

Aberystwyth University

*Lasso penalisation identifies consistent trends over time in landscape and climate factors influencing the wintering distribution of the Eurasian Curlew (*Numenius arquata*)*

Kenobi, Kim; Read, Warren; Bowgen, Katharine M.; Macgregor, Callum J.; Taylor, Rachel C.; Cámara García, Walther C.A.; Hodges, Crona; Dennis, Peter; Holloway, Paul

Published in:
Ecological Informatics

DOI:
[10.1016/j.ecoinf.2023.102244](https://doi.org/10.1016/j.ecoinf.2023.102244)

Publication date:
2023

Citation for published version (APA):
Kenobi, K., Read, W., Bowgen, K. M., Macgregor, C. J., Taylor, R. C., Cámara García, W. C. A., Hodges, C., Dennis, P., & Holloway, P. (2023). Lasso penalisation identifies consistent trends over time in landscape and climate factors influencing the wintering distribution of the Eurasian Curlew (*Numenius arquata*). *Ecological Informatics*, 77, [102244]. <https://doi.org/10.1016/j.ecoinf.2023.102244>

Document License CC BY

General rights

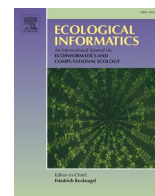
Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk



Lasso penalisation identifies consistent trends over time in landscape and climate factors influencing the wintering distribution of the Eurasian Curlew (*Numenius arquata*)

Kim Kenobi ^a, Warren Read ^b, Katharine M. Bowgen ^c, Callum J. Macgregor ^c, Rachel C. Taylor ^c, Walther C.A. Cámaro García ^{d,e}, Crona Hodges ^f, Peter Dennis ^b, Paul Holloway ^{d,g,*}

^a Department of Mathematics, Aberystwyth University, UK

^b Faculty of Earth and Life Sciences, Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, UK

^c British Trust for Ornithology, BTO, The Nunnery, Thetford, UK

^d Department of Geography, University College Cork, Cork, Ireland

^e MaREL, the SFI Research Centre for Energy, Climate and Marine, Ringaskiddy, Ireland

^f Geo Smart Decisions, Abergynolwyn Tywyn, Gwynedd, UK

^g Environmental Research Institute, University College Cork, Cork, Ireland

ARTICLE INFO

Keywords:

Migratory
Species distribution modelling
Curlew
Winter
Habitat change
Habitat

ABSTRACT

Migratory birds are particularly susceptible to climate change and habitat loss due to their reliance on a global network of ecosystems, with waders and seabirds undergoing significant population decline. The Eurasian Curlew (*Numenius arquata*) is classified as Near Threatened on the IUCN Red List of Threatened Species, but breeding populations in Great Britain and the island of Ireland have declined drastically, with the species on the brink of local extinction. We present a set of models of the distribution of Curlew sightings between November and February in the Great Britain and the island of Ireland over the period 2003 to 2019. Using a model selection process (cross-validated lasso regression), we reduce the fairly large set of CORINE satellite land cover classes to a much smaller set of explanatory variables which we combine with environmental variables and fit binomial Generalized Linear Models to Curlew observation records. This enables us to build up a detailed picture of where and when Curlew are sighted between November and February over the 17 years of the study. Reproducibly, from November to January between 2003 and 2019, the coastal land cover classes, Estuaries, Intertidal Flats, Salt Marshes and Port Areas, feature prominently in the sets of explanatory variables selected by the lasso regression. Moreover, this study represents the first regional scale analysis on the impact of landscape and climate features on wintering curlew distribution, identifying the importance of landscape factors that warrant further research, such as the importance of artificial structures and the importance of February within the migration of the Curlew.

1. Introduction

Coastal ecosystems are being increasingly modified, with an estimated one billion people expected to live within the lower-elevation coastal zone by 2060 (Neumann et al., 2015). Consequently, coastal habitats and the species they support are increasingly under threat from climate change, habitat loss, and urbanization (Holloway and Field, 2020; Russo and Cirella, 2021). The recent Intergovernmental Science Policy Platform on Biodiversity and Ecosystem Services (IPBES) report (Dias et al., 2019) identified that over 1 million species are now threatened with extinction, many within decades. Climate change is

known to exacerbate direct threats from human activities (Munilla et al., 2007), particularly in coastal ecosystems where the combination of sea-level rise and hard-engineering structures causes coastal squeeze, which restricts the geographic distribution of habitat specialists causing local extinctions and/or trophic cascades (Hughes, 2004). Consequently, with climate change and concomitant urbanization leading to species range shifts, range constrictions, and an increasing likelihood of local and global extinction, it is imperative for biodiversity conservation to better understand the drivers of current and past patterns of species distributions to inform mitigation strategies.

Migratory birds are particularly susceptible due to their reliance on a

* Corresponding author at: Department of Geography, University College Cork, Cork, Ireland.
E-mail address: paul.holloway@ucc.ie (P. Holloway).

<https://doi.org/10.1016/j.ecolinf.2023.102244>

Received 3 April 2023; Received in revised form 29 July 2023; Accepted 30 July 2023

Available online 12 August 2023

1574-9541/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

global network of ecosystems (Van Doren et al., 2021), with waders and seabirds undergoing significant population decline (Dias et al., 2019; Manton and Angelstam, 2021). Long-distance migrants are of particular concern globally, yet 91% are estimated to be inadequately protected across their annual cycle (Runge et al., 2014). The Eurasian Curlew (*Numenius arquata*), hereafter Curlew, is classified as Near Threatened on the IUCN Red List of Threatened Species (BirdLife International, 2022), with local populations declining across the world, which is having a consequent adverse effect on the global population of Curlew (Brown et al., 2015).

The UK population is estimated to account for between 19 and 27% of the global breeding population of Curlew (Brown et al., 2015), yet recent population estimates indicate population declines of 60% in Scotland, 73% in Wales and 30% in England, between 1995 and 2021 (Heywood et al., 2023), while in Northern Ireland populations have been estimated to have declined by 80% over the same period (Jones et al., 2022). In the Republic of Ireland, in the late 1980s there were estimated to be approximately 3300–5500 pairs of breeding Curlew; yet in 2019 there remained no more than 150 pairs, which equates to a 98% decline over the last 30 years (O'Donoghue et al., 2019; Servignat and O'Donoghue, 2022).

Human development and interactions within the coastal zone will impinge on many wetlands of international importance. Coastal habitats during the wintering months are of vital importance to the global network of Curlew as they can support large waterbird populations during adverse conditions. Despite this importance, these habitats have been declining across Europe since the 1980s (PECBMS, 2020), and in the UK and Ireland, since 2000 (Frost et al., 2019; Kennedy et al., 2022). Climate may affect such wintering coastal bird populations indirectly via changes in food supply (Carroll et al., 2015) and sea level changes. Coastal squeeze as a result of increased sea level and shoreline hardening may reduce the amount of coastal habitat available, especially at periods of high tide. Other direct impacts include extreme weather, including flooding events that increase rates of mortality (McDonald et al., 2015; Oro, 2014; Sydeman et al., 2015). Such effects will be compounded by other drivers including land use along the coast, as well as in the marine environment (fisheries), pollution, marine infrastructure projects and disease (Burthe et al., 2014; Oro, 2014).

While the main demographic driver of Curlew decline has long been noted as their low rate of breeding success (Cook et al., 2021; Grant et al., 1999), Curlew mortality is higher in winter months, particularly in juveniles, linked to a range of aforementioned factors (Taylor and Dodd, 2013). The Curlew has therefore been the focus of a significant amount of research in recent years to support conservation efforts (Young et al., 2020), with a lot of research centred on the breeding distribution (Berg, 1992; Colhoun et al., 2015; Henderson et al., 2002; Johnstone et al., 2017; Ławicki and Wylegała, 2011; O'Brien et al., 2002; Valkama et al., 1998). Fewer studies have subsequently explored the drivers of the winter distribution (Burton et al., 2002a); however, with such significant global efforts to support the breeding success of Curlew, there remains a need to fully understand the environmental drivers across their full annual distribution, and not just their breeding season. If successfully fledged chicks do not return after their winter migration, then measures taken to improve their breeding success are futile.

Habitat degradation and loss is well cited as one of the primary drivers of Curlew decline (Young et al., 2020). For example, Johnstone et al. (2017) used a combination of statistical models to identify breeding abundance over a 15-year period, related to land cover type, strategy, and vegetation, identifying that while breeding was declining over time, it was highest when configurations were a mixture of moorland and improved farmland. The consensus so far is that for breeding distributions Curlew tend to avoid forested areas (Berg, 1992; Brown et al., 2015; Johnstone et al., 2017) with a preference for bog and grasslands (Henderson et al., 2002; Ławicki and Wylegała, 2011), and an avoidance of human-dominated landscapes including roads (Burton et al., 2002b) and construction (Burton et al., 2002a). However, most (if

not all) of the aforementioned studies focus on a relatively local scale (e.g., farmland, estuary) over a specific time-period (e.g., 10 years) to explore the impact of land cover type on abundance and/or density of birds (Berg, 1992; Burton et al., 2002a, 2002b; Duan et al., 2020; O'Brien et al., 2002; Valkama et al., 1998; Yang et al., 2011) with land cover data collected as part of the study, which while able to provide fine scale information as to the importance of land cover has implications for scaling up the research to a regional level.

Subsequently, the use of remotely sensed imagery, and particularly the Copernicus derived CORINE land cover has been widely employed to support understanding of bird distributions (Portaccio et al., 2021; Radović et al., 2011). An advantage of CORINE is the fine spatial resolution of the data set (minimum mapping unit of 25 ha for areal features and 100 m for linear features), despite challenges associated with the broad thematic categorisation (Koma et al., 2022; Lambert et al., 2021). However, despite its wide implementation in ecology, few studies have linked CORINE land cover categories to regional/continental distributions of Curlew. This is particularly surprising given the variety of habitats Curlew utilise over the various stages of their life cycle, and the importance of understanding species-landscape relationships at a national level given their substantial decline.

One challenge with CORINE land cover is the number of categories, and the need to reduce the number of variables in the model selection phase. There are a plethora of approaches in existence, including entropy (e.g. Estrada and Real, 2021), principal components analysis (e.g. Cruz-Cárdenas et al., 2014), expert opinion, and Lasso penalisation (e.g. Iyer et al., 2019). The latter has been the least used approach within species distribution modelling, offering an opportunity for further testing.

Several drivers of Curlew distributions have been invoked, yet to date these have not resolved the debate surrounding favourable land cover, particularly for winter distributions at a national extent. Moreover, given the conservation status associated with Curlew globally and within the UK and Ireland, determining the drivers of their distribution will have important implications for developing effective and long-lasting conservation and management strategies. Here we focus on the important land cover and climatic drivers of Curlew distribution in the UK and Ireland, considering the wintering seasons (defined as November to February), over a 17-year period. In addressing this we explore to what extent the land cover and climate variables contribute consistently across multiple temporal windows, considering months, years, and CORINE windows. Finally, we evaluate the efficacy of Lasso penalisation as a model selection technique within species distribution modelling of Curlew.

2. Materials and methods

2.1. Curlew observation data

Data for Curlew observations across Britain and Ireland were obtained from multiple sources: these included both public-domain and request-only data. In the former category, data came from collated submissions to the United Kingdom's National Biodiversity Network (NBN) and from the Birds of Ireland data set held by Ireland's National Biodiversity Data Centre (NBDC). Request-only Curlew data came from submissions to the eBird app published by Cornell Lab of Ornithology, from submissions to BirdTrack and Wetland Bird Survey (WeBS), made available by the British Trust for Ornithology (BTO), and from submissions to Ireland Wetland Bird Survey (I-WeBS), made available by BirdWatch Ireland (BWI) (see Supplementary Table S1).

We sought to collect and aggregate as many positive sightings as possible, in order to achieve maximal coverage in assigning binary values to each 1 km square in the British (Ordnance Survey) and Irish (Ordnance Survey Ireland) grids; data from Northern Ireland are resolved to the latter. Note, the binary values here represent at least one Curlew being sighted in a particular grid square in a given month (coded

1) and no Curlew being sighted in a particular grid square in a given month (coded 0). Where multiple sightings are reported within an individual grid, values are resolved to a single value (1 or 0) in our response dataset, with no duplication subsequently incorporated in the models. We cannot distinguish in the data between the situation where Curlew are genuinely absent and the situation where they are present but are not observed. The latter situation could arise either because Curlew are a cryptic species and may not be seen by birdwatchers, or because particular grid squares are simply not visited by birdwatchers in certain months.

We reprojected all data to give us binary data as explained above based on 1 km squares in the Corine equal-area grid. For data originating from the WeBS and I-WeBS surveys, the survey areas themselves are restricted to defined wetland areas only, so will take no account of Curlews occupying 1 km squares which do not intersect these areas, which themselves are typically irregular polygons. Data from other BTO surveys (constituting ~90% of observations contributed within the UK), and from BirdWatch Ireland, are already resolved to 1 km centroids in the relevant grids, so no finer level of detail is available for the bulk of Curlew observations. Records from eBird come in with precise latitude and longitude co-ordinates, but this still comes with a level of uncertainty around the distance and direction of the observed bird(s) from the viewer; co-ordinates were mapped to the appropriate 1 km centroid on the Corine grid.

Summaries by month and source across the years in each CORINE window are given in Supplementary tables S1 to S3.

2.2. Explanatory variables

In this study, we used a combination of land cover class and climate/digital terrain model (DTM) variables over a 1 km × 1 km grid across the whole of Britain and Ireland. Full land cover classification descriptions are available at <https://land.copernicus.eu/>, while descriptions of the 23 reported as important through our models are provided in Table A1.

The land cover class variables were calculated using CORINE data for 2006, 2012 and 2018. CORINE land cover products are the only contemporary land cover product that has a historical component available in Ireland (and until quite recently the land cover map with the finest spatial resolution). Therefore this product was used to allow consistency across the international borders. Since the source raster provided by CORINE is at a 100 m × 100 m (1 ha) resolution, we had 100 of these centroids in each of the 1 km × 1 km squares. For each of the 37 land cover classes present within Britain and Ireland, the number of 100 m × 100 m centroids within each 1 km × 1 km square classified as that land cover class was stored as a discrete raster layer (at 1 km × 1 km resolution). Each such layer therefore gives the approximate integer percentage of every 1 km × 1 km square covered by its (the layer's) respective class of land cover. We used the native CORINE equal-area projection for these data.

We used a set of 5 climate variables, covering the period between January 2003 and February 2019, based on precipitation and temperature: monthly dry days with precipitation lower than 0.05 mm; monthly dry days with precipitation lower than 0.5 mm; monthly frost days (minimum temperature less than 0 °C); monthly days with a minimum temperature lower than 5 °C; and monthly total precipitation. Our two definitions of 'drought' were considered to account for administrative and geographic differences across international boundaries, as well as to incorporate the ecological requirements of the Curlew. Among those requirements is the accessibility of prey species: for example, the depth of invertebrates in the soil layer depends on soil penetrability, which depends in turn on precipitation (Barnett and Facey, 2016; Davis et al., 2006). A shortage of precipitation (< 0.5 mm) might therefore represent a qualitatively different change in habitat from a more severe reduction in precipitation (< 0.05 mm). As such, both precipitation thresholds were incorporated alongside total precipitation in the analysis to investigate any differences across the wintering distribution, at a

monthly resolution. Regarding the two temperature thresholds, alongside sub-zero days, we used days below 5 °C as a separate explanatory variable specifically in order to account for the significant increase in energetic costs for foraging Curlew when exposed to temperatures below 5 °C—the Lower Critical Temperature (values for Curlew from Bowgen, 2017). All 5 variables were derived from the Met Office Gridded land surface climate observations on the British side and from gridded climate observations and reanalysis data provided by the Met Eireann-research TRANSLATE Project team on the Irish side (O'Brien and Nolan, 2023). In line with the land cover classes, the British and Irish data sets were merged and projected over the 1 km × 1 km grid using the native CORINE equal-area projection.

The DTM and slope percent variables we used, again at a 1 km × 1 km resolution, were derived from Ordnance Survey (OS data ©Crown copyright and database right 2022).

2.3. Lasso penalisation

Since there are 37 land cover classes for Britain and Ireland in the CORINE data set, it is desirable to fit models using only a subset of these variables. In order to find a reasoned way to select a relevant subset of the explanatory variables in our logistic generalized linear models, we use a technique called lasso (least absolute shrinkage and selection operator) regression. Lasso regression was first introduced in geophysics (Santosa and Symes, 1986) and later by Robert Tibshirani (Tibshirani, 1996) who coined the term.

The basic idea behind lasso regression is to apply a penalization term to the L1-norm (i.e. the sum of absolute values) of the set of coefficients in the regression model. It can be used in the context of linear modelling, but also in the context of generalized linear modelling. For illustrative purposes, consider a linear model with q explanatory variables with the form

$$Y_i = \beta_0 + \sum_{j=1}^q \beta_j X_{ji}$$

For a given value of the penalization term, λ , lasso regression identifies the values of the coefficients in the above model, subject to the constraint that

$$\sum_{j=1}^q |\beta_j| < t,$$

where the exact relationship between t and λ is data-dependent. (Note that the intercept term, β_0 , is not included in the constraint.) Explicitly, we wish to find

$$\min_{\beta_0, \beta} \left\{ \sum_{i=1}^N (Y_i - \beta_0 - X_i^T \beta)^2 \right\} \text{ subject to } \sum_{j=1}^q |\beta_j| \leq t.$$

In the case of generalized linear models, negative log-likelihood contributions take the place of the squares of residuals in the above formula. Including weights, w_1, w_2, \dots, w_N , for a given value of the penalty parameter, λ , lasso regression calculates

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda \|\beta\|_1,$$

where $l(y_i, \beta_0 + \beta^T x_i)$ is the negative log-likelihood contribution for observation i and $\|\beta\|_1 = \sum_{j=1}^q |\beta_j|$, i.e. the sum of absolute values of the regression coefficients (not including the intercept), or the so-called L1 norm.

There is an R package called glmnet (Friedman et al., 2010) that enables the user to fit lasso penalized generalized linear models for a grid of values of the penalization parameter, λ . There is an online introduction to glmnet (Hastie et al., 2021) in which the details of the algorithm and appropriate plots can be found.

2.3.1. Applying lasso to the Curlew data

Our observation data consist of binary data for Curlew sightings in 1 km \times 1 km squares across the whole of the British Isles for January, February, November and December for each year between 2003 and 2018 as well as January and February for 2019 (a total of 66 months). Again, we emphasise that a 1 means that at least one Curlew was sighted in that grid square in that month, and a 0 means that no Curlew was sighted, either because there were none, or because they were there but not observed. The explanatory variables we use to fit our generalized linear models are of three types. Firstly, for the land cover classes, we use counts (out of a total possible of 100) of the different land cover classes in 100 m \times 100 m grid squares within each 1 km \times 1 km square in data we have obtained from CORINE (European Environment Agency (EEA), 2006, European Union, 2012, 2018). For a given year of the CORINE data, we consider it to be applicable to sightings data from three years prior to the CORINE date through to two years after. Hence we use CORINE 2006 data as explanatory variables for Curlew sightings data obtained between the start of 2003 and the end of 2008, we use CORINE 2012 for sightings between 2009 and 2014, and CORINE 2018 for sightings from 2015 onwards. To facilitate communication, we label these intervals ‘CORINE windows’.

The second type of explanatory variable is climate. We have monthly data (again at a 1 km \times 1 km resolution) for the following five climate variables for each of the 200 months in the study: dry days less than 0.05 mm rain; dry days less than 0.5 mm rain, total precipitation, number of frost days (minimum temperature less than 0 °C), and number of days with minimum temperature less than 5 °C.

Finally, we also include two explanatory variables that are fixed across the whole of the time period of the study: a digital terrain model (DTM) and a variable, slope percent, that gives an indication of the average slope associated with each raster cell.

The algorithm we have used to generate individual GLMs for each of the 66 months consists of two steps. The first is a filtering stage that enables us to decide, for each month and CORINE year combination (i.e. 12 = 3 \times 4 combinations), which of the land cover classes should be included in the individual monthly models within the relevant CORINE window. (Note that we only apply the L1-penalisation in the lasso function to the land cover class explanatory variables, and keep the remaining seven explanatory variables, i.e. the five climate variables and the two DTM variables, unconstrained when we use the lasso filtering. These latter seven variables are averaged across the years in a particular CORINE window for each of the months.)

For the filtering, we assume that due to annual patterns in the behaviour of Curlew, within a particular CORINE window, the probability of seeing a Curlew in a given grid square in a given month is fixed. Suppose for a particular month there are N_m years’ worth of data for that month in a given CORINE window. Then, if we write p_{ij} for the (fixed) probability for grid square (i, j) for the same month (where for ease of notation the dependence on month has been dropped), then the number of times that we see Curlew in that square in that month across the CORINE window has a binomial distribution with parameters N_m and p_{ij} . We can therefore model the probability of seeing a Curlew in a particular grid square in a particular month for a given CORINE window using a binomial family generalized linear model with a logit link function. We end up with 12 models (Four months for each of three CORINE data sets.)

Using the R package `glmnet`, the lasso model (only penalising the land cover variables) is fitted for a grid of λ values. In order to select a value of this penalization parameter, we make use of 10-fold cross-validation, using the `cv.glmnet` function in the `glmnet` package. This works by splitting the data set (randomly) into 10 approximately equal subsets. Each of these subsets is independently used as a test data set, with the remaining 90% of the data constituting the training data. For each value of the parameter λ , this cross-validation process enables us to obtain a point estimate with confidence intervals for the deviance of the fitted GLMs. We have elected to use the `lambda.1se` option within the `cv.glmnet` function, corresponding to the most regularized (or penalized) model such that the cross-validated error is within one standard error of the minimum. Since cross-validation is a stochastic process, we ran it using 31 different pseudo-random numbers as starting points (random seeds) for the computation. To give the reader an insight into how the lasso regression works in the context of a binomial GLM for the Curlew data, consider Fig. 1.

In the left-hand figure of Fig. 1 we see the fitted models (in terms of the land cover coefficients) at a range of values of the regularization parameter, λ . As we move from left to right, the models are progressively less penalized, since the L1 norm is plotted on the x – axis, and this is inversely related to the λ parameter. Each of the coloured lines corresponds to a particular explanatory variable. To read off the coefficients at a particular value of λ (or equivalently the L1 norm), imagine a vertical line at a particular x –value. Where this vertical line crosses the coloured line for a particular explanatory variable gives the value of the corresponding coefficient in the penalized binomial GLM.

In the right-hand figure of Fig. 1 we show the results of a cross-validation procedure. At each value of $\log\lambda$, we see the point estimate (in red) and a confidence interval for the deviance of the fitted GLMs obtained by 10-fold cross-validation. The dotted vertical line at the left of this figure shows the models at the value of the regularization parameter for which the GLM deviance is minimized. The other vertical line shows the `lambda.1se` value of $(\log)\lambda$, i.e. the maximum value of λ such that the cross-validated error is within one standard error of the minimum.

We apply this cross-validation step 12 times (January, February, November and December for 2006, 2012 and 2018) using the same set of 31 random seeds for each month-CORINE year combination. For each of these 334 models, we select only the land cover classes with non-zero coefficients (see second stage below). The signs of these coefficients do not change as the random seed varies, although for different land cover class variables, the proportion of the coefficients that are non-zero varies. We select the land cover class variables for which at least 20 of the 31 random number seeds (about 64%) resulted in a non-zero estimate of the coefficient. Note that $P(X \geq 20 | X \sim \text{Bin}(31, 0.5)) \approx 0.075$.

The second step in the algorithm generates individual monthly models using the lasso fits on the CORINE windows described above. Having selected the land cover variables that meet the threshold outlined above, we combine these with the five climate and two DTM variables and then fit an unconstrained logistic GLM to a training set generated from the observational data for that specific month. The training set is a randomly selected set of half of the raster cells with the constraint that it contain exactly half of the cells in which an observation occurred and exactly half of the cells in which no observation occurred. The remaining half of the raster cells are used as a test data set for the monthly model.

The remaining half of the raster cells are used as a test data set for the monthly model.

2.4. Model validation

In order to assess the validity of the fitted models, we make use of two metrics, the area under the receiver operator characteristic curve (AUC) (see, e.g. Swets, 1988) and a slightly modified form of the Boyce index, the original form of which is described in Hirzel et al., 2006.

The receiver operator characteristic curve (ROC) is a plot of true positive rate against false positive rate for all possible values between 0 and 1 of a threshold parameter. The AUC is a threshold-independent metric that gives a measure of the efficacy of a model as a classifier, and takes values between 0 (worst) and 1 (best). Values above 0.5 indicate a better-than-random classifier.

For the Boyce index, the idea is to consider the vector of predicted values from the model, to split this into a number of bins, and for each bin, to look at how the proportion of presences actually found in the cells with predicted values in that window compares with the expected proportion if the presences were randomly distributed across all predicted values. This ratio, the predicted-to-expected (P/E) ratio, is called F_i in

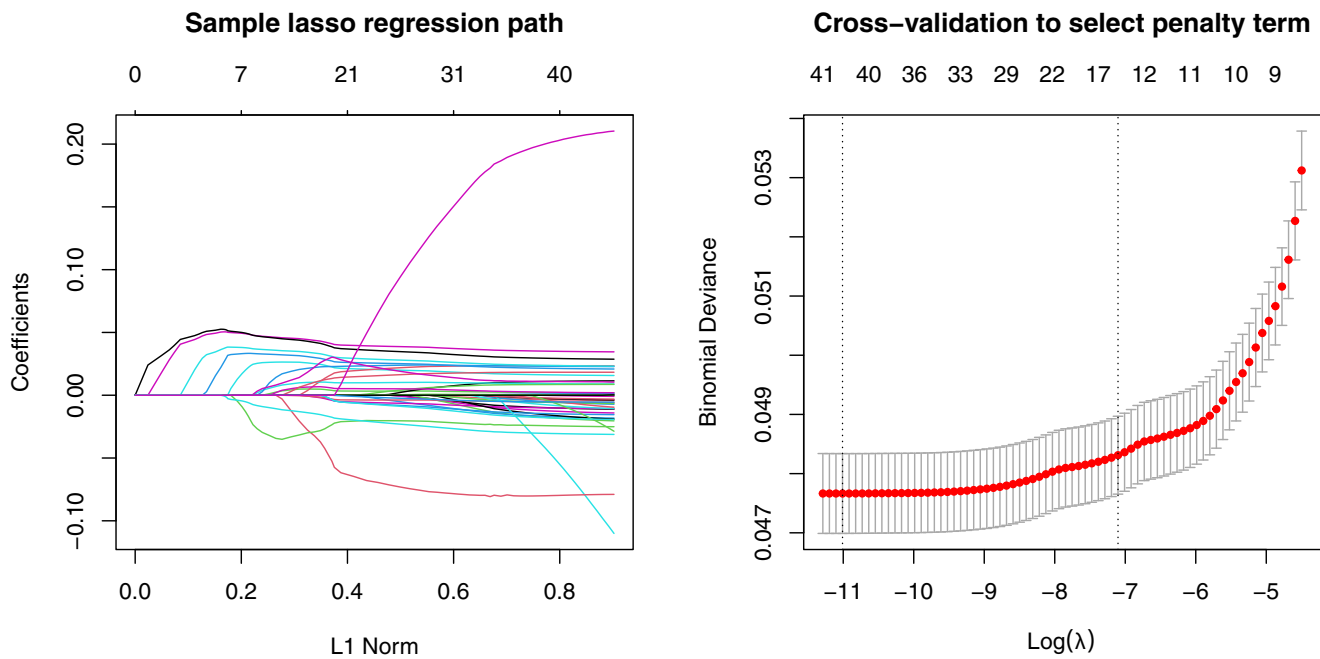


Fig. 1. An example of the coefficients in a lasso penalisation path and crossvalidated deviance values.

Hirzel et al., 2006. For a well-fitting model, one would anticipate that the ratio of predicted proportion to expected proportion would increase for higher predicted values. The Boyce index (Boyce et al., 2002; Hirzel et al., 2006) is simply the Spearman correlation between the mean/median bin value and F_i . It takes a value between -1 and $+1$, with a value tending toward $+1$ indicating good to perfect predictions, values around 0 indicating predictions no different from those obtained by chance, and values toward -1 indicating counter-predictions, i.e. observing presences in low suitability classes and observing absences in high suitability classes (Hirzel et al., 2006).

The modification of the Boyce index we use is simply to split the range of the fitted probabilities into bins, rather than the whole of the interval $[0,1]$.

3. Results

In Fig. 2 we can see the average signs (across the 31 random seeds) of the land cover class coefficients for each of the 12 combinations. Those coefficients whose average sign is positive are shown in green, negative in brown, while those whose average sign is zero are shown in white. Note that land cover classes for which the coefficient is zero in all combinations are not included in the figure. For the complete results for the four months, November to February, across all 31 random seeds and all three CORINE windows, we refer the reader to Supplementary Figs. S1 to S4. These are the matrices that were averaged for each of the four month-CORINE combinations to obtain Fig. 2.

As described in the Methods section, we apply a threshold of 20 out of the 31 random seed runs (about 64%) to select the land cover variables that are included in the unconstrained GLM. The consensus thus obtained for each of the 12 month-CORINE year combinations is shown in Fig. 3.

There are a number of features of note in Fig. 3. Firstly, across all three of the CORINE windows, the models for the months November, December and January include only a small number of CORINE land cover variables. Within these months, the coefficients associated with those land cover variables are predominantly positive (shown in green in Fig. 3). For example, for 2006, Estuaries, Intertidal flats and Salt marshes have positive coefficients for all four months. In addition to these three

variables, Beaches, dunes and sands and Port areas have positive coefficients for December through February. Inland marshes feature (again with positive coefficients) in January and February with Sport and leisure facilities also featuring for February. The only negative (brown) coefficient in the 2006 CORINE window is for Non-irrigated arable land, in January 2006. In 2012, Estuaries, Intertidal flats, Salt marshes and Inland marshes have positive coefficients across all four months. Coastal lagoons have a positive coefficient in November 2012 and Port areas feature with a positive coefficient in November, December and February. Many other variables feature in February 2012, the most complex model selected in the 12 month-CORINE year combinations. The ones with positive coefficients are Water bodies, Water courses, Beaches, dunes and sands, Natural grasslands, Sport and leisure facilities, Dump sites and Industrial or commercial units. The land cover classes Principally agriculture, Pastures and Non-irrigated arable land have negative coefficients. In the 2018 models, Estuaries, Intertidal flats, Salt marshes, Inland marshes, Sport and leisure facilities and Port areas emerge with positive coefficients across all four months in the study. Non-irrigated arable land has a negative coefficient in all four months. In November and February 2018, Coastal lagoons have a positive coefficient, and in February 2018, Water bodies and Beaches, dunes and sands also have positive coefficients.

In order to assess the efficacy of the method described above for selecting an appropriate subset of land cover class explanatory variables in the binomial GLMs for the individual monthly models, we used the area under the receiver operator characteristic curve (AUC) and the modified Boyce index as our model evaluation metrics (see Materials and Methods for details). In both cases, these metrics were evaluated on the test data set (i.e. that half of the raster cells which remained after the training data were removed, subject to the constraint that the proportion of cells with a Curlew sighting be constant across training and test data sets).

In Fig. 4 we see values of the AUC above 0.85 (horizontal line in Fig. 4) for months November through to February in most of the monthly models. This is an indication that the predictive power of the models is high.

In Fig. 5 we see the modified Boyce index for all of the monthly models. The majority of months have a value above 0.75 for this metric

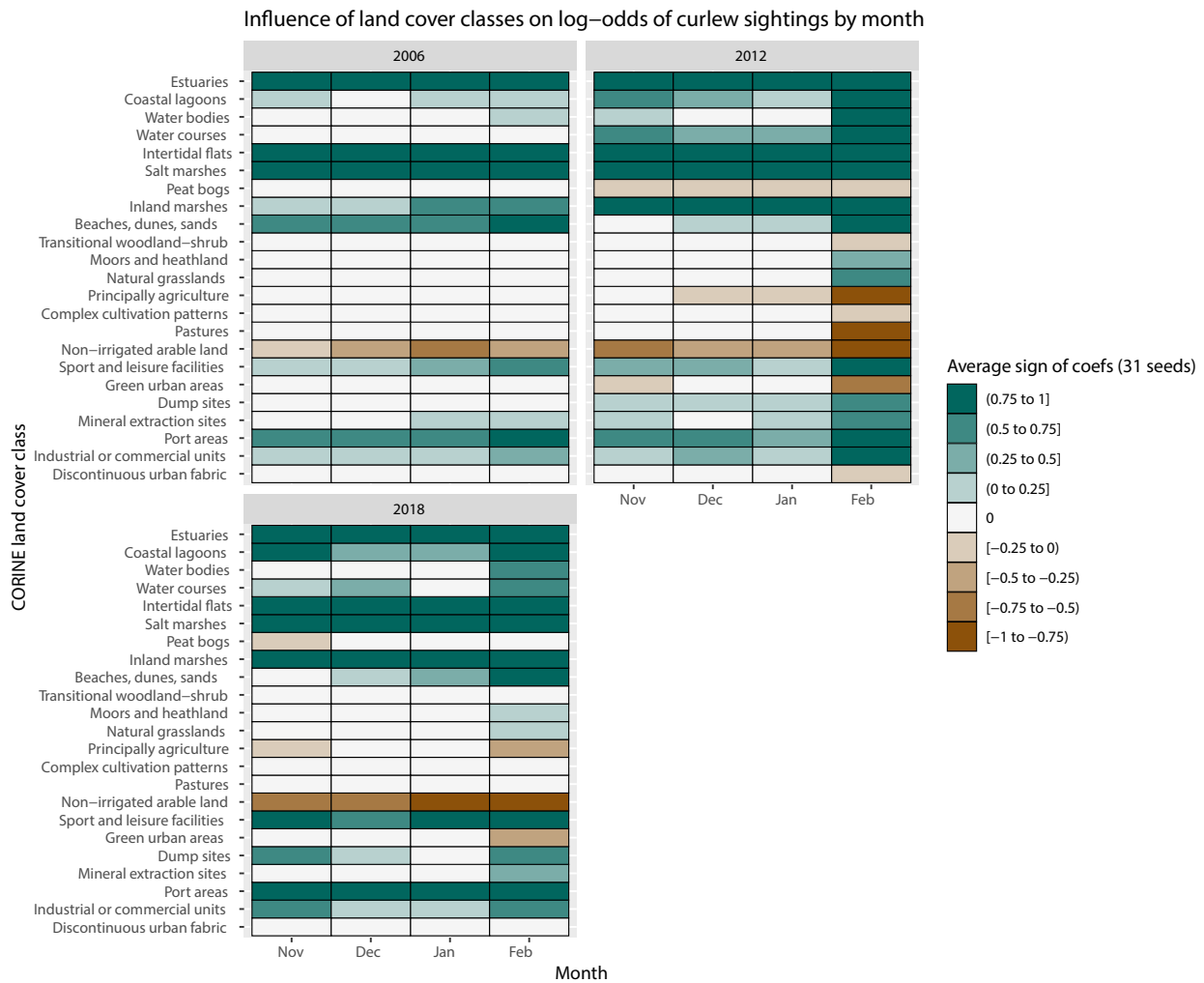


Fig. 2. The mean signs (negative = brown, 0 = white, positive = green), for 31 random seeds, of the coefficients obtained using lasso penalisation of the land cover variables in the 12 month-CORINE combinations. Note the x-axis runs from November to February. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(horizontal line in Fig. 5). The months where the modified Boyce index is lower than 0.75 are January 2008, 2009 and 2016, November 2003, 2004, 2005 and 2010, and December 2012 and 2013. Note, originally the Boyce index was constructed as a metric for models involving presence-only data. We have fitted logistic models where we have coded grid squares where no Curlew were observed as 0 s. As we note in the Introduction and Discussion, we cannot distinguish between true absences and grid squares in which Curlew were present but not detected in particular months. Nevertheless, the modified Boyce index offers a quantitative evaluation of the extent to which cells with a high predicted probability correspond to actual Curlew observations in the data sets.

3.1. Model predictions

For each month in our study, we use the model obtained from the training data set to make predictions of the probability of a Curlew sighting across all of the raster cells for that month. Hence, for each month-CORINE window combination, we have a number of rasters of predicted probabilities. For example, for January-CORINE2006, we have a predicted probability raster for each of January 2003, 2004, 2005, 2006, 2007, 2008. In Figs. 6, 7, and 8 we see the means and standard deviations of the predicted probabilities of Curlew sightings across Britain and Ireland for the period November to February for CORINE windows 2006, 2012 and 2018 respectively.

For all three of the CORINE windows, (Figs. 6, 7, and 8), we observe a distinctly coastal distribution (across both Ireland and Britain) of regions with a high predicted probability of Curlew sightings between November and February (the red regions on the yellow and red maps in the Figures).

When we consider the variability of the fitted probabilities by looking at the standard deviations across the years of the CORINE windows (the blue and white maps in Figs. 6, 7, and 8), we observe that there is low variability across highland regions in Britain and Ireland for the months November to February for all three CORINE windows. This is to be expected since the predicted probabilities are generally lower in these regions at these times of year. For the CORINE 2006 and 2012 windows, variability is fairly uniform across Ireland for November through to February. The eastern and south-eastern parts of Britain show high variability November to February for all three CORINE windows.

3.2. Comments on coefficients of climate and terrain variables in models

In the Supplementary Material, Figs. S5 to S16 show the coefficients obtained in the fitted models for the months November through to February for each of the three CORINE windows. The two variables corresponding to number of dry days (less than 0.05 mm rain and less than 0.5 mm rain) do not show a clear relationship with the log-odds of a Curlew sighting. In some month-CORINE combinations, both

Influence of land cover classes on log-odds of curlew sightings by month

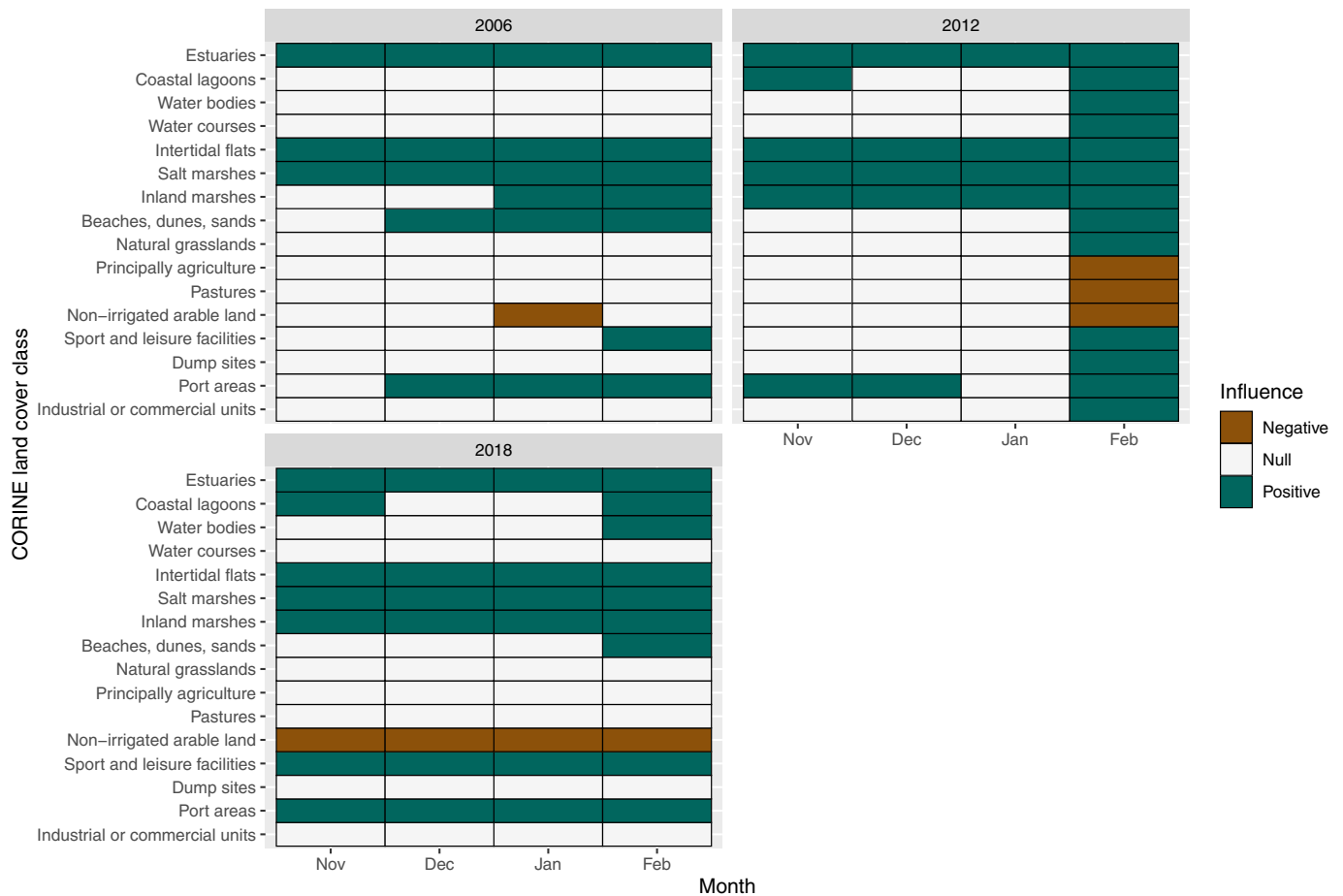


Fig. 3. The consensus signs (negative = brown, 0 = white, positive = green) of the coefficients obtained using lasso penalisation of the land cover variables in the 12 month-CORINE combinations. Note the x-axis runs from November to February. Here we only show coefficients for which the number of non-zero entries across 31 random seeds is at least 20. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

coefficients are close to 0; in others, one or both of the coefficients are significantly different from 0, but with a lot of variability in the pattern.

However, the number of frost days in a particular grid square is influential on the log-odds of a Curlew sighting, with the majority of models including a significant negative coefficient for the frost days variable. (Although this is not uniformly the case across all months in all of the CORINE windows, the pattern is nevertheless striking enough to conclude that Curlew are in general less likely to be seen in colder regions.) The total precipitation is not a particularly influential variable in any of the models.

In the period November to February, the DTM variable is often negative and the slope percent variable is often strongly positive. These two variables are correlated, and a balance between the two offers flexibility in the models. The inclusion of these two variables achieved a considerable improvement in the fit of the models, resulting in a reduction in the residual deviance of the order of 1000s. (Data not shown.)

4. Discussion

In order to assess the distribution of Curlew sightings through the winter (November to February), we have combined a CORINE-derived land cover map with environmental variables. Lasso penalisation applied to the 37 CORINE land cover classes across the the island of Ireland and Great Britain generated models with a fairly small number of land cover class variables (in the range 3–8) for the winter months

November, December and January (Fig. 3). This suggests that the Curlews utilise a specific set of habitats through their wintering distribution. Estuaries, Intertidal flats, Salt marshes, Beaches, dunes and sands and Port areas feature in most of the selected models. This concurs with the known coastal distribution of Curlew in Britain and Ireland (Humphreys et al., 2020), with models parameterised at a national extent reporting high accuracy values (Figs. 4, 5).

February appeared to be the most complex month in terms of variables selected by the lasso regression (at least in 2012 and 2018) compared to the other months of November to January (Fig. 3). However, the locations of habitat suitability did not visibly change across the four months or three CORINE windows (Figs. 6, 7, and 8). Land cover categories of significance in February 2012 include coastal lagoons (+ve), water bodies (+ve), water courses (+ve) natural grasslands (+ve), agriculture (-ve), pasture (-ve) non-irrigated arable land(-ve), dump sites (+ve) and industrial and commercial units (+ve). That waders avoid artificial structures during the breeding season is well established (Wallander et al., 2006), but our results suggest a slight contradiction to this for the wintering distribution, especially in February. February is a month in the UK and Ireland when farmers begin to prepare the ground for crops, which could provide access to additional invertebrate resources in freshly disturbed fields, which could be increasingly important for the birds due to potentially harsher weather conditions at the coast. Populated coastal areas have a legacy of numerous solid waste disposal sites (Nicholls et al., 2021), as well as a high proportion of industrial and commercial units, including airport

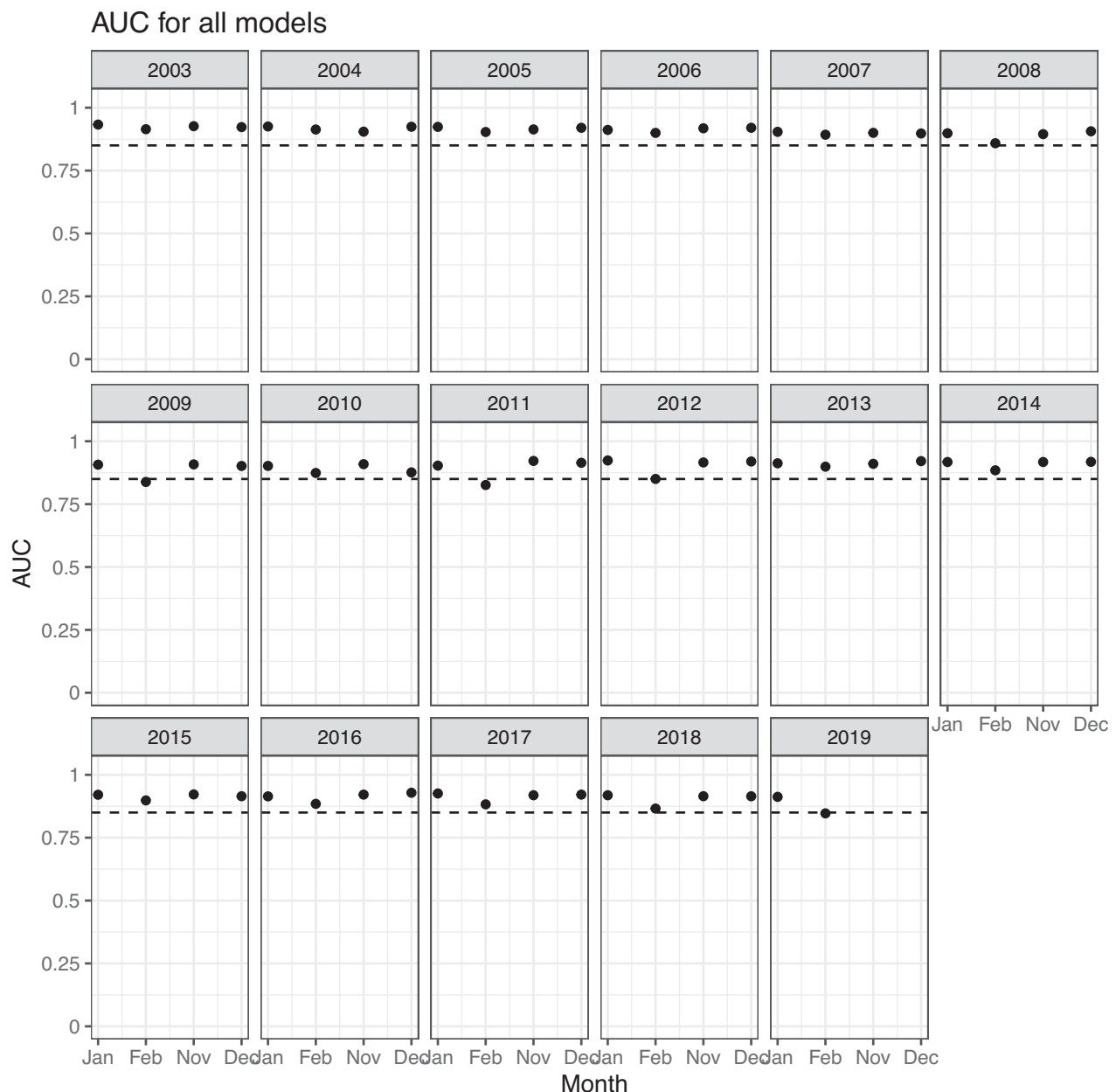


Fig. 4. The area under the curve (AUC) for individual monthly models in January, February, November and December. The top row (2003–2008) corresponds to the CORINE 2006 window, the middle row (2009–2014) to the CORINE 2012 window and the bottom row (2015–2019) to the CORINE 2018 window. The horizontal line is at 0.85.

installations and shopping centres (Flood and Sweeney, 2012; Paprotny and Terefenko, 2017). The inclusion of these features in the models may be nothing more than a proxy for coastal habitats; however, these locations are often associated with lower (human) population densities (Dikshit et al., 2001) and can transition to areas with unused and (relatively) abandoned green spaces (Hackworth, 2014) which could create a preferential feeding opportunity for waders. However, such inference needs to be considered critically with regard to reported sightings, as these areas may represent a form of sampling bias whereby people have more access to these locations than surrounding private agricultural land, and consequently increase the observations compared to other habitats nearby.

Moreover, while natural grasslands were reported to have a positive relationship with Curlew in February 2012, we would have expected to have identified a positive relationship with improved grasslands most

associated with pastures and grass-based agriculture where their predominant food source is earthworms (Navedo et al., 2020). We note here that BirdTrack records are usually submitted from places birdwatchers find interesting to visit. Agriculture, pasture and non-irrigated arable land are all likely to be low-reward environments for birdwatching during the winter months (compared to, e.g. wetland or estuarine sites). This lack of recording effort, combined with potential access issues concerning farmland may contribute to the observed negative associations. This spotlights the importance of tracking technology (e.g., Global Positioning System – GPS) that can remotely monitor the birds in hard-to-access locations and that affords new insights into their habitat use and distribution (Demsar et al., 2015). To date, tracking data have been less readily employed within SDM due to a variety of methodological challenges including spatial and temporal autocorrelation and small geographic distributions (Holloway and Miller, 2017), but opportunities

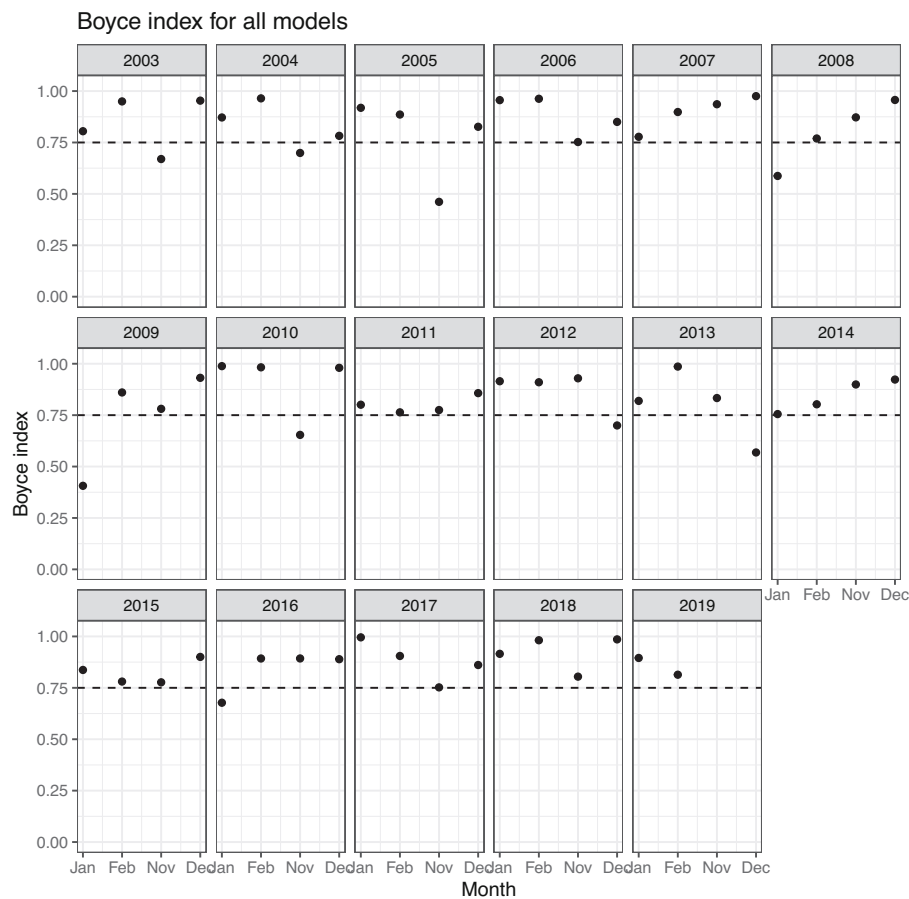


Fig. 5. The (modified) Boyce index for all individual monthly models for months January, February, November and December. Horizontal line is at 0.75.

exist to inform on regional patterns of wintering distribution. Alternatively, the lack of grassland and pasture reported in the models in Fig. 3 may represent the overwhelming absence of Curlew on these habitats inland. For example, Ireland is estimated to be 58.4% grassland (CSO, 2020), with much of this in the centre of the country. Due to the coastal distribution of the Curlew in wintering months, at a national scale these habitats are largely being avoided by the Curlew, except within coastal regions (Figs. 6, 7, and 8). This suggests that finer-scale case-studies are required to parse out the influence of coastal landscape configurations that are important to wintering Curlew, and consequently an area of future research. It also highlights the importance of considering the wider spatial configuration of the landscape and not simply the proportion of land cover. Several studies (Portaccio et al., 2021; Radović et al., 2011) parameterise species distribution with the percentage land cover in an area, but perhaps the importance of focal neighbourhoods are important here. The winter time importance of habitats to Curlew is not purely determined by land use but is a function of land use and other parameters such as proximity to estuarine/coastal locations, communal roosts, etc. which are hard to capture in a classic SDM.

The more 'complex' pattern of February in land covers (2012, 2018) may reflect a shift in the temperature, as January and February usually make up the two coldest months of the year (Met Office, 2016; Walsh, 2012). Here, birds may be required to travel to suboptimal habitats in search of food or to avoid the harsher conditions at the coast. For example, Navedo et al. (2020) identified Curlew abundance increased in fields that had the lowest food availability, feeding preferentially on the smallest class (less than 32.5 mm) of epigeic earthworms. These authors imply a trade-off between energy intake rate and other fitness components, or alternatively it could be a competitive interaction among other

waders or gulls, or again low quality farmland is less disturbed by people, animals, and machinery than higher quality grassland. Alternatively, birds in February may be preparing for migration and may make short journeys inland or up and down the coast, bringing them into contact with a wider range of land uses than in the immediate vicinity of their main wintering grounds. Regardless, our results certainly suggest that Curlew in the UK and Ireland begin to utilise a wider range of habitats in February (consistent across 2012–2018).

There is strong evidence that as the number of frost days in a particular grid square in a given month increases, the probability of seeing a Curlew in that square tends to decrease. The numbers of models with negative coefficients for the frost variable are 13/16 (81%) for November models, 12/16 (75%) for December models, and 16/17 (94%) for the January and February models, with little discernible pattern in the variation between different CORINE windows. Abundance patterns for Curlew have been noted to drop during cold winters, with Woodward et al. (2021) noting milder winters may increase abundance and possibly affect distributions. As the number of frost days increases, the substrates into which Curlew are probing get harder with the potential of damaging the birds' beaks. In addition, the invertebrates move to deeper levels and become inaccessible to the Curlew.

While the compilation of different data sources may impact the results of the models, recent research has identified that data quantity is more important than spatial bias in the predictive performance of SDMs (Gaul et al., 2020). Due to the high data requirements of parameterizing monthly SDMs, we opted for data quantity over spatial bias and uncertainty introduced by combining multiple sampling structures. Fitting temporally explicit species distribution models that capture seasonal distributions is one of the research frontiers of SDM (Milanesi et al.,

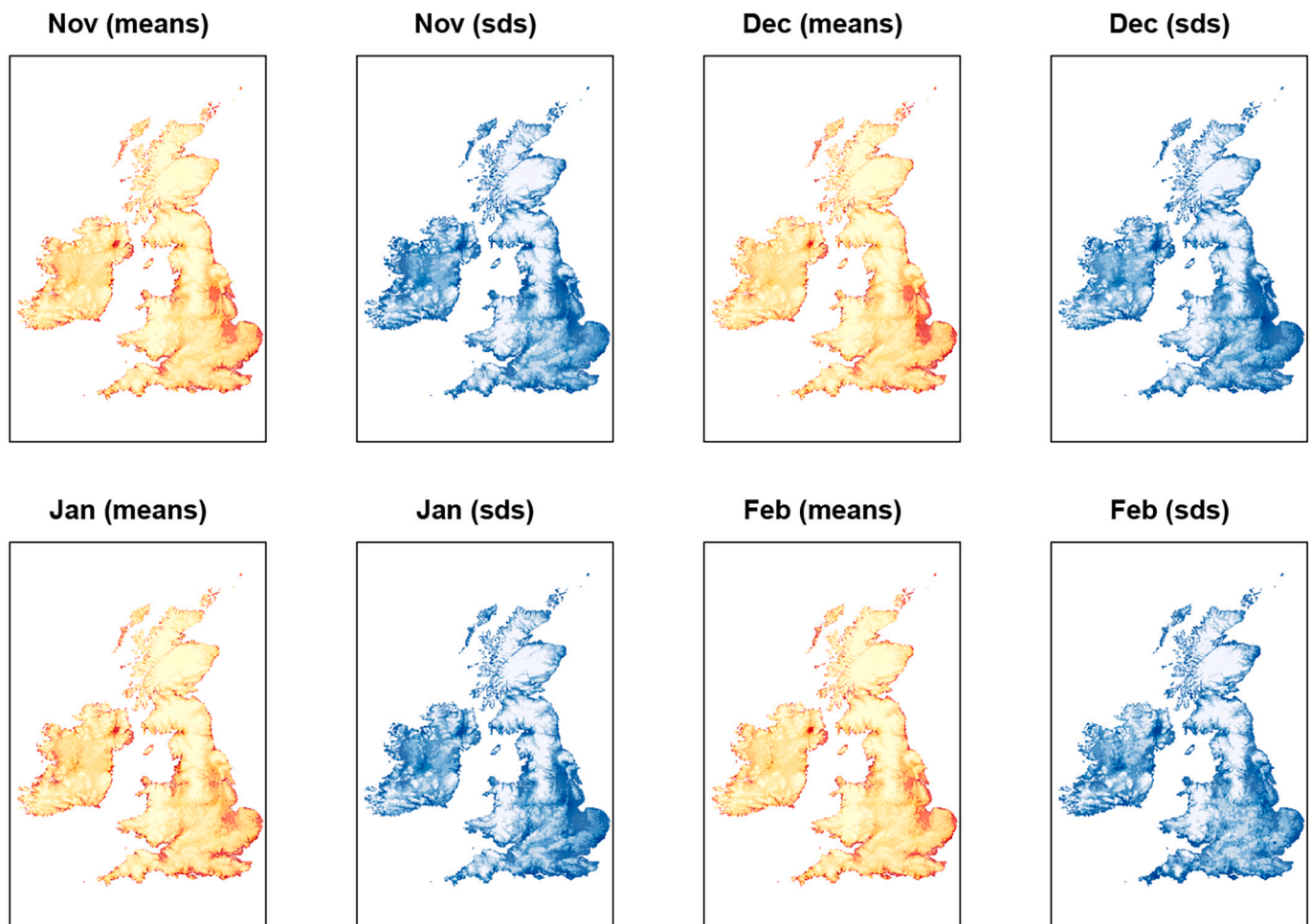


Fig. 6. The yellow and red plots show the mean predicted probabilities for the models (fitted on training data sets consisting of half of the observations) for November through to April for the CORINE2006 window. The redder the points, the higher the probabilities. The blue and white plots show the standard deviations of the predicted probabilities. The darker the blue, the higher the standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2020); coupled with the high accuracy of our models (Figs. 4 and 5) and ecologically realistic outputs (Figs. 6 to 8), we are confident this decision has not significantly biased the results. Moreover, when coupled with Supplementary Tables S1 to S3, it is clear that such an approach was needed to develop monthly models of temporally explicit distributions. We should also note here that we investigated a binary representation of presence and pseudo-absence. As such, there persists a need to explore habitat suitability for wintering Curlew in conjunction with other population measures, such as abundance and density. New methods, such as Poisson point process models, are emerging as suitable tools to predict the distribution and abundance of species from presence-only data (Schank et al., 2017). Recent research has shown that that cold winter weather and increased bird population density negatively affected wintering Curlew survival rates in the UK (Cook et al., 2021), suggesting that future studies should focus on developing predictions of both the distribution and density of Curlew to best inform biodiversity conservation.

To provide consistency across the international borders for the entire temporal period of study, we used three CORINE time windows with the assumption that land cover was consistent for those months. Temporally explicit SDMs provide novel insight into the dynamic distributions of species (Milanesi et al., 2020). Developing temporally explicitly monthly SDMs was a primary aim of our research, but we do acknowledge that some of the land covers may have changed. A recent study by

Cole et al. (2022) noted that only 1.16% of the UK changed between 2012 and 2018, with a previous report documenting an increase in intertidal flat coverage from 274,419 ha in 2006 to 274,597 ha in 2012, with 127 ha of land changed from salt marshes to intertidal flats and a further 15 ha of salt marshes converted to either industrial and commercial units (9 ha) or construction sites (6 ha) (Cole et al., 2018). These numbers are small, and when considered at the national scale of UK and Ireland, we are confident that the monthly models have not been significantly compounded by unobserved land cover conversions. However, these changes do indicate that there are certain anthropogenic processes that may affect Curlew overwintering habitats, and as such future research should explore how short-scale land cover conversions may impact Curlew behaviour, particularly in relation to the winter distribution.

5. Conclusion

Using lasso penalisation as a model selection procedure in logistic generalized models has enabled us to identify a small set of land cover variables that, when combined with climate variables and a digital terrain model, have good explanatory power in terms of Curlew overwintering distributions on the island of Ireland and Great Britain (Figs. 4 & 5). Reproducibly, from November to January between 2003 and 2019, the coastal land cover classes, Estuaries, Intertidal Flats, Salt Marshes

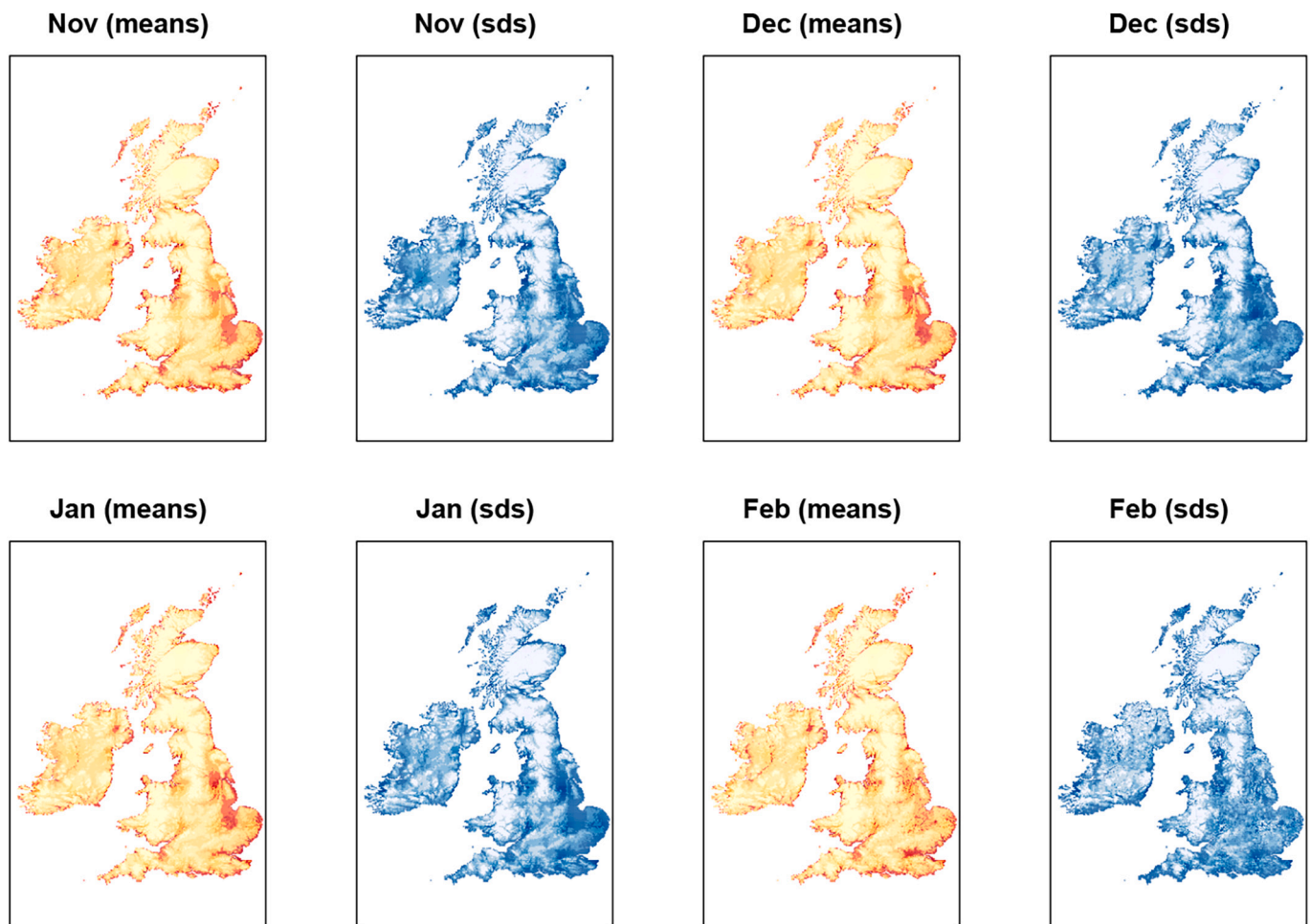


Fig. 7. The yellow and red plots show the mean predicted probabilities for the models (fitted on training data sets consisting of half of the observations) for November through to April for the CORINE2012 window. The redder the points, the higher the probabilities. The blue and white plots show the standard deviations of the predicted probabilities. The darker the blue, the higher the standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and Port Areas, feature prominently in the sets of explanatory variables selected by the lasso regression. Models fitted on training data sets including half of the grid squares coded 1 (Curlew observed) and half of those coded 0 (Curlew not observed) show higher predicted probabilities of Curlew sighting around the coasts over the 17 years of the study.

This research has demonstrated that for modelling Curlew distributions, satellite land cover data are a useful adjunct to environmental data. Further, when a robust model selection procedure is used (in our case cross-validated lasso regression), it becomes possible to observe consistent patterns over time in terms of the habitat types on which Curlew are observed. The winter distribution of Curlew sightings we observe is congruent with existing studies, while offering a more detailed understanding of where and when Curlew are observed. Moreover, this study represents the first regional scale analysis on the impact of landscape and climate features on wintering curlew distribution, identifying the importance of landscape factors that warrant further research, such as the importance of artificial structures and the importance of February within the migration of the Curlew.

Data availability

Observation records for Eurasian curlew were obtained from the following sources: eBird from Cornell Lab of Ornithology: Curlew records for the United Kingdom, the Republic of Ireland, and the Isle of

Man [1]. NBN Atlas for the UK National.

Biodiversity Network: Curlew records for the United Kingdom and the Isle of Man [2]. Ireland National Biodiversity Data Centre (NBDC): Birds of Ireland data set [3]. British Trust for Ornithology (BTO): Bird-Track [4]. British Trust for Ornithology (BTO): WeBS [5]. Bird-WatchIreland (BWI): I-WeBS [6].

eBird data requests require online application; requests are typically granted within hours. NBN and NBDC data are in the public domain and available through a formal request. Accessing both BTO and BWI data entail a formal process of request submission and the completion of an online form (one per data source). Curlew observations from the above sources recorded during the period January 2003 to were amalgamated and reprojected to the European Lambert azimuthal equal area coordinate reference system used by Corine land cover. Presences were recorded per square kilometre, per month, where an observation was recorded in any of the above data sources. In square kilometres where no observation was recorded in any of our data sources, we inferred that curlew were absent. We are conscious that this includes poorly-surveyed areas and accept the possibility that there are a non-negligible quantity of false absences in our derived data set. In combining multiple data sets, we have simply sought to minimise false absences as much as possible. [1]

eBird. 2021. eBird: An online database of bird distribution and abundance [web application]. eBird, Cornell Lab of Ornithology, Ithaca,

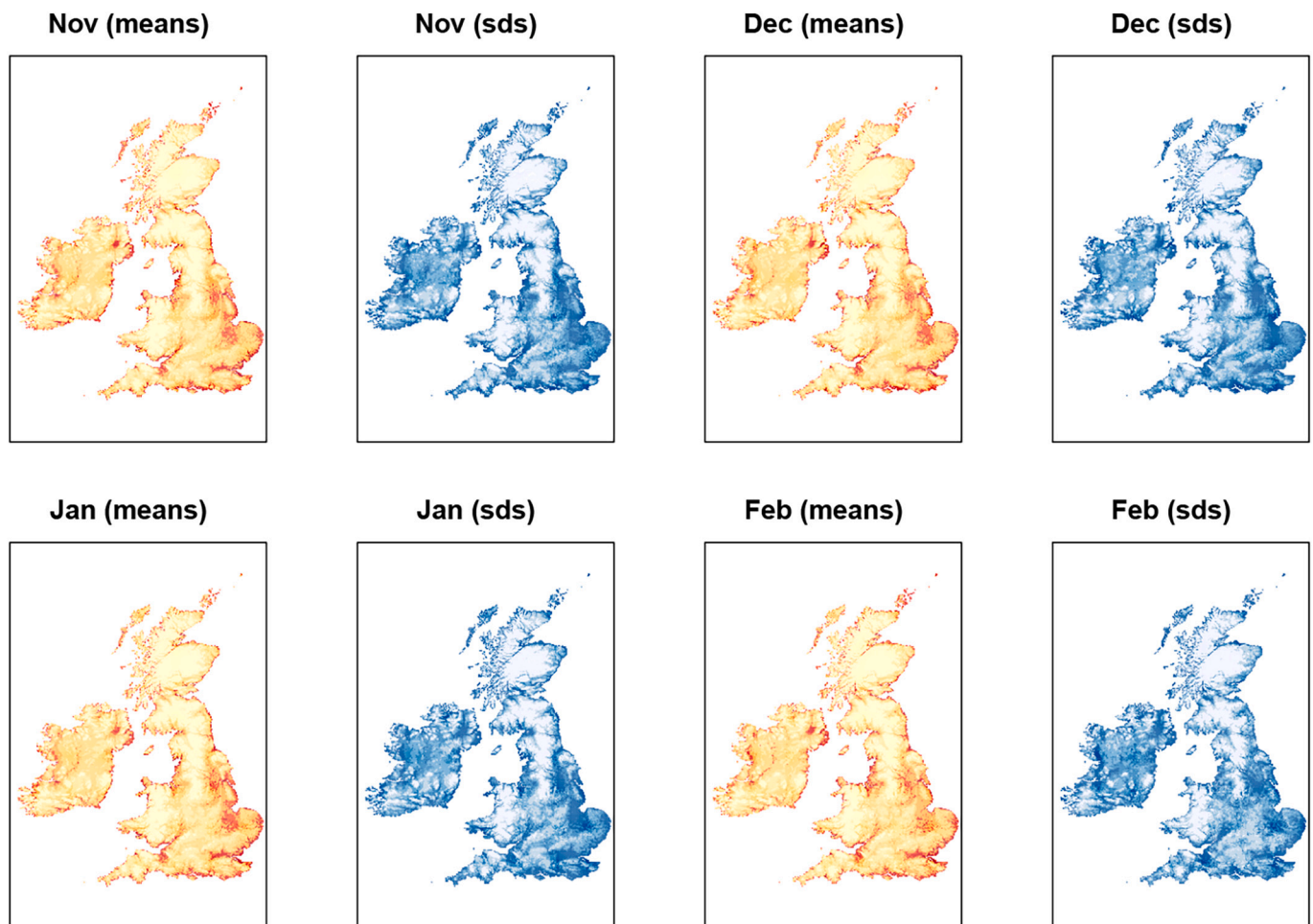


Fig. 8. The yellow and red plots show the mean predicted probabilities for the models (fitted on training data sets consisting of half of the observations) for November through to April for the CORINE2018 window. The redder the points, the higher the probabilities. The blue and white plots show the standard deviations of the predicted probabilities. The darker the blue, the higher the standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

New York. Available: <http://www.ebird.org>. (Accessed: November 12, 2021).[2]

See Supplementary Material[3]

Birds of Ireland. 2021. Birds of Ireland online data set. Birds of Ireland, National Biodiversity Data Centre. Available: <https://maps.biodiversityireland.ie/Dataset/155>. (Accessed: November 12, 2021).

[4] British Trust for Ornithology, online Data Request form. Available: <https://app.bto.org/data-request/new-data-request.jsp>. (Submitted, requesting BirdTrack data: October 14, 2021).[5]

British Trust for Ornithology, downloadable WeBS Data Request Form. Available: <https://www.bto.org/our-science/projects/wetland-bird-survey/data/submit-data-request>. (Completed and sent: October 14, 2021).[6]

BirdWatch Ireland, downloadable I-WeBS Data Request Form. Available: <https://birdwatchireland.ie/publications/i-webs-data-request-form/>. (Completed and sent: December 8, 2021).

Code and guidance for reproducible analysis are available at: https://github.com/echoes-sdm/curlew_SDM

Acknowledgements

This research is part of the ECHOES project. ECHOES is funded through the Ireland-Wales Programme 2014–2020 and part-funded by

the European Regional Development Fund through the Welsh Government. We would like to thank Enda O'Brien and the rest of the Met Éireann funded TRANSLATE project for the provision of Irish climate projections. We would also like to thank the reviewers and editors for their comments and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2023.102244>.

References

- Barnett, Kirk L., Facey, Sarah L., Aug. 2016. Grasslands, invertebrates, and precipitation: a review of the effects of climate change. *Front. Plant Sci.* 7, 1196.
- Berg, Å., 1992. Habitat selection by breeding curlews *Numenius arquata* on mosaic farmland. *Ibis* 134 (4), 355–360.
- BirdLife International, 2022. IUCN Red List for Birds. <http://www.birdlife.org>. Accessed: 2022-12-15.
- Bowgen, Katharine, 2017. Predicting the effect of environmental change on wading birds: insights from individual-based models. PhD thesis. Bournemouth University in collaboration with HR Wallingford.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. *Ecol. Model.* 157, 281–300.
- Brown, D., Wilson, J., Douglas, D., Thompson, P., Foster, S., McCulloch, N., Phillips, J., Stroud, D., Whitehead, S., Crockford, N., Sheldon, R., 2015. The Eurasian curlew—the most pressing bird conservation priority in the UK. *Brit. Birds* 108 (11), 660–668.

- Burthe, S., Wanless, S., Newell, M.A., Butler, A., Daunt, F., 2014. Assessing the vulnerability of the marine bird community in the western North Sea to climate change and other anthropogenic impacts. *Mar. Ecol. Prog. Ser.* 507, 277–295.
- Burton, N.H., Armitage, M.J., Musgrove, A.J., Rehfish, M.M., 2002a. Impacts of man-made landscape features on numbers of estuarine waterbirds at low tide. *Environ. Manag.* 30 (6), 0857–0864.
- Burton, N.H., Rehfish, M.M., Clark, N.A., 2002b. Impacts of disturbance from construction work on the densities and feeding behavior of waterbirds using the intertidal mudflats of Cardiff Bay, UK. *Environ. Manag.* 30 (6), 0865–0871.
- Carroll, M.J., Heinemeyer, A., Pearce-Higgins, J.W., Dennis, P., West, C., Holden, J., Wallace, Z.E., Thomas, C.D., 2015. Hydrologically driven ecosystem processes determine the distribution and persistence of ecosystem-specialist predators under climate change. *Nat. Commun.* <https://doi.org/10.1038/ncomms8851>.
- Cole, B., Smith, G., Balzter, H., 2018. Acceleration and fragmentation of CORINE land cover changes in the United Kingdom from 2006–2012 detected by Copernicus IMAGE2012 satellite data. *Int. J. Appl. Earth Observ. Geoinform.* 73, 107–122.
- Cole, B., Smith, G., de la Barreda-Bautista, B., Hamer, A., Payne, M., Codd, T., Johnson, S. C., Chan, L.Y., Balzter, H., 2022. Dynamic landscapes in the UK driven by pressures from energy production and forestry—Results of the CORINE Land Cover map 2018. *Land* 11 (2), 192.
- Colhoun, K., Mawhinney, K., Peach, W.J., 2015. Population estimates and changes in abundance of breeding waders in Northern Ireland up to 2013. *Bird Study* 62.3, 394–403.
- Cook, A.S., Burton, N.H., Dodd, S.G., Foster, S., Pell, R.J., Ward, R.M., Wright, L.J., Robinson, R.A., 2021. Temperature and density influence survival in a rapidly declining migratory shorebird. *Biol. Conserv.* 260, 109198.
- Cruz-Cárdenas, G., López-Mata, L., Villaseñor, J.L., Ortiz, E., 2014. Potential species distribution modeling and the use of principal component analysis as predictor variables. *Rev. Mex. Biodivers.* 85 (1), 189–199.
- CSO, 2020. <https://www.cso.ie/en/releasesandpublication/ep/p-eii/environmentalindicatoreland2020/landuse/>.
- Davis, Craig A., Austin, Jane E., Buhl, Deborah A., June 2006. Factors influencing soil invertebrate communities in riparian grasslands of the Central Platte River floodplain. *Wetlands* 26 (2), 438–454.
- Demsar, U., Buchin, K., Cagnac, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D., Weibel, R., 2015. Analysis and visualisation of movement: an interdisciplinary review. *Mov. Ecol.* 3 (1), 1–24.
- Dias, M.P., Martin, R., Pearmain, E.J., Burfield, I.J., Small, C., Phillips, R.A., Yates, O., Lascelles, B., Borboroglu, P.G., 2019. Threats to seabirds: a global assessment. In: Croxall, J.P. (Ed.), *Biological Conservation*, 237, pp. 525–537.
- Dikshit, A.K., Padmavathi, T., Das, R.K., 2001. Locating potential landfill sites using geographic information systems. *J. Environ. Syst.* 28 (1), 43–54.
- Duan, H., Xia, S., Jackson, M.V., Zhao, N., Liu, Y.U., Teng, J., Meng, Z., Yu, X., Shi, J., 2020. Identifying new sites of significance to waterbirds conservation and their habitat modification in the yellow and Bohai seas in China. *Glob. Ecol. Conserv.* 22, e01031.
- Estrada, A., Real, R., 2021. A stepwise assessment of parsimony and fuzzy entropy in species distribution modelling. *Entropy (Basel)* 23 (8). <https://doi.org/10.3390/e23081014>.
- European Environment Agency (EEA), 2006. European Union. Copernicus Land Monitoring Service.
- European Union, 2012. Copernicus Land Monitoring Service.
- European Union, 2018. Copernicus Land Monitoring Service.
- Flood, S., Sweeney, J., 2012. Quantifying impacts of potential sea-level rise scenarios on Irish coastal cities. In: *Resilient Cities 2*. Springer, Dordrecht, pp. 37–52.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. <https://www.jstatsoft.org/v33/i01/>.
- Frost, T.M., Austin, G.E., Calbrade, N.A., Mellan, H.J., Hearn, R.D., Robinson, A.E., et al., 2018. Waterbirds in the UK 2017/18: The Wetland Bird Survey. <http://www.bto.org/volunteersurveys/webs/publications/webs-annual-report>. BTO, RSPB and JNCC, in association with WWT. British Trust for Ornithology, Thetford, UK.
- Gaul, W., Sadykova, D., White, H.J., Leon-Sanchez, L., Caplat, P., Emmerson, M.C., Yearsley, J.M., 2020. Data quantity is more important than its spatial bias for predictive species distribution modelling. *PeerJ* 8, e10411.
- Grant, M.C., Orsman, C., Easton, J., Lodge, C., Smith, M., Thompson, G., Rodwell, S., Moore, N., 1999. Breeding success and causes of breeding failure of curlew *Numenius arquata* in Northern Ireland. *J. Appl. Ecol.* 36 (1), 59–74.
- Hackworth, J., 2014. The limits to market-based strategies for addressing land abandonment in shrinking American cities. *Prog. Plan.* 90, 1–37.
- Hastie, T., Qian, J., Tay, K., June 2021. An Introduction to Glmnet. <https://glmnet.stanford.edu/articles/glmnet.html>.
- Henderson, I.G., Wilson, A.M., Steele, D., Vickery, J.A., 2002. Population estimates, trends and habitat associations of breeding lapwing *Vanellus vanellus*, curlew *Numenius arquata* and snipe *Gallinago gallinago* in Northern Ireland in 1999. *Bird Study* 49 (1), 17–25.
- Heywood, J.J.N., Massimino, D., Balmer, D.E., Kelly, L., Noble, D.G., Pearce-Higgins, J. W., Woodcock, P., Wotton, S., Gillings, S., Harris, S.J., 2023. The Breeding Bird Survey 2022. BTO Research Report 756. British Trust for Ornithology, Thetford.
- Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* 199, 142–152.
- Holloway, P., Field, R., 2020. Can rock-rubble groynes support similar intertidal ecological communities to natural rocky shores? *Land* 9 (5), 131.
- Holloway, P., Miller, J.A., 2017. A quantitative synthesis of the movement concepts used within species distribution modelling. *Ecol. Model.* 365, 91–103.
- Hughes, R.G., 2004. Climate change and loss of saltmarshes: consequences for birds. *Ibis* 146, 21–28.
- Humphreys, E.M., Austin, G.E., Frost, T.M., Mellan, H.J., Boersch-Supan, P., Burton, N.H. K., Balmer, D.E., 2020. Wader populations on the United Kingdom's open coast: results of the 2015/16 non-estuarine Waterbird survey (NEWS-III) and a review of population trends. *Bird Study* 67 (3), 371–384.
- Iyer, S.G., Banerjee, A.K., Bhowmick, A.R., 2019. Making choices that matter – use of statistical regularization in species distribution modelling for identification of climatic indicators – a case study with *Mikania micrantha* Kunth in India. *Ecol. Indic.* 98, 92–103.
- Johnstone, I., Elliot, D., Mellenchip, C., Peach, W.J., 2017. Correlates of distribution and nesting success in a Welsh upland Eurasian curlew *Numenius arquata* population. *Bird Study* 64 (4), 535–544.
- Jones, K.A.B., O'Connell, P., Wolsey, S., Carrington-Cotton, A., Noble, D.G., McCulloch, N., Calladine, J.R., 2022. Loss of breeding waders from key lowland grassland sites in Northern Ireland. *Irish Birds* 44, 11–22.
- Kennedy, J., Burke, B., Fitzgerald, N., Kelly, S.B.A., Walsh, A.J., Lewis, L.J., 2022. Irish Wetland Bird Survey: I-WeBS National and Site Trends Report 1994/95–2019/20. BirdWatch Ireland Waterbird Report to the National Parks and Wildlife Service. Tech. rep. Birdwatch Ireland, Wicklow.
- Koma, Z., Seijmonsbergen, A.C., Grootes, M.W., Nattino, F., Groot, J., Sierdsema, H., Foppen, R.P., Kissling, W.D., 2022. Better together? Assessing different remote sensing products for predicting habitat suitability of wetland birds. *Divers. Distrib.* 28 (4), 685–699.
- Lambert, L., Cawkwell, F., Holloway, P., 2021. The importance of connected and interspersed urban green and blue space for biodiversity: a case study in Cork City, Ireland. *Geographies* 1 (3), 217–237.
- Lawicki, L., Wylegala, P., 2011. Recent data on the declining breeding population of Eurasian curlew *Numenius arquata* in western Poland. *Wader Study Group Bull.* 118, 14–17.
- Manton, M., Angelstam, P., 2021. Macroecology of north European wet grassland landscapes: habitat quality, waders, avian predators and nest predation. *Sustainability* 13 (15), 8138.
- McDonald, A., Heath, M.R., Edwards, M., Furness, R.W., Pinnegar, J.K., Wanless, S., Speirs, D.C., Greenstreet, S.P.R., 2015. Climate-driven trophic cascades affecting seabirds around the British Isles. *Oceanogr. Mar. Biol. Annu. Rev.* 53, 55–79.
- Met Office, 2016. UK Regional Climates. Summaries of the Climate Characteristics of 11 Regions in the UK. <https://www.metoffice.gov.uk/research/climate/maps-and-data/regional-climates/index>. Accessed: 2022-12-16.
- Milanesi, P., Della Rocca, F., Robinson, R.A., 2020. Integrating dynamic environmental predictors and species occurrences: Toward true dynamic species distribution models. In: *Ecology and Evolution*, 10.2, pp. 1087–1092.
- Munilla, I., Díez, C., Velando, A., 2007. Are edge bird populations doomed to extinction? A retrospective analysis of the common guillemot collapse in Iberia. *Biol. Conserv.* 137 (3), 359–371.
- Navedo, J.G., Gutiérrez, J.S., Salmón, P., Arranz, D., Novo, M., Díaz-Cosín, D.J., Herrera, A.G., Masero, J.A., 2020. Food supply, prey selection and estimated consumption of wintering Eurasian curlews feeding on earthworms at coastal pastures. *Ardea* 107 (3), 263–274.
- Neumann, B., Vafeidis, A.T., Zimmermann, J., Nicholls, R.J., 2015. Future coastal population growth and exposure to sea-level rise and coastal flooding—a global assessment. *PLoS One* 10 (3), e0118571.
- Nicholls, R.J., Beaven, R.P., Stringfellow, A., Monfort, D., Le Cozannet, G., Wahl, T., Gebert, J., Wade, M., Arns, A., Spencer, K.L., Reinhard, D., 2021. Coastal landfills and rising sea levels: a challenge for the 21st century. *Front. Mar. Sci.* 8, 710342.
- O'Brien, E., Nolan, P., 2023. TRANSLATE: standardised climate projections for Ireland. *Front. Clim.* 5, 1166828.
- O'Brien, M., Tharme, A., Jackson, D., 2002. Changes in breeding wader numbers on Scottish farmed land during the 1990s. *Scott. Birds* 23 (1), 10–21.
- O'Donoghue, B.G., Donaghy, A., Kelly, S.B.A., 2019. National survey of breeding Eurasian curlew *Numenius arquata* in the Republic of Ireland, 2015–2017. *Wader Study* 126, 43–48.
- Oro, D., 2014. Seabirds and climate: knowledge, pitfalls, and opportunities. *Front. Ecol. Evol.* 2, 79.
- Paprotny, D., Terefenko, P., 2017. New estimates of potential impacts of sea level rise and coastal floods in Poland. *Nat. Hazards* 85 (2), 1249–1277.
- PECBMS, 2020. Trends of Common Birds in Europe, 2020 Update. Full text, EBCC, Prague.
- Portaccio, A., Basile, M., Favaretto, A., Campagnaro, T., Pettenella, D., Sitzia, T., 2021. The role of Natura 2000 in relation to breeding birds' decline on multiple land cover types and policy implications. *J. Nat. Conserv.* 62, 126023.
- Radović, A., Bukovec, D., Trtković, N., Tepić, N., 2011. Corine land cover changes during the period 1990–2000 in the most important areas for birds in Croatia. *Int. J. Sust. Dev. World* 18 (4), 341–348.
- Runge, C.A., Martin, T.G., Possingham, H.P., Willis, S.G., Fuller, R.A., 2014. Conserving mobile species. *Front. Ecol. Environ.* 12 (7), 395–402.
- Russo, A., Cirella, G.T., 2021. Urban ecosystem services: new findings for landscape architects, urban planners, and policymakers. *Land* 10 (1), 88.
- Santosa, F., Symes, W.W., 1986. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comput.* 7 (4), 1307–1330. <https://doi.org/10.1137/0907087>.
- Schank, C.J., Cove, M.V., Kelly, M.J., Mendoza, E., O'Farrill, G., Reyna-Hurtado, R., Meyer, N., et al., 2017. Using a novel model approach to assess the distribution and conservation status of the endangered Baird's tapir. *Divers. Distrib.* 23 (12), 1459–1471.
- Servignat, H., O'Donoghue, B.G., Nov. 2022. Curlew Conservation Programme Annual Report 2022. En. Tech. rep. National Parks & Wildlife Service.

- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.
- Sydeman, W.J., Poloczanska, E., Reed, T.E., Thompson, S.A., 2015. Climate change and marine vertebrates. *Science* 350, 772–777.
- Taylor, R.C., Dodd, S.G., 2013. Negative impacts of hunting and suction-dredging on otherwise high and stable survival rates in curlew *Numenius arquata*. *Bird Study* 60 (2), 221–228.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58 (1), 267–288.
- Valkama, J., Robertson, P., Currie, D., 1998. Habitat Selection by Breeding Curlews (*Numenius arquata*) on Farmland: The Importance of Grassland. *Annales Zoologici Fennici*, pp. 141–148.
- Van Doren, B.M., Conway, G.J., Phillips, R.J., Evans, G.C., Roberts, G.C., Liedvogel, M., Sheldon, B.C., 2021. Human activity shapes the wintering ecology of a migratory bird. *Glob. Chang. Biol.* 27 (12), 2715–2727.
- Wallander, J., Isaksson, D., Lenberg, T., 2006. Wader nest distribution and predation in relation to man-made structures on coastal pastures. *Biol. Conserv.* 132 (3), 343–350.
- Walsh, S., 2012. A Summary of Climate Averages 1981–2010 for Ireland, Climatological Note No. In: 14. Met Éiraeann, Dublin.
- Woodward, I.D., Austin, G.E., Boersch-Supan, P.H., Thaxter, C.B., Burton, N.H.K., 2021. Assessing drivers of winter abundance change in Eurasian curlews *Numenius arquata* in England and Wales. *Bird Study* 68 (3), 289–301.
- Yang, H.Y., Chen, B., Barter, M., Piersma, T., Zhou, C.F., Li, F.S., Zhang, Z.W., 2011. Impacts of tidal land reclamation in Bohai Bay, China: ongoing losses of critical Yellow Sea waterbird staging and wintering sites. *Bird Conserv. Int.* 21 (3), 241–259.
- Young, J.C., McCluskey, A., Kelly, S.B., O'Donoghue, B., Donaghy, A.M., Colhoun, K., McMahon, B.J., 2020. A transdisciplinary approach to a conservation crisis: a case study of the Eurasian curlew (*Numenius arquata*) in Ireland. *Conserv. Sci. Pract.* 2 (7), e206.