars have made it possible to study aggregation of many individuals or flocks (Gudmundsson and van Kreveld 2006; Nanni and Pedreschi 2006; Vieira et al. 2009, Buchin et al. 2011). Trajectories of moving objects have also been mined to identify information about what locations will be visited in the future (Monreale et al. 2009; Ying et al. 2011). Prediction of future visits are a particularly important aspect for commercial applications. In our application scenario, the path followed by the Goose is usually quite predictable. However, extracting important locations and interesting patterns for directed conservation efforts remains a challenge. In this regard, comprehensive frameworks that provide generic models and approaches for geographic data discovery (Laube et al. 2005; Guo et al. 2005; Dodge et al. 2008) are useful in that they extract important information about the moving objects and their environment. For example, automated detection of patterns from GPS tracks provides a concise overview of the behavior and relative motion of moving objects. A common approach to detecting behavior is to use pattern matching techniques, e.g. Laube et al. (2005) use their framework “REMO” to detect relative motion patterns in objects. Giannotti et al. (2007) also use the concept of pattern matching to extract trajectory patterns from source trajectories. Finally, as an object moves through its interaction with the environment, it produces artifacts in its trajectories. The detection of such artifacts is non-trivial and provide important information about object-environment interactions (Orellana et al. 2010). In this article, we build upon the Periodica algorithm proposed by Li et al. (2010) and test its utility for highlighting patterns of behavior in Svalbard Barnacle Goose locational data. We emphasize its robustness as it can detect periodicity in a data set and can deal with noisy and complicated cases. We also include directional information, and suggest that periodicity and directionality can be combined to study habitat use by the geese. The periodicity in the movement coupled with the directionality of movement provides important information about the of the Geese and sheds light upon the environment that it interacts with.

3 Methodology

We utilized elements of the Periodica algorithm (Li et al. 2010) to extract the periodic behavior embedded in a movement dataset. This was an important first step towards understanding how well the Periodica algorithm can be utilized to extract species-wide preferences.

3.1 Detecting Periodic Behaviors

The first part of the Periodica algorithm extracts the periodic patterns contained in the dataset (Li et al. 2010) and was utilized for this article as well. The part of the algorithm used is enumerated below:

Periodica Algorithm

1. Detecting periods:
   (a) Input: Movement sequence LOC = loc_1, loc_2, loc_3, . . .
   (b) Find the spots where the movements are concentrated using Kernel Density Estimation (KDE) Methods. The hotspots are O = {o_1, o_2, o_2, . . . , o_d}
   (c) Convert LOC to a binary sequence according to whether a particular point falls within o_i or not. Note that the result of this step is a total of *d* binary sequences, 1 for each hotspot
   (d) Discrete Fourier Transform (DFT) is performed on each of the sequences to find out the periods of movements in the sequence.
   (e) Any periodic behavior which is above a specific threshold is considered significant and to correspond to real periods.

Next, Li et al. (2010) suggested the use of Kernel Density Estimate (KDE) to determine the hotspots. However, KDE tends to have a smoothing effect and thus draws attention to spatially larger hotspots. Moreover, the output of KDE is not characterized by statistical significance measures such as z and p values. This implies that given the same output, application of different thematic thresholds in KDE will produce different results. Thus, we substitute KDE with Getis-Ord Gi* (Getis and Ord 1992) to come up with statistically significant hotspots.

Getis-Ord Gi* compares the local sum with the sum of all features in order to determine whether the local value is significantly different from the global one to be considered a hotspot. When the local sum in a zone is very different from the expected local sum and the difference is too large to be accounted for by random chance, a statistically significant z-score is produced (Mitchell 2005). Thus, the z-score based on the randomization hypothesis gives a objective measure of clustering of points. When building automated systems for detection of spatial patterns, the objective measures provided by the z-score remove subjectivity in identifying the areas of high activity.

Fourier analysis is a mathematical technique for uniquely describing a time series in terms of periodic constituents. Fourier Transforms are commonly used in Signal Processing to decompose a signal to a finite composition of sinusoidal waves. By transforming the original signal from the time domain to the frequency domain, it is easier to interpret the signal in terms of the frequencies present in it. Discrete Fourier Transforms are used in the domain of Digital Signal Processing to take into account the discrete nature of the data. Discrete Fourier Transforms (DFT) can be calculated using Fast Fourier Transform (FFT) in O (*nlogn*) time and hence used instead of DFT which is O (n^2) (O being a measure of time complexity). Not only does Fourier analysis give the components of a time series of data, it also gives special importance to the relative strength of its constituents. The output of the FFT can be plotted on Periodograms to discern the frequencies that are important. In a Periodogram, the power of each frequency is plotted against the frequencies themselves to understand the strength of a repeating signal in the data at that frequency. Thus, visualizing Peridograms makes it easy to find the frequencies of importance.

As pointed out by Li et al. (2010), trying to mine the periodic behavior of the animals from the entire dataset at once is futile. Hence the concept of hotspots is used, as it filters out the noise and converts the problem of period detection from a 2-dimensional space to a 1-dimensional (binary) space. Thus, the problem is greatly simplified, as instead of having to deal with points located in a two-dimensional space, now just a sequence of 0's and 1's capture the visits by an animal to the hotspots.

Once the hotspots have been found using Getis Ord Gi*, the location sequence can be converted to a binary sequence such that:

$$x_i = \begin{cases} 1, \textit{ if } loc_i \textit{ is within hotspot } o_j \\ 0, \textit{ otherwise} \end{cases} \tag{1}$$

This binary sequence B = $b_1, b_2, \ldots, b_n$ is then processed using DFT to yield the sequence of complex numbers $X_1, X_2, \ldots, X_n$. The power frequency $F_k = X_k^2$ is plotted to get the periodogram which gives a visualization of the frequency components present. However, to determine which frequencies are important, a threshold value needs to be determined. The threshold value of 99% confidence is obtained by taking 100 random permutations of the sequence B and then taking the 99th highest power from the highest power of each round as the threshold value.

3.2 Extending Periodica

In Li et al. (2010), movement sequences are interpolated with a constant time gap to fill in any missing data in a raw sequence. This is done to ensure calculation of periodicities from a continuous data source. However, we avoid using this step as interpolating the sequence has the effect of increasing any periodic behavior that is inherent to the dataset. Hence, frequencies that were not significant might also get amplified to an extent where they start becoming important, leading to false positives.

Further, the success of the Periodica algorithm lies in the fact that hotspots are used as the basis of finding periodicities. Kernel Density Estimate (KDE) was used to find these hotspots by Li et al. (2010). KDE is a popular statistical data smoothing technique used in GIS to characterize important locations (Silverman 1986). However, application of KDE in the context of animal tracking data has been highly criticized in the literature. One of the assumptions of KDE is that the points are generated independently whereas movement trajectories consist of non-independent point (Downs 2010). KDE is also sensitive to the shape of the point pattern of the tracking data (Downs and Horner 2008). Moreover, KDE is actually a simple density calculation based on user inputs of search radius and a raster cell size; and the hotspot outputs provided by KDE are generally visually more attractive. Kernel Estimates that are generally used for calculation include Gaussian, Quartic and Triangular Kernels (Smith and Bruce 2008). All the inputs required by KDE are highly subjective in nature and hence the results obtained are also subjective. To sidestep the limitations of KDE, techniques based on spatial statistical tests should be used for detecting hotspots. In this work, we use Getis-Ord Gi* instead of KDE as an objective way of identifying the most visited locations or "hotspots" for the Goose. Calculation of Getis-Ord Gi* statistics falls under the umbrella of inferential spatial pattern analysis techniques grounded in probability theory. By repeatedly generating random permutations of the data, the chances of getting a result as extreme as the one observed is calculated (Mitchell 2005). Thus, the z-score obtained is based on the randomization null hypothesis. If the z-score is large enough, that is, the chances of obtaining the observed pattern is small by random chance, it can be objectively stated that a detected hostspot is indeed statistically significant.

Getis-Ord Gi* works by looking at the value of each feature in the dataset in the context of its neighbour's values. The Hotspots reported by Getis-Ord Gi* have a p-value and z-value associated with them, thus ensuring that the hotspots are statistically significant. To be reported as a statistically significant hotspot it is not enough for a particular feature to have a high value, it has to be surrounded by other features of high value as well. The local sum for a feature and its neighbours is compared proportionally to the sum of all features; when the local sum is very different from the expected local sum, and that difference is too large to be the result of random chance, a statistically significant z-score results. Formally Getis-Ord Gi* is defined as (Esri 2012):

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \tag{2}$$

where x_j = attribute value for feature j, $w_{i,j}$ = spatial weight between features i and j, n = total number of features and

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{3}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2} \tag{4}$$

The value G^* returned by Equation (2) is the z-score itself indicating the statistical significance of the hotspot found.

3.3 Case Study: Svalbard Barnacle Geese

The data used here for the case study to infer the kinds of animal movement on which the algorithm works well were obtained from the Movebank website (www.movebank.org).

The primary data set obtained from Movebank was that of "Barnacle goose (*Branta leucopsis*) (Griffin.)". Twenty-two birds were tracked using 30 g or 45 g solar GPS ARGOS PTT tags (Microwave Telemetry Inc.) for approximately three years to understand their migratory patterns. The data were collected mainly to establish full details of the international migratory flyway, including the spring and autumn staging and breeding areas, as well as additional wintering sites.

The datasets used in this case study are based on location information collected at regular intervals every one to two hours. However, due to lack of battery power for these solar tags on some occasions, due to daylight and/or weather conditions, e.g. particularly during mid-September and mid-winter, some location readings were missing from the data sequence, and these gaps have been ignored. Provided that the animal is tracked over a long period of, in this case, two years, it is suggested that a few missing data points are just minor noise in the data set.

The migratory pattern of Svalbard Barnacle Geese is well known, and also quite evident from Figure 1. It spends the summer months in the cold High Arctic regions of Svalbard, and then migrates south for the winter to coastal Norway and then the Solway Firth, UK.

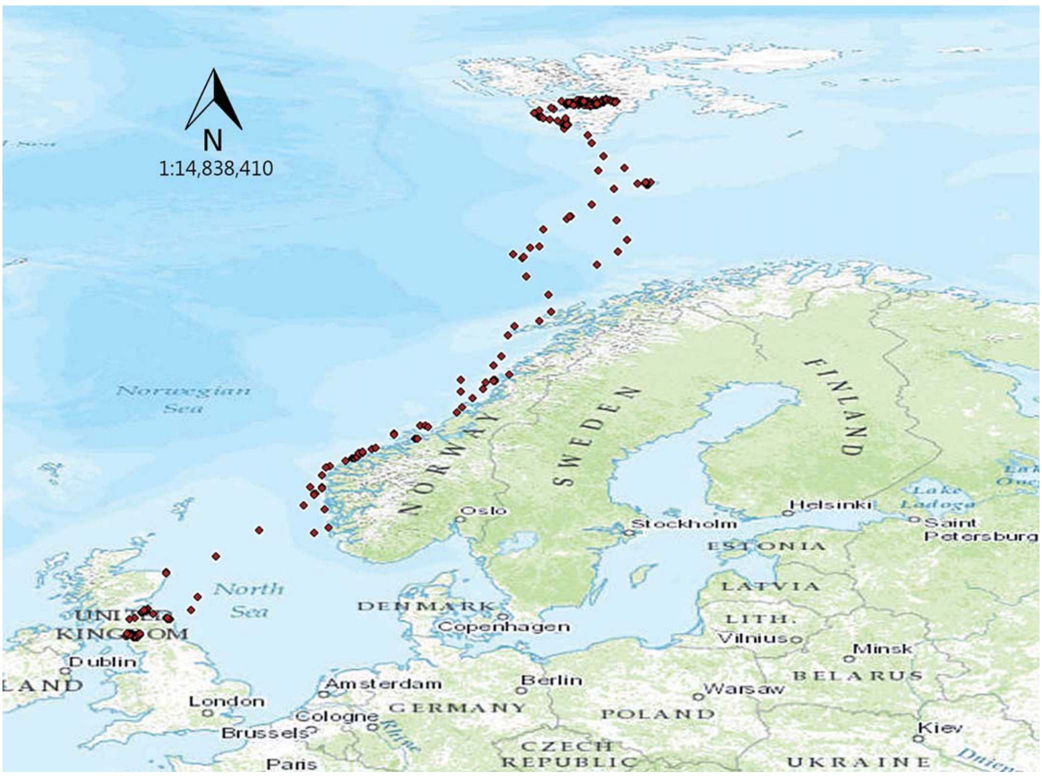


Figure 1 Return migratory GPS tracks of an individual Barnacle Goose from April 2007 to April 2009 (with ID 70563_1)

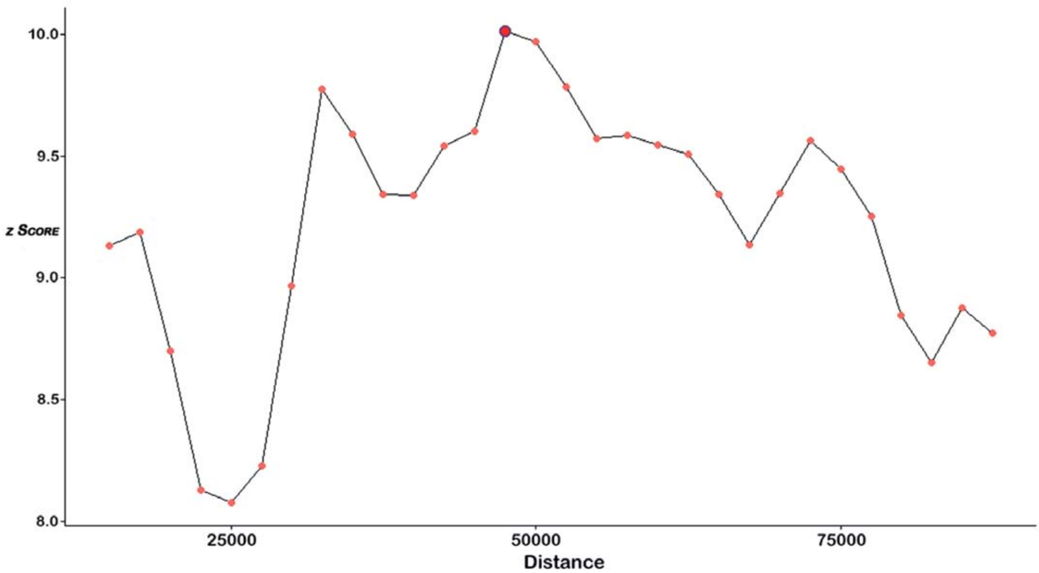


Figure 2 Graphical illustration of variation of autocorrelation with distance (measured in meters) as calculated by Global Moran's I. The highest peak is highlighted

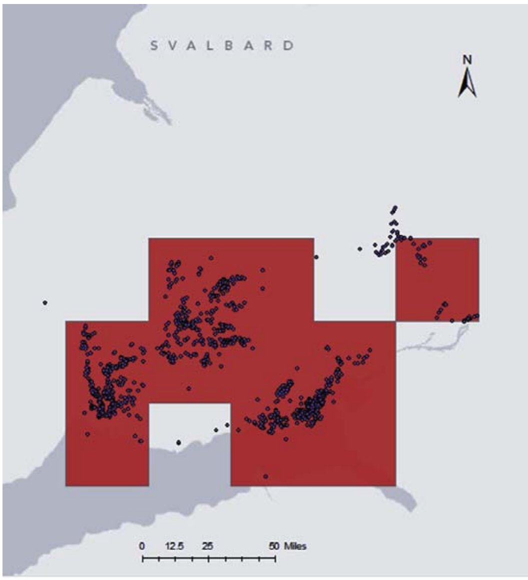
Location points for one individual (with ID 70563_1) were chosen from the dataset for analysis as the case study animal. This is important as using too many individuals at a time can lead to incomprehensible results. This is particularly important for datasets that have individuals from different herds or flocks.

Getis-Ord Gi* statistics were used to find the hotspots and the distance threshold was set to 50 km. To ensure that the scale of analysis is relevant for hotspot detection Global Moran's I Statistics was calculated for a range of distance thresholds to determine the distance at which spatial autocorrelation peaked. Plotting the results show that the autocorrelation peaks at about 50 km (Figure 2). Only the locations with a z-score greater than two were retained as hotspots to ensure 98% statistical confidence of the hotspots. Figures 3 shows the Hotspots that were obtained.

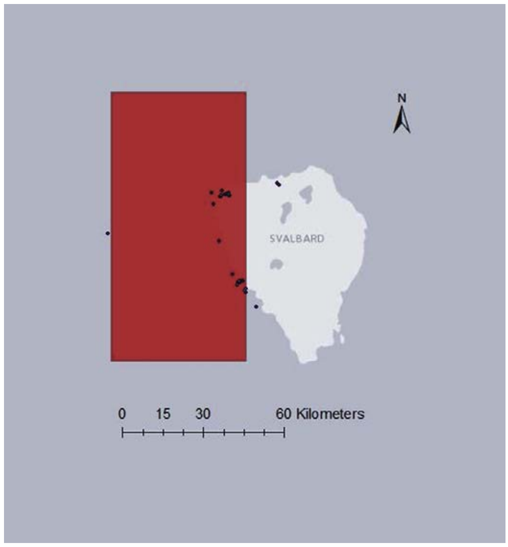
On running Periodica on each of the hotspots, the periodograms corresponding to each of them were obtained (Figures 4).

The Hotspots numbered 1 and 6 correspond to locations in the breeding grounds on Svalbard and wintering grounds on the Solway Firth, UK, respectively. The migration patterns in the case of Barnacle Geese are well defined and correspond to a periodicity of one year. This is evident from the periodograms of Figures 4a and f. The largest spike in the periodogram in both cases corresponds to a periodicity of exactly 334 days. However, it is to be noted that powers corresponding to several other frequencies are also above the threshold. This is probably because of the finite nature of the location sequence. Given a longer sequence these could probably have been nullified and the power corresponding to a period of 334 days (indicating a yearly migration pattern) would have been the only frequency above the threshold. Li et al. (2010) proposed a method to eliminate some of the frequencies above the threshold value. This method could also potentially eliminate unnecessary periods.

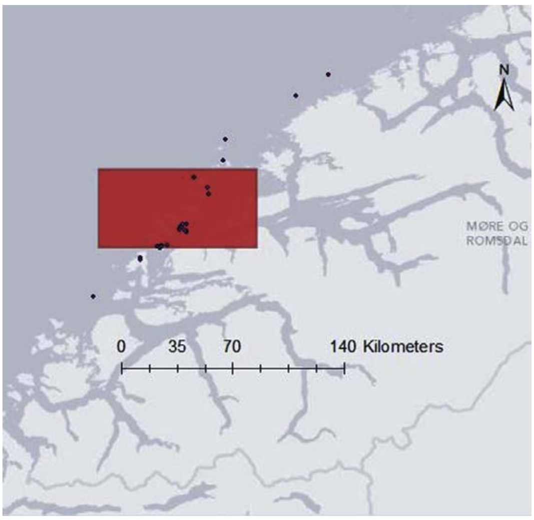
It is also worth noticing that in Hotspot locations 2, 3, 4 and 5, no dominant periodic behaviors can be seen. A look at the attribute table also shows that these locations were only



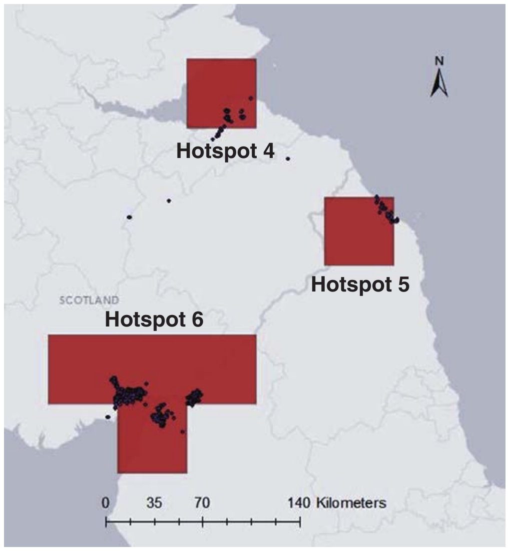
(a): Hotspot (sites in Van Mijenfjorden, Svalbard)



(b): Hotspot 2 (Björnöya, Svalbard)



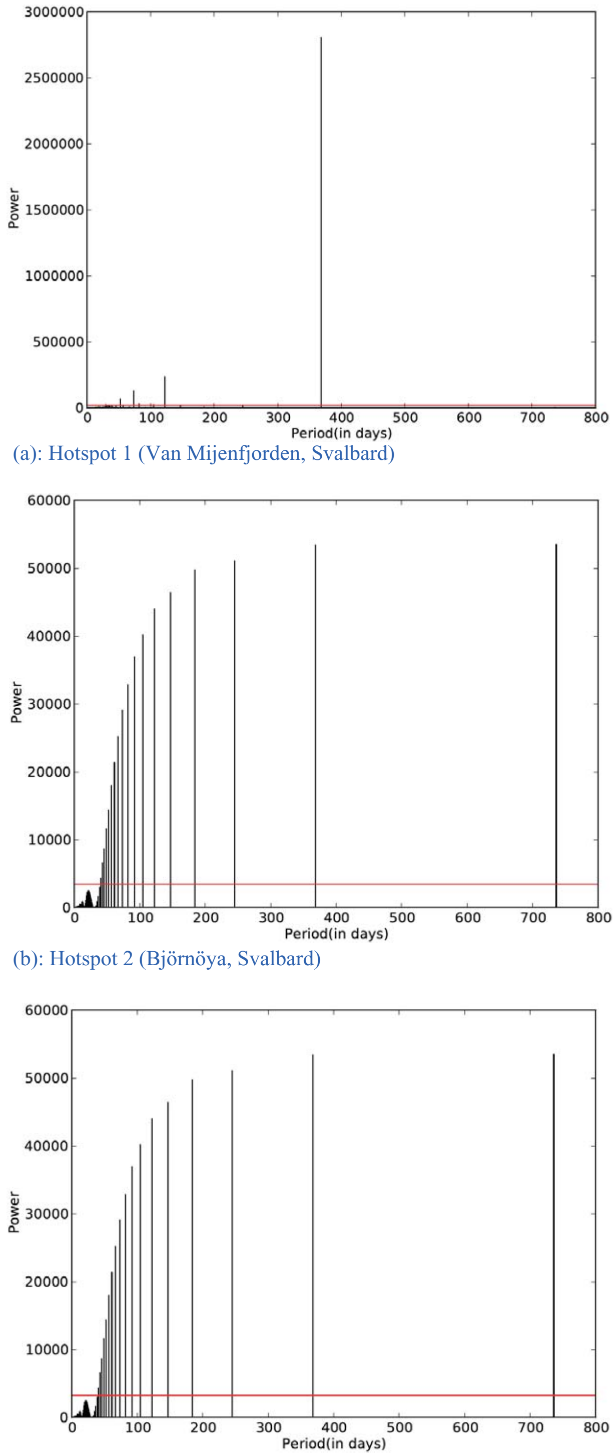
(c): Hotspot 3 (Ålesund, Norway)



(d): Hotspot 4, 5 and 6 (Aberlady, Lindisfarne and the Solway respectively)

Figure 3 Hotspot locations with 98% confidence and above for Barnacle Goose 70563_1

visited in one year and several points were generated because the bird stayed there for a few days, thus making the area qualify as a legitimate hotspot. However, looking at the Periodograms we can see that several powers are larger than the threshold shown by the red line. It is possible that the threshold value being generated is not high enough to eliminate the unnecessary periods. It will be interesting to see whether with considerably larger sequences these areas will still qualify as a hotspot, and if they do, whether the periods associated with these hotspots will be eliminated. This is one proposed area of future work.

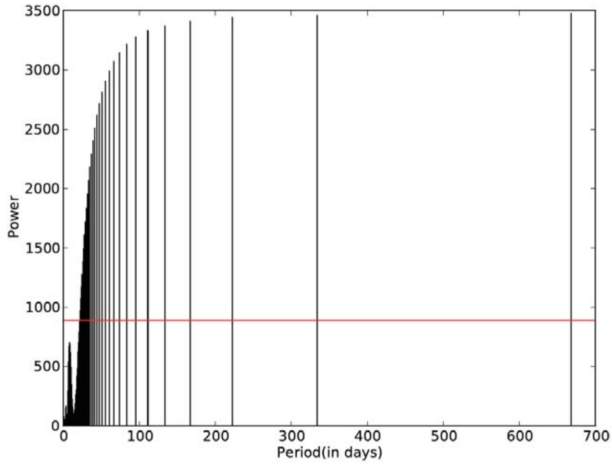


(a): Hotspot 1 (Van Mijenfjorden, Svalbard)

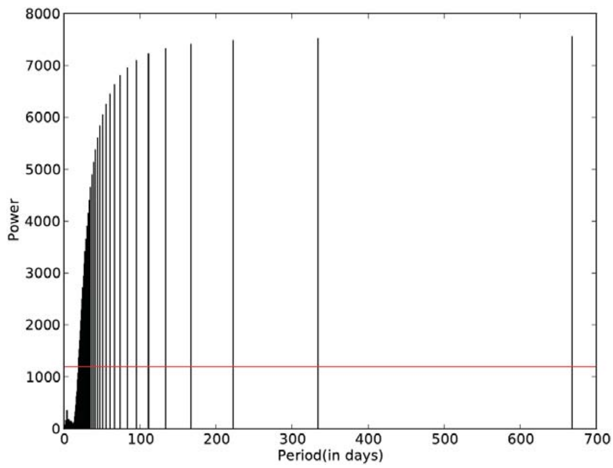
(b): Hotspot 2 (Björnöya, Svalbard)

(c): Hotspot 3 (Ålesund, Norway)

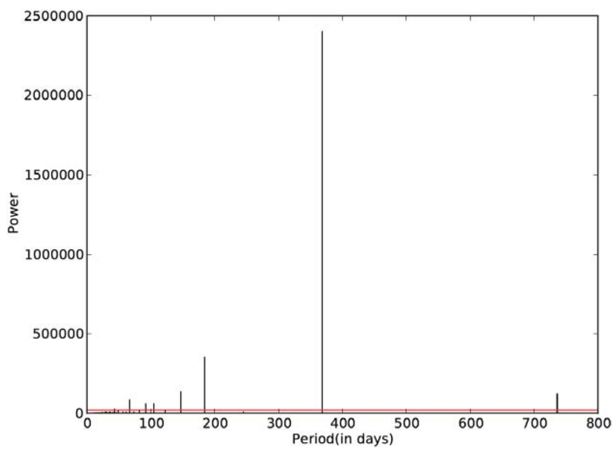
Figure 4 Periodograms corresponding to the hotspots



(d): Hotspot 4 (Aberlady, Scotland)



(e): Periodogram for Hotspot 5 (Lindisfarne)



(f): Periodogram for Hotspot 6 (Solway)

Figure 4 Continued

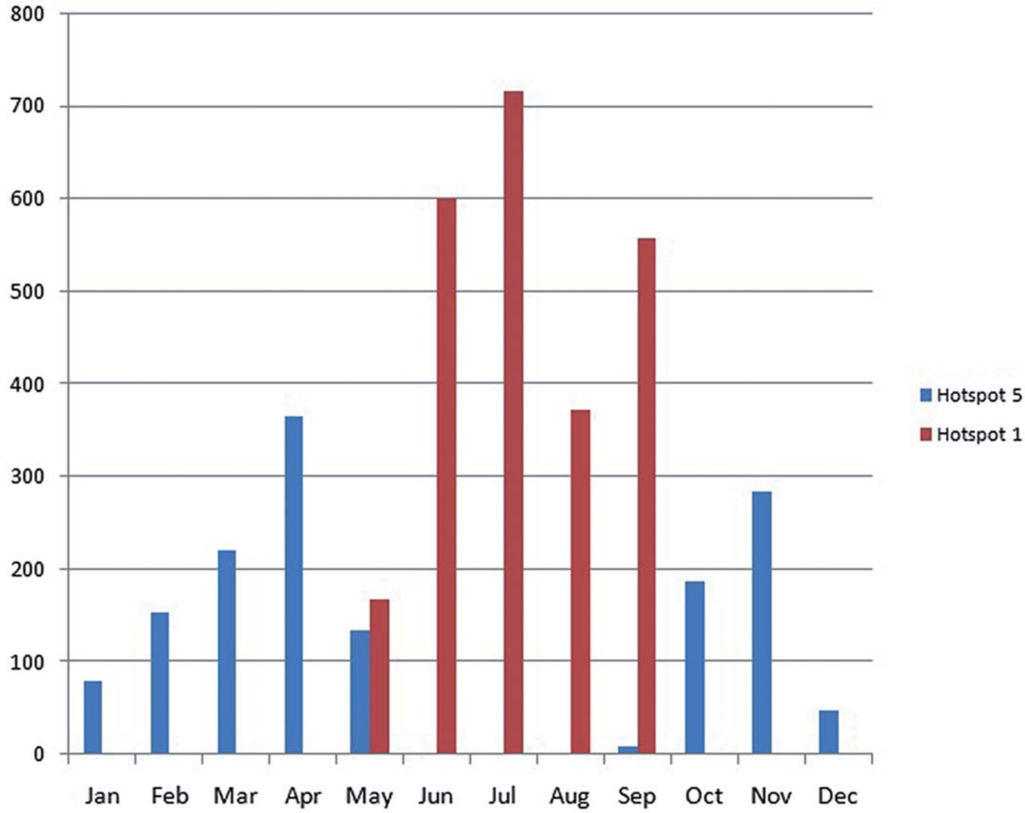


Figure 5 Location of the goose at different times of the year. The y axis shows the number of points within the hotspot

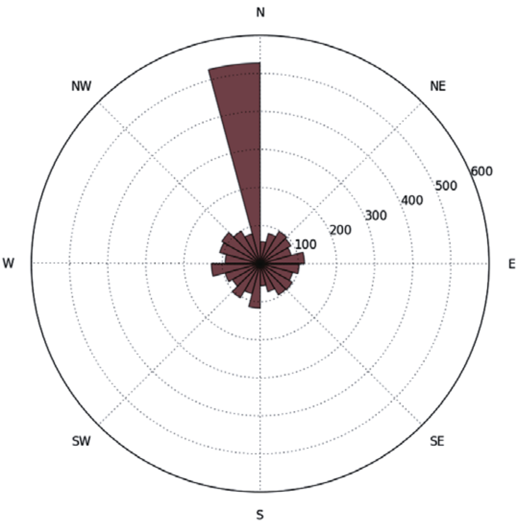
The location of the goose at different times of the year was also calculated and is shown in Figure 5. The distribution in Figure 5 further highlights the migratory behavior of the goose and explains the periodicity obtained from the periodograms of Hotspots 1 and 6.

3.4 Directionality

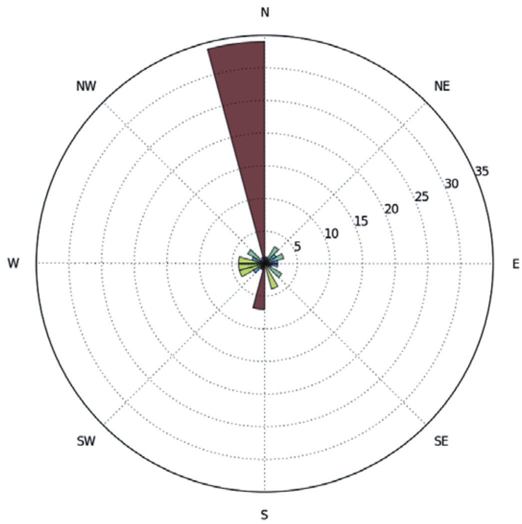
Rose diagrams have been used extensively to explain the frequency of lineations in a given orientation, including marine ecological applications (Torres et al. 2013). This concept can be used to understand the directionality in the movement patterns of the animal when they are inside and outside the hotspots.

The rose diagram constructed for this purpose considers the angle between subsequent locations recorded by a GPS tag. The directions are classified into 15^0 groups and then plotted on the rose diagram. Altering the size of the classification bin has the effect of changing granularity at which the information is represented in the rose plot. Smaller bin sizes capture slight differences in movement direction whereas a larger bin size gives a broad overall picture.

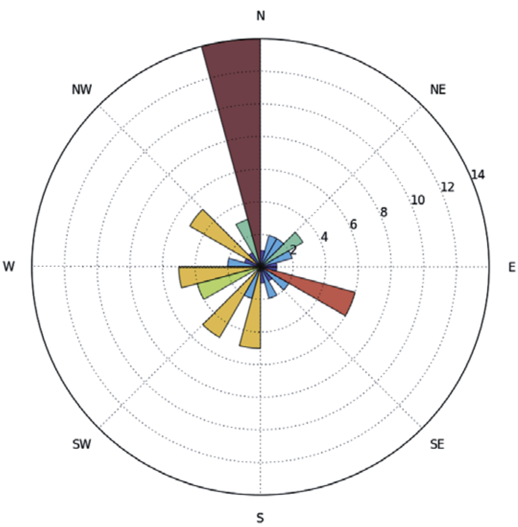
The rose diagrams corresponding to each of the six hotspots for Barnacle Goose 70563_1 are shown in Figures 6, which highlights the fact that there is a strong directionality in the movement of this goose inside the hotspots. Most of the hotspots in Figure 5 show strong



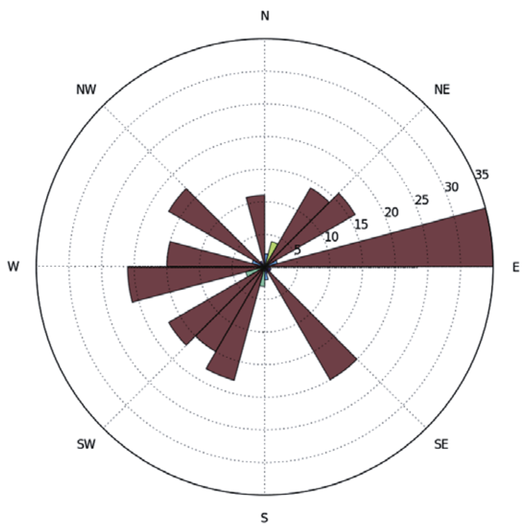
(a): Hotspot 1 (Ålesund, Norway)



(b): Hotspot 2 (Björnöya, Svalbard)



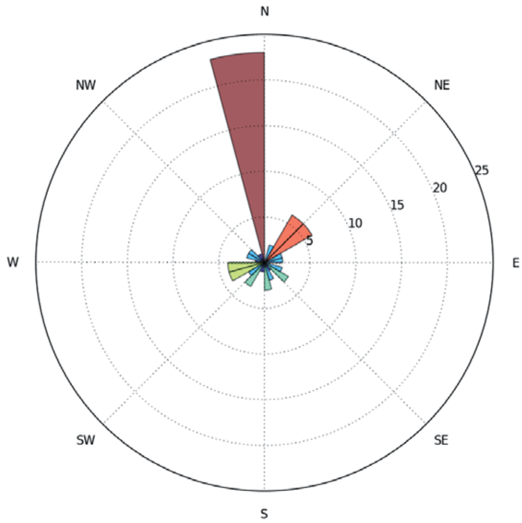
(c): Hotspot 3 (Ålesund, Norway)



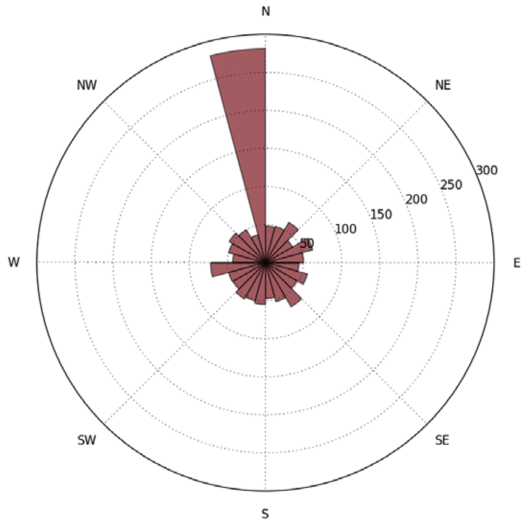
(d): Hotspot 4 (Aberlady, Scotland)

Figure 6 Rose diagrams corresponding to the various hotspots

Northern movement. It is interesting to note that the strong North-South movement pattern of the migratory route is mostly preserved in the local movements as well. In fact, if we exclude the points falling inside any of the hotspots (which denotes local movement) and use only the other points (denoting large-scale displacements) to make the rose diagram, then the strong North-South directionality in the overall movement of the goose gets captured, as shown by Figure 7. Thus, both local and large-scale displacements of the goose confirms earlier observations that geese usually keep to well defined routes that differ slightly in spring and autumn unless displaced by bad weather (Owen and Gullestad 1984). And this north-south pattern of movement holds true even for local movements.



(e): Hotspot 5 (Lindisfarne)



(f): Hotspot 6 (Solway)

Figure 6 Continued

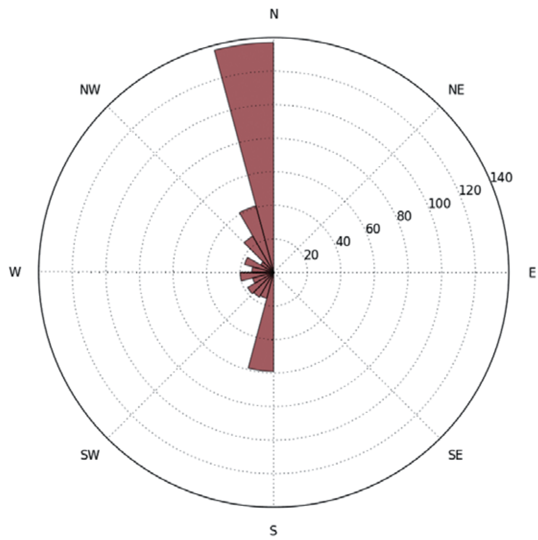


Figure 7 Rose diagram of GPS points outside hotspots

The constrained North-South movement of the goose inside the hotspot as pointed out by the algorithm is evidently counter intuitive. There should be no biological reason for a limited movement within the hotspots as birds are simply feeding or moulting or roosting at those locations. Hence the movements inside the hotspots should ideally be random, i.e. distributed evenly in all compass directions. However, closer examination of the hotspot sheds light on the North-South movement trend inside the hotspots. The hotspots corresponding to the Norwegian islands and the one at Svalbard were used as "stepping stones" in the annual migration in different years, i.e. the "stepping stone" at Norwegian Island was only used in 2008 and the one at Svalbard in 2007. Hence, the progressive Northward movement in these locations manifested itself as strong directional patterns in the rose diagrams corresponding to these locations. Moreover, in the hotspot located at Solway, the movement from the mudbank to the feeding site also happened to be oriented North-South. The chance location of these two sites at such an orientation leads to the strong directionality present in the movement inside this hotspot.

This validation further highlights the robustness of the technique. Data mining approaches when used on large datasets may point out patterns that might prove interesting and useful on further investigation (Clifton 2013). The constraint directionality of movement of the Goose inside the hotspots in this case has proven interesting on further analysis.

Overall, this indicates an ability of the diagrams to capture and show directionality for both local movements as well as large-scale displacements due to migration. This coupled together with periodicity identifies sub-movements within large scale displacements of migratory animals as a useful technique for analyzing animal movement information. Future work will need to tie these analysis techniques to specific landscape characteristics, and to understand the factors that introduce such periodicity and directionality in movements.

4 Conclusions

The framework presented in this article highlights the fact that there is strong periodicity and directionality in animal movement data which can be captured by an automated processes. This is particularly important as periodicity and directionality are two important drivers in the understanding of animal behavior movement patterns under the broad umbrella of movement ecology (Nathan 2008). With the wide deployment of GPS tags on a variety of organisms, huge amounts of data are captured every day, which makes it impossible to analyze them manually. Automated frameworks like the one presented can help in distilling large datasets of GPS tracking data. It can help identify frequently visited places, periodicities associated with these places (if any), and the directionality present in the movement data at the macro as well as micro scale. Furthermore, this method can be used to gain further insights even for hotspots which do not show significant periods associated with them. These areas may perhaps indicate failed nesting sites to which the bird no longer returns. Locating such sites by visual inspection of data can be a tedious task, and automating the process helps in quickly isolating regions for further analysis. Similarly, hotspots that have strong directionality associated with them can also be probed for landscape characteristics that influence the movement. For, example the North-South movement associated with the hotspots were not known apriori, these were detected through implementing the algorithm and expert field knowledge corroborated the reasons for it. Detection of such hidden patterns become more and more important with the increasing use of GPS tags. It is impossible to manually go through the huge datasets searching for interesting, unexpected patterns and automated detection of interesting patterns becomes important.

The results of the framework can also help to succinctly summarize the movement of the animals in terms of periodicity (Li et al. 2010) and directionality. For example, from the example of Barnacle Goose 70563_1 we can conclude that:

1. There are two specific hotspots (regions), one in the UK and one in Svalbard which it visits annually.
2. There is a tendency for north-south movement inside these hotspots.

3. Overall, the migratory pattern of the Barnacle Goose has a strong North-South directionality, which it maintains inside most of the hotspots due to the orientation of the resources located there.

5 Discussion

Owen and Gullestad (1984) indicated that Barnacle Geese migrate over a short period of time and tend to follow very well defined pathways, observations in Scotland in autumn indicating two regular routes followed by the geese: one that was almost directly North to South through Shetland/Orkney and into NE mainland Scotland, and the other from the East-North-East that passes over Northern England, but noting that frequent deviations occurred because of climatic conditions. To an extent this is repeated in reverse in the spring, although there are more concentrated exit points from the North-East UK coast closer to the Solway, whereas in autumn there are more diffuse arrival points into the UK because during the long sea crossing from Norway, wind drift has an opportunity to displace birds north or south of where they might be navigating for.

This was confirmed by the results obtained using the Periodica algorithm and the rose diagrams. A periodicity of 334 days indicates an annual revisit by an individual of the species, with little time spent en route. Further, the directionality indicated that the individual trended almost north-south, indicating that this member of the species followed the typical migration corridor specified above (barring poor weather conditions). Interestingly, it exhibits this specificity in directionality even for local movements. Thus, the geese rarely deviate from their fixed trajectories, except in the light of competitive pressures that lead to new migration strategies (Eichhorn et al. 2006, 2009). The techniques presented above thus corroborate field surveys and other understandings of the movement patterns of the Barnacle Geese used as a case study. It also serves to highlight the potential for future non-subjective inspection and summary of movement behavioral data for other species as well.

If these techniques are to be used for other species, one should note that the Periodica Algorithm works best for animals that exhibit migratory behavior, because migratory patterns have well defined hotspots which are repeatedly visited at regular time intervals. Moreover, the large spatial separation between the hotspots makes them easy to find. Further, if the animal has well defined periodic behaviors within a small (local) range, the scale of analysis needs to be chosen carefully to extract the hotspots accurately. Moreover, it must be ensured that supporting spatial data (e.g. habitat information) is also available at that scale to justify any further analysis that may be performed.

It is also to be noted that the Periodica Algorithm works best when considerable amounts of data are available for the individuals. Continuous GPS recording for a few consecutive years is required to definitively identify the periods and eliminate irrelevant powers from the periodogram. For example, consider the periodograms for hotspots numbered 2, 3, 4 and 5 for the Barnacle Goose. A look at the attribute tables for the points falling inside these hotspots reveal that these hotspots contain entries for a single year only. Thus, these areas should not have shown any periodic behavior at all. But, despite using threshold generated from the data itself, several powers associated with certain frequencies are above the threshold. This means that using the proposed techniques on datasets of animals that do not show migratory patterns and are of limited temporal duration (less than one year) may not adequately determine the frequency of revisit and the directionality of movement.

The fact that Bernacle Goose migrate in a N-S direction was already well known. However, finding important stopover locations remained a difficult task. Sites that have many points in them may not be the most important locations in the course of its migration. Many sites are detected as 'hotspots' due to the large number of points that fall inside them. However, a closer look reveals that this site may just have been visited during a single year and the bird stayed there for a while generating a lot of location samples. The birds in the course of their migration do not follow the exact same path every year. The locations which were visited only once highlight such deviations from the usual route. On the other hand, the locations that are repeatedly visited by the goose are of particular importance from the perspective of conservation. The advantage of knowing the directionality of movement within these location gives ecologists a dashboard view of what the bird might have been doing there. The strong N-S directionality inside hotspot in Solway on further investigation was revealed to be primarily due to the orientation of the feeding site and mudbank in that location. Some hotspots which are routinely revisited may have no dominant directionality, indicating that these are locations where the birds usually stop over for foraging.

6 Future Work

We aim to create an analysis framework that can be used to mine behavioral rules which are not specific to certain individuals or groups of animals. We want to extract rules from several individuals from different groups residing in different areas. These rules can then be compared against each other to extract the traits that hold for the entire species as a whole. Thus, we can reject certain behavioral rules that a particular group has adapted to survive in its particular environment and get only the preferences that are applicable to the species as a whole. This principle has guided the work presented in this article. The main aim was to automate the major part of the procedure as it needs to be reused and run several times to data mine multiple GPS tracks.

We aim to determine the similarity between hotspots that have the same periodicity. The hotspots of the same periodicity mined from different GPS tracks will be annotated with important spatial characteristics, and then grouped together using similarity measures. The problem of finding similarity in this situation is clearly one that cannot be solved by the Supervised Learning techniques of Machine Learning, due to the absence of the concept of training sets. However, unsupervised learning techniques can be used to understand the similarity between the hotspots. Several techniques of unsupervised learning may be tried on the datasets to see which yield better results. Some of the techniques that can be used include k-means clustering and Locality Sensitive Hashing (LSH). The k-means algorithm is one of the common algorithms used to group similar items together. But the disadvantage is that the number of groups required must be specified a priori. LSH is an algorithm for solving the near and approximate Near Neighbor Search in high dimensional spaces (Andoni and Indyk 2006). LSH closely models the clustering and k-nearest neighbour search. The main idea behind LSH is to put items in buckets (hashing) so that the items that are similar to each other are put together in the same bucket with a high probability. Given the high dimensionality of our problem, it seems that LSH based approaches will be appropriate for the future work.

References

Andoni A and Indyk P 2006 Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Proceedings of the Forty-seventh Annual IEEE Symposium on Foundations of Computer Science (FOCS '06)*, Pittsburgh, Pensylvania: 459–68

Beyer H L 2012 Geospatial Modelling Environment (Version 0.7.2.0). WWW document, http://www.spatialecology.com/gme

Bonnell T R, Sengupta R, Chapman C A, and Goldberg T L 2010 An agent-based model of red colobus resources and disease dynamics implicates key resource sites as hot spots of disease transmission. *Ecological Modelling* 221: 2491–500

Bonnell T R, Dutilleul P, Chapman C A, Reyna-Hurtado R, Hernández-Sarabia R U, and Sengupta R 2013 Analysing small-scale aggregation in animal visits in space and time: the ST-BBD method. *Animal Behaviour* 85: 483–92

Burdett C, Niemi G J, and Mech L D 2007 Defining space use and movements of Canada Lynx with Global Positioning System telemetry. *Journal of Mammalogy* 88: 457–67

Buchin K, Buchin M, Gudmundsson J, Löffler M, and Luo J 2011 Detecting commuting patterns by clustering subtrajectories. *International Journal of Computational Geometry and Applications* 21: 253–82

Chetkiewicz C-L B, St. Clair C C, and Boyce M S 2006 Corridors for conservation: Integrating pattern and process. *Annual Review of Ecology, Evolution, and Systematics* 37: 317–42

Clifton C 2013 Data Mining. *Encyclopædia Britannica Online* (available at http://www.britannica.com/EBchecked/topic/1056150/data-mining)

Dodge S, Bohrer G, Weinzierl R, Davidson S C, Kays R, Douglas D C, Cruz S, Han J, Brandes D, and Wikelski M 2013 The environmental-data automated track annotation (Env-DATA) system: Linking animal tracks with environmental data. *Movement Ecology* 1: 3

Dodge S, Weibel R, and Lautenschütz A K 2008 Towards a taxonomy of movement patterns. *Information Visualization* 7: 240–52

Downs J A 2010 Time-geographic density estimation for moving point objects. In Fabrikant S I, Reichenbacher T, van Kreveld M, and Schlieder C (eds) *Geographic Information Science: Sixth International Conference, GIScience 2010, Zurich, Switzerland, September 14–17, 2010 Proceedings*. Berlin, Springer Lecture Notes in Computer Science 6292: 16–26

Downs J A and Horner M W 2008 Effects of point pattern shape on home-range estimates. *Journal of Wildlife Management* 72: 1813–18

Eichhorn G, Drent R H, Stahl J, Leito A, and Alerstam T 2009 Skipping the Baltic: The emergence of a dichotomy of alternative spring migration strategies in Russian Barnacle Geese. *Journal of Animal Ecology* 78: 63–72

Eichhorn G, Afanasyev V, Drent R H, and van der Jeugd H P 2006 Spring stopover routines in Russian Barnacle Geese Branta leucopsis tracked by resightings and geolocation. *Ardea* 94: 667–78

Esri 2012 How hot spot analysis (Getis-Ord Gi*) works. WWW document, http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//005p00000011000000

Getis A and Ord J K 1992 The analysis of spatial association. *Geographical Analysis* 24(3): 127–45

Giannotti F, Nanni M, Pinelli F, and Pedreschi D 2007 Trajectory pattern mining. In *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Jose, California: 330–39

Griffin L 2008 Identifying the pre-breeding areas of the Svalbard Barnacle Goose Branta leucopsis between mainland Norway and Svalbard: An application of GPS satellite-tracking techniques. *Vogelwelt* 129: 226–32

Griffin L, Rees E, and Hughes B 2011 *Migration Routes of Whooper Swans and Geese in Relation to Wind Farm Footprints: Final Report*. Slimbridge, UK, The Wildfowl and Wetlands Trust

Gudmundsson J and van Kreveld M 2006 Computing longest duration flocks in trajectory data. In *Proceedings of the Fourteenth annual ACM International Symposium on Advances in Geographic Information Systems*, Arlington, Virginia: 35–42

Guo D, Gahegan M, MacEachren A M, and Zhou B 2005 Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science* 32: 113–32

Johnson D and Ganskopp D 2008 GPS collar sampling frequency: Effects on measures of resource use. *Rangeland Ecology and Management* 61: 226–31

Kolodzinski J J and Resources N 2010 Effects of GPS sampling intensity on home range analyses. In *Proceedings of the Annual Meeting of the Southeastern Association of Fish and Wildlife Agencies*, Biloxi, Mississippi: 13–7

Laube P, van Kreveld M, and Imfeld S 2005 Finding REMO: Detecting relative motion patterns in geospatial lifelines. In Fisher P F (ed) *Developments in Spatial Data Handling* (pp. 201–215). Berlin, Springer: 201–15

Laube P and Purves R S 2011 How fast is a cow? Cross-scale analysis of movement data. *Transactions in GIS* 15: 401–18

Laube P, Wolle T, and Gudmundsson J 2007 Movement patterns in spatio-temporal data. In Shekhar S and Xiong H (eds) *Encyclopedia of GIS*. Berlin, Springer-Verlag: 726–32

Li Z, Ding B, Han J, Kays R, and Nye P 2010 Mining periodic behaviors for moving objects. In *Proceedings of the Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC: 1099–108

Long J and Nelson T 2013 Measuring dynamic interaction in movement data. *Transactions in GIS* 17: 62–77

Long J, Nelson T, and Wulder M 2010 Regionalization of landscape pattern indices using multivariate cluster analysis. *Environmental Management* 46: 134–42

Long J, Nelson T, and Wulder M 2013 A review of quantitative methods for movement data. *International Journal of Geographical Information Science* 27: 292–318

Merrill S and Mech L 2003 The usefulness of GPS telemetry to study wolf circadian and social activity. *Wildlife Society Bulletin* 31: 947–60

Mitchell A 2005 *The Esri Guide to GIS Analysis: Spatial Measurements and Statistics* (Volume 2). Redlands, CA, Esri Press

Miller H J and Han J (eds) 2009 *Geographic Data Mining and Knowledge Discovery*. Boca Raton, FL, CRC Press

Monreale A, Pinelli F, Trasarti R, and Giannotti F 2009 Wherenext: A location predictor on trajectory pattern mining. In *Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France: 637–46

Nanni M and Pedreschi D 2006 Time-focused clustering of trajectories of moving objects. *Journal of Intelligent Information Systems* 27: 267–89

Nathan R 2008 An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences* 105: 19050–51

Orellana D, Wachowicz M, De Knegt H, Ligtenberg A, and Bregt A 2010 Uncovering patterns of suspension of movement. In *Proceedings of the Sixth International Conference on Geographic Information Science (GIScience 2010)*, Zurich, Switzerland

Owen M and Gullestad N 1984 Migration routes of Svalbard Barnacle Geese Branta leucopsis with a preliminary report on the importance of the Bjørnøya staging area. *Norsk PolarInstitutt Skrifter* 181: 67–77

Patterson T A, Basson M, Bravington M V, and Gunn, J S 2009 Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology* 78: 1113–23

Rodgers A and Anson P 1994 Animal-borne GPS: Tracking the habitat. *GPS World* 5: 20–32

Ropert-Coudert Y and Wilson R 2005 Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment* 3: 437–44

Rubenstein D R and Hobson K A 2004 From birds to butterflies: animal movement patterns and stable isotopes. *Trends in Ecology and Evolution* 19: 256–63

Silverman B W 1986 *Density Estimation for Statistics and Data Analysis*. London, Chapman and Hall

Smith S C and Bruce C W 2008 Kernel density estimation. In Smith S C and Bruce C W (eds) *CrimeStat III User Workbook*. Washington, DC: National Institute of Justice: 62–77

Tang W and Bennett D A 2010 Agent-based modeling of animal movement: A review. *Geography Compass* 4: 682–700

Torres L G, Sagar P M, Thompson D R, and Phillips R A 2013 Scaling down the analysis of seabird-fishery interactions. *Marine Ecology Progress Series* 473: 275–89

Vieira M R, Bakalov P, and Tsotras V J 2009 On-line discovery of flock patterns in spatio-temporal data. In *Proceedings of the Seventeenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, Washington: 286–95

Ying J J C, Lee W C, Weng T C, and Tseng V S 2011 Semantic trajectory mining for location prediction. In *Proceedings of the Nineteenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, Illinois: 34–43

