

Biomathematics and Statistics Scotland

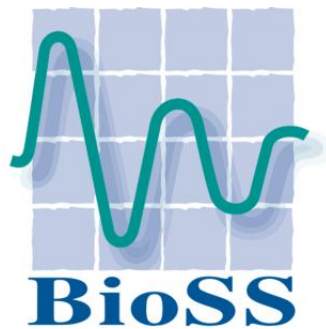
**Prediction of New Colonies – Seabird Tracking
Data (Under Agreement C10-0206-0387)**

CONTRACT No: C10-0206-0387

Report submitted to:

Joint Nature Conservation Committee

November 2012



Authors:

Mark J Brewer, Jackie M Potts,

Elizabeth I Duff, David A Elston

CONTENTS	PAGE
1. Non-Technical Summary	3
2. Introduction	4
3. Data	6
4. Methodology	14
5. Results	17
6. Prediction Maps of Usage	29
7. Discussion	42

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

(i) BioSS Terns Report II – Results Appendix;

(ii) BioSS Terns Report II– Software;

and also ancillary files:

(i) Spreadsheet files of grid predictions for each of the thirteen species/colony combinations for unsurveyed colonies;

(ii) R code files for: ordination; fitting models to a combination of sites; cross-validation; grid predictions

(iii) cleaned and standardised versions of the data files (for survey data and grids).

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 June 2012 BioSS was tasked with making predictions of usage and preference for Arctic, common and Sandwich terns for colonies lacking visual tracking data.

This work forms part of Phase II of a larger tern project JNCC are undertaking, and follows on from a previous project completed by BioSS earlier in 2012 as part of Phase I of the tern project. The Phase I work we undertook previously used visual tracking data to learn about important associations between terns' usage/preference and environmental covariates, and to map usage/preference for each tern species for those colonies with tracking data. The methodology developed from the Phase I project – essentially a flexible weighted logistic regression model – would be used in this new, Phase II project.

Predictive models were defined for all new colonies which combined data from all relevant colonies for each species separately. An evaluative procedure (employing assessment by a form of cross-validation) determined that in terms of the colonies with data, better predictions were obtained by combining data in this way rather than using ecological consideration or multivariate analysis of environmental data to suggest subsets of “similar” colonies for prediction. As part of the cross-validation analysis, we discovered that the most important predictors are distance to colony, distance to shore, bathymetry and chlorophyll concentration.

Based upon the above analysis, predictions and prediction maps were produced for all requested unsurveyed colonies.

2. Introduction

2.1 Background – Previous Phase

This project represents Phase II of analysis of data sets on four species of tern in several colonies in UK offshore waters. Phase I was concerned with developing models specifically designed for the type of tracking data available. The report for Phase I (Brewer et al., 2012) will be referred to throughout this document as “the Phase I report”.

The Phase I analysis determined that a weighted logistic regression was appropriate for analysing the data; the data itself was formed of individual tracks of known foraging instances (forming cases) with sets of randomly generated perturbations of those tracks (forming controls). The analysis thus took a case-control form. It was found that hierarchical (or random effect, mixed) models were not required, as had been used in previous tracking analysis work by Aarts et al. (2008) and Wakefield et al. (2010) – the difference being that for JNCC’s dataset, there were no known repeat observations per individual. In this framework, the cases represent “presence” and the controls represent “absences”. A number of explanatory variables were included in the regression, representing the environmental conditions at different locations, but also including the measures “distance to colony” and “distance to shore”.

Different forms of weighted logistic models were considered during Phase I analyses, using a range of facilities in R. Both GLMs and GAMs were considered with various model selection strategies where appropriate. Spatial autocorrelation of the response data was addressed, both by the weighting in the regression and via a spatial correlation network derived using the INLA (Integrated Nested Laplace Approximation) package in R (INLA, 2012). Different models were appropriate depending on the purpose of the modelling – for example, whether the aim was to identify significant relationships with the environmental covariates or to make predictions of usage and/or preference by each species in each location.

Further details can be found in the Phase I report itself.

2.2 Second Phase – Colonies Without Tracking Data

JNCC wish to provide predictions for a number of colonies which have no tracking data available (Phase II). The task at hand is to use data from surveyed colonies, using models such as those developed in Phase I, to make predictions for these new colonies.

The new colonies as specified in the invitation to tender and subsequently modified by JNCC (with agreement from BioSS) are the following (with colony names we shall use in the rest of this report in **bold**):

Table 1. Colonies without tracking data available, for which predictions are to be made.

Common tern	Dungeness
	Foulness (Greater Thames)
	Breydon Water (Norfolk)
	Liverpool Bay (The Dee estuary; Ribble & Alt estuaries)
	Strangford Lough (N Ireland)
	Carlingford Lough (N Ireland)
	Farne Islands (Northumberland)
Sandwich tern	Isle of May (Firth of Forth)
	Liverpool Bay (Duddon Estuary)
	Carlingford Lough (N Ireland)
Arctic tern	Strangford Lough (N Ireland)
	Isle of May (Firth of Forth)

These colonies supplement the list of colonies in the Phase I report; however, we provide a list here of colonies with data, as some new colonies (indicated with *) have been added for this Phase II work:

Table 2. Colonies with tracking data available.

Common tern	Coquet and Farne* Islands (Northumberland)
	Larne Lough (Northern Ireland)
	Glas Eileanan / South Shian (Mull area, west Scotland)
	Leith Docks (Firth of Forth)
	Cemlyn (Anglesey)
Sandwich tern	Coquet and Farne Islands (Northumberland)
	Larne Lough and Cockle* Island (Northern Ireland)
	Sands of Forvie (Aberdeenshire)
	Cemlyn (Anglesey)
Arctic tern	Coquet and Farne Islands (Northumberland)
	Copeland / Cockle Islands (Outer Ards , Northern Ireland)

The question of how to determine which of the colonies with survey data should be used to predict which of the colonies without is addressed in the methodology Section 4. This required us to determine suitable metrics for comparing models (within species) fitted using different subsets of colonies and different selected covariates.

3. Data

3.1 Data Summary

The environmental covariates for this phase are as for the first part: see Section 2 of the Phase I report for full details. As part of this second phase, we were required to conduct a deeper inspection of the data in order to justify the “extrapolation” required in producing predictions and maps for the new colonies. Boxplots were used to compare the ranges of the environmental covariates between colonies; this is discussed in Section 4.2. Such differences in ranges were not of concern in Phase I as each colony was analysed separately; only when multiple colonies are considered together is range mismatch a potential problem.

Boxplots were also used to identify outliers and variables which have a skewed distribution. Section 3.2 which follows contains a discussion of outliers in some of the environmental variables; this follows up a recommendation made by us in the discussion (Section 6) of the Phase I report. We also considered whether we could use logged versions of chlorophyll concentrations and wave and current shear stresses; on the basis of our new findings, we would recommend that this transformation could have been applied during the Phase I analysis.

Some of the covariates considered in Phase I of the project were not considered further in Phase II. Eastness, northness, slope and sand were not considered because they were not selected in any of the Phase 1 models. (There was one exception where slope was selected by the AIC criterion, but was not significant). The interannual standard deviation of probability of a frequent thermal front in spring and summer were also excluded from the model selection process for Phase II. This was because even though they were selected in some Phase I models, it did not seem biologically realistic to suppose that the birds would respond to these variables while not responding to the probability of a frequent thermal front itself. We would recommend excluding these from the Phase I models also.

3.2 Variable Inspection – Outliers

The boxplots in Figure 1 illustrate the range of values for each environmental covariate; as the predictive grids for Sandwich (out to 55km from the colony) are different from those for the other three species (out to 31km from the colony), there is a separate plot having only the colonies relevant to Sandwich terns. The variable name is indicated by the y-axis, and the key for the colony codes on the x-axis is as follows:

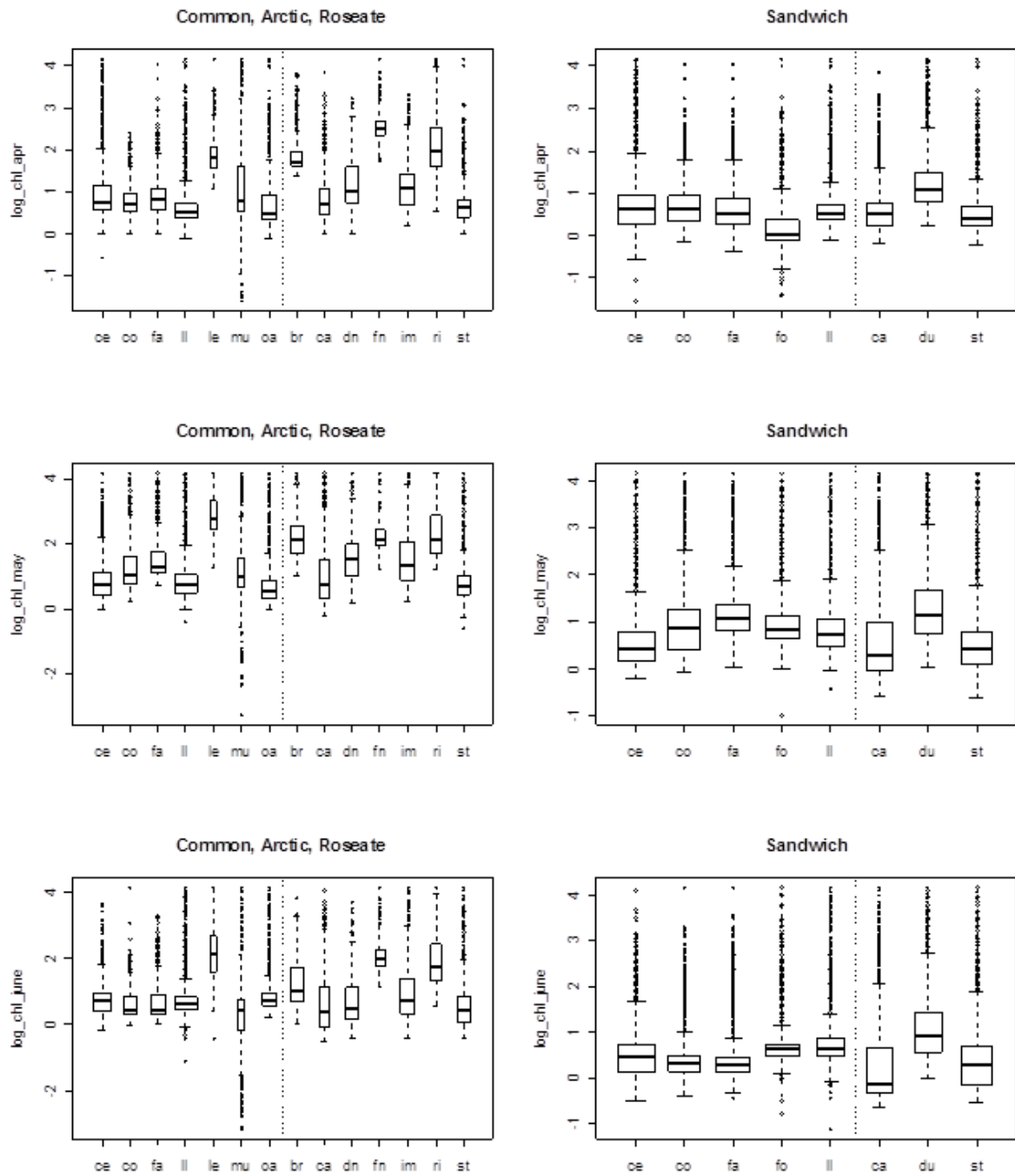
ce	Cemlyn
co	Coquet
fa	Farnes
ll	Larne Lough
le	Leith
mu	Mull
oa	Outer Ards
br	Breydon
ca	Carlingford Lough
dn	Dungeness
fn	Foulness
im	Isle of May
ri	Ribble

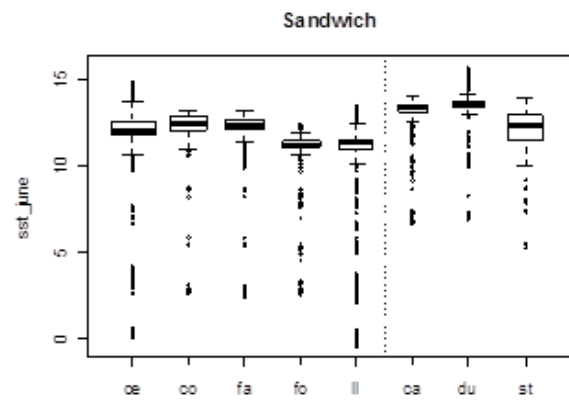
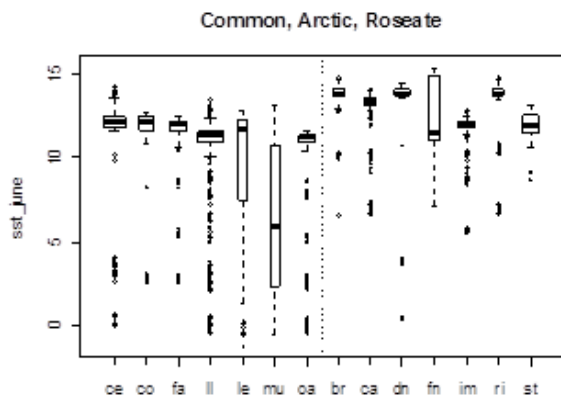
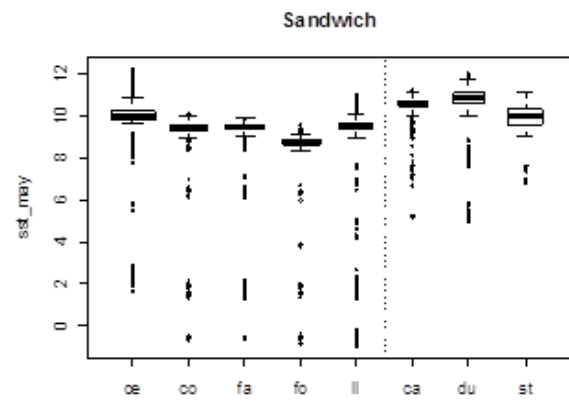
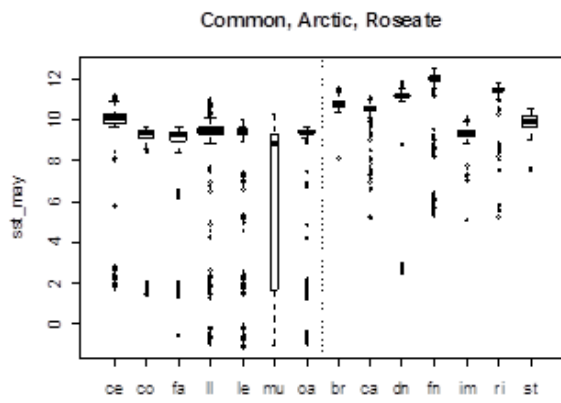
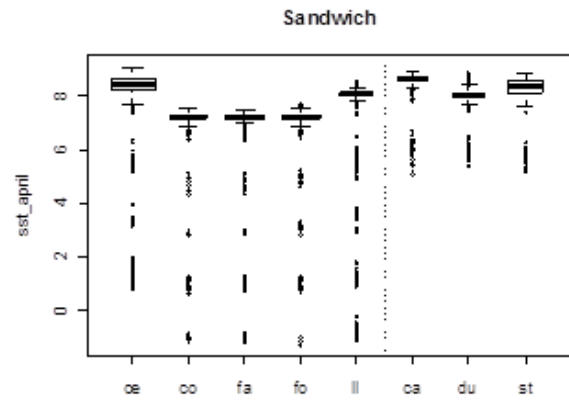
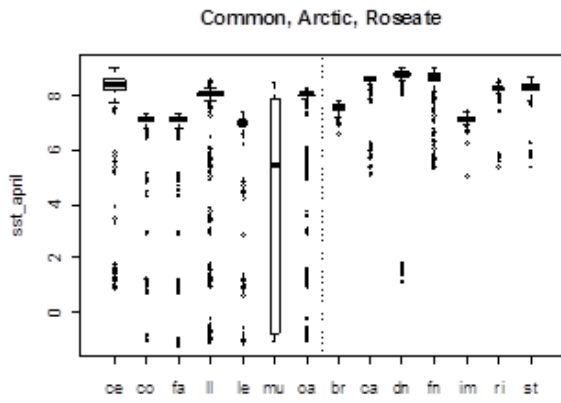
st	Strangford Lough
fo	Forvie
du	Duddon

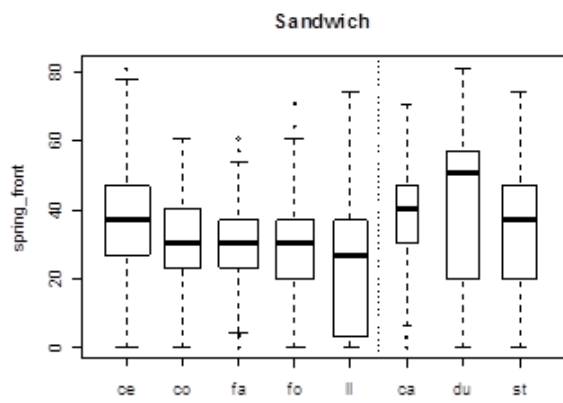
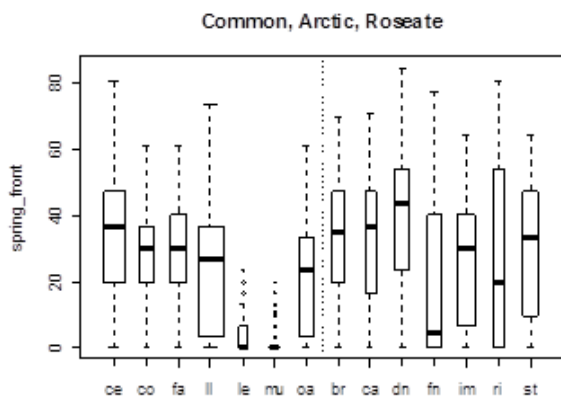
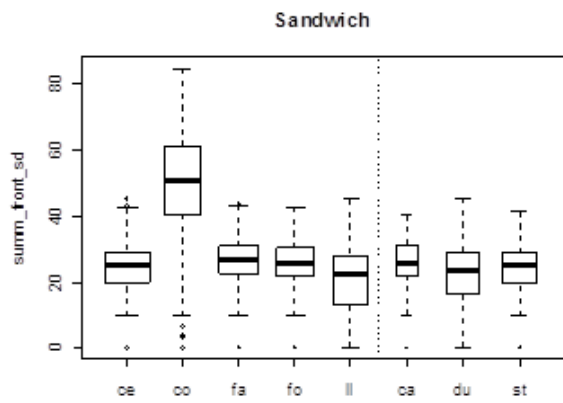
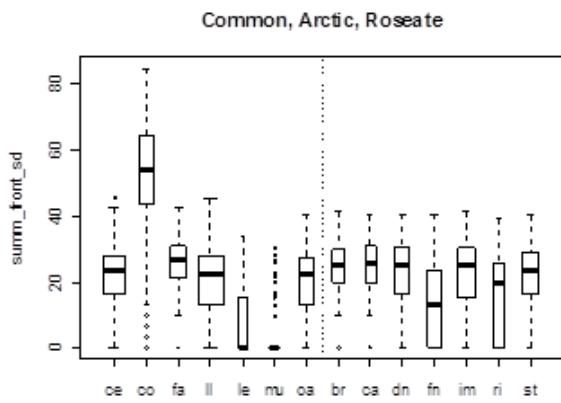
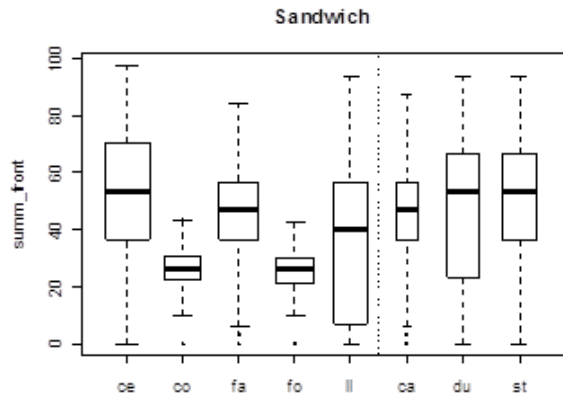
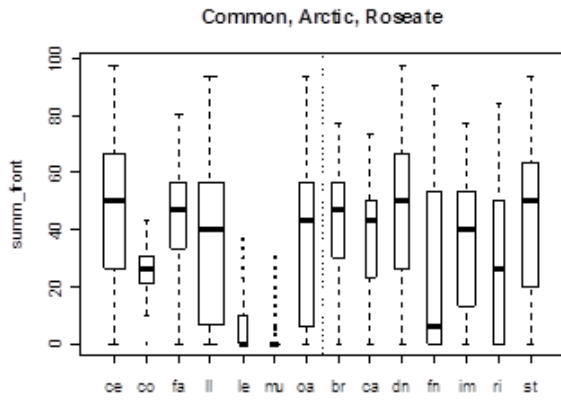
The boxplots show a negatively skewed distribution for sea surface temperature, with low values occurring near the shore. The extent to which these data are reliable is uncertain. Removal of the values that were considered unreliable would have resulted in considerable loss of data, particularly around the shore, so sea surface temperature was excluded from the analysis instead.

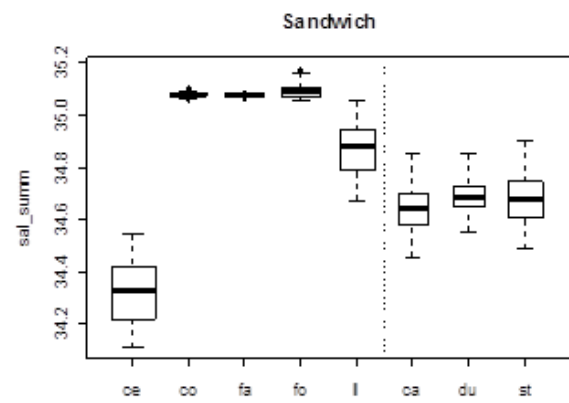
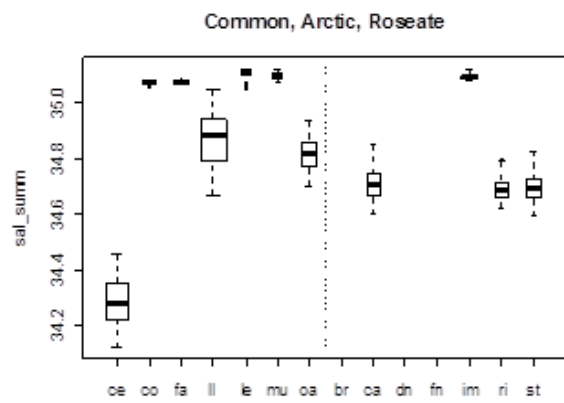
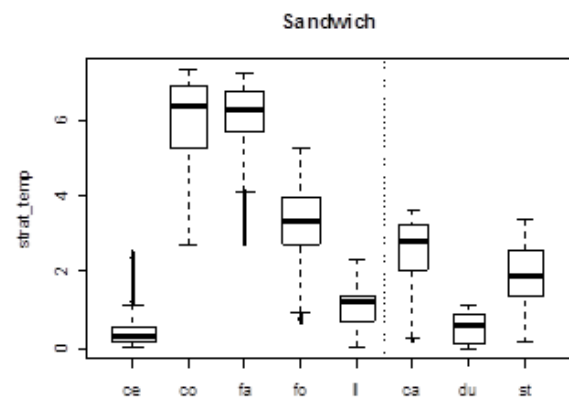
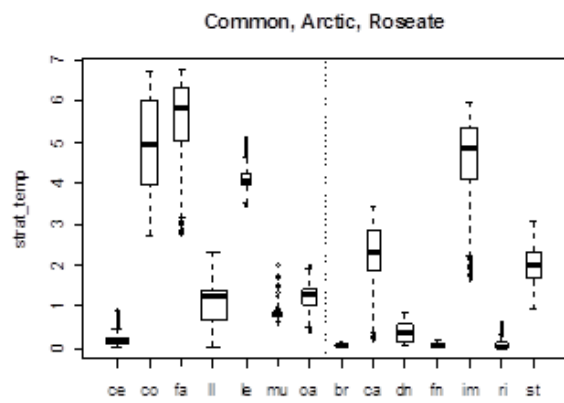
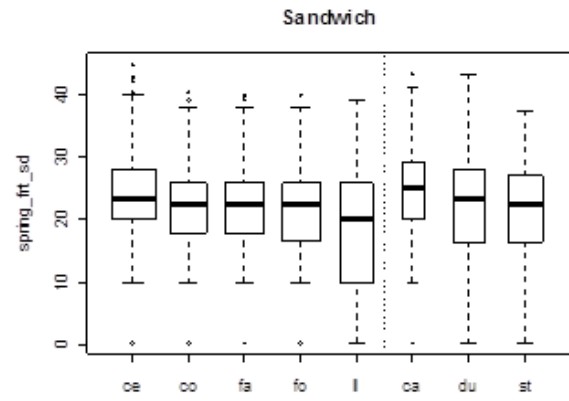
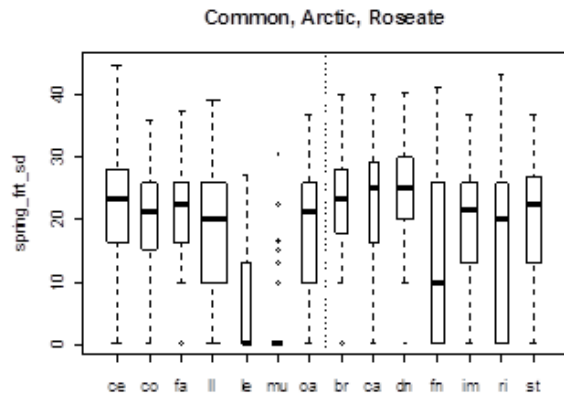
Chlorophyll concentrations and wave and current shear stresses had highly positively skewed distributions and were therefore log-transformed prior to further analysis. (The log-transformed versions are shown in the boxplots below). There was no reason to question the reliability of these, as lognormal distributions frequently arise for variables such as chemical concentration which have low mean values, high variances, and cannot be negative. On the other hand, some of the sea surface temperature values, particularly those below 0°C, seemed unrealistic.

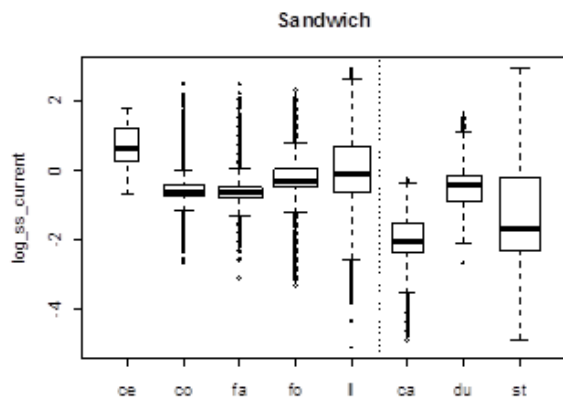
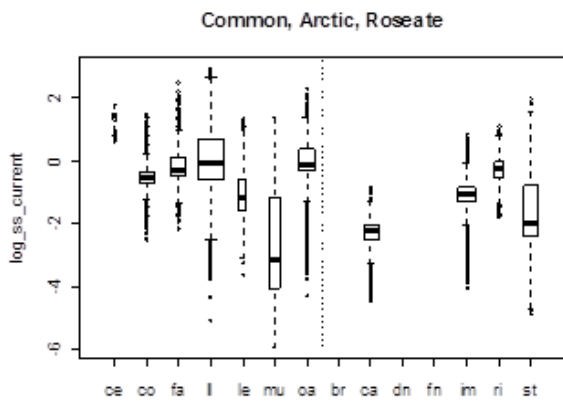
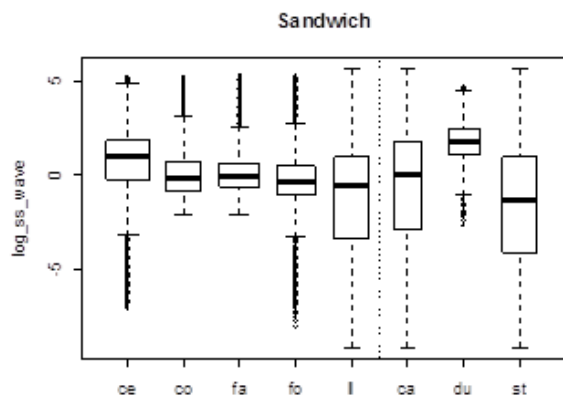
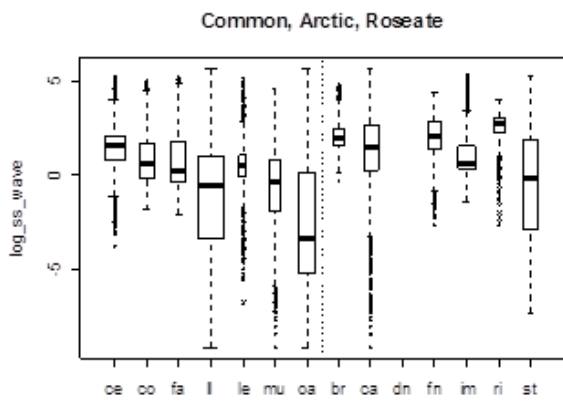
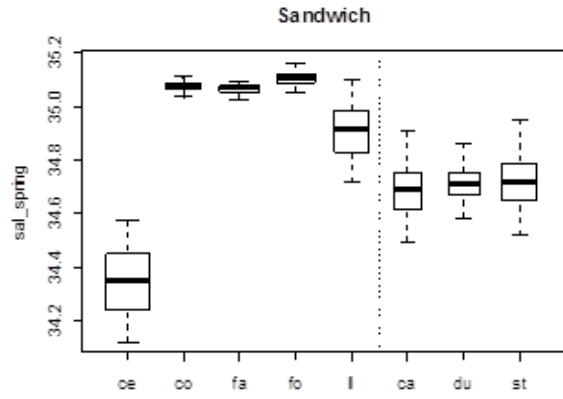
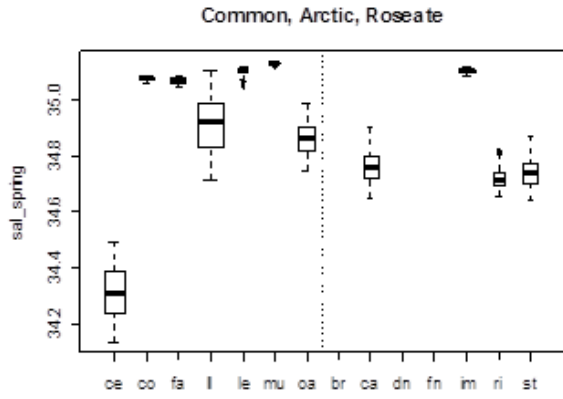
Figure 1. Boxplots of environmental covariates. Colonies to the left of the vertical line are those with tracking data and those to the right are those without.

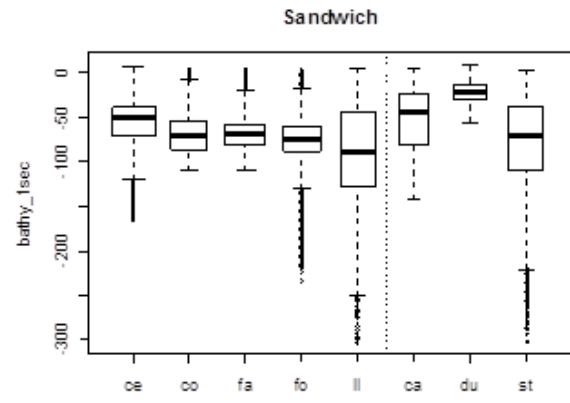
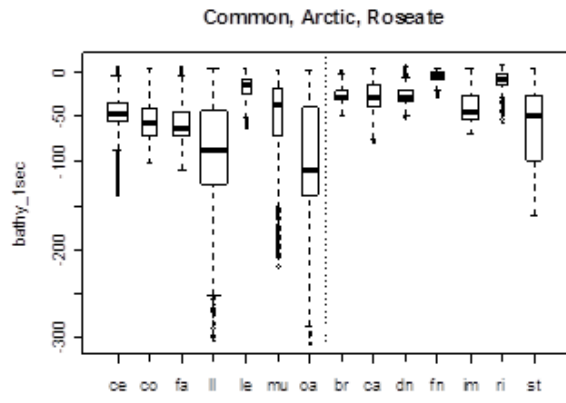












4. Methodology

4.1 Weighted Logistic Regression via a Case-Control Design

As noted earlier, the form of statistical model used for analysing the tern tracking data was a weighted logistic regression based on a case-control design. Full details of the modelling procedure and the generation of the control data can be found in the Phase I report. We did not include INLA in Phase II, as it was only used for model checking in Phase I and not for making predictions.

4.2 Comparisons of Environmental Data Between Colonies

One extremely important aspect of this project is to determine which colony or colonies can be used to build models to make predictions for new colonies lacking tracking data. For each species, we decided to compare the similarity or otherwise across colonies (or, more specifically, the foraging ranges of colonies) of the environmental covariates used in the modelling. The reasoning for this is that if a set of colonies appears to contain approximately the same environments, this might be justification for using a model from one or more colonies within the set to obtain predictions for another. On the other hand, colonies which are well-separated in multivariate environment space may present radically different environments to terns, and therefore a model from one such colony may not be suitable to predicting for another.

We compare the environmental data between colonies in two ways: firstly, we look at simple boxplot summaries (presented in Section 3.2) for each environmental covariate in turn; secondly, we use a principal component analysis (PCA) to study the combination of information from all covariates simultaneously. Principal component analysis takes a set of variables and replaces them with a smaller number of new variables (the principal components) in such a way that as much as possible of the information in the original variables is retained in the new ones. This allows us to plot the data in a concise way, for example by plotting the second principal component (PC2) against the first principal component (PC1). Colonies which are close together in this plot will be similar in terms of the original set of environmental covariates. This exploratory analysis will then help us in selecting suitable subsets of colonies with which to build models for making usage and preference predictions for the new colonies.

Visual inspection of the boxplots in Section 3.2 can indicate which variables may be unsuitable for extrapolating from one colony to another. We found two such variables: (i) Salinity is a significant covariate at Cemlyn, but the boxplots show that the distribution of salinity at Cemlyn is very different from that at other colonies; (ii) wave and current shear stress are significant covariates at Outer Ards, but the distribution of wave shear stress was different from that at many of the other colonies.

The set of variables to be considered in the PCA was:

bathy_1sec , strat_temp , summ_front , spring_front , log_chl_apr , log_chl_may ,
log_chl_june , log_ss_wave , log_ss_current , sal_spring , sal_summ.

However, some of the environmental variables are not available for some of the new colonies. For example, at Dungeness `ss_wave`, `ss_current`, `sal_spring` and `sal_summ` are entirely missing. The PCA function in R will remove entirely any row that contains a missing value; as such, trying to use all the

above variables can result in *all* data for one or more colonies being removed. The offending variables are the final four in the set above; hence, for each species, we conduct a PCA both on the above set of covariates (All Variables) and the following smaller set (Reduced Set of Variables):

bathy_1sec , strat_temp , summ_front , spring_front , log_chl_apr , log_chl_may ,
log_chl_june .

4.3 Cross Validation for Selecting Predictive Models for Colonies/Species

JNCC supplied us with suggested groupings for prediction purposes and were taken into consideration in the cross-validation exercise. These are summarised briefly as follows and were based loosely on geographical similarities. Some of these such as the close similarity between Coquet and Farne Islands were confirmed by the PCA.

Table 3. Suggested colony groupings

Common tern

Group	Model	Prediction
1	Coquet Island, Farne Islands (very little data)	Farne Islands, Isle of May
2	Coquet Island, Farne Islands (very little data), Cemlyn	Dungeness
3	Larne Lough	Strangford Lough, Carlingford Lough
4	Larne Lough Cemlyn	Strangford Lough, Carlingford Lough
5	Coquet Island, Farne Islands, Leith Docks	Foulness , Breydon Water
6	Larne Lough , Glas Eileanan / South Shian (Mull), Cemlyn	Liverpool Bay (Ribble)

Arctic tern

Group	Model	Prediction
1	Coquet Island, Farne Islands	Isle of May
2	Outer Ards	Strangford Lough

Sandwich tern

Group	Model	Prediction
1	Larne Lough, Cockle Island	Carlingford Lough, Strangford Lough
2	Larne Lough, Cockle Island, Cemlyn	Duddon Estuary, Carlingford Lough, Strangford Lough

The suggested ecological groupings and the PCA exercise indicated which colonies might be similar in terms of environment and resulted in a series of colony groupings. Data from each colony within each resulting grouping were combined to produce models that could be used to make predictions for new colonies.

Cross-validation was used to select which colonies to use for prediction. This was done by assessing the fit of predictions to the tracking data from a particular colony from (i) a model developed using the remaining colonies in a proposed grouping and comparing this with (ii) a model developed using data from all the remaining colonies. For example, it was suggested that data from Coquet and Farnes might be used to predict Arctic terns at the Isle of May. We therefore tested whether Farnes was better predicted using a model developed for Coquet alone, or a model using all available Arctic tern data (Coquet and Outer Ards together). The assessment was carried out on the tracking data (observations and controls) rather than on the grid data because we did not have presence-absence data in the form of a grid.

Two scores were used to assess quality of predictions (fitted to the tracking data):

(1) The sum of squared errors $\sum (y_i - p_i)^2$

If this quantity is divided by the number of observations, it gives the mean squared error, also known as the Brier score when applied to probabilistic predictions (Brier, 1950);

(2) A score related to the log-likelihood $\sum (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$

where y is the binary variable indicating foraging behaviour and p is the predicted probability.

The intercept is arbitrary for case-control data as it depends on the ratio of controls to cases, which we have chosen, and which has no biological meaning. An adjustment was therefore made to the intercept for each model before calculating the two scores. A constant was added to the intercept to ensure that the sum of the predicted probabilities was equal to the sum of the values of the binary variable.

There are many other measures that could have been used; see Liu *et al.* 2011 for a review. For example, the area under the receiver operating characteristic (ROC) curve, known as the AUC, is widely used, although it has received some criticism (Lobo *et al.*, 2008). It is unlikely that the overall conclusions would have changed had we used a different metric – the results in the end were clear and consistent in terms of prediction assessment, and predicted maps tended to vary only slightly for the better-fitting models in any case.

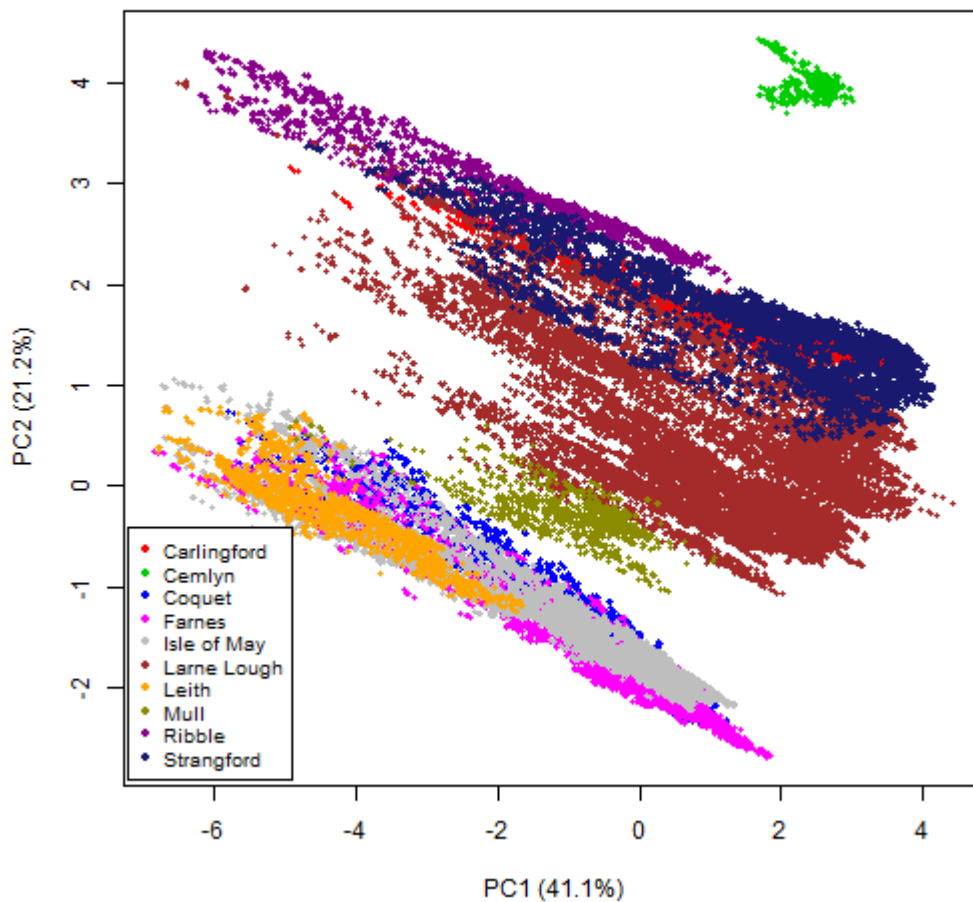
Results and interpretation from this analysis are found in Section 5. The predictions themselves can be found in supplemental spreadsheets while maps of predictions are presented in Section 6.

5. Results

5.1 Principal Component Analysis - Comparisons of Environmental Data

5.1.1 Common Tern – All Variables

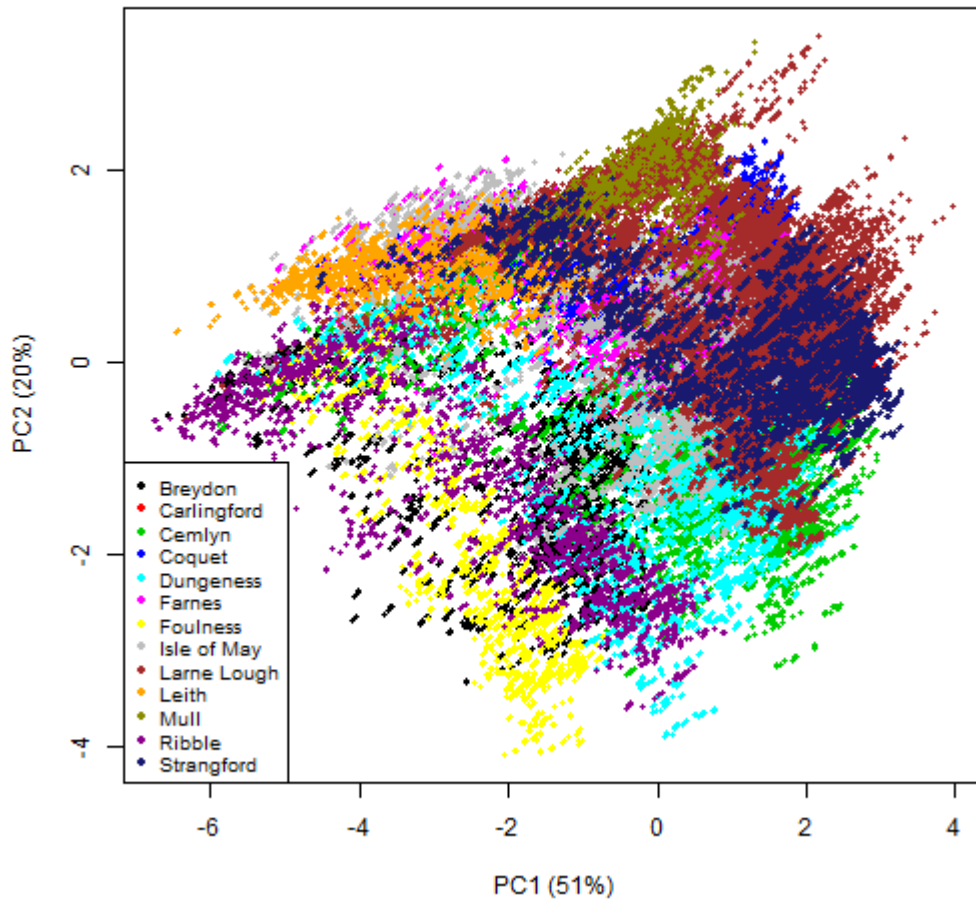
With the full set of PCA variables, it can be seen that Farnes, Isle of May, Coquet and Leith all occupy the same space in PC1 and PC2, suggesting that these colonies are similar in terms of the major sources of variation in environmental conditions. Strangford and Carlingford Loughs appear to lie between (but overlapping) Ribble and Larne Lough. Cemlyn seems to be something of an outlier here, but note that there are a large number of missing values for the variable `ss_current`, which removes most of the data points for that colony.

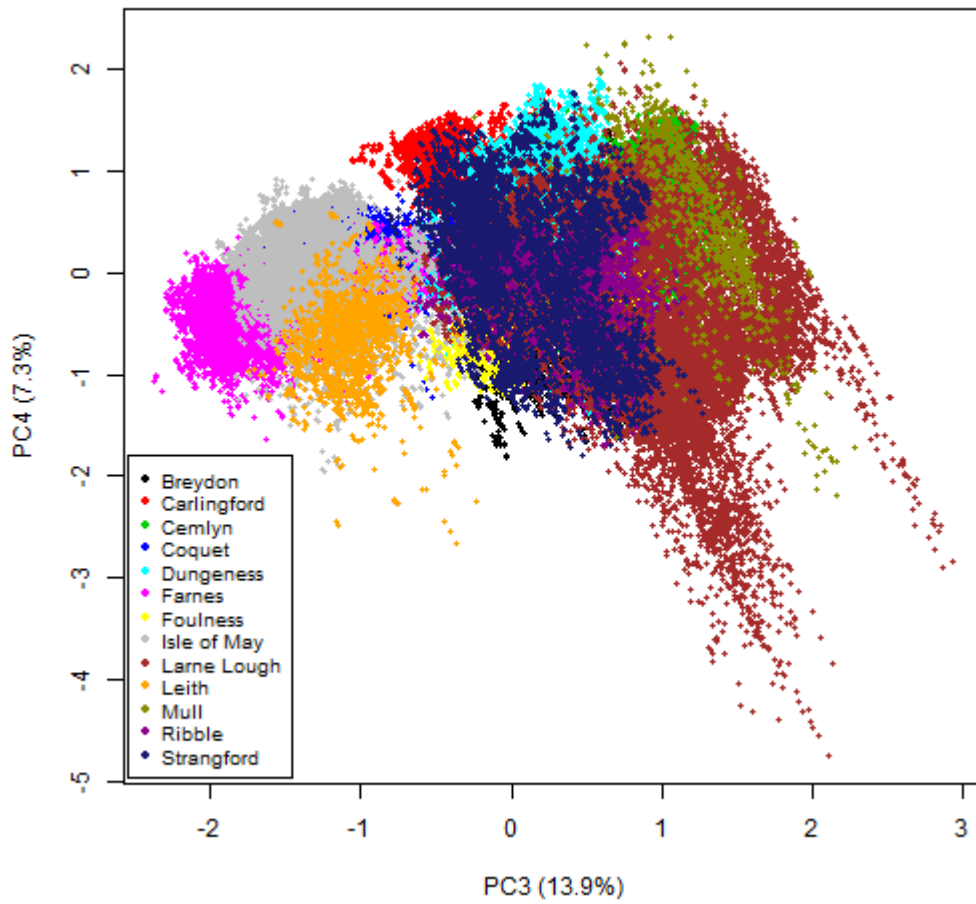


5.1.2 Common Tern – Reduced Set of Variables

With the reduced set of PCA variables, the picture changes dramatically. From the plot of PC2 vs PC1, the colonies do not seem to separate out at all well. The next plot – showing PC4 against PC3 – shows that we need to go to the third principal component before we start getting clear colony distinction. This in itself suggests that differences between colonies are not the major source of

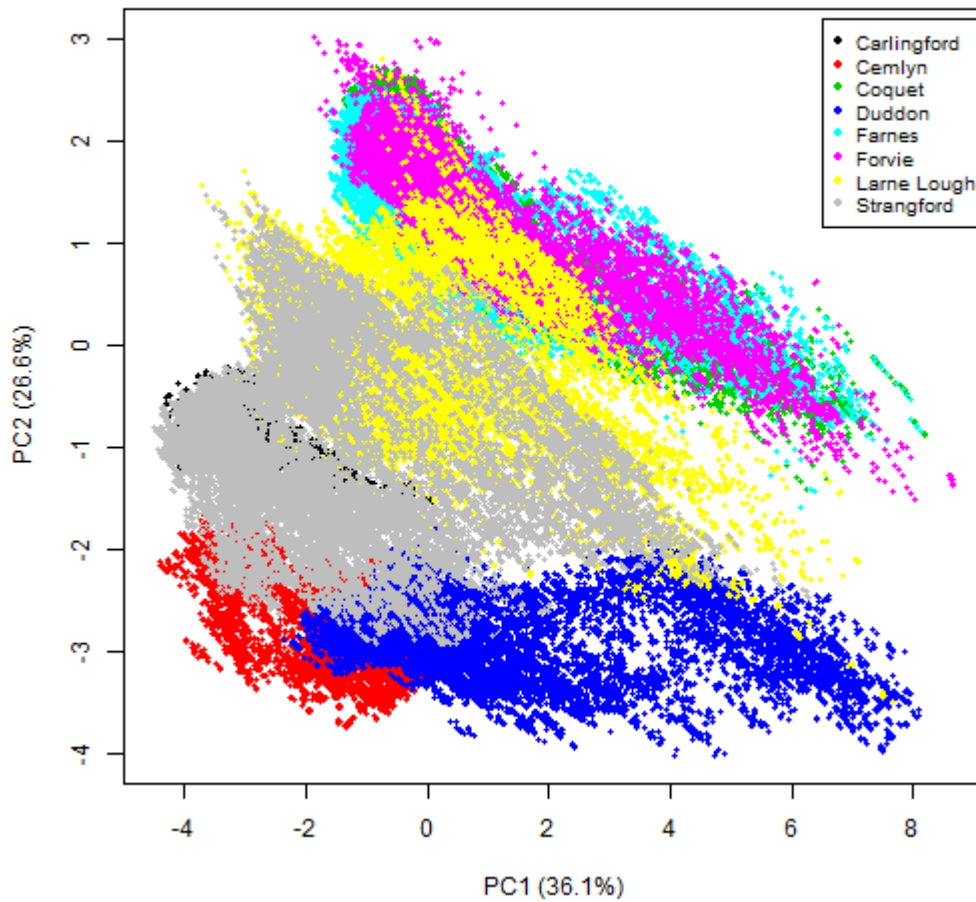
variability in environmental conditions. In the PC4 vs PC3, the patterns are similar compared with the plot in Section 5.1.1, but note that there are now more colonies included – those with all missing values in the excluded variables. Interestingly, Dungeness seems to sit well with the Irish Sea colonies, although we should stress again that from the PC2 vs PC1 plot, Dungeness does not appear noticeably different from other colonies. In either plot, Foulness and Breydon seem similar to each other; they resemble Ribble and Dungeness most closely in PC2 vs PC1, but are linked with Coquet in PC4 vs PC3.





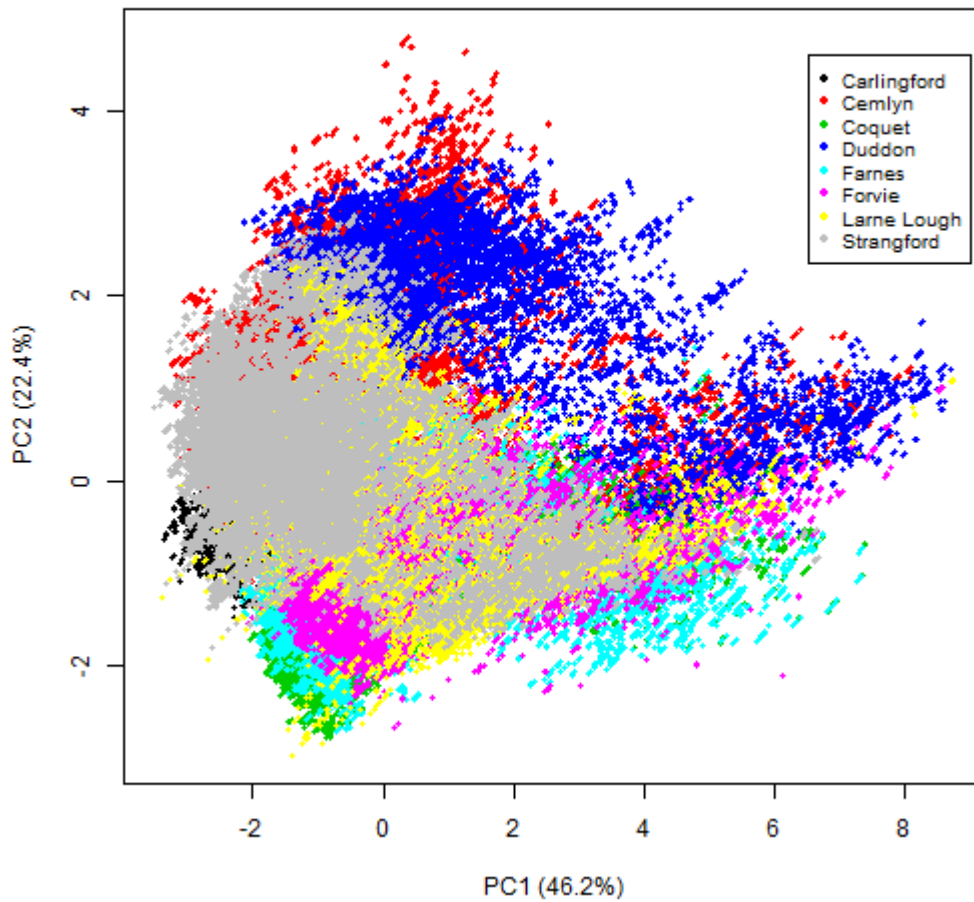
5.1.3 Sandwich Tern – All Variables

With the full set of PCA variables, we see that Coquet, Forvie and Farnes are similar, and that Duddon overlaps Cemlyn. Larne Lough, Carlingford and Strangford lie in between these two groups.



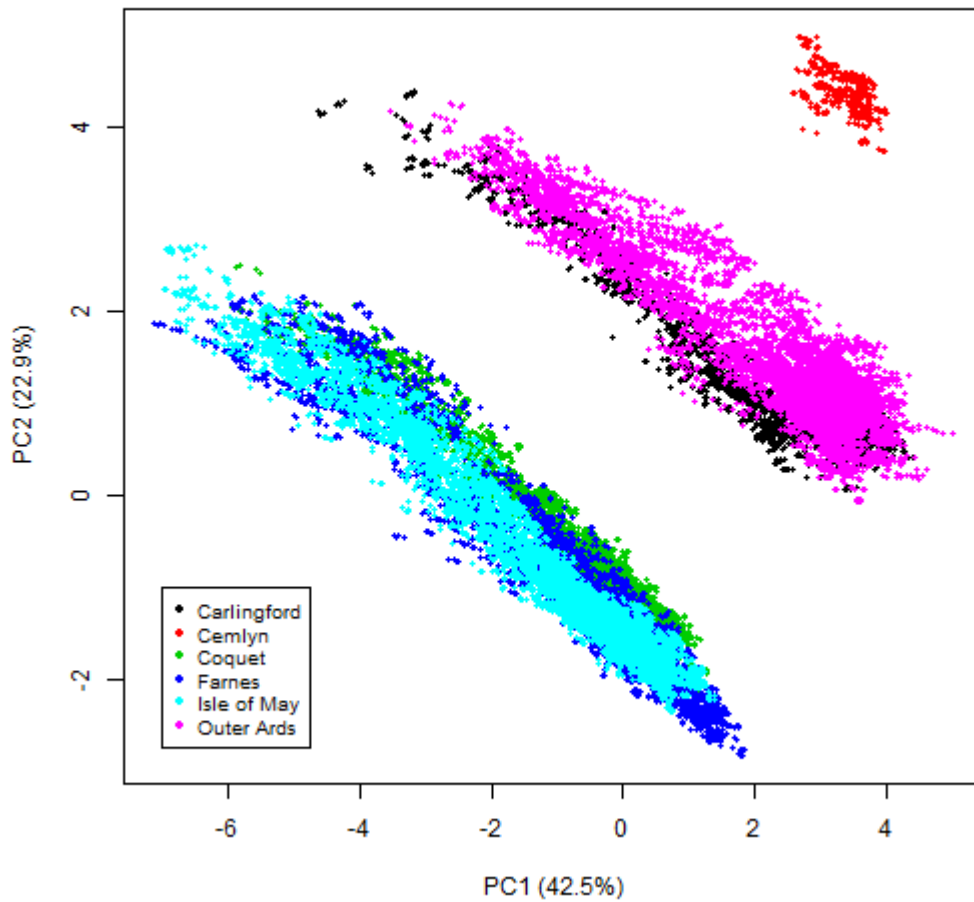
5.1.4 Sandwich Tern – Reduced Set of Variables

With the reduced set of PCA variables, as with Common Terns we see a much less clear separation of colonies. What we do see is that Duddon now overlaps Cemlyn very well, and that the Loughs Larne, Strangford and Carlingford lie between Duddon/Cemlyn and Coquet/Farnes/Forvie.



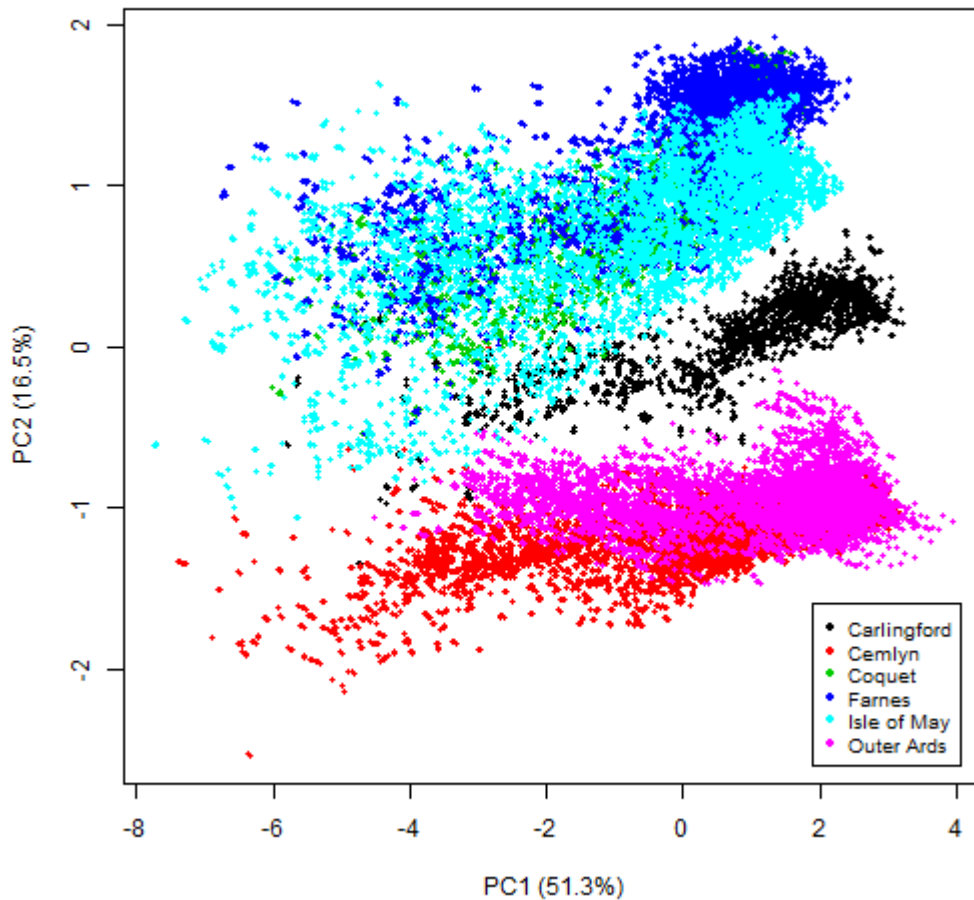
5.1.5 Arctic Tern – All Variables

With the full set of PCA variables, there is very clear separation into groups. Coquet, Farnes and Isle of May form one group, while Carlingford and Outer Ards form another. Cemlyn is something of an outlier, but as noted for Common Terns, very many missing values for one variable means most observations are deleted.



5.1.6 Arctic Tern – Reduced Set of Variables

With the reduced set of PCA variables, the group separation is less clear than with the full set, but still apparent. Coquet, Farnes and Isle of May still overlap strongly, whereas Cemlyn, now less of an outlier, overlaps Outer Ards. Carlingford Lough lies between these two overlapping groups.



5.2 Cross-Validation for Selecting Predictive Models for Colonies/Species

The aim was to find a set of variables that were consistent predictors across the different colonies for which we have data, as it is more likely that these will be successful at making predictions for new colonies. We took this approach, rather than considering all variables when selecting a model for a combination of different colonies, because the latter approach would have tended to select variables that explain a difference in intercept between colonies (which is of no interest), as well as those which explain the pattern of foraging within a colony. In theory, as we have used a ratio of 12 controls to each data point we would expect the intercept to be the same for each colony. However, in practice, it differs because points have been excluded where control tracks fell on land and where there are missing covariate values.

The variables that are consistently selected are `dist_col`, `dist_shore`, `bathy_1sec` and `chl_june`. When considering models for combinations of sites in the cross-validation exercise the candidate variables were reduced to this set. The variable that most commonly appeared to have a nonlinear effect in the Phase I models was `dist_col`. We therefore considered GAM models with a nonlinear term in `dist_col` as possible candidate models, but constrained the other terms to be linear.

Removal of some variables and log-transformation of others, as discussed in Section 3.2, led to some changes in the models for single colonies developed in Phase I of the project. New models selected using either AIC (Akaike's Information Criterion) or likelihood ratio tests (LRT) are shown below. This is to demonstrate that `dist_col`, `dist_shore`, `bathy_1sec` and `chl_june` were being selected consistently; where AIC selects additional variables that are not on this list we present only the results for LRT.

Arctic terns

Coquet: `dist_col`, `chl_june`, `bathy_1sec` (AIC and LRT)

Farnes: `dist_col`, `dist_shore`, `sal_spring` (AIC; LRT omits `dist_shore`)

Outer Ards: `dist_col`, `chl_june`, `ss_wave`, `ss_current` (AIC and LRT)

Common terns

Cemlyn: `dist_col`, `bathy_1sec` (AIC and LRT; salinity excluded)

Leith: `dist_col`, `dist_shore`, `chl_may`, `chl_june`, `sal_summ` (LRT)

Coquet: `dist_col`, `bathy_1sec`, `chl_june` (LRT)

Larne Lough: `dist_col`, `dist_shore`, `chl_june`, `bathy_1sec` (LRT)

Sandwich terns

Cemlyn: `dist_col`, `dist_shore`, `chl_apr`, `chl_june` (AIC; LRT omits `dist_shore` and `chl_june`)

Coquet: `dist_col`, `dist_shore` (LRT)

Farnes: `dist_shore`, `sum_front`, `spring_front`, `bathy_1sec`, `sal_summ` (AIC)

Forvie: `dist_shore`, `strat_temp` (AIC and LRT)

Larne Lough: `dist_col`, `dist_shore` (LRT – after excluding covariates with large numbers of missing values)

Cockle Island: `dist_col`, `chl_june`, `ss_current` (AIC and LRT)

Cross-validation results are shown in Table 4 below. Note that due to the large number of missing chlorophyll values for Larne Lough, chlorophyll was excluded when making predictions for Larne Lough, and from any models in which Larne Lough is the sole colony used to make the predictions.

In general, predictions are better when data from all available colonies for that species are combined. There are some cases where predictions based on a single colony are slightly better than those based on all colonies combined, but they can be considerably worse. In the final models we have therefore chosen to use data from all available colonies for each species, to provide a consistent approach. The use of GAM models with a non-linear term for distance to colony sometimes makes predictions worse when the model is applied to another colony. Chakraborty et al. (2011) note in general that GAMs can be poor for out-of-sample prediction. Linear terms only were therefore used in the final predictive models.

The following covariates were used for each species in the final models:

Arctic terns: distance to colony and bathymetry

Common terns: distance to colony, distance to shore and bathymetry

Sandwich terns: distance to colony, distance to shore, bathymetry and June chlorophyll concentration. (June chlorophyll concentration was omitted for Strangford and Carlingford Loughs due to the large number of missing values).

Full details of the models are presented in the Results Appendix.

Table 4. Results of cross-validation. Models with lower values of the sum of squared errors and higher values (i.e. lower absolute values) of the LL score are better; the best model in each case is shown in bold type. The notation s(dist_col) indicates a GAM with a nonlinear term in distance to colony.

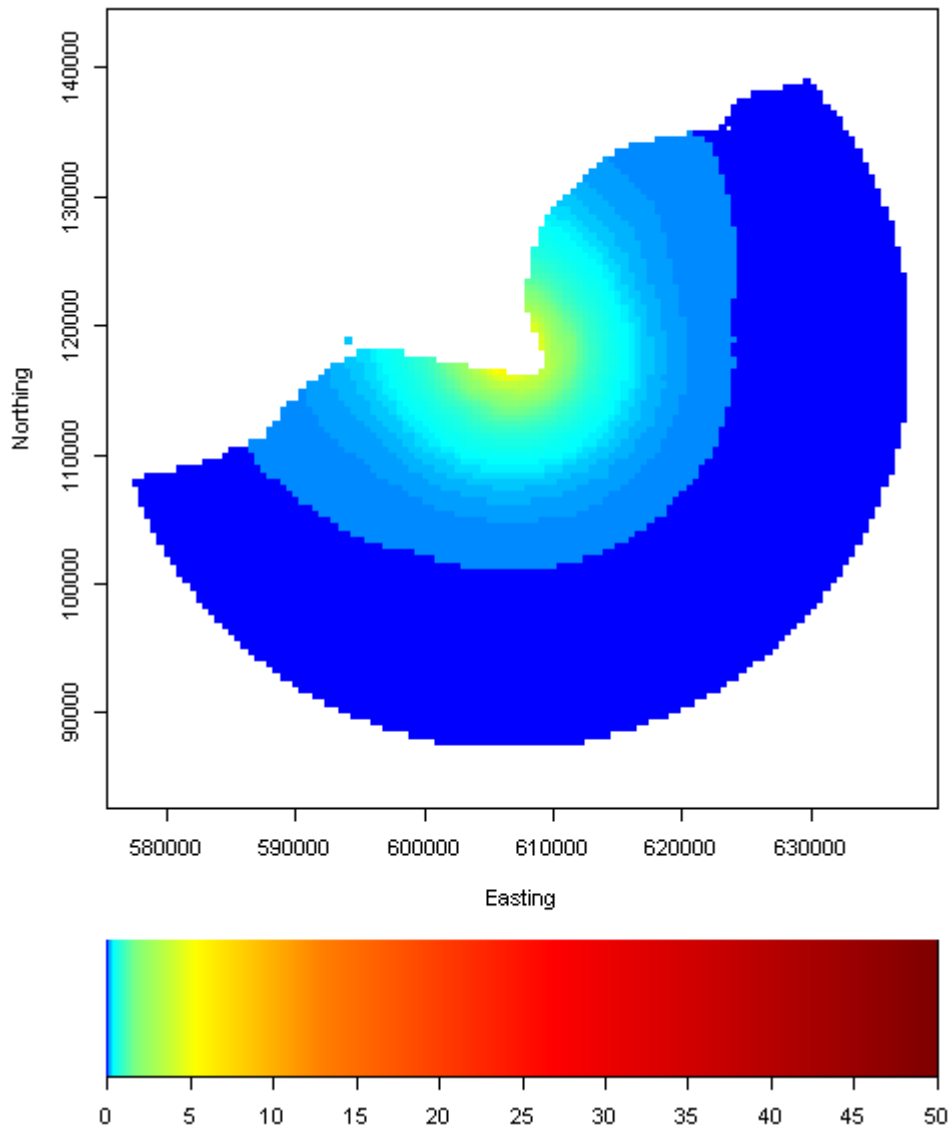
Species	Colony to predict	Model developed for	Covariates	LL Score	Sum of Squared Errors
Arctic	Coquet	Farnes	dist_col	-9612	2582
		Farnes, Outer Ards	dist_col, bathy_1sec	-9560	2558
Arctic	Farnes	Coquet	dist_col, chl_june, bathy_1sec	-4641	1306
		Coquet	s(dist_col), bathy_1sec, chl_june	-4770	1332
		Coquet, Outer Ards	dist_col, chl_june, bathy_1sec	-4297	1217
		Coquet, Outer Ards	dist_col, bathy_1sec	-4132	1180
Common	Coquet	Cemlyn	dist_col, bathy_1sec	-7657	1970
		Leith	dist_col, dist_shore, chl_may, chl_june, sal_summ	-5979	1749
		Leith	dist_col, chl_june	-5868	1744
		Cemlyn, Leith, Larne Lough, Mull, Farnes	dist_col, bathy_1sec	-6165	1769
		Cemlyn, Leith, Larne Lough, Mull, Farnes	dist_col, bathy_1sec, dist_shore	-6108	1761
		Cemlyn, Leith, Larne Lough, Mull, Farnes	s(dist_col), bathy_1sec	-6123	1772
	Cemlyn	Coquet	dist_col, bathy_1sec, chl_june	-3026	916
		Coquet	s(dist_col), bathy_1sec, chl_june	-3021	929
		Larne Lough	s(dist_col), bathy_1sec, dist_shore	-4497	1326
		Coquet, Leith, Farnes	dist_col, bathy_1sec, chl_june, dist_shore	-3213	971
		Coquet, Leith, Farnes	s(dist_col), bathy_1sec, dist_shore	-3325	1049
		Coquet, Leith, Larne Lough, Mull, Farnes	dist_col, bathy_1sec, dist_shore	-3331	1007
		Coquet, Leith, Larne Lough, Mull, Farnes	dist_col, bathy_1sec, chl_june, dist_shore	-3324	1004
	Leith	Coquet	dist_col, bathy_1sec, chl_june	-16574	4514
		Coquet	s(dist_col), bathy_1sec, chl_june	-16091	4422
		Coquet, Cemlyn, Farnes	dist_col, bathy_1sec, chl_june	-17024	4627
		Cemlyn, Coquet, Larne Lough, Mull, Farnes	dist_col, dist_shore, bathy_1sec	-14806	4139
		Cemlyn, Coquet, Larne Lough, Mull, Farnes	s(dist_col), bathy_1sec	-14761	4134

	Larne Lough	Cemlyn Coquet, Farnes, Leith, Mull Cemlyn, Coquet, Leith, Mull, Farnes Cemlyn, Coquet, Leith, Mull, Farnes	dist_col, bathy_1sec dist_col, bathy_1sec,dist_shore dist_col, bathy_1sec,dist_shore s(dist_col), bathy_1sec, dist_shore	-9214 -5034 -5046 -5185	1477 1373 1375 1415
Sandwich	Larne Lough	Cockle Island Cockle Island, Cemlyn Cockle Island, Cemlyn Cockle Island, Cemlyn, Coquet, Farnes, Forvie Cockle Island, Cemlyn, Coquet, Farnes, Forvie	dist_col, dist_shore dist_col, dist_shore, bathy_1sec s(dist_col), dist_shore dist_col, dist_shore, bathy_1sec s(dist_col), dist_shore, bathy_1sec	-2692 -2523 -2942 -2063 -2250	794 736 879 595 697
	Cemlyn	Larne Lough, Cockle Island Larne Lough, Cockle Island, Coquet, Farnes, Forvie Larne Lough, Cockle Island, Coquet, Farnes, Forvie	dist_col, bathy_1sec dist_col,bathy_1sec,dist_shore s(dist_col), dist_shore, bathy_1sec	-6742 -6729 -6606	2016 1834 1834
	Cockle Island	Cemlyn, Larne Lough Cemlyn, Larne Lough, Coquet, Farnes, Forvie	dist_col,bathy_1sec,dist_shore,chl_june dist_col,bathy_1sec,dist_shore,chl_june	-5990 -4760	1387 1282

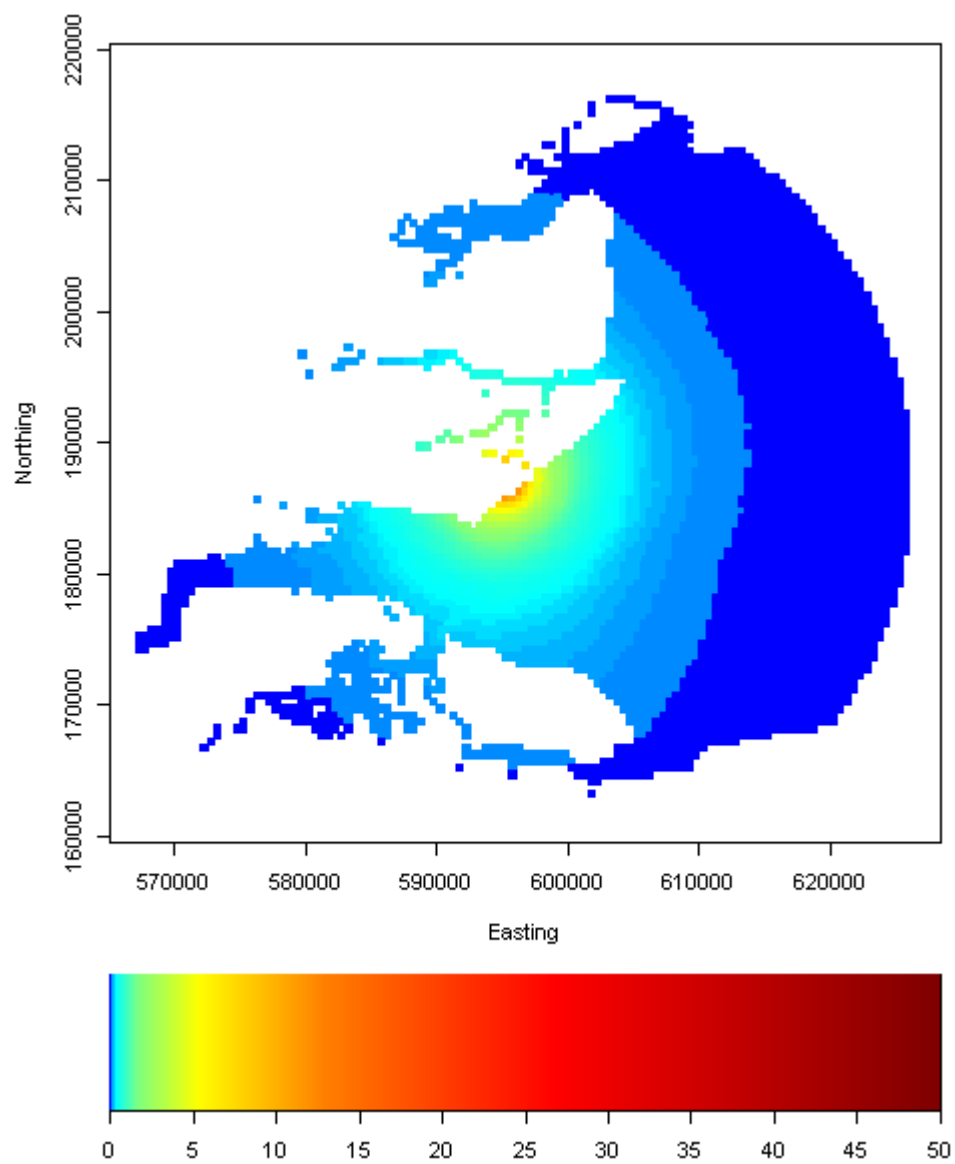
6. Prediction Maps of Usage

To calculate usage, preference is divided by distance to colony and multiplied by a scale factor which ensures that the probabilities sum to one. For mapping purposes, the probabilities have been multiplied by 1000. A very small number of points closest to the colony were removed if this gave a value greater than 50.

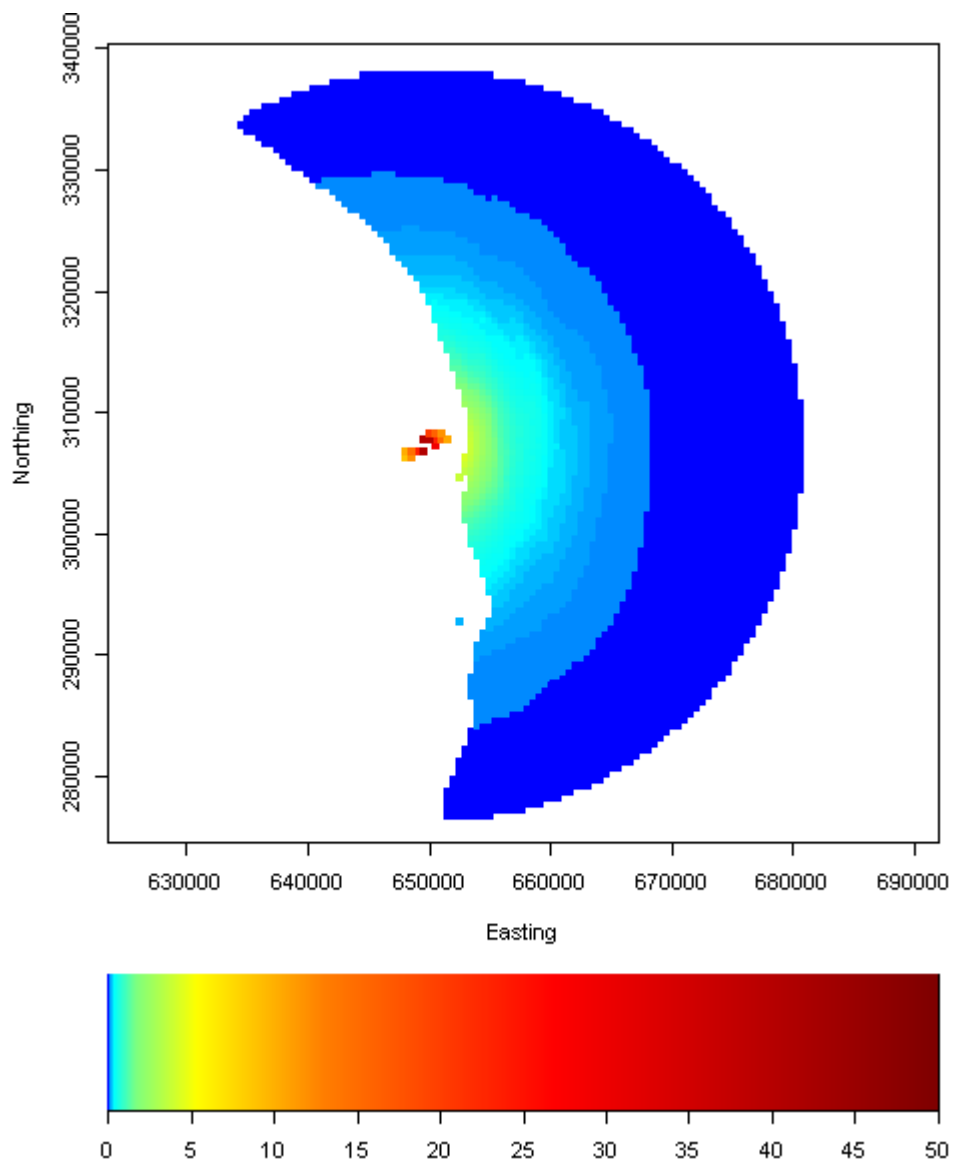
Common Tern, Dungeness



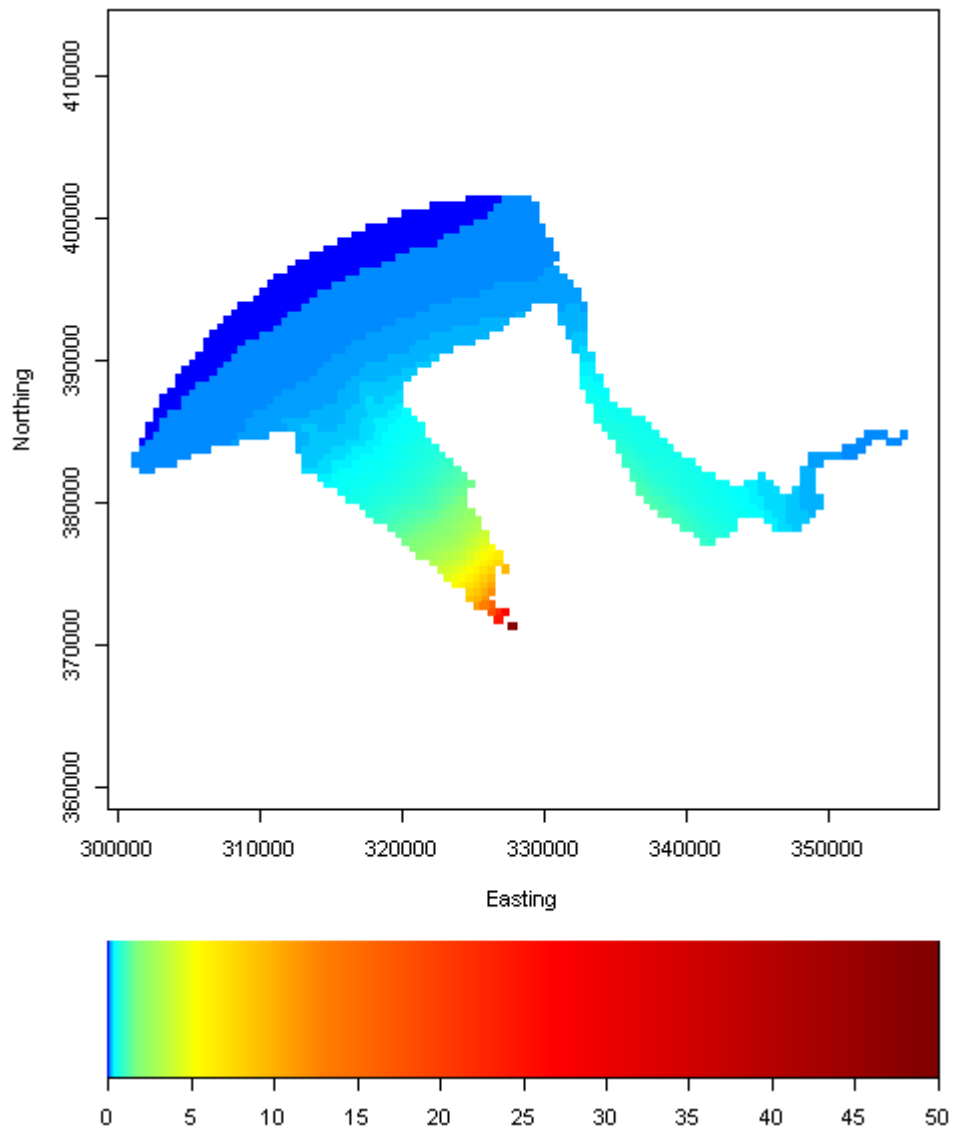
Common Tern, Foulness



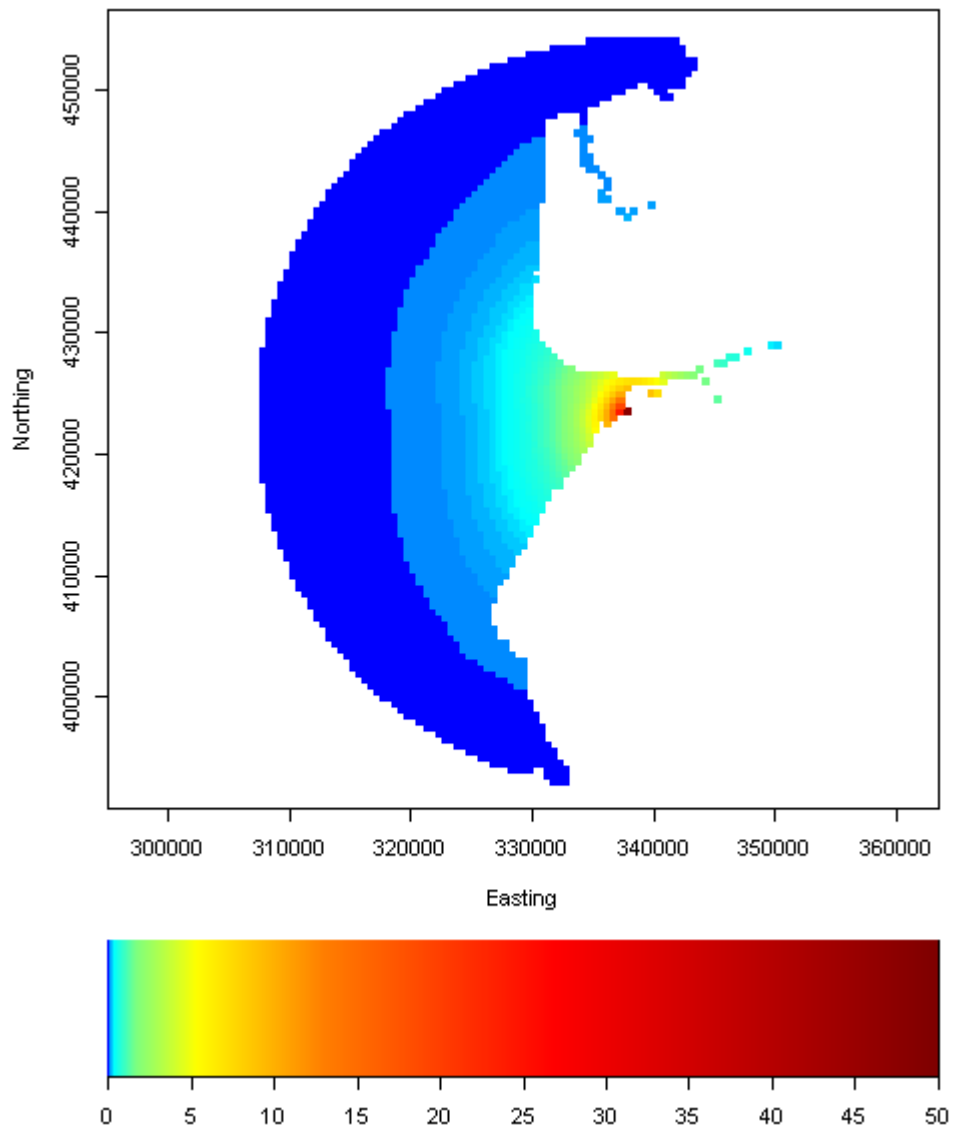
Common Tern, Breydon Water



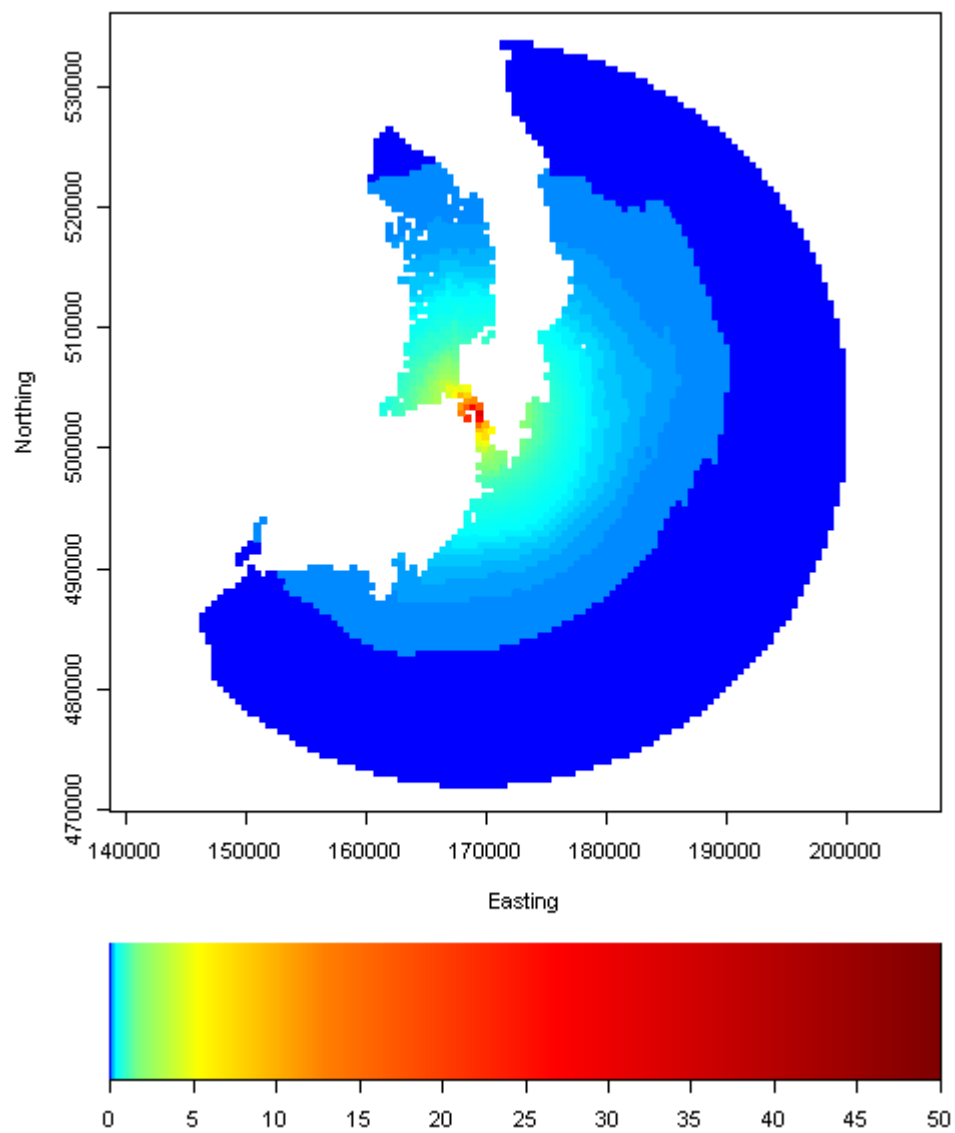
Common Tern, Dee estuary



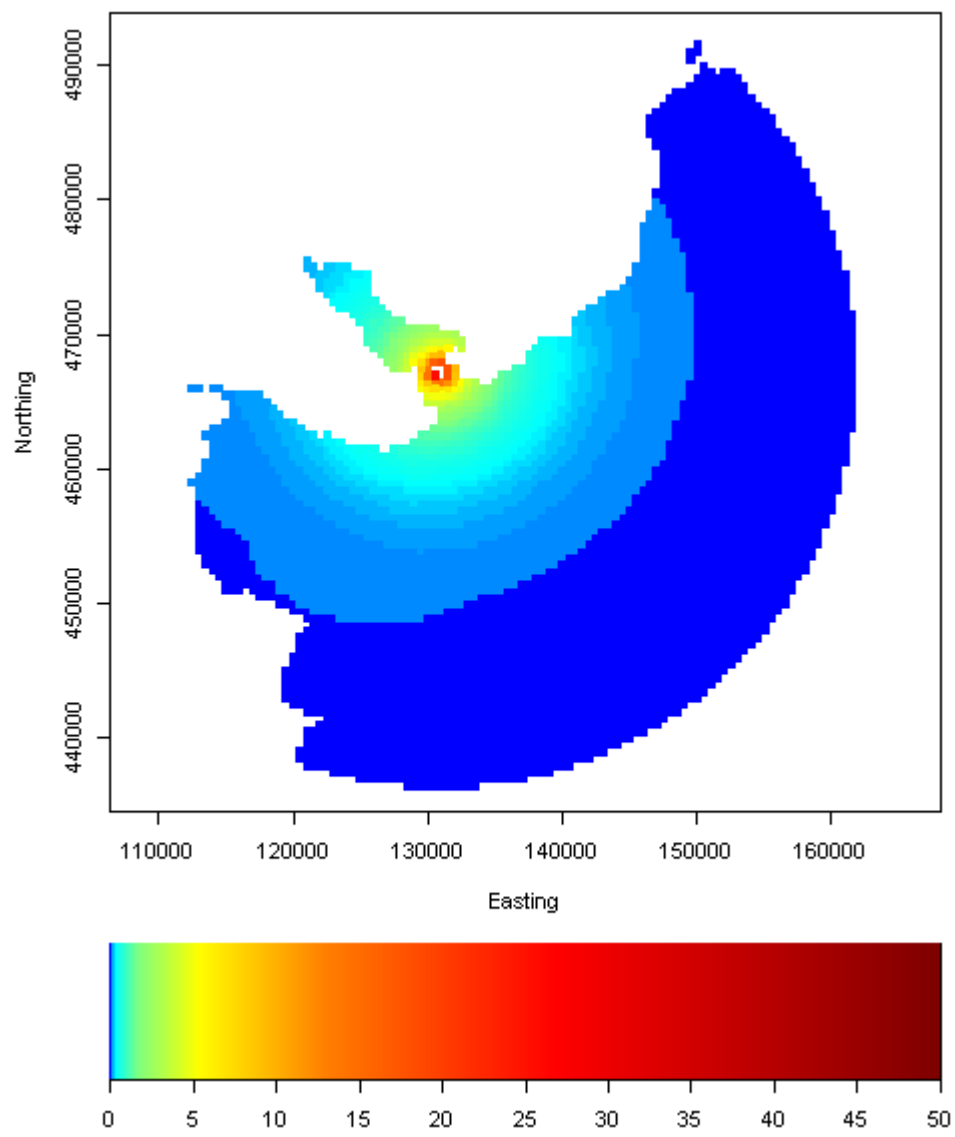
Common Tern, Ribble estuary



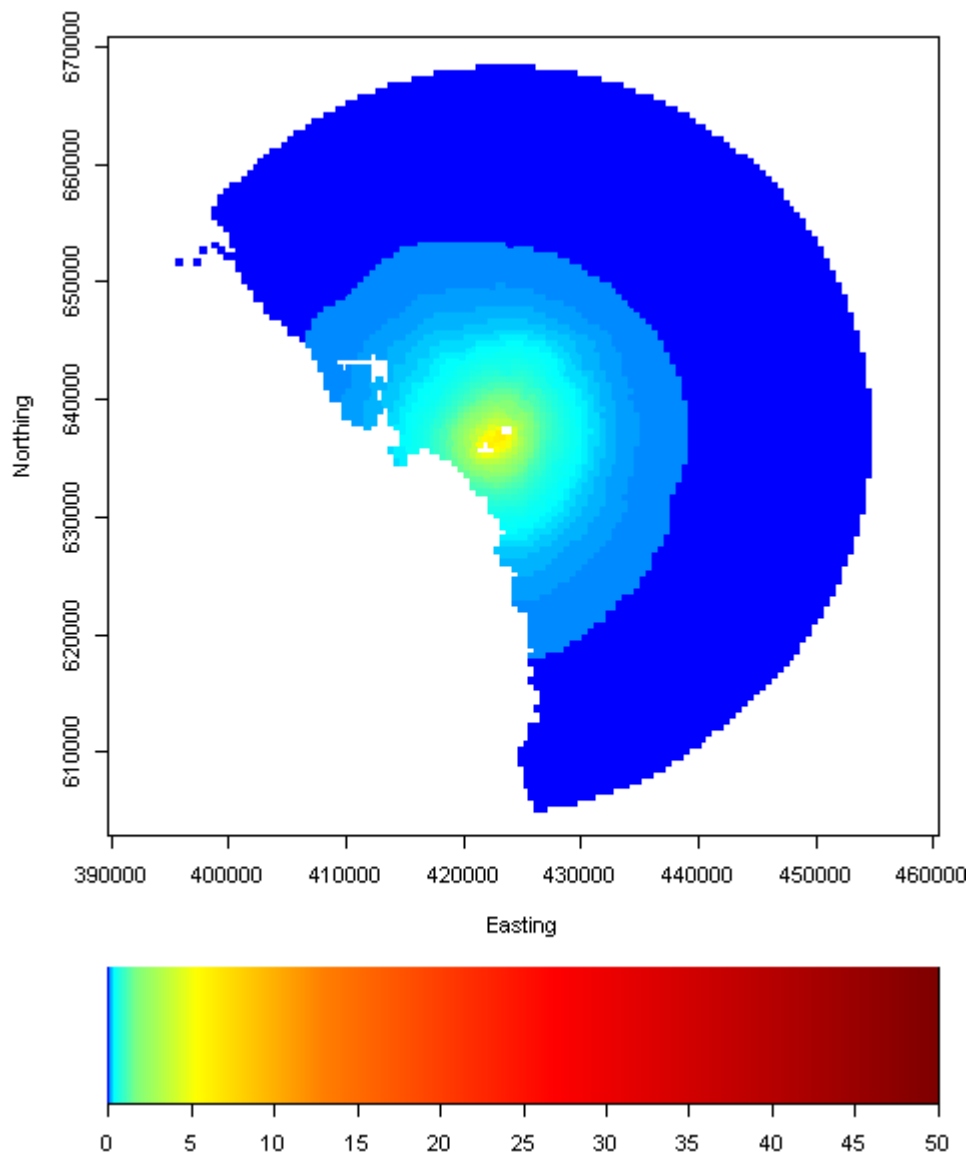
Common Tern, Strangford Lough



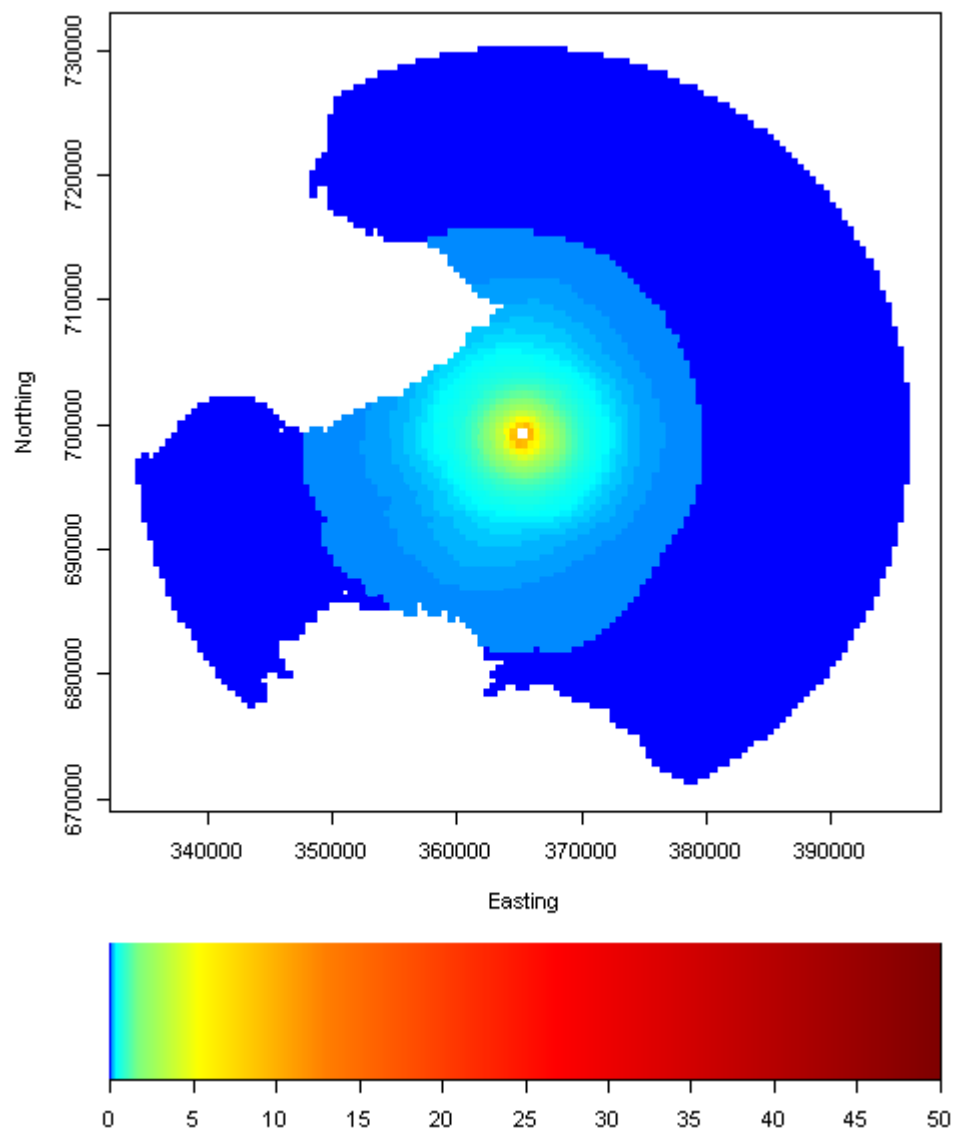
Common Tern, Carlingford Lough



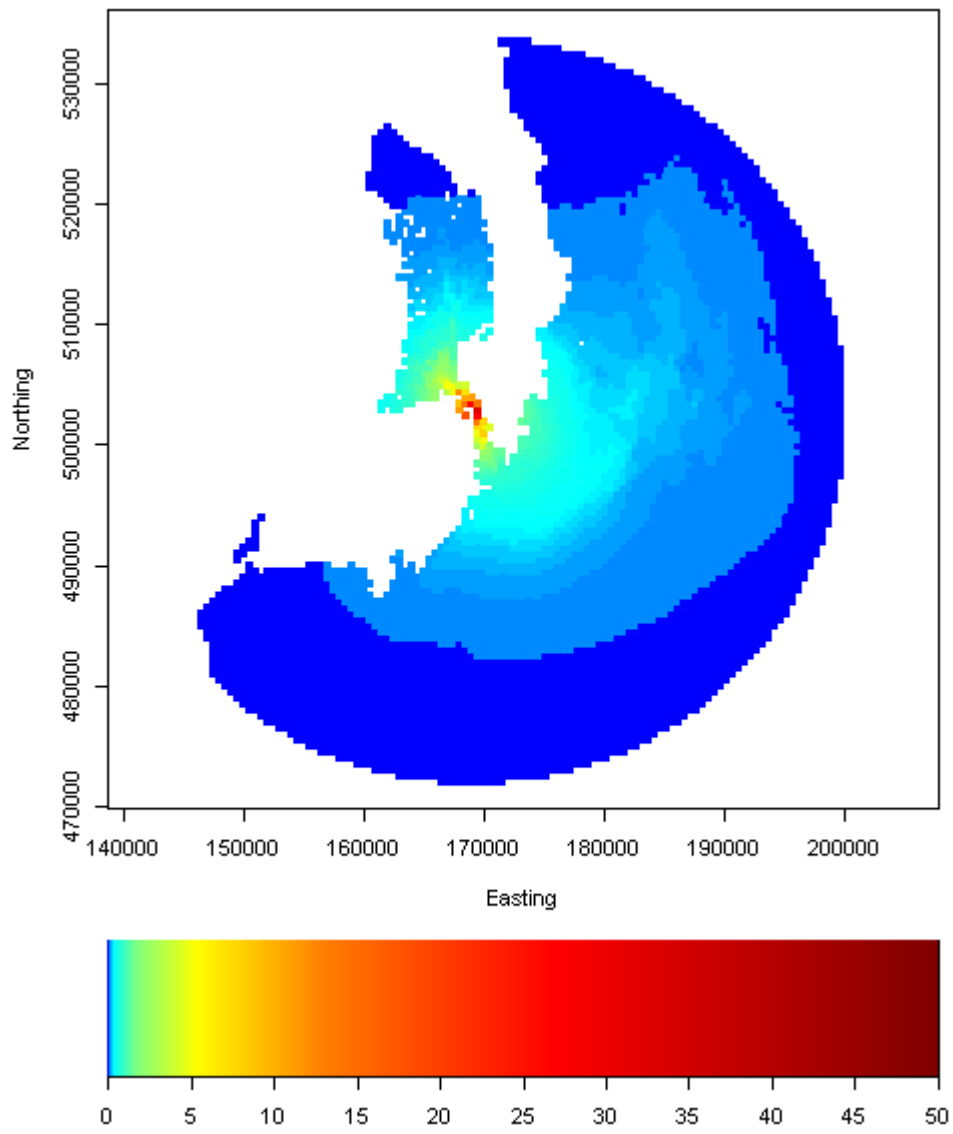
Common Tern, Farne Islands



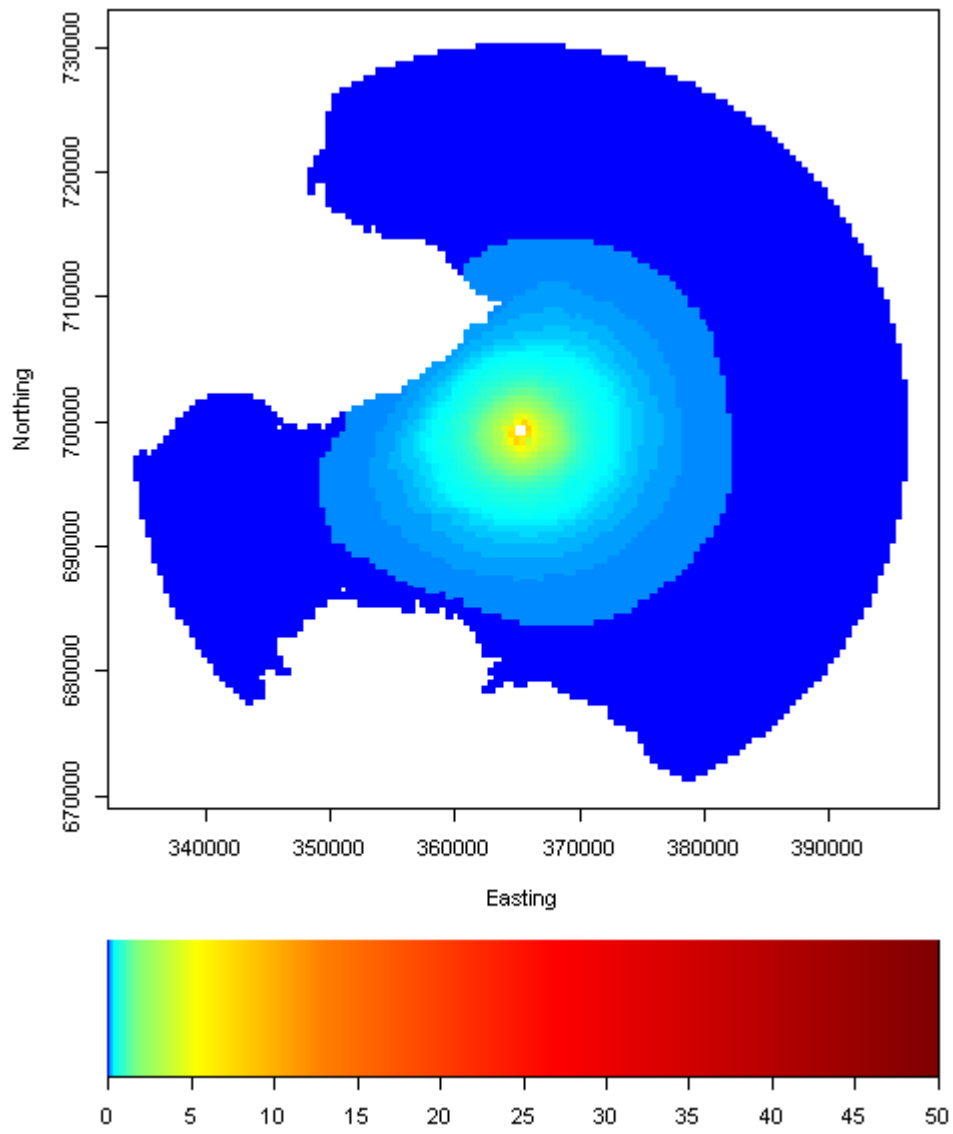
Common Tern, Isle of May



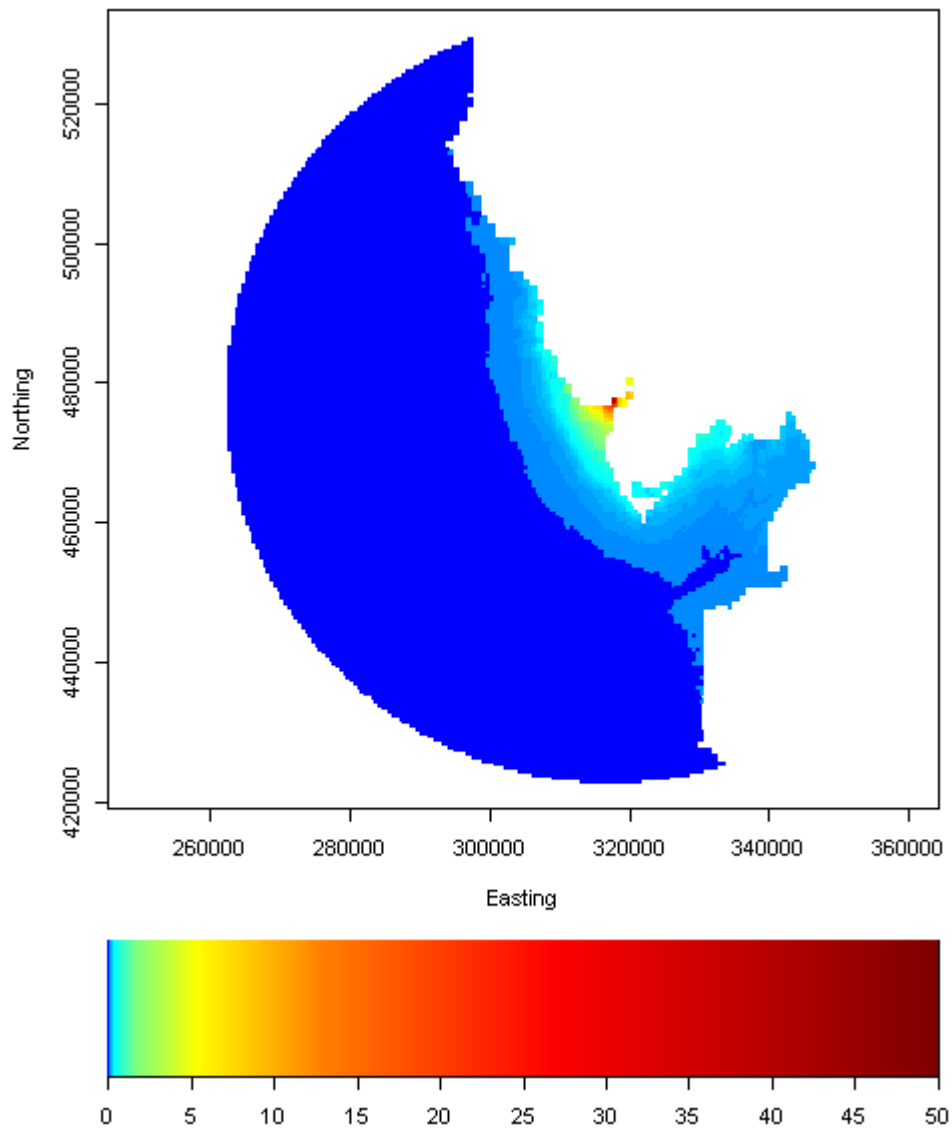
Arctic Tern, Strangford Lough



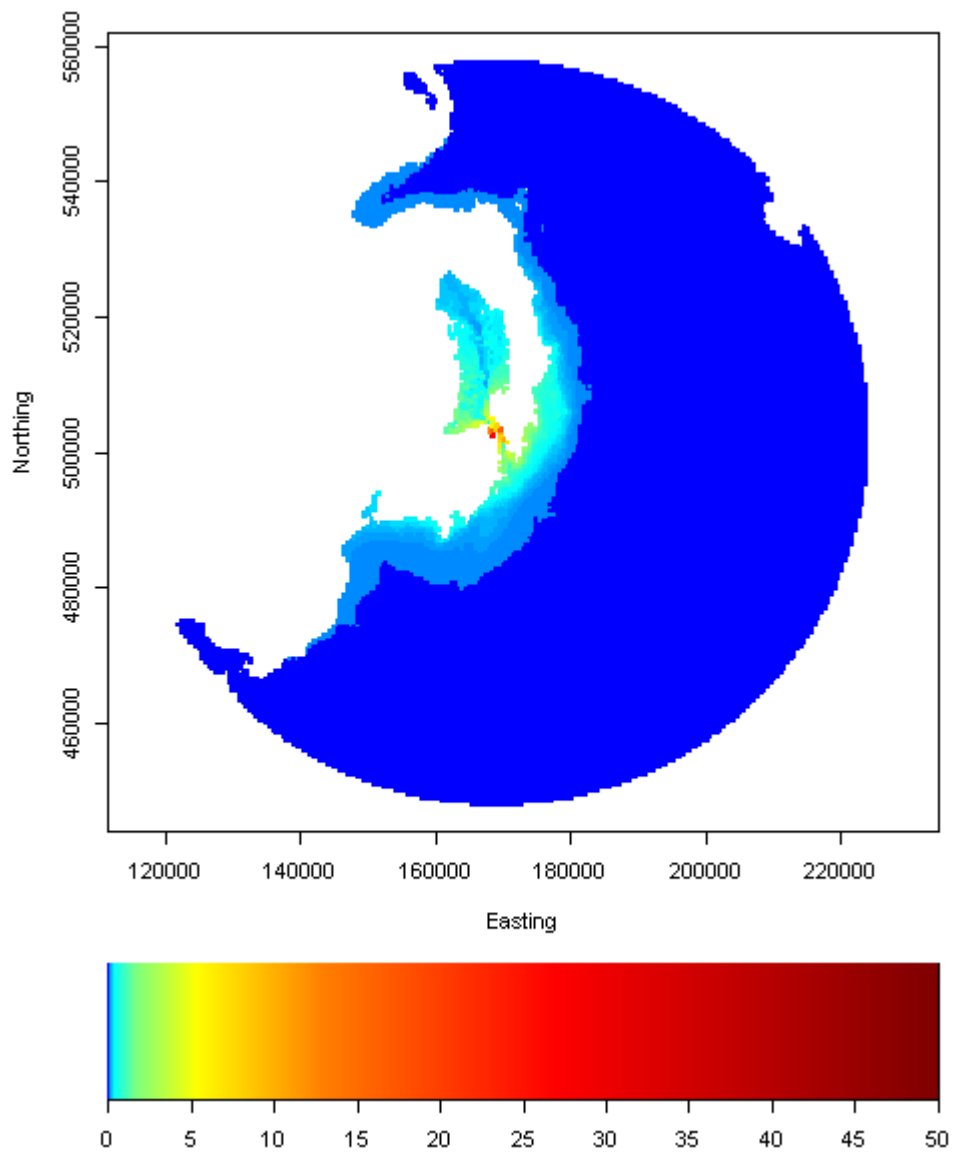
Arctic Tern, Isle of May



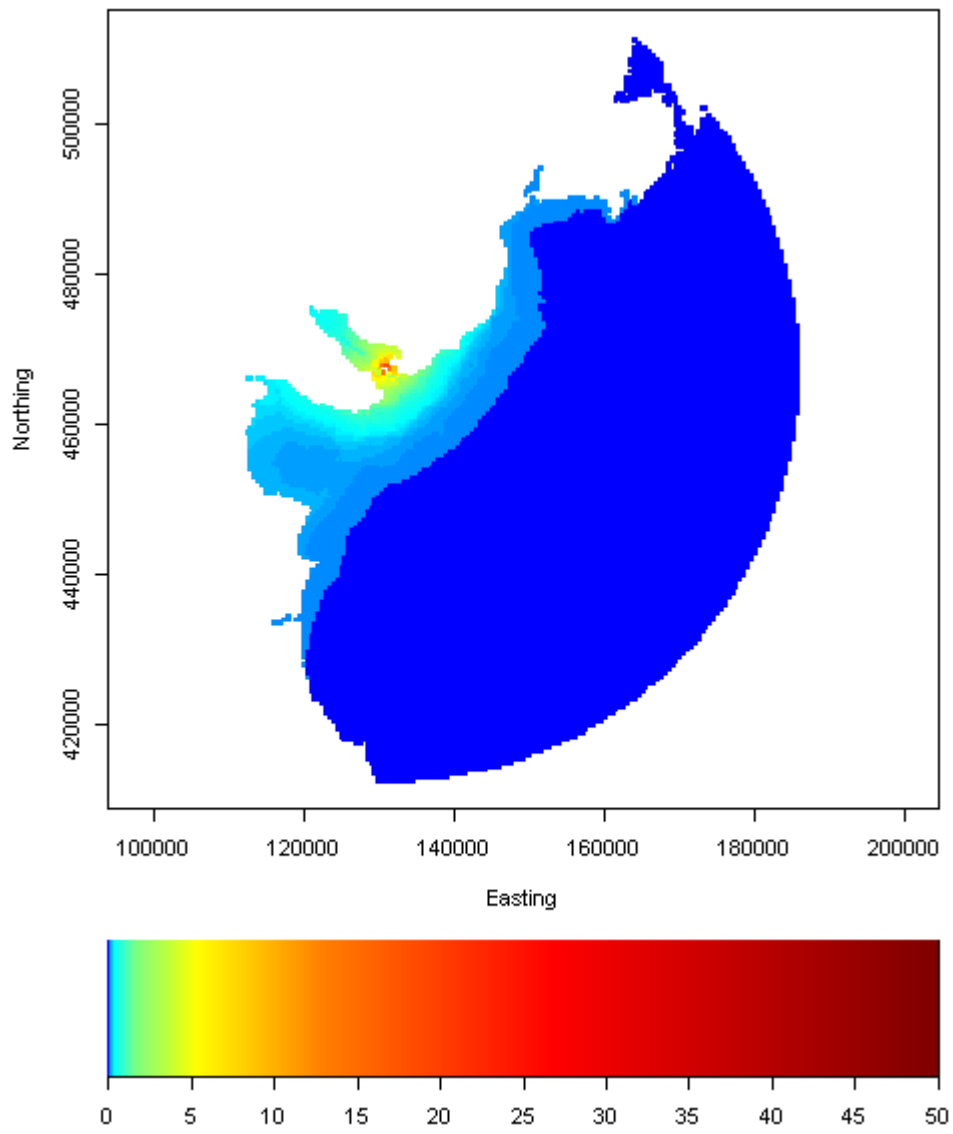
Sandwich Tern, Duddon estuary



Sandwich Tern, Strangford Lough



Sandwich Tern, Carlingford Lough



7. Discussion

Cross-validation has shown that it is generally better to combine all available data for a tern species when making predictions for a new colony, rather than basing predictions on a colony or colonies that appear to be ecologically similar. This means that the model developed for each species is more robust, because it is based on data from a larger number of tracks, and covering a wider range of environments. It might be possible to give the colonies differing weights, but it is unclear how such weights should be chosen, as they would need to take account of the amount of data available for each colony, as well as its ecological similarity to the colony being predicted which is difficult to measure in relation to a terns assessment of its environment. The analysis has also shown that whereas the best models for predicting the available data from a colony may involve many covariates and non-linear terms, simple linear models with a small number of variables (in particular distance to colony, distance to shore, bathymetry and chlorophyll concentration), are better for extrapolating to a different colony.

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.
- Brewer M.J., Potts J.M., Duff E.I. & Elston D.A. (2012). To carry out tern modelling under the Framework Agreement C10-0206-0387. Report submitted to: Joint Nature Conservation Committee.
- Brier, G.W. (1950) Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**, 1-3.
- Chakraborty, A. Gelfand, A.E., Wilson A.M., Latimer A.M. and Silander J.A. Point pattern modelling for degraded presence-only data over large regions. *Applied Statistics* **60**, 757-776.
- INLA (2012) R-Package, <http://www.r-inla.org/>
- Liu C., White, M. and Newell G. (2011) Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography* **34**, 232-243.
- Lobo J.M., Jiménez-Valverde, A. and Real, R. (2008) AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145-151.
- 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.