methods

Rob van Bemmelen

2023-11-09

## Distance sampling analysis

*Distance sampling* is a statistical technique to account for the lower detection of birds that are at greater distance from the observer or, in the case of ship-based and aerial surveys of seabirds at sea, the transect line (Buckland et al. 1993, 2001). Therefore, observers are required to note the distance of (each flock of) birds from the transect line. The expected pattern is that there are less detections of birds at greater distance from the transect line. This relationship can be described by a curve, the *detection curve*. To account for the non-detected birds in calculating bird densities, either the count or the surveyed area can be adjusted. In the latter case, the *effective strip width* (ESW) needs to be calculated from the detection curve. Detection curves (and the resulting ESWs) were estimated for each species using the *Distance* package (Miller et al. 2019, Miller and Clark-Wolf 2023) in R version 4.3.1 (2023-06-16) (R Core Team 2023).

Distances from the transect line were usually recorded in distance bins. Whereas the ESAS protocols prescribe the use of four bands between 0 and 300m from the transect line (with boundaries at 50m, 100m and 200m), several survey programs used different boundaries of the distance bins. Therefore, separate models were fitted for each combination of distance bin boundaries.

Per data set, several models were constructed and compared based on the Akaike Information Criterion (AIC). Models were fitted with a half-normal or hazard-rate base function, and either without covariates but with a cosinus adjustent term, or with no adjustment term and seastate as a covariate. The model with the lowest AIC value was selected as the final model, from which ESWs were calculated and used to calculate the effectively surveyed area for all segments in the original survey data used for the spatial modelling.

## Survey data preparation

Survey data was prepared for spatial modelling in several steps. The appropriate selection of data was made. For each species, only survey campaigns were included were the species was included in the target taxa (for various reasons, species are occasionally excluded). For some species, identification is not always straightforward and when not identified to species level, they are registered as a species group. For example, Razorbill and Common Guillemot are similar and are often registered as being one of the two. The same applies to large gulls. The percentage of individuals not identified to species level can be substantial. Excluding unidentified birds would therefore lead to underestimation of the focal species. Therefore, unidentified birds were devided according to their relative abundance among identified individuals of that species group recorded on that same date in the same survey.

Transects were divided in shorter segments, but this differed between survey campaigns and methods. Ship-based surveys with intervals shorter than 5 minutes were resampled to 5-minute intervals. Aerial surveys with short intervals were resampled to 1-minute intervals. The effectively surveyed area per transect segment was calculated as the segment length multiplied by the effective strip width multiplied by the number of sides of the ship or airplane where was counted.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake (1993). Distance sampling: Estimating abundance of biological populations. Boek/rapport, Chapman; Hall, London, UK.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas (2001). Introduction to distance sampling. Boek/rapport, Oxford University Press, Oxford, UK.

Miller, D. L., and T. J. Clark-Wolf (2023). [Distance: Distance sampling detection function and abundance estimation](https://github.com/DistanceDevelopment/Distance/).

Miller, D. L., E. Rexstad, L. Thomas, L. Marshall, and J. L. Laake (2019). [Distance sampling in R](https://doi.org/10.18637/jss.v089.i01). Journal of Statistical Software 89:1–28.

R Core Team (2023). [R: A language and environment for statistical computing](https://www.R-project.org/). R Foundation for Statistical Computing, Vienna, Austria.