

Received Date : 21-Feb-2013
Revised Date : 21-Mar-2013
Accepted Date : 26-Mar-2013
Article type : Application
Editor: Robert B. O'Hara

Corresponding author email id: martin.cox@aad.gov.au

`nupoint`: An R package for density estimation from point transects in the presence of non-uniform animal density

Martin J. Cox¹, David L. Borchers² and Natalie Kelly^{3,1}

¹Australian Antarctic Division, Channel Highway, Kingston, Tasmania 7050,
Australia

²Centre for Research into Environmental and Ecological Modelling, University of St Andrews,
St Andrews, Fife KY16 9LZ, Scotland

³CSIRO Mathematics, Informatics and Statics and Wealth from Oceans National Research
Flagship, Castray Esplanade, Hobart, Tasmania, 7000, Australia.

Summary

1. The R package `nupoint` provides tools for estimating animal density from point transect surveys in which the conventional point transect assumption of uniform animal distribution in the vicinity of the point is violated.
2. It includes tools for plotting, model selection, goodness-of-fit testing, and simulation.
3. This paper describes the main features of the package and illustrates its use by application to two different kinds of survey dataset.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.12058

© 2013 The Author. Methods in Ecology and Evolution © 2013 British Ecological Society

Key-words: Point transect, distance sampling, non-uniform distribution, road survey, shore survey, density gradient

INTRODUCTION

Distance sampling and capture-recapture methods are the two most widely used methods for estimating animal abundance. Powerful general software packages for distance sampling have existed for many years (Thomas et al., 2010) and mark-recapture (White et al., 1999) and this has greatly facilitated their use in addressing ecological and management problems. These packages have been supplemented in recent years by packages implementing new developments in these fields (Efford, 2012, Fiske and Chandler, 2011) but even so, the rate of new method development is outstripping the rate at which the authors of major packages in these areas can update their software. To allow ecologists and managers to take advantage of new developments soon after their advent as possible, there is clearly utility in releasing more specific packages that implement new methods as they are developed. These may or may not subsequently be incorporated into the major packages in these fields, but meanwhile state-of-the art methods are available for ecologists and managers to use.

This paper describes one such package `nupoint`, now available on sourceforge, which implements a variant of distance sampling that was first developed in 2010 (Marques *et al.*, 2010) and has been adapted and extended for a number of different applications since then. We draw together and generalize the original method and subsequent extensions, into a single package. More specifically, the paper deals with point transect sampling in the presence of non-uniform animal distribution that varies with some observable environmental feature.

The background to new the method is as follows: Point transect sampling is one of the two main forms of distance sampling and involves observers located at a number of randomly-located points, searching for animals and recording the distances to detected animals. Random location of points is crucial for conventional point transect methods because estimation requires that animals are independently uniformly distributed in the searched areas around the points - and by distributing points randomly (according to a uniform distribution), the uniformity assumption is met. But it is not always possible to distribute points randomly, and when points are located on, or with reference to, some environmental feature, the assumption of uniform distribution of animals in the searched areas is likely to be violated. Common examples include locating points along paths or roads and locating points on shore when surveying marine fauna. In such cases conventional point transect methods will result in biased estimation of density (Marques et al., 2010).

The R package `nupoint` implements point transect density estimation for situations in which there is, or may be, an animal density gradient in the searched areas. The package accommodates two kinds of density gradient. The simpler of the two is one in which animal density contours are parallel to some linear feature on which points are located. To do deal with this case, the package incorporates the methods of Marques et al. (2010), which were developed to draw inferences from point transect surveys on roads, assuming animal density contours are parallel to the road, and the methods of Cox et al. (2011), which do the same but were developed for point transect surveys on the sea surface (and which we expand upon below). The less simple kinds of density gradient accommodated by the package are those in which animal density contours are not parallel but are aligned with some observable (typically nonlinear) environmental variable. The package incorporates the

methods of Arranz et al. (unpublished data) to deal with this case (we expand on this application below).

The key features of the package are as follows:

1. By using detection angles as well as distances it allows estimation of both a conventional point transect radial distance detection function and (unlike conventional methods) non-uniform animal density with respect to some observable environmental feature within the searched area (such as distance from road or from shore, or altitude, or depth).
2. Key outputs are estimates of density itself, density function parameters, detection function parameters and associated coefficients of variation and confidence intervals.
3. The package also provides some tools for plotting and for model diagnosis, AIC statistics for model selection and goodness-of-fit test statistics.
4. Finally, a facility to investigate estimator properties by simulation is also provided.

We illustrate the package by applying it to a multibeam hydroacoustic survey (Cox et al., 2011), and a subset of the shore-based survey data of beaked whale surfacings Arranz *et al.* (unpublished data). The similarity of the two problems is illustrated in Figure 1.

As well as addressing quite different biological problems, the two surveys differ in the kinds of density gradients involved and the kinds of supplementary data available. Before getting into the analysis of the data, we give a brief intuitive explanation of how the method works, which we hope will aid understanding of the subsequent analyses.

Consider the echosounder survey in Figure 1 and suppose it surveyed a full 180 degrees.

Suppose also that krill density changed with depth – i.e. the contours of animal density were parallel to the surface. And suppose that the echosounder is as good at detecting animals at all angles from vertical. At 90 degrees to vertical (horizontally just under the surface) it is searching along a contour of constant density and so there are on average as many animals at all distances from it. As a result, the drop-off in the number of animals detected as distance from the echosounder increases reflects detectability only (i.e. the shape its detection function). But in a vertical direction animal density changes as distance from the echosounder changes, so the change in the numbers of animals detected with increasing distance reflects a combination of detectability and density.

If the echosounder only searched vertically you would not be able to separate these two effects and hence be unable to estimate animal density. But because it also searches horizontally (in this illustrative scenario) along a contour of constant density, you can separate them. The key result is that by searching a range of angles relative to animal density contours, and recording both angles and distances, the detection gradient and density gradient can be separated. (In reality you don't just search at two angles, but the same idea applies no matter what range of angles you search and the mathematics of the method takes care of the details of how density and detectability are separated – see Marques *et al*, 2010 and Cox *et al*, 2011 for mathematical details.)

ESTIMATION WHEN DENSITY CONTOURS ARE PARALLEL

A multibeam echosounder surveys a wedge below the vessel to which it is attached. By collapsing the wedge along the vessel's path onto a vertical plane, the survey can be viewed

as a point transect survey with the observer on the sea surface, facing downwards (see Figure 1). As it is known that krill density varies with depth, we expect krill density contours to be parallel to the sea surface and the density model is a function of depth alone.

The key data for analysis of this survey are the pairs of coordinates giving the location of each detected krill swarm, and details of the area searched (the pie slice illustrated in Figure 1). And the key model parameters to be estimated are the parameters of the krill density model (as a function of depth, i.e. in the y-dimension). The detection function parameters are nuisance parameters that have to be estimated in order to estimate the density model parameters.

The model is fitted by calling function `nupoint.fit`. Arguments `det.type` and `grad.type` specify the detection function and animal density function forms. The detection function determines how detection probability changes with distance from the observer, while the animal density function is a probability density function that determines how animal density changes with distance from the linear feature on which points are located (or, when animal density varies with respect to some environmental variable as described below, how it changes as that environmental variable changes). In the second line, `pars` contains starting values for their parameters and bounds for them (`lower.b` and `upper.b`). The locations of detections are contained in `sight.x` and `sight.y`. The last line of the call below specifies the maximum range (`w`) and maximum absolute value of angle (`theta.max`) searched (where an angle of zero is perpendicular to the x-coordinate axis).

```
mbe.fit <- nupoint.fit.(det.type=norm.det.type, grad.type=norm.grad.type,
  pars=norm.pars, lower.b=norm.LB, upper.b=norm.UB,
  sight.x=krill$x, sight.y=krill$y,
  w=100, theta.max=pi/3)
```

There is a variety of options for the form of the detection function and the animal density function and as the model is fitted by maximum likelihood, the most appropriate model can be chosen using AIC statistics. Table 1 shows AIC values for a range of density models available in the `nupoint` library (using a half-normal detection function in each case), from which it is apparent that the Normal and Beta forms have similar support from the krill survey data (the former slightly more than the latter) and the others do not.

The `nupoint` library provides a function `nupoint.boot` for nonparametric bootstrap estimation of variance and confidence intervals.

Goodness of fit of a model can be evaluated by calling the `nupoint.gof` function. This function calculates the observed and expected detection frequencies in each of a number of intervals in the dimension perpendicular to the linear feature on which points are located, and uses these to conduct a standard Chi-squared goodness of fit test. Doing this with our chosen model (that with the Normal density model) gives a Chi-squared goodness-of-fit statistic of 11.07 with 7 degrees of freedom and p-value of 0.14. The code for doing this is shown below.

```
norm.gof=nupoint.gof(y.obs=krill$y, pars=norm.fit$par, w=100,
  grad.type='NORM', det.type='HNORM',
  n=NULL,intervals=11, theta.max=pi/3)
```

The only argument not previously described is `intervals`, which either specifies the number of depth intervals to use for the test, or is a vector specifying the depth interval break points . The function call generates the following output (truncated here for brevity):

```
-----
nupoint: 1D Chi-squared Goodness-of-Fit results
-----

  bin.min bin.max  mids expected observed Chisq
1   10.74   18.69 14.71     0.95         2  1.15
2   18.69   26.63 22.66     3.96         4  0.00
.   .....   .....   .....   .....         .   .....

11  90.20   98.15 94.17     8.35         4  2.27
-----

Chi-squ. statistic  = 11.06848
Number of parameters = 3
Chi-squ. df         = 7
Chi-squ. GoF p-value = 0.1356599
-----
```

ESTIMATION WHEN DENSITY VARIES WITH AN ENVIRONMENTAL VARIABLE

We now consider a survey with a less regular survey region, and in which density contours are neither parallel nor straight lines, but vary with ocean depth. The survey involved searching for whale surfacings from a high point on shore, recording the location of all detected surfacings. This survey is quite naturally viewed as a point transect survey (see Figure 1).

Because our density model is to be parameterized as a function of an environmental variable (depth) rather than Cartesian coordinates, and because depth is not a simple

function of Cartesian coordinates, we specify the relationship between depth and location by giving depths on a grid of points spanning the survey area (with NAs in areas that were not searched). For this survey the grid's unique Cartesian coordinates are contained in the objects `sightings$x` and `sightings$y` and the depths are in the matrix `sightings$z.mat` in the code below, which created Figure 2:

```
image.plot(x=sightings$x,y=sightings$y,z=sightings$z.mat,xlab='X',ylab='Y')
points(sightings$obsx,sightings$obsy,cex=3,pch=19)
points(sightings$sighting.mat$x,sightings$sighting.mat$y,col="black",pch=19,cex=0.5
)
```

The first line above creates the depth colour plot. The second puts a black dot at the observer's location, while the third adds dots at the location of every sighting.

In addition to the detection locations (shown in Figure 2), we need in this case to pass the estimation function the depths and derivatives of depth with respect to offshore distance. To understand why derivatives are needed, consider the echosounder example given at the end of the Introduction again. In that case, we could interpret the drop-off in detection frequency when searching at 90 degrees to the vertical (parallel to and just under the sea surface) as a drop off in detectability only because we knew that the rate of change of animal density was zero in this direction (i.e. its derivative was zero). But in the shore-based survey example we don't have any straight-line direction in which the rate of change of density is zero, because density contours are not straight lines. The derivatives of the environmental variable with respect to offshore distance effectively specify the curves along which the rate of change of density is zero, and what the rate of change of the

environmental variable is in other directions. Without this the estimator could not separate the animal density gradient from the detection probability gradient. The model is fitted using the function `nupoint.env.fit`, as follows:

```
env.fit<-nupoint.env.fit (det.type=norm.det.type, grad.type=norm.grad.type,  
  pars=norm.pars, lower.b=norm.LB, upper.b=norm.UB  
  z=sighting.mat$z, rd=sighting.mat$r, dzdy=sighting.mat$dzdy,  
  z.mat=z.mat, dzdy.mat=zGradmat, rd.mat=rd.mat,  
  minz=minz, wz=wz, wx=wx, wy=wy)
```

The first line of this call specifies the detection function form (`det.type`) and animal density form (`grad.type`). In the second line, `pars` contains starting values for their parameters while `lower.b` and `upper.b` specify the bounds for the parameters. The third line gives the depths (`z`), radial distances (`rd`) and derivatives (`dzdy`) of depth with respect to offshore distance (`y`) at the location of each sighting. The fourth line contains matrices specifying the depth (`z.mat`), derivative (`dzdy.mat`) and radial distance from observer (`rd.mat`) of all points on the grid. The last line just sets bounds in the depth, alongshore and offshore dimensions.

We again fitted a variety of animal density models (with a half-normal detection function form). AIC statistics are shown in Table 2. In this case the normal and the mixture of two normals have substantial support, the former about double the support of the latter.

Figure 3 shows the observed and predicted frequency of detections with respect to depth. To conduct a goodness-of-fit test, it is necessary to define depth bands within which the observed and expected number of detections are calculated for the test. These are shown in

the top right plot in Figure 3 as well as along the axis of the bottom right plot in Figure 3.

The associated Chi-squared goodness-of-fit statistic is 9.86 with 9 degrees of freedom and p-value of 0.44. These statistics and the plot shown in Figure 3 were obtained by calling the function `nupoint.env.gof` as follows:

```
gof <- nupoint.env.gof(pars=env.fit$par,  
  grad.type=norm.grad.type, det.type=norm.det.type,  
  r.mat=rd.mat, z.mat=z.mat,  
  z.obs=sighting.mat$z, dzdy.mat=zGradmat  
  minz=minz, wz=wz,  
  intervals=13, plot=TRUE)
```

The only argument not previously described is `plot`, which determines whether or not to do a plot like that in Figure 3.

The `nupoint` library provides a function `boot.nupoint.env` for nonparametric bootstrap estimation of variance and confidence intervals.

It also provides a function `nupoint.env.simulator` for simulating surveys with specified animal density functions and detection functions. Figure 4 shows an example of simulated detections, animal distribution and detection function obtained using this function.

CONCLUSION

The `nupoint` R library makes newly-developed estimation methods for point transect surveys with non-uniform animal distribution accessible to ecologists and other wildlife professionals. We expect that the library will have two main uses:

- Accepted Article
1. It provides a means of estimating density and its dependence on environmental variables from point transect surveys in which points have not been located according to a random survey design. It should be noted that it does not allow robust estimation of density, or the relationship between environmental variables and density, outside of the searched areas. But nor does any other method – there is no robust way to draw inferences beyond the range of explanatory variables the sampled area from non-random surveys.
 2. When points are located randomly along some one-dimensional feature (like the sea surface when searching underwater, a coastline or a road), it provides a means of estimating density and how density depends on environmental variables in the vicinity of the feature, as far out as the detection function reaches.

ACKNOWLEDGEMENTS

The beaked whale data presented in this package were collected by the University of La Laguna (ULL) with funding from 2003 to 2010, of the Woods Hole Oceanographic Institution (WHOI). In 2004, the studies were co-funded by the Department of the Environment of the Canary Islands, through an agreement made with the Ministry of Defence and in 2010, by the Department of the Environment of the Government of Spain, through the national project CETOBAPH. These data can not be used for analysis, databases or to be published in scientific journals, without the prior consent of the University of La Laguna (contactsnaguilar@ull.es, arranz@ull.es). Multi-beam instrument was loaned by J. Condiotty of Simrad USA. Multi-beam data were collected in association with an NSF-funded (grant #06-OPP-33939) investigation of the Livingston Island nearshore

environment. Logistical support was provided by US Antarctic Marine Living Resources Program, and engineering design and support by Sea Technology Services. The multi-beam instrument was deployed using funds provided by the Royal Society. MJC is funded by Australian Research Council grant FS110200057. We are grateful to Francois Aucoin for allowing us to use the functions mle and distr from the FAMle package.

REFERENCES

- Cox, M.J., Borchers, D.L., Demer, D.A., Cutter, G.R. and Brierley, A.S. (2011) Estimating Antarctic krill (*Eupausia superba*) density from multibeam echosounder observations using distance sampling methods. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **60**: 301-316.
- Efford, M. secr: An R Package for Spatially explicit capture-recapture, version 2.3.2. (2012) <http://cran.r-project.org/web/packages/secr/index.html>.
- Fiske, I and Chandler, R. (2011). unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, 43(10), 1-23
- Marques, T.A., Buckland, S.T., Borchers, D.L., Tosh, D. and McDonald, R.A. (2010) Point transect sampling along linear features. *Biometrics*. 66:1247-1255.
- Thomas, L., Buckland, S.T. , Rexstad, E.A. , Laake, J. L. , Strindberg, S. , Hedley, S. L., Bishop, J. R.B. , Marques, T. A. , and Burnham, K. P. (2010) Distance software: design and

analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology* **47**: 5-14.

White, G.C. and Burnham, K. P. (1999) Program MARK: Survival estimation from populations of marked animals. *Bird Study* **46** Supplement: 120-138

TABLES

Table 1: AIC values for a variety of animal density models for the multibeam echosounder data. AIC is Akaike's information criterion; Δ AIC is the model's AIC less the lowest AIC among the models.

Density model	AIC	Δ AIC	AIC weight
Normal	-3.02	0.0	0.53
Beta	-2.45	0.6	0.40
2-normal mixture	1.19	4.2	0.06
Lognormal	5.08	8.1	0.01
Uniform	51.74	54.8	0.00

Table 2: AIC values for a variety of animal density models for the whale data. Δ AIC is the model's AIC less the lowest AIC among the models.; "2-normal mixture" is a mixture model comprising two normal distributions.

Density model	AIC	Δ AIC	AIC weight
Normal	3893.2	0.0	0.66
2-normal mixture	3894.5	1.3	0.34
Beta	3903.6	10.4	0.00
Lognormal	3903.8	10.6	0.00

Uniform	3937.5	44.3	0.00
---------	--------	------	------

FIGURE CAPTIONS

Figure 1: Multibeam echosounder survey of krill (left) and shore-based visual survey of whales (right). Cartesian (x,y) as well as polar (r, θ) coordinates are shown. In the left figure, the vessel is moving vertical to the page. The “pie slice” outlined by dashed lines is searched. Krill density changes with depth so that a krill density model with parallel straight-line density contours (horizontal in the figure) is appropriate. Whale density depends on bottom depth, and as depth contours are not parallel straight lines, a more flexible density model is required in this case.

Figure 2: Whale detections (dots) superimposed on bottom depths (colours) from the shore-based survey. Note that, unlike Figure 1, the observer is at the bottom of the figure (the black semicircle), looking “upwards”.

Figure 3: Goodness-of-fit plots. The top left plots shows depth, the top right shows the depth bands that define the intervals within which each observed and expected value was calculated for the goodness-of-fit test. The bottom left shows predicted density of detections and the bottom right the predicted (black) and observed (red) detection frequencies in each depth band.

Figure 4: A simulated detection function (colours in top plot), simulated animal density (colours in bottom plot) simulated animals (black dots and white dots) and simulated detections (white dots) obtained from the function `nupoint.env.simulator`.





