Biomathematics and Statistics Scotland

To carry out tern modelling under the Framework Agreement C10-0206-0387

CONTRACT No: C10-0206-0387

Report submitted to:

Joint Nature Conservation Committee

March 2012

Authors:

Mark J Brewer, Jackie M Potts, Elizabeth I Duff,

David A Elston

CONTENTS PAGE

In addition to this report, there are two further documents associated with this project:

(i) BioSS Terns Report – Results Appendix;

(ii) BioSS Terns Report – Software;

and also ancillary files:

(iii) Spreadsheet files of grid predictions for each of the twelve species/colony combinations;

(iv) R code files for: data exploration; model fitting; grid predictions

(v) cleaned and standardised versions of the original data files.

1. Non-Technical Summary

The Joint Nature Conservation Committee (JNCC) is working on the identification of important marine areas around the UK that are used by five species of tern during the breeding season. For the four larger tern species (Arctic, common, roseate and Sandwich terns), data are available from boat surveys, using both visual tracking and transect survey methods.

Following a competitive tendering process, in December 2011 BioSS was tasked with analysis of the visual tracking data for the four larger species of tern.

The analysis was to use the observed tern data in combination with data simulated from the whole of the sample area, with statistical models attempting to relate the locations of terns with background environmental variables. From these statistical models, predictions were to be made on a grid of data and then mapped, showing areas which are preferred by terns.

We found during analysis that the existing statistical methodology was inadequate for analysis, and that a procedure which accounts for the fact that birds are tracked for different lengths of time was needed.

The environmental data contained sets of variables which were very highly related to each other, and so the analyses can only assess associations, rather than identify drivers. We found that for most species, the most important predictor was the distance to the colony itself, but above and beyond that, other variables were also important. For both the Arctic and common terns, the average sea surface temperature in spring proved an important predictor. With only one colony of roseate terns, we could not draw conclusions of consistency across colonies, but chlorophyll may play a part in addition to distance to the colony and sea surface temperature. For Sandwich terns, the clearest predictor was distance to shore, and while other variables seemed important for particular species, there seemed to be nothing consistently so.

2. Introduction

The problem at hand is thus: given the locations of terns, recorded by tracking (as described in Section 3), how can we learn about the preference of the birds in respect to how they select where to forage (based on the explanatory environmental variables we have), and how can this information be used to make predictions about locations frequented by different tern species?

As we are not sure as to the kind (shape) of relationships between tern preference and the environmental variables – that is, whether they will be linear or non-linear – we will need to consider a class of models which allow flexible regression relationships (Generalised Additive Models, GAMs) as well as those which are less flexible (Generalised Linear Models, GLMs).

The first difficulty is in the form of the data. Since birds fly around and since we wish to distinguish between foraging and commuting behaviour, we cannot simply take a random sample of locations and assess whether these locations were visited by terns, which would be a straightforward logistic regression analysis. Instead, the data represent, in essence, "presence-only" data. Ecologists, taking a lead from Boyce and McDonald (1999), who in turn borrowed from the medical statistics literature, started using so-called case-control designs; in medicine, these studies compare individuals having a condition (the "cases") with "control" individuals who do not have the condition. In spatial ecology, this equates to the observed presences being the cases, but unless absence can be determined reliably (as might be possible, for example, if we were studying nesting sites), we lack controls. A partial solution has been to sample from the entire possible range for the population at hand, and to use these as controls; this represents a so-called use-availability design. These controls represent pseudo-absences rather than true absences and because the number of controls is determined arbitrarily, we can only estimate relative preference and not absolute preference.

The second problem is that the data consist of repeated observations for a single bird, and that the number of repeats can vary enormously, depending on the length of the trip each bird made and how long the observers were able to track it. If not accounted for properly, this can bias the results. However, thinning the data set to a single observation per individual is not sensible, given the amount of data loss involved and the fact that we are interested in location-specific covariates, although a certain amount of thinning (equally applied to all individuals) was necessary to make the data set manageable for analysis. Previous work in the area (e.g. Aarts *et al.*, 2008) applied "mixed models" to account for between-individual variation, but does not seem to have fully addressed the problem of serial correlation within a track (see later). The tern data set differs from the one studied by Aarts *et al.* (2008) in that we have only one track per individual, so we cannot distinguish variability between individuals from variability between trips. Thus rather than simply applying existing methodology, there was also to be an element of method development.

Another difficulty relates to the spatial nature of the data. Technically there is a temporal element too, but this only applies within a single track, so we consider the spatial aspect. It is well understood (Beale *et al.*, 2010) that failing to deal with spatial autocorrelation – that is, the fact that data from locations close in space are typically more similar than data further apart – can lead to incorrect statistical inferences; in particular, significant associations can be claimed which are in fact spurious. Modelling spatial autocorrelation can be tricky, especially for the class of binary data we will be dealing with. This ultimately required the use of software which is currently undergoing development in the R statistical package, the software of choice for this project.

The R software (R Development Core Team, 2012) contains numerous packages for fitting the binary response data we have. Our task is to experiment with different models and packages to obtain appropriate functions for fitting the kinds of models required to suit our data. As our background environmental variables are themselves highly cross-correlated, finding important associations can be difficult. Appropriate application of model selection methods is one strategy for reducing multicollinearity, although multicollinearity is less of a problem when the purpose is prediction rather than explanation (Shmueli, 2010). For this reason we will want to compare a range of model selection options.

3. Data

The data received from JNCC consisted of locations where tracked birds were recorded to be foraging; locations where birds were commuting were removed from the data prior to analysis. Foraging observations were thinned to every 10th observation prior to analysis. Environmental covariates provided are shown in Table 1; all except sediment, which is a categorical variable used to derive sand, were considered as potential covariates in the analyses.

Table 1 – Environmental covariates

Parameter	Data set	Source	Date collected	Processing	Original scale and projection	Data type
Chl_month	Chlorophyll-a concentrations, mg/m^{-3}, monthly	PML	2009	Images taken at 1.2km square, re-mapped to 1km square	Approx 1.1km Transverse Mercator	Continuous average concentration values by month
Dist_col	Distance to nearest con-specific colony (m)	Nearest colony identified from JNCC tern colony maps	N/A	Hawths Tools distance between points function in ArcGIS	1km^2 grid cells OSGB 1936 Transverse Mercator	Continuous distance values (metres)
Dist_ shore	Distance to nearest mainland coast (ie shortest distance to coast)	Nearest coastline identified from an Ordinance Survey high water polygon	N/A	Joins and Relates in ArcMap to store distance to closest shore for each point in the environmental layers grid.	1km^2 grid cells OSGB 1936 Transverse Mercator	Continuous distance values (metres)
Sal_spring	Sea surface salinity in spring (‰)	Proudman Oceanographic Laboratory	10 year simulation	Bilinear interpolation, and Inverse distance weighted interpolation to fill in missing values near the coast	0.01^2 decimal degrees GCS WGS 1984	Continuous salinity values derived from simulation of POLCOMS
Sal_summ	Sea surface salinity in summer (‰)	Proudman Oceanographic Laboratory	10 year simulation	Bilinear interpolation, and Inverse distance weighted interpolation to fill in missing values near the coast	0.01^2 decimal degrees GCS WGS 1984	Continuous salinity values derived from simulation of POLCOMS
sst_month	Mean surface temperature by month (ºC)	PML	2006-2010	Images taken at 1.2km square, re-mapped to 1km square	Approx 1.1km Mercator	Continuous average temperature values by month

Strat_temp	Surface to seabed temperature difference in summer (ºC)	Proudman Oceanographic Laboratory	10 year simulation	Bilinear interpolation, and Inverse distance weighted interpolation to fill in missing values near the coast	0.01^2 decimal degrees GCS WGS 1984	Continuous temperature difference values derived from simulation of POLCOMS
Northness_1s	Seabed aspect from -1 (south) to 1 (north)	Derived from Defra digital elevation model data.	NA	Aspect function followed by transformation to radians and trigonometric cosine function, in ArcGIS Spatial Analyst	Approx. $30m^2$ grid cells, varies slightly across extent of data. GCS WGS 1984	Continuous values from -1 to 1.
Eastness_1s	Seabed aspect from -1 (west) to 1 (east)	Derived from Defra digital elevation model data.	NA	Aspect function followed by transformation to radians and trigonometric sine function, in ArcGIS Spatial Analyst	Approx. $30m^2$ grid cells, varies slightly across extent of data. GCS WGS 1984	Continuous values from -1 to 1.
Bathy_1s	Seabed depth (m below lowest astronomical tide)	Defra digital elevation model.	NA	Triangulation with linear interpolation	Approx. $30m^2$ grid cells, varies slightly across extent of data. GCS WGS 1984	Continuous depth values
Sediment_250	Seabed sediment/substrata	British Geological Survey (DigSBS250)	NA	Simplification of DigSBS250 Folk categories supplemented by additional data	Vector dataset GCS WGS 1984	Dominant sediment type categories: mud and sandy mud, sand and muddy sand, mixed sediments, coarse sediments, rock
Sand	Seabed sediment: ‘sandy’ category (1) or ‘other’ category (0)	Based on Sediment_250 with folk triangle.				

Slope_1s_deg	Seabed slope (º incline between adjacent grid cells)	Derived from Defra digital elevation model data.	NA	Slope function in ArcGIS Spatial Analyst	Approx. 30m^2 grid cells, varies slightly across extent of data. GCS WGS 1984	Continuous slope values
Ss_ currents	**Shear stress: currents** Maximum tidal force (Newtons/m^2)	Defra funded Plymouth Marine Laboratory project	June-August from 1998 to 2008.	Inverse distance weighted interpolation, derived from proWAM 12km wave model	0.003^2 decimal degrees GCS WGS 1984	Continuous values
Ss_ waves	**Shear stress: waves** Maximum wave force (Newtons/m^2)	Defra funded Plymouth Marine Laboratory project	June-August from 1998 to 2008.	Inverse distance weighted interpolation, derived from POLCOMS model.	0.003^2 decimal degrees GCS WGS 1984	Continuous values
Spring_ front	Probability of a frequent thermal front in spring. Ratio of strong thermal fronts to observations, averaged over all years.	Defra funded Plymouth Marine Laboratory project	June-August from 1998 to 2008.	Bilinear interpolation	Approx 1.2km^2 GCS WGS 1984	Probability from 0 to 1.
Spring_ frt_sd	Interannual standard deviation of probability of a frequent thermal front	Defra funded Plymouth Marine Laboratory project	June-August from 1998 to 2008.	Bilinear interpolation	Approx 1.2km^2 GCS WGS 1984	
Summ_front	Probability of a frequent thermal front in summer. Ratio of strong thermal fronts to observations, averaged over all years.	Defra funded Plymouth Marine Laboratory project	June-August from 1998 to 2008.	Bilinear interpolation	Approx 1.2km^2 GCS WGS 1984	Probability from 0 to 1.
Summ_front_sd	Interannual standard deviation of probability of a frequent thermal front	Defra funded Plymouth Marine Laboratory project	June-August from 1998 to 2008.	Bilinear interpolation	Approx 1.2km^2 GCS WGS 1984	

4. Methodology

This section provides brief details of and justification for the methodology used during this project.

4.1 Case-Control Design

We use a case-control approach as described in Aarts *et al.* (2008). It should be noted that this is actually a use-availability design (Keating and Cherry, 2004) rather than a case-control design, since the controls represent pseudo-absences rather than true absences. Logistic regression is used to model a response variable which takes the value 1 for the observations and 0 for the control (available environment) points. The exponential function of the linear predictor is then proportional to the expected density of observations (Aarts, 2012). Warton and Shepherd (2010) demonstrate that in the case of pseudo-absences that are regularly spaced or located uniformly at random over the region, the logistic regression slope parameters (but not the intercept) converge to those of the corresponding inhomogeneous Poisson point process model as the number of pseudo-absences increases. The observed data set does not consist of all locations where terns forage, or even all locations where terns are foraging at one specific point in time; instead it is locations where terns were recorded to be foraging. The logistic regression approach models the probability that a point is a presence not a pseudo-absence. This probability has no physical meaning and tends to zero as the number of control points increases; it is the intensity of the presences rather than the probability of occupancy that is of interest and a Poisson point process model therefore has a more natural interpretation. Existing software does not allow a spatial Poisson point process modelling approach to the current problem to be fitted routinely. We therefore take the approach of generating control samples and using logistic regression to approximate the point process model. However, future developments in the INLA (integrated nested Laplace approximation) package (INLA, 2012) should allow the point process model to be fitted directly without the need to generate control samples.

As we wanted to consider a range of modelling options, we proposed generating the control samples so that each set resembled the foraging locations in the original case data (particularly in relation to exhibiting autocorrelation). The simplest way of implementing this seemed to be to generate sets of initial “track starts” – one for each track in the case data – and then applying the rest of the case track movements between foraging locations to each. If a control “track start” fell on land, it was replaced with another random starting location; subsequent points falling on land were simply omitted. This, in effect, creates a set of control samples which are the original case tracks relocated randomly throughout the suggested range of the species – see Figure 1. Having control tracks (rather than just control points as per Aarts et al., 2008) allowed us to think sensibly about using different forms of random effect structure when modelling. However, given the approach eventually chosen did not include random effects, future work could revert back to using control points rather than control tracks; these will provide better spatial coverage per sample point than tracks (therefore reducing the sample size and processing time needed), but will not enable small-scale assessment of spatial correlation.

Control locations (the initial track starts) were generated by taking a random angle (uniformly distributed between 0 and 2π radians) and then a random distance from the colony (also uniformly distributed between zero and the maximum foraging range). This gives a greater density of control points closer to the colony than if points were uniformly distributed over a circle. Aarts *et al.* (2008) recommend that control points should be selected in proportion to accessibility; this would certainly mean a greater density closer to the colony, although it does not necessarily imply a uniform distribution of distance. Selecting points in proportion to accessibility means that the model outputs provide estimates of preference, which is defined as the ratio of use to availability. Estimates of usage can then be obtained by multiplying preference by accessibility. On the other hand, if the control points were uniformly distributed over the circle this would mean that the model outputs would provide direct estimates of usage. The number of control points needs to be sufficiently large to ensure that estimates of slope parameters converge to stable estimates. Exploratory analysis by JNCC suggested that up to five or six control tracks may be required. To be cautious, twelve replicate control tracks were generated for each case track. Further research is needed to determine the number of control points required to achieve stability for different sample sizes as this will not necessarily be a multiple of the number of observations. This is likely to be a complex problem, as the number of control points required could potentially vary hugely between colonies – it will be a function of colony size, colony density, topography of nearby land masses, interactions with other species, availability and distribution of food sources, and so on.

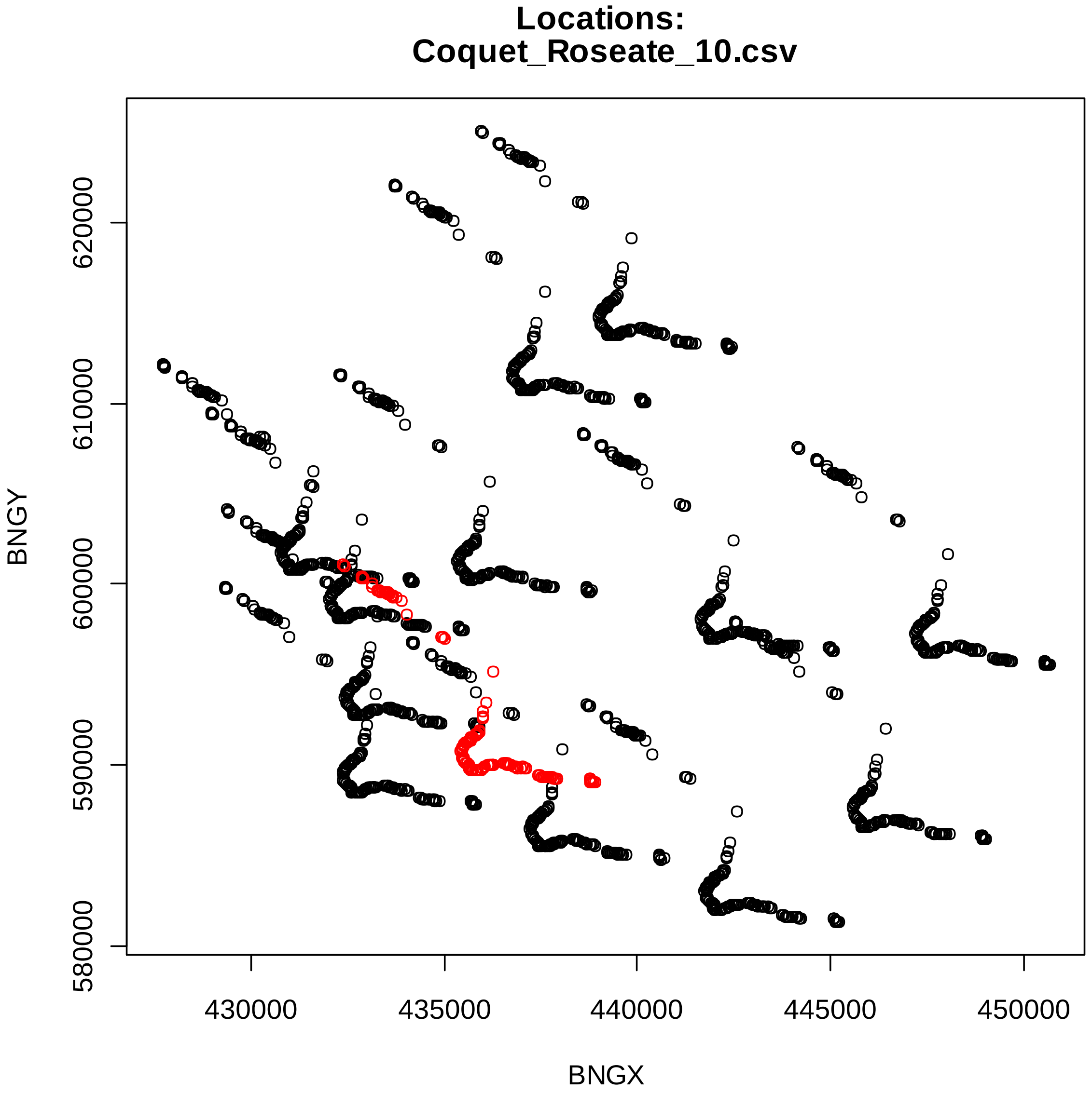


Figure 1 – Example (for roseate tern at Coquet) of the pseudo-absences (in black) generated by shifting the original track (in red)

The environmental information described in Table 1 was attached to each case and control point. To make grids for interpolating the preferences (and usage) to the whole foraging range of each species at each colony, environmental information for points spaced every 500m were provided for the area within the maximum foraging range for each species, around each colony. This was either 30km (Arctic, common and roseate terns) or 60km (Sandwich terns).

4.2 Statistical Model and R Packages

The modelling needs to take account of the fact that points on a given track are not independent (i.e. they are repeated measures), as birds generally move only a short distance from one observation to the next. Failure to account for the lack of independence between observations within track – a form of pseudoreplication – leads to underestimation of the variance of parameter estimates and might therefore result in some environmental variables being wrongly regarded as significant. The simplest way to deal with this is to weight each observation by the reciprocal of the length of the observed track. This has the effect of treating each track as a single sampling unit, instead of treating the individual observation as the sampling unit, but is preferable to averaging data over tracks as individual covariate values are retained for each point. Thus, the basic statistical model used in our analysis is a weighted binomial generalised linear model (GLM) or generalised additive model (GAM) with a logit link. As the data are binary the dispersion parameter should be fixed at 1.

Existing methodological approaches (for example, Wakefield *et al.*, 2011) fit a generalised linear mixed model (GLMM) to account for between-individual variation, although it is unclear exactly how these authors deal with the control points when specifying the random effects. We considered several approaches to the tern analysis using random effects. The standard approach to repeated measures would be to fit a separate random effect for each case or control track. However, this is not appropriate in our situation, because the case tracks consist entirely of presences while the control tracks consist entirely of absences, leading to biased estimates of the fixed effects of the covariates. An alternative approach that we considered, which was probably the one used by Wakefield *et al.*, is to fit a random effect for each combination of a case track with its corresponding control tracks. Unfortunately, this approach is not suitable either, as within each level of the random effect there are units from the same control set that are highly correlated and units from different control sets which are uncorrelated; the model incorrectly assumes that these correlations are equal. For smaller numbers of control sets many of the entries in the correlation matrix are between units from the same track, and as in the previous approach parameter estimates are biased. It is likely that a variation of such a model accounting for spatial autocorrelation would help reduce this bias. As the number of control tracks increases, so entries in the correlation matrix that correspond to units in different control sets start to dominate; this means that the variance of the random effect tends towards zero as the number of control sets increases and the results approach those for an unweighted GLM that does not account for dependence within a track/control set combination. As described above this gives unbiased estimates of the parameters, but estimates of the standard errors are biased (i.e. the parameter estimates falsely precise) because they fail to take account of the dependence within a track. Under-estimation of the standard errors would have been even greater if we had used control points instead of control tracks.

We felt that none of the approaches we explored for specifying the random effects were appropriate, thus we chose to use a fixed effects only weighted model. Weighting avoids biasing the results towards the longer tracks, but would still be necessary even if all the tracks were of the same length because of the autocorrelation between observations in the same track. As there is not time for birds to fly far between successive foraging observations and as the environmental covariates are spatially autocorrelated, observations from the same track tend to occur in similar environmental conditions. This does not necessarily indicate differences in individual preference between birds; multiple tracks per bird would be needed to investigate this.

Our model exploration work led us to three different functions within R, representing different levels of complexity, but all fitting variants of the same basic model:

1. Function glm() in the base “stats” package;
2. Function gam() in the “mgcv” package (Wood, 2011); and
3. Function inla() in the developmental INLA package (INLA, 2012).

All three functions can fit the weighted binary logistic regression required. The first function, glm(), is the GLM workhorse of R, is extremely stable and reliable and has a large suite of generic helper functions. The function gam() in package mgcv provides a simple, flexible interface for fitting smooth functions for the effect covariates, in case some relationships are more complex than straight lines. Finally, we looked for a suitable method to account for residual spatial autocorrelation (i.e. the remaining autocorrelation after the effect of the environmental covariates has been accounted for). Not accounting for this autocorrelation can lead to underestimation of the standard errors for parameter estimates, and this in turn implies some variables may wrongly be declared significant as a consequence; for more detail see Beale *et al.* (2010). The track structure of the data means that a simple rectangular grid would not be appropriate, and the data sets are too large, in general, for modelling large covariance matrices based upon distances between observations (note standard methods consider all pairwise differences). The inla() function in the INLA package solves this problem by defining a “mesh” to represent the spatial autocorrelation structure based upon the locations of the observations; in this way, one observation is linked to a small number of its closest neighbours, and the modelling is intelligent enough to account for these between-neighbour distances. Happily, this method has allowed us to fit full spatial models with whole-species data sets. For an example of the mesh structure produced by the INLA package, see Figure 2.

In summary, the analysis takes three distinct stages; firstly, we fit generalised linear models and perform model selection. The high level of correlation between explanatory variables suggests that we need to be careful about model selection and use different methods depending on whether the aim is prediction or explanation. Because of this concern, we in fact use three different approaches to selection on the basis of: AIC (Akaike’s Information Criterion); BIC (Bayesian Information Criterion); and significance of individual terms via likelihood ratio tests (LRTs). The AIC is known to produce large models, with many explanatory variables which may not be statistically significant, but is optimal in terms of linear models for prediction. The BIC contains a penalty (against adding variables to a model) which is stronger than that for AIC and which is a function of number of observations. BIC is consistent in the sense that if the true model is among the candidate models then the probability of selecting it approaches 1 as the sample size increases, but it is sub-optimal for prediction. As our “number of observations” (which includes the pseudo-absence data) is typically very large, the penalty is itself very large and hence the models selected by BIC tend to be much smaller than those chosen by AIC.

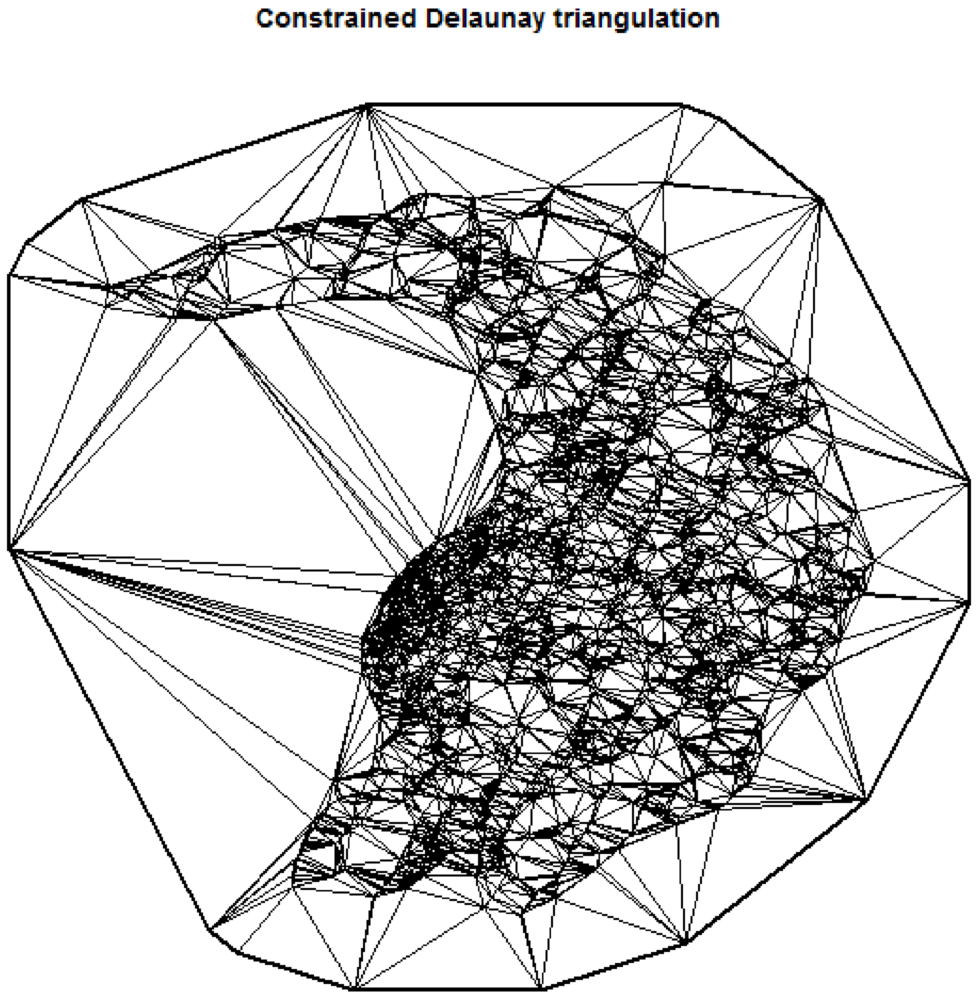


Figure 2 – Mesh to represent spatial correlation structure of response data; example shown for Forvie colony and Sandwich terns

The LRT method is preferred for identification of important associations between tern presence and the covariates, as each term in a selected model must be individually statistically significant; we use a significance level of 10% here as we wish to explore feasible models thoroughly. Note, however, that if two covariates are highly correlated and both are included in the model, neither may appear significant, even though either one of them would be significant if fitted individually. Even in the absence of correlation between the potential covariates association does not necessarily indicate causality as the significant variables may be correlated with a missing, causal, covariate. At this point we also check for consistency of covariate effects across different years, for those species/colony combinations which have more than one year's worth of data; in essence, we fit interaction terms between year and the LRT-selected covariates and conclude effects are not temporally consistent if the interaction is significant.

These generalised linear models will not account for non-linear relationships between response and covariates. For this, we fit generalised additive models (GAMs) which use "spline functions" to describe these response curves – these are non-linear, non-parametric functions which can take any shape, limited only by a number of "basis functions" which control how "wiggly" the line is allowed to be. We chose three basis functions so that the response-covariate relationships would be fairly smooth – we do not believe *a priori* that the relationships should be too complex. For GAM models, as the AIC is only approximate, model selection was instead performed by assessing the significance of smooth terms by comparison to a chi-squared distribution. Next, the INLA method is used to assess the effect of any spatial autocorrelation in the response data, over and above that explained by the environmental covariates in the model. INLA is applied to the model chosen by the GAM model selection; most models chosen across the species/colonies contained some smooth terms, and we wished to have a consistent approach. INLA allows us to account for spatial autocorrelation within the model and to identify those parameters which may have been falsely identified as significant.

Finally, we use the fitted models to produce maps of the estimated preferences and estimated usage, as detailed in Aarts *et al.* (2008). Again, for reasons of consistency we aimed to use the selected GAM model where possible, although variables which the INLA model suggests are no longer significant after allowing for spatial autocorrelation could be removed. The maps of preference show which "habitats" the terns might prefer, to a degree, but note this interpretation is not strictly applicable if either "distance to colony" or "distance to shore" are in the selected model.

If there was no preference for particular habitats, the odds ratio would be equal to the ratio of the number of observations to the number of controls. To calculate preference the odds ratio is adjusted by multiplying by the number of controls per observation, which is equivalent to taking the exponential transformation of the linear predictor and multiplying by the number of controls per observation. This does not purely reflect preferences for environmental variables because of bias in our accessibility model, which assumed a uniform distribution of distance from colony. To correct for this the linear predictor could be broken up into two parts: a component due to distance from colony and a component due to environmental variables alone.

The maps of usage adjust for accessibility, using the same model as that for generating the control sample locations – i.e., a uniform distance distribution from the colony. Equation (5) of Aarts *et al.* is:

$$f_u(X_s) = \frac{h(X_s)}{1-h(X_s)} r f_a(X_s)$$

where $f_u(X_s)$ is the spatial probability density function for usage, $h(X_s)$ is the predicted value from the fitted model, r is the number of controls per observation and $f_a(X_s)$ is the probability density function for accessibility. In our case

$$f_a(X_s) \propto 1/\text{distance}$$

so preference is divided by distance to colony and multiplied by a scale factor which ensures that the probabilities sum to one. These results could then be multiplied by the number of birds in the colony.

4.3 Verification of the Weighted Model

How do we know the weighted regression model is appropriate? This is easily verified by thinking about what the model would look like if there were only a single observation per bird, rather than the tracks that we have currently. There would then be no basis for random effects or weights, and we could fit a simple logistic regression (preferably accounting for spatial autocorrelation in some way – but that is not relevant to the point we're trying to make so we'll ignore it for now). We have tested the "single point" model by sampling randomly a single point from each track, and then running this simple analysis. By generating multiple single sets, we were able to gauge how variable the "single point" results were. Although the results will be quite variable, they provide the perfect yardstick for the "full track" analysis. The weighted logistic model parameter estimates were very similar to the mean estimates from the single point analysis and the standard errors were also similar, as was to be expected, since the weights are effectively accounting for the track lengths; in addition the results were very insensitive to the number of control sets used.

Note that the "single point" analysis represents the maximum amount of thinning that could be applied to the data. In consequence, the single-point analysis described above also (at least partially) resolves the issue of thinning; as long as the data are not thinned too much – where "too much" would imply that possible relationships with covariates are lost, perhaps by obtaining a track on a coarser scale than the covariates – the actual level of thinning is unimportant.

4.4 Other Modelling Approaches / Packages Explored

When the use of a mixed effects modelling approach was investigated initially (before it was found unsuitable), a variety of functions /packages were explored. For fitting a GLMM, both glmer() in package lme4 and inla() again appeared good possibilities. The function glmmadmb in package glmmADMB did not perform well and, in any case, does not appear to allow a weights option. Function glmmPQL in package MASS was not suitable for two reasons: the PQL (penalised quasi-likelihood) fitting process is known to produce biased results; and there does not seem to be a possibility of fixing the dispersion parameter. The gamm() function in mgcv appears to use PQL and therefore suffers from the same problems as glmmPQL.

4.5 Discussion

Our tender suggested exploring a mixed models approach, as this appeared to have been successful in earlier work (e.g. Aarts *et al.*, 2008; Wakefield *et al.*, 2011). In Aarts *et al.* (2008) there are multiple observed tracking instances of the same bird – this would seem to be the difference between their data and the current terns data, and may explain why a mixed model approach may have worked for them and not for us. However, it also appears that these earlier studies may have ignored the fact that whereas observations within a track are serially correlated, there is no correlation between an actual observation and the corresponding control observations. This may have led to standard errors being under-estimated in these studies. We might speculate that a weighted analysis may have been beneficial in these earlier works as well as with our current analysis.

5. Results

This results section contains a summary discussion of the results obtained for the analyses of the twelve data sets. This section is ordered by colony within tern species. The full R output for each model is contained in the "Results Appendix" accompanying this report. The sections in that document mirror those here – for example, section 5.1.1 below summarising the Coquet Arctic tern results is mirrored in section A.1.1 in the Results Appendix document. This should enable straightforward cross-referencing where required.

Note that the models presented are the ones chosen via model selection procedures. Owing to the very large correlations between the environmental covariates, it would be wrong to infer *causal* relationships from the *correlative* relationships discovered; we are finding association rather than proving what is driving habitat selection. Nonetheless, this does not preclude obtaining good predictions. In other words, association rather than causation can provide the information we require to assess which areas are suitable for foraging behaviour.

5.1 Arctic Terns

5.1.1 Coquet

This analysis illustrates the harsher penalty from the BIC for model selection; that produces a model containing only distance to colony, whereas AIC selects distance to colony, sst_may, summ_front_sd, ss_wave and ss_current. LRT selects the same model as AIC, and the effects across the three years' worth of data were consistent.

The GAM model selected also uncovered distance to colony, sst_may, summ_front_sd, ss_wave and ss_current; an illustration of the relationships is shown in Figure A.1.1.4. From this it can be seen that preference increases with increasing values of sst_may and ss_current, and decreases with increasing values of ss_wave and summ_front_sd, although the relationship with sst_may seems to be largely driven by a small number of outliers. Of those four, the relationship with ss_wave seems strongest ($p=0.0125$). The relationship with distance to colony shows that preference decreases steadily as you move further away from the colony (to about 20km) and then the relationship levels off. This is the strongest effect of all ($p<0.001$).

The INLA model suggests that spatial autocorrelation may be causing the suggestion of significance for some of the variables noted above; from the INLA output, we have credible intervals rather than p-values, and we regard a variable as important if the interval does not overlap zero. In fact, only sst_may (0.0994, 3.2528), ss_wave (-0.0804, -0.0129) and the smooth distance to colony relationship remain as important covariates once spatial autocorrelation is taken into account. However, the smooth relationship for distance to colony contains a seemingly unlikely upturn for large distances, likely an artefact caused by a lack of case data at such distances (see Figure A.1.1.5).

5.1.2 Farnes

Here AIC selects distance to colony, sst_april and summ_front_sd, while BIC leaves out sst_april of the three. LRT again agrees with AIC. With only one year's worth of data, there was no need to check for consistency across years. GAM also selected distance to colony, sst_april and summ_front_sd, and as seen in Figure A.1.2.4 the relationships with distance to colony and sst_april are decreasing whereas that with summ_front_sd suggests higher values correspond to higher preference. For this data set, the INLA analysis supports the model selected as none of the credible intervals overlap zero. The covariate relationships were sufficiently linear for INLA to dispense with the GAM element.

5.1.3 Outer Ards

AIC selects distance to colony, distance to shore, chl_apr, chl_jun and sst_may; BIC on the other hand finds only chl_apr and sst_may, ignoring the distance variables completely. The LRT selection adds distance to colony to the BIC selection. There was no evidence of covariate effects being different across years. Interestingly, the GAM model selects distance to colony, chl_apr and sst_april, differing from the LRT selection by switching sst_may and sst_april. Figure A.1.3.4 shows that for all three selected GAM covariates, higher values are associated with lower preference. The spatial INLA analysis provides three credible intervals, none of which overlap zero, so there is no evidence on this occasion that any of the covariate effects found are due solely to spatial correlation. Again, the covariate relationships were linear and so INLA ran a spatial GLM rather than a spatial GAM.

5.2 Common Terns

5.2.1 Coquet

Here AIC selects distance to colony, chl_june, sst_april and ss_wave, but as with the Arctic terns, BIC selects only distance to colony. The LRT model selection matches the AIC choice, and note here there is very slight evidence of the effect of distance to colony varying across years; however, we note that p=0.07443 and hence the evidence is weak; we assume the effects are consistent across years, therefore. The GAM model chosen contains the same covariates as the AIC and LRT selections. Figure A.2.1.4 suggest two linear relationships, with chl_june (increasing) and ss_wave (decreasing), and two non-linear relationships. With distance to colony, the relationship is clearly decreasing, but the rate of decrease increases further away from the colony. The relationship with sst_april is also decreasing, but perhaps levelling out for the upper end of the range of temperatures. The spatial INLA analysis does not contain any intervals overlapping zero, so this reaffirms the significance of the four selected variables. The plot for sst_april in Figure A.2.1.5 is interesting; it suggests that the negative slope is due to the outlying values of temperature at the bottom end of the range, and that in fact the relationship is an increasing one for the bulk of the data in the region around 7 ºC. Further work should consider re-running the model excluding these outliers.

5.2.2 Larne Lough

AIC selects distance to colony, distance to shore, sst_april and bathy_1sec. BIC finds three of these, dropping sst_april. LRT and GAM agree with AIC. There is no evidence of these covariate effects being different across years. The relationship with distance to colony is negative, whereas for distance to shore and sst_april the relationships are positive (see Figure A.2.2.4). The relationship with bathy_1sec is “cup-shape”, having lowest preference in the -150 to -100 metres range, but higher when deeper than 150 metres or shallower than 100 metres. The spatial INLA analysis supports the importance of all the variables except bathy_1sec; note the very wide interval estimates in Figure A.2.2.5.

5.2.3 Leith

For Leith AIC selects a large number of variables: distance to colony, distance to shore, chl_may, chl_june, sst_may, summ_front, spring_front, sal_spring, bathy_1sec, slope_1s_deg. However, the large size of the data set means that the penalty on adding variables with BIC is huge, and therefore only distance to colony is selected. The selection with LRT is large, but not quite as large as for AIC; the variables here are distance to colony, distance to shore, chl_may, chl_june, sst_may, sal_spring, bathy_1sec and slope_1s_deg. With so many variables, it is not surprising that one (sst_may) displays weak evidence ($p=0.0604$) of varying between years, but again we choose to assume a consistent effect across years. (Note that the consequence of not doing so means that we can then only make predictions on a per-year basis, and only for the years for which we have data.)

The GAM model selected matches the LRT model. It can be seen clearly in Figure A.2.3.4 that most of the variables selected have quite small effects on preference relative to the effect of distance to colony, by far the strongest effect ($p<0.001$). The variables chl_may, chl_june, sst_may and sal_spring all have very small positive relationships with preference. The effect of slope_1s_deg is a little greater, although there is uncertainty for high values owing to the relative sparseness of data. Distance to shore shows a moderately negative association with preference. There is a slight non-linear relationship for bathy_1sec, with an increasing relationship for lower depths levelling out for higher values. Distance to colony has a negative relationship which tends to get steeper for higher distances.

The spatial INLA analysis suggests the relationships with chl_may and sal_spring may simply be due to spatial correlation in the response data, as their credible intervals overlap zero, (-0.0007,0.0314) and (-1.6119, 2.0057) respectively.

5.2.4 Mull

For Mull, the AIC selected chl_apr, chl_may, sst_april, sst_may and ss_wave. BIC however selected only chl_apr. LRT was different again, finding chl_apr, chl_may and ss_wave. And GAM was different too – the variables here were chl_apr, chl_may and bathy_1sec. It should be noted here that the Mull colony appears to be surrounded by a lot of land on all sides, with the sea (or lochs) forming narrow strips around the colony. It is not too surprising in particular that distance to colony is not significant – there are areas of water close to the colony which have land in between; it may not have always been possible to continue tracking terns which flew over such land, hence a number of possible areas close to the colony “as the tern flies” may not have been sampled, and thus may have caused bias in the results. We do note that chl_apr and chl_may do occur consistently in chosen models.

Figure A.2.4.4 suggests that all three relationships from the GAM model are non-linear. For both chl_apr and chl_may, preference increases with increasing chlorophyll for low concentrations, then levels off for higher values. For bathy_1sec the relationship is generally negative, levelling out or even increasing slightly for higher values (shallower depths). The spatial INLA analysis broadly supports the GAM model.

5.3 Roseate Terns

5.3.1 Coquet

The AIC chooses a large model here; the variables are distance to colony, distance to shore, chl_may, sst_june, sst_may, summ_front_sd, strat_temp and ss_current. The BIC model is much smaller, having only distance to colony, sst_june and sst_may. The LRT model is in between, being composed of distance to colony, chl_may, chl_apr, sst_june, sst_may and summ_front_sd, and there was no evidence of differential year effects. The GAM model selected matches the LRT model. As shown in Figure A.3.1.4, the effects are all linear, with the strongest being for sst_june and sst_may (both $p<0.001$), although the relationship is positive for sst_may but negative for sst_june. There are weaker positive relationships with chl_apr and summ_front_sd, and weaker negative relationships with distance to colony and chl_may.

For the spatial INLA analysis, the credible interval for chl_apr is (-0.0034,0.6158), which (only just) overlaps zero, casting doubt on the importance of that variable in terms of association with preference.

5.4 Sandwich Terns

5.4.1 Coquet

For Coquet, the AIC chose distance to colony, distance to shore, chl_june and sst_april. BIC only chose the two distance variables. LRT also chose distance to colony, distance to shore, chl_june and sst_april, and spotted a potential cause for concern re: interactions with year – the interaction of distance to shore and year was marginally significant with $p=0.04994$. Again we choose to ignore this interaction, so the maps produced (and presented in Section 7) work on the basis of average effects across all years (produced from the model without interaction term).

The GAM model matches the AIC and LRT models. Figure A.4.1.4 shows the four relationships, which all have different characters. The relationship with chl_june is very small and positive. That between preference and sst_april is non-linear, being flat for lower temperatures but decreasing for higher temperatures. There is a reasonably strong negative relationship with distance to shore, but the strongest effect of all (if not the most significant) is for distance to colony, generally decreasing, but faster with longer distances. Note that it is the high uncertainty with the distance to colony effect – characterised by the wide interval estimates in Figure A.4.1.4 – that cause distance to colony to have a higher p-value than distance to shore, despite having ostensibly a stronger relationship with preference.

With the spatial INLA analysis, the credible interval for chl_june overlaps zero – (-0.0439,0.1972) – which suggests it is not really important, and was only found to be in the GLM/GAM analyses due to spatial autocorrelation in the response data.

5.4.2 Farnes

Here the AIC selects distance to colony, distance to shore, summ_front, spring_front and ss_wave. BIC picks only distance to shore. LRT agrees with AIC, but the GAM model chosen drops distance to colony, despite the four remaining variables displaying linear relationships. As can be seen in Figure A.4.2.4, summ_front and ss_wave have small positive relationships with preference, whereas spring_front has a small negative one. The strongest relationship (and with $p<0.001$) is for distance to shore, very strongly negative, suggesting the birds like staying closer to shore.

There is no evidence in the credible intervals from the spatial INLA analysis that spatial autocorrelation influenced the model selection.

5.4.3 Forvie

For Forvie, the AIC selects distance to shore, sst_june and strat_temp. BIC drops sst_june, but the LRT selection matches that of AIC. The GAM selection agrees with BIC, composed of distance to shore and strat_temp. The effect of both of these variables is strongly negative (see Figure A.4.3.4). The spatial INLA analysis does not cast doubt on the earlier model selection.

5.4.4 Larne Lough

This data set proved very problematic; there were so few combinations of covariates for the case data, it was only feasible to use a very limited set of covariates. In fact, we chose distance to colony and distance to shore, as these were most likely to be important and were at least likely to be measured accurately. Both variables were indeed selected by *all* the different criteria. There was no evidence of the effects varying between years, and the spatial INLA analysis did not provide credible intervals overlapping zero.

From Figure A.4.4.4, it can be seen that the relationship between preference and both distance to colony and distance to shore is negative, and that the effect of distance to shore is stronger (with $p=0.002$).

5.5 Overall Comments

Looking across the colonies for each species, we find a number of consistent features. Note, however, that these are only associations; we cannot state that the selected covariates are causal drivers of foraging behaviour. Of course, it may be that the true drivers are in fact unrecorded.

For Arctic terns, distance to colony is an important predictor, and each analysis unearths one of the sst_ variables. Note that owing to correlations between the covariates, it is not possible to disentangle these effects further. Distance to colony is of course related to accessibility, but it is noticeable that most analyses find significant effects over and above this variable.

The situation for common terns would seem to be similar to that for Arctic terns, in that both distance to colony and an sst_ variable are selected consistently. Mull is an exception here, but as noted earlier the geography surrounding that colony is problematic, in that is has likely caused biases in the data collected.

There is only one colony for roseate terns, which adds chl_ variables to the distance to colony and sst_selections from the previous two species.

For Sandwich terns, the only consistent explanatory variable chosen was distance to shore; the individual colonies mostly produced other significant effects, but these varied quite widely.

6. Discussion

The methodology used presents an advance on earlier methods, with the weighted regression accounting for the differing numbers of observations in each of the recorded tracks. This method is different from that developed by Aarts *et al.* (2008), which does not explicitly use weights and hence implicitly ascribes more weight to tracks with more locations along them. We have made this alteration because the high correlation between locations on the same track makes track a better sampling unit than location

The maps in Section 7 illustrate clearly that the models describe some but not all of the predictability for preference of search/foraging area. With some maps – for example, the common tern for Coquet, the Sandwich tern for Forvie – the red shaded areas (representing higher preference) coincide well with the observed data. For others, however, the observed data lie on a mixture of red and blue areas, suggesting that the given explanatory variables are not sufficient to explain well the preferred foraging areas for those species/colonies; examples here include the common tern for Larne Lough or the Arctic tern for Outer Ards. However, note that we would expect to see some observations in the low preference areas as they are spatially more extensive.

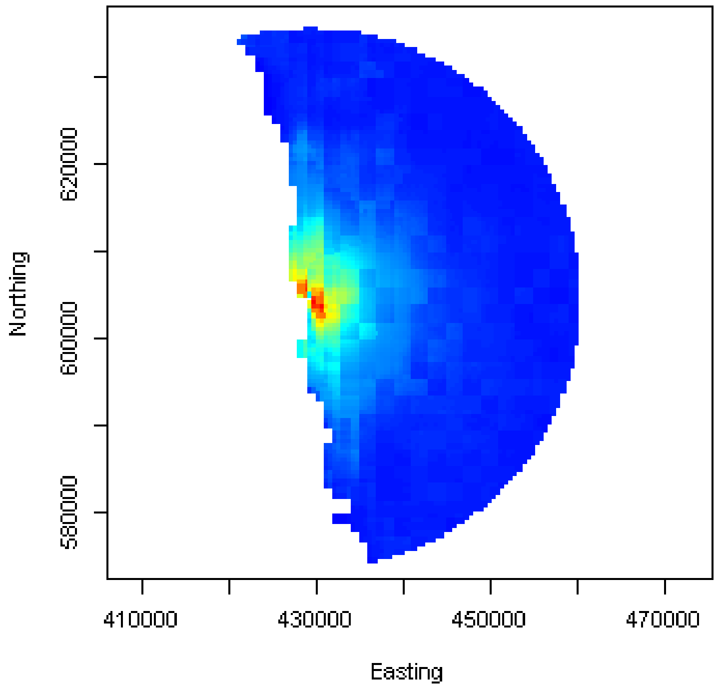
Some of the large sample sizes effectively mean the analysis should be very sensitive to genuine relationships between tern presence and the environmental covariates. However, there is also a real possibility of uncovering spurious relationships, that is, relationships which may be due to a very small actual effect, caused by only a small number of observations that are outlying in covariate space. Further work should investigate the effect of removing outliers in the environmental covariates that may be unreliable. Despite this caveat, there was some evidence of consistent effects between colonies within species of tern. It may be possible that such differences within species but between colonies are due to the birds reacting to local variability. Future work could look at the extent to which it is possible to make reliable predictions for one colony using a model developed for another colony.

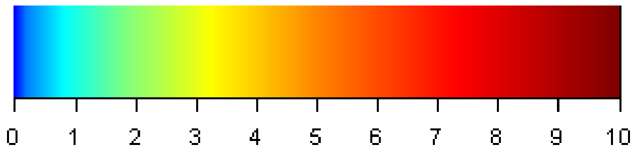
7. Maps Appendix

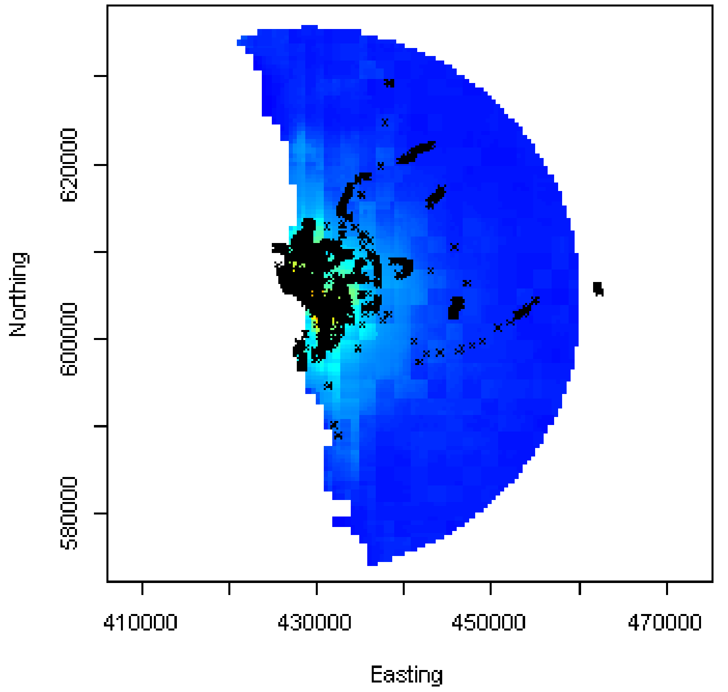
The maps in this appendix have been produced on the basis of the selected GAM models described in Section 5 and in the Results Appendix. Some outliers in the covariate values were removed before making predictions. For each species/colony combination, there are two pages of maps. The first page contains mapped preference, the second mapped usage. For ease of comparison with actual observations, each predicted map has been plotted twice – once by itself (top) and once with the observations added as black crosses.

For the usage maps, the probabilities have been multiplied by 1000.

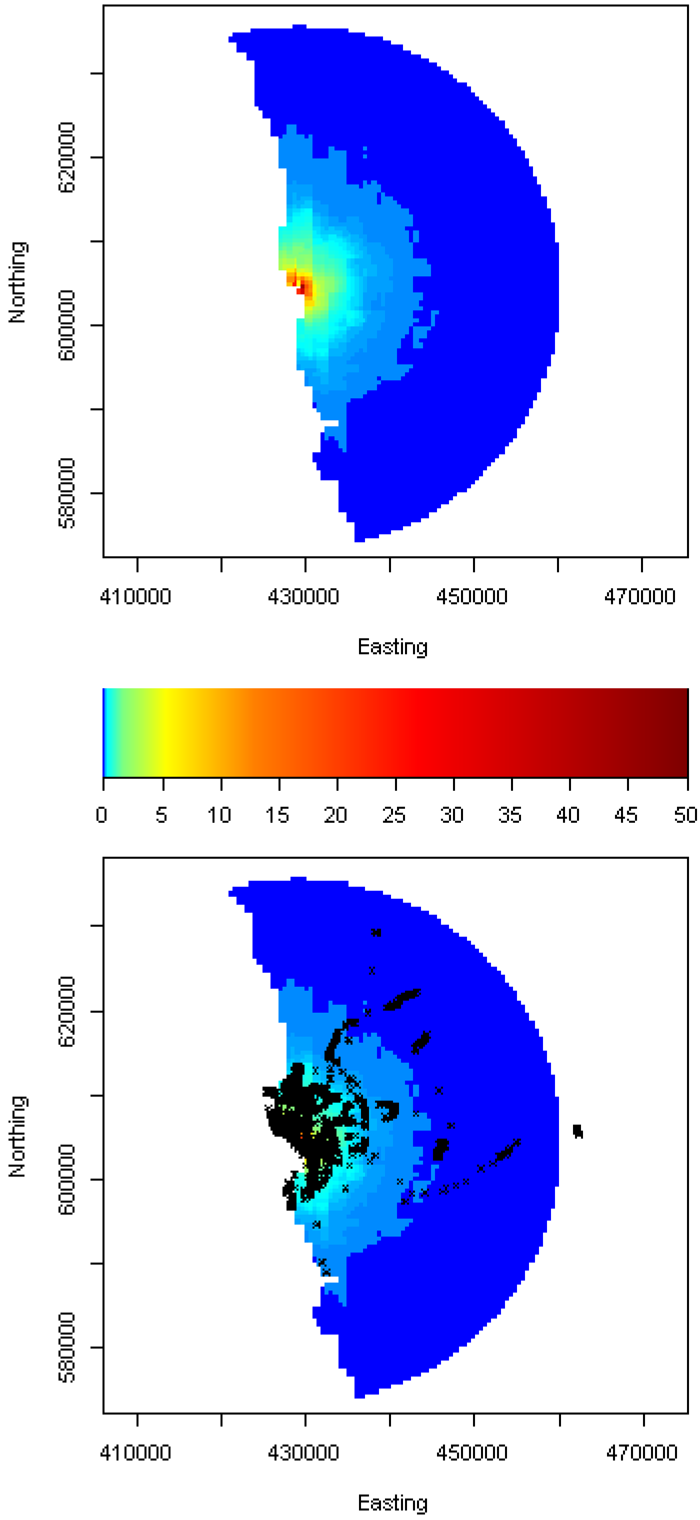
Arctic Tern, Coquet Colony – Preference



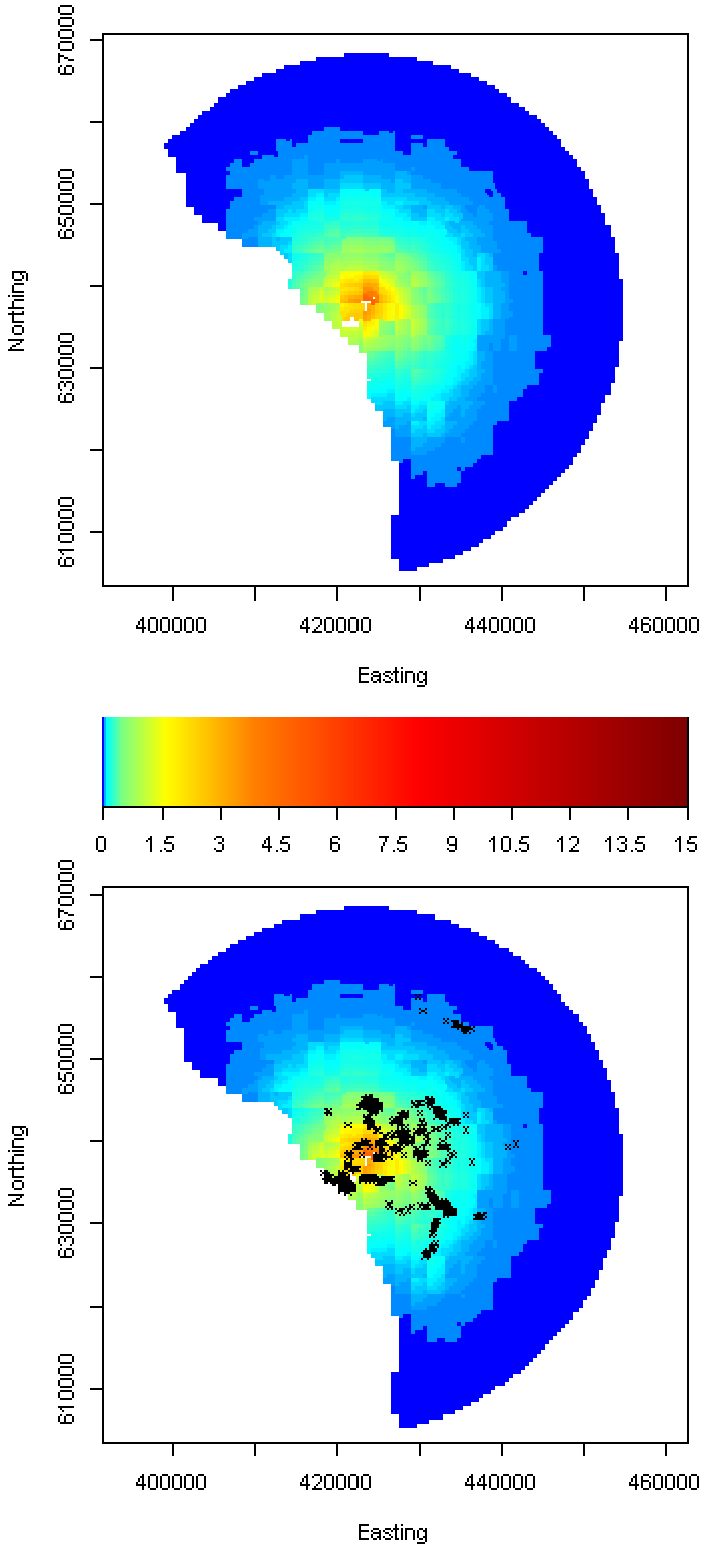




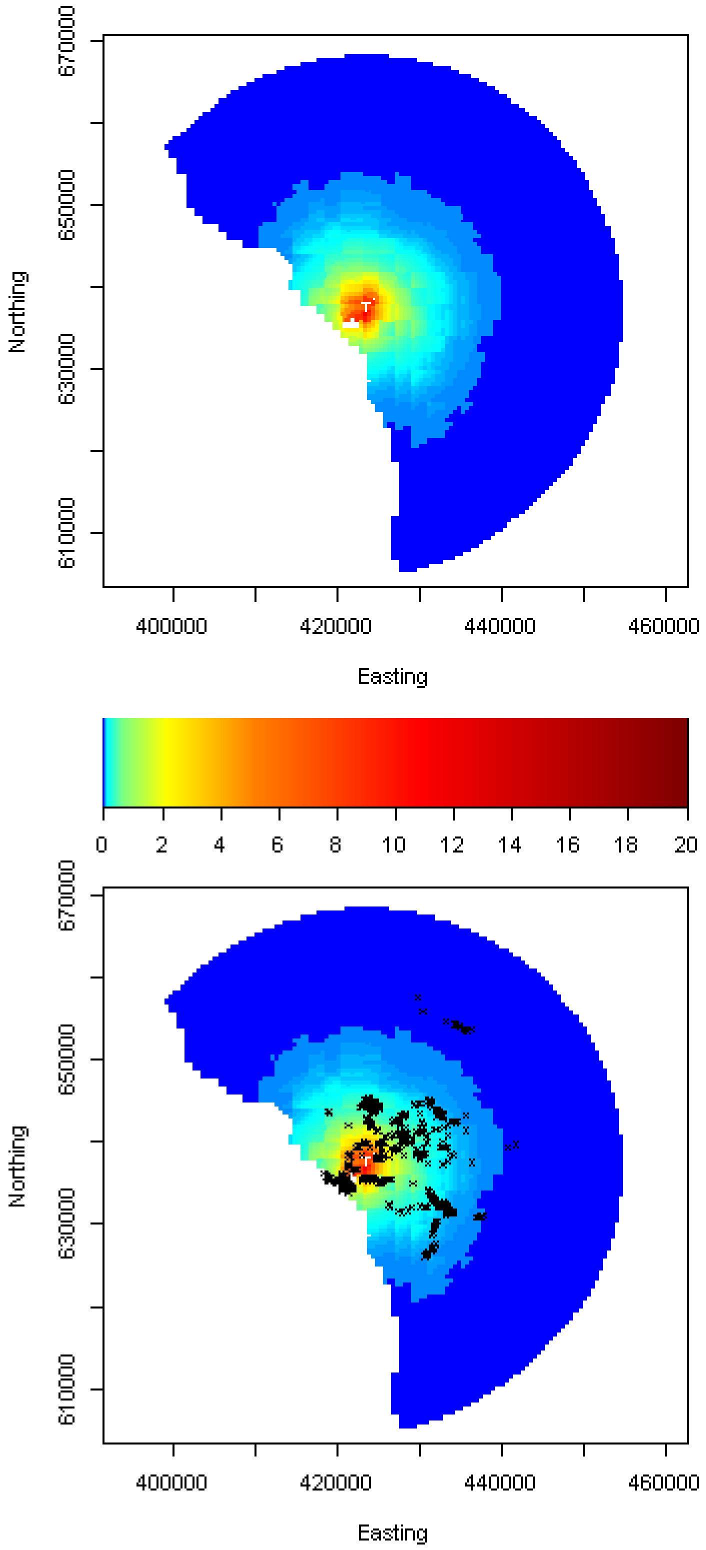
Arctic Tern, Coquet Colony – Usage



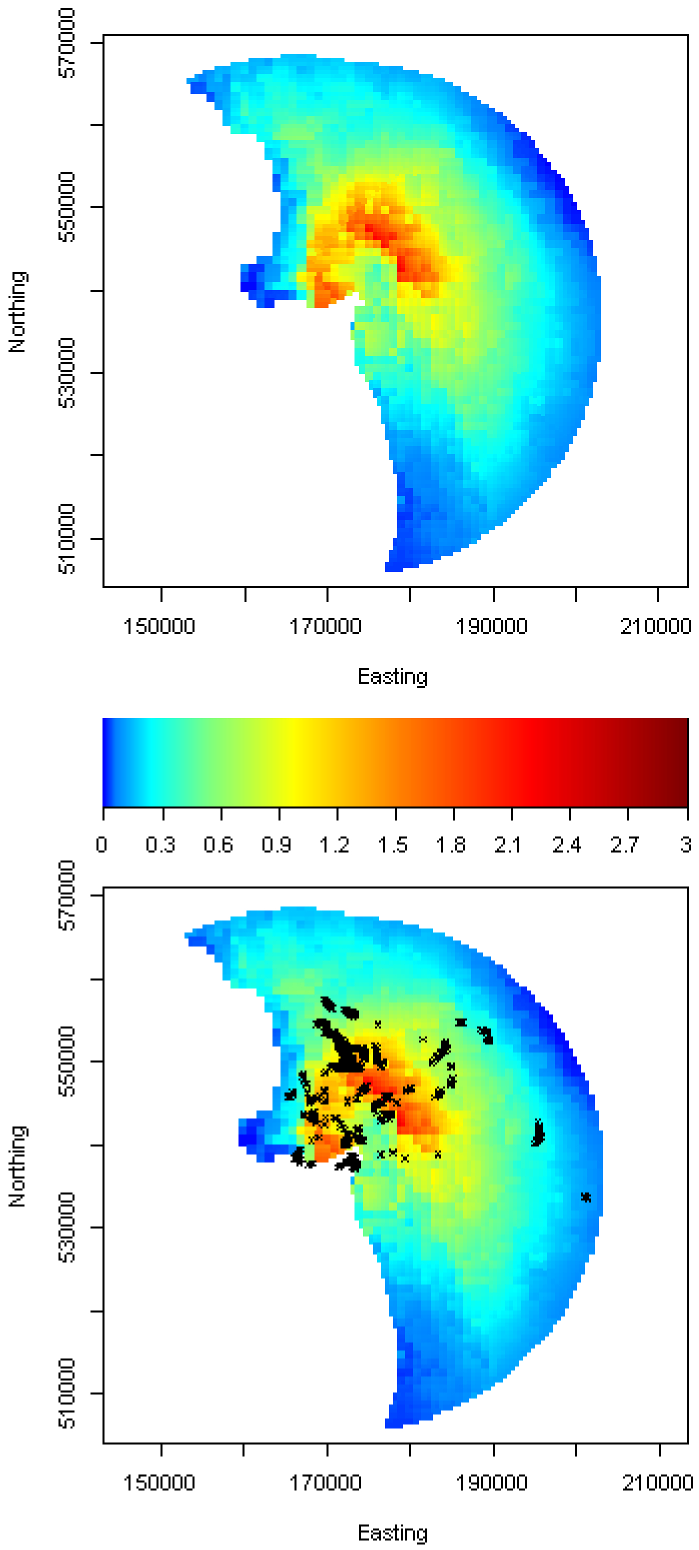
Arctic Tern, Farnes Colony – Preference



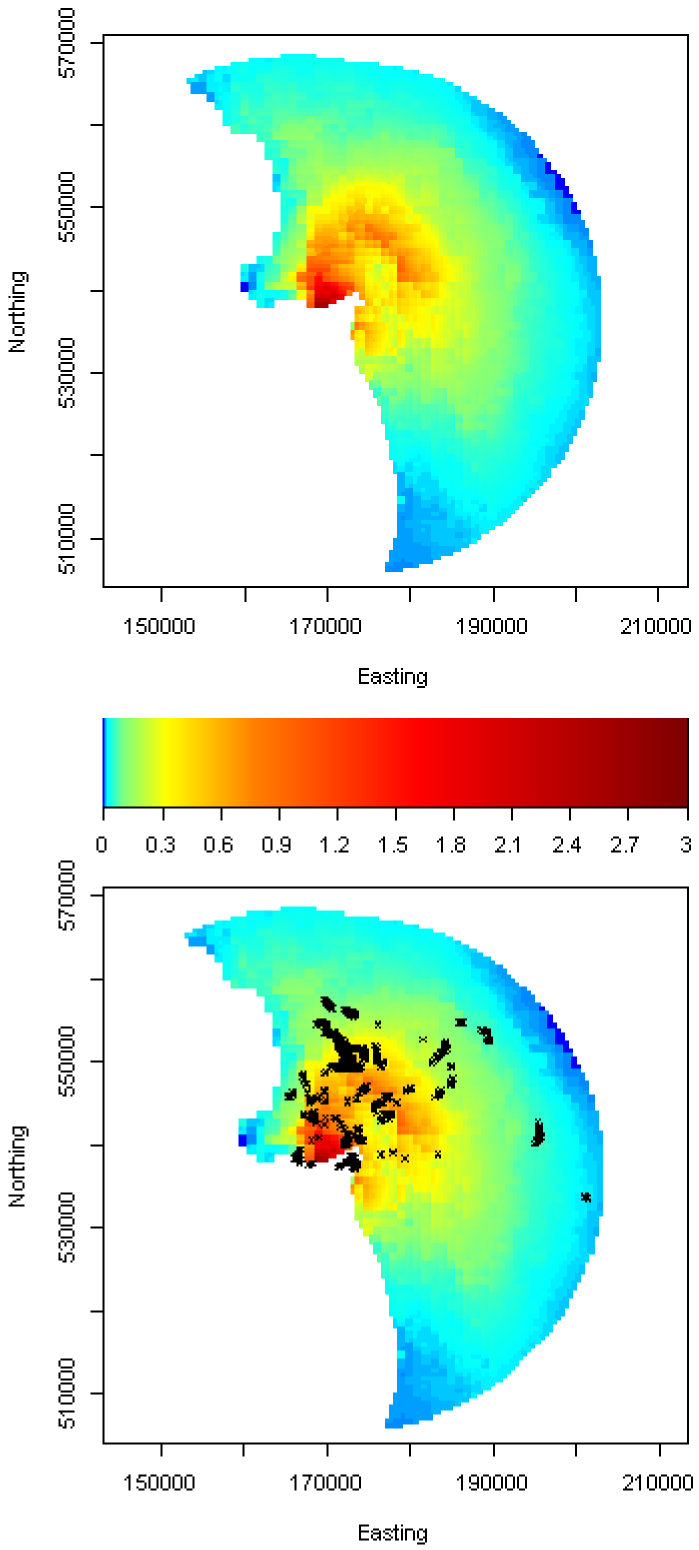
Arctic Tern, Farnes Colony – Usage



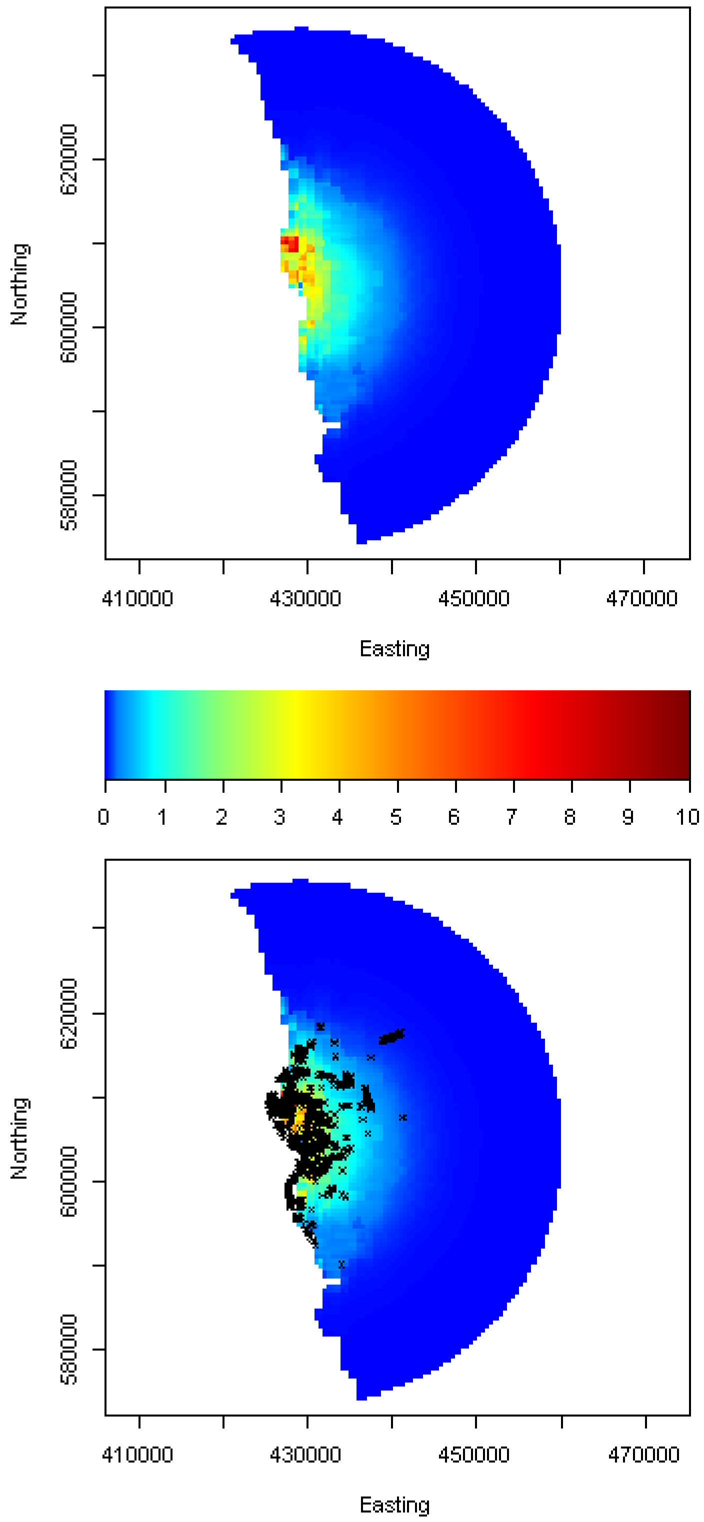
Arctic Tern, Outer Ards Colony – Preference



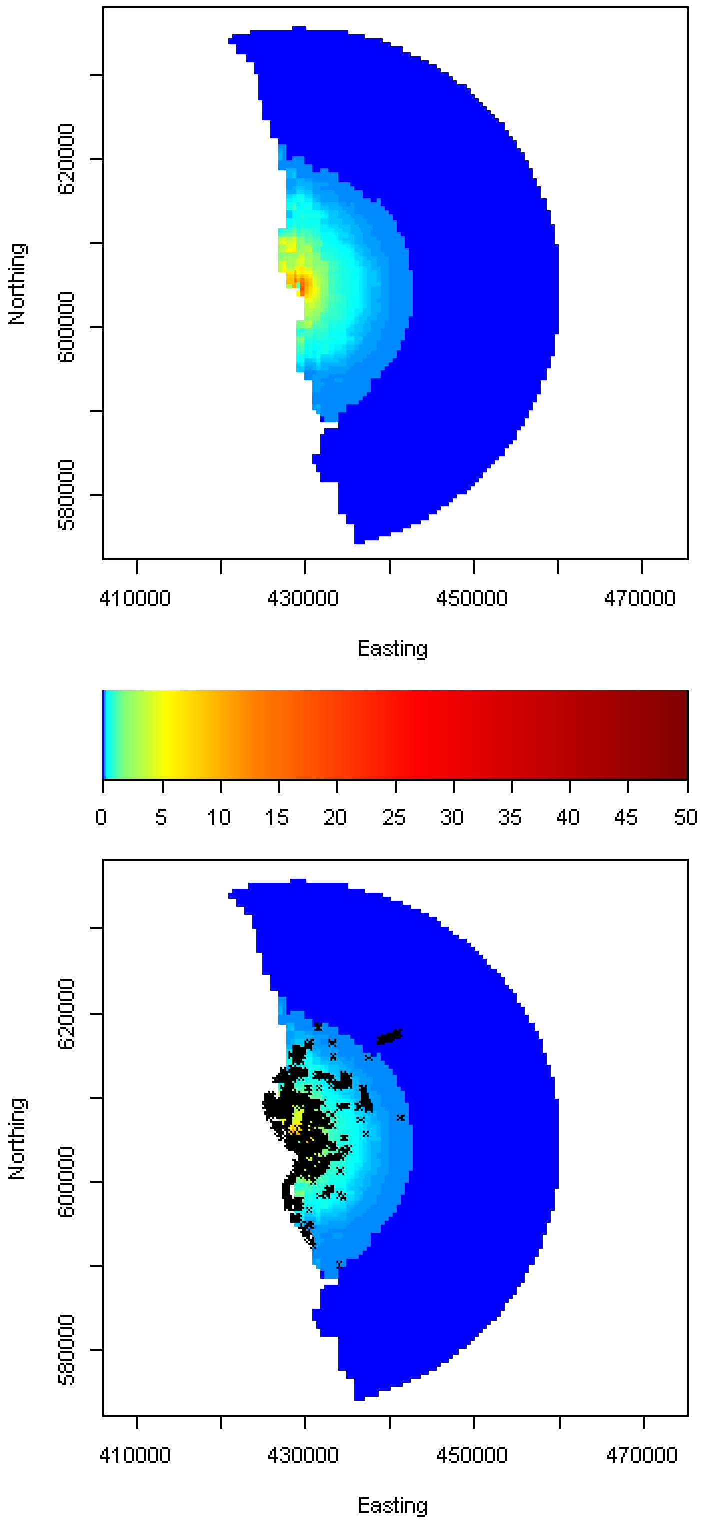
Arctic Tern, Outer Ards Colony – Usage



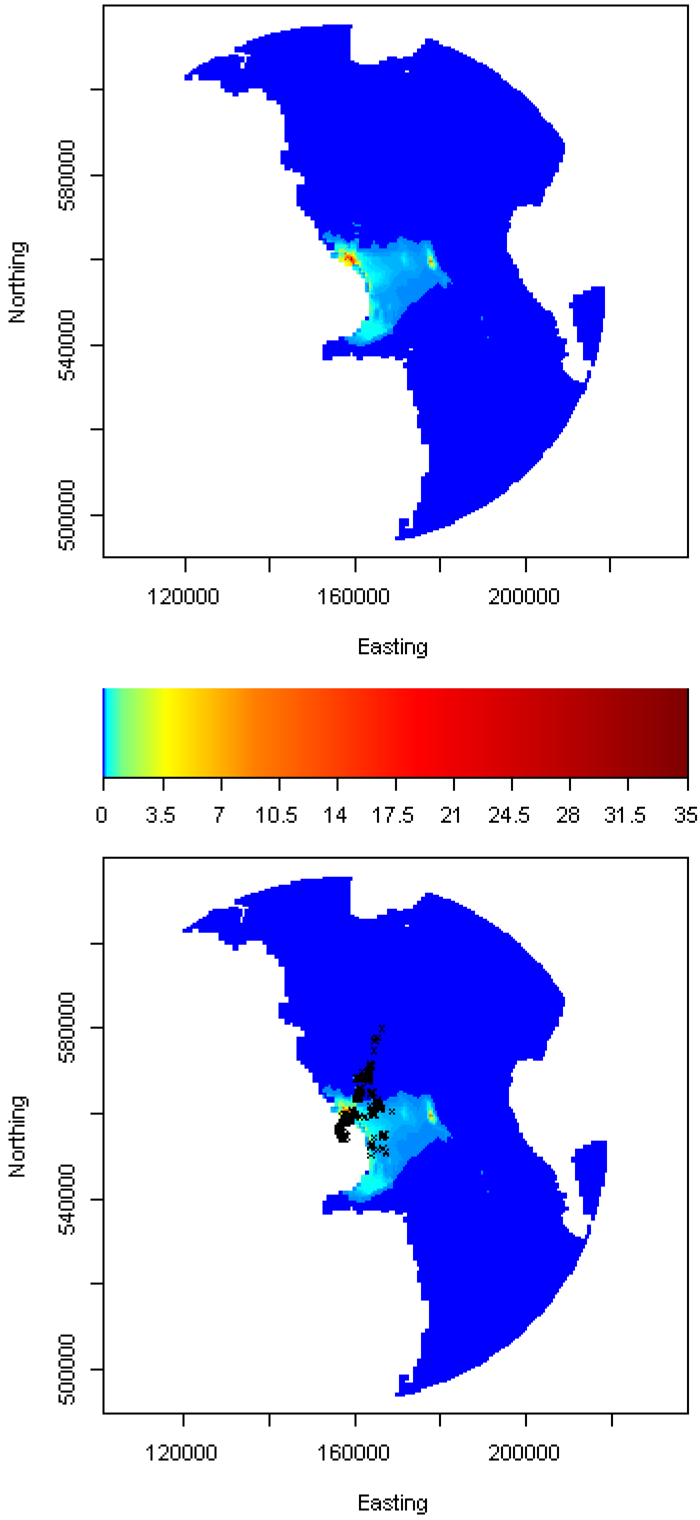
Common Tern, Coquet Colony – Preference



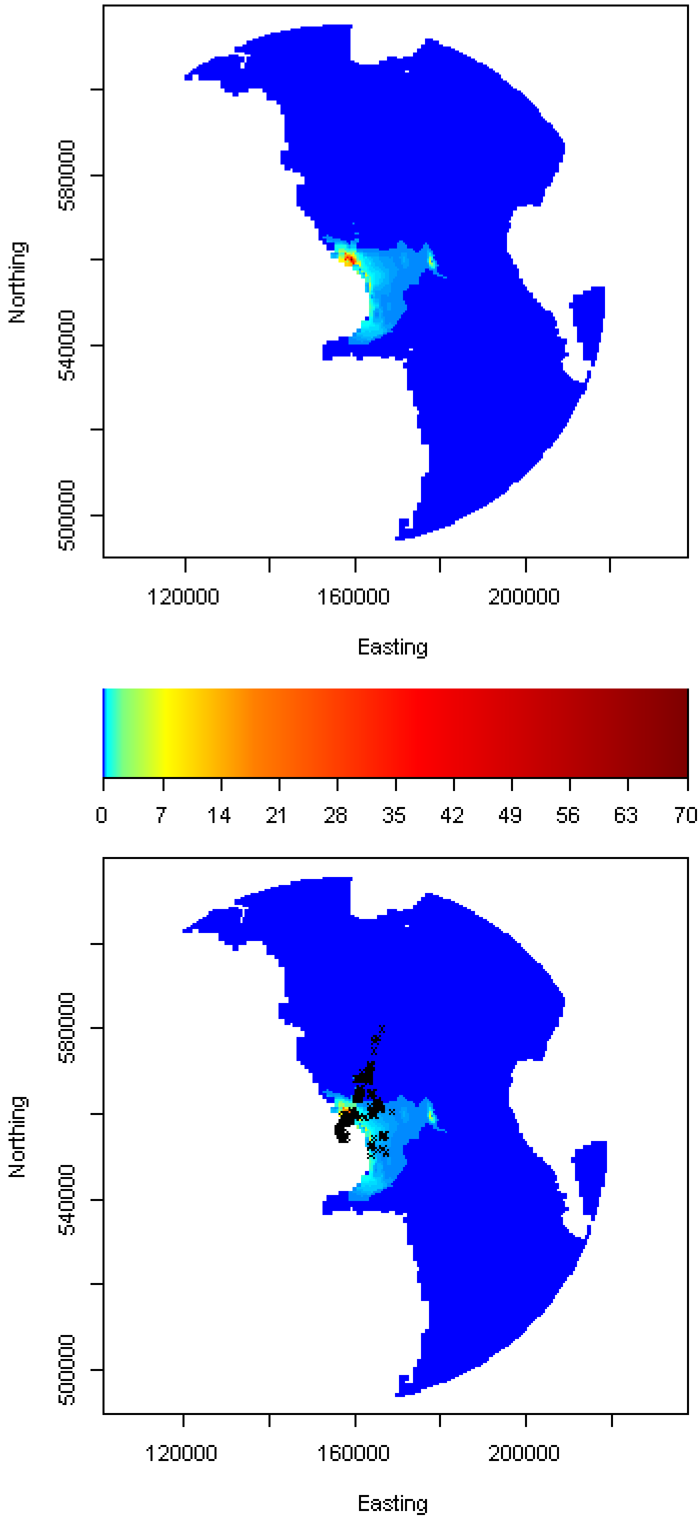
Common Tern, Coquet Colony – Usage



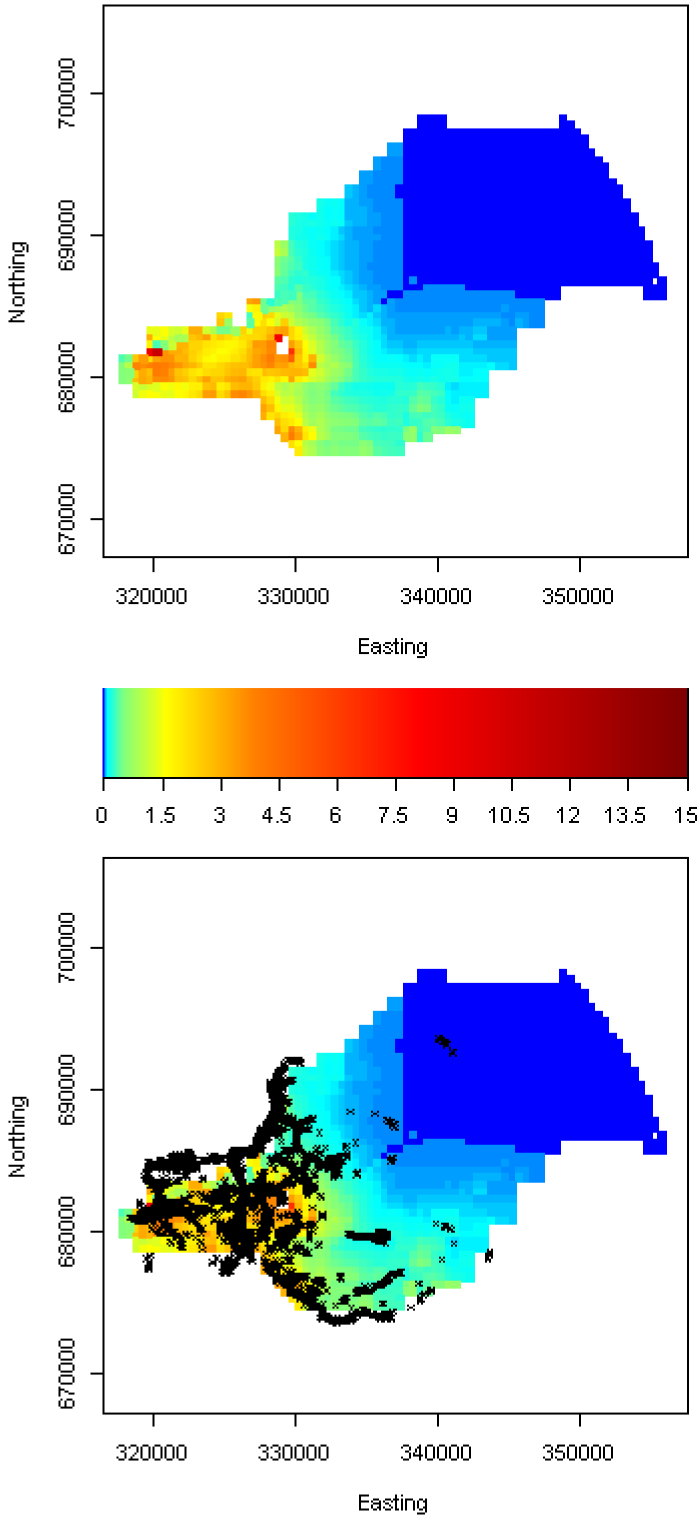
Common Tern, Larne Lough Colony – Preference



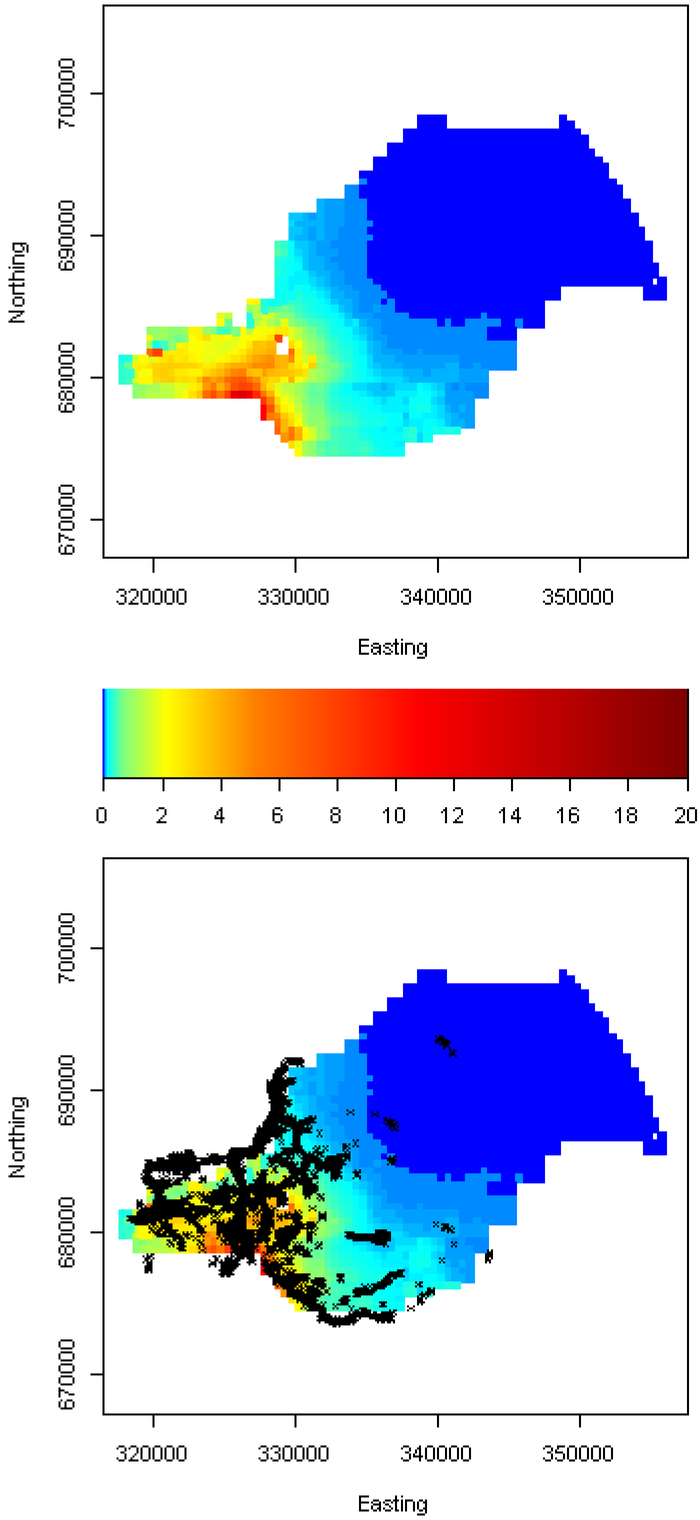
Common Tern, Larne Lough Colony – Usage



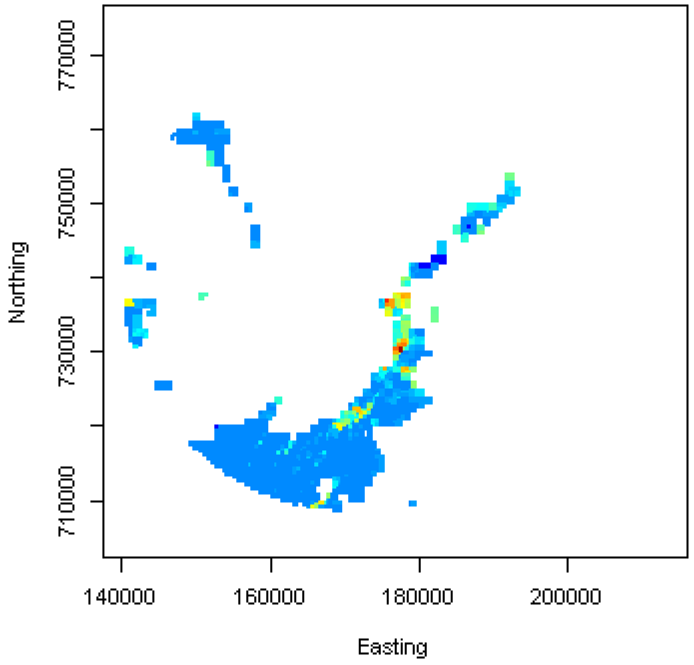
Common Tern, Leith Colony – Preference

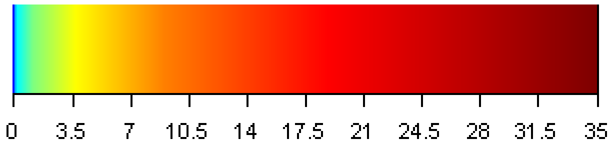


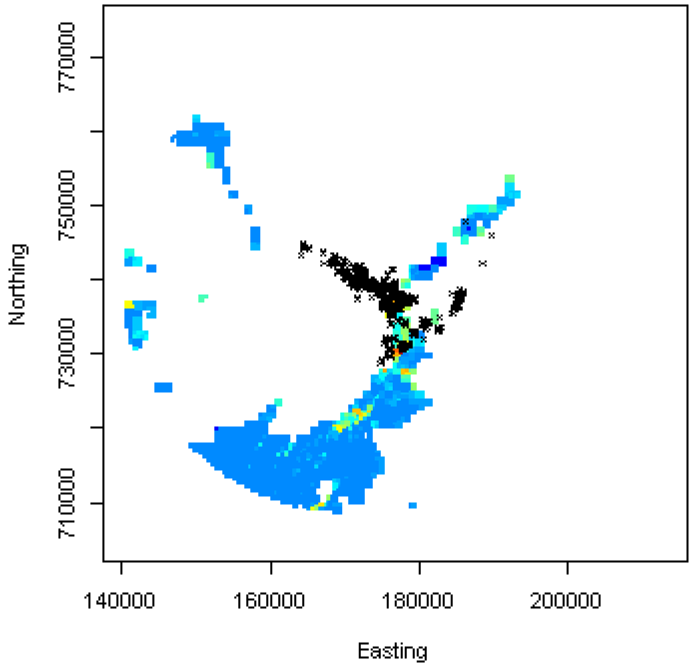
Common Tern, Leith Colony – Usage



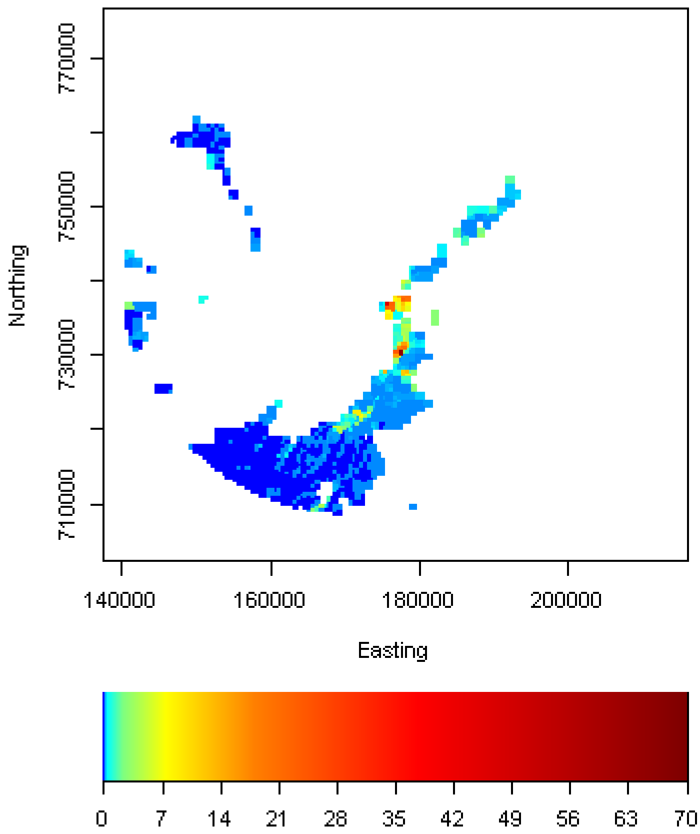
Common Tern, Mull Colony – Preference

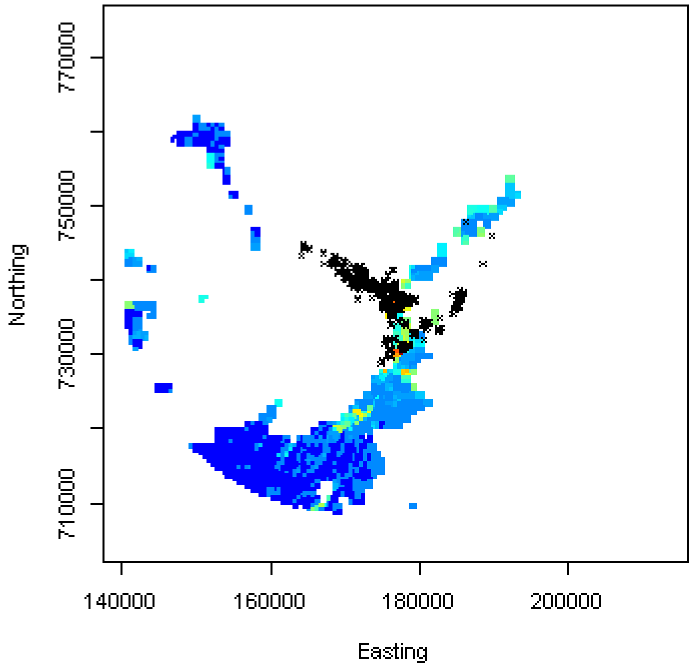




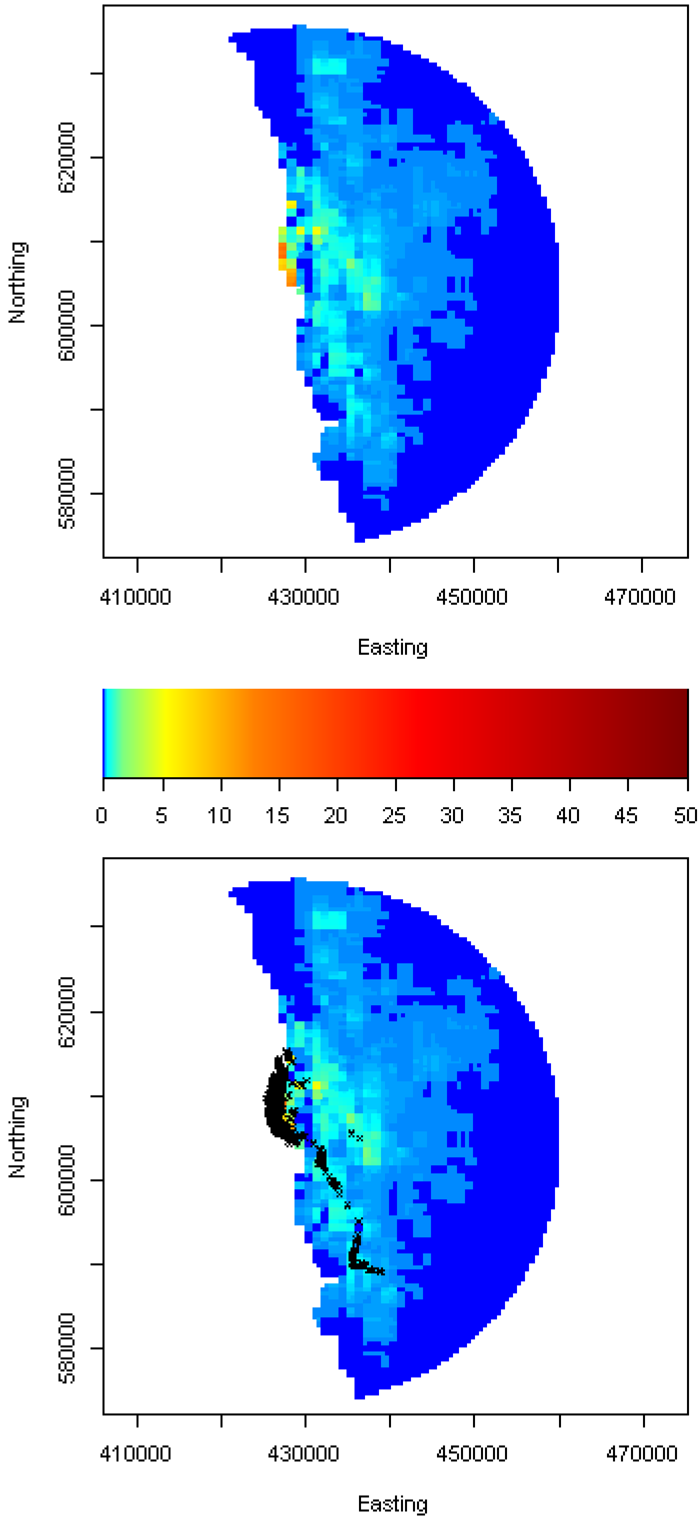


Common Tern, Mull Colony – Usage

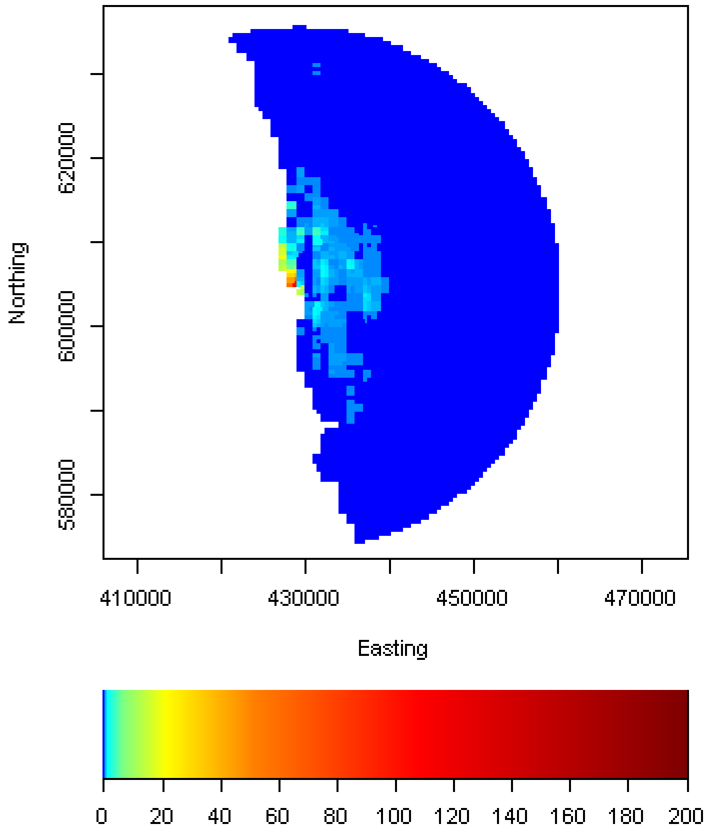


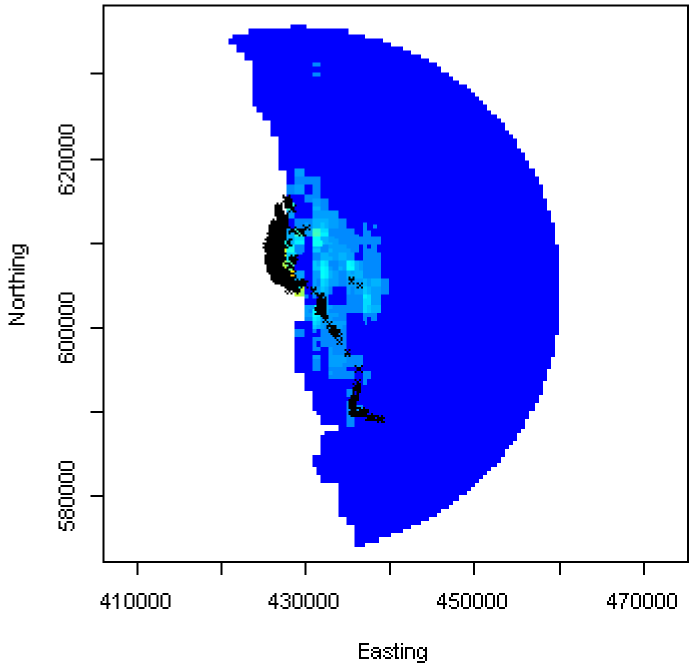


Roseate Tern, Coquet Colony – Preference

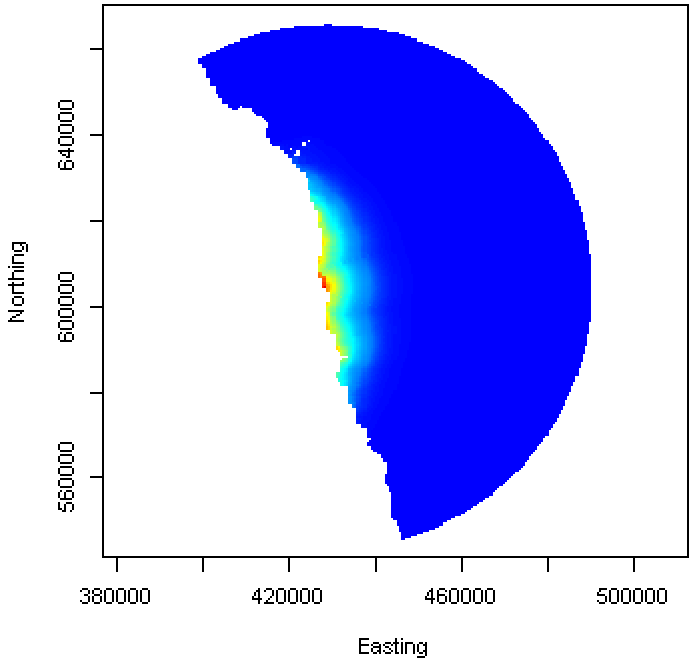


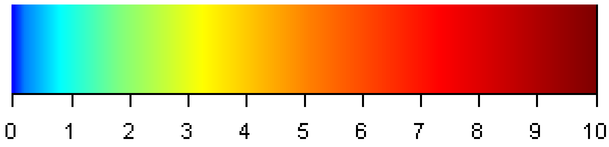
Roseate Tern, Coquet Colony – Usage

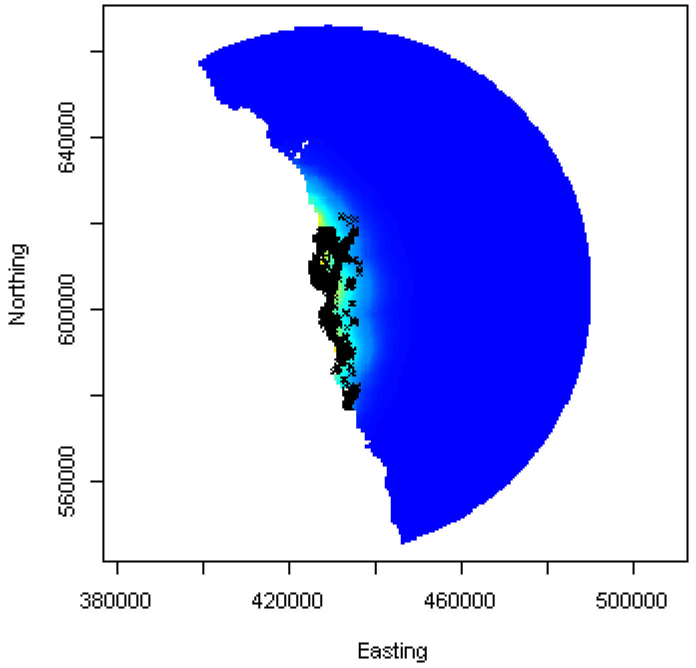




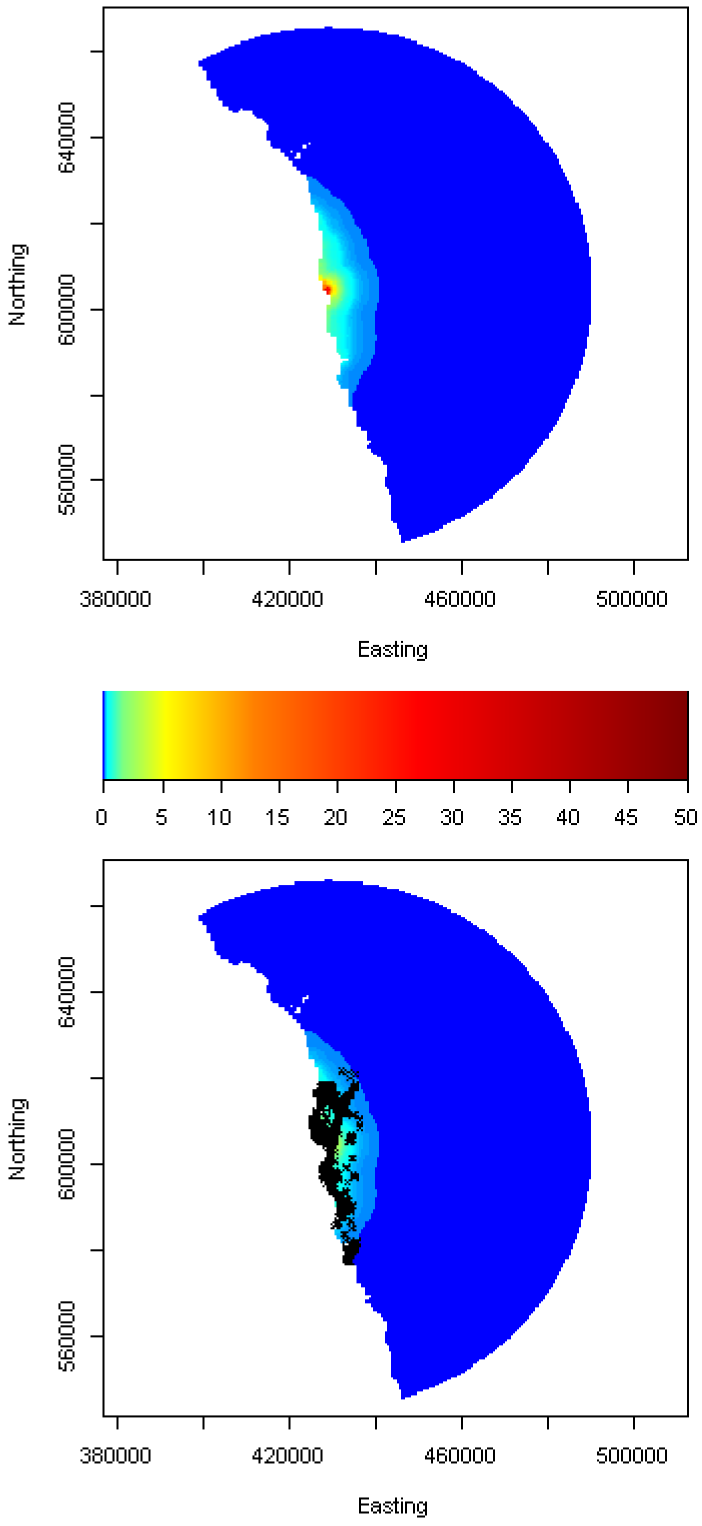
Sandwich Tern, Coquet Colony – Preference



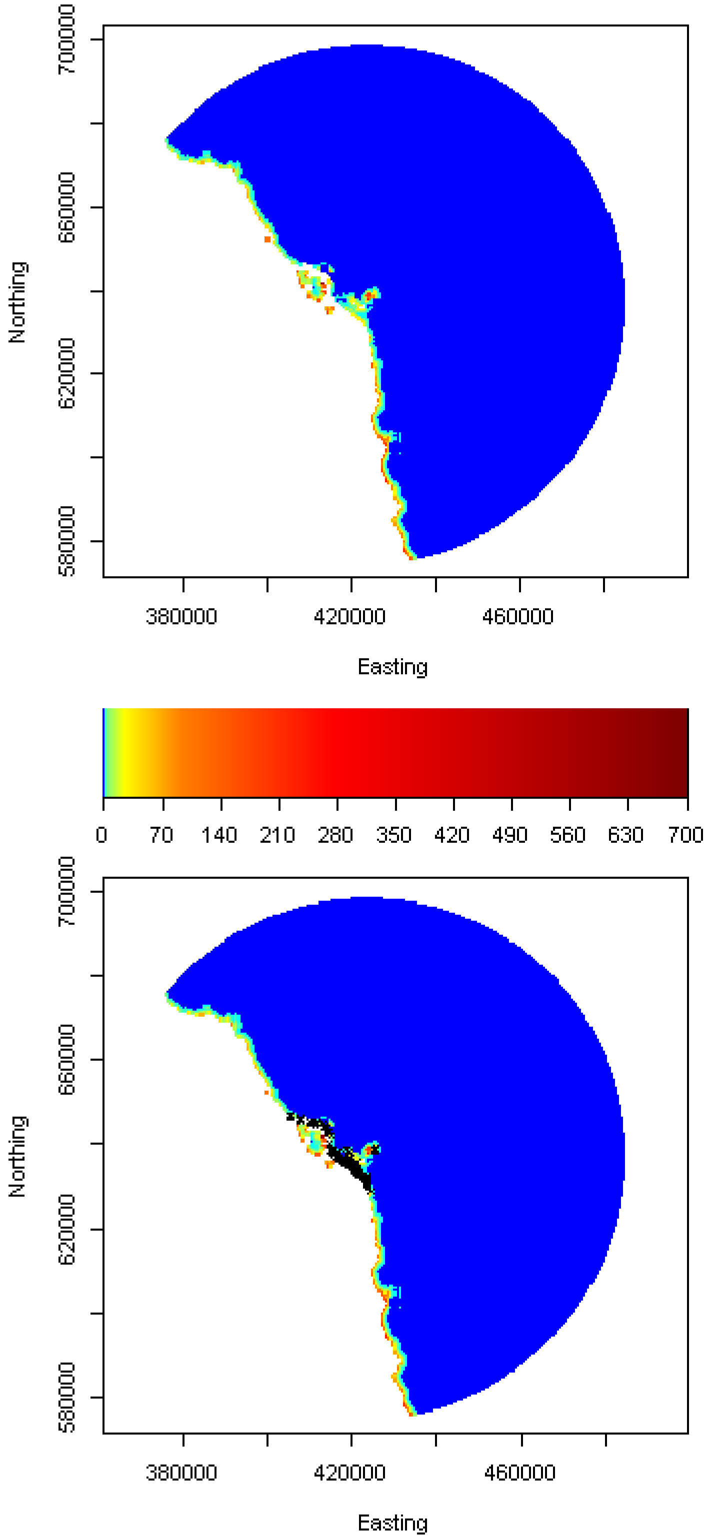




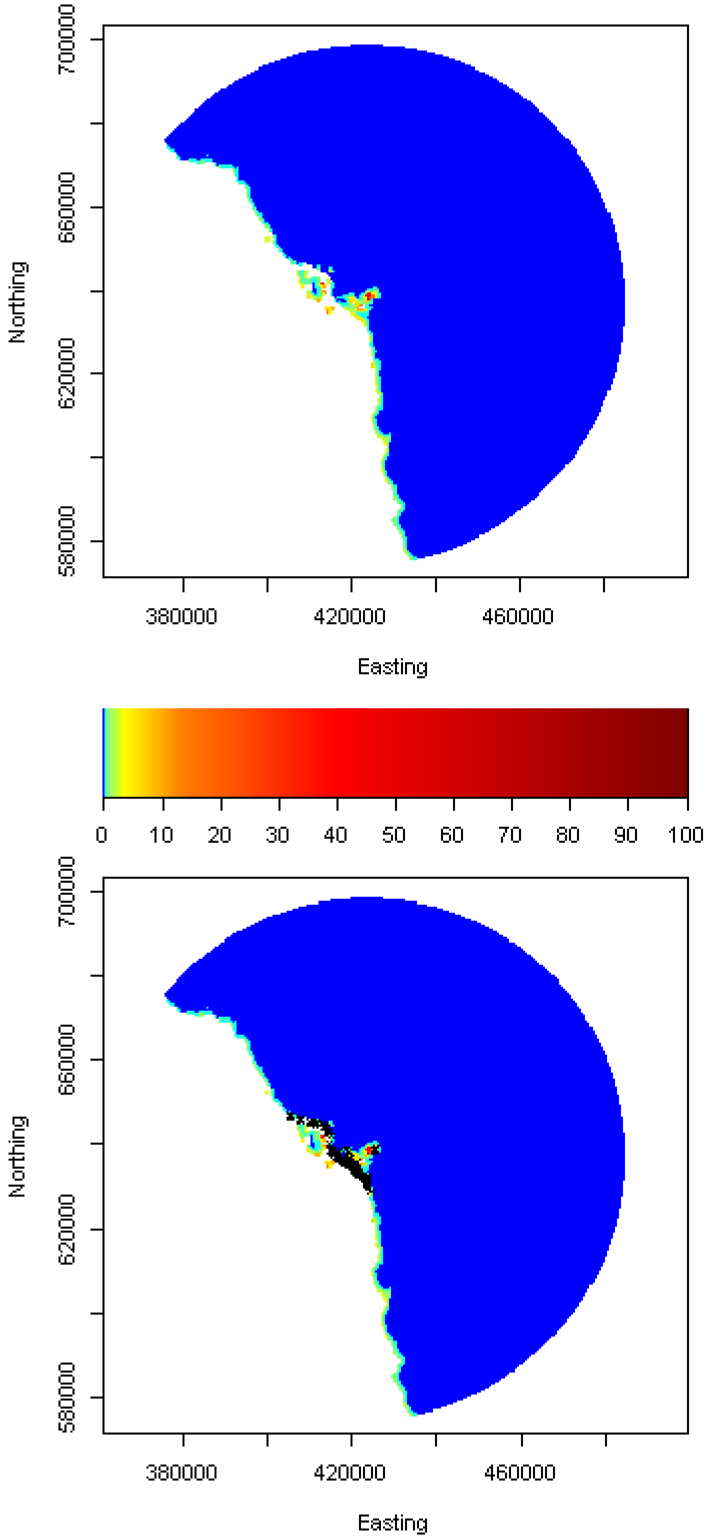
Sandwich Tern, Coquet Colony – Usage



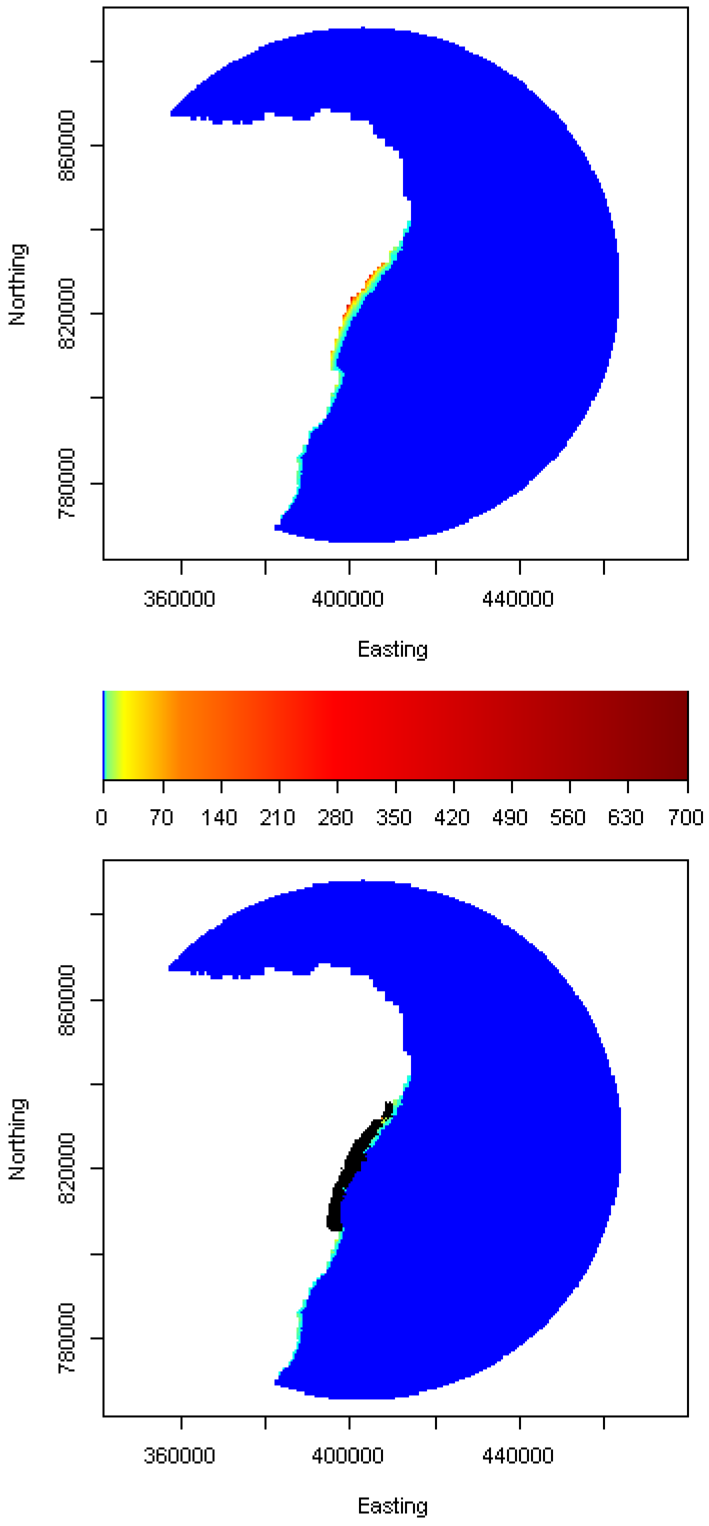
Sandwich Tern, Farnes Colony – Preference



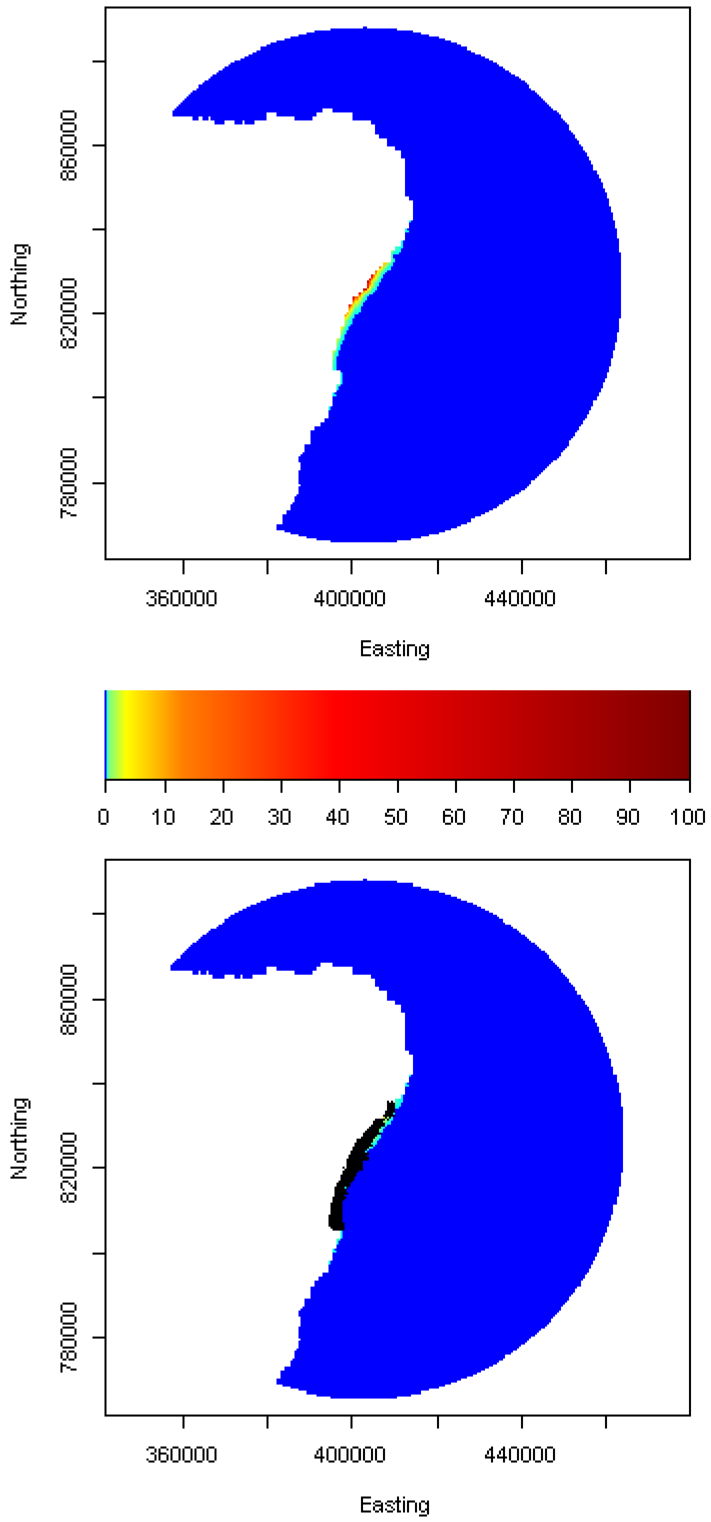
Sandwich Tern, Farnes Colony – Usage



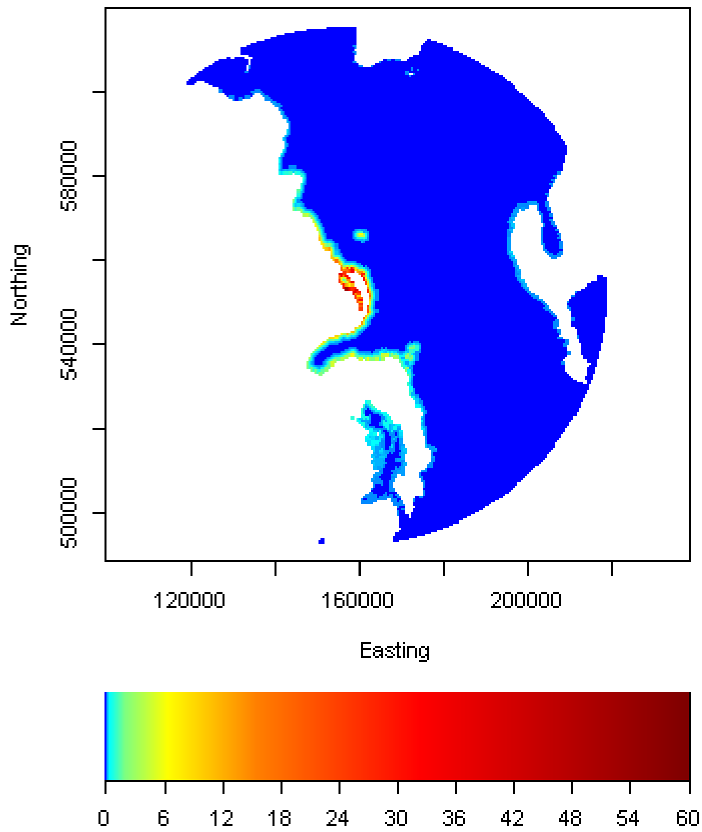
Sandwich Tern, Forvie Colony – Preference

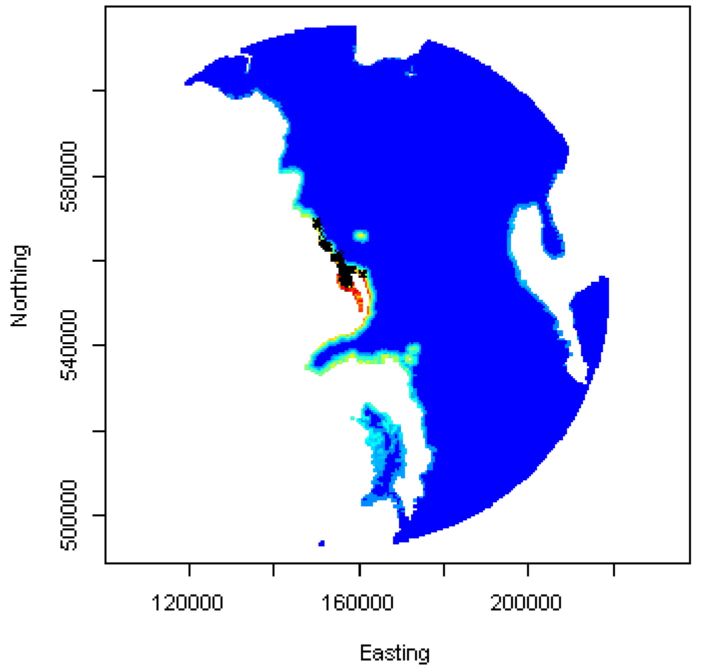


Sandwich Tern, Forvie Colony – Usage

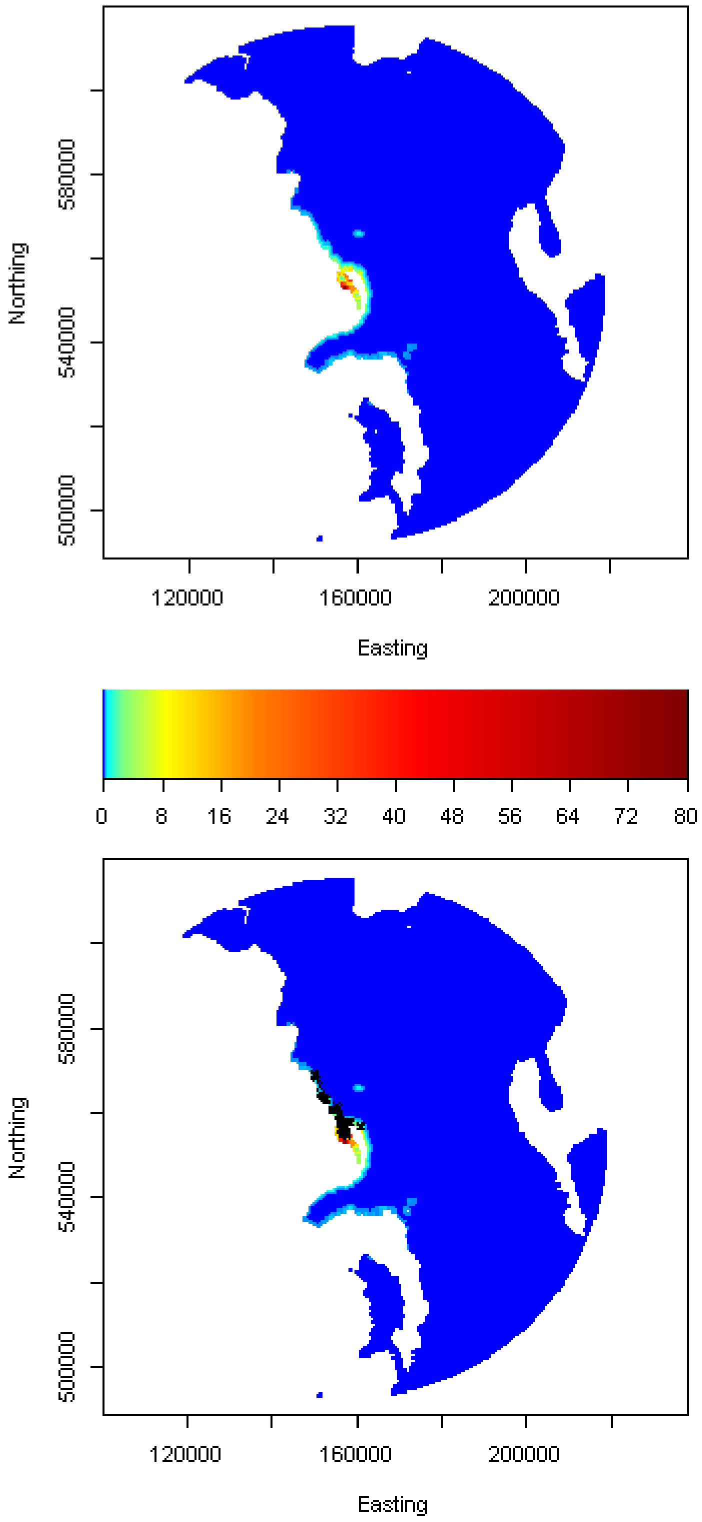


Sandwich Tern, Larne Lough Colony – Preference





Sandwich Tern, Larne Lough Colony – Usage



References

Aarts, G., MacKenzie, M., McConnell, B., Fedak, M. and Matthiopoulos, J. (2008) Estimating space-use and habitat preference from wildlife telemetry data. *Ecography,* **31**, 140-160.

Aarts, G., Fieberg, J. and Matthiopoulos, J. (2012). Comparative interpretation of count, presence–absence and point methods for species distribution models. *Methods in Ecology and Evolution*, **3**, 177-187.

Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J. and Elston, D.A. (2010) Regression analysis of spatial data. *Ecology Letters*, **13**, 246-264.

Boyce, M.S. and McDonald, L.L. (1999) Relating populations to habitats using resource selection functions. *Trends in Ecology and Evolution*, **14**, 268-272.

INLA (2012) R-Package, http://www.r-inla.org/ .

Keating, K. A. and Cherry, S. (2004) Use and interpretation of logistic regression in habitat selection studies. *Journal of Wildlife Management,* **68**, 774-789.

R Development Core Team (2012) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/ .

Shmueli, G. (2010) To explain or to predict? *Statistical Science*, **25**, 289-310.

Wakefield, E.D., Phillips, R.A., Trathan, P.N., Arata, J., Gales, R., Huin, N., Robertson, G., Waugh, S.M., Weimerskirch, H. and Matthiopoulos, J. (2011) Habitat preference, accessibility, and competition limit the global distribution of breeding Black-browed Albatrosses. *Ecological Monographs,* **81**, 141-167.

Warton, D.I. and Shepherd, L.C. (2010) Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Annals of Applied Statistics,* **4**, 1383-1402.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society B,* **73**, 3-36.

Biomathematics and Statistics Scotland

To carry out tern modelling under the Framework Agreement C10-0206-0387

CONTRACT No: C10-0206-0387

RESULTS APPENDIX

Submitted to:

Joint Nature Conservation Committee

March 2012

v. 21.12.2012 18:10

A. Results

This appendix contains the text and graphical output from the analysis of the terns tracking data. The section match those in the results section of the main report, ordered by tern species and then by colony within species.

The sections within each species/colony sections are: (i) output from the AIC selection for GLM; (ii) output from the BIC selection for GLM; (iii) output from the likelihood ratio (LRT) selection for GLM, with checking for consistent Year effects if more than one year of data exists for that species/colony; (iv) output from the GAM selection, including plots of smooth covariate relationships; and (v) output from the INLA analysis applied to the selected GAM model, with plots of smooth covariates if linear terms were insufficient.

Proper significance assessment of GLMs is best performed by studying the Analysis of Deviance table, hence this is presented along with the standard "regression-style" table.

A.1 Arctic Terns

A.1.1 Coquet

A.1.1.1 GLM Output using AIC for Selection

```
AIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_col + sst_may + summ_front_sd +

    ss_wave + ss_current, family = "binomial", data = complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.96383  -0.03654  -0.01648  -0.00803   1.92561


Coefficients:

               Estimate Std. Error z value Pr(>|z|)

(Intercept)   -14.90185    7.39603  -2.015   0.0439 *

dist_col       -0.20735    0.03635  -5.704 1.17e-08 ***

sst_may         1.53373    0.80628   1.902   0.0571 .

summ_front_sd  -0.02659    0.01193  -2.229   0.0258 *

ss_wave        -0.03882    0.01675  -2.318   0.0204 *

ss_current      1.14923    0.54748   2.099   0.0358 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 542.78  on 47731  degrees of freedom
Residual deviance: 419.43  on 47726  degrees of freedom
AIC: 53.701
Number of Fisher Scoring iterations: 8
```

A.1.1.2 GLM Output using BIC for Selection

```
BIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col, family = "binomial",
    data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.79352  -0.03705  -0.01715  -0.00702   1.93099

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.89725    0.18258  -4.914 8.91e-07 ***
dist_col    -0.21127    0.02803  -7.537 4.81e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 542.78  on 47731  degrees of freedom
Residual deviance: 429.21  on 47730  degrees of freedom
AIC: 46.903
Number of Fisher Scoring iterations: 8
```

A.1.1.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
```

```
-0.96383  -0.03654  -0.01648  -0.00803   1.92561


Coefficients:

               Estimate Std. Error z value Pr(>|z|)

(Intercept)   -14.90185    7.39603  -2.015   0.0439 *

dist_col       -0.20735    0.03635  -5.704 1.17e-08 ***

sst_may         1.53373    0.80628   1.902   0.0571 .

summ_front_sd  -0.02659    0.01193  -2.229   0.0258 *

ss_wave        -0.03882    0.01675  -2.318   0.0204 *

ss_current      1.14923    0.54748   2.099   0.0358 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 542.78  on 47731  degrees of freedom

Residual deviance: 419.43  on 47726  degrees of freedom

AIC: 53.701

Number of Fisher Scoring iterations: 8
```

Analysis of deviance output for reliable assessment of significance:

```
Model:

SEARCH_FORAGE ~ dist_col + sst_may + summ_front_sd + ss_wave +

    ss_current

              Df Deviance     AIC    LRT  Pr(>Chi)

<none>             419.43  53.701

dist_col       1   478.95 111.221 59.520 1.211e-14 ***

sst_may        1   423.48  55.750  4.048   0.04422 *

summ_front_sd  1   424.88  57.153  5.451   0.01955 *

ss_wave        1   425.56  57.833  6.131   0.01328 *

ss_current     1   423.71  55.978  4.277   0.03863 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now checking for interactions with Year:

```
Model:
SEARCH_FORAGE ~ Year * (dist_col + sst_may + summ_front_sd +
    ss_wave + ss_current)
                   Df Deviance    AIC     LRT Pr(>Chi)
<none>                  409.63 77.486
Year:dist_col       2   412.74 76.597 3.11156   0.2110
Year:sst_may        2   410.04 73.895 0.40969   0.8148
Year:summ_front_sd  2   409.75 73.604 0.11875   0.9424
Year:ss_wave        2   410.61 74.468 0.98264   0.6118
Year:ss_current     2   409.71 73.568 0.08231   0.9597
No significant Year interactions.
```

A.1.1.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):
Family: binomial
Link function: logit
Formula:
SEARCH_FORAGE ~ s(dist_col, k = 3) + s(sst_may, k = 3) + s(summ_front_sd,
    k = 3) + s(ss_wave, k = 3) + s(ss_current, k = 3)
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.8379     0.2935  -13.08   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                   edf Ref.df Chi.sq  p-value
s(dist_col)      1.726  1.925 31.177 1.49e-07 ***
s(sst_may)       1.000  1.000  3.664   0.0556 .
s(summ_front_sd) 1.000  1.000  3.591   0.0581 .
s(ss_wave)       1.000  1.000  6.240   0.0125 *
s(ss_current)    1.000  1.000  2.897   0.0887 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
R-sq.(adj) =  0.196   Deviance explained = 23.5%

REML score = 213.08  Scale est. = 1         n = 47732
```

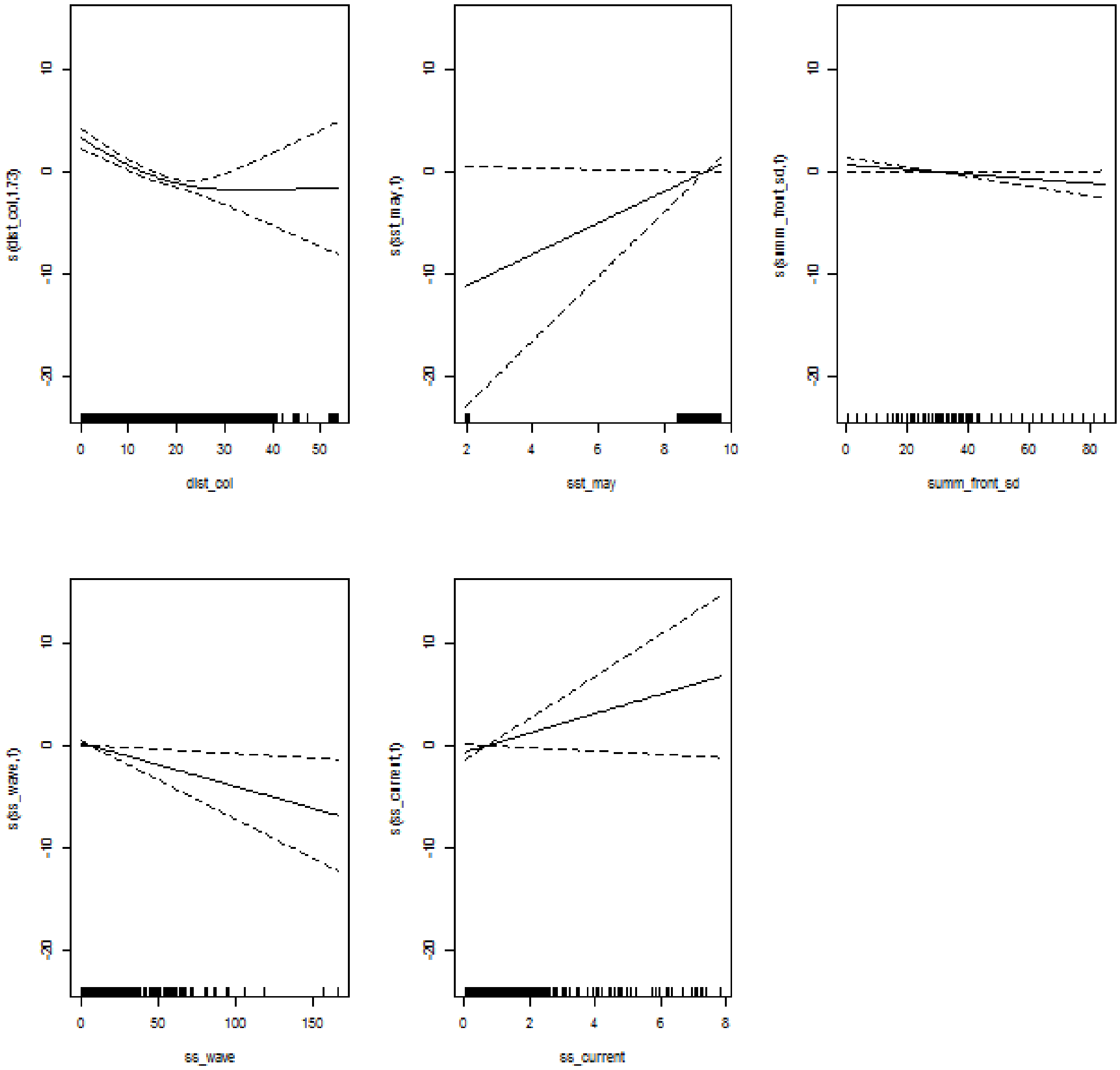


Figure A.1.1.4: GAM covariate plots for Arctic Terns in Coquet

A.1.1.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                    mean     sd 0.025quant 0.5quant 0.975quant    kld
```

```
(Intercept)    -15.3423 7.3532   -30.1945 -15.1883    -1.3298 0.0022

sst_may          1.6314 0.8038     0.0994   1.6149     3.2528 0.0023

summ_front_sd   -0.0220 0.0121    -0.0465  -0.0217     0.0009 0.0014

ss_wave         -0.0451 0.0172    -0.0804  -0.0445    -0.0129 0.0017

ss_current       0.9403 0.5578    -0.1918   0.9532     1.9998 0.0027

dist_col1       -9.2197 1.7094   -12.9319  -9.0862    -6.2210 0.0837

dist_col2       -0.0356 3.1457    -7.1358   0.3443     5.1207 0.0013


Model hyperparameters:

                           mean   sd     0.025quant 0.5quant 0.975quant

T.0 for mesh.points-basisT  7.547  3.169  1.368      7.527   13.816

Expected number of effective parameters(std dev): 7.001(0.001588)

Number of equivalent replicates : 6818.26

Marginal Likelihood:  -211.91
```

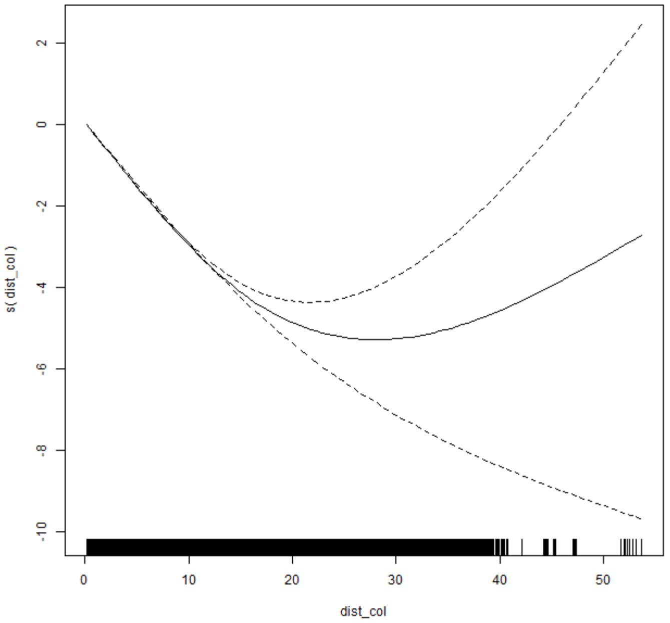


Figure A.1.1.5: INLA GAM covariate plots for Arctic Terns in Coquet (non-linear terms only)

A.1.2 Farnes

A.1.2.1 GLM Output using AIC for Selection

```
AIC Selected Model:

Call:
```

```
glm(formula = SEARCH_FORAGE ~ dist_col + sst_april + summ_front_sd,
    family = "binomial", data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.68797  -0.04791  -0.01884  -0.00604   1.58280

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.51958    1.12428   1.352 0.176505
dist_col      -0.30400    0.06477  -4.693 2.69e-06 ***
sst_april     -0.47563    0.17774  -2.676 0.007451 **
summ_front_sd  0.04579    0.01281   3.574 0.000351 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 213.73  on 17470  degrees of freedom
Residual deviance: 159.82  on 17467  degrees of freedom
AIC: 15.356
Number of Fisher Scoring iterations: 8
```

A.1.2.2 GLM Output using BIC for Selection

```
BIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col + summ_front_sd, family = "binomial",
    data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.67055  -0.04865  -0.02003  -0.00692   1.54020

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.41777    0.37156  -3.816 0.000136 ***
dist_col      -0.29305    0.06254  -4.685 2.79e-06 ***
```

```
summ_front_sd  0.03561    0.01171   3.040 0.002366 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 213.73  on 17470  degrees of freedom
Residual deviance: 166.41  on 17468  degrees of freedom
AIC: 13.721
Number of Fisher Scoring iterations: 8
```

A.1.2.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.68797  -0.04791  -0.01884  -0.00604   1.58280

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.51958    1.12428   1.352 0.176505
dist_col      -0.30400    0.06477  -4.693 2.69e-06 ***
sst_april     -0.47563    0.17774  -2.676 0.007451 **
summ_front_sd  0.04579    0.01281   3.574 0.000351 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 213.73  on 17470  degrees of freedom
Residual deviance: 159.82  on 17467  degrees of freedom
AIC: 15.356
Number of Fisher Scoring iterations: 8
```

Analysis of deviance output for reliable assessment of significance:

```
Model:

SEARCH_FORAGE ~ dist_col + sst_april + summ_front_sd

              Df Deviance    AIC    LRT  Pr(>Chi)

<none>            159.82 15.356

dist_col       1   206.86 60.396 47.039 6.958e-12 ***

sst_april      1   166.41 19.951  6.595 0.0102292 *

summ_front_sd  1   174.27 27.804 14.447 0.0001441 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A.1.2.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):

Family: binomial

Link function: logit

Formula:

SEARCH_FORAGE ~ s(dist_col, k = 3) + s(sst_april, k = 3) + s(summ_front_sd,

    k = 3)

Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  -4.1271     0.5445  -7.579 3.48e-14 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:

                 edf Ref.df Chi.sq  p-value

s(dist_col)        1  1.000  22.02 2.69e-06 ***

s(sst_april)       1  1.001   7.16 0.007464 **

s(summ_front_sd)   1  1.000  12.78 0.000351 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.357   Deviance explained = 25.2%

REML score =  82.13  Scale est. = 1         n = 17471
```

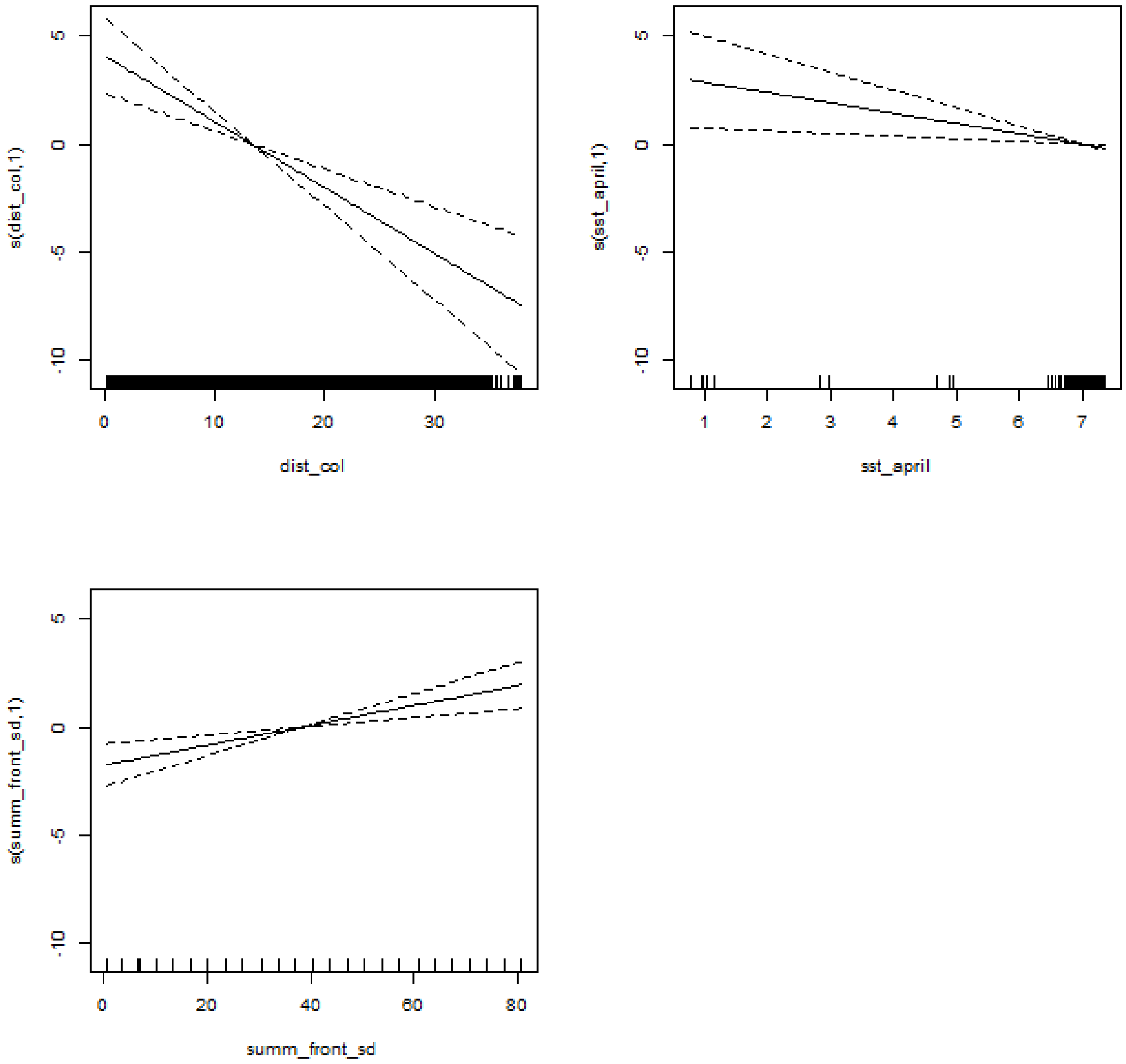


Figure A.1.2.4: GAM covariate plots for Arctic Terns in Farnes

A.1.2.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                  mean     sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)     1.5580 1.1243    -0.6075   1.5433     3.8060 0.0006

dist_col       -0.3095 0.0648    -0.4475  -0.3055    -0.1930 0.0036
```

```
sst_april     -0.4847 0.1777    -0.8406  -0.4822    -0.1428 0.0013
summ_front_sd  0.0471 0.0128     0.0229   0.0468     0.0732 0.0055


Model hyperparameters:
                           mean   sd     0.025quant 0.5quant 0.975quant
T.0 for mesh.points-basisT  7.353  3.162  1.137      7.356   13.557
Expected number of effective parameters(std dev): 4.00(2.882e-05)
Number of equivalent replicates : 4367.37
Marginal Likelihood:  -87.00
```

A.1.3 Outer Ards

A.1.3.1 GLM Output using AIC for Selection

```
AIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore + chl_apr +
    chl_june + sst_may, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.77316  -0.02755  -0.01621  -0.00902   0.69414


Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 17.83795    7.34100   2.430  0.01510 *
dist_col    -0.08642    0.06001  -1.440  0.14988
dist_shore  -0.21399    0.13838  -1.546  0.12200
chl_apr     -1.56145    0.53550  -2.916  0.00355 **
chl_june    -1.07008    0.74078  -1.445  0.14859
sst_may     -1.32740    0.49672  -2.672  0.00753 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 127.053  on 25374  degrees of freedom
```

```
Residual deviance:  94.462  on 25369  degrees of freedom

AIC: 12.482

Number of Fisher Scoring iterations: 9
```

A.1.3.2 GLM Output using BIC for Selection

```
BIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ chl_apr + sst_may, family = "binomial",

    data = complete.data.to.analyse, weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.74926  -0.03023  -0.02159  -0.01476   0.88725


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)   5.0965     2.2975   2.218 0.026531 *

chl_apr      -0.7756     0.3386  -2.290 0.022005 *

sst_may      -0.7022     0.1881  -3.733 0.000189 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 127.05  on 25374  degrees of freedom

Residual deviance: 105.91  on 25372  degrees of freedom

AIC: 6.7872

Number of Fisher Scoring iterations: 7
```

A.1.3.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:

Call:

glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max
```

```
-0.76655  -0.02840  -0.01699  -0.01067   0.63367

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  7.80494    2.86999   2.720 0.006538 **
dist_col    -0.12616    0.05107  -2.470 0.013500 *
chl_apr     -1.20876    0.43496  -2.779 0.005453 **
sst_may     -0.69251    0.20513  -3.376 0.000736 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 127.053  on 25374  degrees of freedom
Residual deviance:  98.209  on 25371  degrees of freedom
AIC: 8.4015
Number of Fisher Scoring iterations: 7
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ dist_col + chl_apr + sst_may
         Df Deviance     AIC     LRT  Pr(>Chi)
<none>        98.209  8.4015
dist_col  1  105.907 14.0991  7.6975   0.00553 **
chl_apr   1  113.517 21.7098 15.3082 9.132e-05 ***
sst_may   1  115.942 24.1342 17.7327 2.542e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now checking for interactions with Year:

```
Model:
SEARCH_FORAGE ~ Year * (dist_col + chl_apr + sst_may)
              Df Deviance    AIC     LRT Pr(>Chi)
<none>             86.958 24.051
Year:dist_col  2   87.638 20.730 0.67937   0.7120
```

```
Year:chl_apr   2   87.542 20.634 0.58306   0.7471

Year:sst_may   2   89.873 22.965 2.91445   0.2329

No significant Year interactions.
```

A.1.3.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):

Family: binomial

Link function: logit

Formula:

SEARCH_FORAGE ~ s(dist_col, k = 3) + s(chl_apr, k = 3) + s(sst_april,

    k = 3)

Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  -3.4964     0.4115  -8.496   <2e-16 ***

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Approximate significance of smooth terms:

             edf Ref.df Chi.sq  p-value

s(dist_col)    1      1  6.504 0.010764 *

s(chl_apr)     1      1  7.885 0.004984 **

s(sst_april)   1      1 11.524 0.000687 ***

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.725   Deviance explained = 22.7%

REML score = 50.408  Scale est. = 1         n = 25375
```

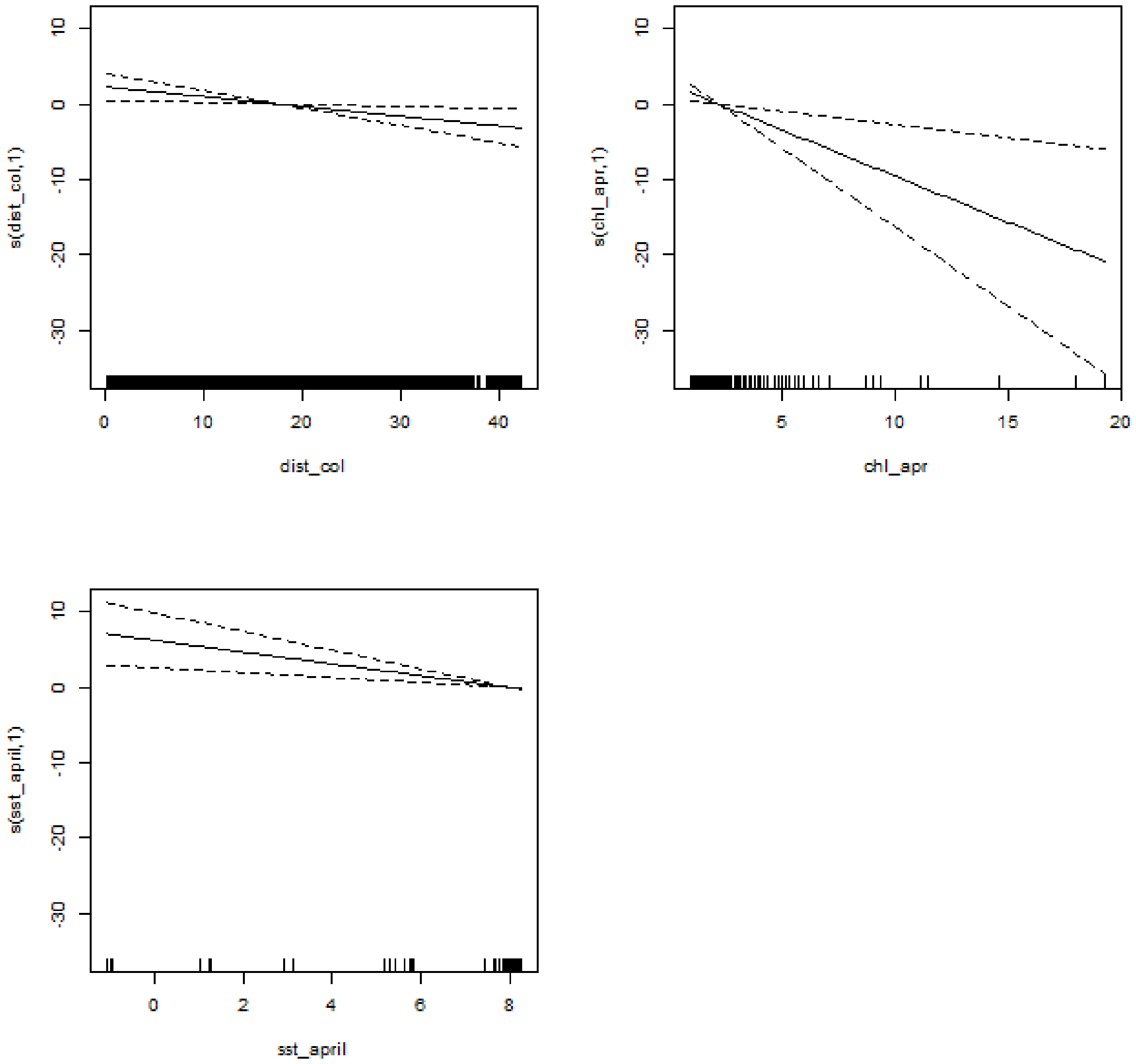


Figure A.1.3.4: GAM covariate plots for Arctic Terns in Outer Ards

A.1.3.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                 mean      sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)   8.1383 2.8296     3.0648   7.9583    14.1855 0.0120

dist_col    -0.1372 0.0510    -0.2441  -0.1348    -0.0437 0.0099
```

```
chl_apr     -1.3165 0.4372    -2.2435  -1.2916    -0.5249 0.0206
sst_april   -0.8058 0.2306    -1.3010  -0.7901    -0.3955 0.0051

Model hyperparameters:
                           mean   sd     0.025quant 0.5quant 0.975quant
T.0 for mesh.points-basisT  8.175  2.588  3.543      7.984   13.741
Expected number of effective parameters(std dev): 4.001(0.000108)
Number of equivalent replicates : 6341.73
Marginal Likelihood:  -52.33
```

A.2 Common Terns

A.2.1 Coquet

A.2.1.1 GLM Output using AIC for Selection

```
AIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col + chl_june + sst_april +
    ss_wave, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.93935  -0.04197  -0.01365  -0.00534   2.21575

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.77050    1.64251   1.687  0.09165 .
dist_col    -0.23482    0.04176  -5.623 1.88e-08 ***
chl_june     0.48865    0.12892   3.790  0.00015 ***
sst_april   -0.67033    0.21828  -3.071  0.00213 **
ss_wave     -0.07805    0.02523  -3.093  0.00198 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 430.38  on 29969  degrees of freedom
```

```
Residual deviance: 319.16  on 29965  degrees of freedom

AIC: 22.509

Number of Fisher Scoring iterations: 8
```

A.2.1.2 GLM Output using BIC for Selection

```
BIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_col, family = "binomial",

    data = complete.data.to.analyse, weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.75936  -0.04582  -0.01874  -0.00808   2.24320


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.92484    0.20783  -4.450 8.59e-06 ***

dist_col    -0.22190    0.03366  -6.592 4.33e-11 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 430.38  on 29969  degrees of freedom

Residual deviance: 340.30  on 29968  degrees of freedom

AIC: 16.789

Number of Fisher Scoring iterations: 8
```

A.2.1.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:

Call:

glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.93935  -0.04197  -0.01365  -0.00534   2.21575
```

```
Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  2.77050    1.64251   1.687  0.09165 .

dist_col    -0.23482    0.04176  -5.623 1.88e-08 ***

chl_june     0.48865    0.12892   3.790  0.00015 ***

sst_april   -0.67033    0.21828  -3.071  0.00213 **

ss_wave     -0.07805    0.02523  -3.093  0.00198 **

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 430.38  on 29969  degrees of freedom

Residual deviance: 319.16  on 29965  degrees of freedom

AIC: 22.509

Number of Fisher Scoring iterations: 8
```

Analysis of deviance output for reliable assessment of significance:

```
Model:

SEARCH_FORAGE ~ dist_col + chl_june + sst_april + ss_wave

          Df Deviance    AIC    LRT  Pr(>Chi)

<none>         319.16 22.509

dist_col   1   377.62 78.975 58.466 2.069e-14 ***

chl_june   1   333.87 35.216 14.707 0.0001256 ***

sst_april  1   328.26 29.611  9.102 0.0025532 **

ss_wave    1   331.82 33.174 12.664 0.0003727 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now checking for interactions with Year:

```
Model:

SEARCH_FORAGE ~ Year * (dist_col + chl_june + sst_april + ss_wave)

               Df Deviance    AIC    LRT Pr(>Chi)

<none>              308.41 40.566
```

```
Year:dist_col   2   313.61 41.762 5.1958  0.07443 .
Year:chl_june   2   308.92 37.076 0.5104  0.77476
Year:sst_april  2   308.77 36.921 0.3549  0.83738
Year:ss_wave    2   311.43 39.589 3.0232  0.22056
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Year effects not consistent - run one year at a time.
```

A.2.1.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):
Family: binomial
Link function: logit
Formula:
SEARCH_FORAGE ~ s(dist_col, k = 3) + s(chl_june, k = 3) + s(sst_april,
    k = 3) + s(ss_wave, k = 3)
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.1675     0.9453  -5.467 4.59e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
               edf Ref.df Chi.sq  p-value
s(dist_col)  1.629  1.862 17.739 0.000115 ***
s(chl_june)  1.000  1.000 12.278 0.000458 ***
s(sst_april) 1.453  1.701  9.607 0.005681 **
s(ss_wave)   1.000  1.000  5.993 0.014362 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) =    0.2   Deviance explained = 26.4%
REML score = 163.61  Scale est. = 1         n = 29970
```

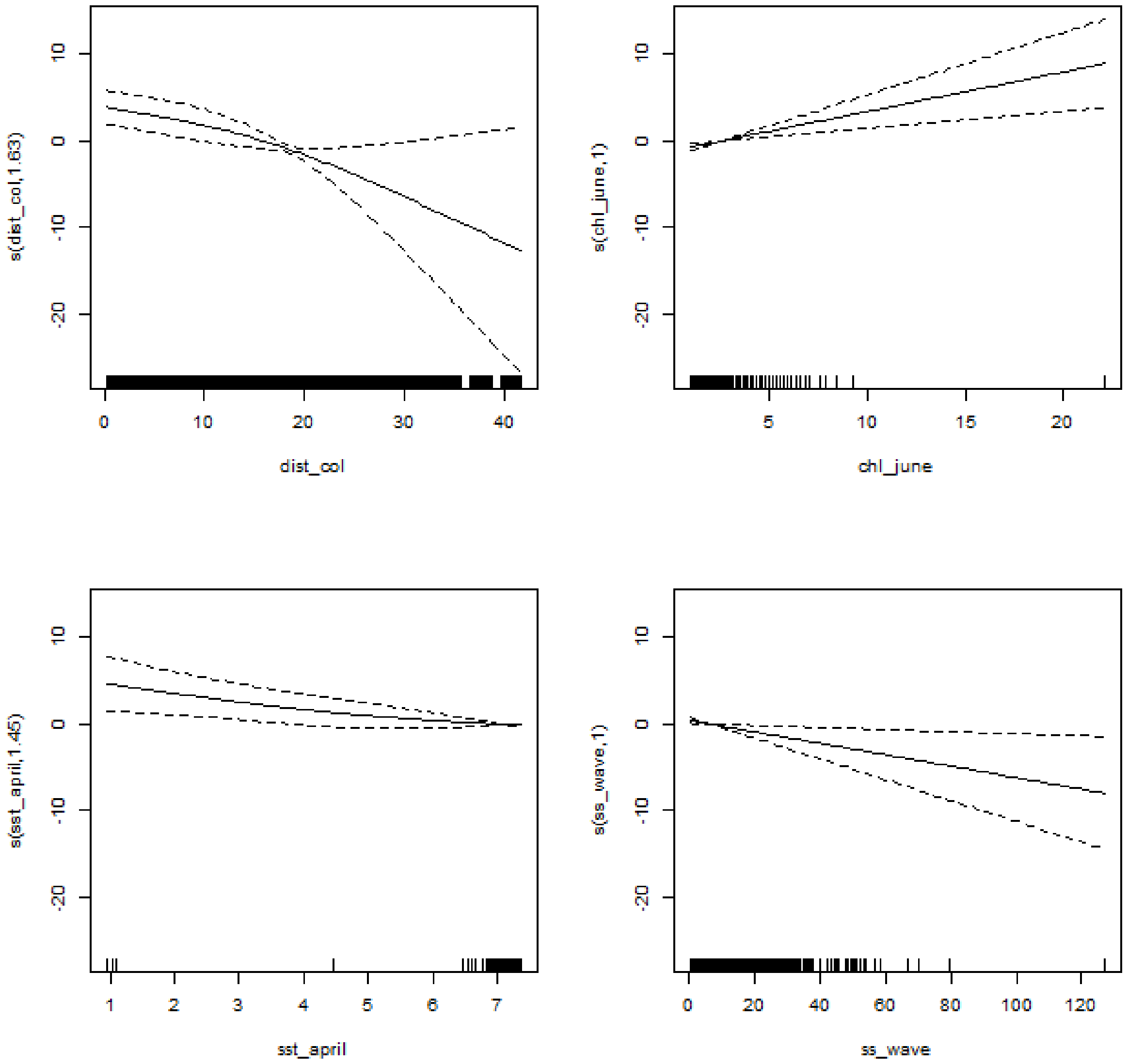


Figure A.2.1.4: GAM covariate plots for Common Terns in Coquet

A.2.1.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                mean      sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)   3.1326  1.9602    -0.3413   2.9926     7.3585 0.0000

chl_june      0.4647  0.1307     0.2149   0.4622     0.7280 0.0024
```

```
ss_wave       -0.0570  0.0270    -0.1140  -0.0555    -0.0077 0.0004

dist_col1    -16.7128  5.4688   -28.9271 -16.1196    -7.5617 0.0502

dist_col2    -20.8654 10.6971   -44.6611 -19.7513    -2.8460 0.0466

sst_april1   -10.1684  3.8209   -18.4069  -9.8953    -3.4036 0.0001

sst_april2    -0.5175  1.5212    -3.5177  -0.5112     2.4481 0.0004


Model hyperparameters:

                           mean   sd     0.025quant 0.5quant 0.975quant

T.0 for mesh.points-basisT  7.344  3.162  1.146      7.343   13.558

Expected number of effective parameters(std dev): 6.84(5.269e-05)

Number of equivalent replicates : 4381.76

Marginal Likelihood:  -155.88
```

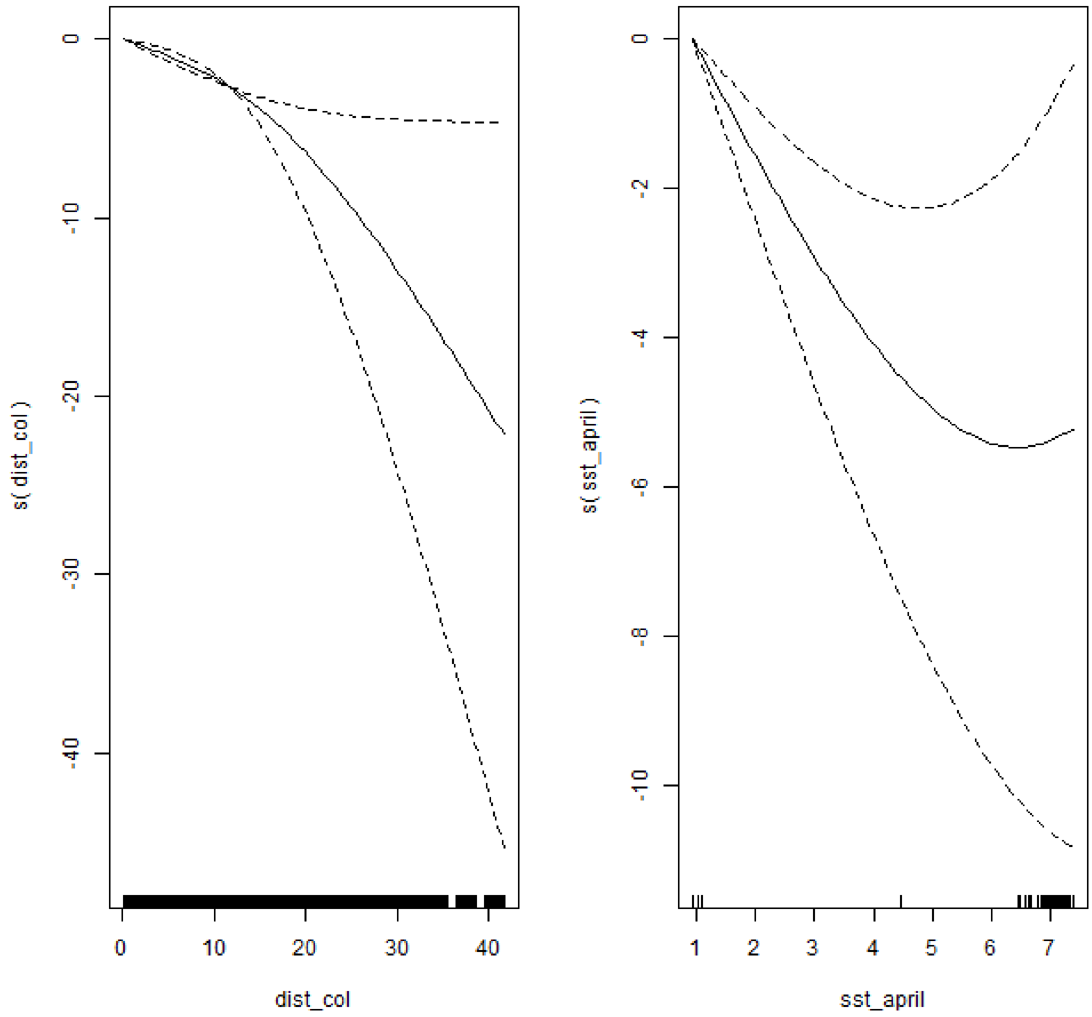


Figure A.2.1.5: INLA GAM covariate plots for Common Terns in Coquet (non-linear terms only)

A.2.2 Larne Lough

A.2.2.1 GLM Output using AIC for Selection

```
AIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore + sst_april +

    bathy_1sec, family = "binomial", data = complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.83858  -0.01706  -0.00799  -0.00033   0.81678


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  0.69907    2.44725   0.286   0.7751

dist_col    -1.02236    0.39887  -2.563   0.0104 *

dist_shore   1.01402    0.42836   2.367   0.0179 *

sst_april    0.54383    0.28704   1.895   0.0581 .

bathy_1sec   0.04195    0.01023   4.101 4.12e-05 ***

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 109.968  on 18056  degrees of freedom

Residual deviance:  56.951  on 18052  degrees of freedom

AIC: 10.953

Number of Fisher Scoring iterations: 10
```

A.2.2.2 GLM Output using BIC for Selection

```
BIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore + bathy_1sec,

    family = "binomial", data = complete.data.to.analyse, weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max
```

```
-1.06182  -0.01760  -0.00960  -0.00083   0.83903


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  2.85919    1.39032   2.056  0.03973 *

dist_col    -0.81257    0.34627  -2.347  0.01894 *

dist_shore   0.91903    0.37410   2.457  0.01402 *

bathy_1sec   0.04069    0.01073   3.792  0.00015 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 109.968  on 18056  degrees of freedom

Residual deviance:  66.235  on 18053  degrees of freedom

AIC: 10.43

Number of Fisher Scoring iterations: 10
```

A.2.2.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:

Call:

glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.83858  -0.01706  -0.00799  -0.00033   0.81678


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  0.69907    2.44725   0.286   0.7751

dist_col    -1.02236    0.39887  -2.563   0.0104 *

dist_shore   1.01402    0.42836   2.367   0.0179 *

sst_april    0.54383    0.28704   1.895   0.0581 .

bathy_1sec   0.04195    0.01023   4.101 4.12e-05 ***

---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 109.968  on 18056  degrees of freedom
Residual deviance:  56.951  on 18052  degrees of freedom
AIC: 10.953
Number of Fisher Scoring iterations: 10
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ dist_col + dist_shore + sst_april + bathy_1sec
           Df Deviance    AIC    LRT  Pr(>Chi)
<none>          56.951 10.953
dist_col    1   89.853 41.855 32.902 9.694e-09 ***
dist_shore  1   71.509 23.511 14.558 0.0001359 ***
sst_april   1   66.235 18.237  9.284 0.0023111 **
bathy_1sec  1   78.572 30.574 21.621 3.323e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now checking for interactions with Year:

```
Model:
SEARCH_FORAGE ~ Year * (dist_col + dist_shore + sst_april + bathy_1sec)
                Df Deviance    AIC     LRT Pr(>Chi)
<none>               54.520 30.587
Year:dist_col    2   54.802 26.870 0.28280   0.8681
Year:dist_shore  2   54.853 26.920 0.33289   0.8467
Year:sst_april   2   54.708 26.776 0.18842   0.9101
Year:bathy_1sec  2   54.550 26.617 0.03036   0.9849
No significant Year interactions.
```

A.2.2.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):
```

```
Family: binomial

Link function: logit


Formula:

SEARCH_FORAGE ~ s(dist_col, k = 3) + s(dist_shore, k = 3) + s(sst_april,

    k = 3) + s(bathy_1sec, k = 3)

Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)   -7.601      1.677  -4.533 5.82e-06 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:

                edf Ref.df Chi.sq  p-value

s(dist_col)   1.000  1.000  7.239  0.00713 **

s(dist_shore) 1.000  1.000  4.060  0.04391 *

s(sst_april)  1.000  1.000  4.744  0.02940 *

s(bathy_1sec) 1.875  1.984 18.665 8.65e-05 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.641   Deviance explained = 54.2%

REML score = 24.223  Scale est. = 1         n = 18057
```

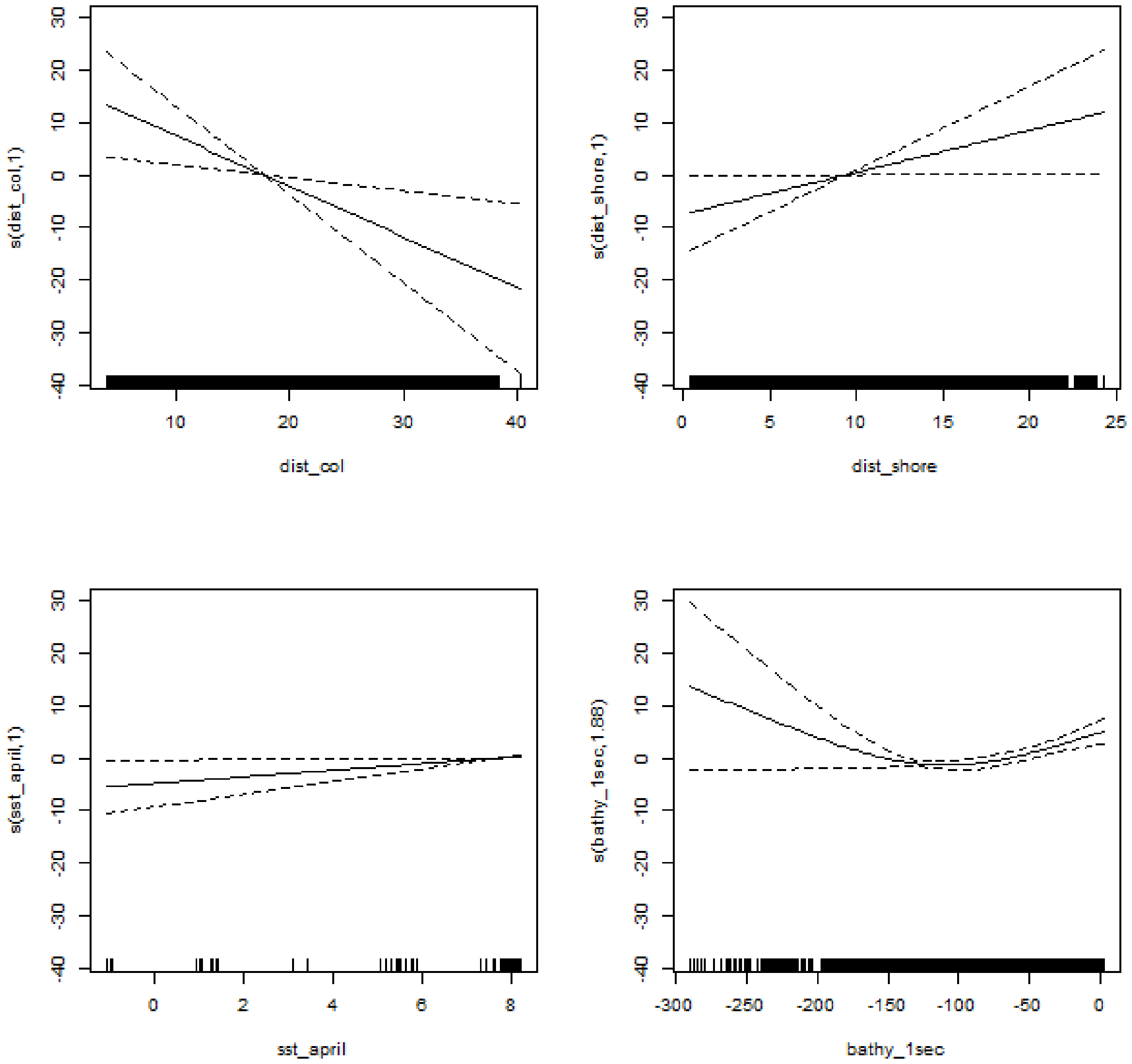


Figure A.2.2.4: GAM covariate plots for Common Terns in Larne Lough

A.2.2.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                mean      sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)  11.0152  8.6669    -7.0725  11.3943    26.9988 0.0034

dist_col     -1.0403  0.3609    -1.8549  -0.9966    -0.4494 0.0214
```

```
dist_shore     0.8705   0.3987     0.2148    0.8233     1.7681 0.0090

sst_april      0.6875   0.2883     0.2161    0.6523     1.3387 0.0239

bathy_1sec1 -27.9916 15.4704   -56.6066 -28.6403     4.2129 0.0003

bathy_1sec2    1.6795   3.0862    -3.8960    1.5006     8.2339 0.0104


Model hyperparameters:

                           mean   sd     0.025quant 0.5quant 0.975quant

T.0 for mesh.points-basisT  7.356  3.172  1.171      7.341   13.611

Expected number of effective parameters(std dev): 5.751(0.0006167)

Number of equivalent replicates : 3139.91

Marginal Likelihood:  -23.28
```

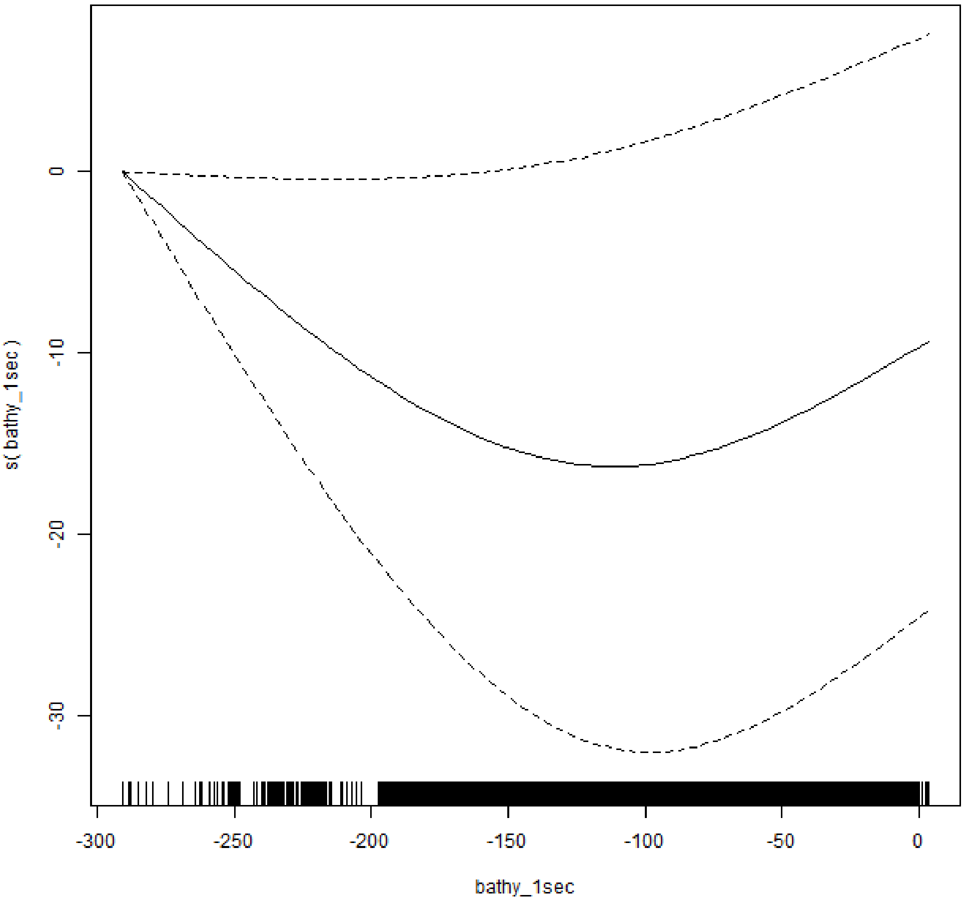


Figure A.2.2.5: INLA GAM covariate plots for Common Terns in Larne Lough (non-linear terms only)

A.2.3 Leith

A.2.3.1 GLM Output using AIC for Selection

```
AIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore + chl_may +

    chl_june + sst_may + summ_front + spring_front + sal_spring +

    bathy_1sec + slope_1s_deg, family = "binomial", data =
complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.85999  -0.04498  -0.02034  -0.00544   1.98457


Coefficients:

               Estimate Std. Error z value Pr(>|z|)

(Intercept)  -5.847e+02  2.060e+02  -2.838  0.00454 **

dist_col     -2.112e-01  4.369e-02  -4.833 1.35e-06 ***

dist_shore   -3.564e-01  1.362e-01  -2.618  0.00885 **

chl_may       1.721e-02  8.495e-03   2.026  0.04281 *

chl_june      2.571e-02  1.212e-02   2.122  0.03386 *

sst_may       1.146e-01  5.737e-02   1.998  0.04576 *

summ_front    6.155e-02  3.372e-02   1.825  0.06793 .

spring_front -1.292e-01  8.593e-02  -1.504  0.13257

sal_spring    1.661e+01  5.880e+00   2.824  0.00474 **

bathy_1sec    4.001e-02  1.830e-02   2.186  0.02880 *

slope_1s_deg  1.557e-01  8.391e-02   1.856  0.06344 .

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 726.30  on 57616  degrees of freedom

Residual deviance: 591.83  on 57606  degrees of freedom

AIC: 74.249

Number of Fisher Scoring iterations: 8
```

A.2.3.2 GLM Output using BIC for Selection

```
BIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col, family = "binomial",
    data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.76723  -0.04666  -0.02832  -0.01118   2.10430

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.87617    0.18912  -4.633 3.61e-06 ***
dist_col    -0.22386    0.03131  -7.149 8.72e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 726.30  on 57616  degrees of freedom
Residual deviance: 629.43  on 57615  degrees of freedom
AIC: 61.737
Number of Fisher Scoring iterations: 7
```

A.2.3.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.80966  -0.04530  -0.02140  -0.00603   2.02993

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.271e+02  2.018e+02  -2.612  0.00900 **
```

```
dist_col     -2.132e-01  4.310e-02  -4.947 7.52e-07 ***
dist_shore   -3.141e-01  1.339e-01  -2.346  0.01899 *
chl_may       1.570e-02  8.337e-03   1.884  0.05962 .
chl_june      2.208e-02  1.157e-02   1.909  0.05627 .
sst_may       1.171e-01  5.700e-02   2.055  0.03988 *
sal_spring    1.496e+01  5.759e+00   2.599  0.00936 **
bathy_1sec    3.496e-02  1.768e-02   1.977  0.04802 *
slope_1s_deg  1.621e-01  8.369e-02   1.937  0.05278 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 726.30  on 57616  degrees of freedom
Residual deviance: 596.51  on 57608  degrees of freedom
AIC: 70.309
Number of Fisher Scoring iterations: 8
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ dist_col + dist_shore + chl_may + chl_june +
    sst_may + sal_spring + bathy_1sec + slope_1s_deg
             Df Deviance     AIC    LRT  Pr(>Chi)
<none>            596.51  70.309
dist_col      1   640.09 111.887 43.577 4.075e-11 ***
dist_shore    1   602.38  74.172  5.862  0.015467 *
chl_may       1   600.08  71.874  3.564  0.059032 .
chl_june      1   599.98  71.771  3.462  0.062798 .
sst_may       1   600.86  72.652  4.343  0.037161 *
sal_spring    1   603.67  75.470  7.161  0.007451 **
bathy_1sec    1   600.85  72.648  4.339  0.037258 *
slope_1s_deg  1   600.07  71.869  3.560  0.059180 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now checking for interactions with Year:

```
Model:

SEARCH_FORAGE ~ Year * (dist_col + dist_shore + chl_may + chl_june +

    sst_may + sal_spring + bathy_1sec + slope_1s_deg)

                  Df Deviance    AIC    LRT Pr(>Chi)

<none>                 575.93 86.618

Year:dist_col      1   576.17 84.866 0.2485   0.6181

Year:dist_shore    1   578.43 87.124 2.5061   0.1134

Year:chl_may       1   577.60 86.291 1.6729   0.1959

Year:chl_june      1   576.83 85.518 0.9000   0.3428

Year:sst_may       1   579.45 88.144 3.5264   0.0604 .

Year:sal_spring    1   576.19 84.879 0.2607   0.6096

Year:bathy_1sec    1   576.69 85.385 0.7670   0.3812

Year:slope_1s_deg  1   576.18 84.874 0.2562   0.6128

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Year effects not consistent - run one year at a time.
```

A.2.3.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):

Family: binomial

Link function: logit

Formula:

SEARCH_FORAGE ~ s(dist_col, k = 3) + s(dist_shore, k = 3) + s(chl_may,

    k = 3) + s(chl_june, k = 3) + s(sst_may, k = 3) + s(sal_spring,

    k = 3) + s(bathy_1sec, k = 3) + s(slope_1s_deg, k = 3)

Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  -4.4153     0.6363  -6.939 3.96e-12 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:

                  edf Ref.df Chi.sq  p-value
```

```
s(dist_col)      1.604  1.843 21.652 1.56e-05 ***
s(dist_shore)    1.000  1.000  6.131   0.0133 *
s(chl_may)       1.000  1.000  3.534   0.0601 .
s(chl_june)      1.000  1.000  3.262   0.0709 .
s(sst_may)       1.000  1.000  3.376   0.0662 .
s(sal_spring)    1.000  1.000  5.481   0.0192 *
s(bathy_1sec)    1.511  1.761  4.944   0.0674 .
s(slope_1s_deg) 1.004  1.009  3.288   0.0707 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
R-sq.(adj) =  0.199   Deviance explained = 18.3%
REML score =  307.6  Scale est. = 1         n = 57617
```

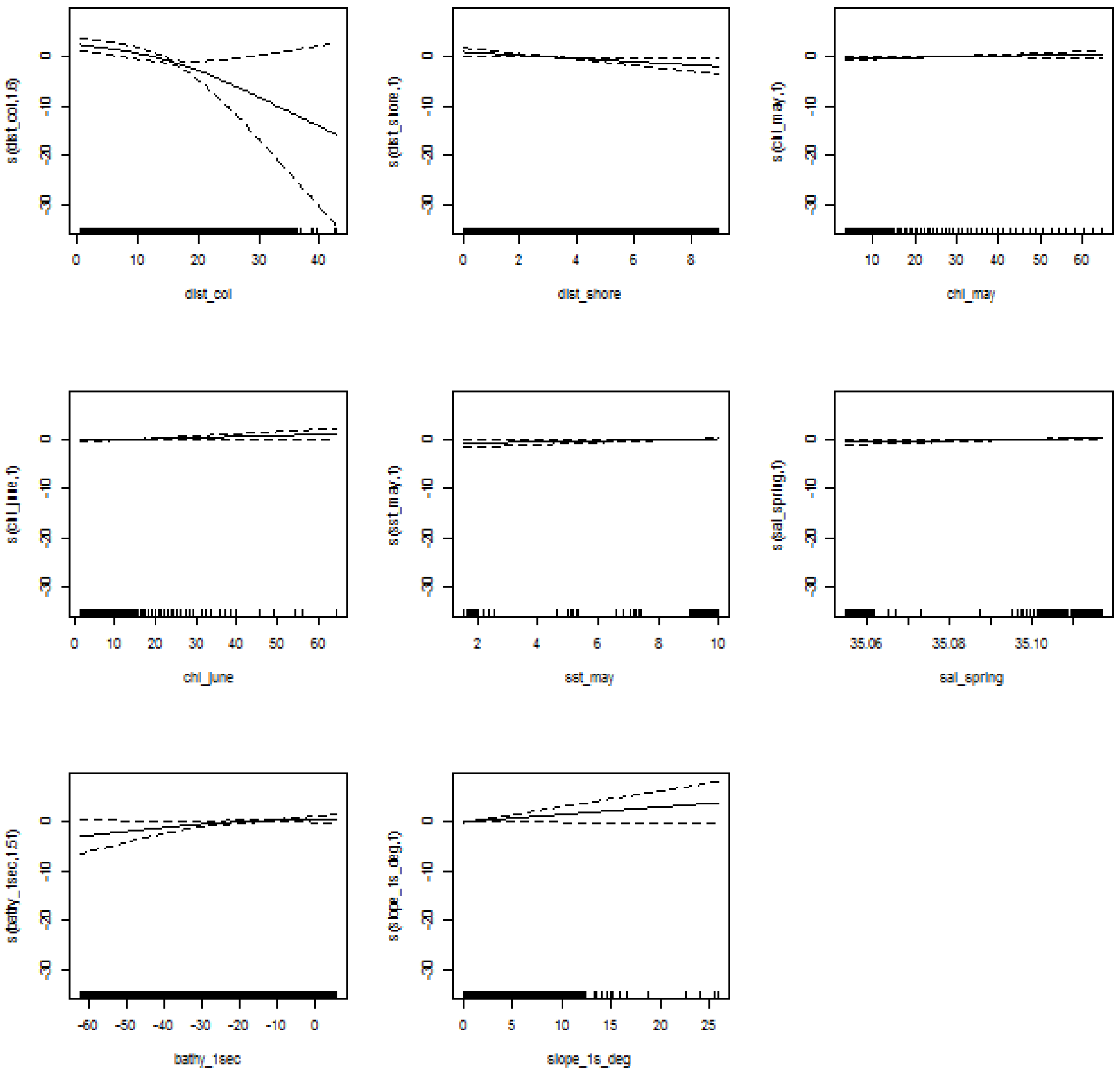


Figure A.2.3.4: GAM covariate plots for Common Terns in Leith

A.2.3.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                 mean      sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)  -15.5270 32.2965   -78.8599 -15.5262    47.8138 0.0000

dist_shore    -0.3160  0.1331    -0.5826  -0.3142    -0.0600 0.0004
```

```
chl_may         0.0153  0.0082    -0.0007   0.0153     0.0314 0.0011
chl_june        0.0246  0.0116     0.0015   0.0247     0.0471 0.0000
sst_may         0.1286  0.0550     0.0231   0.1278     0.2389 0.0001
sal_spring      0.1967  0.9223    -1.6119   0.1966     2.0057 0.0002
slope_1s_deg    0.1728  0.0857     0.0066   0.1721     0.3429 0.0009
dist_col1     -13.3207  4.7528   -23.8067 -12.8663    -5.2110 0.0024
dist_col2     -20.3520 10.6448   -43.6668 -19.4126    -1.9789 0.0024
bathy_1sec1     9.5332  4.6518     1.2654   9.2101    19.5414 0.0056
bathy_1sec2     1.0313  0.8747    -0.6310   1.0118     2.8038 0.0076

Model hyperparameters:
                           mean   sd     0.025quant 0.5quant 0.975quant
T.0 for mesh.points-basisT  7.159  3.001  1.194      7.195   12.961
Expected number of effective parameters(std dev): 9.865(0.0262)
Number of equivalent replicates : 5840.40
Marginal Likelihood:  -305.98
```

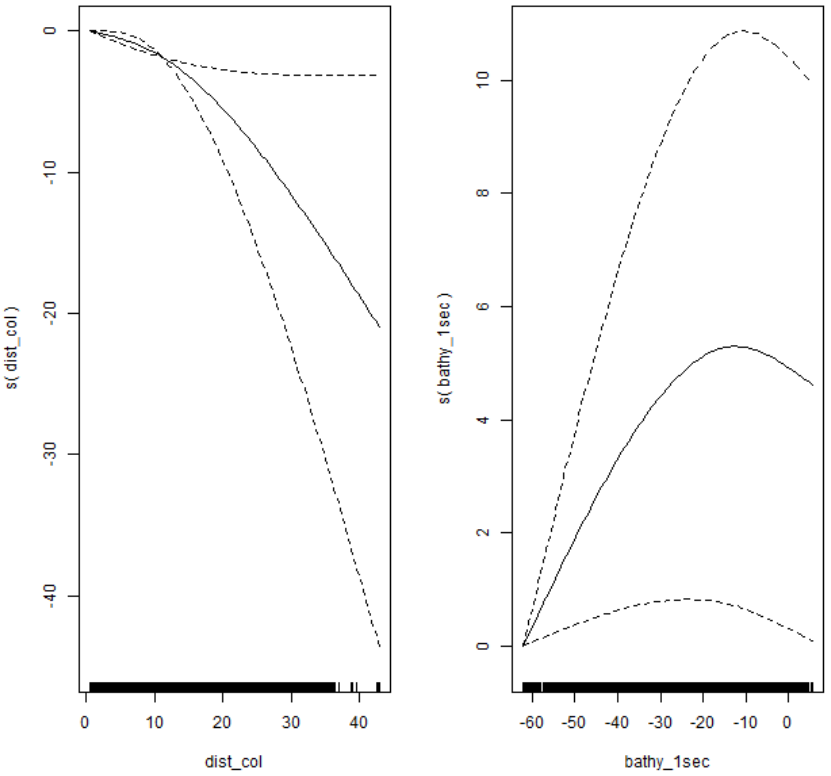


Figure A.2.3.5: INLA GAM covariate plots for Common Terns in Leith (non-linear terms only)

A.2.4 Mull

A.2.4.1 GLM Output using AIC for Selection

```
AIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ chl_apr + chl_may + sst_april +

    sst_may + ss_wave, family = "binomial", data = complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.83221  -0.03256  -0.00769  -0.00056   1.41323


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept) -5.12136    1.76890  -2.895  0.00379 **

chl_apr      0.05560    0.02499   2.225  0.02609 *

chl_may      0.02995    0.01629   1.839  0.06597 .

sst_april   -0.25411    0.18224  -1.394  0.16320

sst_may      0.41777    0.27283   1.531  0.12572

ss_wave     -3.43918    2.66105  -1.292  0.19621

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 91.876  on 7415  degrees of freedom

Residual deviance: 59.726  on 7410  degrees of freedom

AIC: 15.649

Number of Fisher Scoring iterations: 11
```

A.2.4.2 GLM Output using BIC for Selection

```
BIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ chl_apr, family = "binomial", data =
complete.data.to.analyse,

    weights = weights)
```

```
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.69210  -0.03973  -0.02633  -0.01780   1.75873

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.15516    0.58579  -7.093 1.31e-12 ***
chl_apr      0.06450    0.01543   4.180 2.92e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 91.876  on 7415  degrees of freedom
Residual deviance: 72.609  on 7414  degrees of freedom
AIC: 8.593
Number of Fisher Scoring iterations: 7
```

A.2.4.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.87668  -0.03694  -0.01094  -0.00133   1.46745

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.28231    0.82000  -4.003 6.26e-05 ***
chl_apr      0.03785    0.01768   2.141   0.0323 *
chl_may      0.02483    0.01421   1.748   0.0805 .
ss_wave     -2.71678    2.06153  -1.318   0.1876
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 91.876  on 7415  degrees of freedom
Residual deviance: 62.608  on 7412  degrees of freedom
AIC: 11.897
Number of Fisher Scoring iterations: 11
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ chl_apr + chl_may + ss_wave
        Df Deviance    AIC    LRT Pr(>Chi)
<none>       62.608 11.896
chl_apr  1   67.363 14.651 4.7544  0.02922 *
chl_may  1   65.544 12.832 2.9359  0.08663 .
ss_wave  1   67.669 14.957 5.0603  0.02448 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A.2.4.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):
Family: binomial
Link function: logit
Formula:
SEARCH_FORAGE ~ s(chl_apr, k = 3) + s(chl_may, k = 3) + s(bathy_1sec,
    k = 3)
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.5358     0.7937  -5.714  1.1e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf Ref.df Chi.sq p-value
s(chl_apr)    1.741  1.927 11.552 0.00283 **
```

```
s(chl_may)     1.613  1.845  5.461 0.05614 .
s(bathy_1sec) 1.799  1.958  8.269 0.01528 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
R-sq.(adj) =  0.329   Deviance explained = 40.1%
REML score = 31.679  Scale est. = 1         n = 7416
```

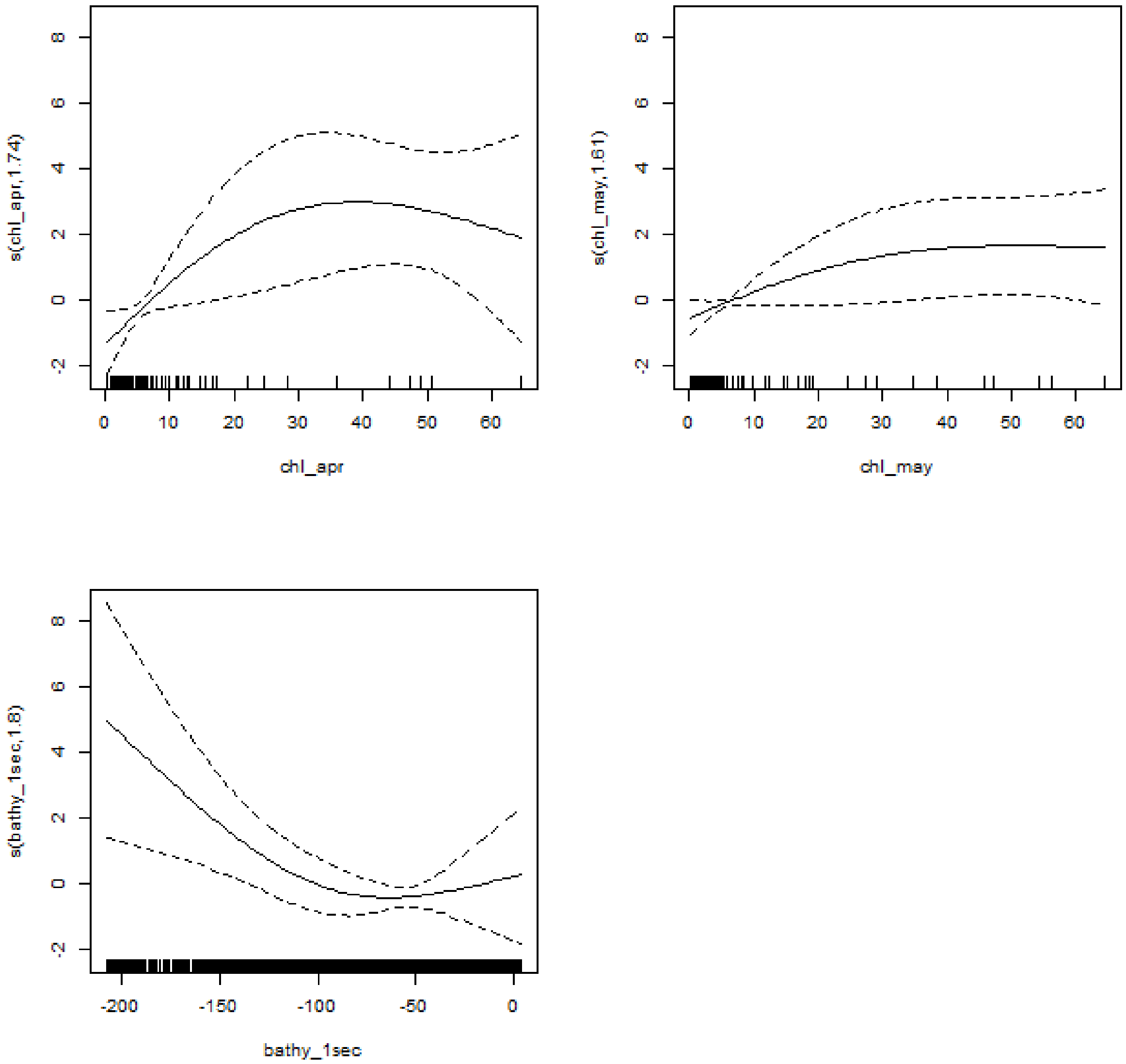


Figure A.2.4.4: GAM covariate plots for Common Terns in Mull

A.2.4.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                mean     sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)  -2.1257 1.6735    -5.7285  -2.0083     0.8434 0.0053

chl_apr1     10.8786 4.2873     3.3763  10.5368    20.1982 0.0213

chl_apr2      1.5918 1.8559    -1.9418   1.5528     5.3451 0.0039

chl_may1      5.3970 2.5115     0.8383   5.2607    10.7080 0.0216

chl_may2      1.6592 1.0339    -0.3856   1.6654     3.6693 0.0163

bathy_1sec1 -12.7551 4.2179   -21.6618 -12.5304    -5.0780 0.0420

bathy_1sec2  -0.2244 1.6295    -3.5639  -0.1747     2.8387 0.0001


Model hyperparameters:

                           mean    sd      0.025quant 0.5quant 0.975quant

T.0 for mesh.points-basisT  7.1401  3.1617  0.9245     7.1389  13.3516

Expected number of effective parameters(std dev): 6.95(1.466e-05)

Number of equivalent replicates : 1066.98

Marginal Likelihood:  -16.99
```

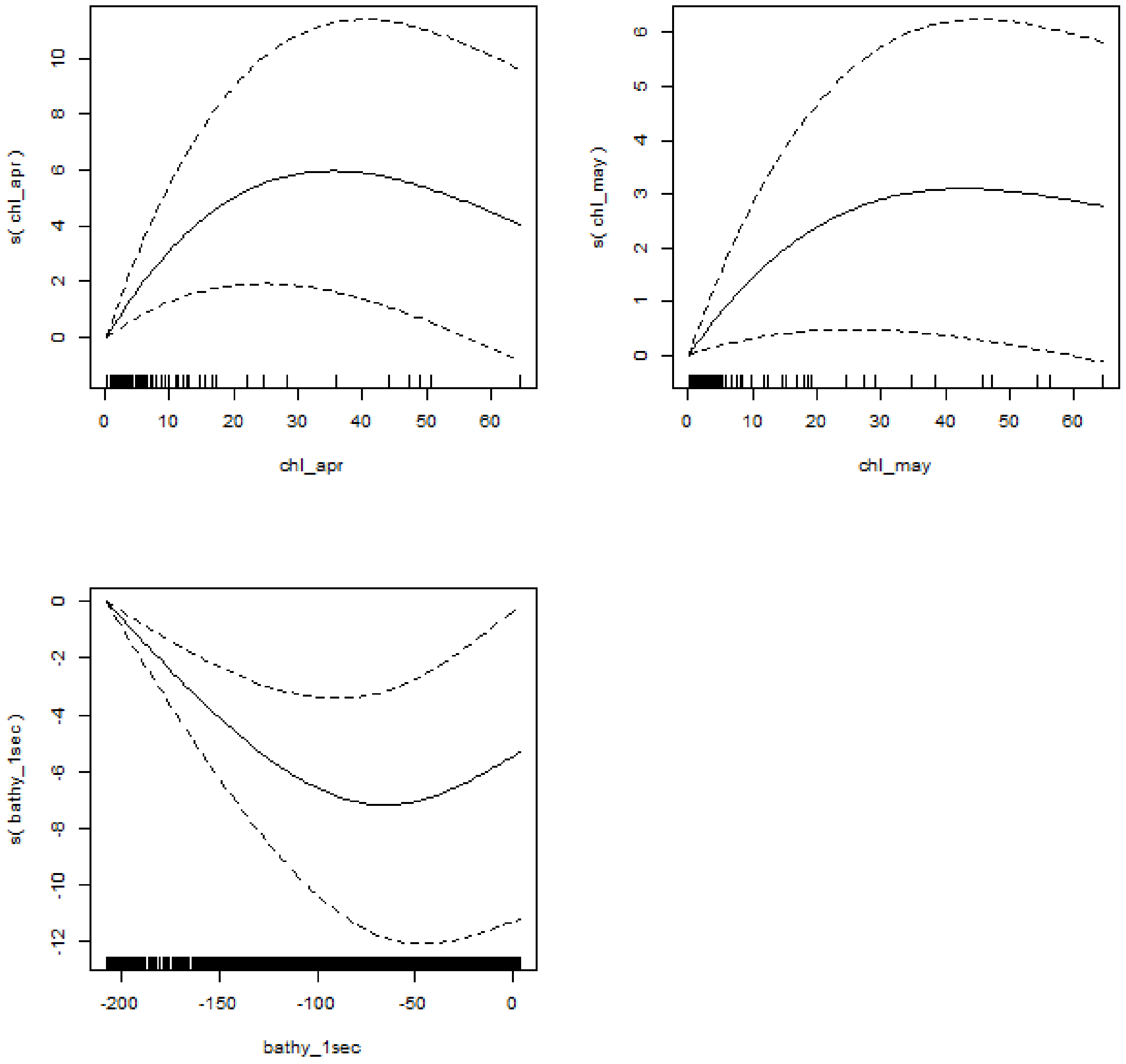


Figure A.2.4.5: INLA GAM covariate plots for Common Terns in Mull (non-linear terms only)

A.3 Roseate Terns

A.3.1 Coquet

A.3.1.1 GLM Output using AIC for Selection

```
AIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore + chl_may +

    sst_june + sst_may + summ_front_sd + strat_temp + ss_current,

    family = "binomial", data = complete.data.to.analyse, weights = weights)

Deviance Residuals:
```

```
     Min        1Q    Median        3Q       Max
-1.01134  -0.01394  -0.00392  -0.00076   1.25435

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.25604    8.35046  -0.749  0.45375
dist_col      -0.38296    0.12664  -3.024  0.00249 **
dist_shore    -0.99891    0.47034  -2.124  0.03369 *
chl_may       -0.13153    0.08050  -1.634  0.10229
sst_june      -5.83586    1.39780  -4.175 2.98e-05 ***
sst_may        6.28284    1.52123   4.130 3.63e-05 ***
summ_front_sd  0.06155    0.03089   1.993  0.04632 *
strat_temp     6.06231    2.92517   2.072  0.03822 *
ss_current    -2.33588    1.67821  -1.392  0.16396
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 207.04  on 24099  degrees of freedom
Residual deviance:  89.69  on 24091  degrees of freedom
AIC: 25.75
Number of Fisher Scoring iterations: 10
```

A.3.1.2 GLM Output using BIC for Selection

```
BIC Selected Model:

Call:
glm(formula = SEARCH_FORAGE ~ dist_col + sst_june + sst_may,
    family = "binomial", data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.09199  -0.01590  -0.00607  -0.00152   1.28785

Coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.36373    3.27227   0.722  0.47008
dist_col    -0.23490    0.07924  -2.964  0.00303 **
sst_june    -4.10827    0.73106  -5.620 1.91e-08 ***
sst_may      4.75048    0.87728   5.415 6.13e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 207.04  on 24099  degrees of freedom
Residual deviance: 102.27  on 24096  degrees of freedom
AIC: 18.038
Number of Fisher Scoring iterations: 9
```

A.3.1.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.18335  -0.01465  -0.00634  -0.00280   1.19303

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    8.53713    4.55846   1.873  0.06109 .
dist_col      -0.23031    0.08506  -2.707  0.00678 **
chl_may       -0.21875    0.09195  -2.379  0.01736 *
sst_june      -6.00578    1.33205  -4.509 6.52e-06 ***
sst_may        6.43205    1.40897   4.565 4.99e-06 ***
summ_front_sd  0.05353    0.02624   2.040  0.04132 *
chl_apr        0.30112    0.15868   1.898  0.05775 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 207.045  on 24099  degrees of freedom
Residual deviance:  91.709  on 24093  degrees of freedom
AIC: 22.587
Number of Fisher Scoring iterations: 9
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ dist_col + chl_may + sst_june + sst_may + summ_front_sd +
    chl_apr
              Df Deviance    AIC    LRT  Pr(>Chi)
<none>             91.709 22.587
dist_col       1  102.701 31.579 10.991 0.0009154 ***
chl_may        1   99.205 28.083  7.495 0.0061857 **
sst_june       1  123.777 52.655 32.068 1.489e-08 ***
sst_may        1  128.140 57.018 36.431 1.582e-09 ***
summ_front_sd  1   95.438 24.316  3.729 0.0534870 .
chl_apr        1   95.190 24.068  3.481 0.0620895 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now checking for interactions with Year:

```
Model:
SEARCH_FORAGE ~ Year * (dist_col + chl_may + sst_june + sst_may +
    summ_front_sd + chl_apr)
                   Df Deviance    AIC    LRT Pr(>Chi)
<none>                  78.172 49.101
Year:dist_col       2   78.598 45.527 0.4260   0.8082
Year:chl_may        2   80.133 47.062 1.9605   0.3752
Year:sst_june       2   81.030 47.959 2.8577   0.2396
Year:sst_may        2   82.214 49.143 4.0415   0.1326
Year:summ_front_sd  2   80.344 47.273 2.1719   0.3376
Year:chl_apr        2   78.421 45.350 0.2487   0.8831
```

```
No significant Year interactions.
```

A.3.1.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):

Family: binomial

Link function: logit

Formula:

SEARCH_FORAGE ~ s(dist_col, k = 3) + s(chl_apr, k = 3) + s(chl_may,

    k = 3) + s(sst_june, k = 3) + s(sst_may, k = 3) + s(summ_front_sd,

    k = 3)

Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  -5.6329     0.8417  -6.692  2.2e-11 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:

                 edf Ref.df Chi.sq  p-value

s(dist_col)        1      1  7.331  0.00678 **

s(chl_apr)         1      1  3.601  0.05775 .

s(chl_may)         1      1  5.659  0.01737 *

s(sst_june)        1      1 20.328 6.52e-06 ***

s(sst_may)         1      1 20.840 4.99e-06 ***

s(summ_front_sd)   1      1  4.163  0.04132 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.766   Deviance explained = 55.7%

REML score = 46.383  Scale est. = 1         n = 24100
```

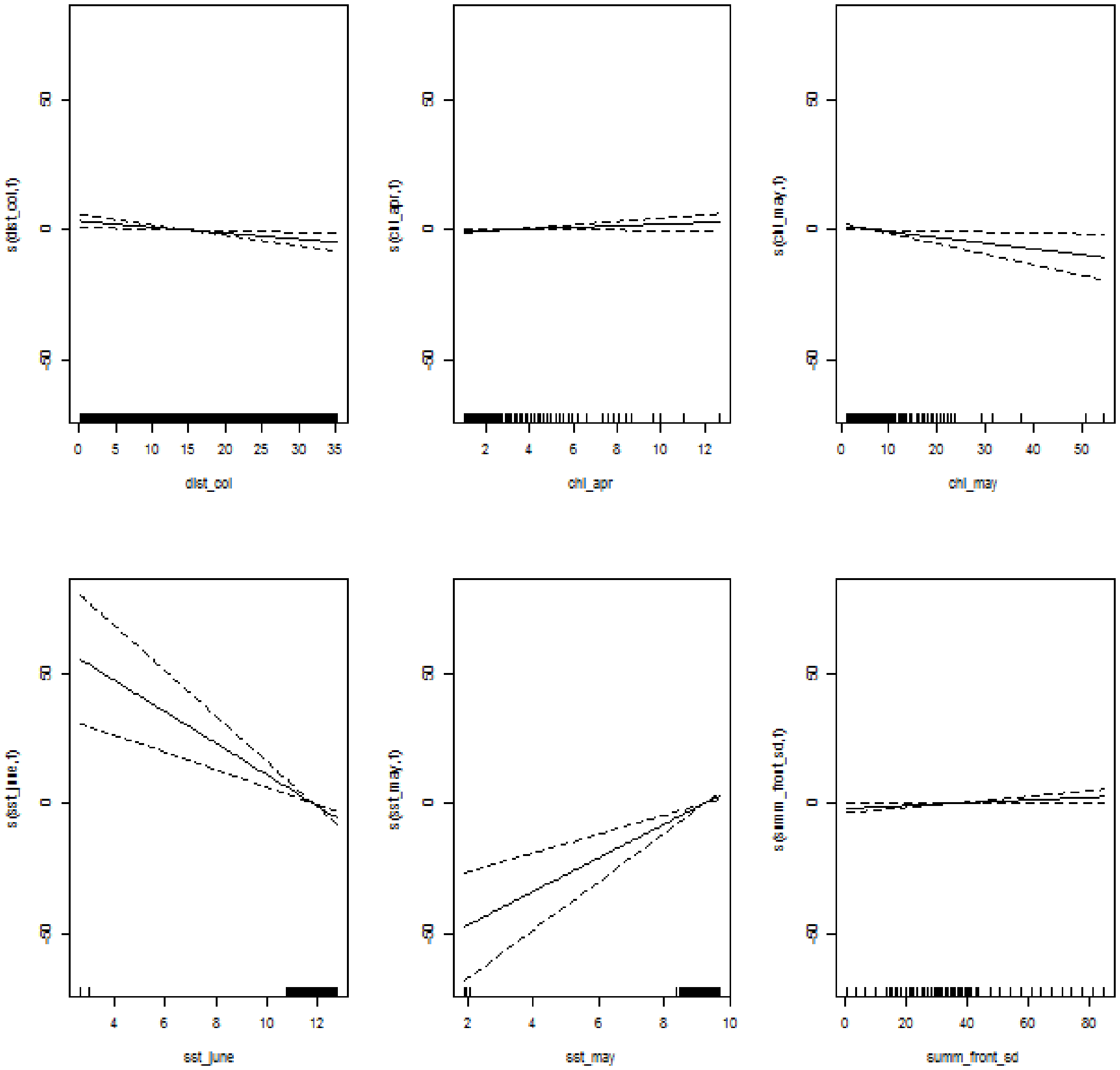


Figure A.3.1.4: GAM covariate plots for Roseate Terns in Coquet

A.3.1.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")


Fixed effects:

                   mean      sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)      7.8512 4.4905    -1.4586   8.0255    16.1874 0.0037

dist_col        -0.2359 0.0845    -0.4180  -0.2299    -0.0860 0.0036
```

```
chl_apr        0.3065 0.1579    -0.0034   0.3066     0.6158 0.0008
chl_may       -0.2116 0.0902    -0.4015  -0.2070    -0.0469 0.0000
sst_june      -5.9285 1.2768    -8.6319  -5.8577    -3.6127 0.0044
sst_may        6.5034 1.3495     4.0569   6.4286     9.3604 0.0221
summ_front_sd  0.0535 0.0258     0.0013   0.0540     0.1028 0.0047

Model hyperparameters:
                           mean  sd    0.025quant 0.5quant 0.975quant
T.0 for mesh.points-basisT 7.424 0.340 6.756      7.424    8.092
Expected number of effective parameters(std dev): 6.969(0.001662)
Number of equivalent replicates : 3458.37
Marginal Likelihood:  -55.62
```

A.4 Sandwich Terns

A.4.1 Coquet

A.4.1.1 GLM Output using AIC for Selection

```
AIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore + chl_june +
    sst_april, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.01383  -0.01093  -0.00155  -0.00016   2.28876

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.75429    0.77899   2.252 0.024322 *
dist_col    -0.09850    0.02570  -3.832 0.000127 ***
dist_shore  -0.25459    0.09668  -2.633 0.008459 **
chl_june     0.10420    0.05848   1.782 0.074752 .
sst_april   -0.38100    0.11822  -3.223 0.001270 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 447.26  on 55926  degrees of freedom
Residual deviance: 261.95  on 55922  degrees of freedom
AIC: 55.537
Number of Fisher Scoring iterations: 10
```

A.4.1.2 GLM Output using BIC for Selection

```
BIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore, family = "binomial",
    data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.93590  -0.00920  -0.00064  -0.00003   2.15522


Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.19595    0.29277   0.669 0.503311
dist_col    -0.08246    0.02221  -3.712 0.000206 ***
dist_shore  -0.44069    0.09469  -4.654 3.26e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 447.26  on 55926  degrees of freedom
Residual deviance: 274.44  on 55924  degrees of freedom
AIC: 57.509
Number of Fisher Scoring iterations: 10
```

A.4.1.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,
```

```
    weights = weights)


Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.01383  -0.01093  -0.00155  -0.00016   2.28876


Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.75429    0.77899   2.252 0.024322 *
dist_col    -0.09850    0.02570  -3.832 0.000127 ***
dist_shore  -0.25459    0.09668  -2.633 0.008459 **
chl_june     0.10420    0.05848   1.782 0.074752 .
sst_april   -0.38100    0.11822  -3.223 0.001270 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 447.26  on 55926  degrees of freedom
Residual deviance: 261.95  on 55922  degrees of freedom
AIC: 55.537
Number of Fisher Scoring iterations: 10
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ dist_col + dist_shore + chl_june + sst_april
           Df Deviance    AIC     LRT  Pr(>Chi)
<none>          261.95 55.537
dist_col    1   288.98 80.567 27.0300 2.003e-07 ***
dist_shore  1   271.63 63.225  9.6879 0.0018549 **
chl_june    1   264.94 56.530  2.9933 0.0836114 .
sst_april   1   272.80 64.384 10.8475 0.0009893 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now checking for interactions with Year:

```
Model:

SEARCH_FORAGE ~ Year * (dist_col + dist_shore + chl_june + sst_april)

                Df Deviance    AIC    LRT Pr(>Chi)

<none>               248.08 72.236

Year:dist_col    2   249.03 69.192 0.9556  0.62015

Year:dist_shore  2   254.07 74.230 5.9939  0.04994 *

Year:chl_june    2   248.40 68.556 0.3202  0.85206

Year:sst_april   2   248.14 68.295 0.0592  0.97084

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Year effects not consistent - run one year at a time.
```

A.4.1.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):

Family: binomial

Link function: logit

Formula:

SEARCH_FORAGE ~ s(dist_col, k = 3) + s(dist_shore, k = 3) + s(chl_june,

    k = 3) + s(sst_april, k = 3)


Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  -13.913      4.196  -3.315 0.000915 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:

                edf Ref.df Chi.sq p-value

s(dist_col)   1.804  1.959  7.345 0.02434 *

s(dist_shore) 1.000  1.000  8.042 0.00457 **

s(chl_june)   1.000  1.000  1.530 0.21619

s(sst_april)  1.736  1.928 10.991 0.00376 **
```

```
---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


R-sq.(adj) =  0.461   Deviance explained = 42.9%

REML score = 132.43  Scale est. = 1         n = 55927
```

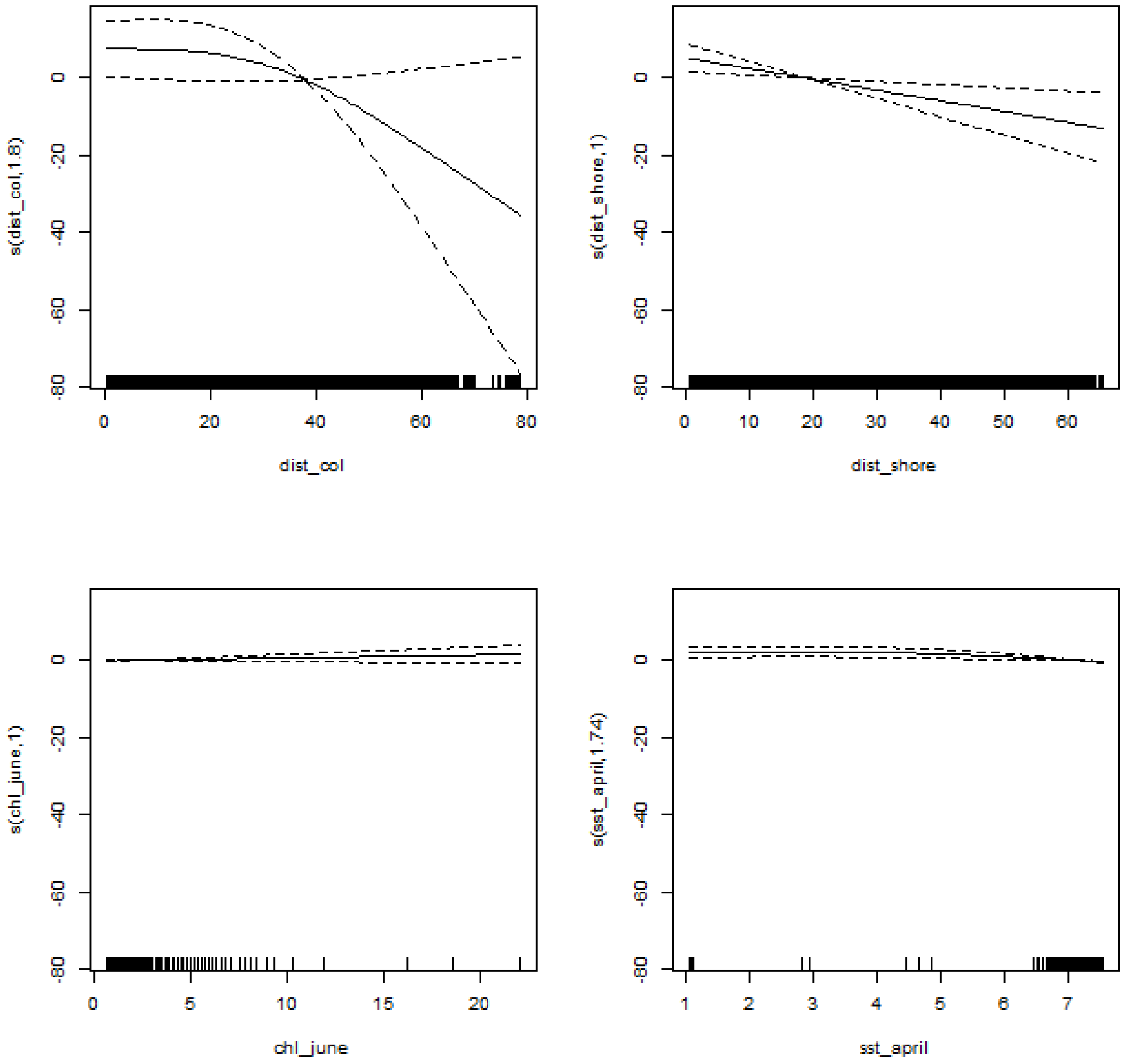


Figure A.4.1.4: GAM covariate plots for Sandwich Terns in Coquet

A.4.1.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")

Fixed effects:
```

```
                mean      sd 0.025quant 0.5quant 0.975quant    kld
(Intercept)   0.4786  0.8183    -1.1047   0.4706     2.1073 0.0028
dist_shore   -0.2600  0.0989    -0.4706  -0.2542    -0.0808 0.0030
chl_june      0.0771  0.0614    -0.0439   0.0773     0.1972 0.0022
dist_col1   -20.7704  8.5412   -37.6322 -20.6781    -4.0885 0.0128
dist_col2   -33.6238 16.2892   -65.5195 -33.5422    -1.7116 0.0154
sst_april1   -1.4888  1.6015    -4.6848  -1.4706     1.6020 0.0002
sst_april2   -3.2350  1.2110    -5.6804  -3.2115    -0.9226 0.0016

Model hyperparameters:
                           mean   sd     0.025quant 0.5quant 0.975quant
T.0 for mesh.points-basisT  8.034  3.144  1.830      8.052   14.162

Expected number of effective parameters(std dev): 6.661(4.235e-05)
Number of equivalent replicates : 8395.72
Marginal Likelihood:  -126.07
```

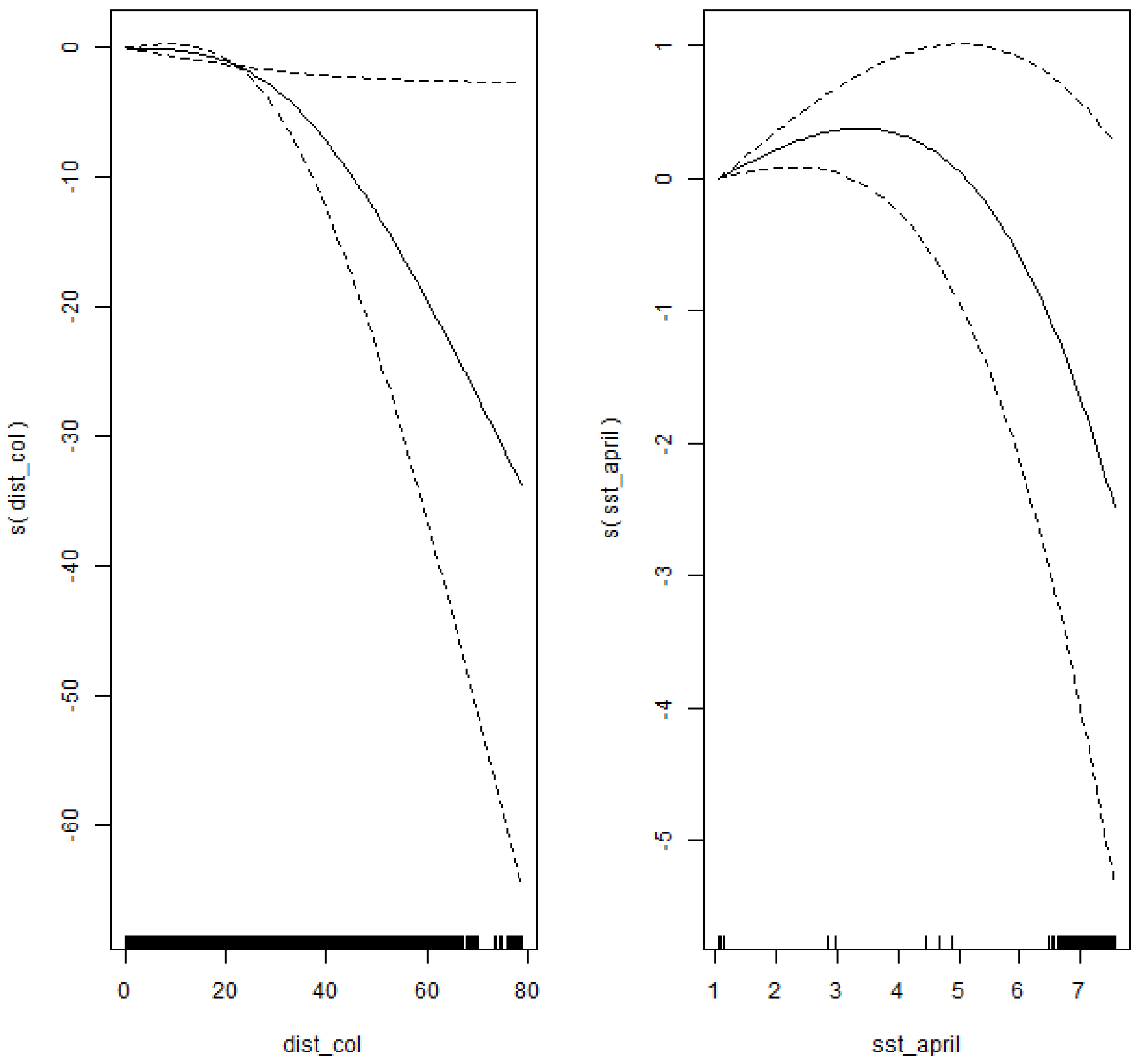


Figure A.4.1.5: INLA GAM covariate plots for Sandwich Terns in Coquet (non-linear terms only)

A.4.2 Farnes

A.4.2.1 GLM Output using AIC for Selection

```
AIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore + summ_front +

    spring_front + ss_wave, family = "binomial", data = complete.data.to.analyse,

    weights = weights)

Deviance Residuals:

    Min       1Q   Median       3Q      Max
```

```
-0.5551   0.0000   0.0000   0.0000   1.3428


Coefficients:

             Estimate Std. Error z value Pr(>|z|)

(Intercept)   1.68802    1.67480   1.008  0.31351

dist_col     -0.16160    0.11039  -1.464  0.14321

dist_shore   -3.51590    0.97701  -3.599  0.00032 ***

summ_front    0.28026    0.11436   2.451  0.01426 *

spring_front -0.46366    0.19019  -2.438  0.01477 *

ss_wave       0.04123    0.01939   2.126  0.03353 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 160.42  on 15038  degrees of freedom

Residual deviance:  30.82  on 15033  degrees of freedom

AIC: 12.472


Number of Fisher Scoring iterations: 14
```

A.4.2.2 GLM Output using BIC for Selection

```
BIC Selected Model:

Call:

glm(formula = SEARCH_FORAGE ~ dist_shore, family = "binomial",

    data = complete.data.to.analyse, weights = weights)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-0.64514  -0.00004   0.00000   0.00000   1.63006


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)   3.1704     0.9034   3.509 0.000449 ***

dist_shore   -3.1710     0.7691  -4.123 3.74e-05 ***

---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 160.421  on 15038  degrees of freedom
Residual deviance:  48.662  on 15037  degrees of freedom
AIC: 4.3857
Number of Fisher Scoring iterations: 13
```

A.4.2.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse,
    weights = weights)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5551   0.0000   0.0000   0.0000   1.3428

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.68802    1.67480   1.008  0.31351
dist_col     -0.16160    0.11039  -1.464  0.14321
dist_shore   -3.51590    0.97701  -3.599  0.00032 ***
summ_front    0.28026    0.11436   2.451  0.01426 *
spring_front -0.46366    0.19019  -2.438  0.01477 *
ss_wave       0.04123    0.01939   2.126  0.03353 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 160.42  on 15038  degrees of freedom
Residual deviance:  30.82  on 15033  degrees of freedom
AIC: 12.472
Number of Fisher Scoring iterations: 14
```

Analysis of deviance output for reliable assessment of significance:

```
Model:

SEARCH_FORAGE ~ dist_col + dist_shore + summ_front + spring_front +

    ss_wave

             Df Deviance    AIC    LRT  Pr(>Chi)

<none>            30.820 12.472

dist_col      1   33.919 13.570  3.098  0.078382 .

dist_shore    1   65.633 45.285 34.812 3.631e-09 ***

summ_front    1   38.920 18.572  8.100  0.004427 **

spring_front  1   40.298 19.949  9.477  0.002081 **

ss_wave       1   35.808 15.460  4.988  0.025527 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A.4.2.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):

Family: binomial

Link function: logit

Formula:

SEARCH_FORAGE ~ s(dist_shore, k = 3) + s(summ_front, k = 3) +

    s(spring_front, k = 3) + s(ss_wave, k = 3)


Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)   -68.53      19.45  -3.524 0.000425 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:

                edf Ref.df Chi.sq p-value

s(dist_shore)     1      1 11.656 0.00064 ***

s(summ_front)     1      1  4.600 0.03198 *

s(spring_front)   1      1  4.771 0.02895 *

s(ss_wave)        1      1  3.722 0.05369 .
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.857   Deviance explained = 78.9%

REML score = 11.781  Scale est. = 1         n = 15039
```

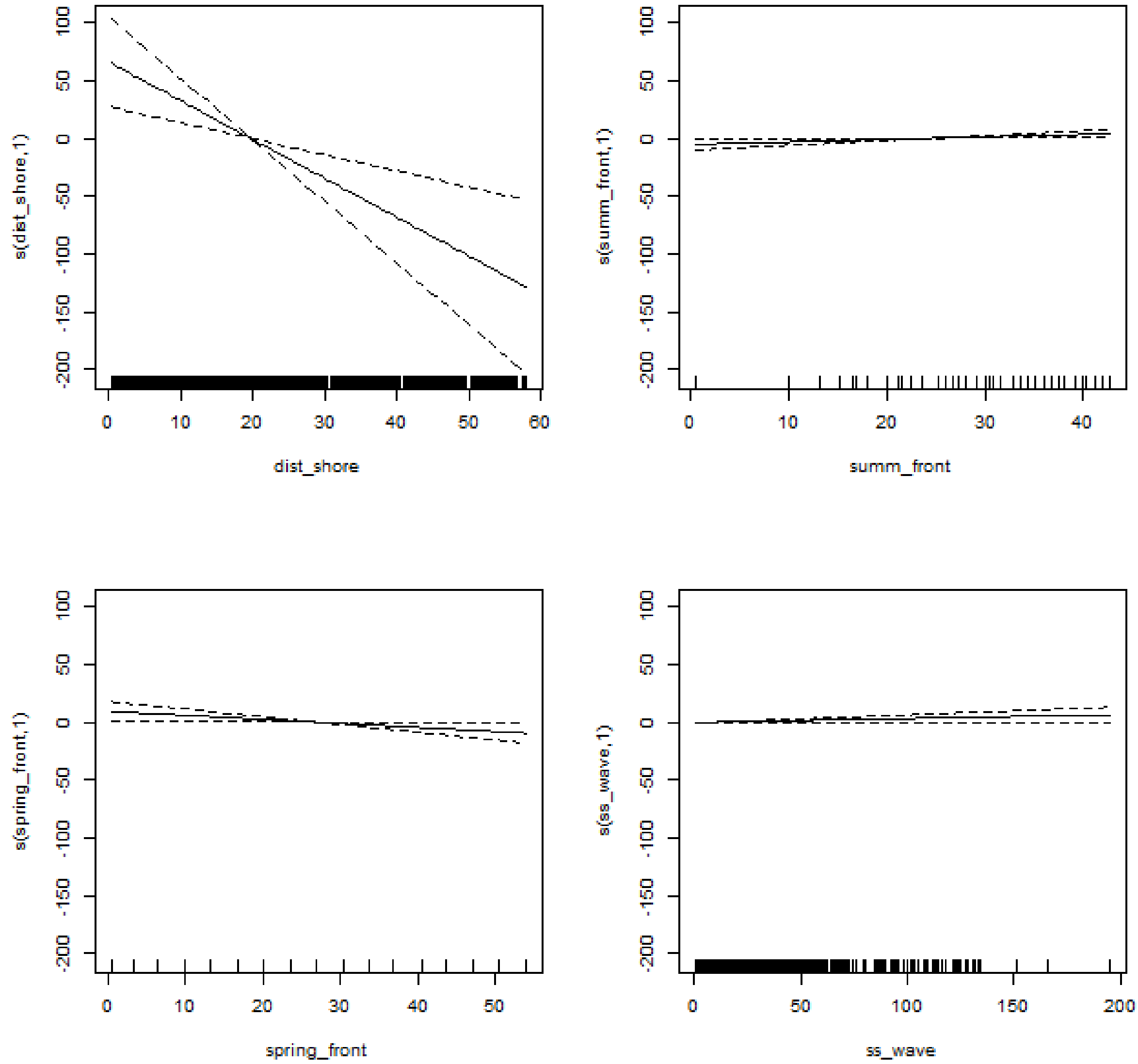


Figure A.4.2.4: GAM covariate plots for Sandwich Terns in Farnes

A.4.2.5 INLA Output

```
Running INLA.


Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")
```

```
Fixed effects:
                mean     sd 0.025quant 0.5quant 0.975quant    kld
(Intercept)   1.3808 1.6987    -1.6641   1.2727     5.0122 0.0100
dist_shore   -3.7858 0.9796    -5.9366  -3.6982    -2.0961 0.0986
summ_front    0.2412 0.1001     0.0609   0.2352     0.4542 0.0355
spring_front -0.3997 0.1590    -0.7426  -0.3883    -0.1185 0.0542
ss_wave       0.0359 0.0177     0.0027   0.0354     0.0723 0.0045

Model hyperparameters:
                           mean   sd     0.025quant 0.5quant 0.975quant
T.0 for mesh.points-basisT  7.994  3.099  2.017      7.952   14.167

Expected number of effective parameters(std dev): 4.994(0.001232)
Number of equivalent replicates : 3011.70
Marginal Likelihood:  -23.58
```

A.4.3 Forvie

A.4.3.1 GLM Output using AIC for Selection

```
AIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_shore + sst_june + strat_temp,
    family = "binomial", data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1334   0.0000   0.0000   0.0000   0.7863


Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -94.133     69.710  -1.350  0.17690
dist_shore    -3.254      1.138  -2.860  0.00424 **
sst_june       9.518      6.536   1.456  0.14535
strat_temp    -7.090      3.533  -2.007  0.04474 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 97.836  on 26516  degrees of freedom
Residual deviance: 25.004  on 26513  degrees of freedom
AIC: 9.2846

Number of Fisher Scoring iterations: 14
```

A.4.3.2 GLM Output using BIC for Selection

```
BIC Selected Model:

Call:
glm(formula = SEARCH_FORAGE ~ dist_shore + strat_temp, family = "binomial", 
    data = complete.data.to.analyse, weights = weights)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-1.5768   0.0000   0.0000   0.0000   0.8352  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)   
(Intercept)   7.8379     3.0822   2.543  0.01099 * 
dist_shore   -2.7111     0.9429  -2.875  0.00404 **
strat_temp   -4.6219     2.5596  -1.806  0.07096 . 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 97.836  on 26516  degrees of freedom
Residual deviance: 28.799  on 26514  degrees of freedom
AIC: 8.4862

Number of Fisher Scoring iterations: 14
```

A.1.3.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:

Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse, 
```

```
    weights = weights)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1334   0.0000   0.0000   0.0000   0.7863

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -94.133     69.710  -1.350  0.17690
dist_shore    -3.254      1.138  -2.860  0.00424 **
sst_june       9.518      6.536   1.456  0.14535
strat_temp    -7.090      3.533  -2.007  0.04474 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 97.836  on 26516  degrees of freedom
Residual deviance: 25.004  on 26513  degrees of freedom
AIC: 9.2846
Number of Fisher Scoring iterations: 14
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ dist_shore + sst_june + strat_temp
           Df Deviance    AIC     LRT  Pr(>Chi)
<none>          25.004  9.285
dist_shore  1   50.042 32.323 25.0380 5.621e-07 ***
sst_june    1   28.799 11.080  3.7953 0.0513980 .
strat_temp  1   38.888 21.169 13.8845 0.0001944 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A.1.3.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):
```

```
Family: binomial

Link function: logit

Formula:

SEARCH_FORAGE ~ s(dist_shore, k = 3) + s(strat_temp, k = 3)


Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)   -58.15      18.89  -3.078  0.00208 **

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Approximate significance of smooth terms:

              edf Ref.df Chi.sq p-value

s(dist_shore)   1      1  8.267 0.00404 **

s(strat_temp)   1      1  3.256 0.07118 .

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.324   Deviance explained = 70.6%

REML score = 8.7164  Scale est. = 1         n = 26517
```

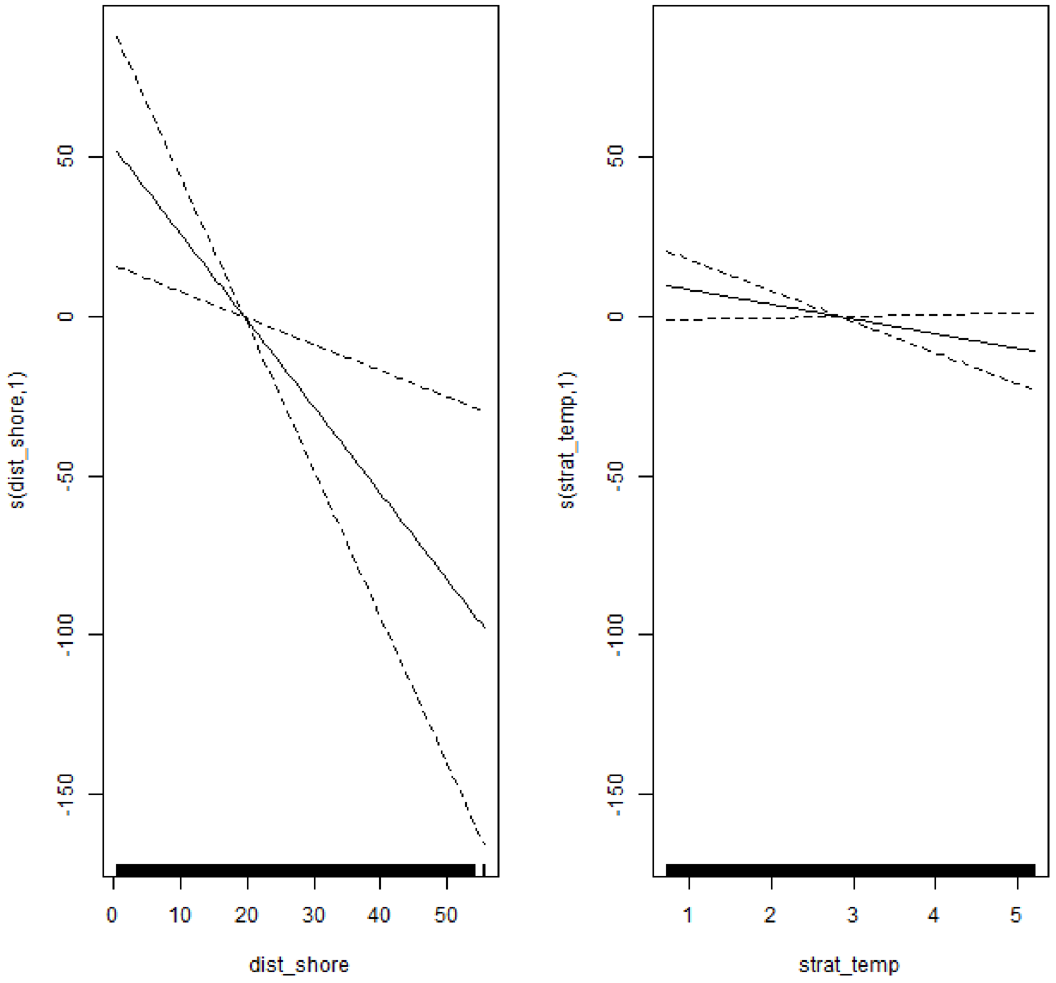


Figure A.4.3.4: GAM covariate plots for Sandwich Terns in Forvie

A.4.3.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")

Fixed effects:

                mean     sd 0.025quant  0.5quant 0.975quant    kld

(Intercept)  8.5092 3.0521     2.9740    8.3420    14.9486 0.0285

dist_shore  -2.8564 0.9388    -4.8381   -2.8066    -1.1474 0.0138

strat_temp  -4.7950 2.5331   -10.3185   -4.5855    -0.3814 0.0037


Model hyperparameters:
```

```
                           mean   sd     0.025quant 0.5quant 0.975quant
T.0 for mesh.points-basisT 10.175  1.743  7.584      9.931   13.822
Expected number of effective parameters(std dev): 2.992(0.002468)
Number of equivalent replicates : 8861.99
Marginal Likelihood:  -13.50
```

A.4.4 Larne Lough

A.4.4.1 GLM Output using AIC for Selection

```
AIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore, family = "binomial",
    data = complete.data.to.analyse, weights = weights)
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.32521  -0.00005   0.00000   0.00000   0.73625

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.64766    0.51868   3.177  0.00149 **
dist_col    -0.13284    0.06301  -2.108  0.03499 *
dist_shore  -3.54580    1.15139  -3.080  0.00207 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 210.722  on 29785  degrees of freedom
Residual deviance:  69.594  on 29783  degrees of freedom
AIC: 7.7562
Number of Fisher Scoring iterations: 12
```

A.4.4.2 GLM Output using BIC for Selection

```
BIC Selected Model:
Call:
glm(formula = SEARCH_FORAGE ~ dist_col + dist_shore, family = "binomial",
```

```
    data = complete.data.to.analyse, weights = weights)
Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-1.32521  -0.00005   0.00000   0.00000   0.73625  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)   
(Intercept)  1.64766    0.51868   3.177  0.00149 **
dist_col    -0.13284    0.06301  -2.108  0.03499 * 
dist_shore  -3.54580    1.15139  -3.080  0.00207 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 210.722  on 29785  degrees of freedom
Residual deviance:  69.594  on 29783  degrees of freedom
AIC: 7.7562
Number of Fisher Scoring iterations: 12
```

A.4.4.3 GLM Output using Likelihood Ratio Tests for Selection

```
LRT Selected Model:
Call:
glm(formula = formula.glm, family = "binomial", data = complete.data.to.analyse, 
    weights = weights)
Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-1.32521  -0.00005   0.00000   0.00000   0.73625  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)   
(Intercept)  1.64766    0.51868   3.177  0.00149 **
dist_col    -0.13284    0.06301  -2.108  0.03499 * 
dist_shore  -3.54580    1.15139  -3.080  0.00207 **
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 210.722  on 29785  degrees of freedom
Residual deviance:  69.594  on 29783  degrees of freedom
AIC: 7.7562
Number of Fisher Scoring iterations: 12
```

Analysis of deviance output for reliable assessment of significance:

```
Model:
SEARCH_FORAGE ~ dist_col + dist_shore
           Df Deviance    AIC    LRT  Pr(>Chi)
<none>          69.594  7.756
dist_col    1   90.483 26.644 20.888 4.869e-06 ***
dist_shore  1   97.727 33.889 28.133 1.133e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Now checking for interactions with Year:

```
Model:
SEARCH_FORAGE ~ Year * (dist_col + dist_shore)
                Df Deviance    AIC     LRT Pr(>Chi)
<none>               65.637 19.885
Year:dist_col    2   66.433 16.681 0.79613   0.6716
Year:dist_shore  2   68.425 18.673 2.78832   0.2480
No significant Year interactions.
```

A.4.4.4 GAM Output using Likelihood Ratio Tests for Selection

```
GAM Model selected (REML output):
Family: binomial
Link function: logit
Formula:
SEARCH_FORAGE ~ s(dist_col, k = 3) + s(dist_shore, k = 3)
```

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -29.854      8.843  -3.376 0.000735 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(dist_col)     1      1  4.444 0.03503 *
s(dist_shore)   1      1  9.481 0.00208 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) =  0.496   Deviance explained =   67%
REML score = 31.248  Scale est. = 1         n = 29786
```

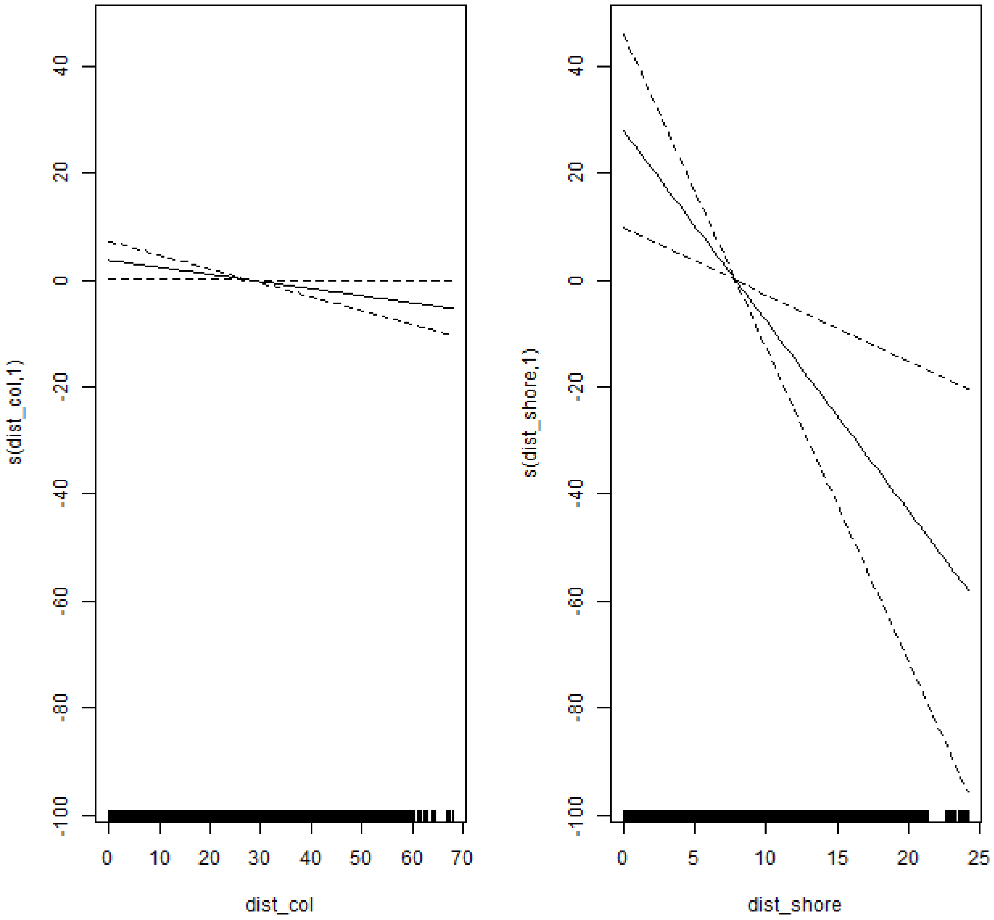


Figure A.4.4.4: GAM covariate plots for Sandwich Terns in Larne Lough

A.4.4.5 INLA Output

```
Running INLA.

Call:

c("inla(formula = formula.inla, family = \"binomial\", data =
complete.data.to.analyse, ",  "    weights = weights, verbose = TRUE)")

Fixed effects:

               mean     sd 0.025quant 0.5quant 0.975quant    kld

(Intercept)  1.8236 0.5229     0.8811   1.7913     2.9399 0.0583

dist_col    -0.1396 0.0630    -0.2808  -0.1325    -0.0350 0.0058

dist_shore  -3.6656 1.1502    -6.0777  -3.5990    -1.5923 0.0059


Model hyperparameters:

                           mean   sd     0.025quant 0.5quant 0.975quant

T.0 for mesh.points-basisT  8.157  3.191  2.050      8.091   14.568

Expected number of effective parameters(std dev): 2.998(0.00142)

Number of equivalent replicates : 9934.32

Marginal Likelihood:  -37.23
```