Locust Analysis Summary

Oliver Lountain

December 8, 2021

1   Introduction

2   Labelling the Shape of a Simulation

2.1   Summary Statistics

We have used a number of summary statistics to quantitatively describe the shape of the locust swarms. The statistics were chosen to capture the qualitative differences that can be observed in the distinct swarms, and in many cases simple statistics were used to maintain interpretability. Each statistic was computed on the swarm of locusts that were present at the end of the simulation.

2.1.1   Ratio of x-spread to y-spread

The first statistics that was used was the ratio of the x-spread to the y-spread of the swarm. That is, the distance between the smallest and largest x-coordinates of the locusts in the swarm, divided by the distance between the smallest and largest y-coordinates of the locusts in the swarm. The motivation for this arose from the clear distinction that could be seen between fan-shaped swarms, in which the y-spread was greater than the x-spread; comets, in which the x-spread was slightly greater than the y-spread; and streams and columns, in which the x-spread was substantially greater than the y-spread.

2.1.2   Number of Occupied Squares

The number of occupied squares was used as a measure of the total spread of the swarm. This could be used to distinguish the swarms which had all locusts concentrated in a small number of grid squares from those which were more spread out.

2.1.3   Skewness of Density Distribution

The skewness of the density distribution was used to distinguish the swarms which had a more even distribution of locusts than those which had locusts concentrated at the front or back of the swarm. In retrospect, it may have been better to look at the skewness of the density of x and y coordinates separately, or possibly just the x-coordinates.

2.1.4 Fractal Dimension

The fractal dimension of the swarm was also calculated. This was a less interpretable statistic but was investigated as it had been used in previous analysis. In the end it was used as it seemed to be a useful predictor in the model.

Unused Summary Statistics

2.1.5 Proportion of Squares Containing a Threshold Number of Locusts

2.1.6 Rotation Determined by PCA

2.1.7 Persistent Homology

2.2 Training a Model

3 Predicting Shape from Simulation Parameters

3.1 Training a Second Model

4 Discussion