TIDE Tool: 
Guidelines for the waterbird habitat analysis methodology

Authors: A. Franco

Institute of Estuarine and Coastal Studies (IECS), 
University of Hull, UK

March 2013
GUIDELINES FOR THE WATERBIRD HABITAT ANALYSIS METHODOLOGY

1 Tool outline

The distribution of waterbirds in estuarine habitats and the identification of the main factors affecting bird habitat use has been investigated within the TIDE report “Determinants of bird habitat use in TIDE estuaries”.

A statistical approach has been proposed for this type of study combining bird count data to the characterization of environmental conditions (including natural habitat areas, water quality parameter and indicators of anthropogenic disturbance) in multivariate analysis (community distribution models) and species-habitat regression models to identify a series of habitat requirements for different bird species. The aim of the analysis is to identify/investigate some of the basic bird habitat requirements along an estuary/tidal river length, incorporating data from a few external drivers.

Using the methods applied to one of the case studies, namely the study of wader assemblages in the Humber Estuary, the present tool aims at providing general guidelines on the methodological approach to the analysis of environmental determinants of bird habitat use in estuarine areas.

The tool structure guides the reader through the main steps of the methodological process. Each phase is briefly introduced in general terms, with description of the most relevant issues and caveats, and details on the methods that can be applied is given by using the example of their application to the case study.
2 Analysis routine

2.1 Choice of the response variable

A first step is the choice of the response variable to be analysed in relation to environmental determinants in the estuary, i.e. bird count data. Bird count data would ideally be from a dataset which offers a reasonable time series as well as good spatial coverage of the studied estuary.

Low water data would be preferable, as a direct relationship with intertidal habitat use could be established. However, there is often a limited availability of these data (e.g., low-tide counts are carried out with a lower frequency and spatial coverage compared to high-tide counts), and this would reduce the size of the analysed dataset, with consequent limitations on the statistical methods that can be applied. This is the case, for example, for most of the low water count data available in the UK and in Germany. In these circumstances, the most comprehensive data set is usually for usage captured around high water, mainly related to bird roosting sites. These data can therefore be used as a response variable in the analysis, with the caveat that the interpretation of the results (in particular the relationships with resources available in the intertidal area) would be less straightforward and it should take into consideration the uncertainty given by the possible mismatch between the habitat use at high tide and low tide. It should be noted that the interpretation of the results of any analysis that is carried out would be constrained by the characteristics of the data that have been analysed as well as by the assumptions made during the analysis.

Example from case study (waders analysis in the Humber Estuary)
Maximum annual counts for 18 wader species on high-tide roosts in estuarine spatial units were analysed. The main focus of the analysis was on the spatial distribution of bird species, but also temporal variability was accounted for by including data collected in different years.

In the Humber Estuary, data for 11 units (WeBS sectors) covering the North bank of the estuary (Fig. 1) were available between 1991 and 2011 (WeBS national survey (leading to a dataset size of 183 observations per species, as some data were missing for certain sectors in certain years).

In order to allow comparison between units of different size, count data were standardised to densities (ind/km²) before any analysis, based on the area of each sector. Sector area at low tide was considered as the basic spatial unit for the analysis, in order to allow characterisation of the habitat available to birds for use in each sector. The possible use of sector length instead was explored, but this was not considered a good indicator of the habitat area available within a sector, due to the fact that the width of intertidal habitat was not proportional to the sector length (wider mudflats are often located in the outer estuary, whereas mudflats are narrower in the inner areas).

High water counts are carried out when many of the preferred intertidal feeding habitats are temporarily unavailable, although in some locations for some species and on some tides, foraging is still available in the upper shore, birds simply moving up shore with the tide. As such, whilst some species will be
roosting in the area around the count period, others may still be feeding. Furthermore, in most instances movement to a roost at this time will be relatively short and often still within a count sector and as such, there is still considered to be a reasonable correlation between the assemblage present at high water in a sector, and the low water assemblage in the same location. Therefore, although the standardisation of high water bird counts over the wider sector area at low tide might lead to an underestimation of the birds using an area at low tide, particularly for those species which can move inland in search of food when their estuarine food resources are not accessible (e.g., Redshank and Oystercatcher), a certain site fidelity of most species is likely, and thus the obtained density data are considered reasonably representative of the bird use of the wider area at low water.

However, where available (in good time series and spatial coverage), low-tide count data would be preferable as this is likely to allow to obtain better relationships of the actual density with the habitat availability and characteristics in the counting unit areas.

Fig 1 WeBS sectors – Humber Estuary, UK. Sectors coded with letters (NA1 to NK are those used in the analysis).

2.2 Choice of the predictor (explanatory) variables

Predictors (explanatory variables) of the bird use of estuarine habitats should be selected by taking into account the possible sources of variability within an estuary.

In general terms and depending on the data availability within an estuary, explanatory variables should include:
A proxy for the location along the estuarine salinity gradient (e.g., average salinity, salinity class (according to Venice classification), distance of the sector from the estuarine mouth); alternatively, the use of any sectoral management zonation used for the estuary (based on other metrics that are deemed of greater importance than salinity) could be used;

- Proxies for habitat and resources availability (e.g., % coverage area of different habitats (intertidal mudflat, pebble/stone substratum, marsh) in each sector);

- Proxies of possible anthropogenic influence (e.g., presence of man-made structures affecting the natural habitat availability; disturbance due to presence of walking paths, roads; distance from ship way to low water mark, number of ship passages);

- Broader trends of the species to account for any variability in the species abundance over time (for single species analysis only). This might be best at a national level, but could arguably use a NW Europe /East Atlantic Flyway metric;

- Proxies for water quality parameters that are deemed relevant to the bird habitat use in the estuary (through direct or indirect effect).

The actual proxy that will be selected will depend on what is considered to be more meaningful for a particular estuary. For example, the choice of a proxy for anthropogenic disturbance will need to reflect site-specific stimuli importance, e.g., water-based disturbance (like vessel traffic) will be an issue on narrow riverine estuaries whereas shore-based disturbance could be more relevant elsewhere.

The choice of the explanatory variable used as a proxy will depend also on the data availability for the studied estuary. In fact, data for explanatory variables should be associated to each of the observations obtained for the response variable (e.g., for different spatial units and in different years), although some assumptions could be made based on the specific spatial and temporal variability of the explanatory variable (see example below for details). Both continuous variable (e.g., average salinity) and categorical variables (e.g., could be used as predictors, and different sources and techniques (e.g. GIS) could be used to collate the data needed (see example below).

It is of note that the data collated for the different environmental variables might highly differ in terms of spatial and temporal coverage of the birds dataset in the studied estuary, and that the selection of the observations for which all the environmental data are available could lead to a drastic reduction of the dataset available for the analysis of the species-environment relationship, to the point that this might possibly prevent any statistical analysis on it (this was the case, for example, of the datasets on habitats and water quality parameters for the Weser Estuary, as these two dataset were only marginally overlapping, with a total of 6 observations between 1992 and 2003). In this case, one might need to choose to carry out separate analyses for separate environmental datasets. In the case of the Weser, for example, the analysis (assessing the relationship between bird distribution and environmental characteristics) was carried out separately for the two types of abiotic characteristics (habitat data and water quality) in this estuary; the dataset for the Weser for which both bird densities
and habitat areas were available covered all the salinity zones (from freshwater to polyhaline), including 43 observations, between 1984 and 2003, whereas the dataset for which both bird densities and water quality parameters were available included 92 observations, between 2004 and 2009, covering only the freshwater and oligohaline zone. As mentioned before, the results obtained from the different analyses would be highly dependent on the dataset that has been analysed and the results could not be directly comparable. This should be taken into consideration when interpreting the results.

Example from case study (waders analysis in the Humber Estuary)

In order to investigate the relationship of spatial/temporal distribution of bird species/assemblages with environmental changes, data on several environmental variables were collected for the Humber Estuary. These generally covered all sectors with the exception of NA1 and NA2 (hence 9 sectors) and all years between 1975 and 2011 (37 years) (=333 observations). However, given the different temporal/spatial coverage of the datasets for waders (183 observations) and wildfowl (324 observations), subset of the habitat data were selected to match the data availability in the two bird datasets.

Salinity (explanatory continuous variable):
The mean salinity in each sector was calculated based on different sources (Gameson 1982\(^1\), Falconer & Lin 1997\(^2\), HUMBER salinity zonation 2000-2010). Given the data availability, spatial variability was only considered, i.e., the same salinity value was allocated to each sector in different years (Fig 2).

![Average salinity](image)

Fig 2. Average salinity in the WeBS sectors, Humber Estuary.

Habitat coverage (explanatory continuous variable):
Habitat maps are available for the Humber Estuary for the years 1975, 1993 and 2008. The habitats mapped were: Intertidal habitat, Subtidal habitat, Supralittoral (Marsh) and Supralittoral (No-Flood Zone). The area covered by each of these habitats within each sector was calculated by overlapping the sectors areas with the habitat maps. In order to obtain a continuous temporal coverage of the data between 1975 and 2011, a monotonous constant decrease/increase of the area of each habitat between years was considered

---


(based on straight lines connecting 1975-1993 and 1993-2008; data for 2009-2011 inferred based on the trend 1993-2008) and the habitat area was calculated accordingly for the missing years. In this way both spatial and temporal variability was taken into account (Fig 3 and 4). Habitat maps did not cover sectors NA1 and NA2, therefore these sectors were not included in the analysis of bird-habitat relationships.

Fig 3. Habitat area (mean ± SD, km²) in WeBS sectors.

Fig 4. Habitat area (km²) in WeBS sectors in different years (from 1975, darkest colour, to 2011, lighter colour).

Disturbance (explanatory continuous variable):
An index of the frequency of potentially disturbing activities in the sectors was calculated. Scores were given in Cruickshanks et al. (2010)\(^3\) to shore-based, water-based and airborne activities (overall) in each sector, ranging from 1 (Rare) to 5 (Very frequent), with also 0 values possibly allocated (Unknown occurrence/does not occur). The index of (potential) disturbance was calculated by averaging the score values in each sector. Spatial variability was only considered (same value allocated to each sector in different years) (Fig 5). No data were available for sector NA1.

![Disturbance proxy](image)

**Fig 5. Average disturbance index in the WeBS sectors, Humber Estuary.**

**Intertidal hard substrata** (explanatory continuous variable):
The occurrence of hard substrata (either pebbles or man-made vertical substratum) in the intertidal zone within each sector has been measured as % of sector length covered by such substrata. Measurements were made based on the visual inspection of aerial maps obtained from Bing, Google Maps and Google Earth. Spatial variability was only considered (same value allocated to each sector in different years). Hard substrata were recognized only in sectors ND (70% hard-pebbly), NE (5% hard-pebbly, 60% hard-man made) and NF (9% hard-pebbly, 50% hard-man made).

**Benthic abundance** (explanatory continuous variable):
Average abundance of benthic invertebrates in the intertidal habitat within each sector was calculated based on the data reported in Allen (2006)\(^4\), as an indicator of the amount of potential food resources available in the intertidal area. Total benthic abundance (indiv/0.0079 m\(^2\)) was given in different mid shore stations distributed across the sectors annually, from 1989 to 2003 (=13 to 15 observations per sector, as some data are missing for certain sectors in certain years). Both spatial and temporal variability was taken into account (Fig 6). No data were available for sectors NA1 and NA2.


EUNIS 3 intertidal habitat type (explanatory categorical variable):
The dominant intertidal habitat type within each sector was identified based on the EUNIS 3 habitat map given in Hemingway et al. (2008). EUNIS 3 habitats included Littoral mud (LMu), Littoral sand (LSa), Littoral mixed sediments (LMx), Low energy infralittoral rock (LLR) and dominant habitats were identified with a coverage >15% in each sector (Fig 7). Spatial variability was only considered (same value allocated to each sector in different years) and no data were available for sectors NA1 and NA2.

Benthic community type (explanatory categorical variable):
The intertidal benthic invertebrate community type within each sector in different years (between 1989 and 2003) was identified based on the cluster analysis given in Allen (2006), as an indicator of the type of food resources available in the intertidal area. The analysis was carried out on average density data at different stations located at mid and low shore in the northern bank of the estuary, and only stations at mid shore (giving the widest spatial and temporal coverage) were taken into account. Eight main community types (a to h) were

---

distinguished (with a maximum similarity <40%) roughly corresponding to a gradient from inner to outer estuary and to an increase in species richness and abundance (Fig 8). Both spatial and temporal variability was taken into account and no data were available for sectors NA1 and NA2.

![Characteristics species within intertidal north bank site groups derived from cluster analysis (from Allen 2006). Main community types (considering only mid shore samples) are indicated by red boxes.](image)

Bird population trend (explanatory continuous variable):
For selected species (based on the results of multivariate analysis), data on annual total maximum counts for GB were collected from WeBS books\(^6\). Count data were standardised by number of sites counted in the months when the maxima were recorded (Fig 9).

![GB population trend (max annual count/site) of selected wader species (A) and wildfowl species (B) (data from WeBS books).](image)

\(^6\) Waterbirds in the UK Series – The Wetland Bird Survey. Published by BTO, RSPB and JNCC in association with WWT.
2.3 Data exploration and treatment

Once the data have been collated in a single dataset, with all the information (response and explanatory variables) allocated to each observation in it, and before applying any statistical analysis or modelling, an exploration of the dataset should be carried out to identify the main sources of variability in it, its main limitations as well as the best transformation to apply to the data (this might be needed in some cases to fulfil the assumptions on data distribution of statistical analyses). Different data treatment might be required for different analyses. It has been highlighted by Zuur et al. (2009) that one should expect to dedicate up to 50% of the time spent on analysis to data exploration.

Example from case study (wader analysis in the Humber Estuary)
Species were allocated to functional groups (guilds) in order to highlight general patterns in the functioning of wader community (Table 1).

For the purpose of multivariate explorative analysis of bird assemblages distribution in the estuary, bird species density was averaged over 5-year periods per sector (8 periods: 1=1975-1979, 2=1980-1984, 3=1985-1989, 4=1990-1995, etc8) in order to reduce inter-annual fluctuations while still taking into account the general temporal variability. The resulting dataset for waders included 49 observations (over 5 periods only, from period 4 to 8).

Table 1. Guild classification of wader species.

<table>
<thead>
<tr>
<th>Guild</th>
<th>Description</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waders</td>
<td>F specialist</td>
<td>Feeding specialist (usually preys on larger/specific prey)</td>
</tr>
<tr>
<td></td>
<td>Mud</td>
<td>Species showing a loose association with mud</td>
</tr>
<tr>
<td></td>
<td>Mud F</td>
<td>Species predominantly feeding on mudflat (generalist mud feeders)</td>
</tr>
<tr>
<td></td>
<td>Mud R</td>
<td>Species predominantly roosting on mudflat</td>
</tr>
</tbody>
</table>

(∗) Generalist feeder on mud but likes Corophium, between generalist and specialist feeding
(∗∗) Feed on hard substratum cobbles and weed on estuaries

Similarly, explanatory variables were allocated to the 5-year periods by averaging them (continuous variables) or by considering the most frequent category in the 5-year period within each sector (categorical variables). The coefficient of variation (CV, %) was calculated for each species over the different sectors in each period, as an index of the spatial distribution of the species (the lower the CV, the more homogeneous the distribution of the species across sectors). Average CV was compared between species by using Kruskal-Wallis test.

Multivariate analysis was applied to the data in order to explore the main patterns of variation in wader community over space and time. Bird data were forth root transformed and the Bray-Curtis similarity matrix was calculated.

---

2 Period 8 includes only 2010 and 2011.
Cluster analysis was carried out on bird data to highlight similarities in spatial-temporal distribution of different species (R mode analysis) and in the bird community of different sectors-by-period (Q mode analysis). Also ordination analysis (Principal Coordinate analysis, PCO) was carried out to highlight general patterns in species/guilds spatial-temporal distribution across sectors and periods. Analysis of similarity (ANOSIM) between guilds and between sectors and periods (2-way crossed analysis) was carried out to highlight significant patterns.

Based on the above analysis, combined with species distinction into guilds, indicator species were chosen for further detailed habitat modeling, based on their being representative of a particular use mode of the estuarine habitat, and would be generally abundant and frequent in the dataset (to allow further detailed analysis).

Further details on the data exploration specific for certain statistical techniques (e.g., habitat modelling) is provided in the sections below.

### 2.4 Community distribution models

A further step in the analysis is the assessment of the relationships between bird spatio-temporal distribution in the estuary and environmental predictors. These relationships can be investigated both at the community level (multivariate analysis) and at the species level (univariate analysis). This section deals specifically with the multivariate case.

Depending on the data availability, the extraction of a subset of the original dataset might be needed for the analysis (e.g., if predictor variables have been derived only for a subset of the observations available for the bird data), bearing in mind that this will have consequences in the limits of validity and applicability of the obtained results (e.g., if a subset of data is analysed only in one salinity zone, then the results would be valid for that zone in the estuary, but they might not be exportable to other zones, where environmental variables may show other ranges of variability and the association between biota and environment might change or be affected by other factors).

Several multivariate techniques can be used to establish relationships between bird distribution (considering assemblages as a whole) and influent environmental variables (e.g., BIOENV analysis, canonical correlation analysis, canonical analysis of principal coordinates, multivariate multiple linear regression analysis). The choice of the most appropriate technique would depend on the main aim of the analysis (e.g., correlative analysis or cause-effect analysis) and on the characteristics of the data available for analysis. Details on the suitability of different techniques should be explored.

---


based on the dataset characteristics, by using the relevant literature (e.g., references in footnotes).

Example from case study (waders analysis in the Humber Estuary)
The spatial/temporal distribution of bird assemblages was related to the environmental variables. Given that environmental variables are not available for all the observations in the dataset (i.e., sectors-by-period, n=49 for waders, but not all environmental data are available for sectors NA1 and NA2 and after 2004), the analysis was carried out on a reduced dataset. The resulting dataset analysed for waders in the Humber Estuary (bird species densities and all the environmental variables) included 27 observations and covered only sectors in the mesohaline and polyhaline zones, therefore preventing the generalization of the results to other areas in the estuary (e.g., oligohaline zone).

A multivariate multiple linear regression analysis was carried out on bird species (density) data and on continuous explanatory variables in order to identify the main factors affecting the overall bird assemblage spatial-temporal distribution. The multivariate multiple regression full model (including all explanatory variables) was investigated by using distance-based redundancy analysis (dbRDA). DISTLM routine was also applied to identify the best subset of variables explaining wader data variability (best reduced model, selected by backward selection method using AIC criterion). Correlation analysis (Spearman's) was also carried out to identify the main relationships between species densities and environmental variables.

The datasets for the Elbe cover all the salinity zones (from freshwater to polyhaline), including 68 observations for which both bird densities and habitat areas are available (between 1984 and 1998) and 92 observations for which both bird densities and water quality parameters are available (between 2004 and 2009). As the datasets on habitats and water quality parameters are not temporally overlapping, the analysis has been carried out separately for the two types of abiotic characteristics in this estuary. The dataset for the Weser for which both bird densities and habitat areas are available covers all the salinity zones (from freshwater to polyhaline), including 43 observations (between 1984 and 2003). In turn the dataset for which both bird densities and water quality parameters are available includes 92 observations (between 2004 and 2009) covering only the freshwater and oligohaline zone. As the datasets on habitats and water quality parameters are only marginally overlapping (between 1992 and 2003, with a total of 6 observations), the analysis has been carried out separately for the two types of abiotic characteristics also in this estuary.

2.5 Species distribution models

_The final step in the analysis is the assessment of the relationships between the spatio-temporal distribution of single bird species and environmental predictors in the estuary._ With this purpose, multiple regression models can be applied by which the annual species density in different spatial units (dependent variable) is related to a series of explanatory variables (either continuous and categorical) in
order to select the best set of variables (best model) explaining the species density distribution in the estuary.

Before the model calibration, colinearity between explanatory variables in the specific dataset that is being analysed should be investigated and highly correlated variables excluded from the analysis. Statistical models (e.g., generalized linear models GLMs, general additive models GAMs) could be applied to either the species density data or to its presence-absence, depending on the type of data, including fixed effects (explanatory variables, and also interaction terms between variables if deemed important and if the data allow for it) and the best model selected based on backward selection method and best AIC criterion (further details are provided in the example below).

**Example from case study (waders analysis in the Humber Estuary)**

Four wader species have been selected based on their representativity of different guilds, their distribution in the estuary (results above) and local relevance: Dunlin, Golden Plover, Redshank and Bar-Tailed Godwit.

Species density data by sector by year have been used as response variable. Only when the frequency of occurrence of the species in the dataset was <75% (mainly due to a more heterogeneous distribution of the species, with association to certain sectors and absence from others), the probability of presence was modelled as response variable (based on presence-absence data) by using a logistic regression.

Environmental variables taken into account as explanatory variables included salinity, habitats areas (intertidal, subtidal, marsh and supralittoral non-flooded zone), disturbance, EUNIS 3 intertidal habitat type, benthic community type and total benthic abundance. The distribution of intertidal hard substrata in the estuarine sectors was considered not relevant for the selected wader species hence it was excluded *a priori* from the analysis. In order to account for the possible presence of a temporal dependence structure in the bird data, Year was also included as covariate in the species model. The GB population density of each species over different years was also included to account for the influence of wider population trends on the species temporal distribution in the Humber Estuary. No water quality data were available for modelling. Also, no interaction terms were considered between the predictors, in order to allow a simpler interpretation of model results.

Given the reduced availability of data for the different covariates (explanatory variables) compared to the availability of count data for the selected species, the inclusion of environmental variables in the analysis led to a reduction of the size of the dataset which could be analysed (Table 2). Two main reduced dataset were identified for the analysis: R2, the most reduced one, including all the 11 covariates indicated above, and R1, including all covariates except of benthic community type and abundance. Benthic data were not available for years before 1989 and after 2003, hence leading to a further decrease in the dataset size (Table 2) with possible effect on the power of the analysis applied to the

---

data. The analysis was applied first to the most reduced dataset (R2) to explore the effect of all the environmental variables on the species distribution. However, whenever the results highlighted that benthic community variables (community type and total abundance) were not relevant in affecting the species distribution, the analysis was applied also to the larger dataset (R1) in order to confirm the patterns and relationships previously obtained.

Table 2. Number of observations in the Full (species density data only), R2 (species density data + all 11 covariates) and R1 (species density data + all covariates except for benthic community type and abundance) datasets for the different wader species in the Humber Estuary.

<table>
<thead>
<tr>
<th>Species (code)</th>
<th>Datasets</th>
<th>Full</th>
<th>R2</th>
<th>R1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunlin (DN)</td>
<td></td>
<td>183</td>
<td>91</td>
<td>146</td>
</tr>
<tr>
<td>Golden Plover (GP)</td>
<td></td>
<td>183</td>
<td>90</td>
<td>148</td>
</tr>
<tr>
<td>Redshank (RK)</td>
<td></td>
<td>183</td>
<td>91</td>
<td>147</td>
</tr>
<tr>
<td>Bar-Tailed Godwit (BA)</td>
<td></td>
<td>183</td>
<td>92</td>
<td>148</td>
</tr>
</tbody>
</table>

When necessary, data were transformed (square root, forth root or log transformation, whichever the most appropriate) in order to remove the possible effect of outliers, normalise the data distributions and to increase homogeneity of variance.

The number of candidate explanatory variables (or predictors) included in the final model was firstly reduced by removing highly correlated variables. Following Fielding and Haworth (1995\textsuperscript{14}), a Spearman correlation analysis was conducted and variables with high correlation coefficient ($|r_S|>0.7$) were not considered for model calibration, in order to avoid multicolinearity. In addition, given that even a moderate collinearity might be problematic, particularly if the ecological signal is weak (Zuur et al. 2009), variables with $|r_S|>0.6$ were also taken into account and were not considered for model calibration whenever their relationship with the response variable was weak ($|r_S|<0.5$).

Relationships between the species mean distribution and environmental variables were studied by means of generalized additive models (GAM) (Zuur et al. 2007). GAMs allow to model some predictors non-parametrically in addition to linear and polynomial terms (Guisan et al. 2002\textsuperscript{15}), allowing the decision of the response shapes to be fully determined by data. This is achieved by introducing a smoothing function for the continuous predictors. GAMs were fitted by using the ‘mgcv’ library (Wood 2000\textsuperscript{16}) for R software packages (R Development Core Team 2008). This type of model is represented in ‘mgcv’ as penalized GLM, each smooth term of a GAM being represented using an appropriate set of basic functions (Wood and Augustin 2002\textsuperscript{17}). The GAM model-building procedures followed the guidelines of Wood (2000), using penalized regression splines. This allows the degrees of freedom for each smooth term in the model to be chosen

simultaneously as part of model fitting by minimizing the Generalized Cross Validation (GCV) score of the whole model (Wood, 2006). Density data were fitted using a Gaussian family with the canonical identity link, whereas presence–absence data were fitted using a binomial family with the canonical logit link, optimizing the GCV score. Model selection was carried out by means of backward selection using AIC as selection criterion. The resulting best model was validated graphically by examination of possible patterns in the residuals, in order to check that assumptions (homogeneity, independence, normality) were fulfilled (see Zuur et al. 2007 for a detailed guidance on the most appropriate statistical procedures). Single predictor models were also considered and their AIC value was used to rank the importance of each environmental variable in affecting the species distribution.

An example of the detailed results of this modeling procedure applied to one of the wader species (namely, Dunlin) is provided below for better clarity. It is of note that these detailed results were not presented in full in the final report “Determinants of bird habitat use in TIDE estuaries” as the relevant final results were summarised in there in order to make that report easier to read.

The reduced dataset including Dunlin density data and all 11 environmental covariates was analysed (R2, n=91). Dunlin density was forth root transformed to remove the effect of outliers and to reduce heterogeneity and increase normality in the data distribution. A similar transformation was applied to the benthic abundance.

The correlation analysis highlighted a strong correlation (|r|>0.7) between salinity and most of the habitat areas hence it was excluded from the analysis (Fig 10). Similarly, a strong negative correlation was observed between year and the species GB population count, the latter covariate being excluded from the analysis. Also a moderate correlation (|r|>0.6) was detected between intertidal and marsh area (positive correlation) and between supralittoral and subtidal area (negative correlation). Given their weak relationship with the response variable, these collinearities were also removed by excluding marsh and subtidal area from model calibration.

As a result, the full model included 7 explanatory variables (the relationship with the excluded variables being indicated in parenthesis):

1. Year, Y (-DN,GB);
2. Intertidal area, Int (+Sal, +Mar);
3. Supralittoral area, Sup (-Sal, -Sub);
4. Disturbance;
5. Benthic abundance, BAb;
6. Intertidal habitat type (Eunis 3), Eun;
7. Benthic community type, BType.

The selected model was the one ranked as the best model, after backward selection, following the decrease in the AIC and explains 86.4% of the total deviance (Table 3). It includes 4 out 7 variables among those retained in the analysis:

Benthic community characteristics (type and abundance) as well as year were not included as predictors of the density distribution of Dunlin.

Aside from the shape of the relationship between Dunlin and a given predictor, it is possible to use single predictor GAMs to rank the importance of the single predictors (Table 3). The predictor that can best explain Dunlin density is intertidal area, with the highest deviance (70%) and the lowest AIC (184.13). Following predictors are supralittoral area and intertidal habitat type, whereas disturbance is the weakest predictor of Dunlin distribution among those included in the selected model.

All the covariates included in the model showed a significant effect on Dunlin density. The shape of these significant effects of each predictor variable on the model response is reported in Fig 11. The smoother associated to supralittoral area has 3.9 estimated degrees of freedom, indicating a higher non-linearity in the smoother associated to this variable, whereas the smoother associated to intertidal area is almost linear (edf=1). A linear positive relationship was also fitted with disturbance.

According to the selected model, the mean density of Dunlin is expected to be maximum with high intertidal area, with a supralittoral area between 0.2 and 0.3 km2, in the mixed habitat type LMu/LMx/LSa (Littoral mud/ Littoral sand/Littoral mixed sediments) and with high disturbance (Fig 11).

As, according to the above analysis, benthic community variables (community type and total abundance) were not relevant in affecting Dunlin density, the analysis was applied to the wider dataset (R1, n=146) following a similar process as applied above (detailed results not presented here) and the final results were summarised in the report “Determinants of bird habitat use in TIDE estuaries”.
Fig 10. Dunlin (R2, n=91): relationships between the variables in the dataset (Sqrt2DN = Dunlin density (forth-root transformed); Y = year; Sal = salinity; Int = intertidal area; Sub = subtidal area; Mar = marsh area; Sup = supralittoral area (no-flooding zone); Dis = disturbance; Sqrt2BAb = benthic abundance (forth-root transformed); DN.GB = Dunlin GB population counts; Eunis = habitat type; BType = benthic community type). Lower panel = scatterplots; diagonal = histograms; upper panel = Spearman’s correlation coefficient.

Table 3. Dunlin (R2, n=91): GAM fits and AIC for single predictor models, full model and best selected model. Variable abbreviations are as in Fig. 10.

<table>
<thead>
<tr>
<th>Model</th>
<th>% deviance</th>
<th>AIC</th>
<th>rank (AIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y + Int + Sup + Dis + Sqrt2BAb + BType + Eun</td>
<td>87.4</td>
<td>123.22</td>
<td>2</td>
</tr>
<tr>
<td><strong>Best model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int + Sup + Dis + Eun</td>
<td>86.4</td>
<td>115.68</td>
<td>1</td>
</tr>
<tr>
<td><strong>Single predictors models:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int</td>
<td>70.4</td>
<td>184.13</td>
<td>3</td>
</tr>
<tr>
<td>Sup</td>
<td>62.4</td>
<td>203.16</td>
<td>4</td>
</tr>
<tr>
<td>Dis</td>
<td>38.8</td>
<td>234.57</td>
<td>6</td>
</tr>
<tr>
<td>Eun</td>
<td>52.7</td>
<td>217.12</td>
<td>5</td>
</tr>
</tbody>
</table>
Fig 11. Dunlin (R2, n=91): Effect of each explanatory variable on the density of Dunlin (forth-root transformed), measured as contribution on the linear term of the model. The fitted values are adjusted to average zero and the dotted bands indicate 95% pointwise confidence intervals. Tick marks along the x-axis show the location of observations along the variable range. Significance level (p-value) of each term in the model is given in brackets.