Biomathematics and Statistics Scotland

Revised model for roseate terns

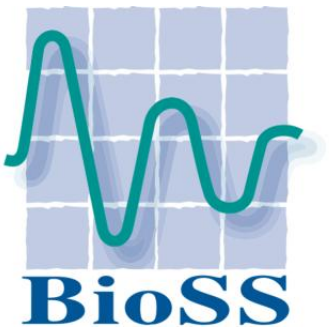
(Under Agreement C10-0206-0387)

CONTRACT No: C10-0206-0387

Report submitted to:

Joint Nature Conservation Committee

May 2014



Authors:

Jackie M Potts, Mark J Brewer

Introduction

This report repeats the analysis described in Potts *et al*. (2013) and Potts & Brewer (2013) for roseate terns at the Coquet colony, omitting chlorophyll, SST and sand variables as candidate covariates. This was because many of the recorded tracks fell in the coastal strip where these variables were missing.

Methods

As in Potts *et al*. (2013) 100 bootstrap samples were formed by sampling with replacement from the tracks. In each case a sample of control tracks was taken by resampling with replacement from the 12 control tracks. The number of times each variable was selected in the model with minimum AIC was counted.

We aimed to find a model that is parsimonious, biologically realistic and with an AIC value close to the minimum. As such, we use the minimum AIC model, but where this model includes covariates which are either selected < 50% of the time in our bootstrapping and/or have biological ranking of > 5 (ranks provided by JNCC), we investigate whether their removal provides an equivalent model (i.e. the AIC difference is <2). If so, we choose to use the model which contains frequently selected and/or high ranking variables rather than the model with the lowest AIC.

Cross-validation was carried out as described in Potts & Brewer (2013). We formed 100 bootstrap samples for each of 10, 20 or 30 tracks in the training set. In each case we held back 30 tracks sampled randomly without replacement for use as a test set and then sampled 10, 20, or 30 tracks randomly with replacement from the training set, together with the corresponding control tracks. Three performance scores were used (the likelihood score, the MSE score and the AUC). An inter-annual cross-validation was also carried out. For 2009 and 2011 we found a minimum AIC model by stepwise selection and fitted the model to the other two years. This was not done for 2010 because only one track was available. Inter-annual cross-validation was also carried out by fitting a model using the covariates selected in the final model based on the full dataset.

Results

The final model is presented in the Appendix. Results from the bootstrapping exercise and the shortlisted models are shown in Tables 1 and 2 below.

Table 1. Frequency with which covariates were selected for 100 bootstrap samples for roseate terns at the Coquet colony, excluding chlorophyll, SST and sand.

Variable	Count
dist_col	85
ss_current	83
dist_shore	66
eastness_1s	61
sal_spring	59
sal_summ	51
spring_front	47
summ_front	46
strat_temp	45
bathy_1sec	19
slope_1s_deg	16
ss_wave	16
northness_1s	13

Table 2. Shortlisted models and AIC statistics for roseate terns at the Coquet colony; the proposed final model is indicated in bold.

Terms	AIC
dist_col, strat_temp, eastness, ss_current	169.80
dist_col, strat_temp, ss_current	171.57
dist_col, strat_temp, eastness	170.77
dist_col, strat_temp	**171.35**
dist_col	212.09

The minimum AIC model involved distance to colony, temperature stratification, shear stress current and eastness. Although both eastness and shear stress current were selected in >50% of the bootstrap samples, they both have ranks >5. Removal of either eastness or shear stress current , or both together, increased the AIC by <2. When temperature stratification was then removed (because it was selected in <50% of the bootstrap samples) from the most parsimonious of these alternative modes, the AIC increased by >2. The final model chosen was therefore one involving distance to colony and temperature stratification.

The average performance scores for the bootstrap samples are shown in Table 3.

Table 3. Average performance scores for bootstrap sample sizes of 10, 20 or 30 tracks.

Colony	Species	Bootstrap Sample Size	Likelihood Score	MSE Score	AUC
Coquet	Roseate	10	-0.272	0.057	0.905
		20	-0.229	0.053	0.917
		30	-0.224	0.053	0.918

The results from the inter-annual cross-validation are shown in Tables 4 and 5.

Table 4. Performance of the minimum AIC model fitted to one year's data when tested on data from other years.

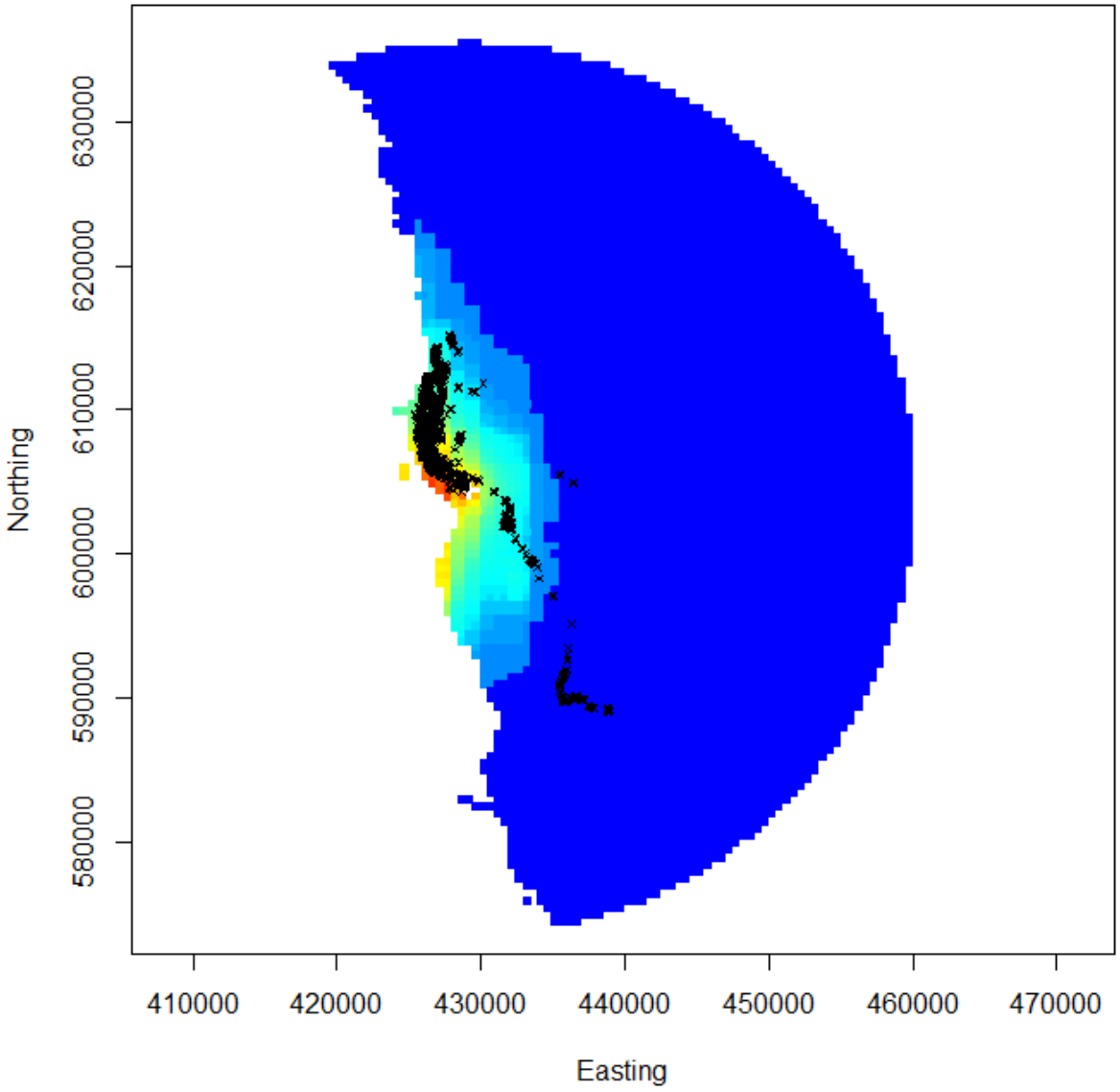
Training Year	Test Years	Minimum AIC Model	Likelihood Score	MSE Score	AUC
2009	2010, 2011	strat_temp, eastness	-0.187	0.055	0.900
2011	2009, 2010	dist_col, dist_shore, spring_front, sal_spring, ss_current	-0.804	0.077	0.878

Table 5. Performance of a model using the covariates selected in the final model using the full dataset (see Table 2) when trained and tested on data from other years.

Training Year	Test Years	Likelihood Score	MSE Score	AUC
2009	2010, 2011	-0.182	0.050	0.919
2011	2009, 2010	-0.342	0.073	0.882

Discussion

Fig. 1. Usage Map for Roseate terns at the Coquet colony.



It should be noted that, with the exception of one very long track recorded in 2010, the terns generally tended to forage to the north rather than to the south of the colony. The model predicts highest usage to the north, but it also predicts quite high usage along the coast to the south of the colony.

However, in the cross-validation study the AUC scores were around 0.9, indicating very good performance.

References

Potts, J. M. Brewer, M.J. & Duff, E. I. (2013). *Refinements of tern Sterna sp. tracking data modelling (Phase 1) – Seabird Tracking Data (Under Agreement C10-0206-0387).* Contract Report to JNCC, September 2013

Potts, J. M. & Brewer, M.J. (2013). *Cross-Validation of tern Sterna sp. tracking data modelling (Phase 1) (Under Agreement C10-0206-0387)*. Contract Report to JNCC, December 2013

Appendix – Results from final model

```
Call:

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

    weights = weights)


Deviance Residuals:

     Min        1Q    Median        3Q       Max

-1.41458  -0.01553  -0.00109  -0.00001   1.64157


Coefficients:

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

(Intercept) 16.48309    3.21534   5.126 2.95e-07 ***

dist_col    -0.09358    0.06876  -1.361    0.174

strat_temp  -5.65131    1.05812  -5.341 9.25e-08 ***

---

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


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 281.72  on 27022  degrees of freedom

Residual deviance: 165.35  on 27020  degrees of freedom

AIC: 30.595


Number of Fisher Scoring iterations: 9
```

```
Single term deletions


Model:
SEARCH_FORAGE ~ dist_col + strat_temp
           Df Deviance    AIC    LRT  Pr(>Chi)
<none>          165.35 30.595
dist_col    1   167.43 30.678  2.083     0.149
strat_temp  1   208.09 71.332 42.737 6.261e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```