Shorebird Data Analysis: thought and data processing

Emily Hamel

April 26, 2021

Background

These data sets were taken from a study conducted in Northern Europe, specifically the Dutch Wadden Sea island of Vlieland. The purpose of this study was to understand the impact of both short and long interval disturbance to the energy budget of roosting Eurasian Oystercatchers (*Haematopus ostralegus*). The data analysis in this document seeks to recreate this study's results, in order to develop a more comprehensive understanding of analyzing ecological, bird-related data sets - a skill that will be necessary for my future professional research.

The following code and graphical information outlines this attempt to recreate the study's analysis and the thought processes involved.

Short-term Disturbance: initial data analysis

The first piece of data to analyze in this study is the short-term impact on bird movement from different types of disturbance. In order to visualize the data I first created a plot (Figure 1) showing the immediate displacement resulting from different disturbance categories (namely: air force activity, non-anthropogenic events, and people walking past the roosting Oystercatchers). The "pre-disturbance" element of this plot seeks to show how disturbance related movements compare to the birds' activity prior to their displacement. The comparable plot in the literature is a box-plot, but I felt that the points provided a clearer picture of movement frequency and the related distance distribution - this choice, however, is simply a personal preference.

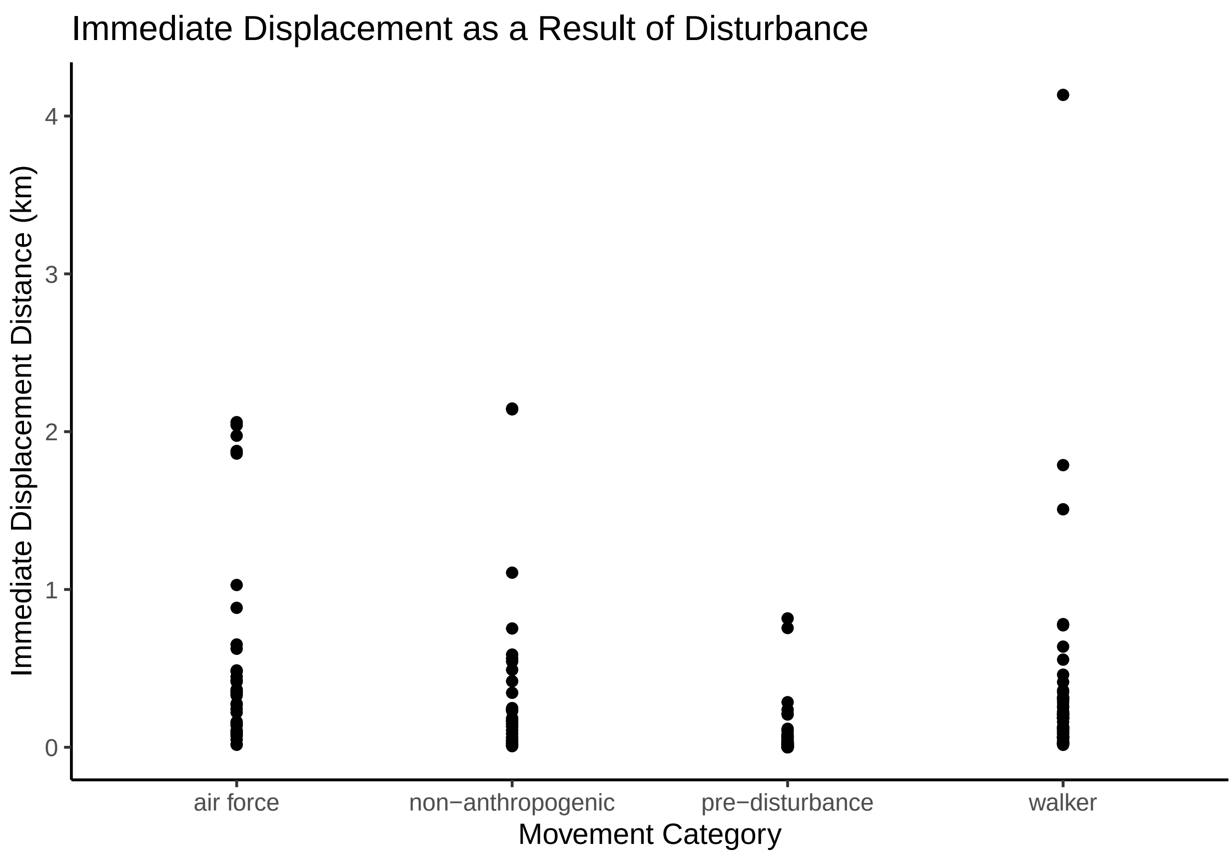

Figure 1: Depiction of immediate displacement (in kilometers) resulting from different disturbance categories. The points are indicative of each displacement event of the 240 observations.

From this data visualization it became important to run two statistical tests to understand if disturbance results in statistically significant movement. First, I ran a Kruskal-Wallis rank sum test on the immediate displacement data set, comparing distance and movement category (as depicted in Figure 1). This test was chosen due to its reference in the research paper and based on its ability to analyze significance across multiple groups. This test was preformed simply by writing:

```
kruskal.test(distance ~ movement_category, data = disturbance)
```

The output of this test displays an overall p-value of $7.3555356 \times 10^{-29}$. Such a low p-value indicates that the relationship between disturbance events and Oystercatcher movement is very significant.

From here I was hoping to understand the significance of the data when compared against itself. This meaning that, I was interested to know if the birds' movements were significantly different when disturbed compared to the pre-disturbance data. The paper I was working from ran a Dunn's test to make this determination, but I chose a pairwise Wilcoxon test - primarily because I was unable to load the package to run the Dunn's test and I felt that a more accessible approach would result in better reproducibility in the long run. In a similar fashion to the Kruskal-Wallis test, this Wilcoxon test was run as follows:

```
pairwise.wilcox.test(disturbance$distance, g = disturbance$movement_category)
```

From this test we observe adjusted p-values using the holm method. After an extensive amount of internet searching, I was still unable to determine how to set the relevant p-values as inline code. Rather than agonizing over it I have pasted the tabular output into this file. The following is the output and pairwise comparisons using this test

```
data:  disturbance$distance and disturbance$movement_category

                  air force non-anthropogenic pre-disturbance
non-anthropogenic 0.117     -                 -
pre-disturbance   1.6e-15   2.2e-11           -
walker            0.043     0.680             < 2e-16
```

What we are interested in looking at is the p-values as indications of significance between pre-disturbance movements and the three categories of post-disturbance movements (again, those are air force activity, walkers, and non-anthropogenic events). We see here that all levels of disturbance are relevant when compared to the birds' state prior to disturbance, with air force activity and non-anthropogenic events at the site causing the most significant amount of movement.

High Tide Interval Disturbance: the next step of analysis

Movement and Water Level

The original study looks to find a relationship between the maximum water level and the high tide travel distance, and how this differs among three different roosting sites. In order to accomplish they utilized a generalized linear mixed model with a gamma response distribution and a log link function. There are no instructions within the study on how to accomplish this, and it is outside my realm of understanding, so I chose to conduct a linear analysis with best-fit lines relevant to the three roosting sites in the study area (referred to in Figure 2 as Mid, North, and South).

This linear model was run using:

```
lm(distance_total ~ waterlevel, data = hightide)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

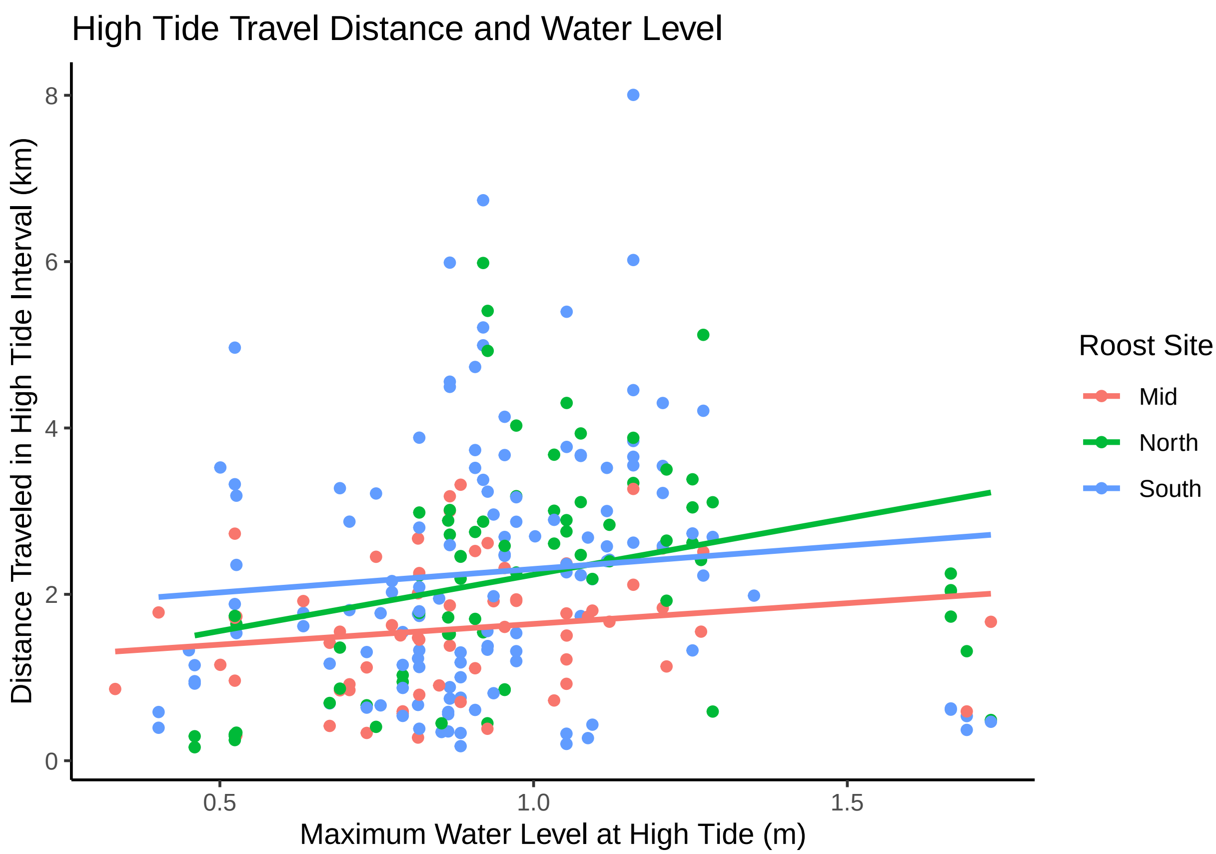

Figure 2: Relationship between the distance birds traveled during intervals of high tide (in kilometers) and the maximum water level (in meters)during these events. Data is divided between three relevant roosting sites: Mid, North, and South (shown in red, green, and blue - respectively).

High Tide Travel and Disturbance Displacement

The relationship between the birds' movement and the related water level is followed by an analysis of the total travel distance within a high tide cycle and the travel distance as a result of the disturbance discussed in Figure 1. The original paper used a similar analysis as discussed in Figure 2, but, again, I chose to simplify the analysis for my benefit and created Figure 3 using a best fit line and a linear model.

Similar to Figure 2, I ran the Figure 3 linear model as:

```
lm(distance_total ~ distance_disturbed, data = hightide)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

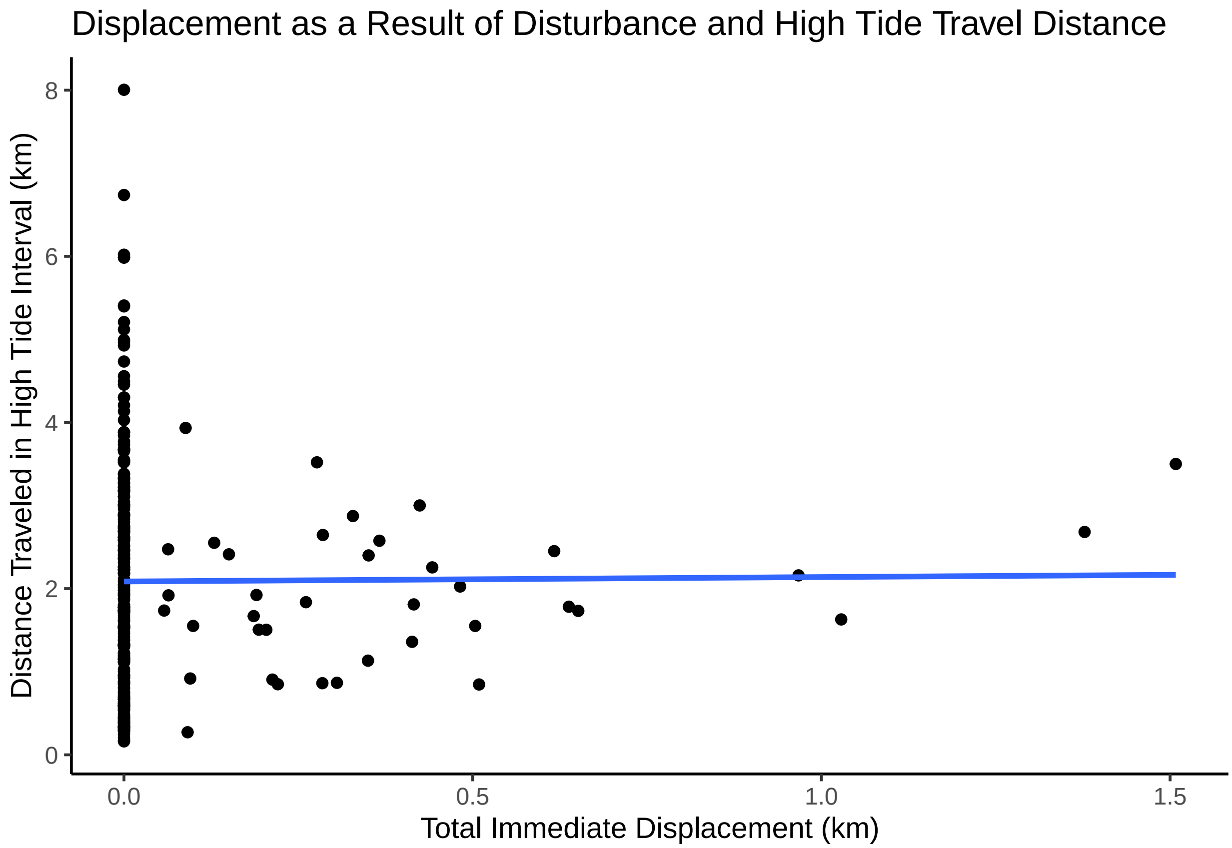

Figure 3: Relationship between the total distance traveled within the high tide period in kilometers (i.e. high tide travel distance) and the total distance traveled in responce to immediate disturbance within the high tide period in kilometers (calculated as the sum of the collected movements from GPS transmitters from disturbance, i.e. total immediate displacement).

Additional Travel Distance and Disturbance Displacement

The research takes this analysis one step further and creates an additional figure to understand if there was a relationship between the additional high tide travel distance (which was defined as the high tide travel distance minus total immediate displacement) and the total immediate displacement. A figure similar to Figure 3 was created to follow this continued analysis. And, again, as was the case for the earlier figures in my analysis, a linear model was used.

```
lm(distance_additional ~ distance_disturbed, data = hightide)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

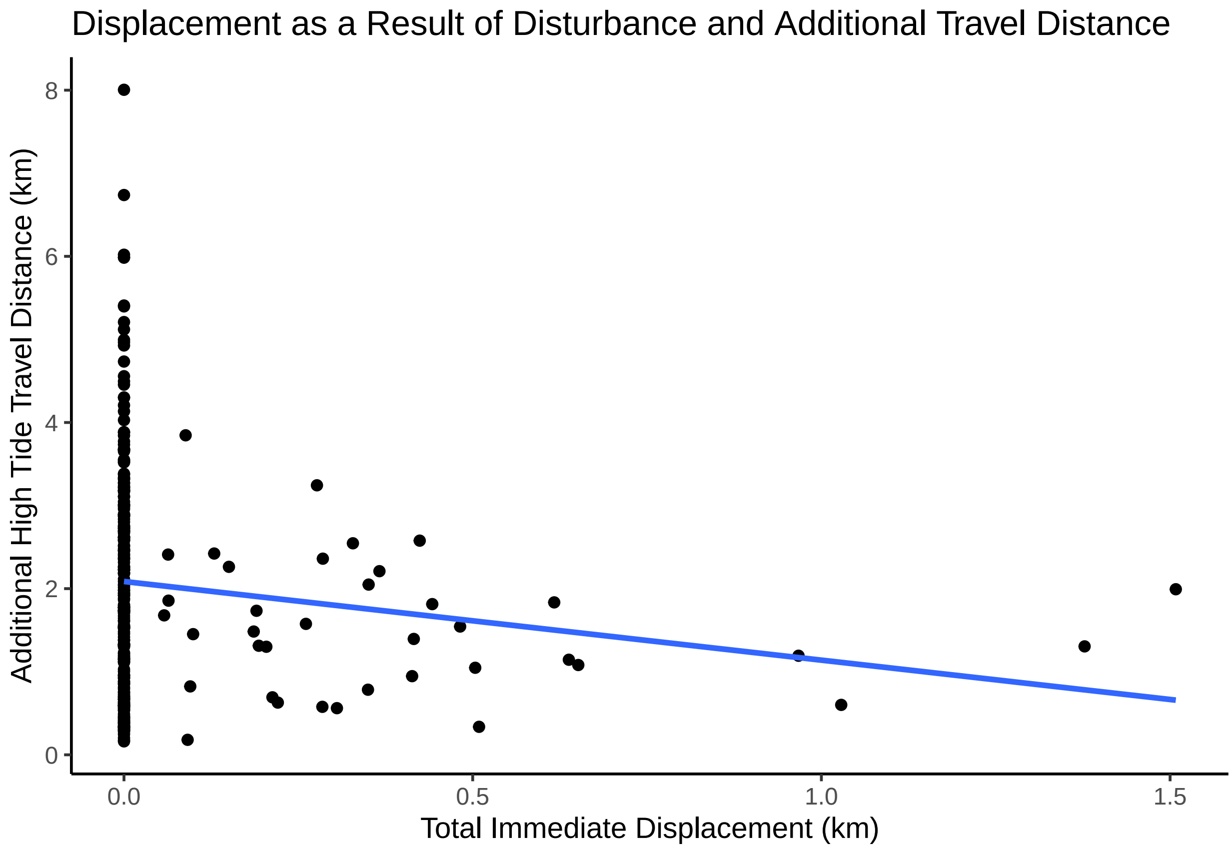

Figure 4: Depiction of the relationship bewteen the additional distance traveled during the high tide interval, measured in kilometers, and the total distance traveled in response to immediate disturbance events, also measured in kilometers.

Energy Requirement

In order to determine the energy expenditure of disturbance events it was first necessary to determine the energy the birds used in flight. The paper outlines an average flight speed of 12 m/s and flight costing a calculated 36 J/s of energy.

```
36 J/s / 12 m/s = 3 J/m = 3000 J/km
```

With this understanding of energy used per kilometer, I then created a table showing the average displacement distance for each disturbance event. This table was further mutated to display the energy use for this average movement, calculated as:

```
average distance (in km) * 3000 J/km
```

This energy use was then compared to the 860 kJ of the birds' daily required energy by subtracting this requirement from the energy loss due to disturbance displacement. This final column shows this impact of displacement from this it is clear that the energy lost to disturbance displacement only poses a minimal threat to the birds' daily energy budget.

Movement Category	Average Displacement Distance (km)	Energy Used (J)	Energy Requirement Minus Loss (J)	Energy Requirement Minus Loss (kJ)
air force	0.5455881	1636.764	858363.2	858.3632

Movement Category	Average Displacement Distance (km)	Energy Used (J)	Energy Requirement Minus Loss (J)	Energy Requirement Minus Loss (kJ)
non-anthropogenic	0.3538641	1061.592	858938.4	858.9384
pre-disturbance	0.0343983	103.195	859896.8	859.8968
walker	0.3347655	1004.297	858995.7	858.9957

Although there was most definitely a tidier way to create these tables, stepping through each one and `mutate()` them in a stepwise fashion allowed me to visualize the data and think through how the researchers in this study were able to assess this relationship.

Final Notes

The purpose of this paper was to step through the process outlined in the 2019 study conducted by Linssen, et al. Although the analysis here is flawed and the stats are not always in line with what was described in the original study (primarily due to my lack of ability to create similar models and determine the appropriate equations), I believe I arrived at similar conclusions regarding the impact of disturbance on the energy budget of these Eurasian Oystercatchers in Northern Europe.

Relevant Literature

Linssen, H., Van De Pol, M., Allen, A. M., Jans, M., Ens, B. J., Krijgsveld, K. L., & Van der Kolk, H. J. (2019). Disturbance increases high tide travel distance of a roosting shorebird but only marginally affects daily energy expenditure. *Avian Research*, 10(1), 1-11. https://doi.org/10.1186/s40657-019-0171-8