Assessing the distribution pattern of African elephant (*Loxodonta africana*) carcasses in Etosha National Park and its implications for management

Lindesay A.S. Scott-Hayward[1] • Monique L. Mackenzie[1] • Cameron G. Walker[2] • Gabriel Shatumbu[3] • J. Werner Kilian[3] • Pierre du Preez[4] • Morgan Hauptfleisch[5, 6] • Claudine Cloete[3]

[1] CREEM, School of Mathematics and Statistics, University of St Andrews, St Andrews, Fife, UK
[2] Department of Engineering Science, University of Auckland, Auckland, New Zealand
[3] Etosha Ecological Institute, Ministry of Environment, Forestry and Tourism, Okaukuejo, Namibia
[4] African Wildlife Conservation Trust, Windhoek, Namibia
[5] Namibia Nature Foundation, Windhoek, Namibia
[6] Unit for Environmental Sciences and Management, North West University, Potchefstroom, South Africa

Correspondence: L.A.S. Scott-Hayward (lass@st-andrews.ac.uk)

Appendix 1: Supplementary Material

1. RAINFALL MODEL

Fit a high dimensional smooth term to 156 locations of annual rainfall from 1999 to 2015 (2016/17 unavailable at the time of modelling) to interpolate values for the presence locations and pseudo-absence grid.

```
require(mgcv)
fit<-gam(meanrain ~ s(x.pos, y.pos,fx = TRUE, k=150), data=rainfall2)
analysisdat$meanrain<-predict(object = fit,
                newdata = data.frame(x.pos = analysisdat$x.pos,
                                     y.pos = analysisdat$y.pos))
```

The results of the interpolation model are shown in Figure 1.

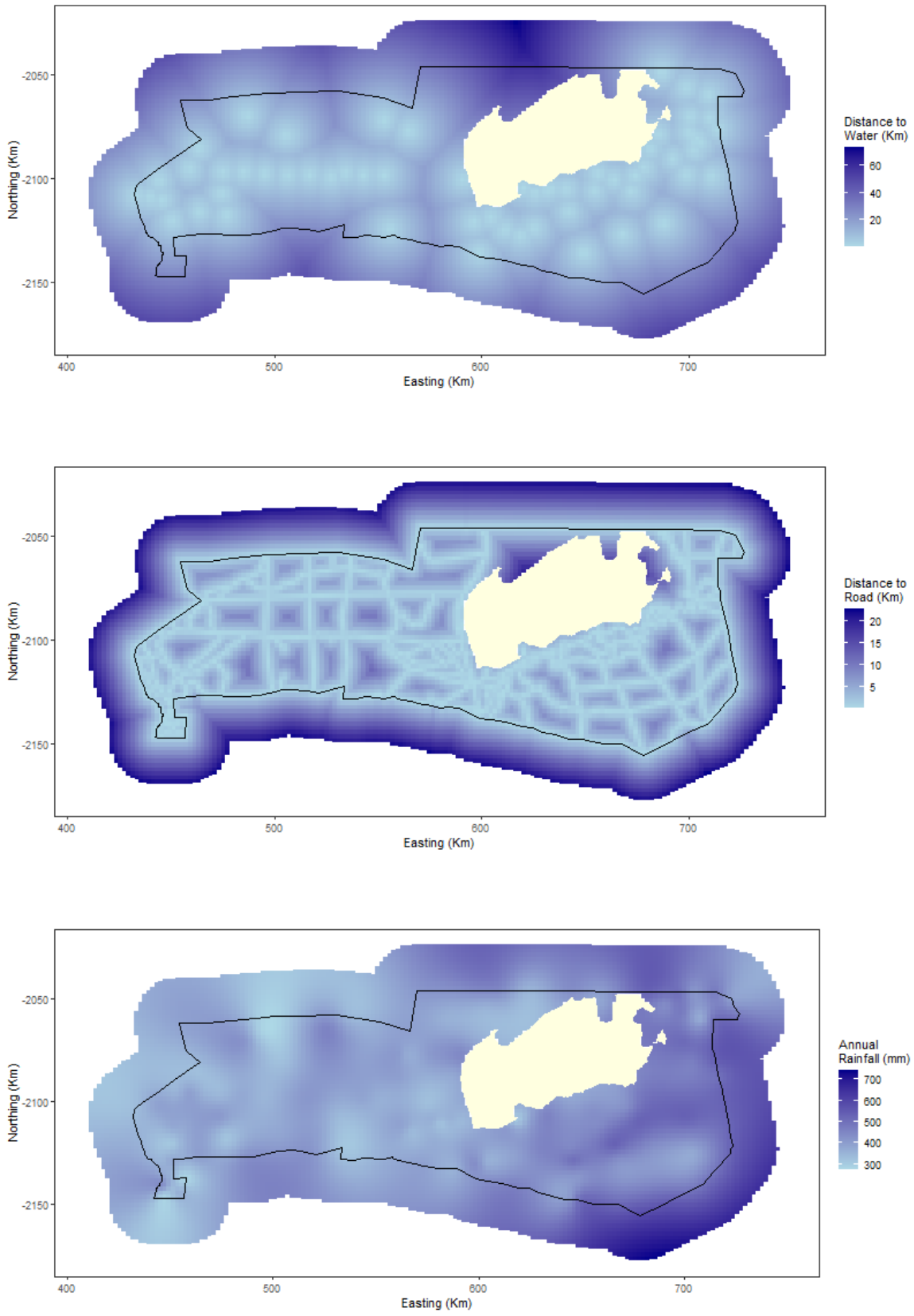

Figure 1 Covariate data in the study region. Distance to water (top), distance to roads (centre) and interpolated annual rainfall (bottom).

2. PSEUDO-ABSENCE SELECTION

- Grid spacings trialled: 5, 4, 3, 2, 1.5, 1.25 and 1 km
- SALSA2D specification
  - knot grid: all non-duplicated presence locations
  - start knot number: 10, 20, 30, 40
  - min knots and max knots equal to start knot number.
  - distance metric: Euclidean
  - basis: Gaussian and exponential
- Fit models for each specification of grid, start knots and basis
- Evaluate the log-likelihood
- Select the coarsest resolution after which, an increase in resolution makes little difference to the likelihood.

Figure 2 shows the log-likelihood scores for the different parameterisations. The vertical dashed line indicates the best grid resolution; $2km^2$.

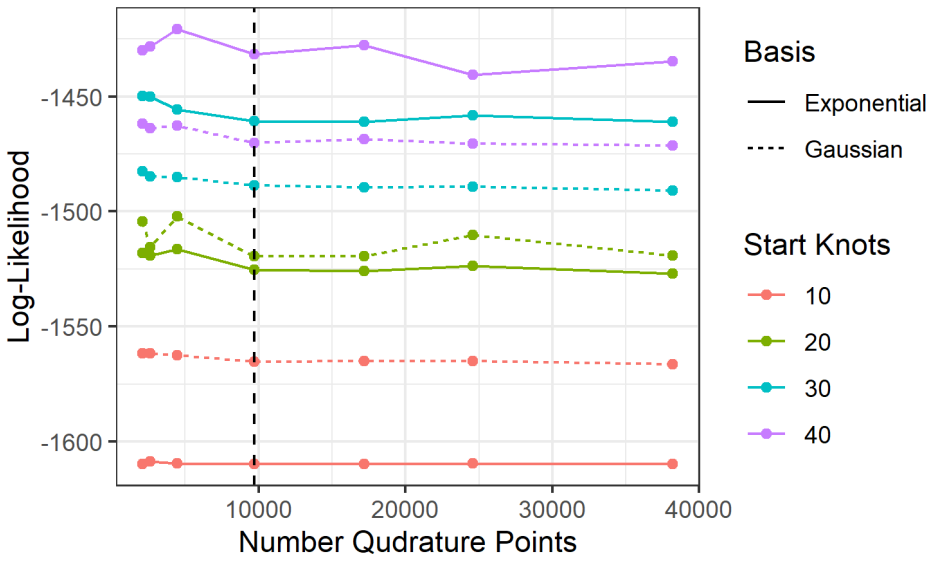

Figure 2 Figure showing the convergence of the log-likelihood for for different spatial resolutions across multiple SALSA2D parameterisations.

3. MODEL AVERAGING VS SALSA2D

3.1 SALSA2D algorithm details

The algorithm that drives SALSA2D has an iterative 3-step structure shown in the pseudo code in Figure 3.

SALSA2D:
Given an n-dimensional set K_l of possible knot locations over the region of interest.
Initialise
Initialise knots, K_s within the points of K_l
Check for convergence
Repeat
Repeat Simplification step *while* ($K > K_{\min}$ and fit measure improves)
Repeat Exchange step *while* ($K < K_{\max}$ and fit measure improves)
Repeat Improvement step *while* (fit measure improves)
While (an improvement in fit measure is made by one of the above steps)

Figure 3 Pseudo-code outlining the structure of SALSA2D (adapted from Figure 1, Walker et al. (2010)), where K is the number of knots used for fitting.

3.1.1 Initialisation

Each observed location, i, is considered a possible location for a knot position. To avoid estimation issues, only unique knot locations are considered giving K_l legal knot locations. The user specifies a starting number of knots, K_s, where $K_s < K_l$, and these are selected from K_l using a space-filling algorithm (Johnson, Moore, and Ylvisaker 1990). This method provides good coverage across the spatial region as a starting position for SALSA2D. Additionally, the minimum number of knots, $K_{\min}$ ($2 \leq K_{\min} < K_s$) and maximum number($K_{\max}$ ($K_s < K_{\max} \leq K_l$) are specified.

To evaluate the basis function, the r_k-value for each basis must also be chosen. The SALSA2D algorithm selects from R possible options for r_k which range from a very local basis to a globally acting basis. The middle option which is neither very local or very global, is chosen to 4nitialize the first model.

To ensure that the initial model fit has converged, there is a drop step component that is activated if the variance of the 4nitialized first model exceeds that of the simpler input model (the variance should not increase with additional parameters/flexibility in the model). If this occurs, knot locations with the largest contributions to the variance are removed one by one until the overall variance of the more complex model is lower than the input model.

3.1.2 The simplify step

Using the fit criteria specified, the simplify step compares the current model with all models obtained by removing an existing knot (as long as this is at least $K_{\min}$). At each iteration, the model with the best fitness measure is retained and the process repeated until there is no further improvement in the fitness measure. This step can be carried out by fixing r_k or by choosing r_k for each basis as each knot is dropped for comparison.

3.1.3 The exchange step

The exchange step increases the extent of the search of model space by enabling a move away from a local minima (of the fit criterion). It uses the maximum Pearson residual from the current fitted model to identify a possible candidate location for a new knot (although in theory other types of residuals could be chosen and we use an alternative metric for the point process models in the next sections). The algorithm then compares the objective fit criteria for these models that result when each of the existing knots in the current model is moved to this new location, and also the fit criteria from the model that results when an additional knot at this location is added to the current model (if this does

not exceed $K_{\max}$). The model with the best fitness measure is retained in this step if it has a better fitness measure than the current model. Evaluation of each of these models can be very quick to return but this process is naturally more computationally expensive, if r_k is also chosen for each basis function for each candidate model. In practice, the algorithm uses the knot locations of the five largest residuals as candidates for an exchange or move.

3.1.4 The improve step

The improve steps allows a more nuanced search of the local minima by allowing small adjustments to the location of each knot. Using the fit criteria specified, the improve step compares the current model with all models obtained by moving an existing knot to one of its five nearest neighbours (determined by the distance metric employed: geodesic or Euclidean). At each iteration, the model with the best fitness measure is retained. As with the exchange step, alternative choices for the (r_k) parameter may be considered when fitting each new model and this process is likely to be swift at this stage.

3.1.5 Determining r_k

This routine considers incrementing or decrementing r_k values in the sequence of R possible values, where the sequence is selected using the method from Scott-Hayward et al. (2014). It can be evaluated either once at the end of the exchange, improve and simplify steps or as part of every decision taken during these steps. The process is done by considering each of the radial basis columns in turn, and incrementing or decrementing the r_k values in the index until there is no improvement in the fitness measure. At each step the r_k-values for the other basis columns are maintained at the current solution. The best of these models is selected as the new current model, and the process iterates until no improvement is made. This process can have a large computational overhead and may significantly prolong the procedure but constitutes a broader search of the model space.

3.2 Model specification

To compare the performance of SALSA2D with model averaging as a model selection approach, models with a two dimensional smoother-based term for geographic locations were fitted to the MIKE data. The comparison involved either the published CreSS method which employs model averaging or model selection using SALSA2D to determine knot number and location.

The CreSS approach fits pure spatial regression models to a set of coordinates $\mathbf{z}_i$ of the form:

$$g(\mathbf{y}_i) = \eta = \beta_0 + s(\mathbf{z}_i) \qquad (1)$$

where g is the link function and η the linear predictor. $\mathbf{s}$ is a two dimensional surface approximated by a linear combination of exponential basis functions bE. The formula for this basis function at observation i and knot location k is:

$$bE_{ki} = \exp^{(-h_{ki}/r_k^2)} \qquad (2)$$

where r_k dictates the extent of the decay of this exponential function with distance between points, and thus the extent of its local nature. Notably h_{ki} indicates a geodesic or Euclidean distance (for some observation i and the k-th knot location). Parameter r_k takes values such that if r_k is small the model will have a set of relatively local basis functions and if r_k is large the model will have a set of relatively global basis functions. The exact values of r_k are dependent upon the range and units of the spatial covariates.

As part of recent work, we have expanded the original CreSS approach to include a Gaussian radial basis to the choice of basis functions available for selection (alongside the existing exponential option). The two bases have different shapes, with the exponential being more peaked at the centre.

These choices allow for more nuanced model fitting, akin to link function or distance metric choice. The Gaussian radial basis, bG, is specified as:

$$bG_{ki} = \exp^{(-(h_{ki}r_k)^2)} \qquad (3)$$

where r_k and h_{ki} are as defined for the exponential (Equation 2) except that for the Gaussian basis, a small value for r_k returns a relatively global basis and a large r_k value returns a relatively local basis.

$$Y_i \overset{indep}{\sim} Poisson\big(\lambda(\mathbf{X}_i)\big)$$

For this method comparison section, we model the intensity as a function of coordinates, **x**, only.

$$\log\big(\lambda(\mathbf{X}_i)\big) = \eta_i = \beta_0 + s(\mathbf{x}) = \mathbf{X}_i^T\boldsymbol{\beta} \qquad (4)$$

where η_i is the linear predictor, consisting of the intercept, β_0, and a smooth function of coordinates, s(**x**). The smooth function is either the exponential or Gaussian basis function.

For both the model averaging and SALSA2D methods, the following specifications were used to return the columns of the design matrix **X** in Equation 4:

- Two basis options: Exponential (bE_{ki}; Equation 2) or Gaussian (bG_{ki}; Equation 3)
- Two distance measures (Euclidean or geodesic) to calculate h in the basis equations; the geodesic distances are calculated using Floyd's algorithm (Floyd 1962) and for more details see Scott-Hayward et al. (2014). In this study, geodesic distances are 'around the salt pan' distances.
- 12 choices of fixed knot number (for the model-averaging approach) and 12 choices of starting knot numbers, K_s for the SALSA2D approach. In each case, the fixed/starting knot set was: {5, 10, 15, ..., 55, 60}. A total of 285 legal knot positions (K_l) were considered. These consisted of all non-duplicated carcass locations (n=245) and 50 space-filled pseudo-absence locations ($\sim 20\%$ of all K_l).
- 10 choices of r_k (also specified as part of Equations 2 & 3)

Additionally, for SALSA2D, $K_{\min}$ and $K_{\max}$ were set to 2 and 100 respectively, for all model specifications.

In keeping with Scott-Hayward *et al.* (2014), the model-averaging CReSS method was governed by AIC_c weights which were used to choose which models to average ($\Delta AIC_c \leq 10$) and their relative contribution to the overall averaged model. In keeping with Walker et al. (2010), the BIC was used to govern SALSA2D model selection regarding the choice of knot number and their locations across the range of combinations of basis type, distance metric, starting knot number and r_k choices (Schwarz 1978). In all cases, the log-likelihood score was calculated for each model to enable comparison between model selection strategies.

The SALSA2D algorithm is implemented inside the MRSea R package (Scott-Hayward *et al.* 2024, R Core Team, 2024) for easy use by practitioners (http://lindesaysh.github.io/MRSea/).

3.3 Finding the largest residual

- Find nearest candidate knot location (of the legal knots remaining and ignoring the already selected knots) to each data point (both presence and pseudo-absence locations). Note that 'nearest' is calculated based on whichever distance metric the model uses. Figure 4 shows the neighbourhood around each knot.
- Sum the observed counts within each knot region

- Make predictions to the pseudo absence grid and sum the estimated intensity within each knot region
- Calculate the absolute residual ($|(O - E)|$)
- Find the 10 knot regions with the largest score. These become the candidates for the exchange/move step.

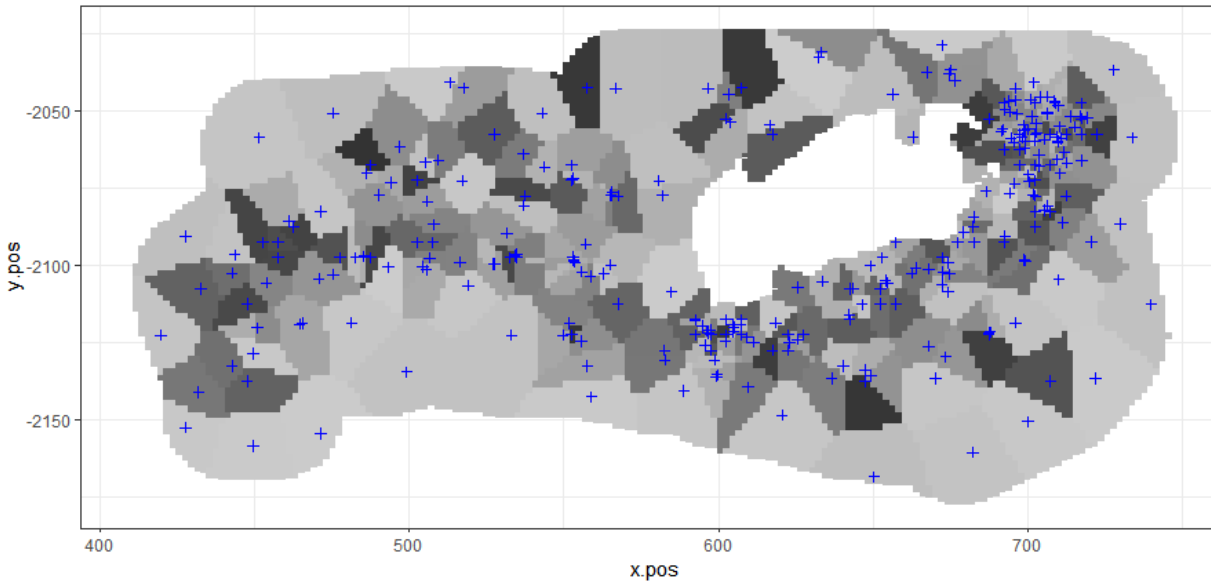

Figure 4 Figure showing the knot locations (blue crosses) and the colour shows the nearest knot on the pseudo absence grid.

3.4 Numerical comparison

The log-likelihood scores returned for the model averaging method were fairly close (maximum difference 14 points) regardless of the basis function and distance metric used in each model (Table 1, Method: `Model averaging'). The geodesic-exponential combination scored the best (largest log-likelihood) of the 4 combinations trialled. Interestingly, this combination chose 11 models with which to average over to obtain this solution, compared with some options that chose far fewer models to use as part of the average calculation. In general, geodesic distances were preferred to Euclidean regardless of basis.

The log-likelihood scores for the SALSA2D based selection are shown for the model with the highest log-likelihood for each of the basis/distance metric combinations (Table 1, Method: SALSA2D). Across the four combinations, the scores were less homogeneous than for the model averaging results and the exponential-Euclidean SALSA2D model (using 41 knots) was the best of all trialled here. In contrast to the averaging approach, there was a preference for the exponential basis with the distance metric secondary. In reality, the user may prefer to select the best model using BIC (as was used for k/r selection). In this case, the order of the four parameterisations was the same (exponential-Euclidean the best and Gaussian-Euclidean the worst) and the best model using BIC was the same as in Table 1 when log-likelihood was used (see Section 3 of Appendix S1 for an expanded version of Table 1).

Table 1 The results of the model averaging and SALSA2D methods of model selection for a given basis type and distance metric used. The `No. Models' indicates the number of models chosen to carry out the model averaging in each case, and the `No. Knots' indicates the number of knots chosen for each model using the SALSA2D selection method. The star indicates the model with the largest log-likelihood (LL) score, and thus the chosen model in each case.

Method	Basis	Distance Measure	No. Models	No. Knots	Log-Likelihood
MA	Exponential*	Geodesic	11	-	-1432.0
	Gaussian	Geodesic	2	-	-1441.5
	Exponential	Euclidean	1	-	-1443.4
	Gaussian	Euclidean	8	-	-1446.3
SALSA2D	Exponential	Geodesic	-	32	-1369.7
	Gaussian	Geodesic	-	32	-1408.3
	Exponential*	Euclidean	-	41	-1301.6
	Gaussian	Euclidean	-	47	-1541.6

Using the 'best' SALSA2D models only, for all but one combination of basis type and distance metric used, all SALSA2D models produced better scores than the model averaging method – sometimes reducing the log-likelihood score by as much as 10%. However, if SALSA2D initialises with too few knots, the algorithm may get stuck in local minima. So long as a large enough number of starting knot locations was selected (~≥ 40), SALSA2D-based selection resulted in superior scores over the model-averaging alternative (Figure 5). This demonstrates that the SALSA2D model selection method can return improved results and at worst, SALSA2D results were almost indistinguishable from the best model averaging-based result.

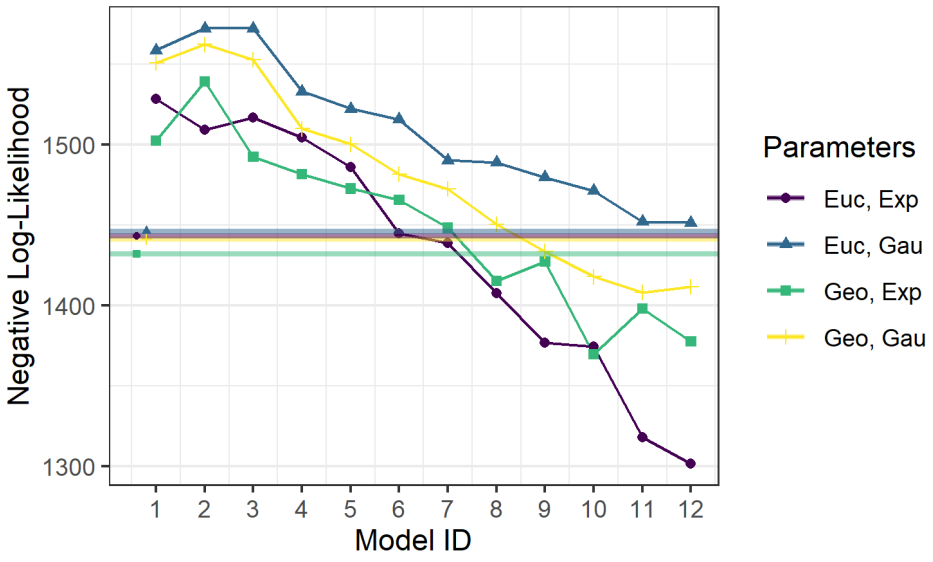

Figure 5 The model identification number (increasing start knots) and the negative log-likelihood score for each of the SALSA2D models resulting from a different start knot number, K_s. The horizontal lines are the scores for the equivalent model averaging result. (Euc - Euclidean, Geo - Geodesic, Exp - Exponential and Gau - Gaussian).

3.5 Visual comparison

Figures 6 & 7 show the best models from the two different methodological frameworks and the four different parameterisations. The best models are selected using BIC for SALSA2D and AIC_c weights for model averaging. Since the two frameworks are compared using the log-likelihood, the figures show the best log-likelihood selected models.

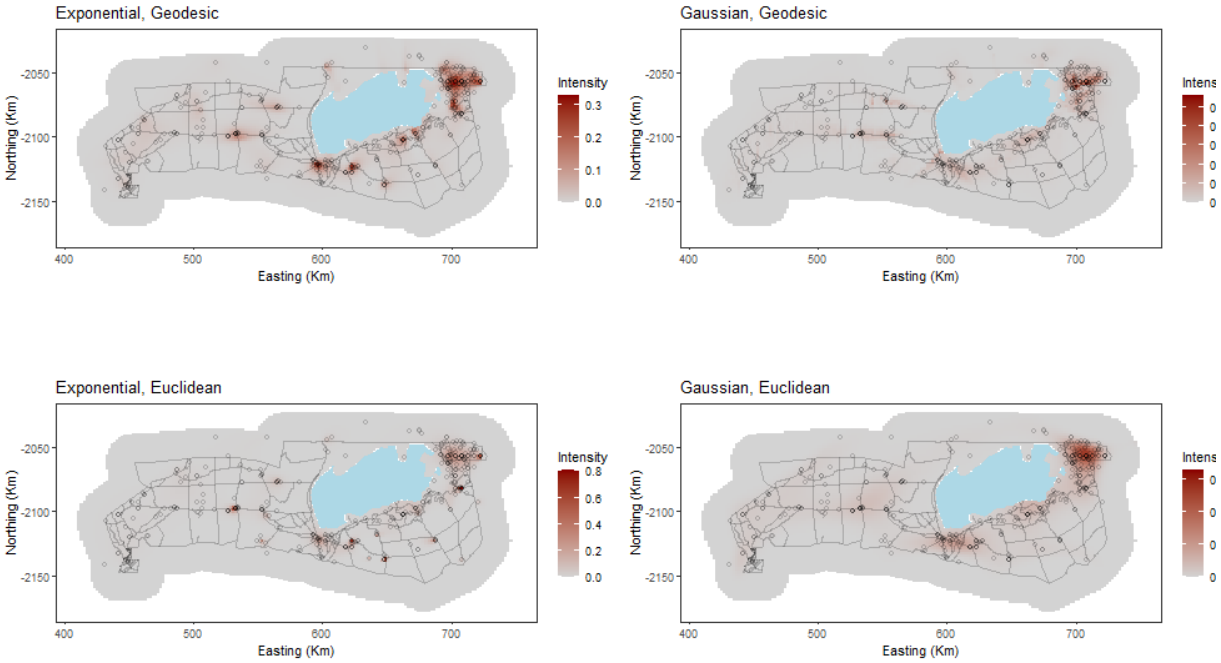

Figure 6 Fitted intensity surfaces for the four SALSA2D models selected using log-likelihood.

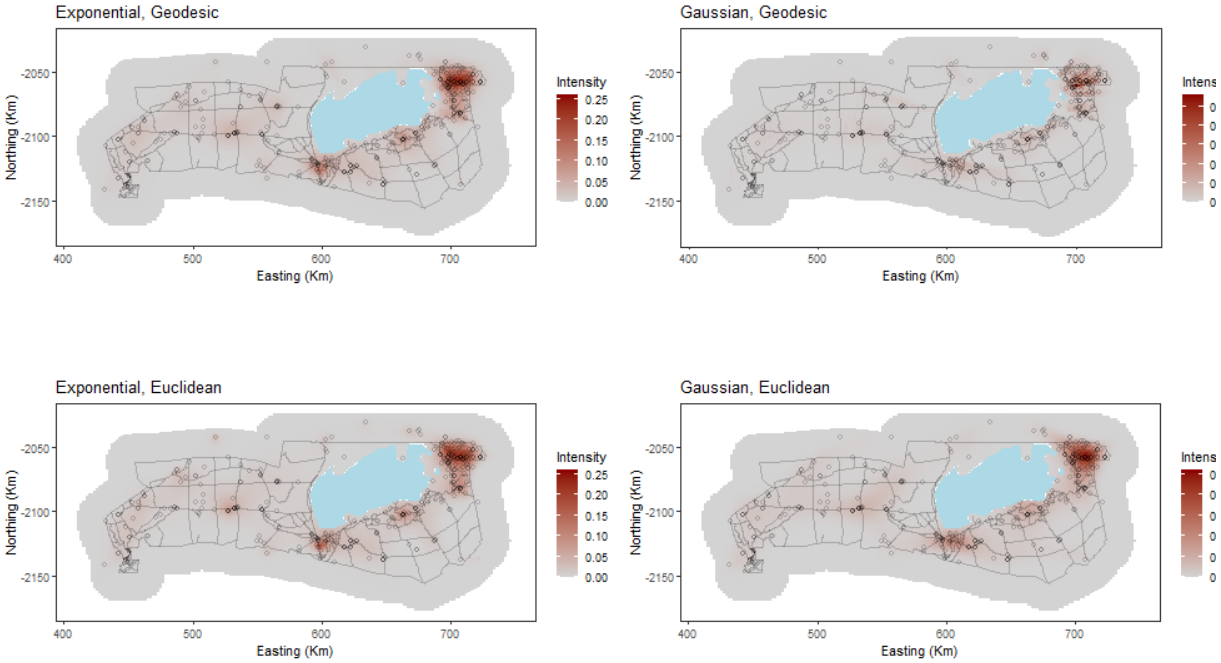

Figure 7 The best model averaged outputs from the four different parametrisations and selected using AIC_c model weights.

Figure 8 shows the selected knot locations and equivalent r parameter from the 11 averaged models (Figure 8 a) and the one best SALSA2D model (Figure 8 b). The averaged knot locations are more difficult to represent but it can be seen that there are multiple r values (ranging from global to very local) across the same locations and occasionally a location where the sign of the coefficient changes between models. The SALSA2D result is more nuanced with very few knot locations selected to the west of the park. For the 41 selected locations, a variety of r's were chosen. It is interesting that the SALSA2D approach found the Euclidean distance metric to be best and it is possible that the more local knots chosen under this method negate the need for the geodesic distances by limiting the possible leakage across the pan.

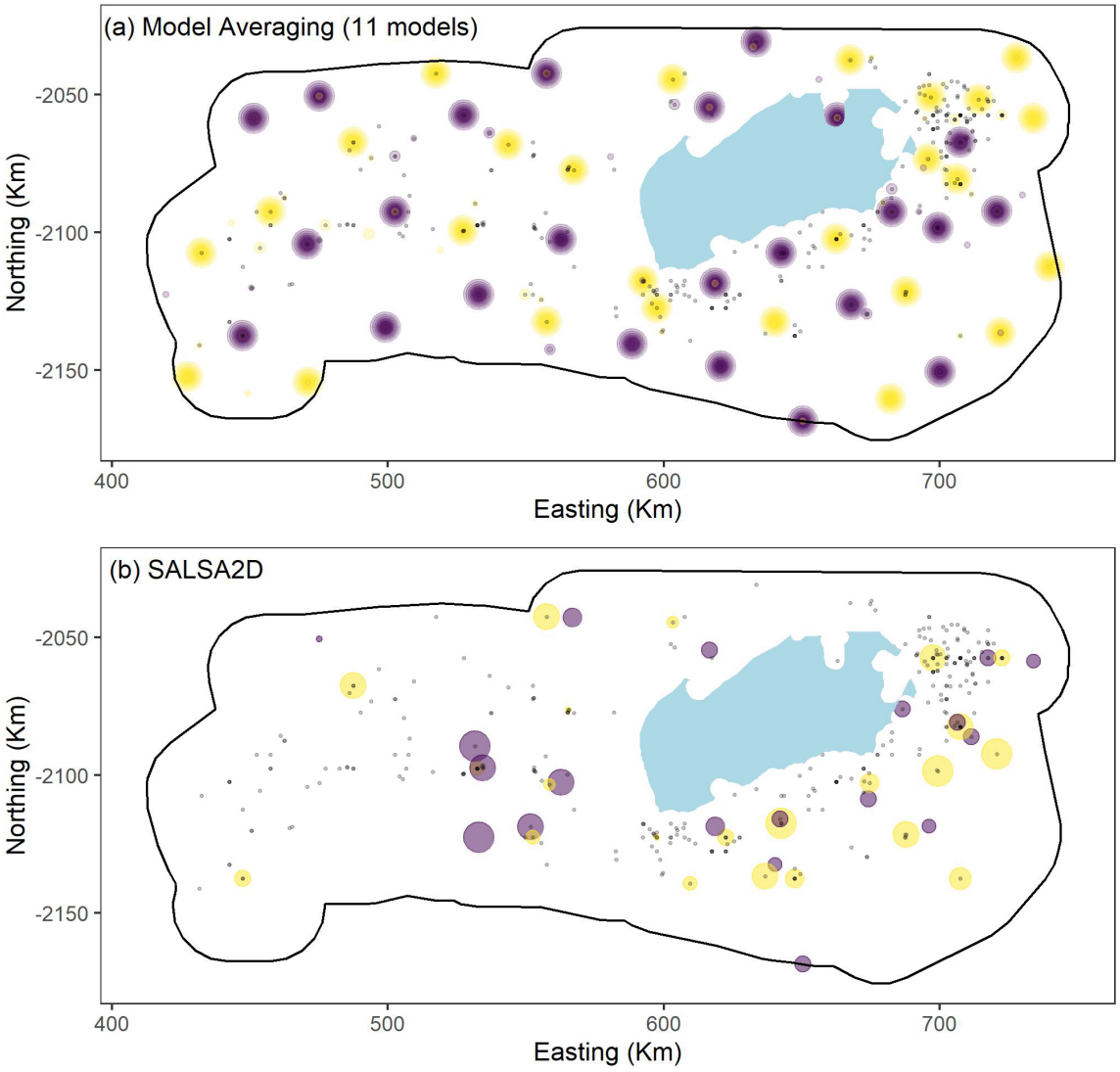

Figure 8 The knot locations and r (effective range of basis function) from the best model averaging (top) and SALSA2D (bottom) models. Yellow is for a positive model coefficient and purple a negative one. The size of the coloured circles is a visual representation of the size of the r parameter. Note that in (a) the concentric rings are from models had the same knot locations with different r. In (b) the colours overlap but each k is in a different location. The carcass locations are shown as grey/black circles. The blue polygon is the Etosha salt pan.

4. REFERENCES

Floyd, R. W. 1962. "Algorithm 97: Shortest Path." *Communications of the ACM* 5: 345.

Johnson, M. E., L. M. Moore, and D. Ylvisaker. 1990. "Minimax and Maximin Distance Designs." *Journal of Statistical Planning and Inference* 26: 131–48.

R Core Team (2024) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6 (2): 461–64.

Scott-Hayward, L. A. S., M. L. Mackenzie, C. R. Donovan, C. G. Walker, and E. Ashe. 2014. "Complex Region Spatial Smoother (CReSS)." *Journal of Computational and Graphical Statistics* 23 (2): 340–60.

Scott-Hayward, L. A. S., M. L. Mackenzie, and C. G. Walker. 2024. "MRSea Package v1.6: Statistical Modelling of Bird and Cetacean Distributions in Offshore Renewables Development Areas." University of St Andrews.

Walker, C. G., M. L. Mackenzie, C. R. Donovan, and M. J. O'Sullivan. 2010. "SALSA - a Spatially Adaptive Local Smoothing Algorithm." *Journal of Statistical Computation and Simulation* 81 (2): 179–91.