**The Challenges in Selecting a Model to Predict the Hare Population Density in the Peak District using Distance Sampling**

1. **Exploratory data analysis**

While distances are presumed to have been measured accurately using a laser range finder, the angles were measured with a compass. Analysis of the angles measured in the field as shown in Figure 1 seems to show that angles have been rounded to the nearest 10 degrees and that angles under 20 degrees have been heaped to 0. Figure 1 also shows that the perpendicular distances are generally close to the transect line and that the distances measured from the observer seem to be heaped at 25m, 50m, 100m, 350m and 400m.

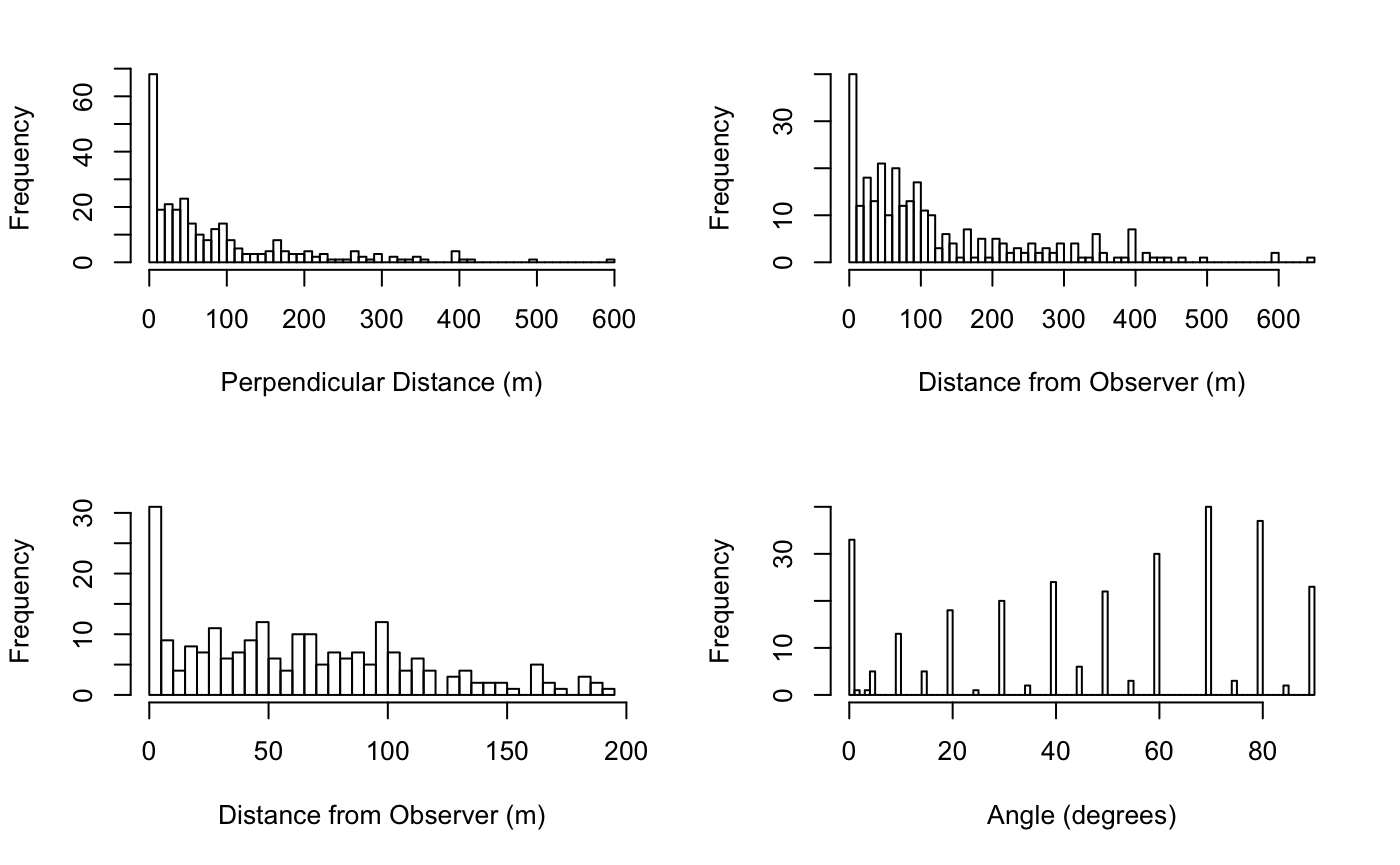


Figure 1 Analysis of measurements taken during survey showing assumed measurement errors. Top left showing perpendicular distance with many observations near the transect line. Top right and bottom left showing heaping in distance measured from observer. Bottom right showing heaping in angles measured.

Buckland (2001:267) suggests five possible solutions for measurement errors:

1. Record distances and angles more accurately
2. Use models for the detection function that always have a shoulder
3. Group the data before analysis
4. ‘Smear’ the data …
5. Use radial distance models [according to the text these have not been demonstrated to be robust].
6. **‘Smearing’ the measured angles**

Smearing, although often criticized, is regularly used in cetacean surveys to reduce the effects of measurement errors in angle readings (Buckland, 2001:269). The angles are adjusted within a given range using a uniform distribution

Sρ = | ρ + υ ∆ρ |

where

ρ = recorded angle,

υ = uniform random number between -0.5 and 0.5,

∆ρ = range of angles to be smeared over.

(Kinzey, 2002) and (Buckland, 2001:269-271).

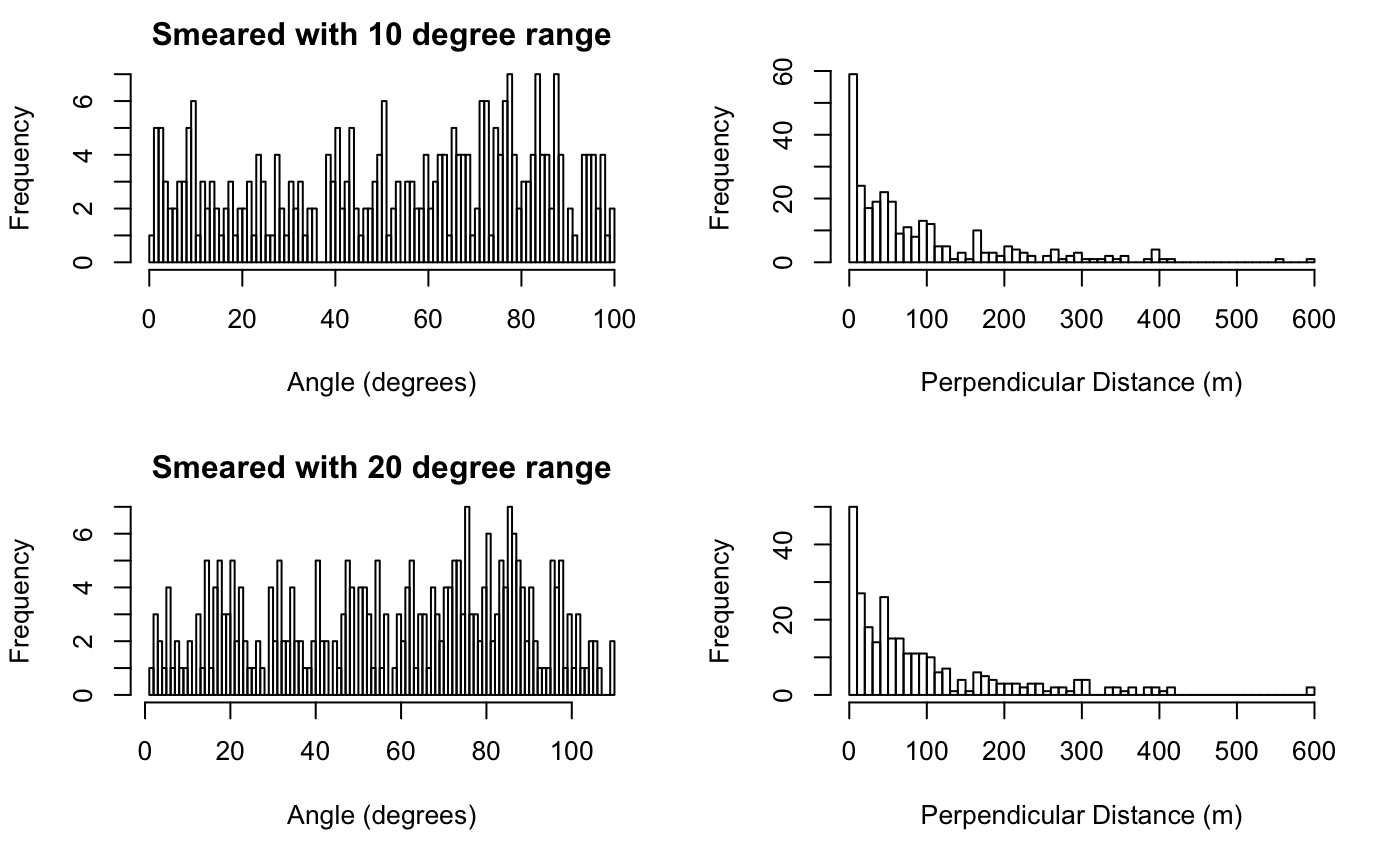


Figure 2 Effects of smearing the angle measured in the field. Top shows the effect of smearing the angle by 10 degrees on both the angle and the perpendicular distance from the transect line. Bottom shows the effect of smearing the angle by 20 degrees.

Figure 2 shows there is very little effect on the perpendicular distance when smearing the angle. Figure 3 shows that the majority of the distances recorded with a small angle were for animals very close to the observer; this explains why the smeared angle has very little effect on the number of observations close to the transect line.

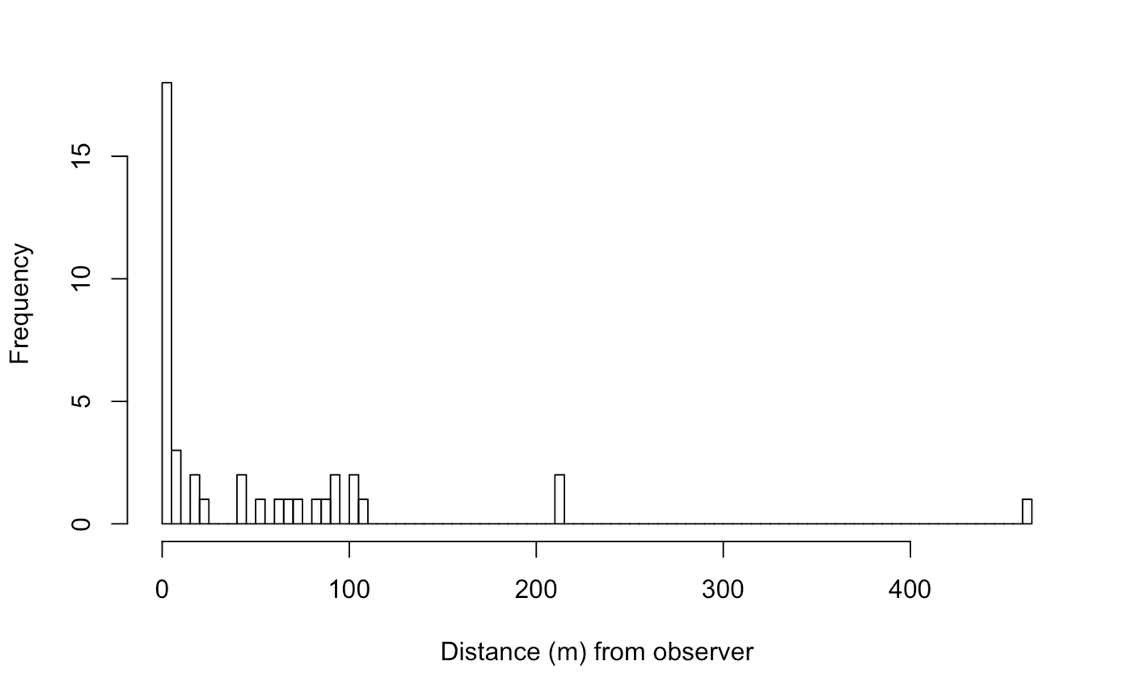


Figure 3 Distance measured from observer where measured angle is less than 10 degrees

1. **Grouping the data**

Buckland (2001:109) states that grouping should be used where there are measurement errors and that appropriate grouping can lead to a better model fit and improved estimates of density. The text suggests that cutpoints should be selected that avoid the rounding distances. When many sightings fall on the transect (which is particularly common when using angles and distances), then a relatively wide first interval should be chosen. Bibby (1998) states that when grouping data to deal with heaping, the first interval chosen should be narrow and fall within the ‘shoulder’ and the other groups should increase with distance from the transect. Buckland (2001:158) suggests that six to eight groups is a reasonable number for resolving heaping.

1. **Shape Criterion**

Buckland (2001:42) discusses the shape criterion and says that a detection function should have a shoulder, which means that the derivative of the detection function on the transect is equal to 0, i.e. the observer detects all animals on and just off the transect. The text goes on to say that this is particularly important when heaping at zero is suspected.

1. **Cluster Size**

Larger cluster sizes are more visible at distance than single animals. This means there is a bias for only recording animals in larger clusters at a further distance. Figure 4 shows that there is some cluster size bias in the hare data.

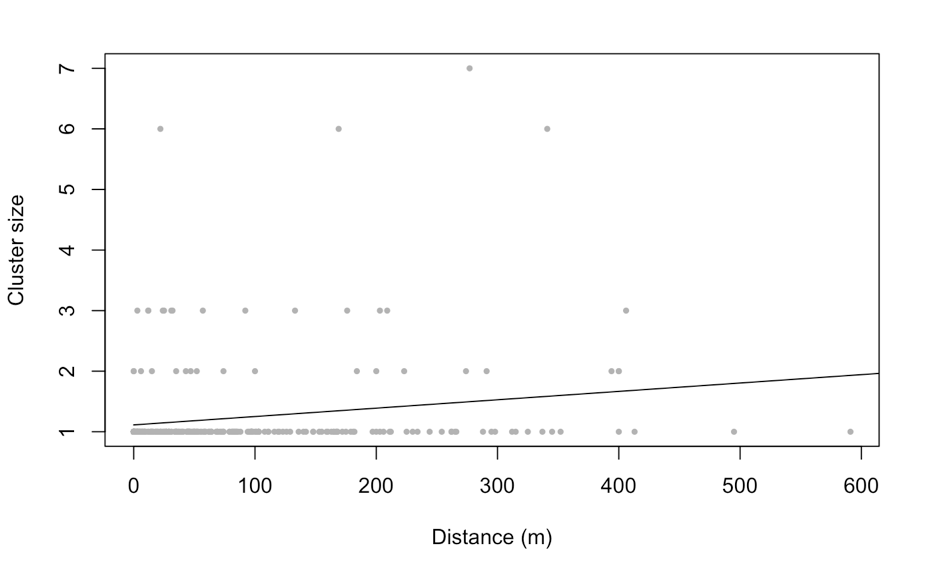


Figure 4 The number of animals observed in each observation against the distance from the transect.

(Buckland, 2001:71-75) Gives the following approaches for dealing with cluster size bias:

1. Truncate the data to ignore large distances where bias occurs
2. Add the cluster size as a covariate to the distance when modelling
3. Calculate the mean cluster size using a regression estimator
4. Replace clusters with individuals. When the number of detected clusters is small (13% of detections were clusters in the hare data), then this approach may perform poorly.

Fewer clusters of 2 or 3 hares appear to have been seen beyond 200m from the transect.

1. **Truncation**

Right-truncating the data may lead to slightly reduced precision, however, losing some outliers will help reduce the number of parameters in the detection function and so reduce bias (Buckland, 2001:107).

Truncation approaches include:

1. Truncating the furthest 5% - 15% of sightings
2. Fitting a preliminary model and truncating any sightings at a distance with less than 15% chance of detection.

(Buckland, 2001:103-104).

1. **Model Selection**

Models examined were uniform with polynomial adjustment, half-normal with cosine adjustment and hazard-rate with hermite-polynomial adjustment, see Appendix 1. In general, trying every combination of detection function with adjustment will lead to very similar density estimates (Thomas, 2010).

The model chosen uses a hazard-rate detection model with no adjustment. It was generated on data truncated at 400m and with cutpoints at 25m, 50m, 100m, 200m and 300m. The first group falls just on the shoulder of the detection function and shows that there was nearly 100% chance of seeing animals on the transect, many other models predicted more than 100% chance of seeing animals on the transect. The chi-squared test gave a P-Value of 0.65581 which is above the acceptance value of 0.05. Although this model has not handled cluster size bias, a similar model that had cluster size as a covariate gave a density estimate that was well inside the lower confidence level of this model.

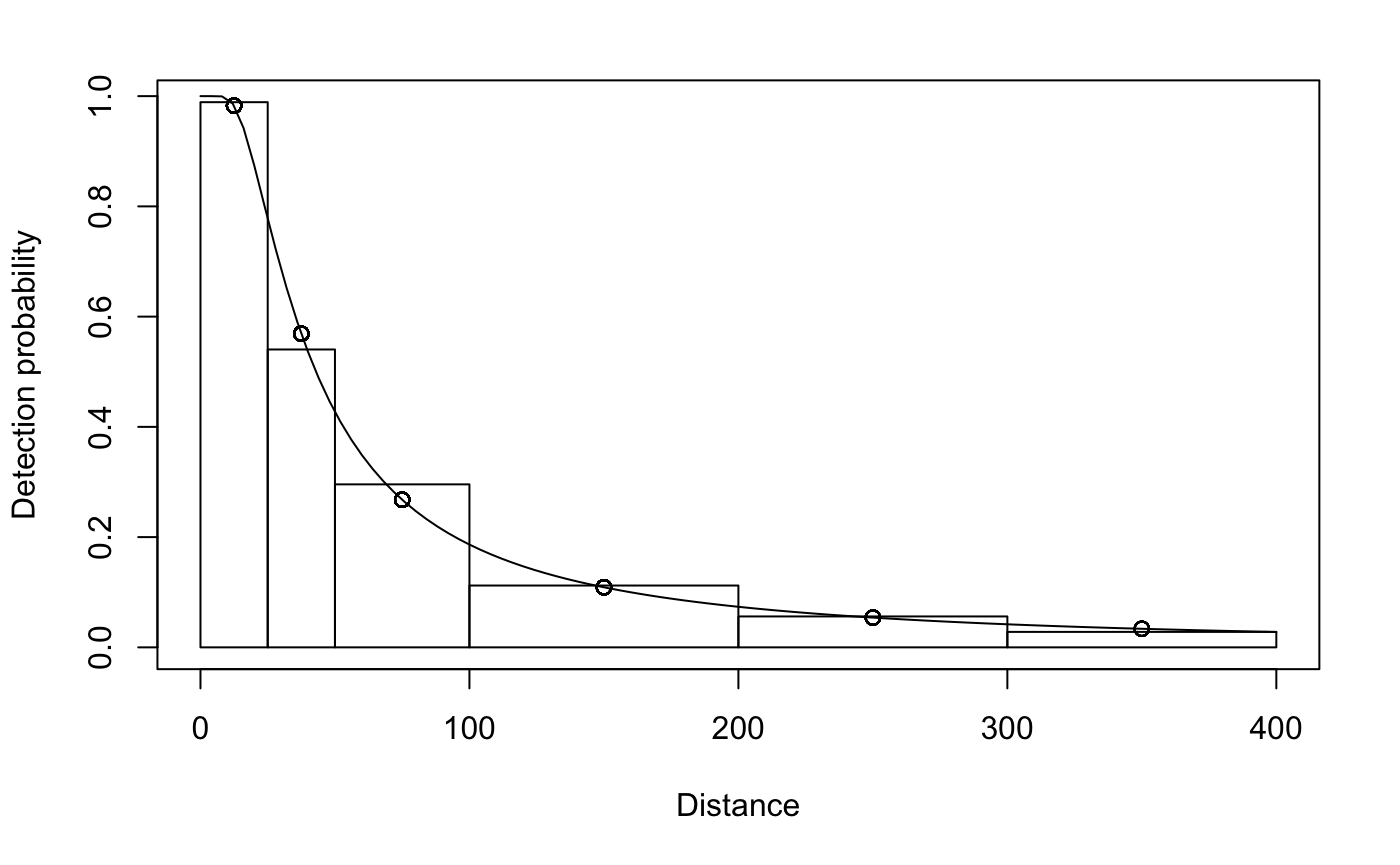


Figure 5 Hazard-rate detection function fitted to grouped data.

1. **Density Estimates**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Density Estimate (per km2)** | **LCL** | **UCL** |
| Bleaklow | 23.67 | 13.07 | 42.88 |
| Margery Hill | 15.51 | 7.5 | 32.08 |
| **Total** | **19.59** | **12.21** | **31.44** |

Table 1 Estimated density of hares at each survey site with lower and upper confidence levels.

**Appendix 1: Models Tested**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model + Adjustment | Size as covariate? | Truncation | Groupings | Angle smearing range | AIC | Chi squared p-value | Cramer von-mises test p-value | Density Estimate | LCL | UCL |
| Uniform + Polynomial | n | n | n | n | 3316.922 |  | 0.0109244 |  |  |  |
| Half-normal + Cosine | n | n | n | n | 3170.053 |  | 0.000207484 |  |  |  |
| Hazard-rate + Hermite | n | n | n | n | 3116.362 |  | 0.000358798 |  |  |  |
| Half-normal + Cosine | y | n | n | n | 3238.901 |  | 0.000294133 |  |  |  |
| Hazard-rate + Hermite | y | n | n | n | 3127.607 |  | 0.00193743 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 200 | n | n | 2527.935 |  | 8.06E-06 |  |  |  |
| Half-normal + Cosine | n | 200 | n | n | 2513.633 |  | 0.00141255 |  |  |  |
| Hazard-rate + Hermite | n | 200 | n | n | 2452.736 |  | 9.31E-06 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 400 | n | n | 3149.395 |  | 0.000527033 |  |  |  |
| Half-normal + Cosine | n | 400 | n | n | 3072.789 |  | 0.0015983 |  |  |  |
| Hazard-rate + Hermite | n | 400 | n | n | 3027.025 |  | 0.000167999 |  |  |  |
| Half-normal + Cosine | y | 400 | n | n | 3134.702 |  | 8.75E-05 |  |  |  |
| Hazard-rate + Hermite | y | 400 | n | n | 3019.606 |  | 6.74E-05 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 200 | n | 20 | 2512.231 |  | 7.76E-05 |  |  |  |
| Half-normal + Cosine | n | 200 | n | 20 | 2498.192 |  | 0.0154844 |  |  |  |
| Hazard-rate + Hermite | n | 200 | n | 20 | 2478.025 |  | 0.0573472 | 30.71705 | 18.78328 | 50.23285 |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 400 | n | 20 | 3138.813 |  | 0.00018442 |  |  |  |
| Half-normal + Cosine | n | 400 | n | 20 | 3066.555 |  | 0.0382431 |  |  |  |
| Hazard-rate + Hermite | n | 400 | n | 20 | 3043.313 |  | 0.101833 | 35.24809 | 21.29131 | 58.35374 |
| Hazard-rate + Hermite | y | 400 | n | 20 | 3041.829 |  | 0.0395731 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 400 | 0,25,50,100,200,300,400 | n | 1015.191 | 0 |  |  |  |  |
| Half-normal + Cosine | n | 400 | 0,25,50,100,200,300,400 | n | 942.62 | 0.00606 |  |  |  |  |
| Hazard-rate + Hermite | n | 400 | 0,25,50,100,200,300,400 | n | 932.8541 | 0.65581 |  | 19.59155 | 12.209739 | 31.43628 |
| Hazard-rate + Hermite | y | 400 | 0,25,50,100,200,300,400 | n | 0.44539 | 0.44539 |  | 17.49231 | 11.20352 | 27.31116 |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 400 | 0,25,50,100,200,300,400 | 20 | 1011.293 | 0 |  |  |  |  |
| Half-normal + Cosine | n | 400 | 0,25,50,100,200,300,400 | 20 | 944.1318 | 0.0038597 |  |  |  |  |
| Hazard-rate + Hermite | n | 400 | 0,25,50,100,200,300,400 | 20 | 932.4087 | 0.92453 |  | 20.30542 | 12.58895 | 32.75173 |
| Hazard-rate + Hermite | y | 400 | 0,25,50,100,200,300,400 | 20 | 930.2722 | 0.75787 |  | 17.76928 | 11.392547 | 27.71524 |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 300 | 0,25,50,90,150,220,300 | n | 954.5297 | 0 |  |  |  |  |
| Half-normal + Cosine | n | 300 | 0,25,50,90,150,220,300 | n | 916.9435 | 0.087036 |  | 17.40578 | 11.159487 | 27.1483 |
| Hazard-rate + Hermite | n | 300 | 0,25,50,90,150,220,300 | n | 910.8166 | 0.83375 |  | 20.09716 | 12.265219 | 32.93018 |
| Hazard-rate + Hermite | y | 300 | 0,25,50,90,150,220,300 | n | 911.0428 | 0.64604 |  | 18.2226 | 11.3641 | 29.22035 |
|  |  |  |  |  |  |  |  |  |  |  |
| Uniform + Polynomial | n | 300 | 0,25,50,90,150,220,300 | 20 | 953.8261 | 0 |  |  |  |  |
| Half-normal + Cosine | n | 300 | 0,25,50,90,150,220,300 | 20 | 909.7893 | 0.090878 |  | 17.67098 | 11.373478 | 27.45543 |
| Hazard-rate + Hermite | n | 300 | 0,25,50,90,150,220,300 | 20 | 903.3648 | 0.92721 |  | 20.95821 | 12.757271 | 34.43106 |
| Hazard-rate + Hermite | y | 300 | 0,25,50,90,150,220,300 | 20 | 904.6242 | 0.7853 |  | 19.39861 | 12.02792 | 31.28604 |
|  |  |  |  |  |  |  |  |  |  |  |
| Half-normal + Cosine | n | 200 | 0,25,50,75,100,150,200 | n | 834.054 | 0.10629 |  | 16.64996 | 10.71608 | 25.86963 |
| Hazard-rate + Hermite | n | 200 | 0,25,50,75,100,150,200 | n | 830.1693 | 0.54693 |  | 19.81103 | 11.904585 | 32.96856 |
|  |  |  |  |  |  |  |  |  |  |  |
| Half-normal + Cosine | n | 200 | 0,25,50,75,100,150,200 | 20 | 838.1544 | 0.0049901 |  | 16.8075 | 10.802045 | 26.15171 |
| Hazard-rate + Hermite | n | 200 | 0,25,50,75,100,150,200 | 20 | 832.3526 | 0.082095 |  | 21.04678 | 12.430237 | 35.63626 |
|  |  |  |  |  |  |  |  |  |  |  |
| Half-normal + Cosine | n | 400 | 0,25,50,90,150,220,300,400 | n | 1013.923 | 0.028326 |  | 16.75545 | 10.740981 | 26.13775 |
| Hazard-rate + Hermite | n | 400 | 0,25,50,90,150,220,300,400 | n | 1005.018 | 0.75468 |  | 19.82216 | 12.29525 | 31.95688 |
| Hazard-rate + Hermite | y | 400 | 0,25,50,90,150,220,300,400 | n | 1003.018 | 0.59158 |  | 17.54047 | 11.209898 | 27.44611 |
| Hazard-rate + Cosine | y | 400 | 0,25,50,90,150,220,300,400 | n | 1003.018 | 0.59158 |  | 17.54047 | 11.209898 | 27.44611 |
| Hazard-rate + Polynomial | y | 400 | 0,25,50,90,150,220,300,400 | n | 1003.018 | 0.59158 |  | 17.54047 | 11.209898 | 27.44611 |

Table 2 A list of the models fitted to different arrangements of the data. Some of the models included the cluster size as a covariate. Models where fitted to either all the data or truncated data. The data was either continuous or grouped. Some of the models were fitted to data that had been manipulated with ‘smeared’ angles. The models are included in blocks above where a set of models is fitted to the same data, the AIC for each model is included. Models fitted to grouped data have the chi-squared p-value reported. Models fitted to continuous data have the Cramer-von-mises p-value reported. Models that have an acceptable p-value also have the total density estimate with the LCL and UCL reported.

**Bibliography**

Bibby, C., Jones M. & Marseden S. (eds.) (1998), *Expedition field techniques, bird surveys.* Royal Geographical Society

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. (2001), *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University Press

Kinzey, D., Gerrodette, T. & Fink, D. (2002), *Accuracy and precision of perpendicular distance measurements in shipboard line-transect sighting surveys*.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R. B., Marques, T. A. & Burnham, K. P. (2010), *Distance software: design and analysis of distance sampling surveys for estimating population size.* Journal of Applied Ecology