# Analysis of Bird Data for the Fall of Warness Area:
# A report prepared for
# AURORA Environmental Limited

DMP Statistical Solutions UK Ltd.
04 September 2006

## Executive Summary

1. The analysis results predict the average number of birds sighted across the entire survey region to be approximately 3 per (daylight) hour. Specifically, we estimate the average lies somewhere between 1.6 and 6.6 per hour (with 95% confidence).

2. There is evidence for differences in bird numbers across sub-areas in the site. Few birds were sighted in sub-areas far from land while as many as 9 birds per hour, on average, were observed close to land.

3. There is also evidence of monthly changes in relative bird abundance. Bird numbers appeared to increase somewhat from October 2005 to March 2006 before appearing to stabilise, though these increases were not statistically significant.

4. Bird abundance was found to be related to a range of environmental factors, and the way in which each variable appears to be related to bird abundance is illustrated in this report.

   Consistent with the earlier analysis, birds were generally more prevalent in the early morning, low winds, and flood/slack tides, slack water flows and when the flow direction was recorded as `North'.

5. As for the earlier analysis, all best and worse case scenarios provided in the report (based on upper and lower confidence limits) were adjusted to correct for the time dependent nature of the data. This is crucial if valid abundance comparisons are to be made across time and if valid comparisons with future results are required.

# Table of Contents

# 1 Introduction

This report gives a review of the statistical analyses of the Fall of Warness bird data provided by Aurora Environmental Ltd. For clarity, technical details of the statistical models are separated from the main results and can be found in §6.

The main objectives of this analysis were:

- assess the distribution of birds across the study site and across time
- provide outputs that summarise their distribution
- ascertain relationships between collected environmental variables and the observed relative bird abundances

The data collection and analysis is thought to be ultimately required to detect changes in the bird usage of the Fall of Warness region through time. This might be in terms of their use of the space, and/or their response to measured environmental conditions e.g. they may shift preferred locations of activity or become active at different times of day. This would form the basis of an environmental impact assessment for the placement of the proposed undersea turbines.

**Initial Caveats:**
The terms *abundance/counts/numbers* will be used throughout this report – this requires initial clarification. The nature of the study (a fixed observation point) means that there is likely to be a lower probability of observing animals the greater the distance from the observation point. For example comparing 2 sites, one far and one near – even with equal numbers of animals, the further site will have an apparently lower abundance due to lower probability of detection.

Due to this, any numbers derived cannot be interpreted as true count estimates without proper account of this detection probability (inestimable for the current data). The terms used here will be synonymous with detected numbers. However, the intended use is for measuring relative change through *time*, which can be ascertained if the study design remains unchanged and detectability remains constant. Relative spatial sightings are similarly affected by detection probability.

It is assumed throughout that inference is restricted to times similar to those sampled i.e. general daylight hours.

The term *bird* is synonymous with the species represented in the dataset provided by Aurora Environmental Limited.

# 2 Data Details

This section briefly outlines the data manipulation, exploratory analysis and final analyses applied. Greater detail is given in the *Technical Details* § 6.

## 2.1 Software

All data analysis was performed in the statistical data package R, with associated addon packages as required[1].

## 2.2 Data Preprocessing

The following are notes relating to the data manipulation performed on the environmental and species count data contained within the files:

Eday seal and bird survey 01-02-06 to 26-02-06.xls
Eday seal and bird survey 01-07-06 to 14-07-06.xls
Eday seal and bird survey 02-03-06 to 31-03-06.xls
Eday seal and bird survey 02-05-06 to 26-05-06.xls
Eday seal and bird survey 02-06-06 to 30-06-06.xls
Eday seal and bird survey 03-04-06 to 29-04-06.xls
Eday seal and bird survey 09_01_06 to 31_01_06.xls

These files were provided to DMP by Aurora Environmental.

**Summary:**

- Tide height for observation 592: assuming this to be Hi.4.0m in keeping with observations before and after.
- Replaced instances of HI with Hi (8 of these)
- Replaced instances of " H" with Hi (one of these)
- Assuming Lo -0.4m are Lo.0.4m (3 instances)
- Changed all tide_time coding to state:hr:min for convenience in text manipuation
- Time entry O942 transferred to 0942 i.e. strictly numeric.
- Removed observation from count data (in obs 837): No zone recorded – can't be exactly inferred from surrounding data
- Observation dropped from count data (in obs 646): Zone recorded as 0-1 – can't be inferred from surrounding data
- Any instances/recordings of Zone 5 were dropped.
- Non-bird observations dropped: 561 instances of these e.g. basking sharks.
- Counts of "100 +" changed to 100
- Observation 967 replaced with 697 (assumed to be an inversion)
- Observation 77 replaced with 777

[1] R Development Core Team (2005) *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

- Multiple observation numbers 826 → replaced with sequence 826-829
- High/Low tide times inferred from one another as required. Assumed 6 hour 12 minute delay between high/low tides.
- High/Low tide heights inferred from one another as required. The empirical distributions for each were derived - the quantile for the observed data was used to infer the unobserved value from the opposing high/low tide distribution.
- Instances of precipitation data requiring alteration:
  - 'CONSTANT' changed to be 'CONSTANT RAIN'
  - 'MIST & RAIN' changed to 'MIST/RAIN',
  - 'NIL' changed to 'NONE'
  - 'OCC RAIN' and 'OCCASIONAL RAIN' changed to 'OCC. RAIN',
  - 'OCC SHOWERS' changed to 'OCC.SHOWERS',
  - 'SNOW  SHOWERS' changed to 'SNOW SHOWERS'
- Several instances of counts over several zones aggregated into one figure – as noted in the "comments" column. These aggregated counts were apportioned evenly over the indicated zones.

## *2.3 Data Available*

All environmental data available were initially considered as candidates for prediction of bird abundance. However, data sparsity meant some variables required exclusion e.g. visibility and sea state were not recorded for a large proportion of the sampling occasions/hours. The environmental variables available for the models of bird counts were therefore:

- **Wind strength**: a score measured from 0--6
- **Precipitation**: 21 categories
- **State of tide**: 6 categories
  - Ebb, Flood, Flood/Slack, High, Low, Slack
- **Water flow speed**: 8 categories
- **Water flow direction**: 5 categories:
  - North, North-West, Slack, South, South-East
- **Cloud cover**: a percentage score
- **Month of year** : 13 months from July 2005 to July 2006
- **Sub-area**: 25 categories as combinations of A, B, C, D, E with 0, 1, 2, 3, 4 which represent the spatial position of the birds sighted (see Figure 10)
- **Time of day**
- **Time at high tide**
- **Time at low tide**
- **Tide height at high tide**
- **Tide height at low tide**

# 3 Modelling Methods

The modelling method used here is very similar to the earlier analysis. Specifically, the method used is flexible and naturally accommodates abundance data collected over time. This helps ensure that the end user obtains a model which returns good quality predictions and realistic confidence about these predictions.

Model flexibility was considered key in this analysis. For example, bird numbers are unlikely to increase (or decrease) at a constant average rate across the day, or with the height at high tide and so flexible curves were permitted for all measurement variables in the model, eg. Cloud cover, time of day and tide height at high tide.

The model selection method used for this analysis was superior to that used for the previous analysis. Specifically, recent advancements in statistical methodology were manually coded to give more defensible model selection results; more details can be found in §6.

Categorical variables (such as precipitation, month and sub-area) entered the model in the traditional way, and a different model coefficient was estimated for each level of these variables. e.g. the number of birds is permitted to be different for each month of each year and for each sub-area.

The confidence attributed to model predictions was also considered important in this analysis. Without identifying a realistic range of plausible bird counts before (and after) any potential impact, it is almost impossible to determine if there has been a real change in bird counts.

Models for data collected over time deserve special consideration. Observations collected during the course of a month are likely to be more similar than observations collected from different months and this similarity/correlation is unlikely to be described by the data variables available in full. Correlation of this sort violates at least one fundamental assumption of traditional analysis methods and ignoring this similarity can result in overconfidence in model predictions. Specifically, one would be more likely to confuse natural fluctuations in bird counts over time with real noteworthy increases or decreases. As a result, statistical methods which naturally accommodate data collected over time were used for this analysis. Further details about the modelling approach can be found in §6.

## *3.1 Modelling Results*

This section gives the results of analyses for all birds combined. There are a total of 18 bird species recorded in the Fall of Warness dataset and of these perhaps 12 are sufficiently abundant throughout the sampling period to permit species specific models. Species specific summary statistics can be found in section §4.

All data variables listed in §2.3 were deemed necessary in the model (using criteria stated in section §6) and a brief description for each variable is provided below.

### 3.1.1 Categorical Variable Results

In each of the following figures, categories are represented along the horizontal axis and model coefficient values on the vertical[2]. In all cases the estimate for a category is given by a small central square, with the 95% confidence bounds represented by vertical lines.

**Plot interpretation:** Higher coefficient values indicate greater predicted numbers of birds. Categories that have confidence bounds that are distinct from one another (i.e. do not overlap when compared horizontally) can be considered statistically different from each other at the 5% level.
Additionally, all interpretations are made assuming all other terms in the model are held constant; e.g. All else being equal, significantly more birds are predicted to be observed in group A than in group B.

[2] Coefficient values are given on the scale of the link function – refer Technical Section §6

- **Wind strength:**

Bird numbers were generally found to decrease with increasing wind strength. Significantly more bird numbers are predicted in wind strengths of *0* compared with categories *2-5* (Figure 1).

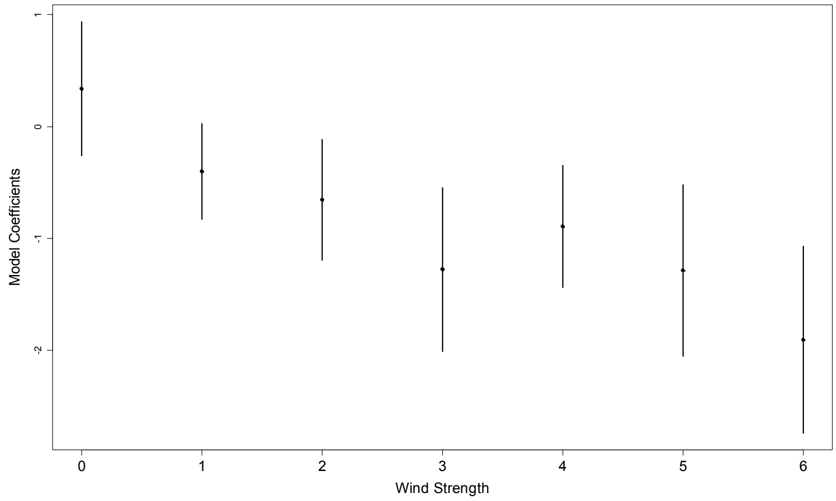


**Figure 1: Relative effects of *Wind Strength* on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

- **Wind direction:**

Bird numbers appeared to be similar across wind direction groups. The estimated number of birds in the *variable* group were fewer on average than in other groups, however this estimate was not statistically distinct from any other groups (Figure 2).

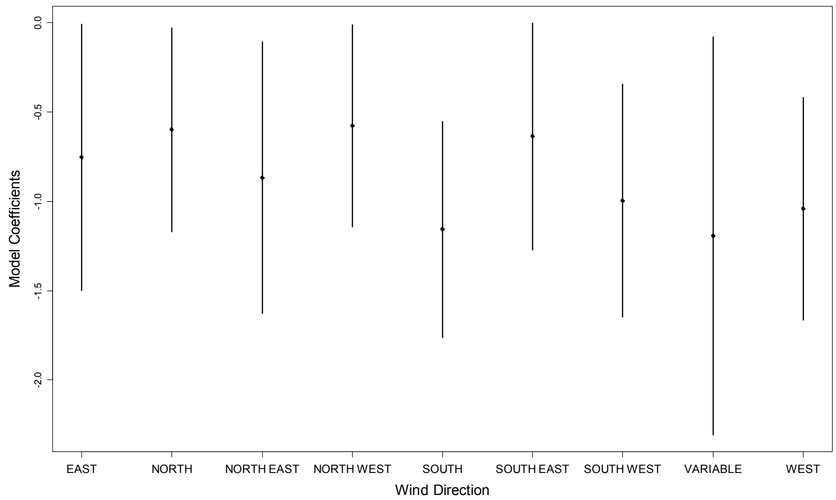


**Figure 2: Relative effects of *Wind Direction* on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

- **Precipitation:**

Bird numbers for precipitation groups were generally indistinct, with only the *Occassional snow* group predicting greater numbers of birds when compared to several other precipitation groups e.g. *hazy* or *mist*. As before, we anticipate visibility issues require consideration when interpreting these results.

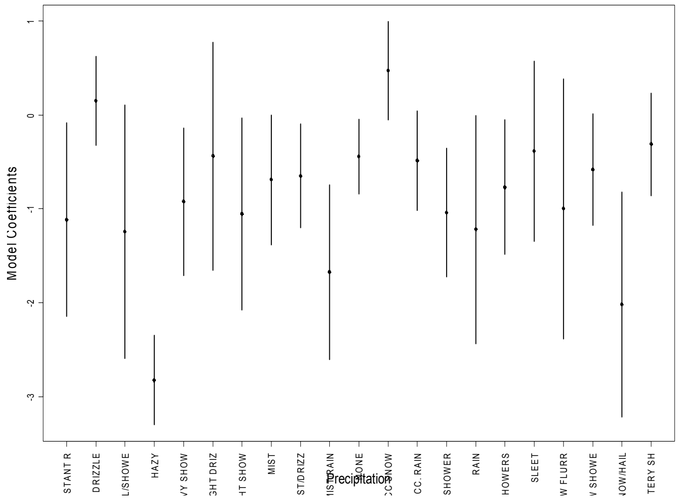


**Figure 3: Relative effects of *Precipitation* group on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

- **Tidal State:**

Bird sightings were most abundant in the *Flood/Slack* tide state (Figure 4) and this estimate was statistically distinct from the *Ebb*, *Low* and *Slack* tide states. Significantly more birds were also sighted during *Flood* tides compared with *Ebb, Low* or *Slack* tide states.

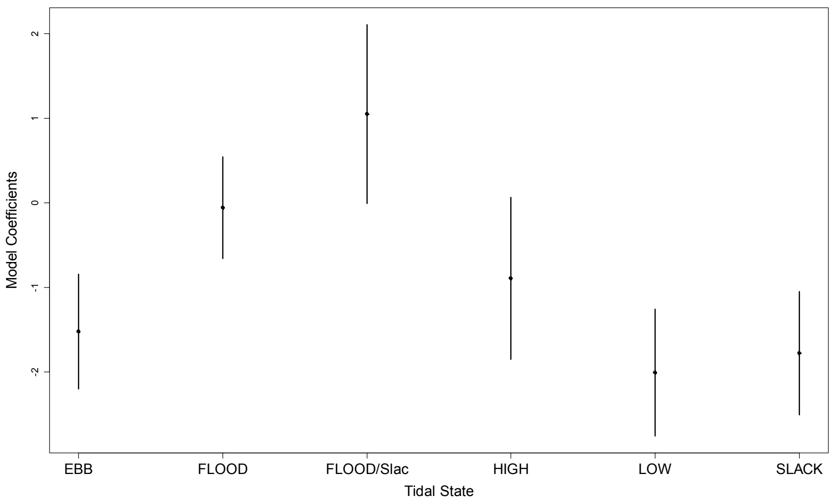


**Figure 4: Relative effects of *Tidal State* on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

- **Water flow speed:**

Only the *Slack* category of *Water Flow Speed* was found to differ significantly from any other levels (Figure 5). This category exhibited significantly higher bird numbers than all other categories.

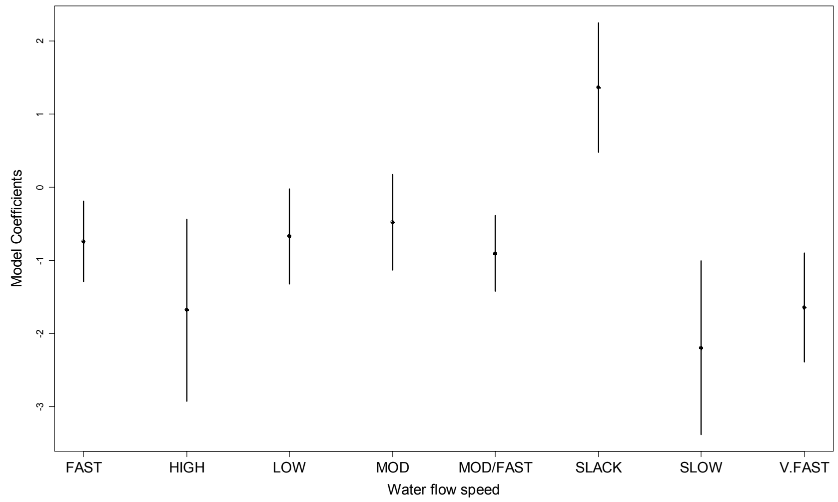


**Figure 5: Relative effects of *Water Flow Speed* on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

- **Flow Direction:**

Significantly more birds were associated with the *Northerly* flow group compared with the *North-West*, *Slack* and *South-East* groups (Figure 6). *NorWesterly* flows give greater numbers than *Slack* or *SouEasterly* flows, while *Slack* and *SouEasterly* flows exhibit equivalently low numbers of birds.

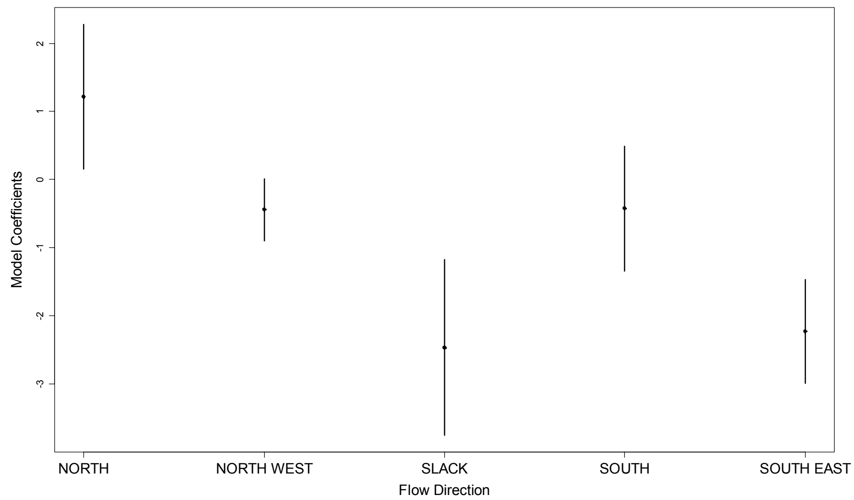


**Figure 6: Relative effects of *Flow Direction* on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

* **Month:**

June 2006 is estimated to have the highest numbers of birds (Figure 7) and significantly more than July 2005 -- February 2006. In contrast, September 2005 was predicted to have the lowest number of birds and significantly fewer birds than those predicted in January -- July 2006.

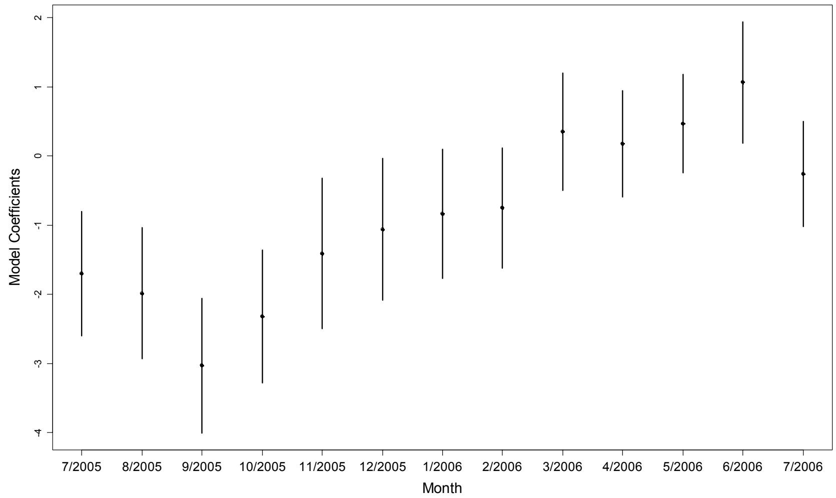


**Figure 7: Relative effects of *Month* on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

* **Sub-area:**

All sub-areas give significantly higher predicted bird numbers than sub-areas *A0* and *A1,* which have the lowest estimates. Additionally, sub-area *A3* is predicted to have significantly more birds than sub-areas *B0*, *B2*, and *E4*. The spatial relationships between sub-areas are better displayed in Figure 9 (following page) and Figure 10 (page 3-14).

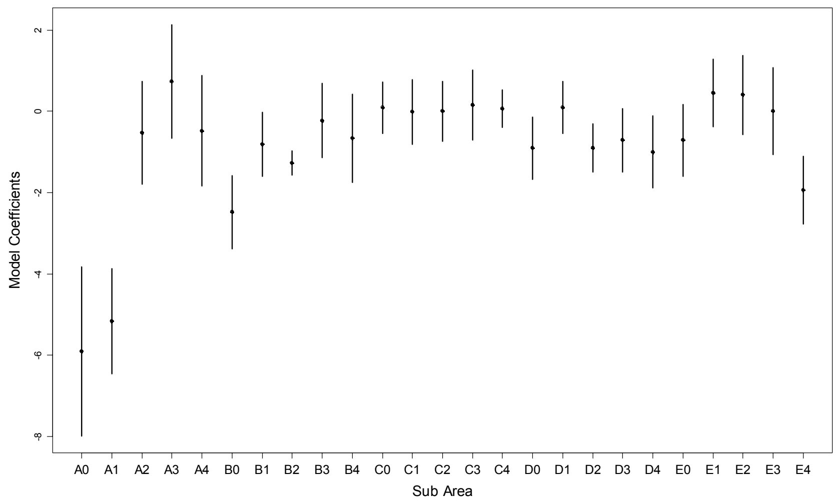


**Figure 8: Relative effects of *Sub Area* on estimated numbers of birds. Dots indicate estimates, vertical lines the 95% uncertainty bounds.**

The predicted average number of birds per hour, by sub-area, are represented in Figure 9 . These are placed in increasing order, with the exception of the predicted average number of birds per hour for the entire survey region (in blue). These values are also shown in Table 1 (page Table 13-17) for each sub-area and for the site as a whole.

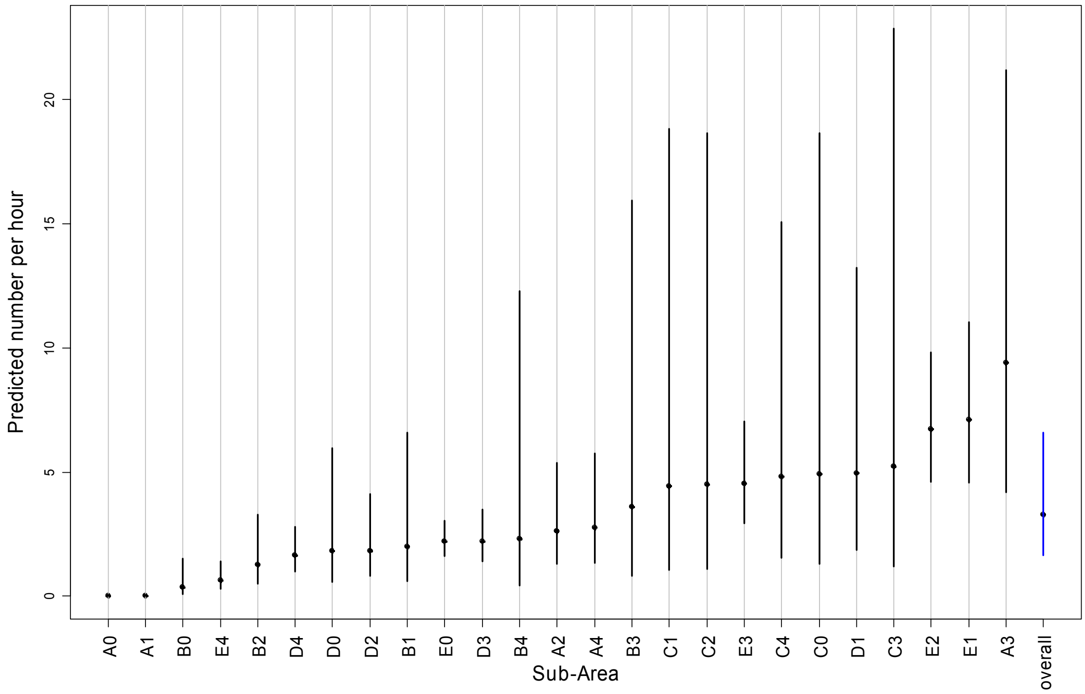


**Figure 9: Predicted numbers of birds by sub-area. Dots indicate estimates, vertical lines the 95% uncertainty bounds. The blue *overall* denotes the overall estimate (i.e. average over all sub-areas combined).**

The following plots (Figure 10-Figure 12) convey the same information as Figure 9 but with their 2D spatial locations made explicit. Note sub-areas in row 4 and column A were not always bounded on the originally supplied map – the outer borders below were subjectively added for plotting purposes.

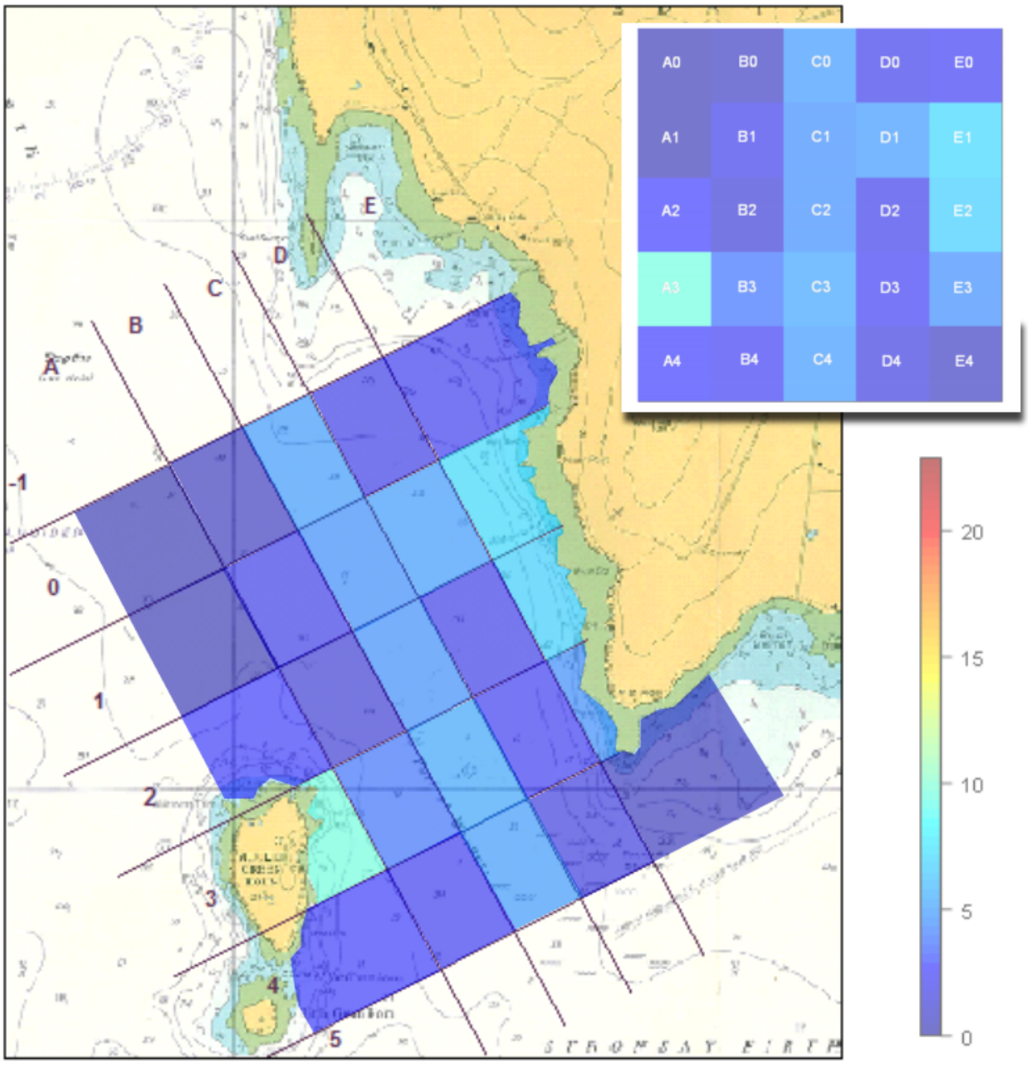


**Figure 10: Estimated relative bird abundance (per observation hour) by sub-area. *Note: estimated abundances are subject to the caveat of page 1-4.* Grid as supplied by Aurora-Environmental in shown on page 6-27.**

The sub-areas of highest estimated hourly counts are those very near the island opposite the survey observation post, and the land adjacent to the survey position. Considering uncertainty in these estimates (Figure 11 and Figure 12) this spatial pattern is still apparent. The caveat relating to detection probability (§1) is reiterated.

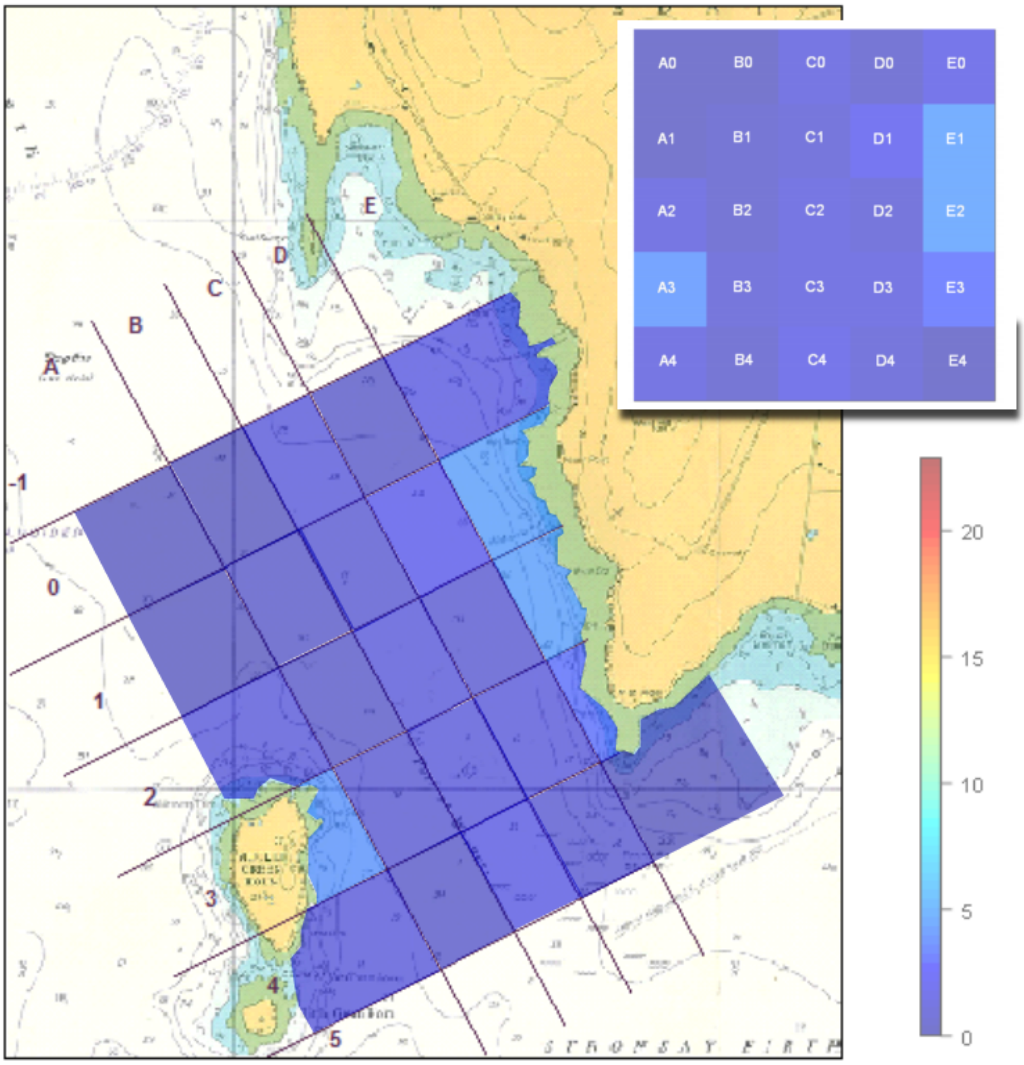


**Figure 11: Lower 95% confidence bound on numbers of birds per hour by sub-areas, relating to estimates in Figure 10**

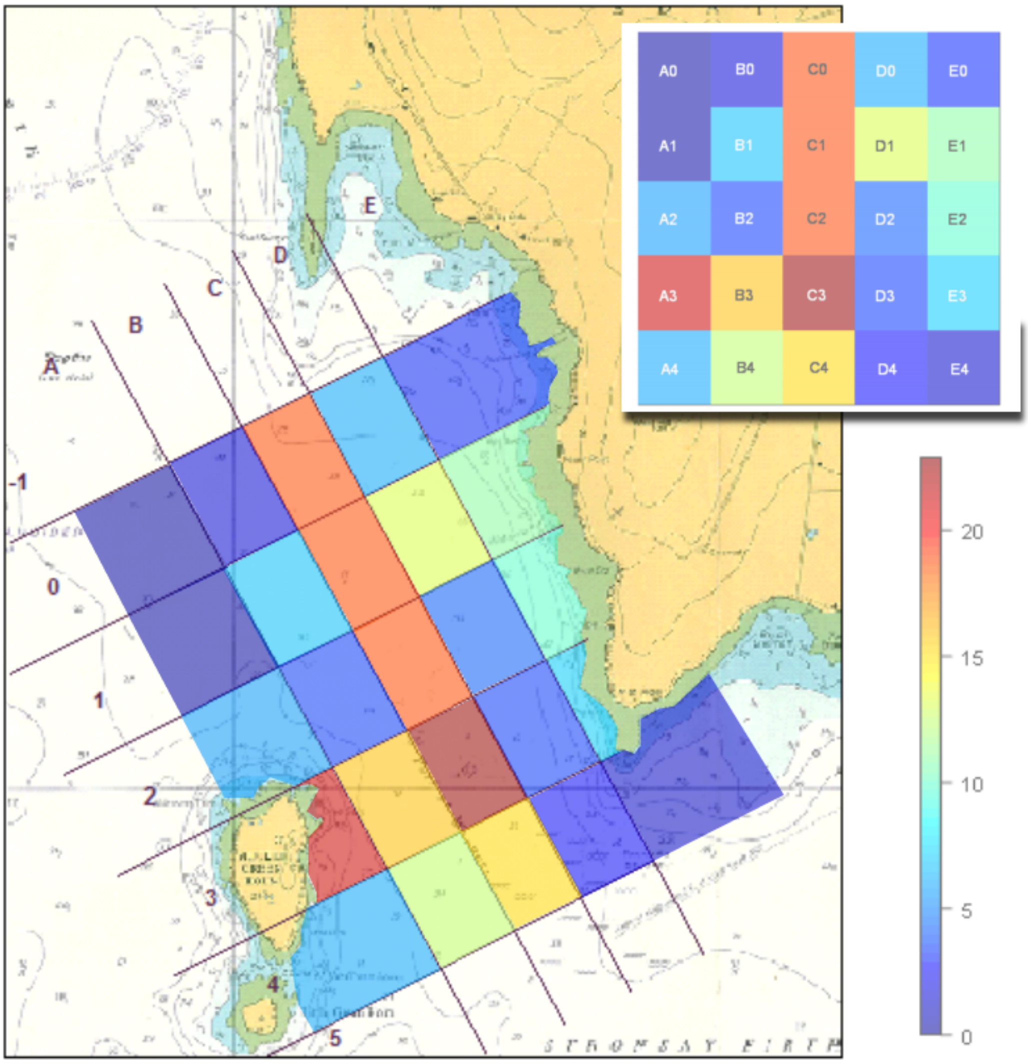


**Figure 12: Upper 95% confidence bound on numbers of birds per hour by sub-areas, relating to estimates in Figure 10**

| Sub-area | Lower limit | Average | Upper limit |
|---|---|---|---|
| A0 | 0.0025 | 0.0122 | 0.0597 |
| A1 | 0.0105 | 0.0256 | 0.0623 |
| A2 | 1.2951 | 2.6369 | 5.3693 |
| A3 | 4.1704 | 9.3999 | 21.1873 |
| A4 | 1.3403 | 2.7796 | 5.7638 |
| B0 | 0.0953 | 0.3775 | 1.4950 |
| B1 | 0.6135 | 2.0099 | 6.5857 |
| B2 | 0.4870 | 1.2650 | 3.2858 |
| B3 | 0.8049 | 3.5812 | 15.9331 |
| B4 | 0.4372 | 2.3185 | 12.2963 |
| C0 | 1.3046 | 4.9323 | 18.6454 |
| C1 | 1.0478 | 4.4411 | 18.8234 |
| C2 | 1.0909 | 4.5100 | 18.6454 |
| C3 | 1.1922 | 5.2216 | 22.8694 |
| C4 | 1.5425 | 4.8230 | 15.0790 |
| D0 | 0.5577 | 1.8241 | 5.9655 |
| D1 | 1.8556 | 4.9520 | 13.2170 |
| D2 | 0.8200 | 1.8375 | 4.1173 |
| D3 | 1.4041 | 2.2160 | 3.4977 |
| D4 | 0.9772 | 1.6537 | 2.7985 |
| E0 | 1.5981 | 2.2060 | 3.0453 |
| E1 | 4.5663 | 7.1022 | 11.0475 |
| E2 | 4.6145 | 6.7281 | 9.8100 |
| E3 | 2.9385 | 4.5421 | 7.0217 |
| E4 | 0.2952 | 0.6481 | 1.4228 |
| | | | |
| **Whole site** | **1.6357** | **3.2818** | **6.5843** |

**Table 1: Predicted average number of birds sighted per hour for each sub-area for all bird species combined. The upper and lower limits for these averages were obtained using the GAM/GEE estimates (see section 6 for details).**

### 3.1.2 Continuous Variable Results

There is some uncertainty about the shape of the relationships between the continuous variables available for these data.

- **Cloud cover:**

Higher bird numbers were predicted at approximately 50% and 100% cloud cover under the model (Figure 13). However, there is great uncertainty about this relationship, thus over-interpretation of the relationship between *Cloud Cover* and the number of birds is ill-advised.

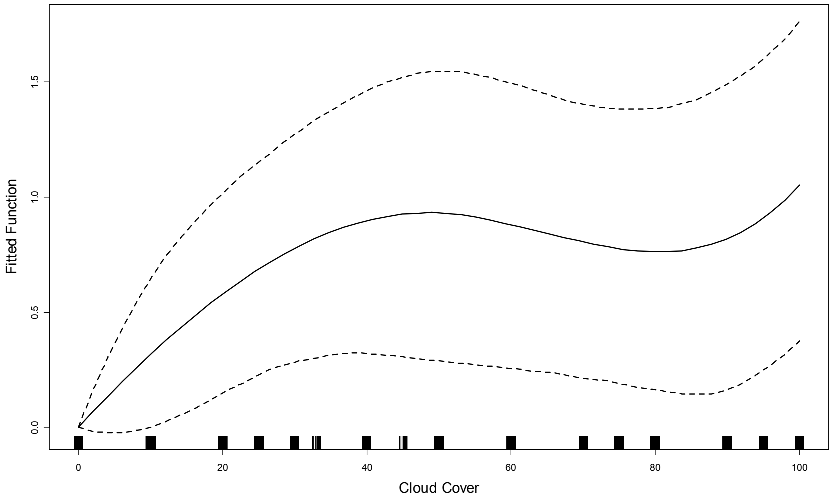


**Figure 13: Effects of *Cloud Cover* on estimated numbers of birds. Solid curve indicates estimated function, dotted curve above and below indicate the 95% confidence envelope (envelope derived from GEE estimates – refer §6).**

- **Time of day:**

Under the model, the greatest numbers of birds are found early in the sampling day, falling away after this point (Figure 14). Naturally, low light levels may have a confounding effect on these results.

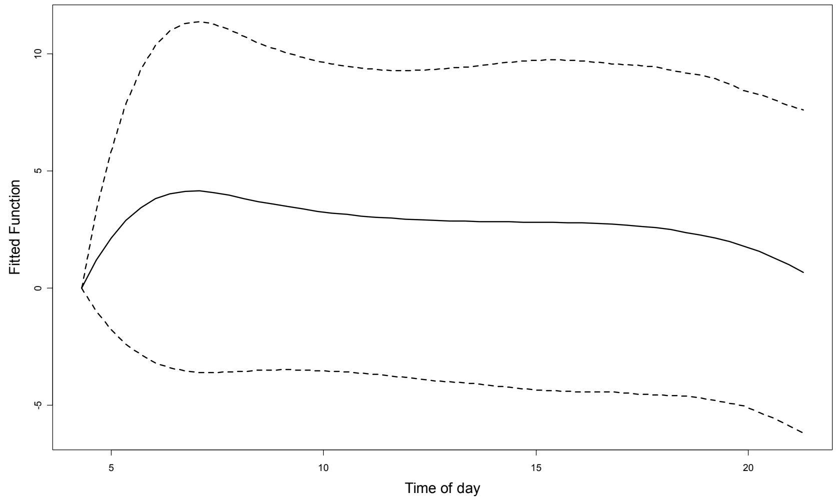


**Figure 14: Effects of *Time of Day* on estimated numbers of birds. Solid curve indicates estimated function, dotted curve above and below indicate the 95% confidence envelope.**

- **High tide height:**

Bird counts are predicted to increase under the model until the high tide height reaches approximately 3m, after which bird numbers are predicted to stay relatively constant (Figure 15). The uncertainty about model predictions for hide tide heights in excess of 3.5m are very unreliable and should be used with caution.

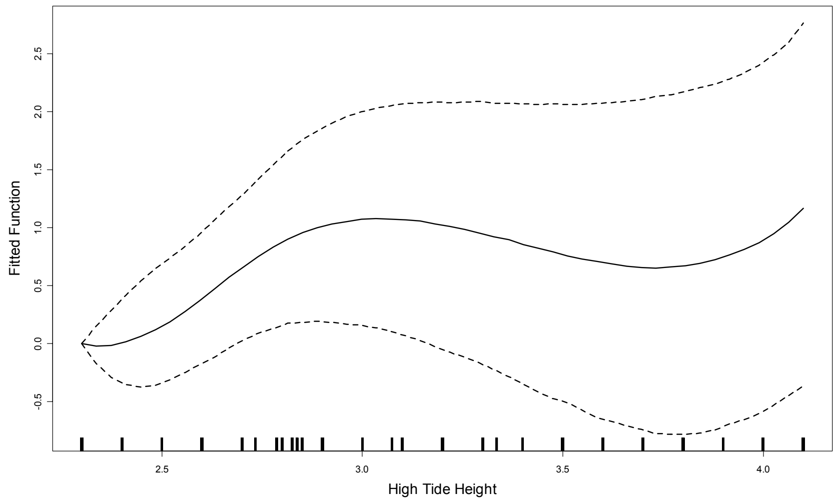


**Figure 15: Effects of *High Tide Height* on estimated numbers of birds. Solid curve indicates estimated function, dotted curve above and below indicate the 95% confidence envelope.**

- **Low tide height:**

Bird counts are predicted to steadily increase for low tide heights in excess of approximately 1m. There is also however, a great deal of uncertainty about this relationship (Figure 16).

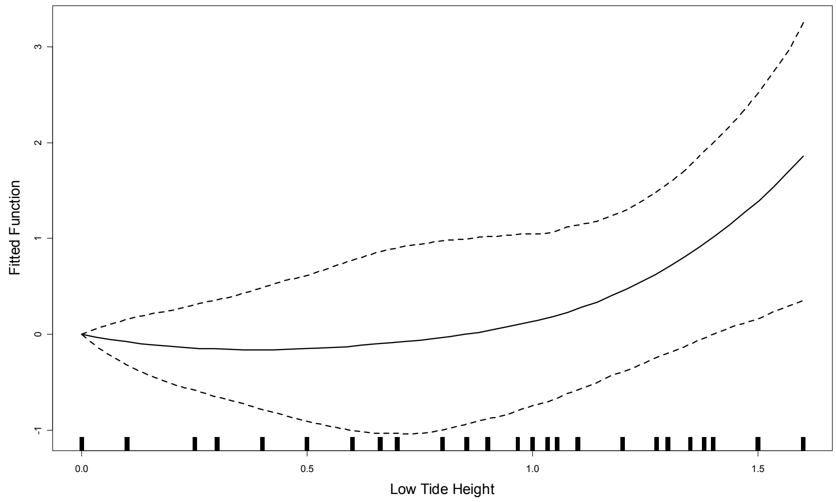


**Figure 16: Effects of *Low Tide Height* on estimated numbers of birds. Solid curve indicates estimated function, dotted curve above and below indicate the 95% confidence envelope.**

- **High tide time:**

Bird counts are predicted to be relatively constant for high tides from early morning until about 10am after which bird numbers are expected to increase and stabilise (Figure 17).

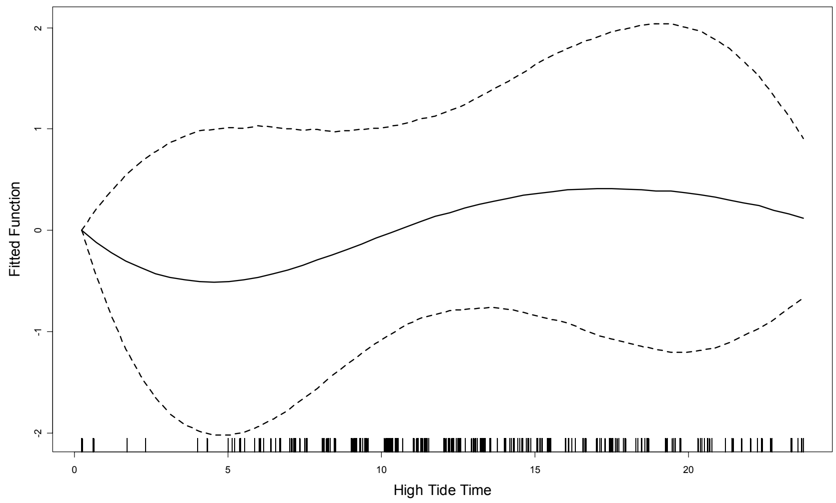


**Figure 17: Effects of *High Tide Time* on estimated numbers of birds. Solid curve indicates estimated function, dotted curve above and below indicate the 95% confidence envelope.**

- **Low tide time:**

Bird counts are predicted to be variable with respect to low tide times (Figure 18). There are tentatively greater counts at low tide times about 3pm however, the edges of spline models are frequently unreliable and so should be interpreted with caution.

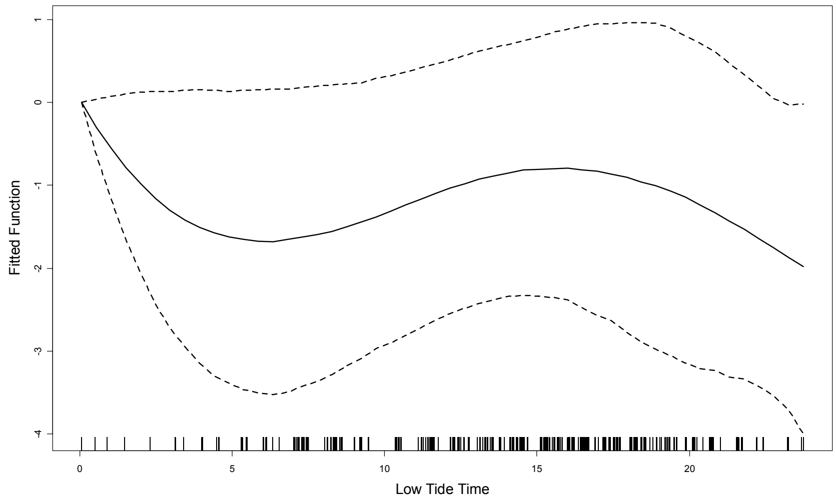


**Figure 18: Effects of *Low Tide Time* on estimated numbers of birds. Solid curve indicates estimated function, dotted curve above and below indicate the 95% confidence envelope.**

# 4 Species Specific Summaries

The mean number of birds sighted per hour are shown in descending order below:

| Species | Average number of birds seen per hour | Conservation Designation | | |
|---|---|---|---|---|
| | | EU Habitats | EU Birds | UK Amber |
| GUILLEMOT | 39.2363 | | | ✓ |
| BLACK GUILLEMOT | 13.571 | | | ✓ |
| EIDER DUCK | 10.8218 | | | ✓ |
| SHAG | 8.1938 | | | ✓ |
| ARCTIC TERN | 3.1503 | ✓ | ✓ | ✓ |
| KITTIWAKE | 2.7326 | | ✓ | |
| PHALACROCORAX SPP (nb: may be in addition to simple Shag counts) | 1.8881 | | | |
| GANNET | 1.5907 | | | ✓ |
| PUFFIN | 1.4788 | | ✓ | ✓ |
| CORMORANT | 0.8249 | | ✓ | |
| RED-THROATED DIVER | 0.5078 | ✓ | ✓ | ✓ |
| RAZORBILL | 0.3596 | | | ✓ |
| GREAT NORTHERN DIVER | 0.1990 | ✓ | | ✓ |
| LONG-TAILED DUCK | 0.0870 | | | ✓ |
| DIVER | 0.0187 | | | |
| RED-BREASTED MERGANSER | 0.0031 | | ✓ | |
| LITTLE AUK | 0.0010 | | | |
| SLAVONIAN GREBE | 0.0010 | ✓ | ✓ | ✓ |

**Table 2: Average number of birds sighted per hour for each bird species. Species are ordered by the average number seen per hour. `EU Habitats' represents birds listed on the EU Habitats Directive Annex I, `EU Birds' represents birds listed on the EU Birds Directive SPA list, and `UK Amber' represents birds listed on the UK Amber list.**

# 5 Conclusions

- This study provides information that may be used to determine bird abundances relatively *through time*. This can provide a basis for temporal monitoring by permitting comparisons across time.

- A sound statistical model has been formulated to allow relative bird abundances to be predicted for the site as a whole under differing environmental conditions, and for each sub-area. Further, special attention has been paid when presenting best and worse case scenarios (using 95% confidence limits) to ensure comparisons made during this analysis (eg. across sub-area) and any future comparisons are reasonable.

- The average number of birds sighted per hour appears to differ across the sub-areas sampled, with greater numbers detected in the sub-areas near land, eg. areas E1, E2 and A3.

  Specifically, about 4.8 birds were sighted per hour on average for sub-areas E1, E2, E3 and E4 combined, while 9.4 birds were sighted per hour on average in area A3. These numbers can be contrasted with sub-areas furthest from the observation point (A0, A1 and B0) which are predicted to have less than 0.4 birds sighted per hour on average.

- The average number of birds sighted appeared to be related to many factors and all variables measured were considered important in the model. In general, more birds were recorded in the early morning, low winds, and flood/slack tides, slack water flows and when the flow direction was recorded as `North'.

- Bird abundance appears to be seasonal. Specifically, more birds were sighted in March-June 2006 when compared with other monthly values, however bird numbers for these months were only statistically distinct from estimates from July-October 2005. Confirmation of seasonal patterns can only be confirmed with several years data.

- All bird species were combined in this analysis, however this method could also be used to consider species level abundances through time for many of the recorded species.

## 5.1 Limitations/Caveats

Absolute estimates of bird abundance are unable to be obtained from the data at hand. For instance, there may be the possibility of repeatedly counting the same individuals through time and this could seriously bias estimates of absolute abundance. However, relative abundance information can be extracted which is suitable for monitoring changes through time.

We continue to have concerns about the apparent relationship between the distance from the observer and observed bird abundance. i.e. predicted bird abundance is highest near land. For instance it is well known that the probability of detecting an animal decreases as the distance from the observer increases and animals are more easily identified against a contrasting background. If detection primarily determines the number of birds recorded by the observer rather than the number of birds present, then this model will not adequately reflect underlying differences in sub-area to sub-area abundances[3]. That said, if the sampling and observer protocol and detection rates stay constant with time, valid comparisons can still be made across time using this approach.

Augmentation of the current survey design would go some ways to addressing these detection concerns. Specifically, additional observers placed at (randomly) chosen locations could provide bird counts concurrent with the observer already in place, and any bias in the current design could be objectively evaluated. Further, this augmented design could also be used to correct for any biases which emerge; counts for affected sub-areas could be inflated or deflated according to the extent and nature of the bias revealed.

More data is required to fully account for seasonality. As more data is collected, longer term seasonality effects could also be tested for inclusion.

[3] this is not an uncommon issue. If the detection function can be modelled then absolute abundance figures can be generated.

# 6 Technical Details

## 6.1 Exploratory Data Analysis

1. Scatterplots and boxplots were used to explore the relationships in the data and to check for outlying values.

2. Variance inflation factors (VIFs) were used to check for collinearity. In this analysis, the VIFs indicated no prohibitive levels of collinearity and all variables were considered as candidates in the model.

## 6.2 Modelling Approach:

Models for the average number of birds sighted per hour were fitted using Generalized Additive Models (GAMs) with log link and Quasi-Poisson errors. Specifically, splines were used to model the continuous explanatory variables while the majority of the variables were fitted as factors.

Generalized Estimating Equations (GEEs) were used to accommodate temporal auto correlation in the errors. Specifically, observations within months were permitted to be correlated and since the data set was of considerable size and GEE standard errors are robust to the nature of the correlation specified, a working independence structure was used.

## 6.3 Variable Selection

Model selection is not straightforward for GEE based models since these are not likelihood based. For this reason, a quasi-likelihood measure (the QIC) was manually coded and used to select a model. The continuous variables could enter the model as smooth or linear terms, while the factor variables (with all levels) were tested for inclusion. In keeping with the last analysis, this QIC retained all candidate variables and all continuous variables were retained as smooth terms in the model.

This model has a large number of parameters and as a consequence is computationally demanding to fit. Despite the size of this model, an objective measure designed for model selection of this type indicated that this level of complexity was justified.

## 6.4 Modelling Results

The effects of accounting for the time dependence in the model residuals were marked. As expected the temporal autocorrelation across time was estimated to be positive and thus, confidence intervals were considerably wider under the GEE scheme compared with those that ignore the temporal correlation.

In most cases, the majority of the factor levels appeared to be statistically distinct from each other, which was not the case when the correlation was incorporated in the model. All confidence bounds presented in this report are based on GEE standard errors.

The plots in section 3.1.1 and 3.1.2 were created using the coefficients on the scale of the link function and these are not adjusted for multiple comparisons.

Predictions for the sub-areas and overall were obtained in the following way:

- The predicted average number of birds sighted per hour for each sub-area were obtained using the estimated model coefficients and the observed values for the explanatory variables. The upper and lower limits for each sub-area were based on the GEE standard errors.

- The predicted average number of birds sighted for the whole area (per hour) was obtained by averaging the predictions across the sub-areas (which was considered valid given the environmental covariates were not sub-area specific). The upper and lower limits for this value were obtained using GEE standard errors.

## 6.5 Model Assessment

Model fit was assessed using the percent of deviance explained[4] while Pearson's residuals were used to identify any ill-fitting behaviour. Cooks distances were used to detect influential points.

The fit of the model was reasonable and there was no evidence of ill fitting behaviour or influential points. Approximately 36% of the deviance was explained by the model.

[4] 1-(Residual Deviance/Null Deviance)x100

## 6.6 Reference Grid

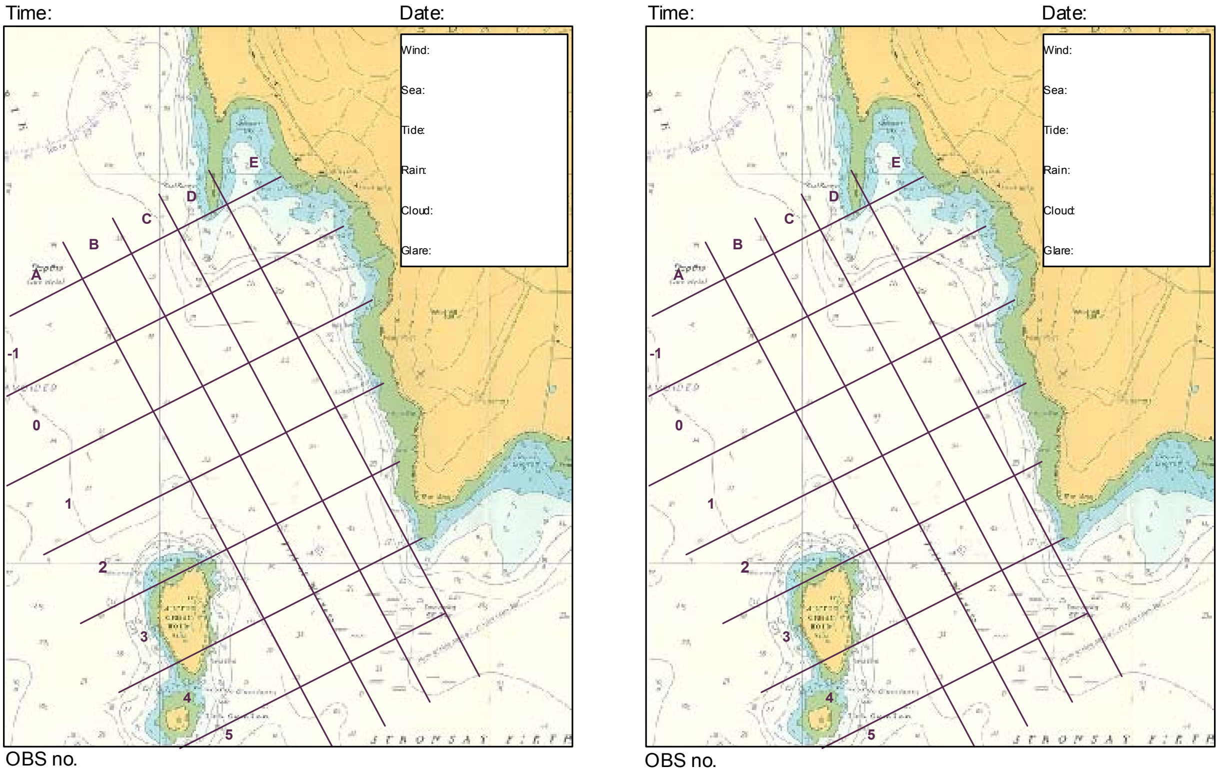


**Figure 19: Eday survey grid as supplied by Aurora Environmental Limited. Alternate versions with colour coded estimates of relative bird abundance are shown in Figure 10 on page 3-14.**