## Appendix III – Approaches to data analysis of visual sightings data from fixed vantage points

### Introduction

Traditionally, data gathered by developers for marine renewable developments has focused on characterising the abundance and distribution of marine mammal species in and around proposed development sites, with the aim of ultimately quantifying relative or (ideally) absolute density of animals and assessing the effects of these developments (significant or otherwise) on animal populations.

Data for such studies has often been collected from land-based vantage point, although survey work in wave and tidal stream development sites can be challenging for various reasons as discussed in the main text. In fact, very few surveys to date have enabled absolute density of animals to be calculated, and most assessments of impacts have been either semi-quantitative, or just descriptive. Furthermore, difficulties can arise when quantifying the actual biological significance of any recorded effect (if statistically significant) and disentangling this from often very large natural fluctuation in animal abundance or distribution (Maclean et al. 2014). This short report aims to briefly discuss some currently available analytical methods for fixed vantage point data. Anyone attempting to use these methods should have some basic knowledge of R as well as some understanding of the Distance sampling theory and software (references). Further instructions and practical pointers to these packages will be added to this document at later stage after further worked up examples are processed and finalised.

### Collecting data that is fit for purpose

Various bodies have attempted to analyse data collected from fixed vantage point, so that it would suit the requirements laid out by regulators, and answer questions on abundance and density in the development site. However, this requires the data to be collected in such a way that the variables affecting detectability can be assessed. Therefore, it is imperative that field data collection is fit for purpose, and data from fixed visual vantage points needs to incorporate information on animal locations and distances from the observer. This can be achieved via photogrammetric methods or by using a theodolite as discussed earlier in the main text body.

The developer should be aware that data collection needs to fulfil some requirements in order to be able to answer these questions and they should make sure that the data they are collecting meets these. Firstly, there needs to be enough data, spanning large enough time period to have some understanding of natural variability of animal distribution and abundance on the site – preferably several years. Secondly, the observations need to be spatially referenced, which typically means taking theodolite angles or photographs/video of the sightings so that sightings locations and distances to the observer can be calculated. Thirdly there needs to be a record of effort. Note where you looked even if you did not see anything.

### Analysing data

Properly collected sightings data can be used to inform developers about animal presence in the target area, habitat use and associated seasonal, tidal or daily patterns. With robust survey design and appropriate statistical methods the effect of development (the “impact”) can be measured and assessed. Methods exist to assess density of animals in a defined area and the associated detection probability through Distance sampling techniques (Buckland et al. 1993; 2001). This comprises a variety of related methods such as line-transect surveys and point sampling, which are based on a concept that even if not all animals can be detected, the proportion of missed animals can be estimates by collecting information about distances to detection.

Traditional Distance sampling relies on strict assumptions about data which must be met, which is often the problem with vantage point surveys because some of the assumptions of distance sampling methods simply do not hold. In particular, the assumption of uniform distribution of animals in the vicinity of the sampling point is not likely to be met (Oedekoven, 2013). This is due to the fact that observations are typically made from a land-based vantage point, such as cliff edge, and it is clear to see that the density of cetaceans might increase with the distance from the shore (due to physical restrictions of shallow depths near shore or specific habitat preferences within the search area). Therefore the ability of the observer to detect the animals from the vantage point will initially increase with distance from cliff, but after a while it will not be possible for the observer to cover the expanding survey area with equal effort and ability as the near shore area and eventually the detectability will decrease and finally drop off completely. Any data resulting from fixed vantage points or models fitted to such data will inevitably be a mixture of the detection process and the underlying distribution of animals in the area (-Mackenzie et al 2013).

For the developer and the regulator this poses a serious obstacle: is a recorded decrease of animals due to observers not detecting them, or the fact that animals actually are not there?

### Methods to deal with imperfect detection and non-uniform data

The data analyses need to be able to assess the detection probability (i.e. the detection function) of the sightings so that we can understand what proportion of the animals are missed by the observer. For standard distance sampling using line-transect data this is best achieved using double platform (or independent observer) data to assess detections missed by primary observer. This still doesn’t solve the question of non-uniform animal distribution but recent developments have allowed some of these to be relaxed through variety of statistical methods such as mark-recapture distance sampling (MRDS) which allows for imperfect detection on the transect line and spatially explicit capture recapture (SECR) which defines density for specifically defined area (Borchers and Efford, 2008) and blends both Distance sampling and mark recapture methods (Marques et al. 2012). DISTANCE software (Thomas et al. 2010) allows its user to deal with these issues as long as distances and radial angles are accurately collected. Some of these techniques can potentially be adapted to suit fixed vantage point data as well but only if certain assumptions about the data are fulfilled. As of yet only initial results from studies using MRDS and SECR techniques for fixed vantage point data have been released and it can be difficult to find in depth information or instruction on how to use these techniques.

### MRSea R package

One analytical method of dealing with vantage point data is the MRSea R package developed at the Centre for Research into Ecological and Environmental Modelling (CREEM) at the University of St. Andrews (Scott-Hayward et al 2013b). The package enables the examination of animal survey data for assessing changes in abundance and distribution from marine renewable or similar development. It models data using spatial and environmental variables to explain distributions of animals. The difference to various other Generalised Additive Modelling (GAM) and Generalised Additive Mixed Modelling (GAMM) packages is that instead of using smoothing functions uniformly distributed over the survey area, MRSea allows these to vary spatially within the survey area using the Spatially Adaptive Local Smoothing Algorithm (SALSA) of Walker et al.

The package can be used to fit detection function to distance sampling data, taking data corrected for imperfect detection (for example by using programme Distance) which has been segmented to I order to allocate covariate values for modelling across the spatial scale. In addition it also allows the modelling of non-distance sampling data, such as vantage point data where no attempt has been made to correct data for animals missed for imperfect detection, although note that no detection function can be fitted if distances to animals are not collected ().

The user guide (ref – Scott-Hayward 2013a) uses two examples, data from distance sampling line-transect survey and another dataset from a near shore vantage point survey. Both of these datasets are included in the package download.

### Practical tips for MRSea users

MRSea is downloadable from <http://creem2.st-andrews.ac.uk/software/>. Contact the software developer to ensure you have the latest version of the package. As with all R packages written using a specific version of R, MRSea may have some problems with newest versions and ensure you install all the other packages upon which MRsea depends on which are detailed in the MRsea package help files (Scott-Hayward et al. 2013b), but the software definitely works with the latest R version 3.1.2 as well as the older 3.0.2. There may be workshops available at CREEM to assist with familiarisation with the software, which are well worth attending especially for basic R users.

Although MRsea itself has been updated (current version is 0.1.2.) the User Guide has not, and users may find some inconsistencies especially in the examples provided to describe the coding. Most of these are easy to identify by even a basic user of R. Don’t blindly copy and paste the code but check that databases are correct and use the help file from R to ensure that parameters for each function are valid. The first issue that some users may come across is with the function called “SALSA1D” which has now been superseded by “SALSA1D\_withremoval”, and the user must check that all the input parameters are correct according to the help file. Issues can appear with SALSA1D if you follow the User Guide, as it has not been updated to include a requirement for making a column for “foldid” which is required for the function to run. Do not follow the guide together without the updated help file version 2. The program will notify the user most of the issues, but new (or very basic) users of R might find this little tricky. Some may find it difficult when flicking between the User Guide and the Manual and finding the right instructions from the manual if following the User Guide step-by-step.

Furthermore, the example datasets are made up of the ‘real’ data of animal sightings as well as a “prediction data sets” – which are used to estimate the number of knots used for each smoother amongst other things. This is a little challenging for the basic user, as they have to create a data frame from the predicted values from the initial models; however help files should be able to assist with this.

The examples within the user guide utilise three different datasets (and associated predictions) to illustrate the use of the software; one which has no change following the impact, second with decreased number of animals and the third which displays a redistribution of animals within the area after the impact. These are very useful in depicting the potential uses of the software and how to deal with fixed vantage point detections.

However the user will have to have some basic knowledge of creating and handling spatial data in R, in order to create the prediction data set as well as producing the kinds of graphs shown in the User Guide – as unfortunately the code for creating these is not included. However the code is simple enough and available from most R guides for spatial use and more advanced user can probably pick the correct code up from help files and manuals to packages like ggplot2.

However, this is the first and only R package with detailed enough instructions to the layperson to be able to deal with vantage point data – providing the data is in the correct format of course. No amount of complicated analyses will be able to deal with poorly collected data. Still, good understanding of Distance software and Distance R packages would probably aid in understanding the coding for the package.

### Nupoint R package

Another software package developed to deal with issues of vantage point data analyses is the R package *nupoint* (Cox et al. 2013) which 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. This software enables estimation of density for data where, in addition to radial distances, angles are taken for each detection. These data can then be used to estimate both a detection function and the gradient in animal densities under certain conditions (Oedekoven et al. 2013). Problem with this package is that it is not actively being maintained or developed further.

### Practical tips for nupoint users

The nupoint data is available for download from <http://sourceforge.net/projects/nupoint/>, it cannot be found through CRAN repository. The package seems to only work with R versions older than 3.0. Although Cox et al. 2013 go through the functions of the package and have helpfully published code for their graphs as well; although just following the paper can be slightly tricky for basic R user without further instructions. As with MRSea it is imperative that all dependencies are installed, otherwise the package simply won’t work, and installing nupoint doesn’t automatically download the required packages, which include at least ‘fields’ and ‘nor1mix’. If using R Studio, remember to tick the box ‘install dependencies’ when installing packages. The example data used in Cox et al. 2013 is not included as data in the package, which is a shame, as viewing data makes it much easier to understand the data requirement formats and also allows one to practice the use of the code.

### Spatially explicit capture-recapture models (SECR)

As mentioned above, SECR models were designed to be used for modelling animal capture-recapture data when it’s collected with arrays of detectors, such as traps or hydrophones, in a defined area where so called ‘edge-effects’ can affect density estimation (Borchers and Efford, 2008). Recently SECR models have been used to assess and correct for the drop-off in detectability of animals with distance in a confined survey area (such as fixed vantage points) when animals are not distributed uniformly. Recent trials by SMRU ltd for MCT/Siemens in Skerries in Anglesey have shown reassuring results that SECR models can be used create detection functions for double observer data when distances are collected accurately (Plunkett et al. 2014, unpublished report). However the current secr package in R is intended for surveys where data is generated from multiple traps within confined area and where individuals can be clearly recognised (capture histories). There are very detailed instructions provided within the secr package for doing this, but the package doesn’t lend itself for vantage point survey data very easily without some modifications and recoding the script, which may be beyond the basic R user. One could modify the script (and input data) so that detectors would be the different observers, but the problem of identifying individuals for small coastal cetaceans remains, unless some assumptions can be relaxed.

### Data requirements for dealing with non-uniform data

For all the packages (MRSea, nupoint and SECR) the sightings need to include associated distance information, and in nupoint the data needs to be in the format of radial distance and angle from observer to the sighting. The observed area or transect lines need to be associated with the correct and appropriate environmental information, such as depth or other variable and the data needs to be divided into segments in case of line transect data and grids in case of vantage point data. If only bearing and estimated distance is recorded for vantage point sightings, the data tends to be lumped into distance bins and can mask effects of distance from observer in the dataset. The actual data requirements and formats for each of these packages are listed on the help file or manual of each package.

### R Versions

On Windows, RStudio uses the system's current version of R by default. When R is installed on Windows it writes the version being installed to the Registry as the "current" version of R. This is the version of R which RStudio runs against by default. You can override which version of R is used via General panel of the RStudio Options dialog. This dialog allows you to specify that RStudio should always bind to the default 32 or 64-bit version of R, or to specify a different version altogether. Note that by holding down the Control key during the launch of RStudio you can cause the R version selection dialog to display at startup.

### Conclusions

Advanced analytical methods are being developed to deal with the datasets from fixed vantage point surveys. However they require some basic understanding of the programme R and knowledge if distance sampling techniques. The most accessible R package is the MRSea, developed by CREEM, St Andrews, with extensive help files and a User Guide. However it will not produce detection functions for data without accurate distances, and therefore it is paramount that data is collected from start fit for purpose.

### Acknowledgements

A big thank you to Dr Lindesay Scott-Hayward for her generous assistance in working out the MRSea package.

### References

Borchers, D.L., Efford, M.G. (2008) Spatially explicit maximum likelihood methods for capture-recapture studies. Biometrics 64:377-385

Buckland, S.T., Anderson, D.R., Burnham, K.P. and Laake, J.L. 1993. Distance Sampling: Estimating Abundance of Biological Populations. Chapman and Hall, London. 446pp.

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, July 2001

Cox,M.J., Borchers, D.L. and Kelly, N. (2013) nupoint: An R package for density estimation from point transects in the presence of nonuniform animal density Methods in Ecology and Evolution 4:589-594

Len Thomas, Stephen T. Buckland, Eric A. Rexstad, Jeff L. Laake, Samantha Strindberg, Sharon L. Hedley, Jon R.B. Bishop, Tiago A. Marques, Kenneth P. Burnham. Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology, February, 2010. 10.1111/j.1365-2664.2009.01737.x

Maclean,I.M.D., Inger,R., Benson,D., Booth,C.G., Embling,C.B., Grecian,W.J., Heymans,J.J., Kate Plummer, Michael Shackshaft, Carol Sparling, Ben Wilson, Lucy J. Wright, Gareth Bradbury, Nadja Christen, Brendan John Godley, Angus Jackson, Aly McCluskie, Rachel Nichols-Lee and Stuart Bearhop (2014) Resolving issues with environmental impact assessment of marine renewable energy installations. Marine Affairs and Policy (in press).

Mann, J., Connor, R.C., Tyack, P. & Whitehead, H. (2000). *Cetacean Societies: Field Studies of Dolphins and Whales*. University of Chicago Press, Chicago and London.

Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti,D.J, Harris, D. and Tyack, P.L., (2012) Estimating animal population density using passive acoustics. *Biol. Rev.* (2012), pp. 000–000. doi: 10.1111/brv.12001

Oedekoven, C.S., M.L.Mackenzie, L.A.S. Scott-Hayward, E. Rexstad (2013). Statistical Modelling of Seabird and Cetacean data: Literature Review.University of St. Andrews contract for Marine Scotland; SB9, (CR/2012/05).

Plunkett, R., Sparling, C. and Kidney, D. (2014) Anglesey Skerries marine mammal monitoring: land-based vantage point survey trial. Unpublished Report. SMRUM-MCT-2014-012. Confidential.

Scott-Hayward, L.A.S., Oedekoven, C.S., Mackenzie, M.L., Walker, C.G. and Rexstad E. (2013a). User Guide for the MRSea Package: Statistical Modelling of bird and cetacean distributions in offshore renewables development areas. University of St. Andrews contract for Marine Scotland; SB9 (CR/2012/05).

Scott-Hayward LAS, Oedekoven CS, Mackenzie ML, Walker, CG and Rexstad E (2013b). MRSea package (version 0.1.2): Statistical Modelling of bird and cetacean distributions in offshore renewables development areas''. University of St. Andrews: Contract with Marine Scotland: SB9 (CR/2012/05)