EXPLORATORY ANALYSIS OF AVIAN POINT PATTERN DATA: APPROXIMATING METHODS OF INTENSITY ON AIRFIELD HABITAT OF INTERIOR ALASKA

By

Emily M. Richmond

A Project Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in

Statistics

University of Alaska Fairbanks

May 2022

APPROVED:

Ronald Barry, Committee Chair
Margaret Short, Committee Member
Scott Goddard, Committee Member
John Rhodes, Department Chair
Department of Mathematics & Statistics

ABSTRACT

The Animal and Plant Health Inspection Service agency of the United States Department of Agriculture began their work on United States Army Garrison Fort Wainwright of Alaska in 2018. In conjunction with airfield personnel, the main objective of this agency is to protect aircraft on Ladd Army Airfield (LAAF) from wildlife hazards and mitigate human-wildlife interactions on Post. The main wildlife hazard for aircraft is of the avian variety. The patterns of avian use on LAAF were examined for the first time using various non-parametric and parametric spatial methods. The main non-parametric technique applied was kernel density estimation of points in two-dimensional contour plots and three-dimensional surfaces. As for parametric means, Poisson point process modeling was used to estimate intensity (points per unit area) of the spatial region in question. Each year displayed a unique pattern of use among density plots that were consistent with an inhomogeneous process upon tests of complete spatial randomness. The baseline estimated intensity (homogeneous process) for years 2018, 2019, and 2020 were 1.609, 0.986, and 1.450 observations per hectare, respectively. Spatial locations as covariates revealed that intensity varies in North-South or East-West directions depending on the year. In addition to fleshing the dataset at hand, I outline theory and steps taken to numerically approximate the likelihood of the inhomogeneous Poisson point process. Logistic regression of observations on a continuous covariate (minimum distance to water) was used to demonstrate that fine pixel approximation yields adequate estimates of intensity.

Contents

1 Introduction **3**

1.1 Objectives . . . 3

2 Methods **3**

2.1 Data Collection . . . 3

2.2 Non-parametric Intensity Estimation . . . 3

2.3 Parametric Modeling . . . 4

2.4 Poisson Point Process Modeling . . . 5

2.4.1 Assumptions . . . 5

2.4.2 Assess Complete Spatial Randomness . . . 5

2.4.3 Models . . . 6

3 Results **6**

4 Discussion **8**

5 Approximating the Poisson Point Process **9**

5.1 spatStat Method . . . 9

5.2 Logistic Regression Approximation . . . 10

5.2.1 Process in R . . . 10

5.2.2 Results . . . 11

6 Appendix **12**

6.1 Figures . . . 12

6.2 References . . . 28

1 Introduction

The Federal Aviation Administration (FAA) reported tens of thousands of collisions (strikes) with aircraft and wildlife in recent years among all civil aircraft costing millions of dollars and threatening human health and safety. Avian species made up 98.8% of 11,605 reported strikes in 2020 in the US [5]. In the interior of Alaska, airports affiliated with the United States Army and Air Force implement bird and wildlife aircraft strike hazard plans to provide instructions for reporting strikes and detailing actions to be taken for management of the airfield. Wildlife Services (WS) of the United States Department of Agriculture provide wildlife hazard management at 33 airports in Alaska: 23 civil, 7 joint-use, and 3 military as of 2018 [1]. In 2018, United States Army Garrison Fort Wainwright (FWW) in Fairbanks enlisted WS to maintain safe operating conditions for aircraft during the spring – fall months on Ladd Army Airfield (LAAF), assisting airfield personnel in the dispersal of wildlife.

1.1 Objectives

The main objective of this project is to raise awareness of wildlife hazard on LAAF through exploratory analyses of avian hazing events for the years 2018-2020. The analysis of this dataset provided by WS was approached from a spatial statistical perspective. Rather than approximating important features of a point pattern by eye, the field of spatial statistics serves to provide summary statistics and model-based explanations of point patterns in a region of interest. I hope my work can be implemented in future seasons to come so that trends through time may be analyzed and FWW can use the information to assess their current management strategies. An additional objective for this project is to learn and demonstrate the techniques of approximating intensity over a region. One such method as described in Spatial Point Patterns: Methodology and Applications with R, is to use logistic regression on a fine pixel grid [4]. All analyses and reported figures were completed in R. [13]

2 Methods

2.1 Data Collection

Technicians sighted wildlife from the wayside of the airfield using binoculars or happened upon them during daily rounds in a vehicle. Upon arrival at the observed threat, technicians would determine the best method of dispersing wildlife in the safest and quickest fashion. Pyrotechnic weaponry and vehicle hazing techniques are most often used in hopes of pushing wildlife in a safe direction off the airfield. The locations of each encounter were recorded along with, but not limited to, species, count, direction of flight, hazing action and weather covariates. Dispersing a group of birds was considered one event. Rounds on the airfield were made Monday-Friday early morning before general airfield activity and at least once in the evening for a minimum of two times a day. The data used is a subset of 968 points that span the years 2018-2020 within a 2.39km^2 (239 hectares) spatial window (Figure 1). Only months April-September were considered as WS had a significant presence on LAAF during this time which includes both spring and fall migrations.

2.2 Non-parametric Intensity Estimation

Exploratory analysis of a dataset serves to highlight features of a spatial pattern in hopes of forming hypotheses to test. Non-parametric density estimation uses the raw data to approximate the probability density when the setting of the data is "not well known" or the data "do not appear to follow a parsimonious parametric form" [9]. Assuming independent

and identically distributed samples, each location is assigned a kernel and the collection of locations approximates the true probability density function [10]. Spatial data requires bivariate density estimation as each point represents a coordinate of two values.

$$\hat{f}(x,y) = \frac{1}{nh^2} \sum_{j=1}^{n} \phi\left(\frac{x - X_{1j}}{h}\right) \phi\left(\frac{y - Y_{1j}}{h}\right)$$

The ggplot2 package of R was used to estimate the density of avian observations on LAAF for all years [11]. Point patterns were laid over the region for each year of the dataset (Figures 2,3,4) and the default methods of density estimation were used within ggplot2 to generate two-dimensional contour plots. Specifically, the gaussian product kernel for bivariate density estimation was implemented with a normal reference bandwidth h [10]. In addition to the two-dimensional contour plots, the lattice package in R provides the tools to express the density of points as a three-dimensional surface [8]. Surfaces were made for all years.

2.3 Parametric Modeling

Loglinear Poisson Regression

A Poisson loglinear model is a generalized linear model (GLM) that assumes our counts are Poisson distributed [2]. Based on the distribution of points in the spatial region, I hypothesized that each year would contain different numbers of observations in the season from April-September. Likewise, due to spring and fall migrations, it was expected that each month within the season would contain different expected frequencies of observations.

Table 1: *A contingency table of observations in each month of a given year.*

Month	2018	2019	2020	Total
April	1	0	28	29
May	37	51	85	173
June	114	58	100	272
July	72	73	77	222
August	123	25	39	187
September	38	29	18	85
Total	385	236	347	968

The models I considered were additive and interactive with time variables Month and Year as explanatory variables to test this hypothesis. A quadratic model was considered for month to test the expectation that observations would be higher in the early and late months of this time frame. Akaike's Information Criterion (AIC) was used to pick the best model [3].

Table 2: *Candidate set of models with time explanatory variables to use for loglinear Poisson regression*

Models
Month
Year
Month + Year
Month + Month^2
Month + Year + Month:Year

Chi-square goodness-of-fit

The cover type at each hazing location was recorded. A chi-square goodness-of-fit test was used to assess whether points in each year and cover type matched their expected frequencies based on the probability of being in each cover type on the airfield. Probabilities of being in each cover type were calculated based on the proportion of each cover type present on the airfield. I chose to include only locations with cover types applicable to the airfield: runway, taxiway, shortgrass, and structure. Classifying the excluded points as "other" was an option considered for a fifth category, however it is not reasonable to quantify "other" cover types as a proportion of the airfield.

,

Table 3: *Contingency table of counts in each year for relevant cover type of the airfield. The probabilities of being in each cover type is the proportion of area that cover type takes up in the region.*

Cover	2018	2019	2020	Total
Runway (0.05)	76	33	68	177
Shortgrass (0.63)	116	90	89	295
Structure (0.02)	13	10	16	39
Taxiway (0.30)	153	98	129	380
Total	358	231	302	891

2.4 Poisson Point Process Modeling

2.4.1 Assumptions

Point Process models allow for statistical modeling of spatial point processes to estimate intensity $\lambda(u)$ (points per unit area) at a given point in the region. In a homogeneous process, the value of intensity $\lambda(u)$ is constant across a region with total area $|W|$ and assumes complete spatial randomness of points (CSR) with the following assumptions[4]:

1. the random variable $\mathbf{X}$ is the number of points in a region of interest and is Poisson distributed

$$\mathbf{X} \sim Poisson(\lambda|W|)$$

2. the intensity is constant across the region with expected number of points equal to $\lambda|W|$
3. The number of points in each disjoint pixel $B_1, B_2, ..., B_j$ of the region are random variables
4. For each disjoint pixel B, points are independent and uniformly distributed in B

2.4.2 Assess Complete Spatial Randomness

If statistical techniques detect departure from any assumption, it would be reasonable to consider the point pattern as not CSR. I hypothesized that my point pattern will violate the assumption of constant intensity (assumption 2), independent disjoint regions of $|W|$ (assumption 3), and uniform distribution of points (assumption 4). I used the empirical K-function to visually assess if my point pattern violates assumptions 2 and 4 [7]. I chose the translation edge correction method to combat the inherent loss of information when counting the average neighboring points per event along edges of a region [6]. I used a Monte Carlo method to test my hypothesis called the DCLF test. The DCLF test in spatStat uses the integrated mean squared error of my data-derived K-estimate and K-estimates from patterns simulated from a CSR process to find the probability of observing a K-estimate as extreme or more extreme than mine [12]. I performed a quadrat test to assess if my point pattern violates assumption 3. I hypothesized that the counts of

observations will depend on the quadrat it is in. For each year of data, I divided my region into a three by three grid for a total of nine quadrats and performed a chi-square test of independence [14].

2.4.3 Models

I constructed baseline models assuming a homogeneous process at all locations u where intensity is directly estimated from the data and is constant over the spatial region:

$$\lambda(u) = \frac{\mathbf{x}}{|W|}$$

Given the previous kernel density plots, I hypothesized the intensity of points over my region is inhomogeneous. There was a lack of relevant covariates in the form spatStat needs to fit such models so I used spatial locations (x,y) recorded in the universal transverse mercator coordinate system (UTMs) as covariates to estimate θ given a baseline intensity $b(u)$ calculated using the equation above:

$$\lambda(u) = e^{b(u)+\theta_1 x+\theta_2 y}$$

All models were performed using the ppm() function in the package spatStat [4].

3 Results

Kernel Density Estimation

The appendix contains the two-dimensional and 3D surfaces of the spatial region for each year (Figure 5,6,7,8,9,10). As expected, each year displayed a unique pattern with high incidents of hazing events adjacent to the runway and along taxiways.

Loglinear Poisson Regression

Models were evaluated using the AIC score. The lowest AIC value corresponded to the saturated model for Month and Year, meaning the proportion of observations in a month of a given year explain the dataset the best.

Table 4: *Loglinear Poisson regression models and AIC score. The model with the lowest AIC was the interaction model between Month and Year.*

Models	AIC
Month	271.2
Year	460.6
Month + Month^2	472.8
Month + Year	470.7
Month + Year + Month:Year	**128.7**

Chi-square Goodness-of-Fit on Cover Type

The observed counts do not follow the expected pattern of use given the probabilities of each cover type. For each year we have an extremely large chi-square test statistic well beyond the distribution mean equal to the degrees of freedom.

Table 5: *The chi-square value, degrees of freedom and p-value for each year's goodness-of-fit test on occupancy of cover type.*

Year	χ^2	df	p-value
2018	292.5	3	≈ 0
2019	87.6	3	≈ 0
2020	299	3	≈ 0

Empirical K-function

The theoretical K-function lies in red as a reference for CSR data (Figure 11). For all years, the estimated K-function lies above the theoretical curve which is consistent with clustering of points. For a typical location, there are more neighboring points than expected at given distance. The DCLF test for each year ran 99 simulations for a CSR process and ranked the estimated K-estimate for each, including the estimate for my data. For each year, the DCLF test returned a p-value of 0.01, indicating my estimated k-function is far different from the k-functions simulated under CSR, giving me evidence to believe my data do not come from a CSR process.

Quadrat Counting

Dividing up the spatial region into nine quadrats and performing a two-sided chi-square test of CSR yielded significant results for evidence of not CSR as seen by the large test statistics and small p-values below.

Table 6: *The quadrat test reports a chi-square statistic, degrees of freedom, and p value. The grid size was the same for all years.*

Year	χ^2	df	p	grid size
2018	59.3	8	≈ 0	3x3
2019	18.8	8	≈ 0	3x3
2020	23.6	8	≈ 0	3x3

To see detailed quadrat information, see Figure 12 in the appendix.

Point Process Models

The software reports the log(intensity) per unit of the area (in m^2). After exponentiating each estimate and multiplying by 10^4 to go from square meters to hectares, the expected number of observations per hectare was 1.609, 0.986, and 1.45 for 2018, 2019, and 2020, respectively.

Table 7: *The baseline intensity (points per square meter) each year is below followed by the expected number of observations per hectare of the region after exponentiation of the coefficient and scaling by 10^4.*

Baseline Models	Coefficient	# Obs.\ hectare
2018	0.0001609	1.609
2019	0.0000986	0.986
2020	0.000145	1.45

The Cartesian models varied for each year as displayed below. In 2018, there were positive estimates for both directions. There was an increase in intensity moving eastward (x-direction) and northward (y-direction) with confidence intervals that did not contain zeros, deeming significance. In 2019, there were positive estimates with only a significant shift in intensity to the north. In 2020, only a significant shift in intensity in the westward direction was reported.

Table 8: *A cartesian model for each year is below. The estimates are not transformed in order to more easily interpret the confidence intervals. Significant explanatory variables have asterisks.*

Cartesian Models	Estimate	Coefficient	CL	CH	Sign.
2018	intercept	-9.643	-9.944	0	***
	x	0.0003	0.0002	0.00045	***
	y	0.0009	0.00049	0.0013	***
2019	Intercept	-9.626	-9.985	-9.267	***
	x	0.000008	-0.00015	0.00016	
	y	0.00088	0.00034	0.0014	***
2020	Intercept	-8.671	-8.944	-8.399	***
	x	-0.00018	-0.00031	-0.000054	**
	y	0.00019	-0.00024	0.00063	

4 Discussion

I conclude there is strong evidence in favor of inhomogeneity and a lack of complete spatial randomness. Observation patterns differed each year and between months of the season in question in the kernel density plots. The preference of the Month*Year interaction in loglinear modeling section is unsurprising. I expected observations to be more frequent in the beginning and end of the season as those are crucial times for spring and fall migration. However that was not the case. This should not be interpreted as lower risk time periods as each observation can represent a great flock of birds. The ranking of cover type use and availability matched in such a way that the most observations recorded were on taxiways which makes up 70% of the region however only 42.6% of observations did indeed occur there. The distribution of points seemed to be more evenly spread than expected. There is indication of particular pockets of clustering based on the jumps seen in the estimated K-functions in Figure 11. It is worth noting that 2018 and 2020 were very similar years while 2019 had less severe clustering. There were in total less observations for 2019, however. The changing of intensity in north-south directions occurred in two of the three years (2018 and 2019). Most observations took place on runways and taxiways. In general, the distribution of cover type is most similar in the east-west direction and different in the north-south direction. Birds could be choosing either long stretch of pavement (taxiway and runway) for foraging or warmth in the early months. A negative shift in intensity to the west could be explained by the fact that most aircraft landings approach from the east as opposed to the west, pushing birds to occupy the opposite end. However only 2020 displayed this pattern.

Future research for this dataset could focus on a measurement error model, species and count related patterns, or incorporating time as a third dimension in the model. There were interesting features in the data worth mentioning (Figure 1). It is clear that the observations are forming a grid in some spots. This could open the door to some sort of measurement error model for the devices used. I am unsure which units were used but it does raise a few questions. Either the device is heavily rounding the location or potentially the observer is recording by estimating locations nearby rather than getting exact coordinates. Gulls and geese pose problems as they often occur in groups. These major flocks of birds were only considered as one observation so incorporating the effect of count would be worth investigating.

From personal observation, gulls on the airfield either have a nest nearby or they are visiting from a common location off of the Chena River. Upon hazing, they often return to the same perching spots just north east of the airfield. Recognizing common areas for the birds to group up could be incorporated in the management strategy.

5 Approximating the Poisson Point Process

Maximum likelihood estimation is a powerful technique used to estimate unknown parameters of a model for distribution of data. Assuming conditions of regularity, maximum likelihood estimators (MLE) are asymptotically efficient and consistent with the true parameter. To estimate the parameters of intensity for a Poisson process, we begin by discretizing the spatial region into m independent and disjoint pixels so that the area a of each pixel is

$$a = \frac{|W|}{m}$$

Recall the first assumption that observations of a homogeneous process are of the following distribution $\mathbf{X} \sim Poisson(\lambda|W|)$. Now each pixel is a Poisson random variable according to assumption 3, with distribution $\mathbf{X} \cap m \sim Poisson(\lambda a)$. To find an estimate for unknown λ, we begin with the probability distribution of a single count - a Poisson random variable. The likelihood function is formed assuming independent points. Finally, we rescale the likelihood relative to the probability for a Poisson process of rate 1, take the logarithm of the function and ignore constants not dependent on the parameter [4].

$$\begin{aligned} f(x;\lambda a) &= \frac{(\lambda a)^x e^{(-\lambda a)}}{x!} \\ L(\lambda) &= \prod_{i=1}^{m} \left(\frac{(\lambda a)^{x_i} e^{(-\lambda a)}}{x_i!} \right) \\ logL(\lambda) &= \mathbf{x} log(\lambda) - \lambda|W| \end{aligned}$$

The solution to the equation that results from setting the derivative of this loglikelihood equal to zero is $\hat{\lambda} = \mathbf{x}/|W|$. It is the number of points divided by the area. We can then obtain the expected number of points by multiplying the estimate by $|W|$, the area of the spatial region. If we assume inhomogeneity, each observed location s has an intensity governed by a function of some covariate with coefficient θ. The loglikelihood for parameter θ becomes

$$logL(\theta) = \sum_{i=1}^{m} log(\lambda_\theta(s_i)) - \int_W \lambda_\theta(u_j)$$

where s_i are observed locations only and u_j is the location of all possible points contained in the region, observed and unobserved (dummy points). This integration is a double integral over the entire area of spatial region. The vector of parameters θ cannot be derived in closed form and must be approximated.

5.1 spatStat Method

In spatStat, the method of estimating parameters of intensity uses the Berman-Turner device[15]. As described above, the spatial region is divided into a grid of pixels, each with area a. This is called a quadrature scheme which can take

many forms, the simplest being an evenly spaced grid of rectangles. The integration is approximated as follows by a weighted sum over u_j distributed in m pixels

$$logL(\theta) = \sum_{i=1}^{n} log(\lambda_\theta(s_i)) - \sum_{j=1}^{m} \lambda_\theta(u_j)w_j$$

with weights $w_j = \frac{|T|}{k}$, when $|T|$ is the area of the pixel and k is the number of quadrature points in that pixel, ie. number of observed points in that pixel plus a single dummy point. The weights for all quadrature points must sum to the total area of the spatial region. Rewriting the loglikehood as follows allows for the estimation of θ using standard GLM software for a weighted Poisson regression on a new response $y_i = I_j/w_i$ where I_j indicates 1 if it is an observed point, 0 if a dummy point and $Y_j \sim Poisson(\lambda_j)$ [17].

$$\sum_{j=1}^{n} (y_j log\lambda_j - \lambda_j)w_i$$

5.2 Logistic Regression Approximation

Alternatively, the Poisson process can be fully approximated by a logistic regression model given a fine grid of disjoint pixels over the spatial region where $P(\mathbf{X} > 1) \to 0$ for a given pixel [16]. If we let area $a \to 0$ of the m^{th} pixel and fix the number of our observed points, it can be mathematically shown that the likelihood of a binomial distribution with the probability of a single point occupying a pixel equal to λa converges to a Poisson likelihood.

5.2.1 Process in R

In this demonstration, I used minimum distance to water as a covariate in an attempt to explain the spatial distribution of observations on LAAF on the combined dataset of all years. The Chena River wraps around the airfield from the north to the east. This water body tends to attract groups of shorebirds and waterfowl for foraging, roosting, and nesting. I did not have access to habitat files of the region to do my calculations. I used the Google Maps API tool provided by Theme Birds on Stamps to coarsely draw a polyline on the Chena River which returned a list of points for me to import into R [18]. Using the dist() function of the stats package, I calculated the minimum distance to water for each observed location and dummy point during the simulations. I created functions using the spatStat and logistic regression method separately. Having the spatStat method to compare the logistic regression results to will show that logistic regression is an appropriate technique to apply. The spatStat method uses the ppm() function with the data and a quadrature scheme of dummy points. Each point (real or dummy) has a value for minimum distance to water. Coefficient estimates are obtained by maximizing the pseudolikelihood in this method. I created a function that simulates the above for the following grid sizes:
Each subsequent size is 1.5 times the former with total dummy points increasing by 1.5^2. In general, the finer the grid, the better the approximation. As for logistic regression, a similar approach was made in that I wrote a function that simulated coefficient estimates using the glm() function. Each actual observation was a 1 in response vector while any pixel with a dummy point only was a 0, giving us our binary dataset. Each point had an associated value for minimum distance to water as our only covariate. For each method, I ran a simulation with increasing grid size as shown in the previous table.

Table 9: *The following grid sizes were used for each method of the simulation. The corresponding dummy points are the value squared.*

Grid Size	Total Dummy Points
60x60	3600
90x90	8100
135x135	18225
202x202	40804
303x303	91809
455x455	207025

5.2.2 Results

The inflation of the number of empty cells represented by the increase in dummy points in the region yields an approximation for the intensity of the region with estimated parameters and p-values that converge to the true value in both methods (Figures 14, 15). Note in Figure 16 the trend of the intercept changes by the log of the change in amount of pixels added. That is, if we started with a 60x60 grid and increase the grid size to 1.5 times the previous to obtain a new approximation result, the intercept will change by $log(1.5^2)$. Due to the size of the augmented data, this method can be computationally intensive, therefore it is not the default method of spatStat. The coefficients for the logistic regression model using the 202 by 202 grid size are displayed below. I chose this grid size over the 455 by 455 in order to save computation time when creating the predicted surface. Figure 13 shows the surface kernel density plot of the actual data on top for all years and the predicted surface given the estimated coefficients of the logistic model using the 202 by 202 grid size.

Table 10: *The table displays the output of regressing intensity on minimum distance to water. The negative sign tells us that as minimum distance increases, we can expect a lower intensity.*

	Estimate	SE	p-value
Intercept	-3.688	0.081	≈ 0
mindist	-0.00008	0.0001	0.475

Given that the estimated coefficient for minimum distance to water is not significant with a p-value of 0.475, I expected the predicted surface to not describe the actual surface well. It does seem to preserve the north to south trend seen in the actual surface. The edges for the actual surface are seen going to zero intensity. This is an issue in general with kernel density in that edges are hard to map. Data could exist just outside the boundary but is not used in the density estimation. The predictive surface has issues as well. It is not respecting the boundary and naively predicting intensity to just increase forever beyond the boundary. This demonstration serves to validate the logistic regression approximation and also to show that relevant distance covariates could be used to predict and describe intensity in future analyses.

6 Appendix

6.1 Figures

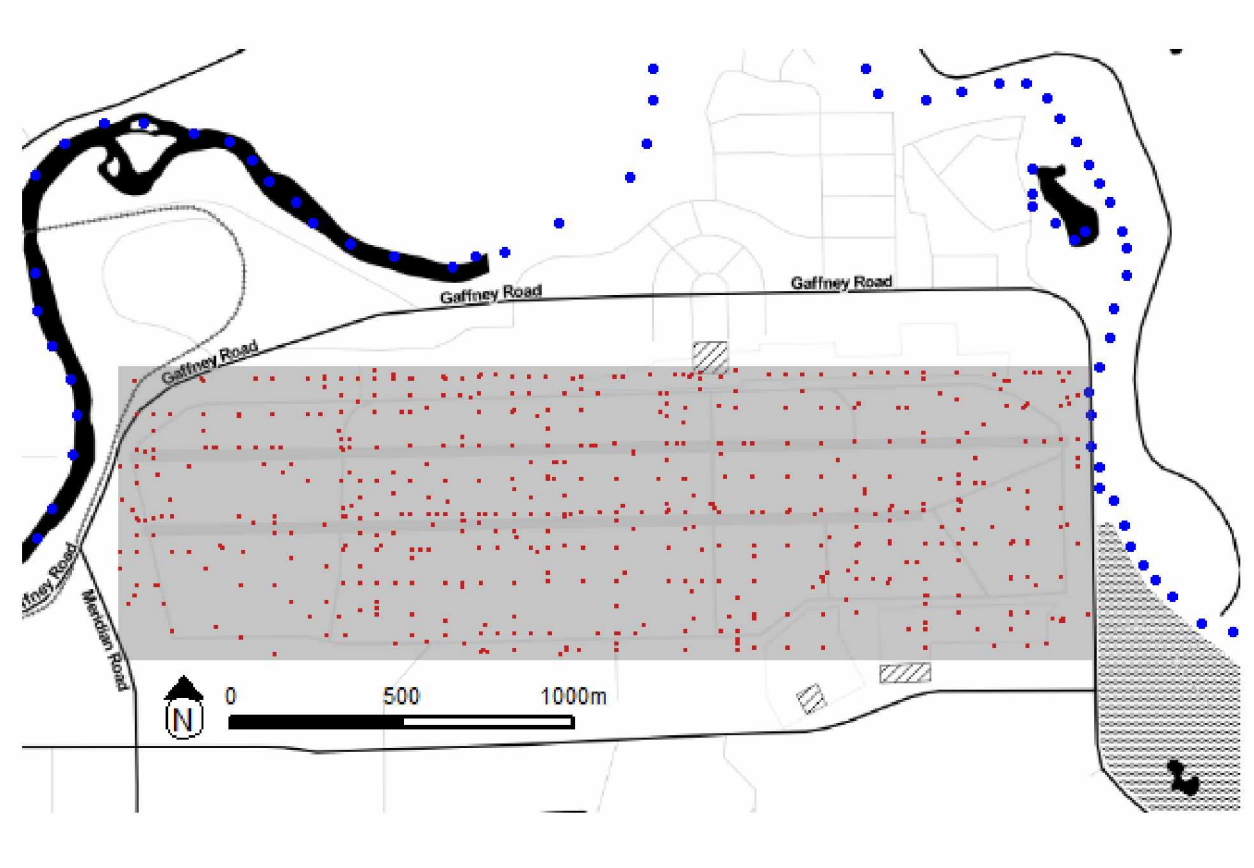

Figure 1: *All observations for the year 2018-2020 are plotted in red. The spatial window used in analyses is the shaded box surrounding the airfield. The blue dots represent water in the surrounding area. The main water feature is the Chena River seen bordering the airfield to east, north and west. There are smaller ponds to the south that are not on this map.*

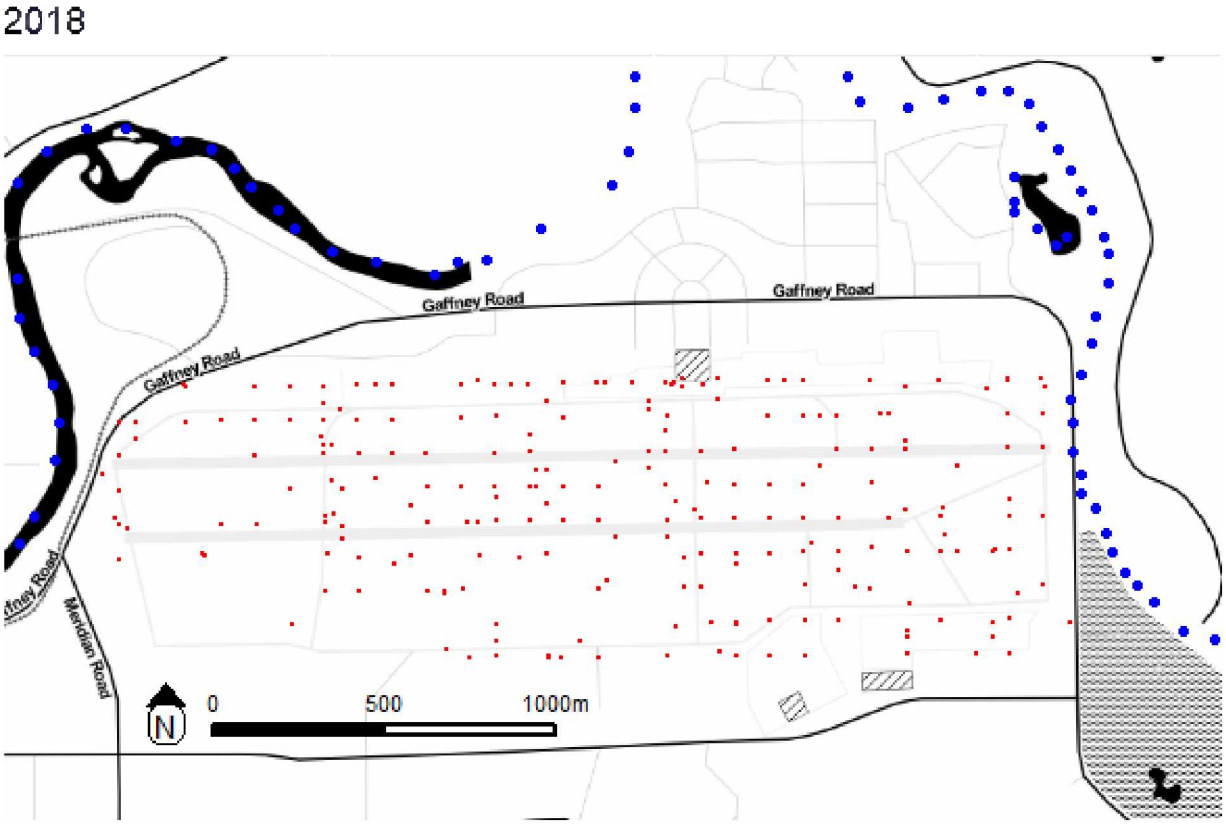

Figure 2: *Observations for 2018 only.*

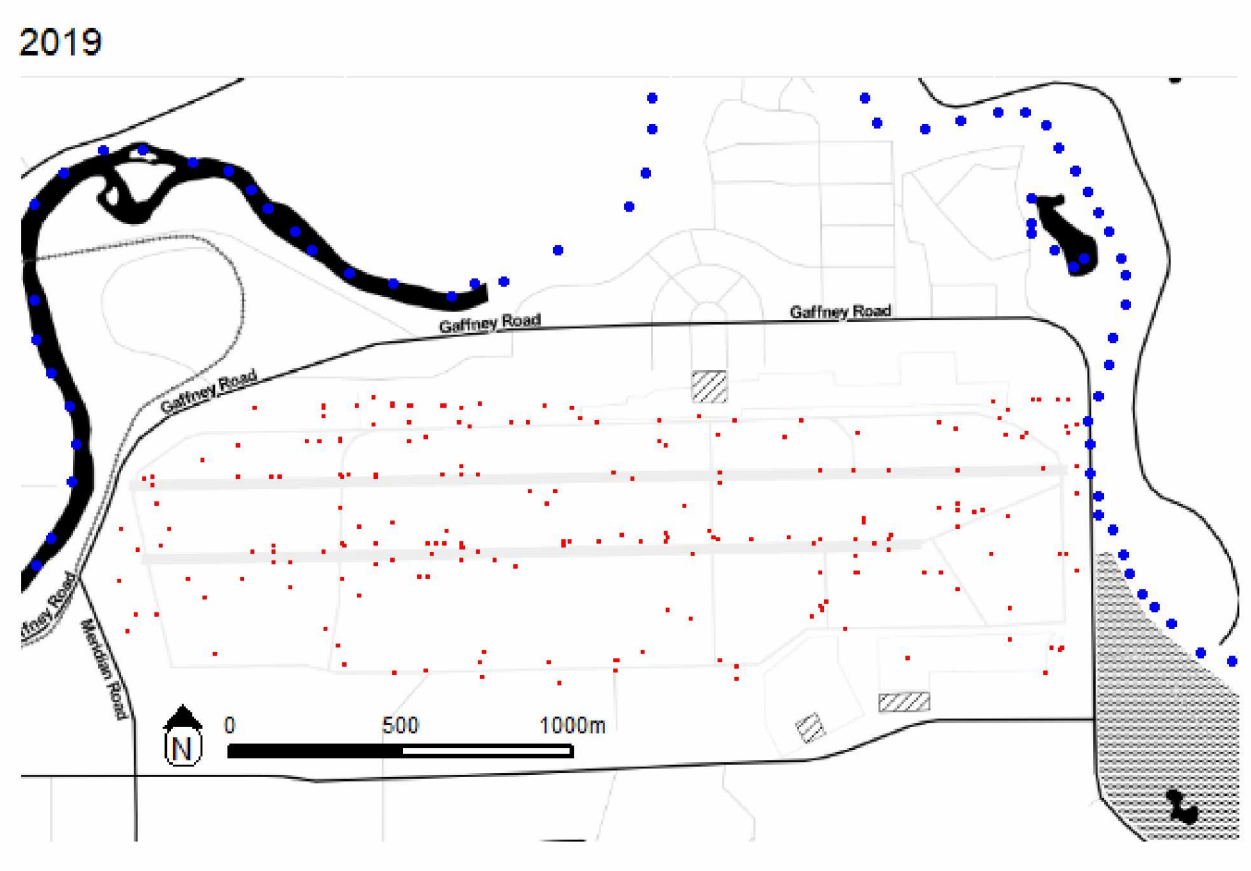

Figure 3: *Observations for 2019 only.*

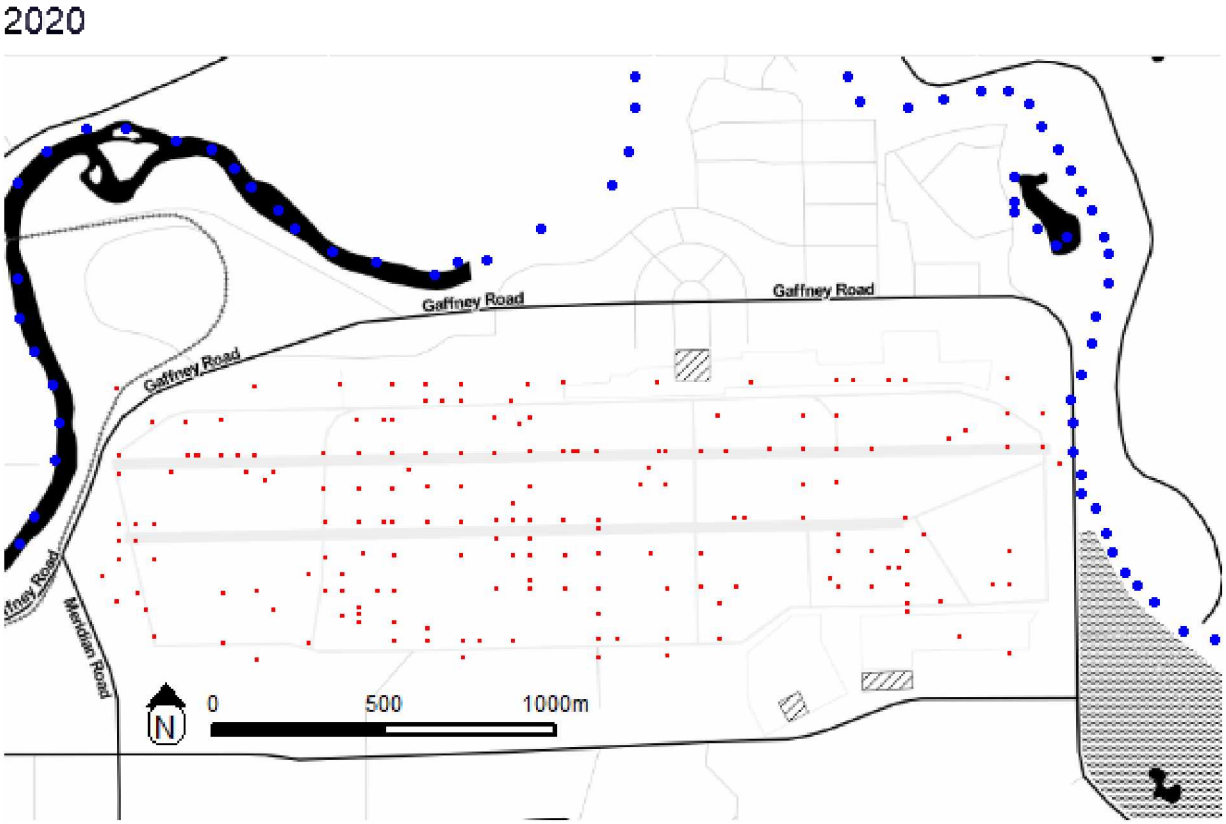

Figure 4: *Observations for 2020 only.*

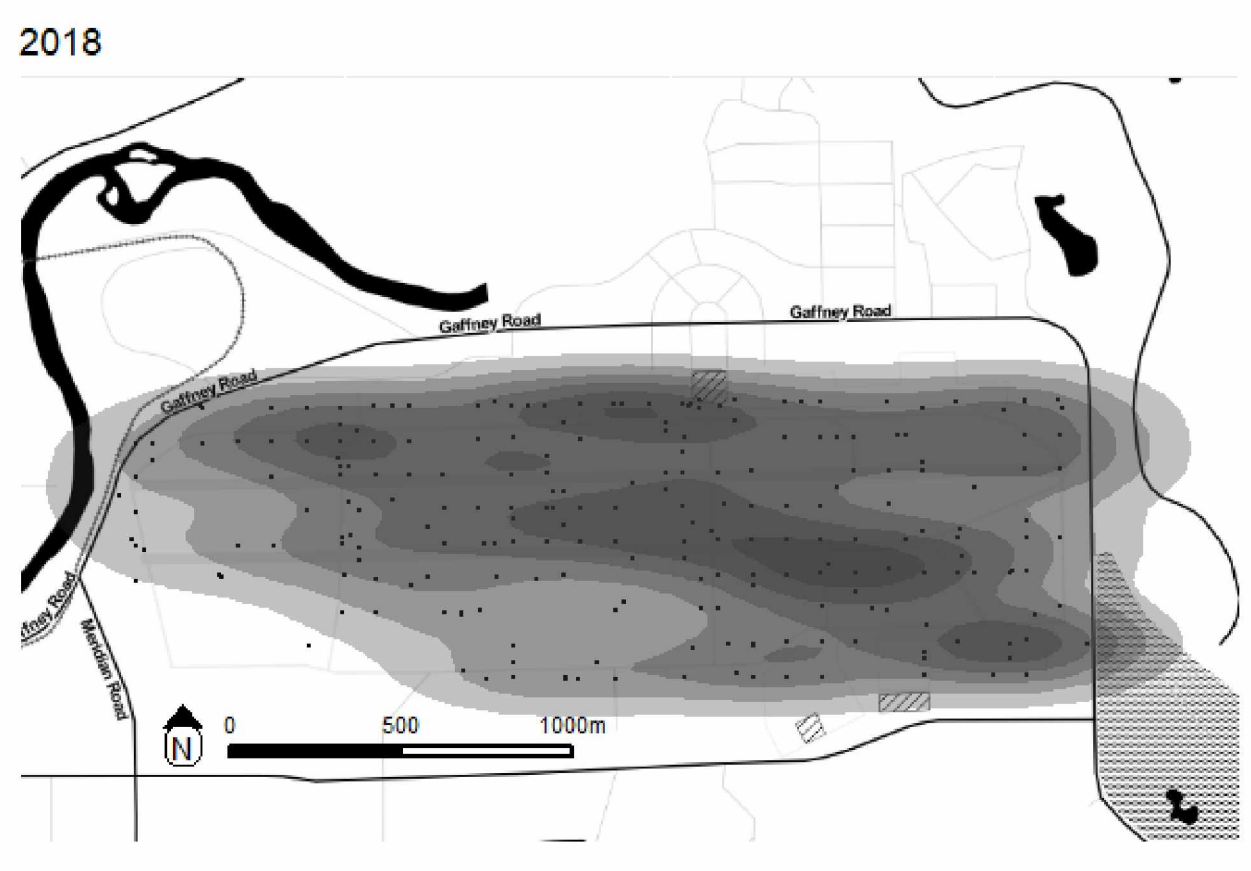

Figure 5: *Two dimensional contour plot of 2018 observations. There appears to be hot spots in the middle eastern part of the region.*

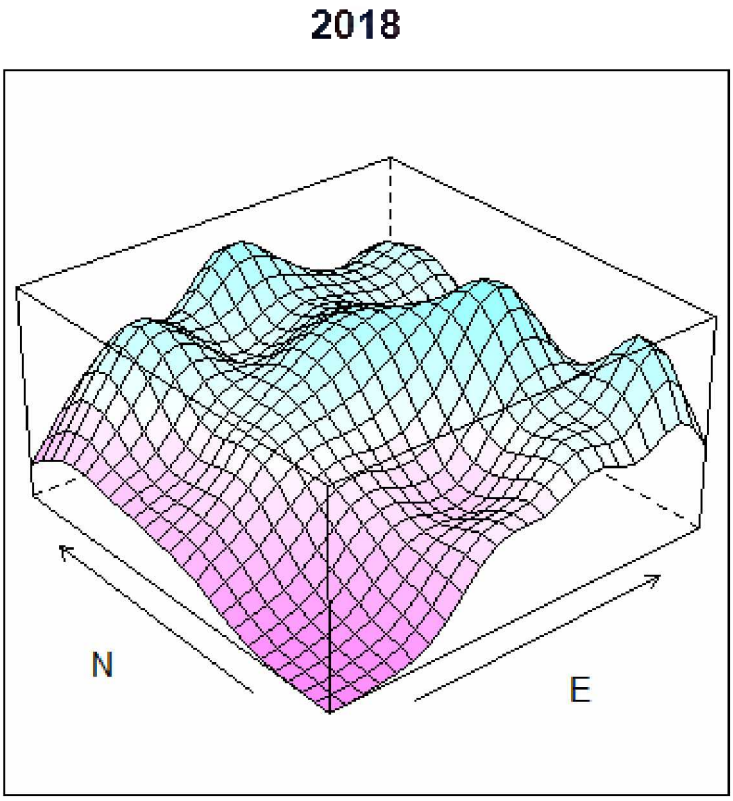

Figure 6: *Three dimensional surface plot of 2018 observations.*

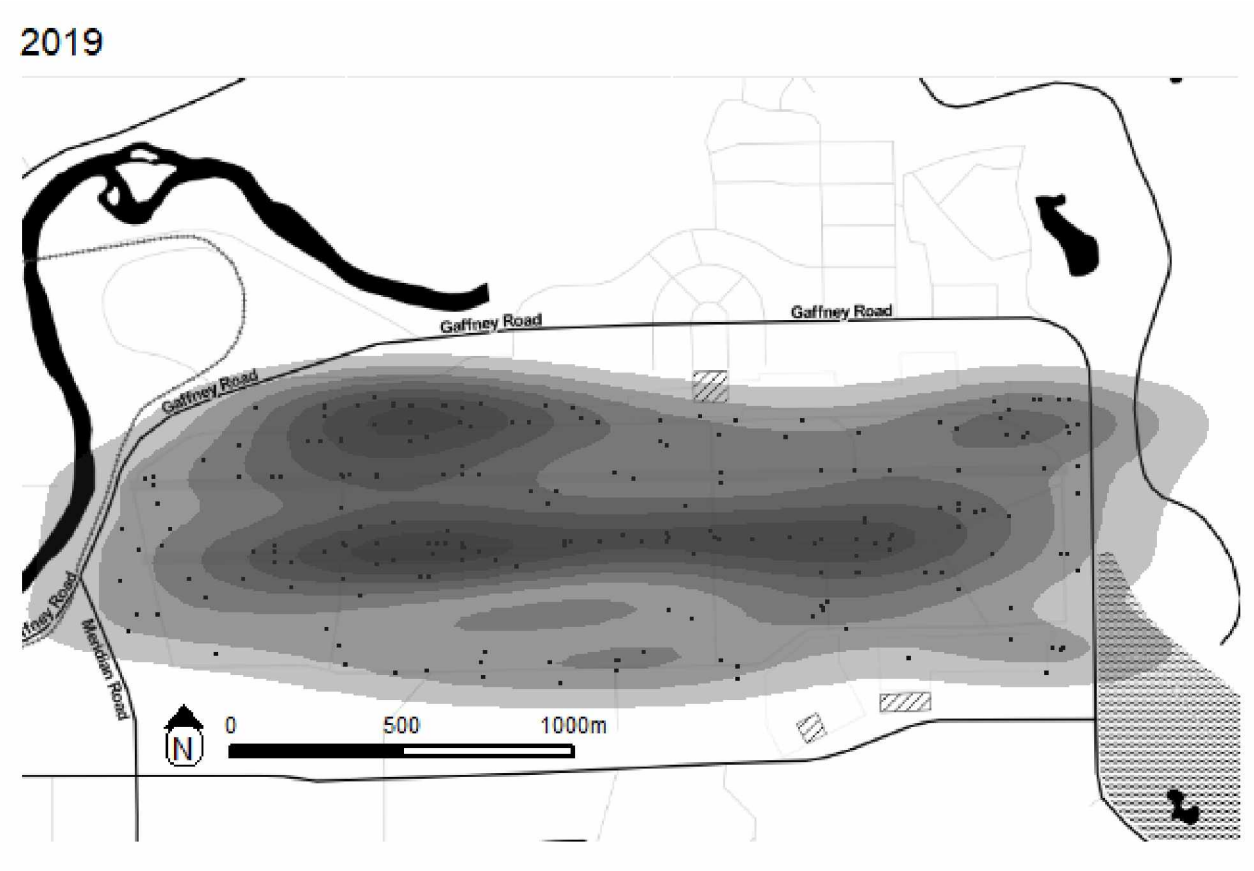

Figure 7: *Two dimensional contour plot of 2019 observations. There appears to be a pattern of observations stretching from the east to west on the longest taxiway of the airfield.*

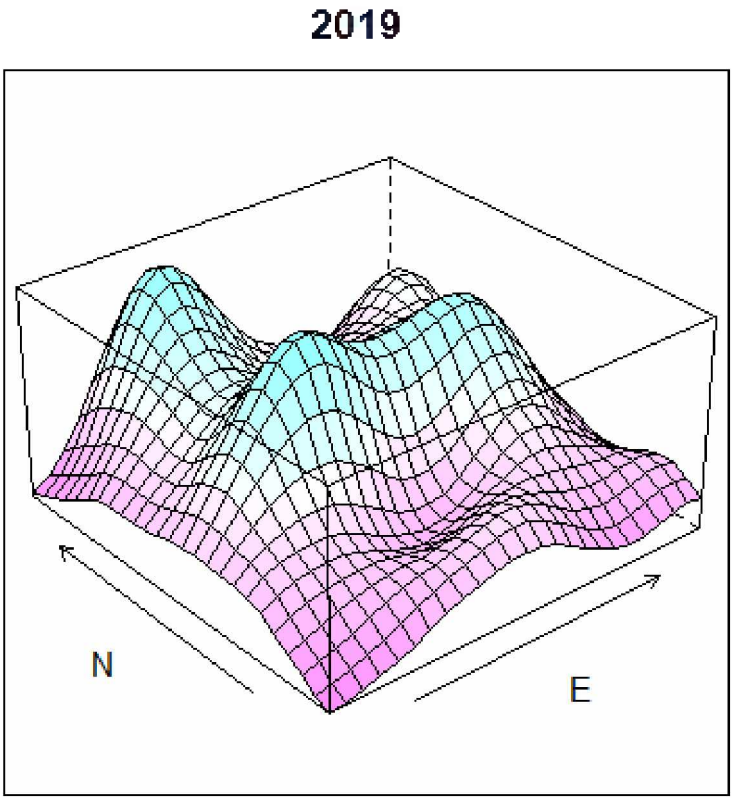

Figure 8: *Three dimensional surface plot of 2019 observations.*

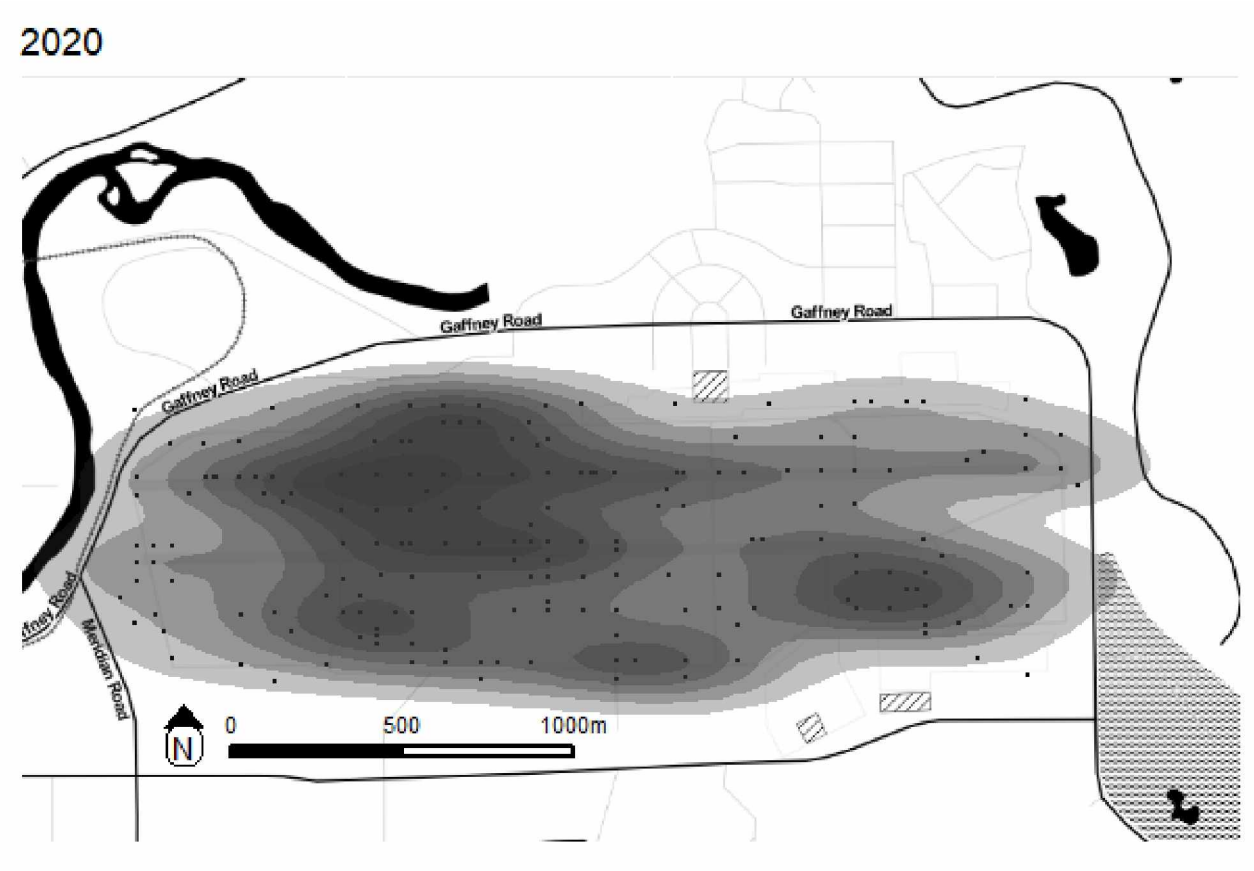

Figure 9: *Two dimensional contour plot of 2020 observations. There appears to be a pattern of observations on the north eastern half of the airfield.*

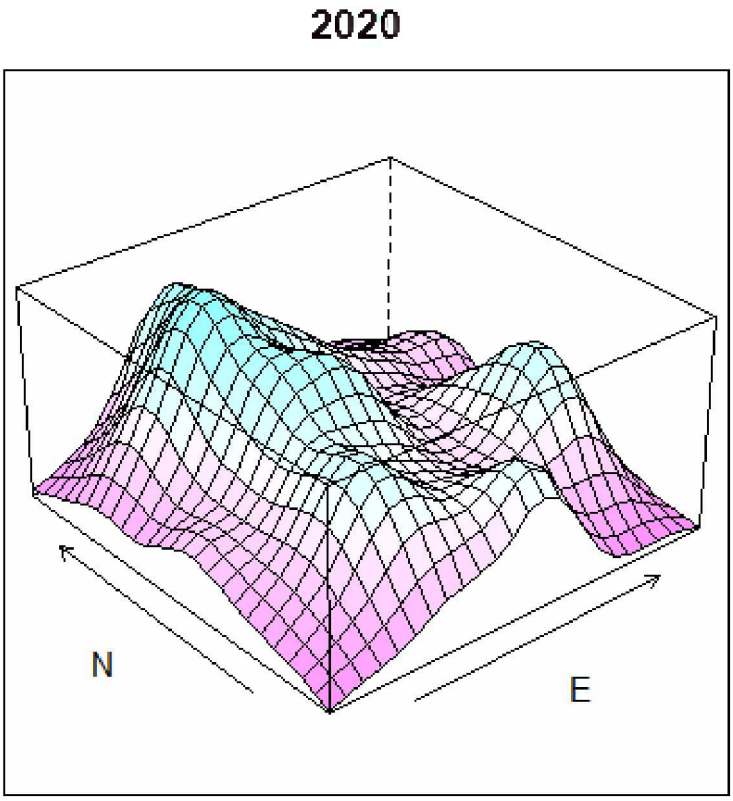

Figure 10: *Three dimensional surface plot of 2020 observations.*

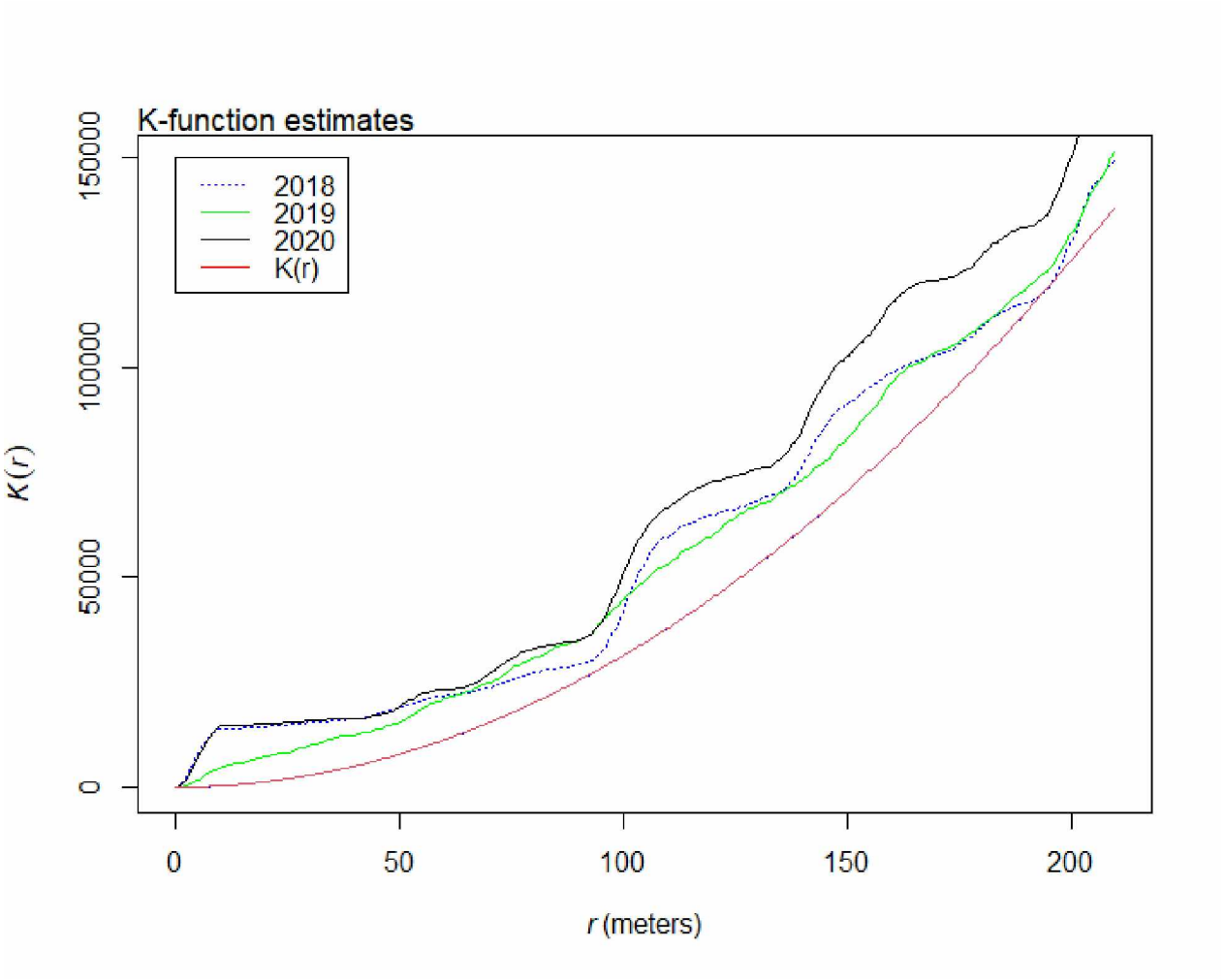

Figure 11: *The estimated K-functions for each year are plotted on the same graph in reference to the theoretical red line. The translation edge correction was applied in all years.*

2018

48 42.8 0.8	59 42.8 2.5	60 42.8 2.6
29 42.8 -2.1	51 42.8 1.3	49 42.8 0.95
3 42.8 -6.1	37 42.8 -0.88	49 42.8 0.95

2019

34 26.2 1.5	31 26.2 0.93	28 26.2 0.35
31 26.2 0.93	31 26.2 0.93	33 26.2 1.3
15 26.2 -2.2	16 26.2 -2	17 26.2 -1.8

2020

47 38.6 1.4	56 38.6 2.8	31 38.6 -1.2
39 38.6 0.072	43 38.6 0.72	25 38.6 -2.2
36 38.6 -0.41	46 38.6 1.2	24 38.6 -2.3

Figure 12: *The above three figures summarize the spatial region of each year into three numbers. The top two numbers represent the observed number of points (left) and expected number of points (right) for that quadrat. The bottom number is the pearson standardized residual which is an indicator for how much a particular quadrat deviates from the expected value. The greater the absolute value of the number, that quadrat contained more (positive value) or less (negative value) points than expected.*

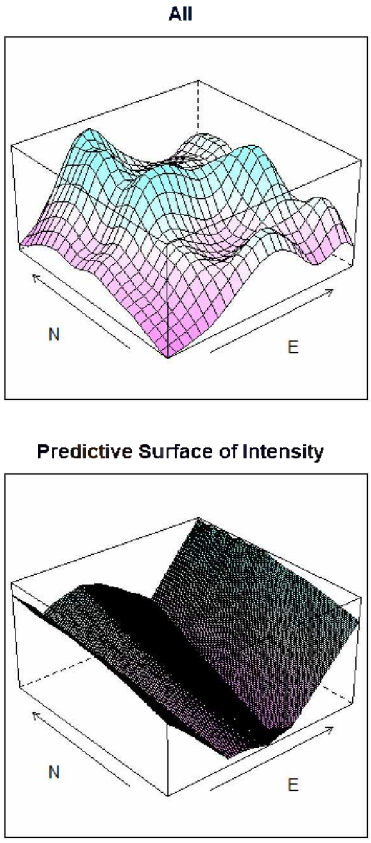

Figure 13: *For demonstration purposes, the kernel density surface for all years is displayed on the top and the predictive surface using the minimum distance to water model is on the bottom.*

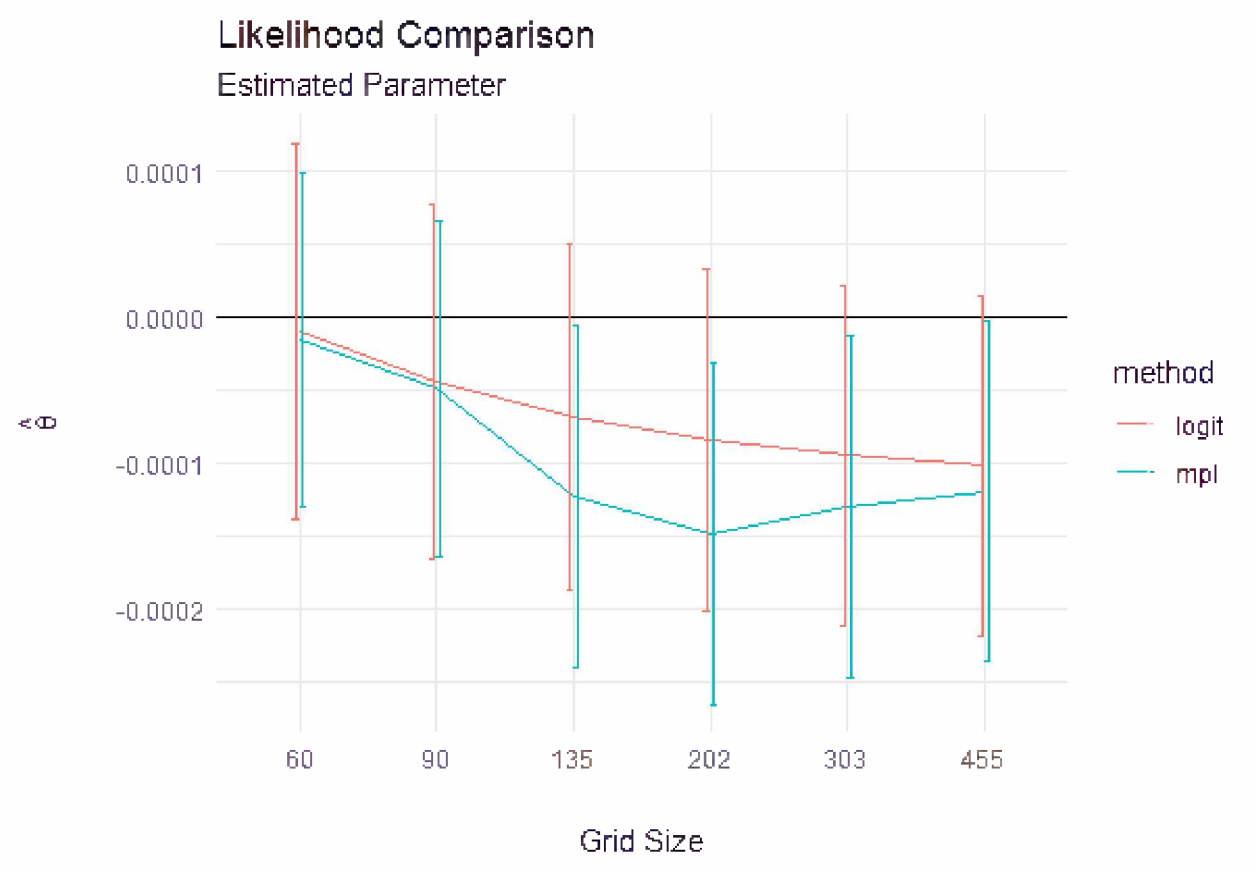

Figure 14: *Two methods were used on the same model. The ppm.quad() function in the spatStat package uses the maximum pseudolikelihood method (mpl;blue) and the glm() function in the stats package uses a likelihood for a logistic regression (logit;red). The estimated thetas for the covariate minimum distance to water are converging to a true value in both cases, albeit in different patterns. The curved lines are running through the reported estimated coefficient by the function and the bars extending are the standard errors. This was done on the entirety of the dataset, years 2018-20.*

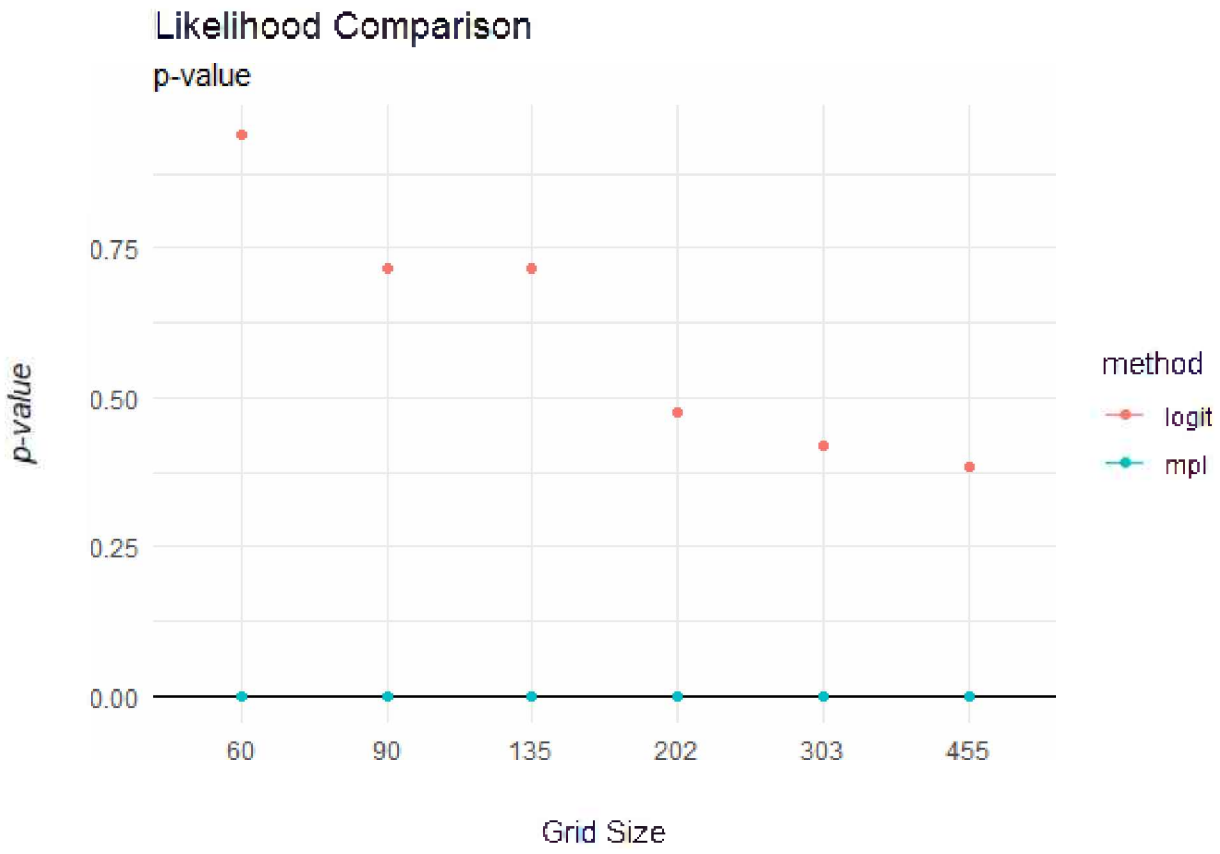

Figure 15: *Two methods were used on the same model. The ppm.quad() function in the spatStat package uses the maximum pseudolikelihood method (mpl;blue) and the glm() function in the stats package uses a likelihood for a logistic regression (logit;red). The p-values for the estimated coefficient of covariate minimum distance to water is shown converging for the logit method. The mpl method did not report and actual p-value, therefore there are no estimates.*

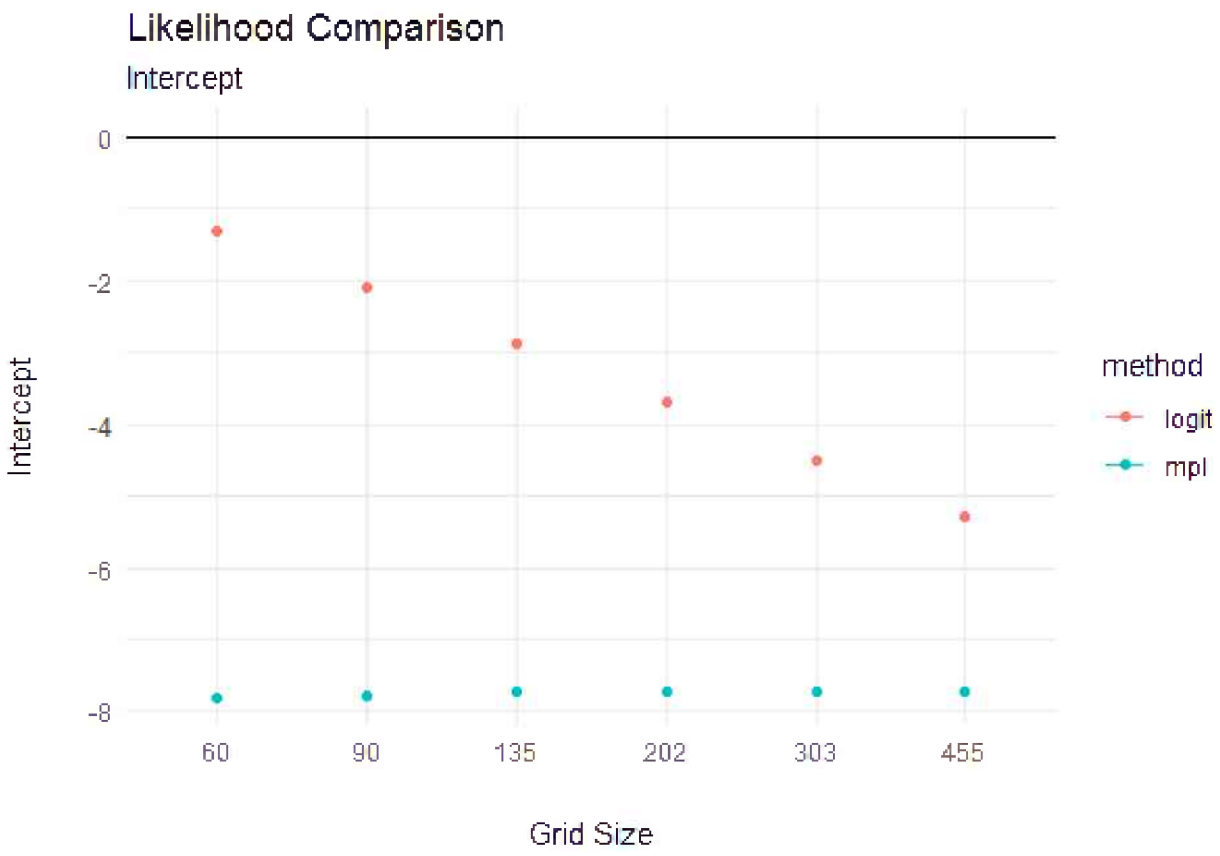

Figure 16: The trend of intercepts for an increasing grid size for the two methods are very different. The logit method decreases by a known amount based on the increase in grid size. The ppm.quad() function does not change all that much in comparison as grid size increases.

6.2 References

[1] Airport Wildlife Hazards Program. (2018). USDA APHIS: Airport Wildlife Hazards Program. https://www.aphis.usda.gov/

[2] Agresti A (2019). An Introduction to Categorical Data Analysis, Third Edition. Published by John Wiley & Sons, Inc.

[3] Akaike H (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & B. F. Csaki (Eds.),Second International Symposium on Information Theory, (pp. 267–281). Academiai Kiado: Budapest.

[4] Baddeley A, Rubak E, Turner R (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London.

[5] Dolbeer R. A., Begier M. J., Miller P. R.,Weller J. R., and Anderson A. L. (2021). Wildlife strikes to civil aircraft in the United States, 1990-2019. U.S. Department of Transportation, Federal Aviation Administration, Office of Airport Safety and Standards, Serial Report No. 26, Washington, DC., USA. (in press).

[6] Ohser J (1983). On estimators for the reduced second moment measure of point processes, Series Statistics, 14:1, 63-71, DOI: 10.1080/02331888308801687.

[7] Ripley B.D. (1977). Modelling Spatial Patterns. Journal of the Royal Statistical Society: Series B (Methodological), 39: 172-192. https://doi.org/10.1111/j.2517-6161.1977.tb01615.x

[8] Sarkar D (2008). Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5, http://lmdvr.r-forge.r-project.org.

[9] Scott D.W. (2004) : Multivariate Density Estimation and Visualization, Papers, No. 2004,16, Humboldt-Universität zu Berlin, Center for Applied Statistics and Economics (CASE), Berlin.

[10] Silverman B.W. (1986). Density estimation for statistics and data analysis (Chapman and Hall, London)

[11] Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

[12] Diggle PJ. (1986). Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. J Neurosci Methods. Oct;18(1-2):115-25. doi: 10.1016/0165-0270(86)90115-9. PMID: 3796038.

[13] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[14] Greig-Smith, P. (1952). The Use of Random and Contiguous Quadrats in the Study of the Structure of Plant Communities, Annals of Botany, Volume 16, Issue 2, Pages 293–316, https://doi.org/10.1093/oxfordjournals.aob.a083317

[15] Berman M., Turner, T. R. (1992). Approximating Point Process Likelihoods with GLIM. Journal of the Royal Statistical Society. Series C (Applied Statistics), 41(1), 31–38.

[16] Besag J., Milne R., & Zachary S. (1982). Point Process Limits of Lattice Processes. Journal of Applied Probability, 19(1), 210–216. https://doi.org/10.2307/3213930

[17] McCullagh, P. & Nelder, J.A. (1989) Generalized Linear Models. 2nd Edition, Chapman and Hall, London.

[18] Scharning, K. http://www.birdtheme.org/.