

## **Modelling impact assessment in renewables development areas using the new R package, MRSea v0.1.1**

**LAS Scott-Hayward<sup>1</sup>, ML Mackenzie, CS Oedekoven**  
Centre for Research into Ecological and  
Environmental Modelling, University of St Andrews,  
Scotland

**CG Walker**  
Department of Engineering  
Science, University of Auckland,  
New Zealand.

### **ABSTRACT**

For both developers and government licensing organisations it is important to have the ability to quantify spatially explicit change in the density and/or distribution of animals in and around marine renewables sites and, in particular, to identify if change occurs near renewables devices [1]. The publicly available MRSea package (Marine Renewable Strategic environmental assessment) [2] has recently been developed for analysing data collected for assessing potential impacts of renewable developments on marine wildlife, although the methods contained in this package have wide applicability.

As a part of work commissioned by Marine Scotland, a number of candidate modelling methods were critically compared and the Complex REgion Spatial Smoother (CReSS) [3] with spatially adaptive knot placement using SALSA [4] was the recommended approach due to its success at locating spatially explicit impact-related change. The CReSS/SALSA approach was coupled with Generalised Estimating Equations (GEEs), which accommodate the spatial and temporal correlation that is generally inherent in baseline monitoring and impact assessment data.

We present the capabilities of MRSea using an example data set from the package, which is based on offshore data collected from an existing renewables development. Specifically, we analyse a scenario where the animals have re-distributed across the study area between two time points, before and after construction of an offshore wind farm. We begin with correcting the observed counts from the survey data for imperfect detection, fit a spatial model with environmental covariates to the corrected counts, assess the fit of the model, run model diagnostics, make predictions and calculate uncertainty about these predictions. Most importantly for these applications, we identify spatially explicit significant differences in animal density before and after the construction.

### **INTRODUCTION**

Previous assessment of the impact of renewable

energy development has focused upon measuring differences in animal abundance prior to and following development. This approach suffered from the disadvantages of a) attributing any potential change to development as the causal agent, b) failing to acknowledge other forces that influence animal abundance and distribution and c) insensitivity to more subtle changes in animal populations, e.g. shifts in animal distribution to areas of habitat quality different than prior to renewable development.

The statistical issues to be addressed in assessing animal population distribution and potential changes to those distributions are subtle and complex. If methods for addressing such questions were straightforward, then methods would be universally in use. However, such methods are at the leading edge of statistical development.

This paper presents a new package MRSea, which was developed specifically to tackle the assessment of potential impacts of renewable developments on marine wildlife, although the methods are applicable to other studies as well. The functions of this package can be used to analyse segmented line transect data and nearshore vantage point data and include model fitting, diagnostic tools and non-parametric bootstrapping to estimate uncertainty.

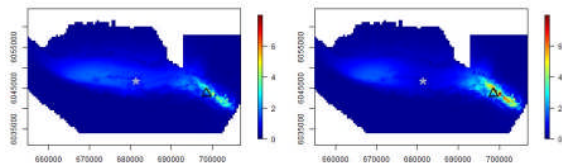
### **DATA**

The data were simulated based on off-shore survey data collected before an impact effect, for example the construction of a wind-farm, hereafter referred to as 'the impact'. The impact effect was then imposed which reduced animal numbers in the impacted area and re-distributed these animals to the south east of the study region (Figure 1).

Observed counts, with imperfect detection imposed, were lifted from this surface in the form of line-transects. This is the data set called `dis.data.re` within the MRSea package.

---

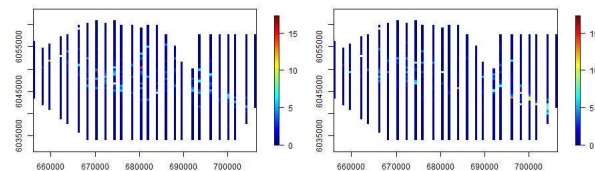
<sup>1</sup> Corresponding author: [lass@st-and.ac.uk](mailto:lass@st-and.ac.uk)



**Figure 1 Simulated densities of birds (per km<sup>2</sup>) before impact (left) and after impact (right). The grey star is the centre point of the impact and the black triangle the centre point of redistribution.**

### CORRECTING FOR IMPERFECT DETECTION

We use distance sampling methods [5] to adjust the observed counts for imperfect detection. Specifically a half-normal detection function was fitted to the raw data. Figure 2 shows the adjusted counts.



**Figure 2 Mean bird counts estimated from a distance sampling analysis before (left) and after (right) the impact.**

### MODEL SPECIFICATION

The data are over dispersed counts and were therefore modelled using a quasi-poisson error structure. The covariates available were depth, season (factor; 1:4), impact (factor; 0/1) and a spatial term (x,y). To allow for a re-distribution of animals a spatial-impact interaction term was also considered. A CReSS-GEE framework was employed to estimate the smooth terms in the model whilst allowing for positive autocorrelation in model residuals (assessed using `runs.test` from the `lawstat` library and `runACF` from `MRSea`)

### MODEL FITTING AND MODEL SELECTION

The `runSALSA1D` function in the `MRSea` package was used to choose the smoothness of the depth relationship. If `runSALSA1D_with_removal` is used then each covariate (specified to be a smooth term) is cycled through and either removed or retained based on k-fold cross-validation. If it is retained in the model it may be linear or smooth (with degrees of freedom between the min and max specified). Note that `runSALSA1D` finds the best smoothness for each covariate but does not consider if the covariate should be linear or removed. The function `runSALSA2D` was employed to determine the smoothness of the spatial term. See the `MRSea` user guide for details on these functions and their use [6].

The models returned from `runSALSA1D` and `runSALSA2D` are of class 'glm' so functions such as `summary`, `update`, `predict` and `fitted` etc are available to the user.

GEE based *p*-values may also be used for model selection. These can be found using the `getPvalues` function and Table 1 shows the results for this model.

**Table 1. GEE based *p*-values for a CReSS model with SALSA knot placement. *p* < 0.05 suggests the covariate should be retained in the model**

Variable	<i>p</i> -value
s(Depth)	<0.0001
as.factor(Season)	<0.0001
s(X,Y)	<0.0001
as.factor(Impact)	0.5468
s(X,Y):as.factor(Impact)	0.0081

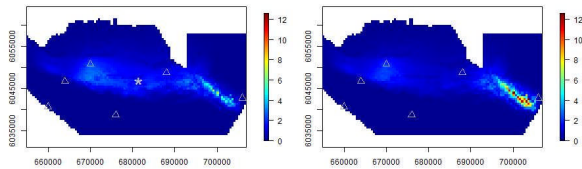
### DIAGNOSTICS

Functions to perform diagnostics in the `MRSea` package include:

- `runPartialPlots`: partial plots for the estimated relationships between each covariate and the response with GEE based confidence intervals (95%).
- `runDiagnostics`: returns a plot of observed vs fitted values to assess model fit (with marginal  $R^2$  and concordance correlation reported in the title) and a plot of fitted values vs scaled Pearson's residuals to assess the mean-variance relationship.
- `plotCumRes`: returns plots of cumulative residuals ordered by covariate value, fitted value and indexed value (temporally) to assess systematic over/under prediction.
- `plotRunsProfile`: returns runs profile plots ordered by covariate value, fitted value and indexed value to assess the correlated nature of residuals.
- `runInfluence`: returns plots of correlated block ID vs PRESS and COVRATIO statistics. These can be used to assess how aspects of the model change when individual blocks are removed from the analysis. This may take some time and the `timeInfluenceCheck` function will establish roughly how long it will take.

### PREDICTION AND INFERENCE

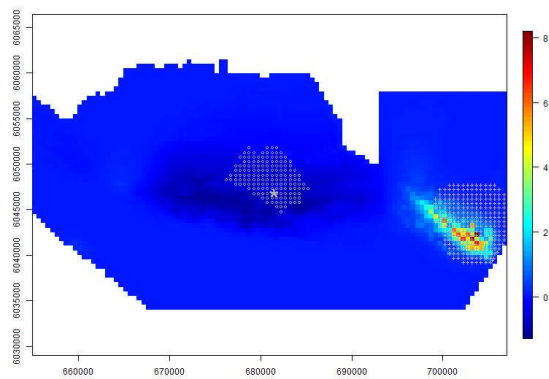
We can make predictions using the `predict` function (Figure 3).



**Figure 3 Predictions of bird density (mean birds/km<sup>2</sup>) from the fitted model for before (left) and after (right) the impact.**

Upper and lower percentile confidence intervals for this predicted surface are found using a bootstrap procedure and combining the uncertainty at the detection function fitting stage and the uncertainty in model parameters at the count model fitting stage. For each bootstrap iteration, the original transect data is resampled with replacement, the detection function and count models are re-fitted, the count model coefficients are sampled from a multivariate Normal distribution and predictions made to the study area using these coefficients. `do.bootstrap.cress` calculates all the new predictions and `makeBootCIs` takes these to calculate the percentile intervals.

In this study we are interested in where in the predicted surface there is a significant change in animal numbers; spatially explicit differences. We can use the `getDifferences` function to find these. This function assesses the before and after predictions for each bootstrap iteration and finds the difference and the 95% interval for the difference. If the interval contains zero, there is no difference before and after. Figure 4 shows the differences for the fitted model with ‘+’ indicating a positive difference and ‘o’ a negative difference.



**Figure 4 Mean differences in predicted bird density (mean birds/km<sup>2</sup>) before and after impact. Positive values indicate more birds post impact and negative values fewer birds post impact. ‘+’ indicates a large positive difference and ‘o’ a large negative one. The grey star is the centre of the impact event.**

In this case there was a large decline in animals around the impact site and an increase in the south east of the study area.

## CONCLUSIONS

This is an example of just some of the functions available in the MRSea package. A full list is available in the reference manual [2] and a full worked example and some additional tips and tricks in the user guide [6]. Both these documents, along with the package can be found at <http://creem2.st-andrews.ac.uk/software.aspx>

## ACKNOWLEDGEMENTS

Thanks to Marine Scotland for funding this research.

## REFERENCES

- [1] Mackenzie, M.L, Scott-Hayward, L.A.S., Oedekoven, C.S., Skov, H., Humphreys, E., and Rexstad E. (2013). Statistical Modelling of Seabird and Cetacean data: Guidance Document. University of St. Andrews contract for Marine Scotland; SB9 (CR/2012/05).
- [2] Scott-Hayward LAS, Oedekoven CS, Mackenzie ML, Walker CG and Rexstad E (2013). MRSea package (version 0.1.1): Statistical Modelling of bird and cetacean distributions in offshore renewables development areas. University of St. Andrews: Contract with Marine Scotland: SB9 (CR/2012/05), URL: <http://creem2.st-and.ac.uk/software.aspx>
- [3] Scott-Hayward, L., Mackenzie, M. L., Donovan, C. R., Walker, C. G., and Ashe, E. (2013). Complex Region Spatial Smoother (CRSS). Journal of Computational and Graphical Statistics.
- [4] Walker, C., Mackenzie, M., Donovan, C., and O’Sullivan, M. (2011). SALSAS - a Spatially Adaptive Local Smoothing Algorithm. Journal of Statistical Computation and Simulation 81, 179-191.
- [5] Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001). Introduction to Distance Sampling. Oxford University Press.
- [6] Scott-Hayward LAS, Oedekoven CS, Mackenzie ML, Walker CG and Rexstad E (2013). “User Guide for the MRSea Package: Statistical Modelling of bird and cetacean distributions in offshore renewables development areas.” University of St Andrews. Contract with Marine Scotland: SB9 (CR/2012/05), URL:<http://creem2.st-and.ac.uk/software.aspx>