Applications of Fractal Analysis Techniques to Animal Positional Datasets and the Field of Ecology

Spencer B. Fradkin

Acknowledgements:

I would like to give a special thanks to Dr. Richard Voss for all of his time and hard work while introducing me to the study of fractal geometry and data analytics. Through this process, I have gained even more respect for abstract fields of mathematics and their applications to real world situations. Without him, this project would have been impossible. I would also like to thank the Yellowstone Wolf Project (Matthew C. Metz) for supplying the data on wolf positions in Yellowstone National Park.

# Contents:

1. Introduction:

      With GPS tracking technology becoming more readily available, there is an abundance of animal tracking data accessible to ecologists, statisticians, and mathematicians alike. The increase of this type of data in the past few decades has been met with a demand for new ways to characterize and analyze it. In light of this, there are numerous ways ecologists analyze animal GPS positional data, each providing a unique perspective on some aspect of the data. Although different in practice, most of these analyses use the same parameters such as the distance between consecutive coordinates, the displacement of an animal from a specific starting point, the straightness of an animal's path, or the time it takes an animal to reach or return to or from a specific location. [1]

      One of the more well-known methods for characterizing animal GPS data is Path Segmentation. [2] Path Segmentation is typically used when GPS data is recorded at smaller intervals, creating a detailed path for each animal that can then be broken into small segments for further analysis. [2] The distances between adjacent coordinates in the GPS data are calculated, and lines are drawn between coordinates creating a path. [2] The accuracy of these individual paths varies based on how often the GPS data was collected. Once the animal positional data is broken into small segments based on what aspects of the animal's path are of interest, ecologists can more clearly see exactly where relevant changes may occur in the animal's movement data. For example, these recorded changes could be influenced by habitat, the time of the day, or seasonal climate. Statisticians can then create plots of segment lengths or directions versus their habitat, the time of the day, or seasonal climate as done by Edelhoff, Signer, and Balkenhol. [2]

      Another method for characterizing animal GPS data is a Space-Use Analysis. Space-Use Analyses use animal GPS positional data to estimate 2-dimensional spaces where certain animals

are most likely to reside. [1] There are two primary ways to perform a Space-Use Analysis. The most common way is the Minimum Convex Polygon method. When using this method, statisticians draw the minimum convex polygon that covers a certain percentage of the points.[1] The percentage of points that is chosen typically depends on the spread of the data. Although very easy to utilize, this method often leaves large gaps of space where animals may not actually be residing. When a more rigorous estimation is required, the Local Convex Hull method (LoCoH) is usually employed.[3] This method entails forming a convex hull around the GPS data and constructing a utilization distribution that shows the probabilities of finding an animal inside the convex hull within a specific time period. [3] Examples of both methods being applied can be found in Seidel, where the differences between these methods are compared using home ranges of zebras. [1]

A less common class of methods for characterizing animal GPS coordinates are fractal methods. These methods are particularly applicable to ecology and the natural environment as most natural phenomena are not well modeled by Euclidean shapes.[7] A typical fractal analysis measures the roughness or complexity of a dataset by looking at how its detail changes while the observational scale changes in time or space. Once this analysis is complete, each dataset will be assigned a fractal dimension or scaling exponent that can be related to other characterizations of interest. This method has been applied multiple times when characterizing the tortuosity of animal trails. In 1988, Marcel Dicke characterized the tortuosity of mite trails on leaves. [4] Additionally, in 2009, Stephen Webb used fractals to characterize the tortuosity of white-tailed deer trails. [5] Although there is evidence of this method being used (as mentioned above), these uses are relatively few and far in between when compared to the more common methods reviewed earlier.
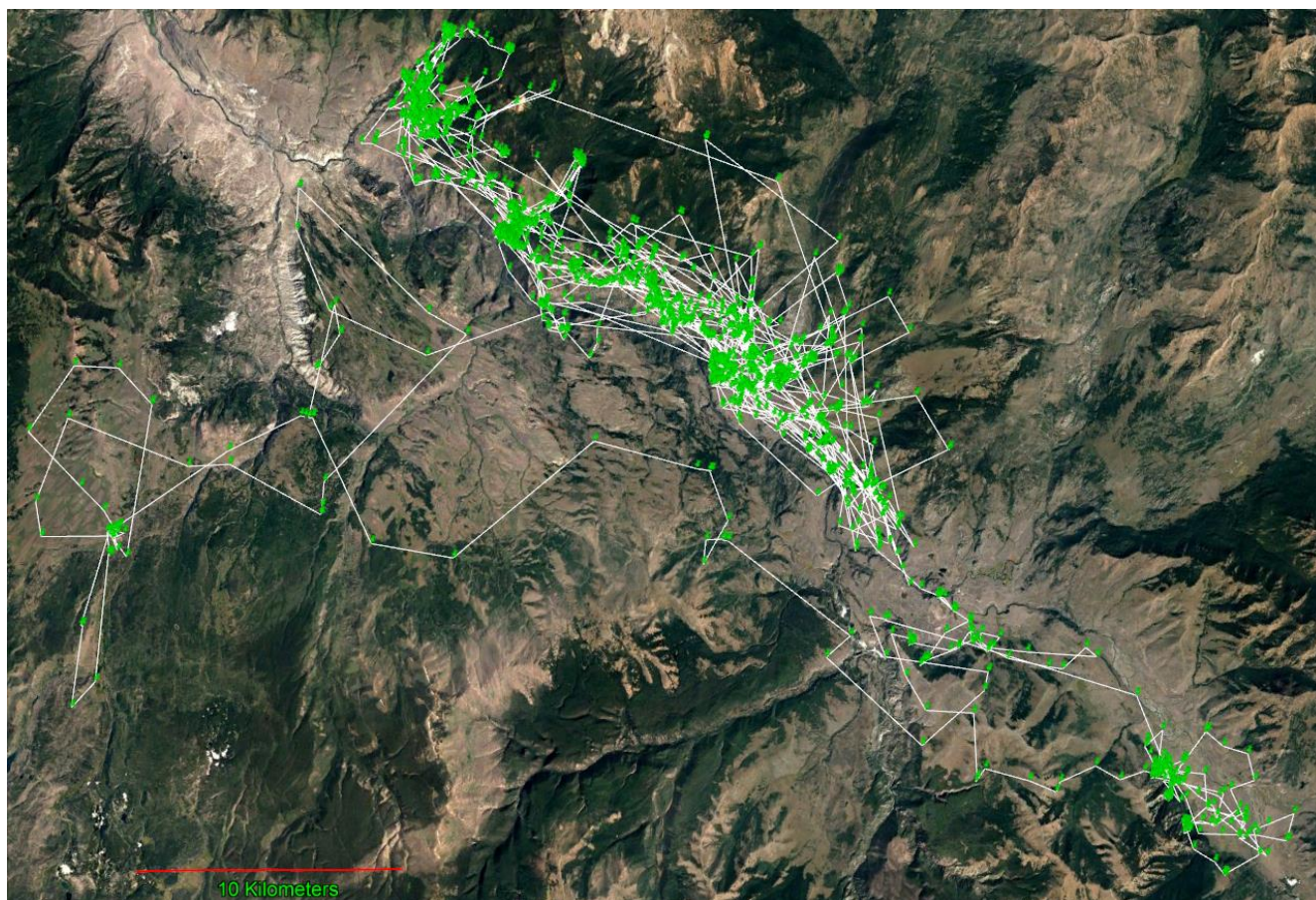
The points on the map above show the locations of a single wolf in Yellowstone National

Park from January to May in 2005. In this paper, we will be discussing various methods of

fractal analysis and applying these methods to datasets of animal positions like the one shown

above. Diverse early work on scaling analysis by Harold Hurst was unified by Benoit

Mandelbrot as a part of fractal geometry and analysis. Fractal analysis techniques have since

been used to characterize seemingly random processes and shapes in nature such as country

coastlines, river discharge, and even musical compositions.[6] Animal movement positional data

exhibits the same type of randomness and irregularity that are common in random fractals, so a

fractal analysis may give us a new way of characterizing the data. Before moving on to the actual

methods of analysis, it is important to understand the mathematics behind concepts related to

random motion and the generation of random shapes. More specifically, we will be reviewing the

difference between Brownian Motion (Bm) and Fractional Brownian Motion (fBm), how traces

of fBm differ from trails of fBm, and the difference between self-similarity and self-affinity.

2. Mathematics of Random Walks and Brownian Motion:

I. Brownian Motion (Bm)

   A random walk is a discrete path of random steps where each step is not dependent on

previous or future steps.[6] For example, consider flipping a fair coin n times and taking a step 1

unit up if the coin is heads and 1 unit down if the coin is tails. This creates a random walk in 1

dimension. The expected displacement over the n steps is 0 as there is an equal chance the

walker goes either up or down. Two simulations of this scenario with n=20 and n=2000 steps are

shown in Figure 2 below.

Figure 2: Simple Random Walks
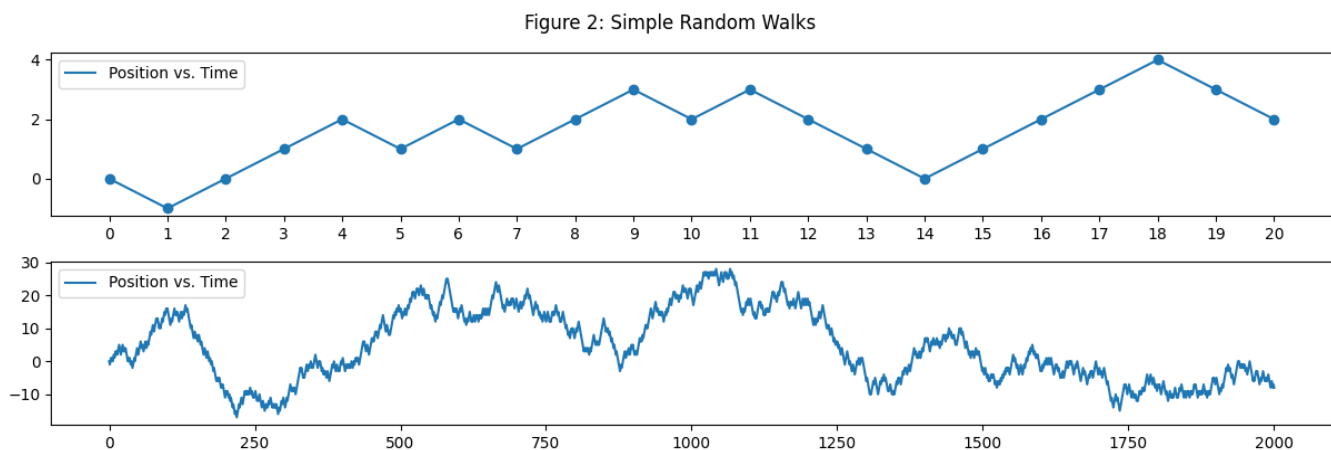
Unlike these examples, Brownian motion is a continuous stochastic process. Additionally, for

times t and s where s<t, Brownian motion has normally distributed independent increments with

mean 0 and variance t-s. Since a Brownian motion at time t is made up of infinitely many

independent increments, we also know that the result of a Brownian motion at time t is normally

distributed with mean 0 and variance t. The key characteristic of Brownian motion is that its

increments are independent, meaning that each increment is not impacted by past or future

increments. We use the Hurst parameter H =1/2 to describe this. The Hurst parameter tells us

how the standard deviation of the Brownian motion scales with time.[6] For example, in the case

of Brownian motion with H=1/2, after n=2000 steps, we would expect the Brownian motion to

be between $\sqrt{n}$ and $-\sqrt{n} \Leftrightarrow \sqrt{2000}$ and $-\sqrt{2000} \Leftrightarrow$ 45 and -45.

In the real world, animal positional data gathered by GPS trackers is typically not

continuous. We can approximate a Brownian motion by piecing together individual random

walks with many steps. [7] If we let each increment of the Brownian motion be one of these

individual random walks, by the Central Limit Theorem, the increments of the Brownian motion

will approach a normal distribution as the number of steps in each individual walk get larger and

larger. Since we are analyzing real world data in this paper, when the term "Brownian motion

(Bm)" or "fractional Brownian motion (fBm)" is used, we are referring to a discrete

approximation of a Brownian motion or fractional Brownian motion that is constructed of

random walks.

II. Trace vs Trail of Brownian Motion (Bm)

Each Brownian motion has a single trace that can be viewed as position in E dimensional

space with respect to time. A Brownian motion in E spatial dimensions has E trace projections

with respect to time.[6] Therefore, if we have a Brownian motion in two spatial dimensions, it

will have two trace projections (x and y). On the other hand, a trail of Brownian motion in two

spatial dimensions shows the points visited by the walker in the two spatial dimensions

independent of time.[6] This can of course be extended to E dimensions. While the trace gives us

a look at what is happening in each dimension with respect to time, the trail gives us insight into

the geometric qualities of the Brownian motion as a whole. The trail (plot (c)) and traces (plots

(a) and (b)) of a Brownian motion taking n=16384 steps in two dimensions are shown in Figure 3

below. Since we have two traces in time, this dataset is two dimensional, so the trail acts as a

"map" of the walker's position starting at the green point and ending at the red point.

Figure 3: Brownian Motion
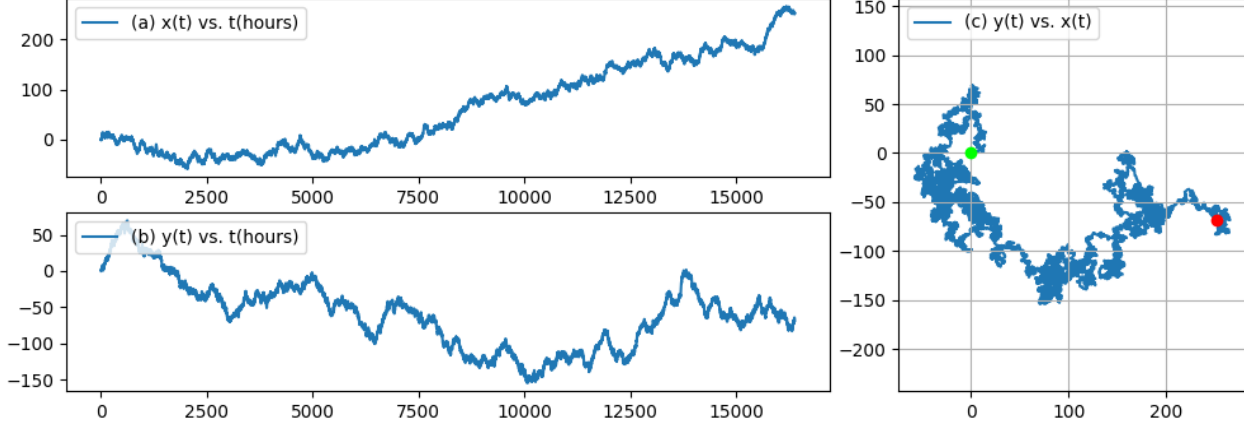
III. Fractional Brownian Motion (fBm)

Fractional Brownian motion (fBm) has normally distributed increments with variance

proportional to $|t-s|^{2H}$ for times t and s.[6] The main difference between Bm and fBm is that

fBm increments are not independent. Specifically, increments of a fBm are either positively or

negatively correlated.[6] When dealing with fBm, we use the Hurst parameter $0 < H < 1$ to

distinguish whether the increments of the fBm are positively or negatively correlated.[6] If H is

greater than 1/2, the fBm is positively correlated, and if H is less than 1/2, then the fBm is

negatively correlated. [6] If H is exactly 1/2, the motion has independent increments and is

classified as Brownian motion.[6] The fractal dimension of a trace of fBm is 2-H[1], and the fractal

dimension of the trail corresponding to that trace is $\frac{1}{H}$[2]. [6] In this paper, we will be

---

[1] Position vs time is not a traditional geometric object. For a more in-depth coverage, see "The relation of D to H for self-affine fractional Brownian motion" in [6].

[2] The dimension D=1/H is referred to as the latent fractal dimension of a trail of fBm. For a more in-depth coverage, see "Trails of fBm" in [6].

demonstrating a way to estimate the Hurst parameter given a trace of fBm and applying it to animal positional datasets.

IV. Self-Similarity vs Self-Affinity

Traces and trails of fBm have unique qualities regarding self-similarity and self-affinity. A shape exhibits self-similarity if it is composed of smaller geometrically similar "copies" of itself. This means that as we change the scale at which we examine the shape, it will always look identical. On the other hand, a shape is self-affine if it is composed of "copies" of itself that are scaled differently in each direction. More rigorously, a set of points S creating a shape in dimension E is self-affine if S is composed of N distinct subsets containing points with E coordinates, and each coordinate is scaled by a different r where $0 < r < 1$.[6] On the other hand, a set of points S creating a shape is self-similar if the set S is composed of N non-overlapping sets of points that are similar to S, but equal to size rS where $0 < r < 1$.[6] Note that each set in N contains points with E coordinates where E is the dimension of the Euclidean space, and each of those components is scaled by the same r (this is not the case with self-affinity). Trails of fBm are statistically self-similar, while traces of fBm are statistically self-affine.[6] Since we are characterizing natural phenomena, we are more concerned with statistical self-similarity and self-affinity, meaning not exactly self-similar or self-affine. The dimension D of a statistically self-similar fractal trails of fBm can be calculated by the equation $D = \frac{\log (N)}{\log (R)}$ where N is the number of congruent components making up the shape (detail) and R is the scale we are examining the shape at.[6]

VI. Fractal Analysis of Brownian Motion (Bm)

We will be analyzing this data both geometrically and temporally. When we analyze the data geometrically, we will be characterizing the walker's position on a two-dimensional plane

that contains points of the walker's position. When we analyze the data temporally, we will be examining the positional data with an added time dimension. We will perform two separate fractal analyses on two separate Brownian motions with step length n=16384. The first Brownian motion is constrained to start and end at the origin, and the second can move freely with no constraints. The constrained Bm particularly resembles animal positional data as animals often return to a starting point during their travels. The trails and traces for each Bm can be seen in Figures 4 and 5 below.

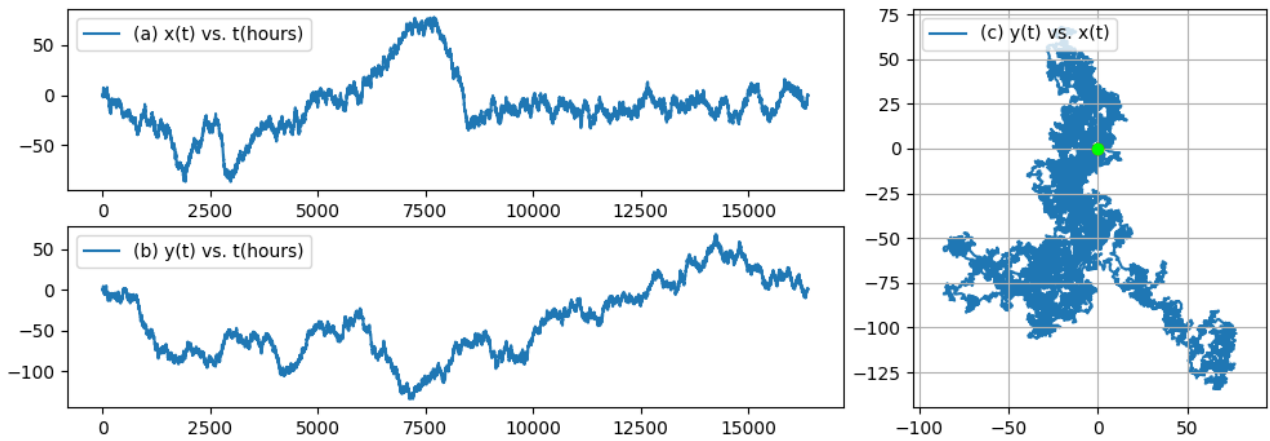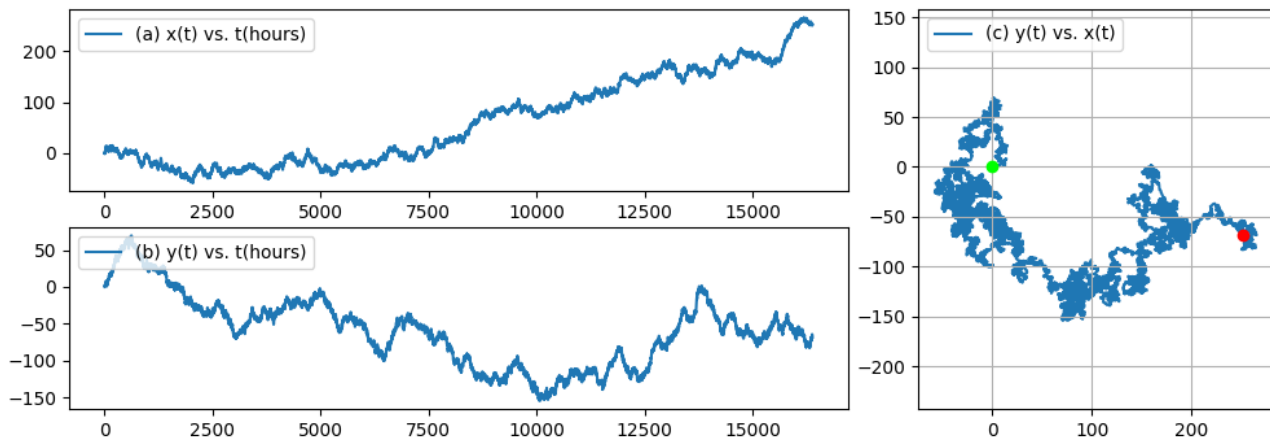
Figure 4: Brownian Motion With Constraint

Figure 5: Brownian Motion Without Constraint

Here, we can see that the trail of Bm with the constraint (Figure 4) looks like a closed shape as it ends exactly at the origin. If we examine the traces, we can also see that both the x trace and the y trace end at the origin. In the case of the completely random walk (Figure 5), the trail looks less like a closed shape as the walker could move freely without any constraints. Keep in mind that although unlikely, it is possible for a completely random walk to end at the origin.

The first method of fractal analysis we will be applying to the Brownian motions in Figure 4 and Figure 5 is called the "D Analysis" and will be used to characterize the datasets geometrically by assigning them each a fractal dimension D. Since we are examining the data geometrically, we will be performing this analysis on only the trails of Brownian motion. The goal of a fractal analysis is to study how a shape's detail changes as we examine it at different scales.[7] A typical method of doing this with a two-dimensional dataset is box counting.[7] This method involves examining how the average number of points inside boxes of size E (detail) changes as a function of E (scale).[6] We can eliminate the need to physically count points inside boxes by examining how N(R), the average number of points within a distance R about a particular point averaged over all points, changes as a function of R. If the trail exhibits any self-similarity, there will be a power law relationship between R and N(R). The slope of this power law relationship on a log-log plot will be equal to the dimension D of the dataset. In other words, $N = R^D \Leftrightarrow \log(N) = D \log(R)$. Figure 6 below shows the dimension of each Brownian motion and the least squares regression to estimate each D.

Figure 6
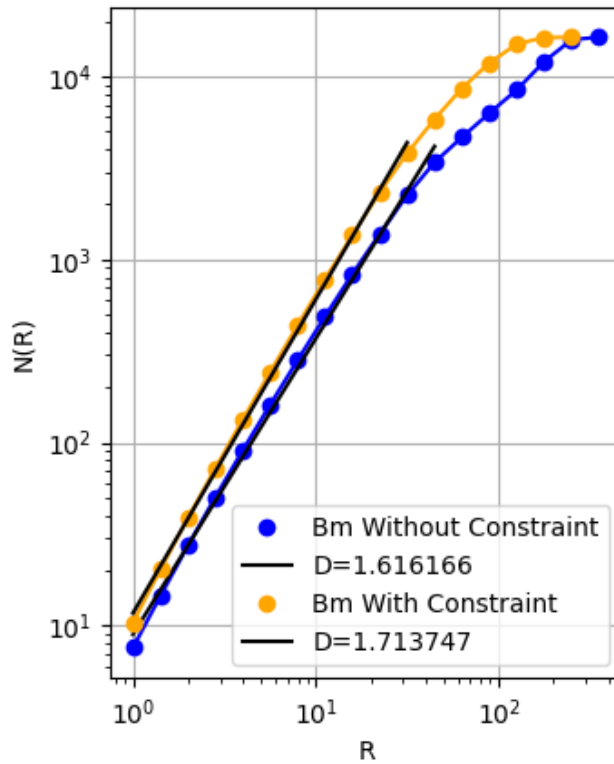
Here, we can see that N(R) values eventually level off in both cases as R increases. This is intuitive because once R gets to a size where all the points have a distance less than R between them, changing R to a higher value does not change the N(R) values.

The next type of fractal analysis we will be applying to the Brownian motions is called the "H Analysis". The goal of the H Analysis is to characterize the trace of a Brownian motion by assigning the Bm trace a Hurst parameter. As mentioned earlier, the Hurst parameter is a useful way of characterizing the correlation of the increments of a Bm.[7] To perform an H Analysis, we check the average displacement of the walker between points separated by a specific time. Since most data is not uniformly spaced, we can organize the data into bins based on time differences and then average the distances in each bin. Looking for a power law relationship between the average time distance in each bin and the average displacement in each bin on a log-log plot will allow us to estimate H for the dataset. The H Analysis for the two Brownian motions is shown in Figure 7 below.

Figure 7
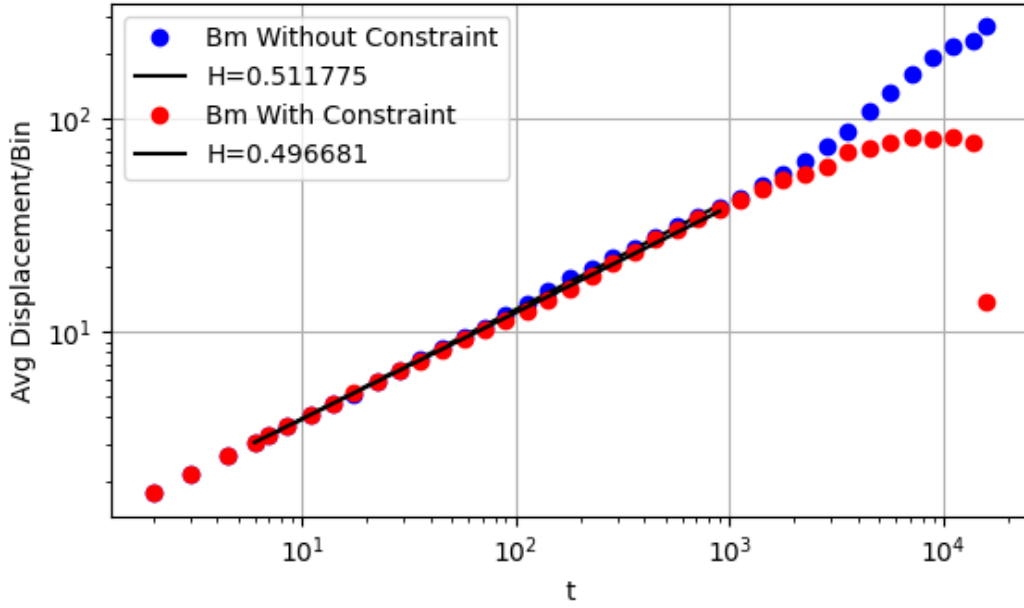
Since both datasets are Brownian motion, we can see that they both have a Hurst estimate of around ½, meaning that their increments are completely independent. Additionally, in the Brownian motion that ends at the origin, we can see that as the difference in time between points increases, the distance between those points decreases after a time of 10000 hours. This is because this Bm is forced to end at the origin, so as we get closer and closer to the end of the Bm, the average displacement per bin must decrease to 0.

VII. Fractal Analysis of Fractional Brownian Motion (fBm)

Here, we will be using the same two fractal analyses we performed in the previous section, except we will be examining two simulated datasets of fBm. The two datasets can be visualized in Figure 8 and Figure 9 below.

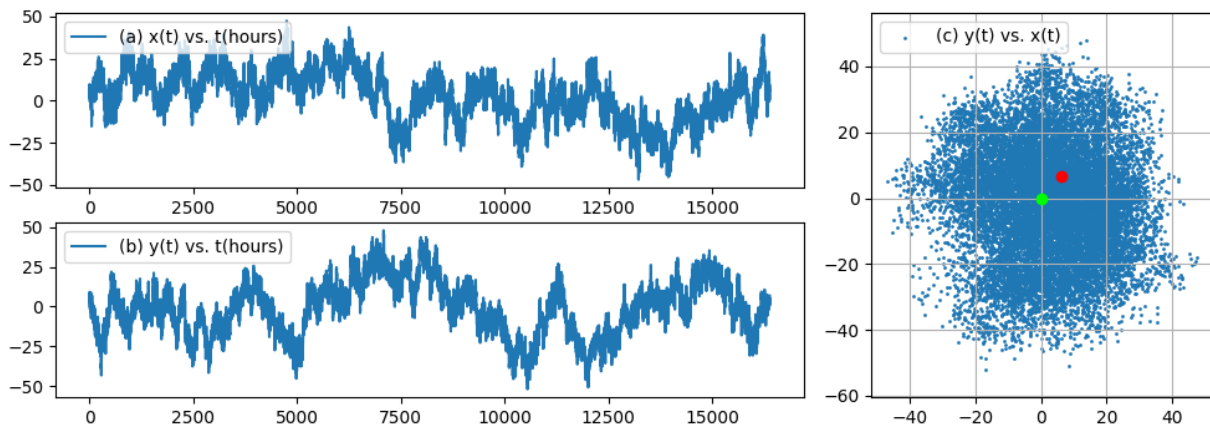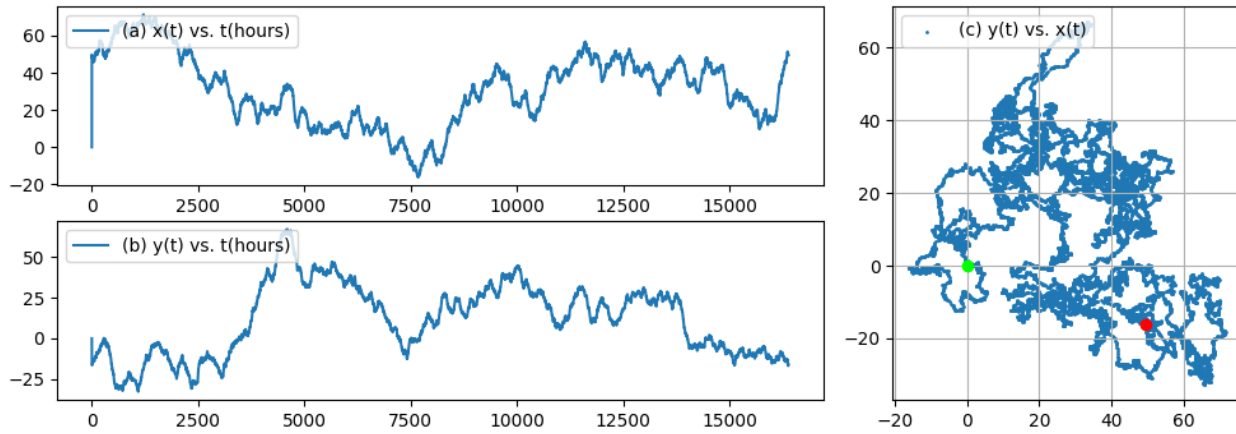Figure 8: Fractional Brownian Motion #1

Figure 9: Fractional Brownian Motion #2

Visually, both simulations of fBm look very different. The trail of fBm in Figure 8 looks significantly denser and more tortuous than the trail of fBm in Figure 9. Additionally, the x and y projections of the trace of fBm in Figure 8 look significantly more jagged than the x and y projections of the trace of fBm in Figure 9. The fractal analysis techniques discussed earlier will give us a way to characterize these datasets mathematically.

First, we will examine this dataset geometrically by applying the D Analysis we discussed when analyzing the Brownian motion. The result is shown below in Figure 10.

Figure 10
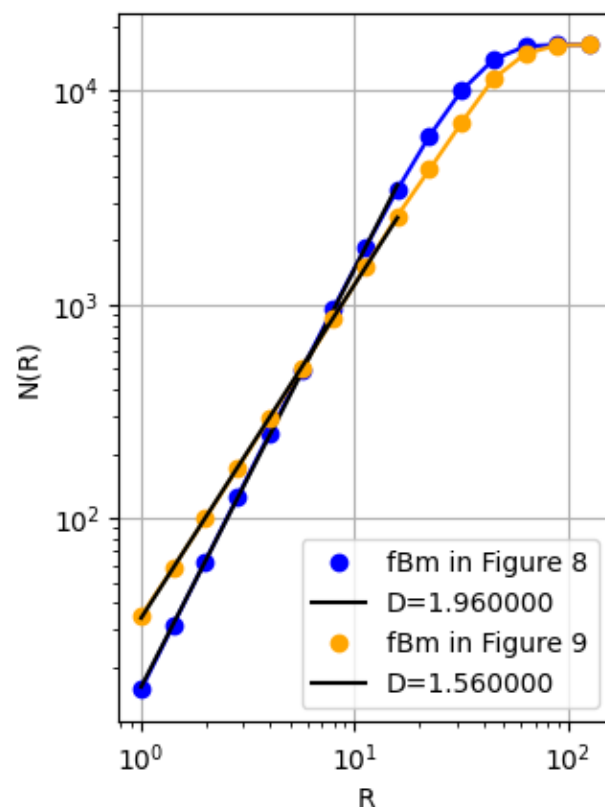
From the fractal analysis, we can see that the dimension of the fBm in Figure 8 is significantly

higher than the fractal dimension of the fBm in Figure 9. This parallels what we noticed visually

as the fractal dimension in a sense is used to depict complexity, and the trail in Figure 8 looks

more complex visually than the trail in Figure 9. [7]

Next, we will apply the H Analysis to characterize each trace of fBm with a Hurst

exponent H. The result of the H Analysis is shown below in Figure 11.


Figure 11
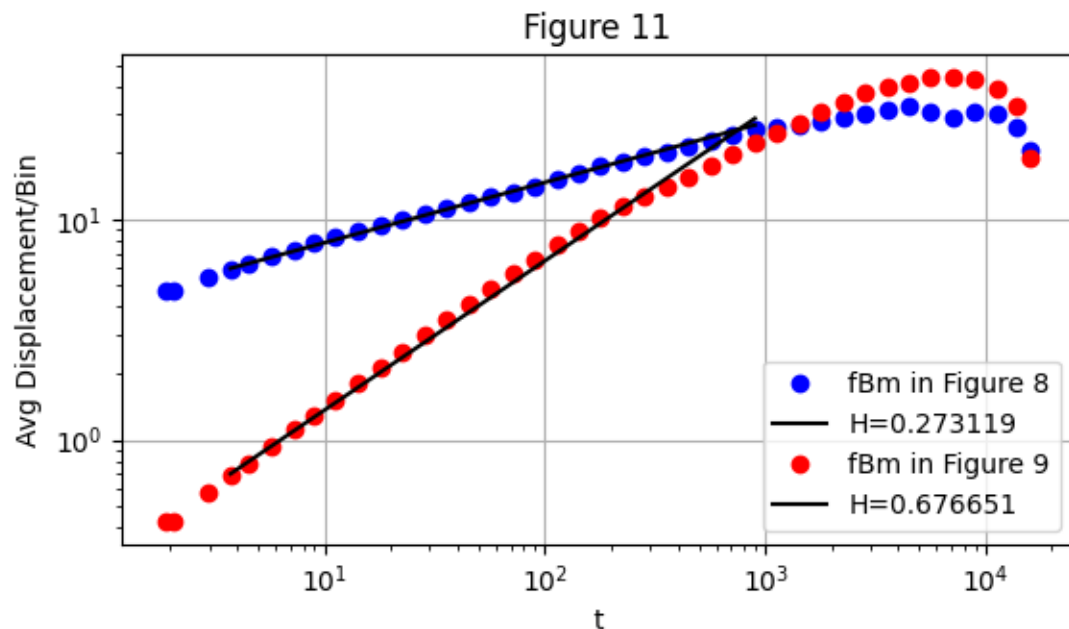
Since we are looking for a power law relationship, the line of best fit for each fBm was estimated

by picking the longest and straightest portion of each plot. Upon inspection of the Hurst analysis,

it is evident that we have been analyzing a positively correlated case of fBm and a negatively

correlated case of fBm. Figure 9 has H=.68, meaning it has positively correlated increments, and

Figure 8 has H=.27, meaning it has negatively correlated increments. When the increments are

positively correlated, the trace is less likely to change direction, meaning it will be less jagged.

When the increments are negatively correlated, the trace is more likely to go in the opposite

direction than it was going previously which contributes to a more jagged and twisted look. A

Hurst parameter of .27 is much less than ½, so it is understandable that the x and y projections of

the trace are extremely jagged and twisted. Now that we have a Hurst parameter associated with

each dataset, we can use this to get another estimation of the dimension of the trail of each fBm.

Using the definition that the dimension of a trail with Hurst H is 1/H, we can estimate the

dimension of Figure 9 to be 1/.68=1.47 and the dimension of Figure 8 to be 1/.27=3.7.[6] Since

the fractal dimension cannot exceed the number of spatial dimensions (2 in this case), we say

that the dimension of Figure 8 is 2.[6] Both of these estimates are relatively close to the

dimensions we calculated in the previous analysis. Since we are only estimating the fractal

dimension, we can expect some variability depending on the method we use. In the next section,

we will apply these types of fractal analysis with the goal of characterizing the movement data of

wolf packs in Yellowstone National Park.

3. Fractal Analysis of Wolf Locational Data in Yellowstone National Park:
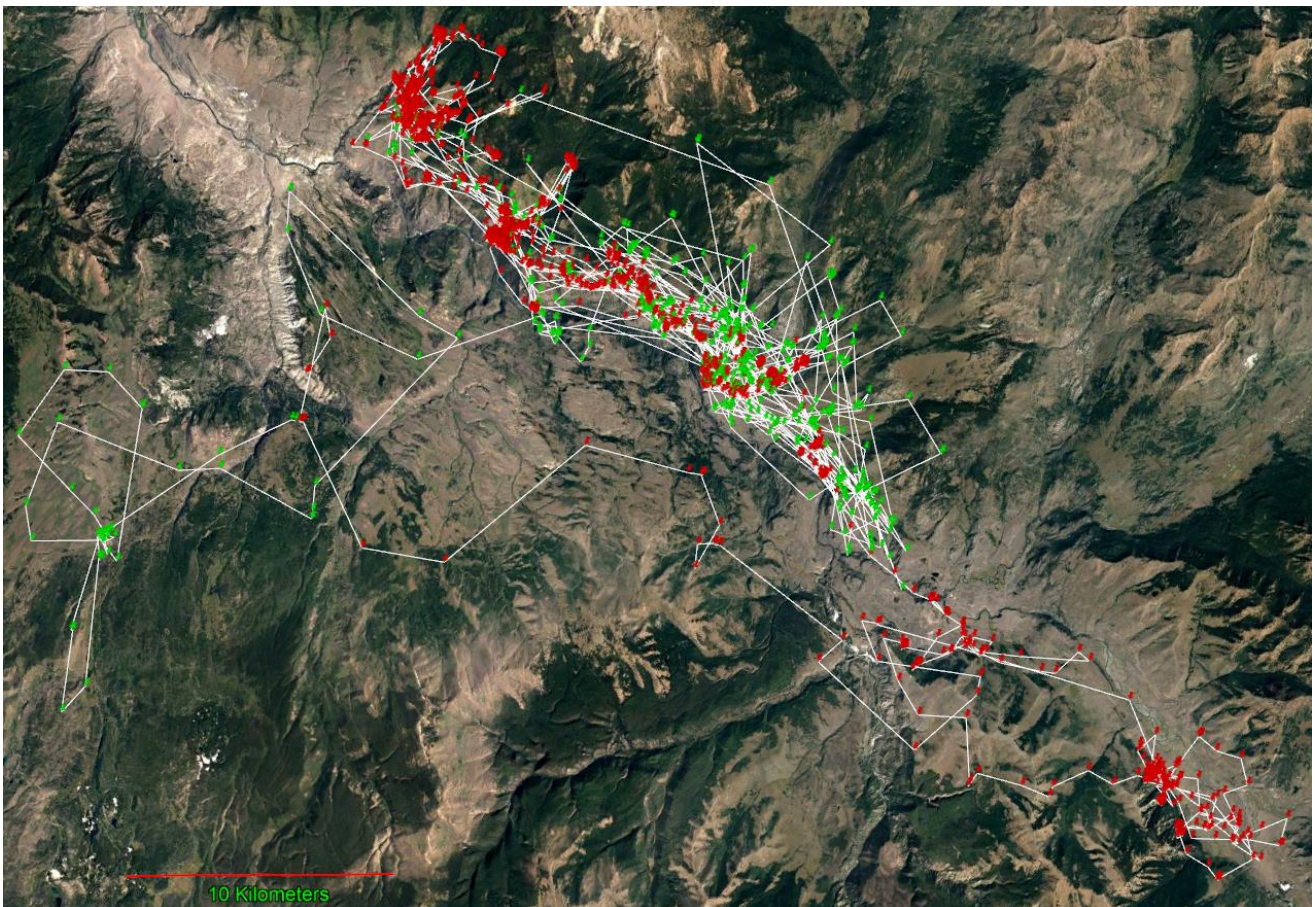
Pictured above are the positions of wolf 0488M from January 2005 to May 2005[3]. The green points represent the position of 0488M when it was part of the Geode Creek wolf pack, and the red points represent the position of 0488M when it left the pack and started traveling alone. The green points were recorded every 3 hours, and the red points were recorded every 30 minutes. Geode Creek flows from a small lake into the Yellowstone River. We will be treating each color pin as its own individual wolf for this fractal analysis. This will allow us to draw conclusions about how 0488M moves while it is traveling with the pack vs traveling alone. Figure 12 and Figure 13 contain the trace projections of wolf 0488M while it is with and without the pack.

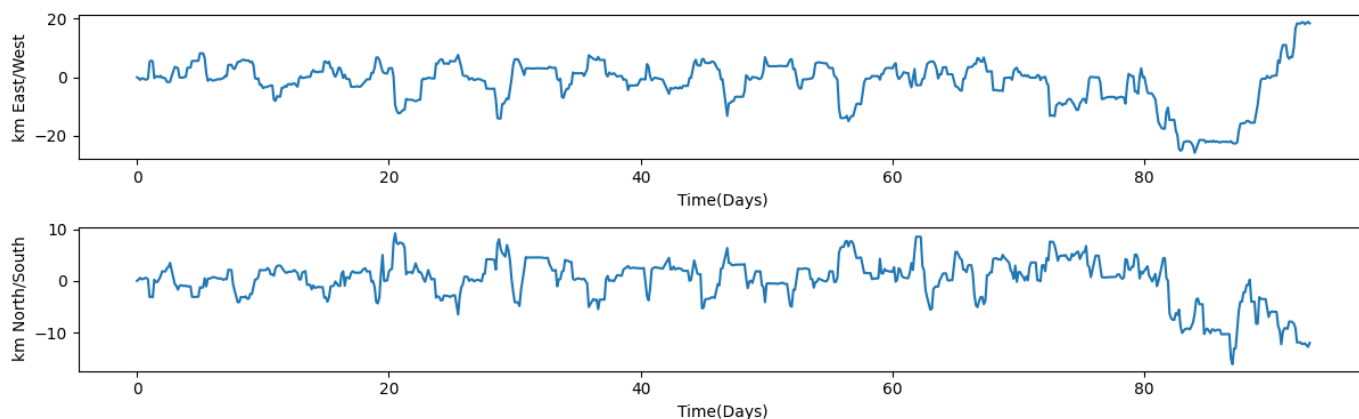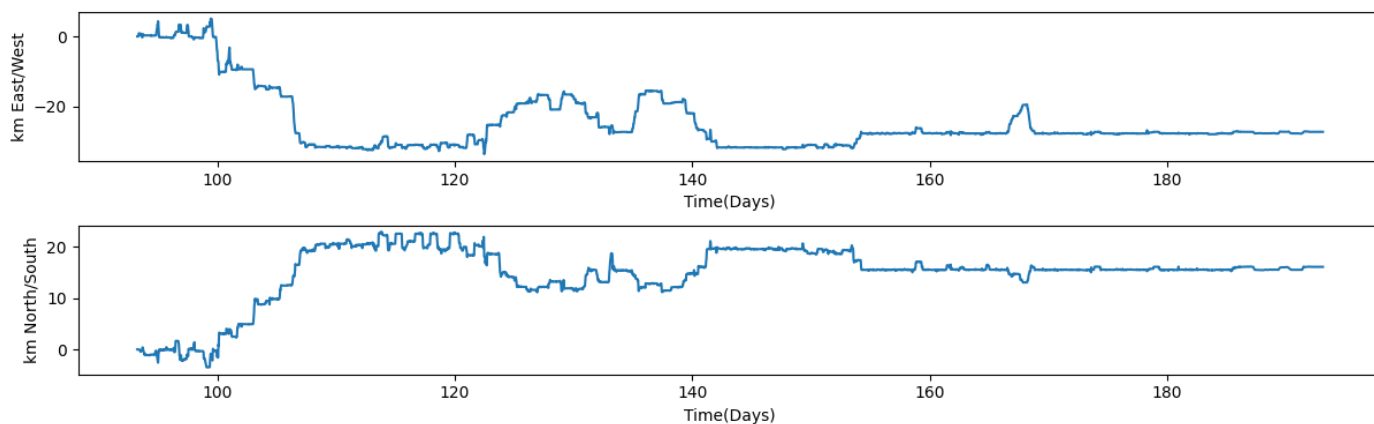Figure 12: 0488M with Geode Creek Pack (Green Points)

Figure 13: 0488M without Geode Creek Pack (Red Points)

[3] Thank you to the Yellowstone Wolf Project (Matthew C. Metz) for supplying the data on wolf positions in Yellowstone National Park.

Looking at the trace projections, we can see that 0488M's movement was more tortuous when he was with the pack. Looking at the trail, we can see that the Geode Creek pack seems to stay relatively close to Geode Creek. Since the pack is roaming over a small area, it is clear that 0488M is making many turns. Additionally, we can see that 0488M left the pack around 90 days after we began tracking him (04/17/2005). Initially, 0488M was much more active than he was when he was with the pack. This activity then lessened significantly around 140 days after we began tracking him (06/16/2005). The trace projections make it very easy to identify time periods of interest such as this one. Next, we will characterize each set of data by performing an H Analysis. In addition to providing us with a Hurst parameter to characterize each set of data, we can also use this analysis to draw conclusions about where we expect 0488M to be after various lengths of time. The H Analysis for when 0488M was with the pack and without the pack are shown in Figure 14 below.


Figure 14
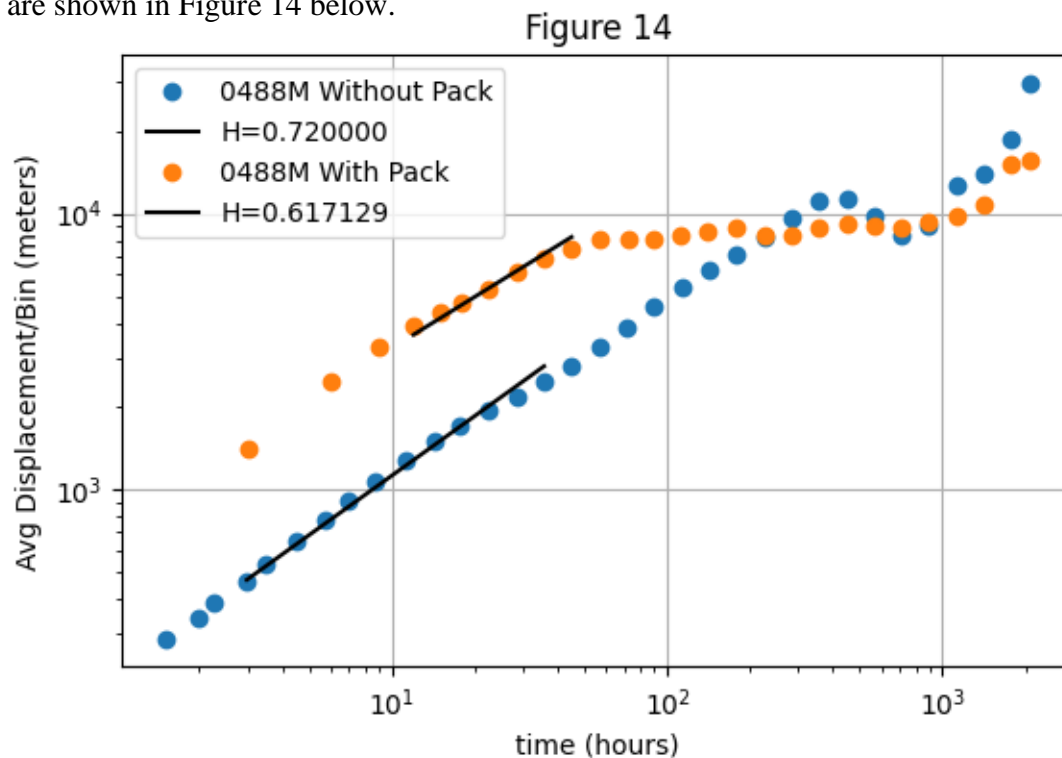
The positions of the wolf resemble fractional Brownian motion because the Hurst parameter associated with each dataset is greater than ½. This implies that 0488M's steps are

positively correlated, meaning that if he is moving in one direction, he is more likely to continue moving in that direction in future steps.[7] Additionally, we can see that 0488M covered more territory while alone than with the pack. This is consistent with the fact that the Geode Creek pack seems to stay relatively close to Geode Creek, and 0488M may need to travel to areas other than Geode Creek to find a mate or a new pack. This analysis also gives us a way to estimate where we expect 0488M to be after different lengths of time. For example, we can expect 0488M to be 10km away from his starting point after about 1000 hours regardless of if he is with the pack. However, when 0488M is not with the pack, we can expect him to be further away from his starting point than he would be with the pack after lengths of time greater than 1000 hours.

Next, we will look at a group of 9 other wolves, some of which are from the same pack and some of which are from different packs. We will characterize each of the wolves using both the D Analysis and the H analysis, then we will compare the behavior of these wolves to the independent wolf 0488M. Pictured below are the trails of each of the 10 wolves.
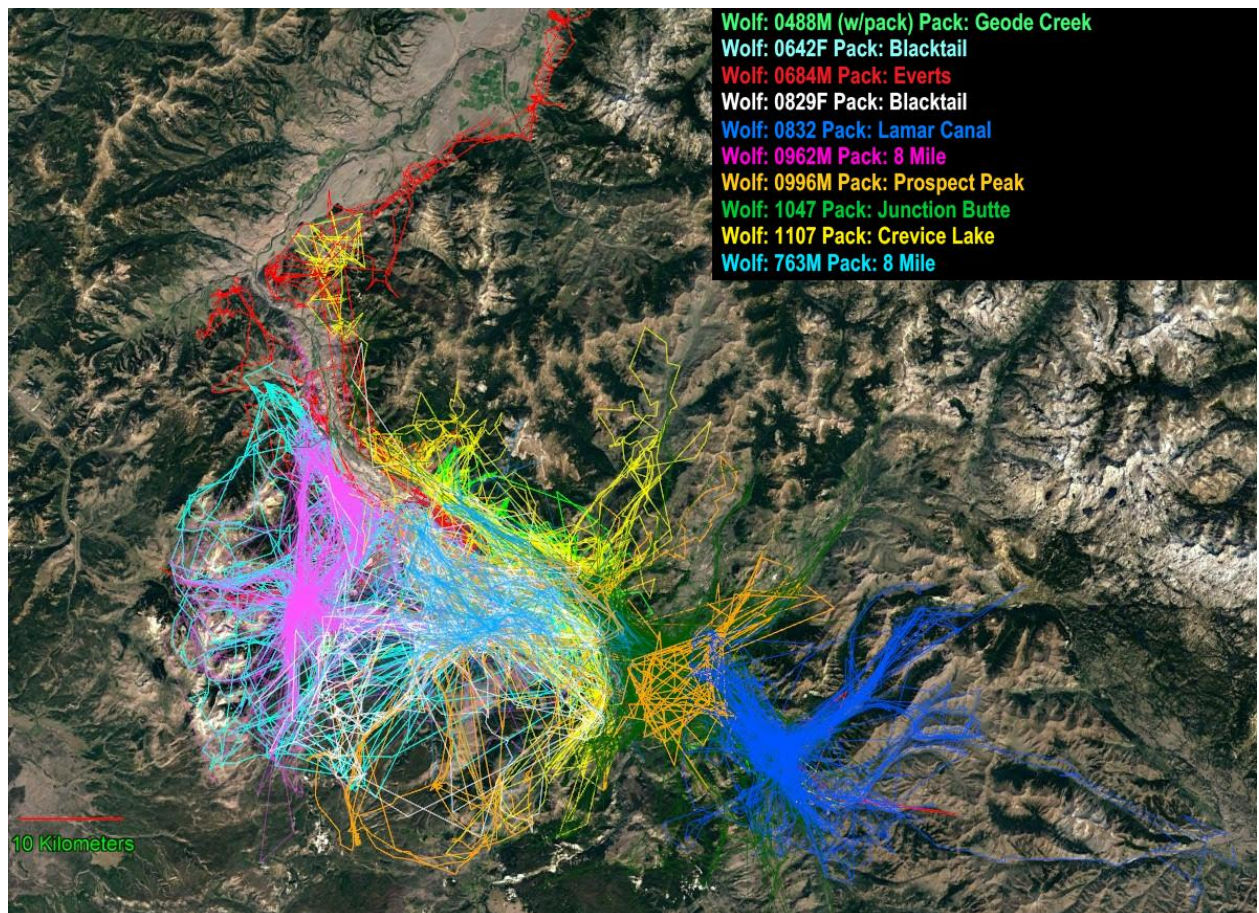
Although the wolves are in different packs, they occupy the same general area. This implies that these packs could be competing for food and territory, and that movement from pack to pack could be correlated.

The H Analysis for each of the 10 wolves including 0488M is shown in Figure 15 below. Each y value was scaled to space the plots away from each other. Doing this preserves the slope associated with each wolf plot but makes the meters on the y axis inaccurate. The H parameter associated with each wolf is the slope of the line of best fit. For example, 0488M has H parameter .72.


Figure 15
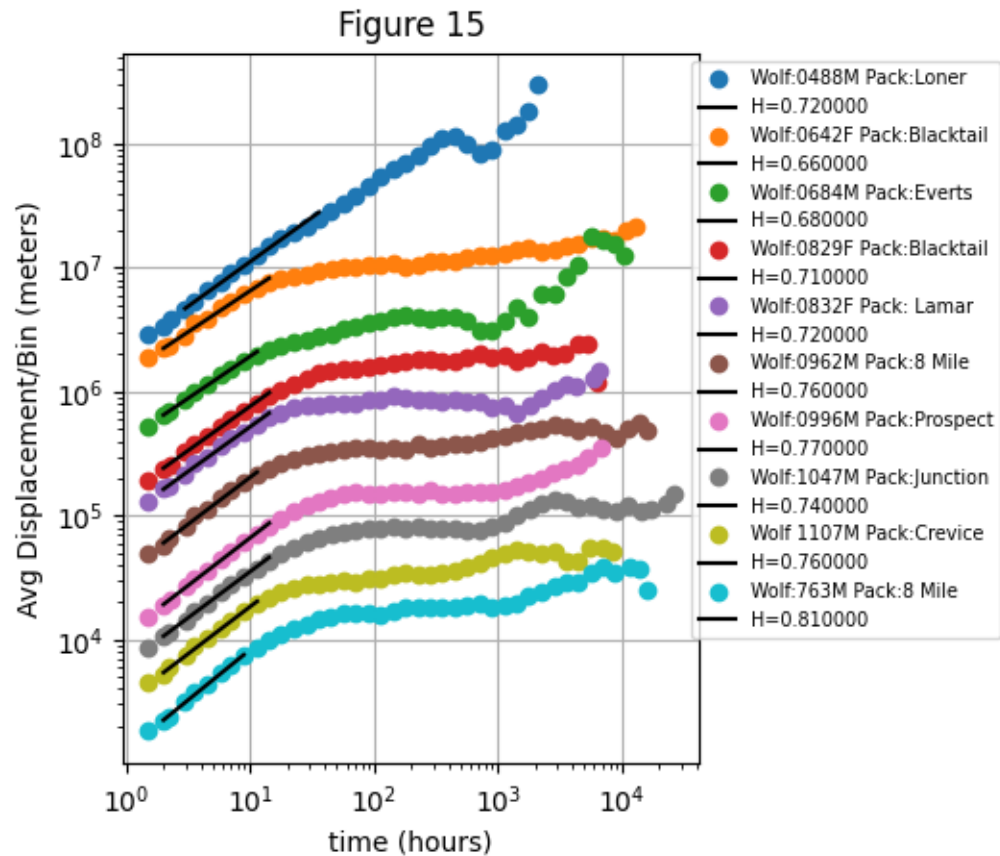
We can see that all the wolves other than 0488M are part of a specific pack, with a few of the wolves sharing the same packs. Each of the 9 wolves have very similar H Analysis results.

We can see that the average displacement of the wolf exhibits a power law relationship with respect to small lengths of time, but as we start looking at larger time differences, we no longer see the same relationship. For example, from time differences of 100 hours to 1000 hours, we see that although time is passing, the wolves in packs are not traveling longer distances. Then from time differences of 1000 hours and above, the wolves seem to start traveling longer distances again. This change could correspond to certain changes in the wolves' environment such as food supply or changes in the weather. We can see that this pattern does not continue with 0488M as the displacement between points more consistently increases as the lengths of time between points increases.

Additionally, none of the wolves have a Hurst parameter equal to 1/2, meaning that their movement is not completely random. Although it is understandable that an animal would not move completely randomly, it is important to note that these datasets do not display an exact power law relationship across all time differences. Since this is the case, the H value we estimated for a subset of the total time differences is not representative of the total data from each wolf. Thus, if we were investigating this further, it would be beneficial to break each wolf's data into pieces. We can then get a more detailed view of H by performing the H Analysis individually on each piece.

The last analysis we will use to characterize the wolves is the D Analysis. Figure 16 below shows the R vs. N(R) plots for each of the wolves and the line of best fit gives us a D estimation for each. Again, each y value was scaled to space the plots away from each other. Doing this preserves the slope associated with each wolf plot but makes the N(R) values on the y axis inaccurate.

Figure 16
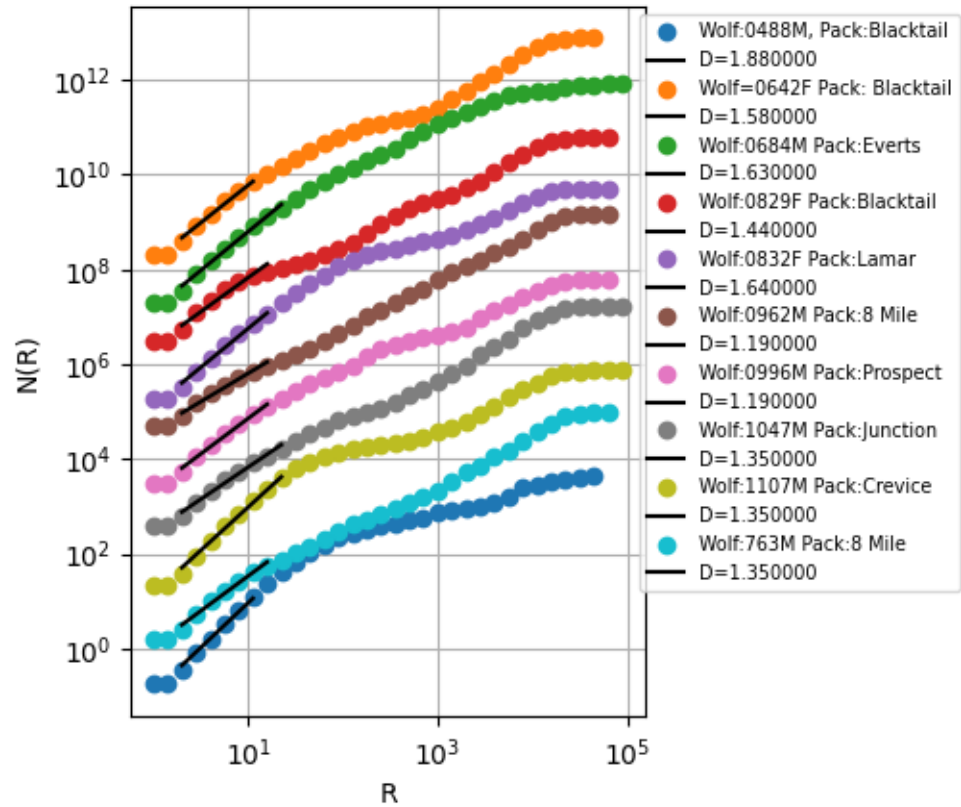
Here, the portion of each plot that showed the closest to a power law behavior was selected to estimate D. In this case, the most linear areas of each plot happened to be when R values were between 10 and 1000. Wolves such as 1107M and 0829F do not have linear plots. Thus, it is difficult to characterize these wolves with a dimension D because there is not a true power law relationship between R and N(R). Some of the other wolves such as 763M and 0962M do show a power law relationship between R and N(R). These two wolves also have nearly identical R vs. N(R) plots. This is expected because they are from the same pack, and packs usually stay together as they move. Comparing this data with the simulated fBm, we can see that there is a significant difference in the linearity of the results. Although we cannot characterize this data entirely accurately with this specific form of analysis, we can still use it to look for geometric similarities from wolf to wolf. For example, since 763M and 0962M have very similar R vs. N(R) plots, their shapes could be similar in look and complexity. Referring to

the map that contains the trails of all the wolves, we can see that the blue trail (763M) and the pink trail (0962M) look nearly identical.

4. Implications / Next Steps

Fractal Analysis techniques provide a unique way to characterize animal positional data that exhibits similar randomness and irregularity to Bm and fBm. In addition to assigning these datasets values of D and H that give us information about the qualities of the shape and the correlations of each step of the animal, these techniques allow us to easily visualize how the animal moves during various lengths of time. In addition to the techniques discussed here, there are various other fractal analysis techniques that can also be used and experimented with to potentially characterize animal positional and time series data further.

With more detailed data, fractal analysis techniques could be used to look for relationships between changes in an animal's environment and the way that it moves. For example, if we had data spanning an entire year, we could examine how seasonal changes impact the animal's trail shape and traces in time. In addition, if we examined animals that typically cross multiple biomes over a period of time, we could investigate how changes in physical surroundings impact the animal's movement characteristics. Mandelbrot's fractal geometry gives us a unique way of characterizing shapes that defy the traits of Euclidean geometry.[7] Assigning D and H values to specific animals to characterize the shape of their movement is a unique trait of a fractal analysis, and with detailed enough data, the applications of this type of analysis will continue to grow. It is now up to ecologists to identify specific areas where these techniques may be particularly applicable and could provide unique and useful results.

References:

[1] Seidel DP, Dougherty E, Carlson C, Getz WM. Ecological metrics and methods for GPS movement data. *Int J Geogr Inf Sci*. 2018;32(11):2272-2293.

[2] Edelhoff, H., Signer, J., & Balkenhol, N. (2016). Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. *Movement ecology*.

[3] Getz WM, Fortmann-Roe S, Cross PC, Lyons AJ, Ryan SJ, Wilmers CC. LoCoH: nonparametric kernel methods for constructing home ranges and utilization distributions.

[4] DICKE, M. and BURROUGH, P.A. (1988), Using fractal dimensions for characterizing tortuosity of animal trails. Physiological Entomology, 13: 393-398.

[5] *Journal of Mammalogy*, Volume 90, Issue 5, 15 October 2009, Pages 1210–1217

[6] Voss, Richard. (2007). Characterization and Measurement of Random Fractals. Physica Scripta. 1986. 27. 10.1088/0031-8949/1986/T13/004.

[7] Mandelbrot, B. B. (1983). River Discharges, Scaling Nets and Noises. In *The Fractal Geometry of Nature*. New York, New York: Freeman.