

# Fine-scale mapping of relative abundance and trends, and extraction of small-area population trends for breeding birds

Jennifer A. Border & Simon Gillings



**ACKNOWLEDGEMENTS:** The BTO/JNCC/RSPB Breeding Bird Survey is a partnership jointly funded by the BTO, RSPB and JNCC, with fieldwork conducted by volunteers. This research was conducted under the development work programme of the BBS. We are grateful to BBS volunteers, without whose data these analyses would not be possible. We thank Philipp Boersch-Supan for statistical advice.

# Fine-scale mapping of relative abundance and trends, and extraction of small-area population trends for breeding birds

Jennifer A. Border & Simon Gillings

BTO Research Report 720

BTO, The Nunnery, Thetford, Norfolk, IP24 2PU

© British Trust for Ornithology 2020

BTO, The Nunnery, Thetford, Norfolk IP24 2PU  
Tel: +44 (0)1842 750050 Email: [info@bto.org](mailto:info@bto.org)  
Registered Charity Number 216652 (England & Wales), SC039193 (Scotland).

ISBN 978-1-912642-08-3



Executive summary .....	5
1. Introduction .....	6
2. Methods .....	6
2.1. BBS count data .....	7
2.2. Environmental variables.....	7
2.3. Evaluation of model fit .....	10
2.4. Producing estimated yearly abundance from the models .....	10
2.5. Extracting trends and comparison with published trends .....	10
3. Results .....	11
3.1. Selecting the final model structure .....	11
3.2. Evaluating model fit.....	12
3.3. Comparison with published trends .....	12
3.4. Predicted abundance through time.....	18
4. Discussion .....	18
4.1. Usage of predictive models .....	21
4.2. Limitations.....	21
4.3. Conclusions.....	23
5. References .....	23
6. Appendix .....	24

## EXECUTIVE SUMMARY

1. The BTO/JNCC/RSPB Breeding Bird Survey (BBS) is the UK's main scheme for monitoring population changes in the UK's common terrestrial breeding birds. Trends are published annually for up to 111 bird species for the UK, England, Scotland, Wales, Northern Ireland and nine regions within England.
2. These trends are useful at showing large-scale population changes but we know that some species show spatially varying trends that are masked by national trends, and cannot be captured by regional trends due to sample size constraints. Previous work has developed maps of abundance for different periods of the BBS time series but these are now out of date and lack the temporal resolution of standard population trends.
3. Here, we update previous work, aiming to produce annual predictions of relative abundance, which can vary spatially and temporally, for the 111 species featured in the published BBS trends from 1994 to 2016. We evaluated two methods, Generalised Additive Models (GAMs) and Geographically Weighted Regression (GWR) to model spatial and temporal variation in species abundance relative to habitat and elevation.
4. The GWR method took too long to run to be a feasible option for routine production of maps for over 100 species.
5. The GAM method worked well, yielding models with acceptable fit metrics and producing annual maps of relative abundance which agreed well with known distribution and abundance patterns from *Bird Atlas 2007–11*.
6. From these maps we extracted trend estimates which were highly correlated with the published long-term and 10-year trends.
7. These maps could be useful for setting local conservation priorities and may be better than the published BBS trends at estimating abundance for areas with poor coverage or rare and localised species. However, further work is needed to validate small area trends derived in this way against independent data.

# 1. INTRODUCTION

The UK Breeding Bird Survey (BBS) (BTO/JNCC/RSPB, Harris *et al.* 2017) is the UK's main scheme for monitoring population changes in the UK's common terrestrial breeding birds. It has been running since 1994, and the trends in breeding birds are published annually (Harris *et al.* 2017). The survey relies upon volunteers surveying a stratified random sample of 1-km squares. In each square, two 1 km long transect lines (ideally 500 m apart) are walked at a slow constant pace and all birds seen or heard are recorded. Each square receives two visits per year, one in April to mid-May and a second in mid-May to the end of June. The maximum square count for each species out of the two visits is selected as the annual measure of relative abundance. To derive the published trends for each species from these data, a Generalised Linear Model (GLM) is used, with fixed effects of year and 1-km square and a weight to account for varying regional survey effort.

Trends are produced for the UK as a whole, for England, Scotland and Wales, and for nine regions within England (see Figure A1). However, current methods require a species to occur in an average of at least 30 1-km squares a year before a trend can be calculated, which prevents the calculating of trends for smaller areas. From the three bird atlases we know that spatial variation in change over years can be substantial (Sharrock 1976, Gibbons *et al.* 1993, Balmer *et al.* 2013), for example due to local differences in habitat or climate (Newson *et al.* 2009, Morrison *et al.* 2010). The Bird Atlases can measure this at a 10-km square resolution but are restricted to measuring change between two points in time 20 years apart; whereas in reality, populations may change rapidly in a much shorter time period. For example, British Greenfinch *Chloris chloris* populations declined by 35% in some high Trichomonosis incidence regions between 2006 and 2007 (Robinson *et al.* 2010). To ensure suitable local conservation measures are put in place in time to prevent substantial population declines, it would be useful to produce spatially explicit trends at a high temporal resolution. This would allow limited funds to be used more effectively and efficiently than policies based on national trends. Producing spatially explicit trends would also allow local trends to be extracted for specific areas. This would be useful, for example, to extract local trends for areas which have undergone substantial urbanisation since 1995 and thereby analyse the effect of this urban development on the species composition.

Spatially explicit trends have previously been generated from BBS data in Massimino *et al.* 2015 to investigate the spatial variation in farmland and woodland bird indicators. But this work was restricted to a sample of 54 indicator species and the spatially explicit maps were produced for two separate time periods (1994–1996 and 2007–2009) and used to extract trends. To date, no one has analysed all years of BBS data together to produce maps that vary both spatially and temporally on an annual basis.

This project aims to update and build on the work of Massimino *et al.* 2015 by producing new maps which vary both spatially and temporally for the entire available time span of BBS data, for all species included in the published BBS trends (111). The next step will then be to assess the accuracy and predictive ability of these trends and determine if they are sufficiently robust to produce reliable trends for bespoke areas of interest, smaller than the current regions for which we can generate trends for. However, it is important to note here that it is difficult to reliably assess the accuracy of these spatially and temporarily varying trends as we do not have an independent trend dataset to assess them against and therefore will have to use the published BBS trends. There is no way of knowing how well the published BBS trends reflect the real-world situation; therefore, it is possible that mismatches between the two might be because the spatially and temporarily varying trends reflect the reality better than the published trends, rather than worse.

Ideally these trends would be generated in such a way as to allow the generation of regular maps of trends and abundance which could be presented alongside the usual published trends. In the long term this process could be automated to run alongside the usual BBS trends each year. Therefore, as well as an accurate predictive model, we also needed a method that could be automated to run for all species and would run within an acceptable time frame comparable to the usual BBS trends.

## 2. METHODS

The key aims of this project can be summarised as follows:

1. Model the relationship between BBS count data and environmental variables to produce individual fine-scale predictions of abundance for 111 species, allowing predictions to vary in space and time;

2. From these predictions produce spatially and temporally varying maps for each of the 111 species;
3. Assess the accuracy and reliability of these maps against independent sources, such as *Bird Atlas 2007–11*;
4. Summarise predictions across years to generate population trend estimates and compare these to those generated by the established BBS GLM method;
5. Make recommendations for how to implement annual spatial modelling of BBS data and future use of model-based trend estimates.

## 2.1 BBS COUNT DATA

Because we wanted modelled predictions of species relative abundance to be comparable with the published BBS trends, we used the same dependent variable: maximum count for each species per 1-km square per year. There is other work ongoing to investigate the feasibility of generating the BBS trends from density estimates, thereby producing trend estimates that control for some aspects of detectability variation. This could be achieved by converting the counts per 1-km square to densities using distance sampling techniques (Buckland *et al.* 2005; Newson *et al.* 2008), but implementing this on an annual basis is still in the development stage. Separate analyses are also investigating whether the production of BBS trends should account for variation in visit timing, variation in species phenology, and variation in the number of visits (Massimino in prep). But, again, this work is still at the development stage and these variables are currently not included in the published BBS trends. Therefore, we did not consider these variables here either, so as to ensure comparability between our results and the published BBS trends. A restricted list of 111 species (Table A1) was selected for modelling and mapping. These are species for which BBS trends were published in 2017 (Harris *et al.* 2017) which in practice means these are species historically detected in at least 40 BBS squares per year in the UK or in 30 squares per year in one of the countries or English regions. By selecting these species we have reduced the likelihood of attempting models for species where the data would be too sparse to achieve a good fit.

## 2.2. ENVIRONMENTAL VARIABLES

Possible environmental variables for inclusion in the models were limited as we wanted to apply the same model structure to all 111 species allowing the possibility of automating the modelling process in future. Additionally, we wanted to limit the number of variables as much as possible to ensure there would be sufficient data per variable to avoid over-fitting (ideally at least 10 data points per variable: Harrell *et al.* 1996) for all 111 target species, including those which were rare and localised. Therefore, we were guided by the results of Massimino *et al.* (2015) and decided to include elevation and broad habitat variables from the Centre for Ecology and Hydrology (CEH) Land Cover Map (LCM) (Table 1). These LCM data are available for a range of periods (1990, 2002, 2007, 2015), but unfortunately data from the different periods are not comparable due to differences in the way the satellite images are analysed. We chose to use the LCM2007 (Morton *et al.* 2011) as a point mid-way through our time series (1994–2016) which minimises the time frame for habitat change during the study period.

The UK is made up of many different islands as well as mainland Britain and Northern Ireland, and we know from bird atlas data that some species' distributions and abundances vary between the mainland and islands, or within island groups (Balmer *et al.* 2013). It is also likely that population trends may vary as species colonise island groups, or if key mammalian predators colonise or are introduced to islands. Therefore, we included a categorical 'island' term in the models. We trialled two versions of this term, one with just two categories: 'island' and 'mainland', and one with seven categories distinguishing the main island groups (i.e. Shetland, Orkney etc.) from Northern Ireland and mainland Britain (Table 1). We would expect the models with the seven-level categorical variable for 'island' to be better at explaining patterns in bird abundance as variation in species composition between islands would be accounted for, but there is a risk that this model might over-fit the data and prevent generalisations being made. We ran the analyses twice, once with each version of the 'island' term.

We also included 'year' so as to model temporal variation, and easting and northing, both to model spatial variation and to account for spatial auto-correlation and spatial variation in survey effort (Wood 2017). The way these variables are included differs depending on the modelling method.

**Table 1. The environmental variables included in the spatial models of bird abundance.**

Variable	Description	Source
% cover semi-natural grassland	Includes rough grassland, neutral grassland, calcareous grassland, acid grassland and fen, marsh and swamp habitats.	Land Cover Map (LCM2007) from the Centre for Ecology and Hydrology (CEH) (Morton <i>et al.</i> 2011).
% cover mountain, heath and bog	Heather, heather grassland, bog, montane habitats and inland rock.	LCM2007 from CEH (Morton <i>et al.</i> 2011).
% cover coastal	Includes supra-littoral rock, supra-littoral sediment, littoral rock, littoral sediment, saltmarsh.	LCM2007 from CEH (Morton <i>et al.</i> 2011).
% cover freshwater	Includes freshwater bodies.	LCM2007 from CEH (Morton <i>et al.</i> 2011).
% cover improved grassland	Includes improved grassland.	LCM2007 from CEH (Morton <i>et al.</i> 2011).
% cover arable	Includes arable and horticultural.	LCM2007 from CEH (Morton <i>et al.</i> 2011).
% cover urban	Includes urban and sub-urban.	LCM2007 from CEH (Morton <i>et al.</i> 2011).
Ratio of coniferous to deciduous woodland	$\text{Log}_{10}((\% \text{ coniferous woodland} + 0.01) / (\% \text{ broadleaved woodland} + 0.01))$ ; the 0.01 was added to all counts to avoid obtaining infinite values when the percentage cover equalled zero.	LCM2007 from CEH (Morton <i>et al.</i> 2011).
% cover trees	Raster of 20 m resolution giving the percentage tree cover per pixel. The data were re-projected from the European ETRS89 grid to the British National Grid and percent cover estimates were derived for each 1-km square. This dataset includes scattered trees and some hedgerows and treelines and was therefore preferred over the 2015LCM which only includes significant blocks of trees.	Copernicus Pan-European High Resolution Layer Tree Cover Density 2012 raster dataset (HRL; <a href="http://land.copernicus.eu/pan-european/high-resolution-layers">land.copernicus.eu/pan-european/high-resolution-layers</a> ).
island or mainland	A categorical term, two versions were trialled: a two-level categorical term of 'island' or 'mainland', and a seven-level categorical term distinguishing groups of islands (Northern Ireland, Orkney, Shetland, Isle of Wight, Lewis/Harris, Uists and Britain).	Manually derived from maps.
elevation	Mean elevation per 1-km square in meters above sea level .	GGIAR-SRTM 90 m raster (Jarvis <i>et al.</i> 2008).
year	1994 to 2016 as a numeric term.	BBS data (Harris <i>et al.</i> 2017).
easting and northing	The centroid of each 1-km survey square.	BBS data (Harris <i>et al.</i> 2017).

### **Spatial model structure**

We tried two alternative methods of modelling spatial and temporal variation in species relative abundance, a Generalised Additive Model (GAM) and a Geographically Weighted Regression (GWR).

### **Generalised Additive Model**

Massimino *et al.* 2015 used a GAM modelling approach. A GAM is a non-parametric extension of a general linear model, which can deal with non-linear relationships by using smooth terms (Guisan *et al.* 2002). The smooth terms in a GAM are additive which allows us to interpret it in a similar way to a traditional linear model. In the R package mgcv (Wood 2017) smooths are implemented using penalised regression splines. A GAM allows modelling of spatial and temporal variation through these smooth terms. Given what we understand about how bird abundance varies, and the types of trends we would like to estimate, we aimed to produce a model that included an 'easting-northing' smooth (allowing bird abundance to vary spatially), a smoothed 'year' term (allowing bird abundance to vary annually) and an interaction between the 'easting-northing' smooth and the 'year' smooth (allowing the relationship between bird abundance and year to vary spatially). Thin-plate regression splines were used to model easting and northing together in a 2D isotropic spatial smooth. An isotropic smooth was used as opposed to a tensor product smooth as easting and northing are on the same scale, i.e. a 1 unit change in easting is equivalent to a 1 unit change in northing as both are measured in metres. We tried various ways of modelling year, as a continuous linear term, as a categorical term and as a continuous variable with a thin plate regression spline. An interaction was fitted between the 2D spatial smooth and the 'year' term. For the models where the 'year' term was a smooth, this made a 3D smooth. As models used count data, we assumed Poisson distributed errors and a log link.

### **Geographically Weighted Regression**

A GWR incorporates spatial variability in a different way to a GAM. Effectively, GWR uses a spatial search window and fits a separate regression model to all the points in the search window each time (Fotheringham *et al.* 2003). Before the model is fitted, calibration is performed to find the appropriate width of this window, called the 'optimal bandwidth'. The bandwidth is defined either by actual geographical distance, 'a fixed kernel' or by the number of nearest neighbours to include, 'an adaptive kernel'. In this way a GWR model would produce a range of parameter estimates for each variable in the model. For example, for the

coefficient year, GWR would produce a range of parameter estimates for the relationship between year and bird count (essentially generating a local population trend estimate). However, a GWR would also allow the relationship between bird count and habitat to vary spatially too, if, for example, it was likely that Willow Warbler responded differently to woodland in Scotland and south-east England. This is also possible to do in a GAM using an interaction term between each habitat variable and the easting and northing smooth but this has several drawbacks: i) it would result in an incredibly complex and potentially over-fitted GAM; ii) the GAM would likely take a very long time to run, iii) at least some of the species would likely have convergence issues, preventing the model from being implemented in automated manner. Here, we tried generating a GWR model (using the spgwr package in R, Bivand & Danlin 2017) both including the habitat, island and topographical terms along with year, and just including the year coefficient. An adaptive kernel was used to account for variation in survey effort.

### **General model structure applied to both methods**

Analyses treated 1-km squares as independent data points. A slight modification was required in the uplands where certain squares are paired with an adjacent square to increase bird detections for these hard to reach locations. The total species counts from both squares were combined and included as a single data point for analysis purposes. To account for the extra survey effort in these upland paired squares, all models included an offset, with this being set to log(1) for the majority of squares and to log(2) for paired squares. This essentially converts all counts to count per 1-km square.

Initially, a weight to account for variation in survey effort between regions was included in the GAMs (this is not possible for GWRs), but it became apparent that this was absorbing local spatial variation that we wanted to include in the model; therefore the weight was removed. As the spatial smooth will account for variation in survey effort in these models it is unnecessary to have both a spatial smooth and a weight to account for survey effort.

Elevation was positively skewed and therefore was normalised by subtracting the minimum value and dividing by the range before inclusion in the models. For the GAMs we also tried fitting a smooth term for 'elevation'. This was because 'elevation' will vary over a larger spatial scale than the habitat variables and will likely represent variation in various other environmental

parameters, such as climate and even habitat to an extent, so we wanted to allow more flexibility in this relationship. The smoothing parameter,  $k$ , was set to three so as to only allow linear or quadratic relationships which were considered biologically plausible. Some of the habitat variables, such as woodland cover, may also have a non-linear relationship with bird abundance for some species, but we did not fit smooths for all these terms so as to avoid over-fitting and impractical model run times. For the GWR, as different regression models are fitted through space, the relationship between bird abundance and elevation or habitat is already very flexible.

### 2.3 EVALUATION OF MODEL FIT

We tested the predictive power of the final model fitted for each species using two methods: agreement with distribution data from the *Bird Atlas 2007–11* (Balmer *et al.* 2013) and cross-validation. For the former, occupancy at the 10-km square level from the *Bird Atlas 2007–11* was used as a coarse verification of the model predictions. Predictions for all 1-km squares in Britain were made with year set to 2010, then split into two groups: those where the parent 10-km square was occupied based on atlas data (squares with possible, probable or confirmed breeding evidence) and those where the 10-km square was unoccupied. A t-test was used to compare these two groups; if the predictions are valid we would expect the predictions for occupied squares to be significantly higher than the predictions for unoccupied squares.

Secondly, we conducted 10-fold cross validation. The final model was fitted using data from a randomly selected 90% sample of 1-km squares ('training data'). This model was then used to predict abundance based on the variable values from the remaining 10% of the 1-km squares ('test data'). This split was done at the square level to ensure independence between test and training data and between folds. The predicted abundances were then compared to the observed abundances from the test data using Spearman's rank correlation coefficient. This process was repeated 10 times using a different 10% of the data for testing each time and the mean correlation coefficient (and 95% confidence intervals) across the 10 replicates was calculated.

### 2.4 PRODUCING ESTIMATED YEARLY ABUNDANCE FROM THE MODELS

The final model for each species was then used to create 1-km square level predictions of the estimated relative abundance of each species each year using

the 'model predict' function and setting 'year' to the appropriate year. A recognised limitation is that we use static habitat data when making predictions as comparable land cover metrics are not available throughout the time series. Predictions then underwent a filtering process which aimed to remove values for which the predictions were likely to be inaccurate:

- Predictions for Shetland were removed due to low reliability as there are no BBS squares in Shetland to inform the predictions. Also, Shetland is on the edge of the spatial smooth so predictions here are sensitive to edge effects (Wood 2017).
- It was apparent that the models could not predict well for large lakes. Squares which are 100% water are not covered by the BBS which is a purely land-based survey, so we removed the predictions for 1-km squares which were 100% freshwater based on the 2007LCM.
- Poisson models cannot predict exact zeros on the response scale which makes it difficult to distinguish extremely low counts from absences. Therefore, for each species we also filtered the predictions using the *Bird Atlas 2007–11* (Balmer *et al.* 2013). The Bird Atlas data provided 10-km resolution presence-absence data for the whole of Britain and Ireland. We used this to find all unoccupied 10-km squares (where there was no evidence of breeding) For all 1-km squares falling in these unoccupied 10-km squares we calculated the median predicted abundance. Finally, any predictions in the UK equal to or lower than this value were set to zero.

These yearly relative abundance estimates were next mapped both in individual year maps for each species and in animated maps showing changes over the years in a GIF format, using the gganimate package in R (version 0.1.0.9000 Robinson 2016).

### 2.5 EXTRACTING TRENDS AND COMPARISON WITH PUBLISHED TRENDS

This project aimed to assess the feasibility of producing reliable local and regional trends, either to allow trends to be produced for bespoke areas, or to provide trends for species in countries or regions where this is currently impossible using the standard square  $\times$  year model. Therefore, we assessed how well estimated trends from our final models matched the published trends

for countries and regions. Though as alluded to earlier, we cannot know which of these modelling methods produces trends that match reality best. To compare our model results to the published BBS results we predicted abundances for the years used in the published trends: 1995, 2006, 2015 and 2016. Next we summed all the 1-km predictions for each region for each of the aforementioned years. The 1995 to 2016, 2006 to 2016 and 2015 to 2016 trends were calculated by  $\log_{10}((\text{latest year} + 0.0001)/(\text{earliest year} + 0.0001))$ . The log ratio of change was also calculated for the published trend estimates by back calculation  $\log_{10}((\text{trend}/100)+1)$ . A scatter plot was generated for each region with our modelled trend for each species plotted against the published BBS trend and a correlation coefficient was calculated. Additionally, we also generated a trend just for the county of Cambridgeshire from our models and compared this to a trend generated for the same region using the standard BBS GLMs of count  $\sim$  1-km square + year.

The previous part of the analysis checks whether there is overall agreement across species between the different methods of calculating trends, but it can conceal large differences in trend estimates for individual species. To consider this we tested whether the trend estimates derived above would have resulted in the same magnitude of change as that based on the standard method. We assigned each species to Green, Amber or Red status based on its 20-year trend between 1995 and 2015: Red list species were those which had declined by 40% or more during this period, Amber list species, those which had declined by between 20–40% and Green list species were those which had declined by less than 20%, remained stable or had increased. The assignation to Red, Amber or Green was done both for our modelled estimates and the published GLM trends and the results compared.

## 3. RESULTS

### 3.1 SELECTING THE FINAL MODEL STRUCTURE

Unfortunately, it was not possible to fit a GWR model. When all variables were included, the bandwidth fitting stage of the model for one species was still running after six days (on a 64 bit i3 processor with 16 GB of RAM) and showing no signs of finishing in the near future. Without significant parallel computing resources it would take over two years just to fit the bandwidth part of the model for each species. Even fitting the bandwidth for a model with just year changing in space, and so not including any of the other environmental

variables, the process still had not finished running after six days. Therefore, a GAM structure was used for all subsequent analyses. In these models, the smoothed 'year' term worked best compared to a categorical 'year' term and a linear 'year' term. The linear term was unable to represent non-linear relationships with time. A categorical term was too complex to allow the model to converge, having up to 23 levels.

The results for the GAMs with the seven-level categorical term for island and with the two-level categorical island term were very similar (Figure 1). However, the seven-level island term models performed marginally better on average both in terms of cross validation (mean correlation coefficients of  $0.359 \pm 0.015$ , as opposed to  $0.358 \pm 0.015$ ) and in terms of run time (a total of 91 hours 38 minutes for all 111 models as opposed to 120 hours 33 minutes) (Table A1). The difference in run-time appeared to be mainly due to Woodpigeon which took 10 hours with a two-level 'island' term and only 45 minutes with a seven-level 'island' term. However, as evident from Table A1 and Figure 1a, the quickest model varied depending on the species selected and surprisingly there was no relationship between run-time and the number of counts greater than zero in the input dataset. When comparing the plots of the GAM trend versus published GLM trend, the two versions (two-level versus seven-level island term) were visually identical, with the exception of Ring-necked Parakeet trends for the UK as a whole. But Ring-necked Parakeet was one of the few species for which it was not possible to create sensible models anyway (see limitations for more discussion of this). The animated maps were also visually identical for the two-level and seven-level 'island' term. Therefore, here we will only present the results from the models with the 7-level categorical 'island' term.

The final model included:

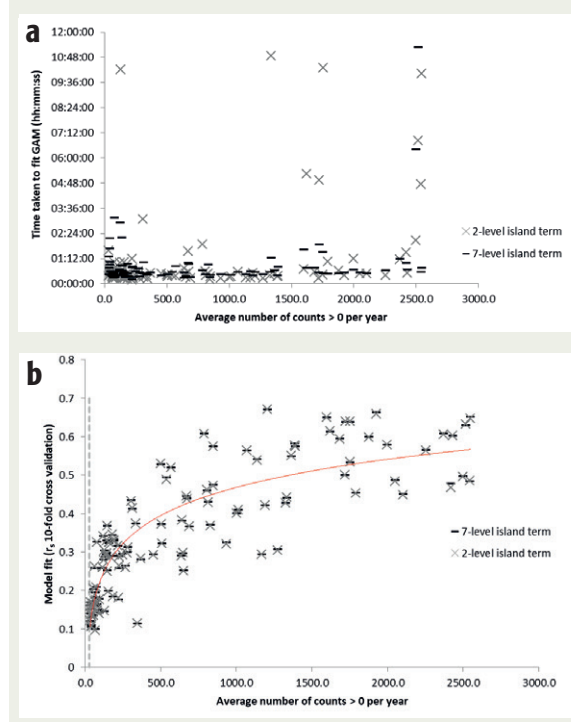
- linear effects for the habitat variables listed in Table 1;
- a seven-level categorical term for 'island';
- a smoothed term of 'elevation' where the smooth was restricted to a maximum k of 3;
- an isotropic 2D 'easting-northing' smooth;
- a 'year' smooth;
- an interaction between 'year' and the 'easting-northing' smooth.

The spatial and temporal smooths (easting, northing and year) in the models did not show over-fitting and therefore we did not specify  $k$ .  $k$  was selected internally within the program based on the data using Generalized Cross Validation; in this way the level of smoothing was allowed to vary between species as needed.

### 3.2 EVALUATING MODEL FIT

The GAMs performed moderately well according to both evaluation methods. The cross-validation correlation coefficients of observed versus predicted values ranged from 0.10 (for Little Grebe) to 0.67 (for Yellowhammer) (Table A1), but the majority of predictions were between 0.26 and 0.36. The cross-validation  $r_s$  value increased as the number of counts for the species (excluding zeros) increased, but followed a diminishing returns curve, with additional counts after 1,000 leading to little increase in model accuracy (Figure 1a).

**Figure 1. Plot of a) the time taken to fit each species' GAM and b) Spearman's rank correlation coefficient 10-fold cross validation results, versus the average number of counts above zero for that species per year (between 1994 to 2016). Results are shown for the GAM with a two-level island term (represented by a grey cross) and the GAM with a seven-level island term (represented with a black dash). The grey dashed line represents 30 counts, the cut-off point used to determine whether or not to publish a BBS trend for a species.**



For all 111 species, the model predictions were significantly higher for occupied 10-km squares (from *Bird Atlas 2007–11*) compared to unoccupied 10-km squares ( $P < 0.001$  for all 111 t-tests).

### 3.3 COMPARISON WITH PUBLISHED TRENDS

Table 2 displays the Spearman's rank correlation coefficients and the estimated slope of the regression line between the estimates from the GAMs and the GLM published trend estimates (slopes closer to 1 show higher agreement between the two modelling methods) and Figure 2 displays the corresponding correlation plots. It should be noted that these regional trends include species for which there are fewer than 30 records within a region, and therefore species for which the regional published GLM trends is considered unreliable. Goodness of GAM fit was not related to correlation to the published GLM trend: there was no relationship between the GAM cross-validation  $r_s$  value and the discrepancy between the GAM predicted trend and the published BBS trend (correlation  $< +0.02$ ).

For the long-term (1995 to 2016) and 10-year (2006 to 2016) trends, there is generally a strong correlation (although variable calibration) between the published trends and those from the GAMs (95–16 mean  $r_s = 0.869 \pm 0.05$  (SE), mean  $\beta = 0.64 \pm 0.1$ , 06–16 mean  $r_s = 0.81 \pm 0.02$ , mean  $\beta = 1.06 \pm 0.2$ ; Table 2, Figure 2). But the one-year trends (2015 to 2016) extracted from the maps were poorly correlated and poorly calibrated with published trends (15–16 mean  $r_s = 0.22 \pm 0.03$  (SE), mean  $\beta = 0.25 \pm 0.2$ ; Table 2, Figure 2).

The GAM method performed poorest for the UK trend, with the regression slope suggesting that the long-term trend was underestimated (95–16  $\beta = 0.04$ ; Table 2) and the 10-year trend over-estimated (06–16  $\beta = 3.60$ ; Table 2). This was due to the outlier value for Ring-necked Parakeet, without Ring-necked Parakeet 95–16  $\beta = 1.07$  and 06–16  $\beta = 1.08$ . This species had to be removed from the plots as trend estimates were so far outside the range of other species, making it difficult to see the general relationship (Ring-necked parakeet GAM estimated 95–16 trend = -8.46 compared to published GLM estimate = 1.20). This is due to the extreme predictions from the GAM for Ring-necked Parakeet, with total annual UK estimated relative abundance ranging from  $1.66 \times 10^4$  to  $1.20 \times 10^42$ , suggesting that despite the reasonable cross-validation value of 0.33, the GAM for this species is unreliable. Yorkshire and Humberside and Wales showed slightly weaker correlations between the published GLM trends and the trends estimated from the GAMs, as did the 1995 to

2016  $\beta$  estimate for the South East region, whereas the West Midlands showed one of the strongest agreements between the published trends and the GAM estimated trends (Table 2; Figure 2). The county trend for

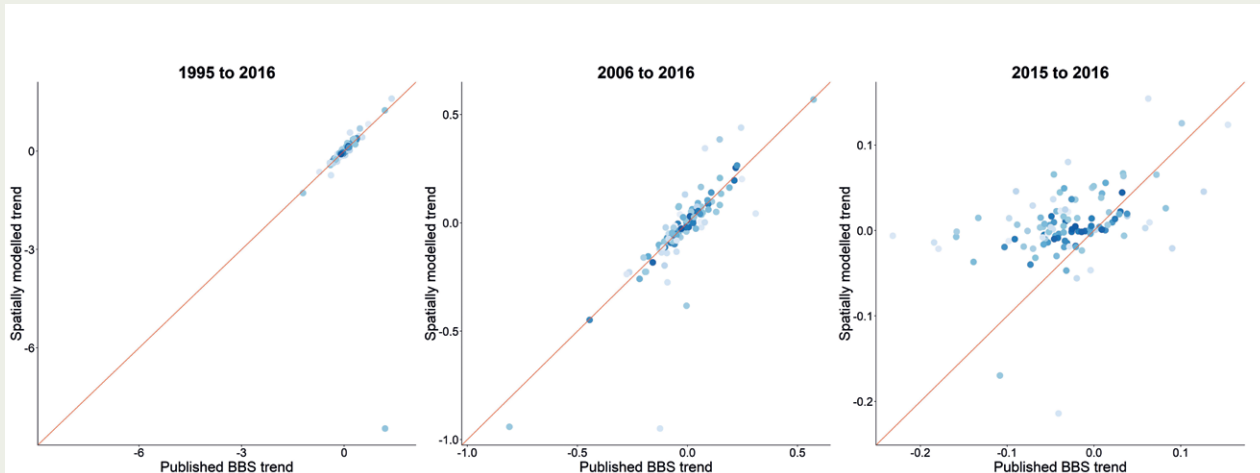
Cambridgeshire shows a weaker relationship between the GLM BBS trend and our spatially and temporarily varying trends (Figure 3;  $r_s=0.69$  and  $\beta=0.23$ ) than for the regional long-term trends, but is still generally better than the one-year regional trends.

**Table 2. Comparison of the species specific 20-year (1995 to 2016), 10-year (2006 to 2016) and 1-year (2015 to 2016) published BBS GLM trends with the trends estimated from predictions from the spatiotemporal GAMs. The correlations coefficient between the two ( $r_s$ ) is measured by Spearman's Rank Correlation Coefficient, and the slope ( $\beta$ ) from a linear regression of GAM estimated trend as a function of BBS GLM published trend.**

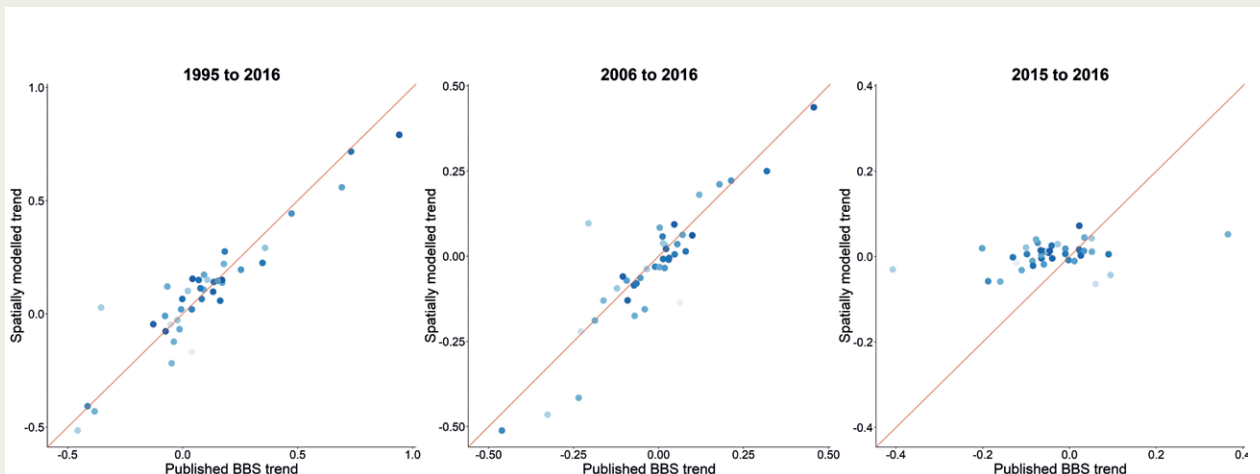
	1995 to 2016		2006 to 2016		2015 to 2016	
Region	$r_s$	$\beta$	$r_s$	$\beta$	$r_s$	$\beta$
UK	0.906	0.042	0.906	3.602	0.445	2.577
Scotland	0.872	0.958	0.770	0.866	0.180	0.064
Wales	0.842	0.170	0.621	0.791	0.199	0.044
Northern Ireland	0.850	0.813	0.796	0.876	0.108	0.066
North West	0.877	0.917	0.801	0.773	0.232	0.052
East England	0.923	0.844	0.834	1.113	0.196	0.108
North East	0.929	0.158	0.855	0.861	0.235	0.041
Yorkshire & Humberside	0.797	0.291	0.747	0.752	0.110	0.052
East Midlands	0.816	1.008	0.800	0.800	0.262	0.063
West Midlands	0.964	0.885	0.927	0.755	0.214	0.032
South East	0.902	0.542	0.901	0.990	0.258	0.008
South West	0.804	1.108	0.721	0.914	0.070	0.038
London	0.818	0.627	0.904	0.660	0.279	0.058

**Figure 2. Plots of species-specific published BBS GLM trend data versus the species-specific trend estimated from predictions from the spatiotemporal GAM. The darkness of the blue dots indicates the cross-validation value (Table A1), darker = higher model predictive ability. One species, Ring-necked Parakeet, had to be removed from the UK trend graph for the 10-year and 1-year plots as it fell well outside the range of the other predictions and therefore made the graph very difficult to read. In each plot the red line shows the 1:1 relationship expected if both trends agreed.**

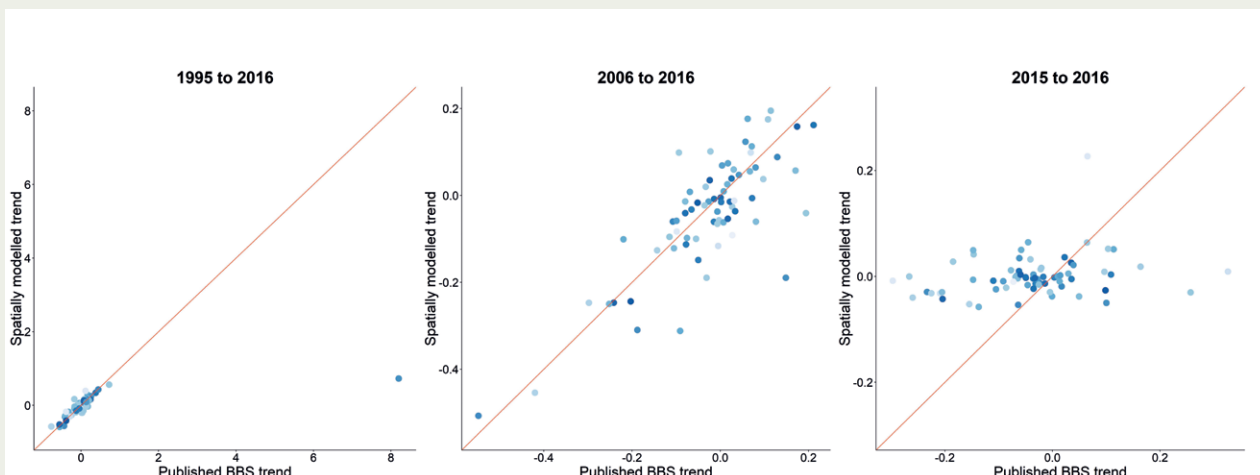
## United Kingdom



## Scotland

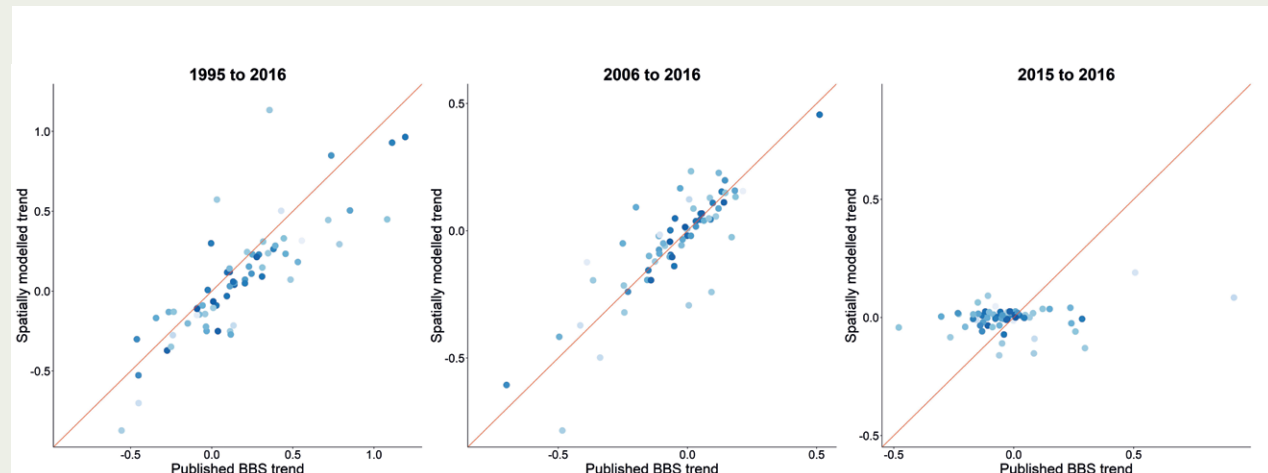


## Wales

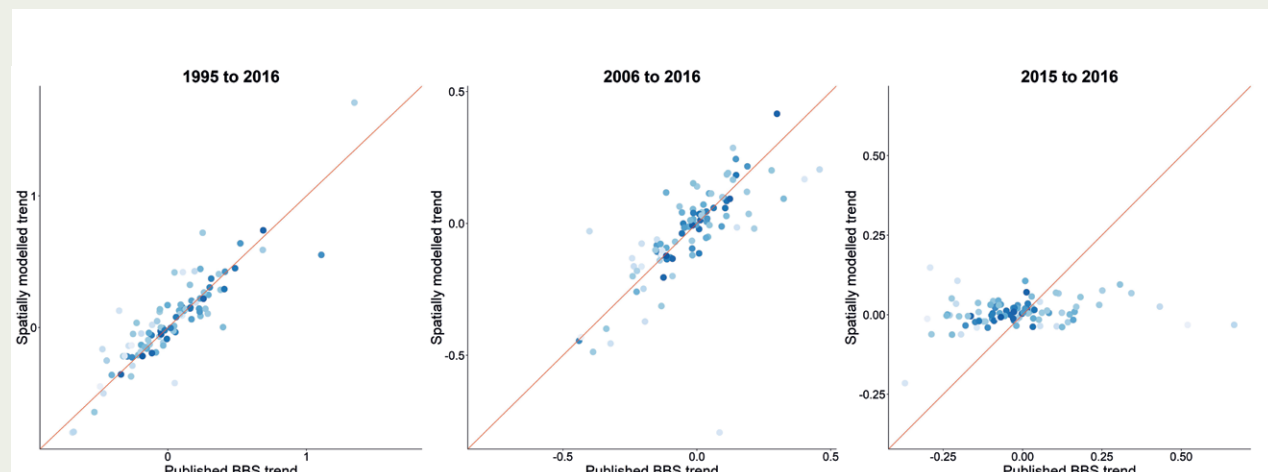


**Figure 2. (continued) Plots of species-specific published BBS GLM trend data versus the species-specific trend estimated from predictions from the spatiotemporal GAM. The darkness of the blue dots indicates the cross-validation value (Table A1), darker = higher model predictive ability. One species, Ring-necked Parakeet, had to be removed from the UK trend graph for the 10-year and 1-year plots as it fell well outside the range of the other predictions and therefore made the graph very difficult to read. In each plot the red line shows the 1:1 relationship expected if both trends agreed.**

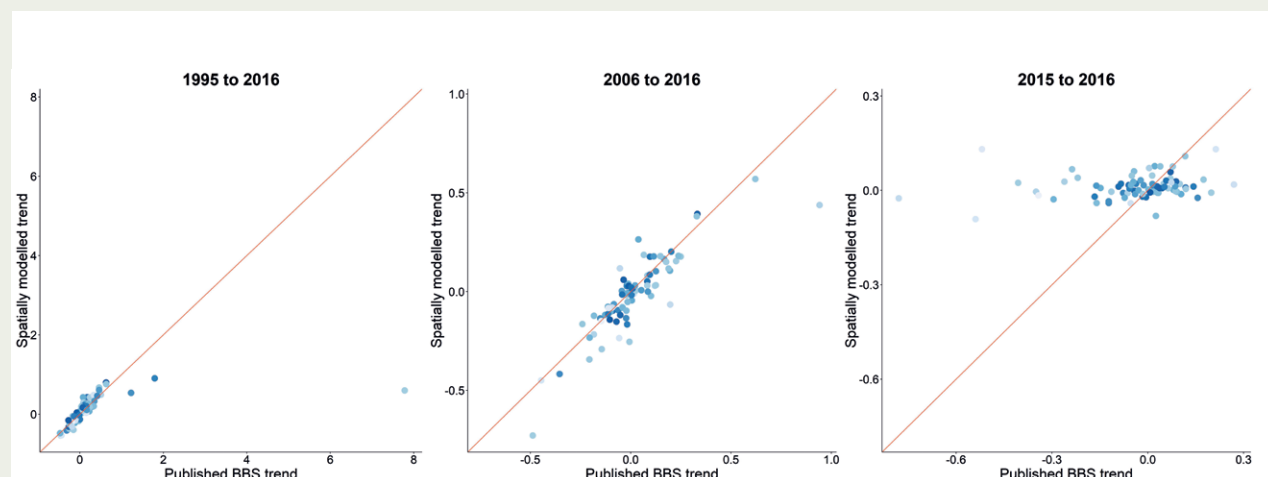
## Northern Ireland



## North West

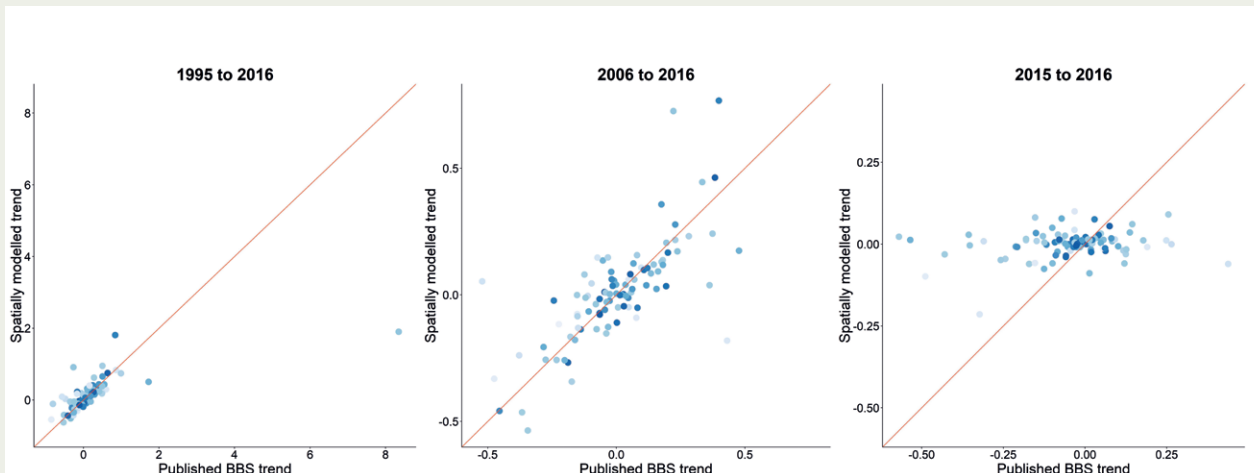


## North East

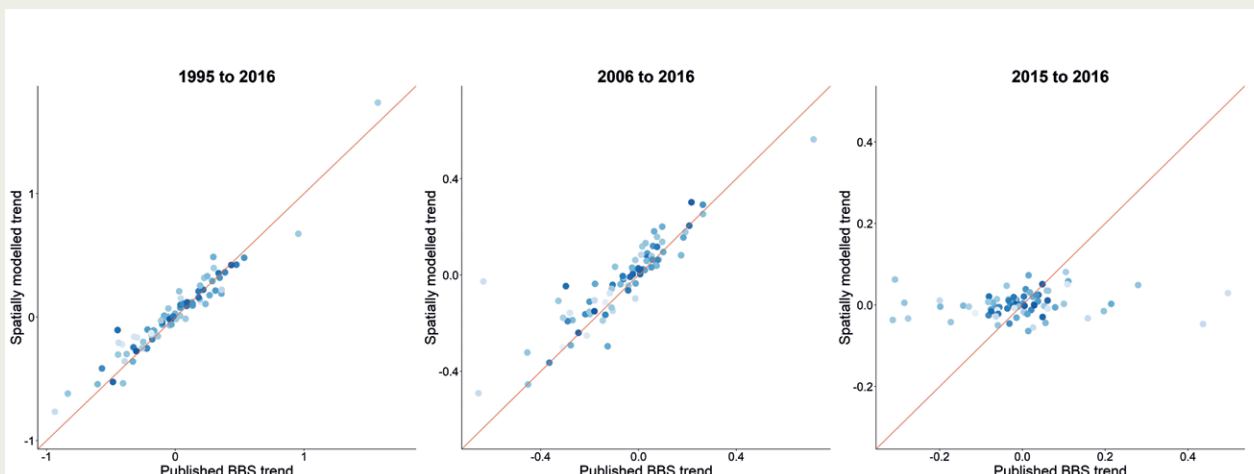


**Figure 2. (continued) Plots of species-specific published BBS GLM trend data versus the species-specific trend estimated from predictions from the spatiotemporal GAM. The darkness of the blue dots indicates the cross-validation value (Table A1), darker = higher model predictive ability. One species, Ring-necked Parakeet, had to be removed from the UK trend graph for the 10-year and 1-year plots as it fell well outside the range of the other predictions and therefore made the graph very difficult to read. In each plot the red line shows the 1:1 relationship expected if both trends agreed.**

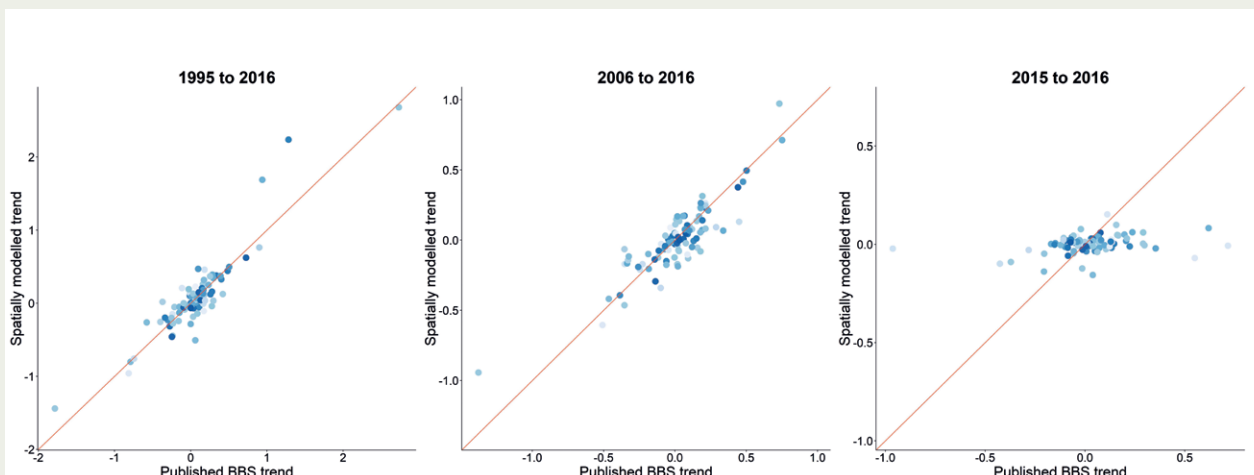
### Yorkshire & The Humber



### West Midlands

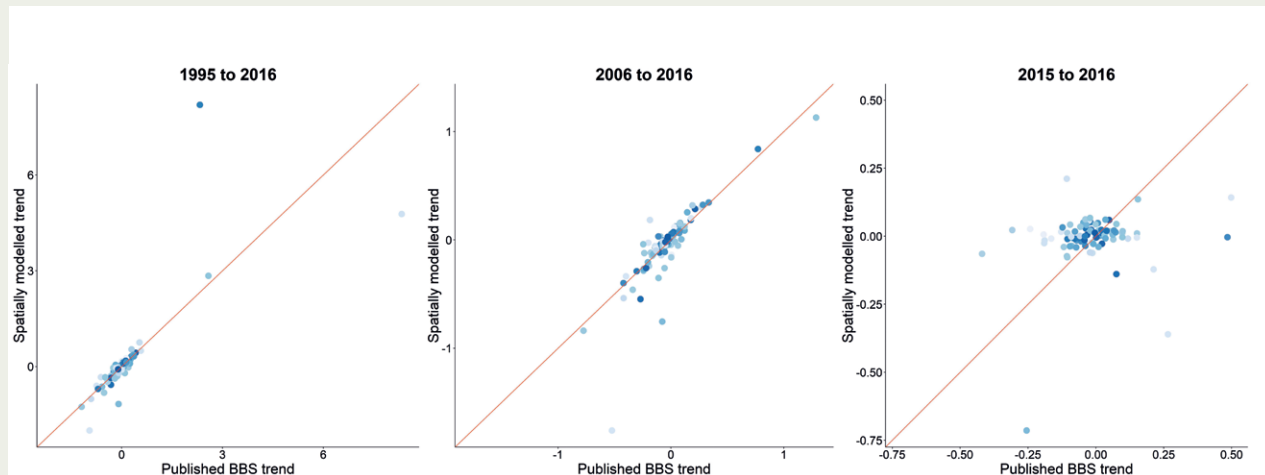


### East Midlands

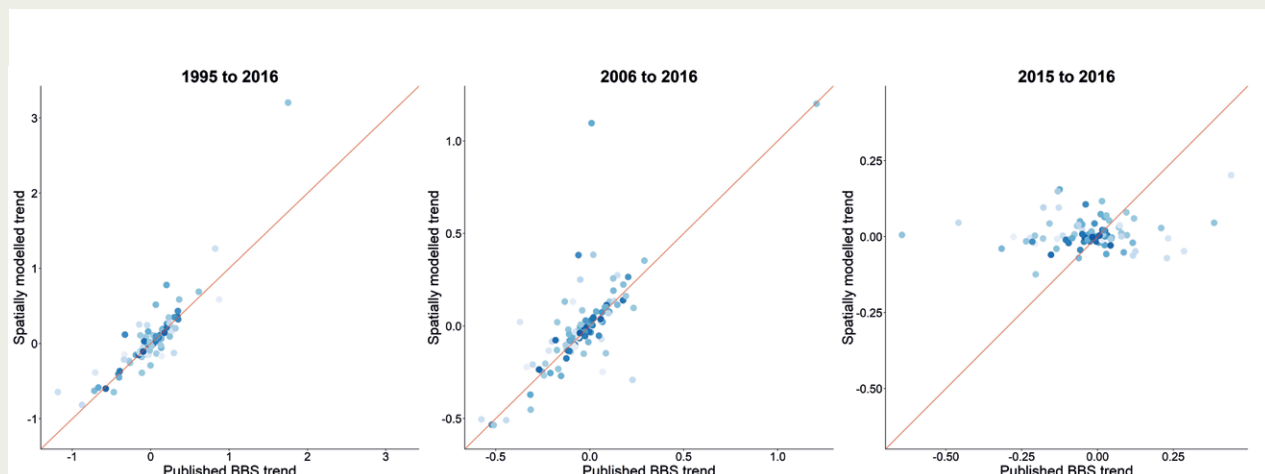


**Figure 2. (continued) Plots of species-specific published BBS GLM trend data versus the species-specific trend estimated from predictions from the spatiotemporal GAM. The darkness of the blue dots indicates the cross-validation value (Table A1), darker = higher model predictive ability. One species, Ring-necked Parakeet, had to be removed from the UK trend graph for the 10-year and 1-year plots as it fell well outside the range of the other predictions and therefore made the graph very difficult to read. In each plot the red line shows the 1:1 relationship expected if both trends agreed.**

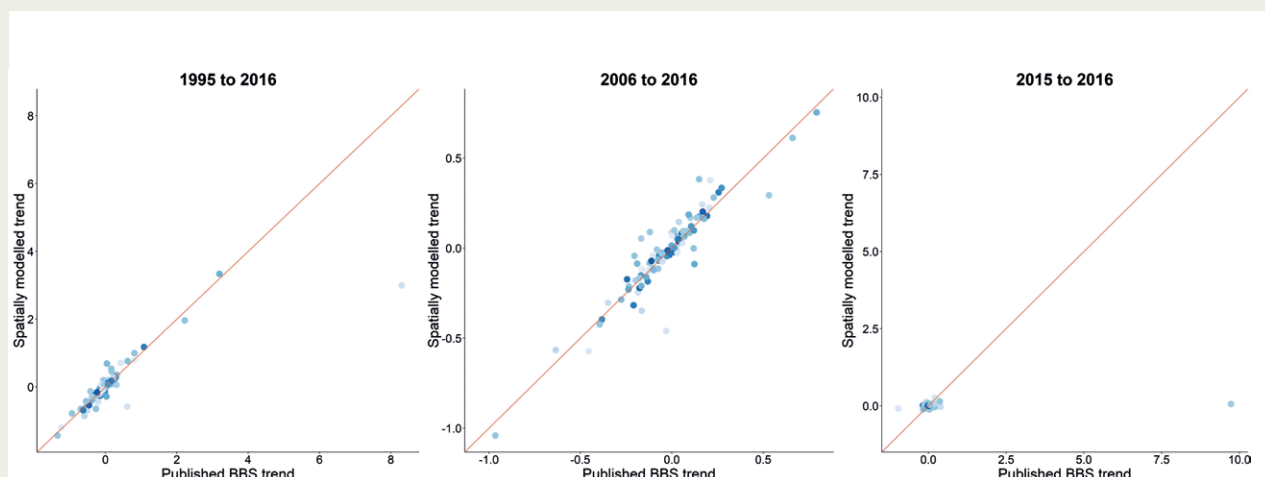
### East of England



### South West

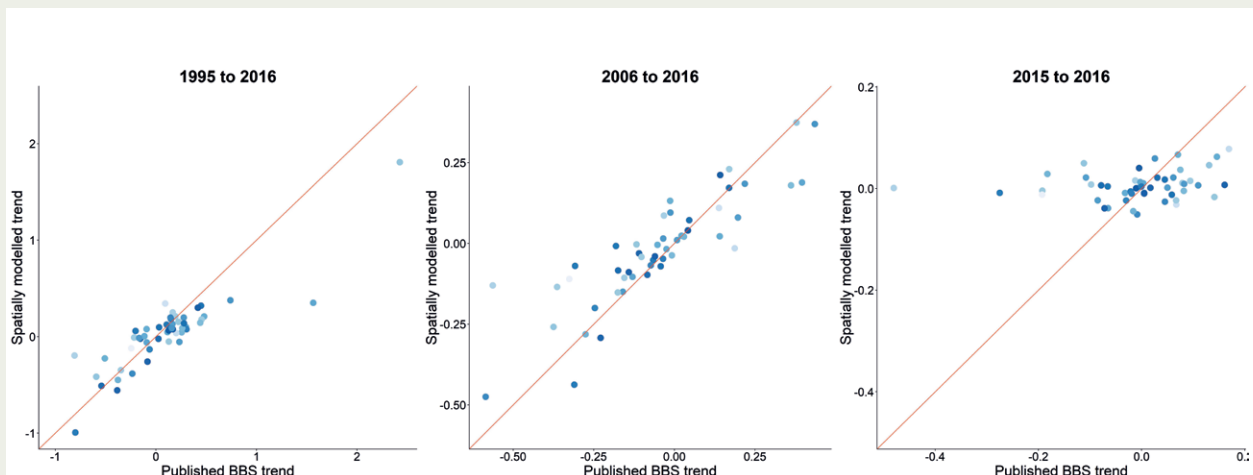


### South East

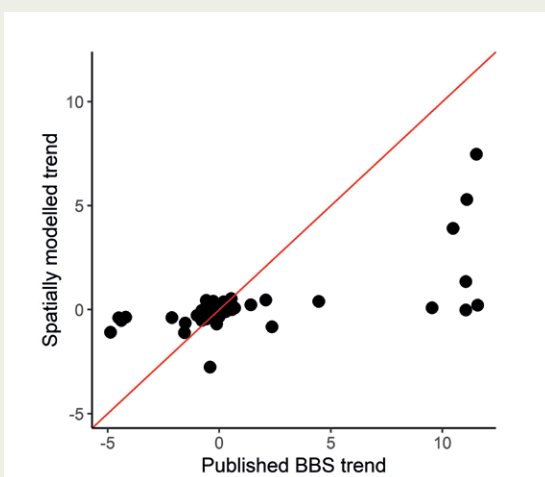


**Figure 2. (continued) Plots of species-specific published BBS GLM trend data versus the species-specific trend estimated from predictions from the spatiotemporal GAM. The darkness of the blue dots indicates the cross-validation value (Table A1), darker = higher model predictive ability. One species, Ring-necked Parakeet, had to be removed from the UK trend graph for the 10-year and 1-year plots as it fell well outside the range of the other predictions and therefore made the graph very difficult to read. In each plot the red line shows the 1:1 relationship expected if both trends agreed.**

## London



**Table 3. Species for which the status assessment (Red, Amber or Green) derived from the GAM estimated trend differed from the published GLM trend derived status assessment.**



was relatively small, <10% and just coincided with the category boundary. For the other seven species the difference between the published GLM and estimated GAM trends were more substantial (Table 3). On average the GAMs had a lower accuracy (as measured by 10-fold cross validation Table A1) for the seven species for which the difference between the published GLM and estimated GAM trend was larger ( $r_s$  range = 0.110–0.327 and mean = 0.187) compared to the seven species for which the difference was smaller ( $r_s$  range = 0.115–0.671 and mean = 0.343).

### 3.4 PREDICTED ABUNDANCE THROUGH TIME

The models were used to make annual maps of predict relative abundance over the whole of the UK which were compiled into animated GIF files for each species. Figure 4 shows an example of the annual maps for Buzzard illustrating how Buzzard abundance first increased in the west of Britain then spread across to populate the east of Britain.

## 4. DISCUSSION

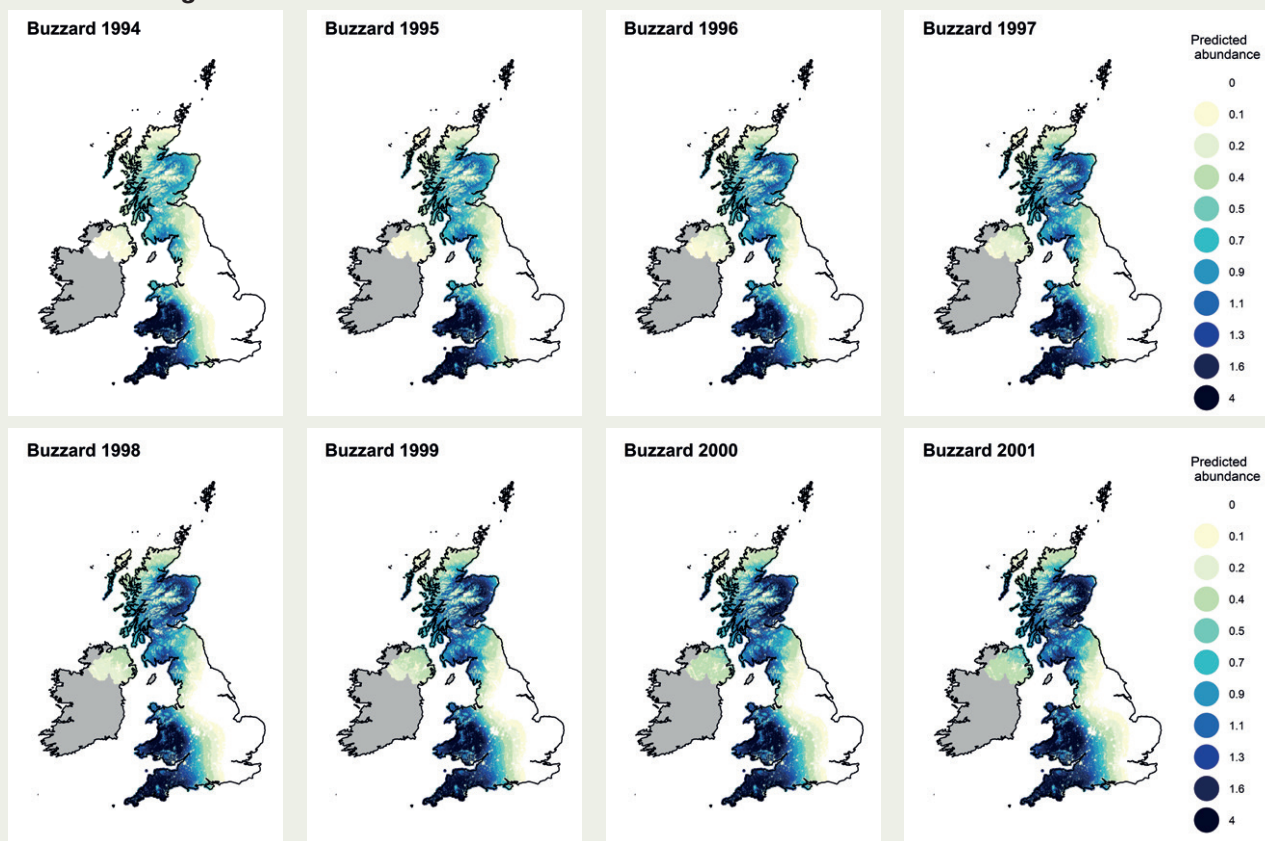
In this project we have produced reliable predictions of relative abundance for many breeding bird species using GAMs with habitat variables and spatial and temporal smooths. This opens up options for extracting localised trends for some species and for visualising changes in species' abundance and distribution through time.

The majority of species (97 out of 111) had the same status (Red, Amber or Green) from the published BBS GLM trends and the GAM-based estimates. Nine of the 14 exceptions (Table 3) were species for which the GAM trend yielded a higher threat status than calculated based on the published GLM. For seven (Corn Bunting, Feral Pigeon, Lapwing, Marsh Tit, Spotted Flycatcher, Sparrowhawk and Yellowhammer) the difference between the published GLM and estimated GAM trends

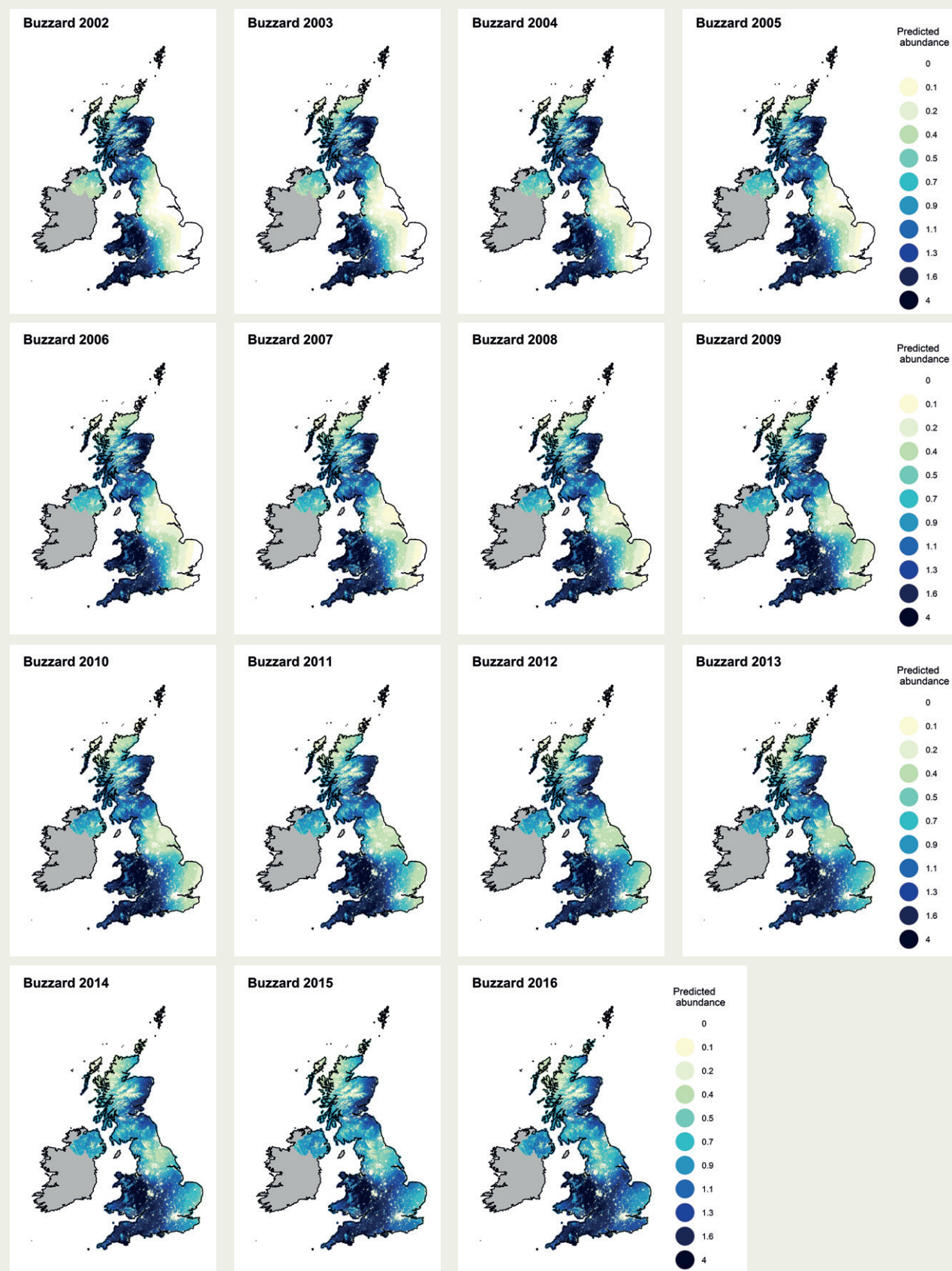
**Table 3. Species for which the status assessment (Red, Amber or Green) derived from the GAM estimated trend differed from the published GLM trend derived status assessment.**

Species	GLM estimated 1995–2015 trend	GLM assigned category	GAM estimated 1995–2015 trend	GAM assigned category
Ring-necked Parakeet	1455.0	Green	-100.0	Red
Common Crossbill	-1.7	Green	-41.0	Red
Grasshopper Warbler	-16.7	Green	-40.6	Red
Great Crested Grebe	9.9	Green	-28.2	Amber
Sedge Warbler	-9.5	Green	-25.0	Amber
Peregrine	-12.6	Green	-24.2	Amber
Spotted Flycatcher	-37.7	Amber	-46.6	Red
Corn Bunting	-33.5	Amber	-41.2	Red
Sparrowhawk	-16.0	Green	-23.3	Amber
Yellowhammer	-15.7	Green	-20.9	Amber
Marsh Tit	-41.2	Red	-38.0	Amber
Lapwing	-42.9	Red	-35.6	Amber
Feral Pigeon	-21.1	Amber	-10.8	Green
Great Black-backed Gull	-36.9	Amber	-9.2	Green

**Figure 4. Maps of the modelled Buzzard relative abundance each year between 1994 and 2016 from the GAM. The colour categories are in deciles.**



**Figure 4. (continued) Maps of the modelled Buzzard relative abundance each year between 1994 and 2016 from the GAM. The colour categories are in deciles.**



## 4.1 USAGE OF PREDICTIVE MODELS

The animated maps are an engaging, intuitive and quick way to show how species abundances have changed over time. They could help us to engage with a much larger audience than our current published BBS tables would, especially in this current age of social media where high volumes of information mean pictures are vital for catching initial interest. For example, tweets with an image lead to a 35% increase in re-tweets ([https://blog.twitter.com/official/en\\_us/a/2014/what-fuels-a-tweets-engagement.html](https://blog.twitter.com/official/en_us/a/2014/what-fuels-a-tweets-engagement.html)). The maps are also much more accessible to a non-science audience than tables of trend figures would be (Siricharoen 2013).

The maps are especially good at showing large-scale changes in distribution or abundance, for example the Buzzard recolonization of eastern England, the spread of Nuthatch northwards and the decline of Curlew and more recently Greenfinch. These examples are patterns we already know about from the BBS trends, Bird Atlases and other bird survey work. But the maps may also be useful to pick up trends that would not be uncovered in these other outputs, such as localised declines which affect too small an area to be picked up in the national and regional BBS trends, or are too recent to be uncovered in the 20-year change maps from *Bird Atlas 2007–11*. An argument can also be made that these GAMs may produce more reliable trend estimates than the site  $\times$  year GLMs for areas with low coverage (e.g. remote upland areas) where non-random coverage and low sample size may have synergistic negative effects as biased data are up-weighted in sparsely surveyed survey strata. The lower correlation between the GAM and GLM county trends for Cambridgeshire compared to the regional trends, may actually represent a better predictive ability for the GAM method in a smaller region. Though, without an independent trend data source we cannot test this theory. The inclusion of habitat variables and elevation in the GAMs means it is possible to infer abundances for areas with sparse data and the spatial smooths will also assist this process by interpolating local patterns of abundance based on squares with data. An example where this may be effective is for habitat-specific population declines, such as the recent declines in upland birds (Balmer *et al.* 2013) or woodland birds (Defra 2018). In the same way, this mapping method may also allow us to map trends in species which have too small a sample size to allow a published BBS trend, but probably only for species with very localised ranges, where the spatial smooths will strongly inform the distribution. It may be possible for conservation managers and local policymakers to use the maps to

uncover localised declines in species and thereby set appropriate conservation priorities, though as we do not have temporally varying habitat data this would be determined by the resolution of the smooth terms.

However, it is important to remember that we do not and cannot know what the actual abundances and trends in species are at a countrywide scale. Both the published BBS trends and the GAMs are predictions based on survey data. Ideally, small areas should be ground truthed with exhaustive surveys and these data compared to the two sets of estimates. However, in practise it is likely to prove difficult to do this over a long enough period of time, even for a small area. Validation of the patterns shown here with other independent survey datasets, such as local county-based atlases, single species surveys, the Breeding Waders of English Upland Farmland survey (<https://www.bto.org/volunteer-surveys/breeding-waders-english-upland-farmland>), or BTO Garden BirdWatch (<https://www.bto.org/volunteer-surveys/gbw>) would help to improve confidence in the GAM predictions.

It would be possible to run these models routinely, though it would be wise to filter the BBS data first to remove uninformative models. This could involve screening for extreme outlier values, and removing species with low sample sizes. A threshold of at least 30 occupied squares per year on average is already applied to determine whether BBS trends are published using the site  $\times$  year GLM method. It would not be necessary to run these GAMs every year, running them every five years would probably be sufficient to identify newly developing local trends and for public engagement purposes.

## 4.2 LIMITATIONS

The models in general yielded reasonably good correlations between observed and predicted values from cross-validation for count data, where the inherent stochasticity in the counts prevents really high correlation values (the highest correlation values from previous projects were around 0.68, Johnson *et al.* 2013, Newson *et al.* 2015, Border *et al.* 2017, Border *et al.* 2019). Long-term and mid-term trends estimated from the maps showed high correlation to the equivalent published BBS trends at national and regional scales. This gives us confidence in the reliability and usability of these predictive GAMs. However, the predicted 1-year trends did not match the published data well. This is likely to be due to the temporal smooth which, as its name suggests, smooths over differences from year to year. This means these models will be unable to reflect

sudden population crashes such as those observed for Long-tailed Tit (>27% reduction) and Wren (17% reduction) from 2017 to 2018 due to the severe winter event in February 2018 (the so-called 'Beast from the East' Harris *et al.* 2019). The spatial and temporal smooth may also affect predictions in coastal areas due to edge effects (Wood 2017). Therefore, we would not recommend relying purely on these predictions for coastal regions.

The few species where the GLM published and GAM estimated trends did not match for long-term and mid-term trends were species with small sample sizes and also often species which had undergone recent colonisation and expansion, such as Ring-necked Parakeet and Red Kite in eastern areas. But in these cases, both the published GLM BBS trends and the trends estimated from the GAM did not produce reliable population trends, giving large confidence intervals and some extreme and unrealistic predictions. This was especially true for Ring-necked Parakeet, with the BBS published long-term trend equalling +1480 (CI: 568–1746), and 1-km predicted values varying between zero and  $1.11 \times 1042$ . This suggests that with the currently available data we are unlikely to be able to achieve a good model for this species. There was also an issue with modelling the distribution of flocking water birds, such as geese. This was partly rectified by removing squares which were 100% freshwater, but, there was still a minority of extremely high counts in the raw data from flocks which led to some extreme predicted values. These issues with extreme counts highlight the more general problem of how to determine which species cannot be effectively modelled. Possibly we could impose a rule of thumb to remove species where predictions exceed an unrealistic magnitude, for example removing species with predicted values 10 times that of observed values. This rule would need to be tested and refined to ensure it worked suitably in practise with all species.

In terms of correctly assigning the threat status of a species, the GAMs did this correctly for the majority of species. For nine species the predicted conservation status threat level from the GAM was worse than the predicted threat level from the published GLM, this may mean extra resources are spent in more detailed assessments of these species' trends. More worrying though is the four species where the status threat level from the GAMs was better than their conservation status based on the published GLM trends. If the GAM models were relied upon in the absence of the GLM published trend this could delay the identification of the species'

true threat status and thereby delay conservation action. However, this is only true if the published GLM trend is more accurate than the GAM estimated trend, and we cannot measure this without the presence of independent survey data.

The models only use one set of habitat data for all years, assuming that habitat has not changed. This is unlikely to be the case over a 22-year time span, and if the same methods are employed in the yearly trend generation then as the time frame increases, it is likely the habitat data will become less relevant. With the current model design, only one set of habitat data can be used per 1-km square. However, it would be possible to update the habitat data for later period trends, the problem of habitat inaccuracies will only persist if the model is used to calculate predictions over a long period. CEH are currently deriving a 1990LCM which will be comparable with the more recent 2015LCM. With comparable datasets through time it would be possible to use the most relevant habitat data for each year. It also would be possible to produce separate models with different LCM datasets and then use these to produce weighted ensemble predictions for each year. In these the predictions from the model using the LCM product nearest in time to the year of interest would get the highest weighting and the models using the other LCM datasets would get progressively lower weightings to reflect the lower relevance of the habitat data. Added complexities such as this would further increase the computation burden.

At present there are no measures of uncertainty associated with the maps or with the trends estimated from them. It would be possible to run bootstraps on the model for each species to obtain confidence intervals around predictions but it would take a long time to run. Fitting a single model currently takes between 12 minutes and 11.5 hours, depending on the species. To do this 100 times for bootstraps would take from 20 hours to 47 days depending on the species. In total, the run-time for all species would be increased to c10,000 hours, and would likely increase further with each new year of data. Although we have access to parallel computing solutions which could reduce the duration over which the work is done, this is still a very significant total computational requirement and the associated carbon costs of running this many models should be considered.

We built the models using observed BBS counts and unlike Massimino *et al.* (2015) we did not attempt to adjust for imperfect detection using distance sampling.

This means the abundance patterns generated cannot be compared in absolute terms across species due to detectability differences. Future iterations of these models should be based on detectability corrected data so that models output predicted densities. This will make maps comparable across species, will allow maps to be summarised across species to assess community metrics, or to be summarised spatially, for example summated to produce population estimates.

### 4.3 CONCLUSIONS

In conclusion, the GAMs developed here are a useful way of visualising spatial and temporal variation in species abundance. They are relatively quick to run and provide accurate results for the majority of species. However, there are a few species – those with small sample sizes, recent colonisers, or flocking waterbirds – for which the GAMs perform poorly. It is also not possible to accurately model trends in Shetland. These issues may be resolved in future years as more data accumulate. In the future it would be useful to develop a quicker method for obtaining confidence intervals.

## 5. REFERENCES

- Bivand, R. and Danlin Yu, D. 2017. spgwr: Geographically Weighted Regression. R package version 0.6-32. <https://CRAN.R-project.org/package=spgwr>
- Border, J.A., Newson, S.E., White, D.C.J. and Gillings, S. 2017. Predicting the likely impact of urbanisation on bat populations using citizen science data, a case study for Norfolk, UK. *Landscape and Urban Planning* **162**: 44–55.
- Border, J. A., Gillings, S. & Newson, S. E. 2019. *Norfolk local bat conservation plan*. Report for Natural England.
- Buckland, S.T., Anderson, D.R., Burnham, K.P. & Laake, J.L., 2005. Distance sampling. *Encyclopedia of Biostatistics*, **2**.
- Department for Environment, Food and Rural Affairs. 2018. *Wild bird populations in the UK, 1970 to 2017*.
- Fotheringham, A. S., Brunsdon, C. & Charlton, M. 2003. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*, Wiley & Sons, New Jersey.
- Gibbons, D.W., Reid, J.B. & Chapman, R.A., 1993. *The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991*. T. & A.D. Poyser, London.
- Guisan, A., Edwards Jr, T. C. & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**: 89–100.
- Harrell, F. E. Jr., Lee, K. L. & Mark, D. B. 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, **15**: 361–87.
- Harris, S.J., Massimino, D., Eaton, M.A., Gillings, S., Noble, D.G., Balmer, D.E., Pearce-Higgins, J.W. & Woodcock, P. 2017. The Breeding Bird Survey 2016. *BTO Research Report* **700**. British Trust for Ornithology, Thetford.
- Harris, S.J., Massimino, D., Gillings, S., Eaton, M.A., Noble, D.G., Balmer, D.E., Procter, D. & Pearce-Higgins, J.W. 2019. The Breeding Bird Survey 2018. *BTO Research Report* **717**. British Trust for Ornithology, Thetford.
- Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. 2008. *Hole-filled SRTM for the globe* Version 4. – Available from the CGIAR-CSI SRTM 90 m Database: <http://srtm.csi.cgiar.org>.
- Johnston, A., Ausden, M., Dodd, A. M., Bradbury, R. B., Chamberlain, D. E., Jiguet, F., Thomas, C. D., Cook, A.S.C.P., Newson, S. E., Ockendon, N., Rehfisch, M. M., Roos, S., Thaxter, C. B., Brown, A., Crick, H.Q.P., Douse, A., Mccall, R. A., Pontier, H., Stroud, D. A., Cadiou, B., Crowe, O., Deceuninck, B., Hornman, M. & Pearce-Higgins, J. 2013. Observed and Predicted Effects of Climate Change on Species Abundance in Protected Areas. *Nature Climate Change*, **3**: 1,055–1,061.
- Massimino, D., Johnson, A., Noble, D. & Pearce-Higgins, J. 2015. Multi-species spatially explicit indicators reveal spatially structured trends in bird communities. *Ecological Indicators*, **58**: 277–285.
- Morrison, C. A., Robinson, R. A., Clark, J. A. & Gill, J. A. 2010. Spatial and temporal variation in population trends in a long-distance migratory bird. *Diversity and Distributions*, **16**: 620–627.
- Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R. & Simpson, I.C. 2011. Final Report for LCM2007 - the new UK Land Cover Map. *Countryside Survey Technical Report* No. **11/07** NERC/Centre for Ecology & Hydrology 112pp. (CEH Project Number: C03259).
- Newson, S. E., Evans, K. L., Noble, D. G., Greenwood, J. J., & Gaston, K. J. 2008. Use of distance sampling to improve estimates of national population sizes for common and widespread breeding birds in the UK. *Journal of Applied Ecology*, **45**: 1330–1338.
- Newson, S. E., Ockendon, N., Joys, A., Noble, D. G. & Baillie, S. R. 2009. Comparison of habitat specific trends in the abundance of breeding birds in the UK. *Bird Study*, **56**: 233–243.
- Newson, S.E., Evans, H.E. & Gillings, S. 2015. A novel citizen approach for large-scale standardised monitoring of bat activity and distribution, evaluated in eastern England. *Biological Conservation*, **191**: 38–49.
- Robinson, D. 2016. *gganimate: Create easy animations with ggplot2*. R package version 0.1.0.9000. <http://github.com/dgrtwo/gganimate>
- Robinson, R. A., Lawson, B., Toms, M. P., Peck, K. M., Kirkwood, J. K., Chantrey, J. & John, S. K. 2010. Emerging infectious disease leads to rapid population declines of common British birds. *PLoS One*, **5**, e12215.
- Sharrock, J. T. R. 1976. *The Atlas of Breeding Birds in Britain and Ireland*. T & A. D. Poyser, Calton.
- Siricharoen, W.V. 2013. Infographics: the new communication tools in digital age. In *The international conference on e-technologies and business on the web* (ebw2013), pp. 169–174.
- Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC.

## APPENDIX

**Table A1. Cross validation results for each model structure. The correlation coefficient is the Spearmans' correlation coefficient between observed and predicted values.**

Species (scientific name)	Average number of yearly counts > 0	Model with two-level island term		Model with seven-level island term	
		Correlation coefficient	Fitting time	Correlation coefficient	Fitting time
Blackbird ( <i>Turdus merula</i> )	2515.7	0.633 (0.618–0.647)	06:50:00	0.630 (0.608–0.652)	11:17:20
Blackcap ( <i>Sylvia atricapilla</i> )	1685.1	0.594 (0.584–0.605)	00:33:29	0.595 (0.578–0.611)	00:44:58
Bullfinch ( <i>Pyrrhula pyrrhula</i> )	637.1	0.290 (0.283–0.296)	00:43:09	0.291 (0.281–0.301)	00:32:02
Barn Owl ( <i>Tyto alba</i> )	47.1	0.161 (0.148–0.173)	00:31:14	0.162 (0.152–0.172)	00:35:43
Blue Tit ( <i>Cyanistes caeruleus</i> )	2368.9	0.606 (0.586–0.626)	01:10:12	0.608 (0.595–0.621)	01:10:55
Buzzard ( <i>Buteo buteo</i> )	1070.2	0.563 (0.557–0.57)	00:33:00	0.564 (0.551–0.577)	00:33:09
Carrion Crow ( <i>Corvus corone</i> )	2420.7	0.468 (0.439–0.496)	01:30:00	0.478 (0.463–0.492)	00:59:03
Cormorant ( <i>Phalacrocorax carbo</i> )	225.5	0.288 (0.268–0.308)	00:43:00	0.289 (0.278–0.300)	00:48:08
Corn Bunting ( <i>Emberiza calandra</i> )	140.8	0.294 (0.276–0.311)	00:32:00	0.294 (0.272–0.315)	00:36:59
Chiffchaff ( <i>Phylloscopus collybita</i> )	1598.5	0.651 (0.639–0.662)	00:41:00	0.650 (0.637–0.663)	01:37:31
Collared Dove ( <i>Streptopelia decaocto</i> )	1363.6	0.551 (0.540–0.563)	00:28:00	0.550 (0.529–0.570)	00:38:37
Canada Goose ( <i>Branta canadensis</i> )	507.9	0.323 (0.301–0.345)	00:25:00	0.322 (0.299–0.344)	00:26:38
Chaffinch ( <i>Fringilla coelebs</i> )	2542.0	0.485 (0.456–0.514)	04:45:00	0.484 (0.460–0.508)	00:33:52
Cuckoo ( <i>Cuculus canorus</i> )	690.9	0.367 (0.359–0.376)	00:16:00	0.367 (0.350–0.384)	00:19:12
Coot ( <i>Fulica atra</i> )	268.8	0.260 (0.246–0.274)	00:36:00	0.265 (0.247–0.283)	00:43:33
Crossbill ( <i>Loxia curvirostra</i> )	57.9	0.205 (0.190–0.219)	00:20:00	0.204 (0.195–0.214)	00:24:58
Common Sandpiper ( <i>Actitis hypoleucos</i> )	66.2	0.190 (0.164–0.217)	00:16:00	0.194 (0.175–0.214)	00:20:22
Coal Tit ( <i>Periparus ater</i> )	849.2	0.472 (0.449–0.494)	00:20:00	0.475 (0.461–0.489)	00:24:44
Curlew ( <i>Numenius arquata</i> )	500.6	0.531 (0.520–0.542)	00:26:00	0.529 (0.512–0.547)	00:26:42
Dunnock ( <i>Prunella modularis</i> )	2102.5	0.448 (0.431–0.466)	00:28:00	0.451 (0.429–0.474)	00:30:22
Dipper ( <i>Cinclus cinclus</i> )	63.5	0.172 (0.154–0.189)	00:21:00	0.172 (0.156–0.187)	00:30:24
Little Egret ( <i>Egretta garzetta</i> )	36.0	0.159 (0.145–0.173)	00:42:00	0.162 (0.147–0.176)	00:45:33
Feral Pigeon ( <i>Columba livia</i> )	675.3	0.438 (0.425–0.452)	00:34:00	0.439 (0.418–0.459)	00:55:50
Green Woodpecker ( <i>Picus viridis</i> )	846.1	0.575 (0.558–0.592)	00:13:00	0.575 (0.562–0.589)	00:18:40
Gadwall ( <i>Mareca strepera</i> )	39.1	0.138 (0.118–0.157)	00:38:00	0.139 (0.116–0.161)	02:09:55
Great Black-backed Gull ( <i>Larus marinus</i> )	83.8	0.149 (0.131–0.166)	00:22:00	0.150 (0.131–0.168)	00:23:18
Goldcrest ( <i>Regulus regulus</i> )	813.0	0.429 (0.410–0.449)	00:24:00	0.431 (0.412–0.450)	00:27:06
Goosander ( <i>Mergus merganser</i> )	42.4	0.136 (0.113–0.160)	00:26:00	0.139 (0.120–0.157)	01:41:28
Great Crested Grebe ( <i>Podiceps cristatus</i> )	71.4	0.175 (0.157–0.192)	00:19:00	0.183 (0.164–0.202)	00:26:59
Grasshopper Warbler ( <i>Locustella naevia</i> )	80.2	0.167 (0.151–0.183)	00:18:00	0.165 (0.150–0.180)	00:18:51
Greylag Goose ( <i>Anser anser</i> )	228.0	0.265 (0.249–0.280)	00:16:00	0.258 (0.250–0.267)	00:18:53
Grey Wagtail ( <i>Motacilla cinerea</i> )	221.5	0.181 (0.169–0.193)	00:11:00	0.177 (0.158–0.197)	00:12:55
Goldfinch ( <i>Carduelis carduelis</i> )	1717.7	0.500 (0.485–0.514)	00:15:14	0.500 (0.486–0.513)	00:32:07
Golden Plover ( <i>Pluvialis apricaria</i> )	64.1	0.263 (0.249–0.278)	00:23:19	0.259 (0.235–0.283)	00:22:42
Greenfinch ( <i>Chloris chloris</i> )	1753.1	0.534 (0.521–0.547)	10:17:14	0.533 (0.520–0.546)	00:29:53

		Model with two-level island term		Model with seven-level island term	
Species (scientific name)	Average number of yearly counts > 0	Correlation coefficient	Fitting time	Correlation coefficient	Fitting time
Great Spotted Woodpecker ( <i>Dendrocopos major</i> )	1136.2	0.539 (0.526–0.552)	00:16:50	0.542 (0.528–0.557)	00:27:38
Great Tit ( <i>Parus major</i> )	2254.7	0.564 (0.550–0.579)	00:23:18	0.566 (0.548–0.583)	00:38:12
Garden Warbler ( <i>Sylvia borin</i> )	452.5	0.293 (0.274–0.312)	00:28:00	0.294 (0.278–0.311)	00:24:26
Grey Heron ( <i>Ardea cinerea</i> )	650.1	0.298 (0.288–0.308)	00:29:00	0.297 (0.280–0.314)	00:22:17
Hooded Crow ( <i>Corvus cornix</i> )	119.5	0.342 (0.325–0.358)	00:21:00	0.342 (0.323–0.362)	00:36:40
House Martin ( <i>Delichon urbicum</i> )	934.1	0.321 (0.302–0.340)	00:16:00	0.325 (0.308–0.342)	00:27:42
House Sparrow ( <i>Passer domesticus</i> )	1621.0	0.615 (0.601–0.628)	05:14:00	0.613 (0.596–0.630)	00:45:30
Hobby ( <i>Falco subbuteo</i> )	43.3	0.106 (0.097–0.115)	00:52:42	0.106 (0.101–0.112)	01:02:18
Jay ( <i>Garrulus glandarius</i> )	807.1	0.462 (0.452–0.472)	00:36:00	0.461 (0.453–0.470)	00:26:21
Jackdaw ( <i>Coloeus monedula</i> )	1790.8	0.453 (0.433–0.473)	01:03:00	0.455 (0.442–0.468)	00:28:40
Kestrel ( <i>Falco tinnunculus</i> )	652.1	0.251 (0.237–0.264)	00:19:00	0.252 (0.243–0.261)	00:18:15
Kingfisher ( <i>Alcedo atthis</i> )	56.0	0.136 (0.123–0.148)	00:30:00	0.135 (0.121–0.149)	00:23:48
Red Kite ( <i>Milvus milvus</i> )	143.7	0.329 (0.318–0.34)	00:37:00	0.329 (0.315–0.342)	00:52:06
Lapwing ( <i>Vanellus vanellus</i> )	671.5	0.445 (0.435–0.455)	01:33:00	0.446 (0.428–0.463)	00:59:52
Little Grebe ( <i>Tachybaptus ruficollis</i> )	68.0	0.097 (0.077–0.117)	00:17:00	0.100 (0.093–0.106)	00:16:39
Linnet ( <i>Linaria cannabina</i> )	1190.4	0.422 (0.407–0.438)	00:31:00	0.421 (0.407–0.434)	00:24:14
Little Owl ( <i>Athene noctua</i> )	92.8	0.177 (0.170–0.184)	00:48:00	0.179 (0.172–0.187)	00:26:01
Lesser Redpoll ( <i>Acanthis cabaret</i> )	168.5	0.286 (0.261–0.312)	00:22:00	0.290 (0.276–0.303)	00:21:46
Long-tailed Tit ( <i>Aegithalos caudatus</i> )	1001.9	0.401 (0.390–0.411)	00:19:00	0.401 (0.392–0.409)	00:24:58
Lesser Whitethroat ( <i>Sylvia curruca</i> )	279.4	0.297 (0.287–0.307)	00:15:00	0.300 (0.289–0.310)	00:20:30
Mistle Thrush ( <i>Turdus viscivorus</i> )	1169.5	0.294 (0.279–0.309)	00:17:00	0.294 (0.276–0.312)	00:34:50
Mallard ( <i>Anas platyrhynchos</i> )	1327.5	0.428 (0.410–0.446)	00:24:00	0.428 (0.410–0.445)	00:33:29
Magpie ( <i>Pica pica</i> )	1926.4	0.659 (0.639–0.679)	00:24:00	0.663 (0.652–0.675)	00:31:07
Moorhen ( <i>Gallinula chloropus</i> )	639.5	0.383 (0.367–0.400)	00:24:00	0.382 (0.363–0.402)	00:17:42
Mandarin duck ( <i>Aix galericulata</i> )	32.5	0.122 (0.110–0.134)	00:22:00	0.121 (0.105–0.137)	00:26:33
Meadow Pipit ( <i>Anthus pratensis</i> )	786.5	0.607 (0.584–0.630)	01:52:00	0.609 (0.595–0.624)	00:37:59
Mute Swan ( <i>Cygnus olor</i> )	248.9	0.293 (0.27–0.316)	00:23:00	0.296 (0.283–0.309)	00:34:35
Marsh Tit ( <i>Poecile palustris</i> )	147.1	0.250 (0.234–0.266)	00:15:00	0.253 (0.235–0.271)	00:24:44
Nightingale ( <i>Luscinia megarhynchos</i> )	32.4	0.150 (0.132–0.167)	01:34:00	0.151 (0.140–0.162)	01:17:41
Nuthatch ( <i>Sitta europaea</i> )	540.5	0.488 (0.472–0.504)	00:23:00	0.494 (0.477–0.511)	00:32:48
Oystercatcher ( <i>Haematopus ostralegus</i> )	307.1	0.431 (0.414–0.449)	03:05:00	0.435 (0.419–0.450)	00:30:11
Grey Partridge ( <i>Perdix perdix</i> )	219.5	0.316 (0.305–0.327)	01:12:00	0.316 (0.305–0.328)	00:45:54
Peregrine ( <i>Falco peregrinus</i> )	41.8	0.109 (0.095–0.123)	00:26:00	0.110 (0.098–0.122)	00:26:04
Pied Flycatcher ( <i>Ficedula hypoleuca</i> )	39.5	0.170 (0.148–0.192)	00:53:00	0.170 (0.157–0.184)	00:55:06
Pheasant ( <i>Phasianus colchicus</i> )	1877.5	0.599 (0.581–0.616)	00:37:00	0.600 (0.583–0.617)	00:30:04
Pied Wagtail ( <i>Motacilla alba</i> )	1273.9	0.304 (0.283–0.326)	00:15:00	0.308 (0.290–0.326)	00:27:38

		Model with two-level island term		Model with seven-level island term	
Species (scientific name)	Average number of yearly counts > 0	Correlation coefficient	Fitting time	Correlation coefficient	Fitting time
Robin ( <i>Erithacus rubecula</i> )	2431.8	0.602 (0.582–0.622)	00:30:00	0.602 (0.590–0.615)	00:39:50
Reed Bunting ( <i>Emberiza schoeniclus</i> )	504.4	0.372 (0.362–0.382)	00:20:00	0.372 (0.362–0.381)	00:26:57
Red Grouse ( <i>Lagopus lagopus</i> )	147.0	0.369 (0.352–0.385)	00:40:00	0.368 (0.351–0.386)	00:35:56
Ring-necked Parakeet ( <i>Psittacula krameri</i> )	77.9	0.325 (0.304–0.345)	01:02:00	0.327 (0.312–0.343)	03:09:37
Redshank ( <i>Tringa totanus</i> )	71.1	0.162 (0.143–0.181)	00:15:00	0.161 (0.137–0.186)	00:18:23
Red-legged Partridge ( <i>Alectoris rufa</i> )	570.4	0.519 (0.506–0.533)	00:21:22	0.519 (0.508–0.529)	00:49:17
Raven ( <i>Corvus corax</i> )	311.5	0.414 (0.401–0.426)	00:18:13	0.413 (0.401–0.425)	01:00:47
Rook ( <i>Corvus frugilegus</i> )	1334.6	0.442 (0.422–0.462)	10:52:22	0.441 (0.426–0.457)	01:15:18
Redstart ( <i>Phoenicurus phoenicurus</i> )	181.9	0.347 (0.330–0.363)	00:22:43	0.345 (0.326–0.363)	00:56:26
Reed Warbler ( <i>Acrocephalus scirpaceus</i> )	123.5	0.280 (0.256–0.303)	00:59:16	0.281 (0.267–0.295)	02:54:16
Skylark ( <i>Alauda arvensis</i> )	1752.0	0.640 (0.621–0.660)	00:26:00	0.639 (0.628–0.651)	01:31:10
Stonechat ( <i>Saxicola rubicola</i> )	136.7	0.291 (0.276–0.306)	00:27:00	0.288 (0.263–0.314)	02:12:39
Stock Dove ( <i>Columba oenas</i> )	830.3	0.369 (0.360–0.378)	00:29:00	0.371 (0.356–0.386)	00:55:31
Spotted Flycatcher ( <i>Muscicapa striata</i> )	185.4	0.184 (0.173–0.194)	00:14:00	0.184 (0.172–0.195)	00:30:26
Starling ( <i>Sturnus vulgaris</i> )	1721.3	0.639 (0.626–0.653)	04:56:00	0.639 (0.630–0.649)	01:52:12
Sparrowhawk ( <i>Accipiter nisus</i> )	346.3	0.115 (0.104–0.125)	00:14:00	0.115 (0.108–0.122)	00:25:50
Swift ( <i>Apus apus</i> )	1014.3	0.410 (0.399–0.421)	00:21:00	0.410 (0.392–0.427)	00:24:40
Siskin ( <i>Spinus spinus</i> )	192.5	0.332 (0.316–0.348)	00:19:00	0.332 (0.315–0.349)	00:22:27
Swallow ( <i>Hirundo rustica</i> )	1996.6	0.579 (0.564–0.594)	01:12:00	0.579 (0.565–0.592)	00:39:25
Sand Martin ( <i>Riparia riparia</i> )	127.5	0.147 (0.135–0.160)	00:19:00	0.145 (0.128–0.162)	00:23:15
Snipe ( <i>Gallinago gallinago</i> )	163.4	0.325 (0.317–0.334)	00:26:00	0.328 (0.311–0.345)	00:18:43
Song Thrush ( <i>Turdus philomelos</i> )	2049.9	0.484 (0.463–0.504)	00:28:00	0.487 (0.469–0.505)	00:31:33
Shelduck ( <i>Tadorna tadorna</i> )	118.2	0.259 (0.241–0.278)	00:34:00	0.259 (0.237–0.281)	00:39:45
Sedge Warbler ( <i>Acrocephalus schoenobaenus</i> )	281.1	0.312 (0.286–0.338)	00:29:41	0.313 (0.297–0.330)	00:31:20
Treecreeper ( <i>Certhia familiaris</i> )	368.5	0.284 (0.267–0.301)	00:30:05	0.281 (0.267–0.296)	00:32:21
Turtle Dove ( <i>Streptopelia turtur</i> )	127.9	0.304 (0.29–0.318)	10:12:38	0.305 (0.292–0.318)	00:52:50
Tawny Owl ( <i>Strix aluco</i> )	93.0	0.139 (0.126–0.152)	00:17:42	0.142 (0.126–0.158)	00:35:56
Tree Pipit ( <i>Anthus trivialis</i> )	144.6	0.297 (0.275–0.320)	00:18:00	0.298 (0.281–0.315)	00:28:37
Tree Sparrow ( <i>Passer montanus</i> )	191.7	0.329 (0.308–0.349)	00:18:00	0.330 (0.322–0.338)	00:18:51
Tufted Duck ( <i>Aythya fuligula</i> )	155.7	0.197 (0.179–0.215)	01:02:00	0.199 (0.179–0.220)	00:33:44
Wheatear ( <i>Oenanthe oenanthe</i> )	335.3	0.373 (0.358–0.389)	00:24:00	0.374 (0.352–0.395)	00:27:04
Whinchat ( <i>Saxicola rubetra</i> )	74.9	0.208 (0.192–0.224)	00:25:00	0.210 (0.197–0.222)	00:32:25
Whitethroat ( <i>Sylvia communis</i> )	1386.8	0.577 (0.562–0.591)	00:20:00	0.582 (0.572–0.593)	00:23:03
Wood Warbler ( <i>Phylloscopus sibilatrix</i> )	51.7	0.167 (0.146–0.187)	00:36:00	0.168 (0.154–0.182)	00:33:38
Woodpigeon ( <i>Columba palumbus</i> )	2545.6	0.647 (0.635–0.659)	10:01:01	0.651 (0.638–0.664)	00:45:15
Wren ( <i>Troglodytes troglodytes</i> )	2497.4	0.496 (0.477–0.516)	02:03:48	0.498 (0.473–0.524)	06:23:53
Willow Tit ( <i>Poecile montanus</i> )	46.8	0.132 (0.123–0.140)	00:25:19	0.131 (0.117–0.144)	00:27:10
Willow Warbler ( <i>Phylloscopus trochilus</i> )	1390.2	0.575 (0.568–0.582)	00:22:41	0.575 (0.560–0.590)	00:47:40
Yellowhammer ( <i>Emberiza citrinella</i> )	1206.0	0.671 (0.657–0.685)	00:22:27	0.671 (0.659–0.683)	00:33:16
Yellow Wagtail ( <i>Motacilla flava</i> )	158.5	0.325 (0.312–0.338)	00:15:46	0.323 (0.302–0.343)	00:23:10

**Figure A1. Map of NUTS regions.**





Images: Liz Cutting / Edmund Fellowes / Tom Streeter. Cover image: Charles Tyler

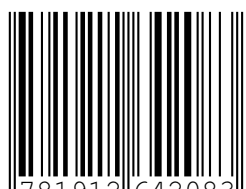
## Fine-scale mapping of relative abundance and trends, and extraction of small-area population trends for breeding birds

The BTO/JNCC/RSPB Breeding Bird Survey (BBS) is the UK's main scheme for monitoring population changes in the UK's common terrestrial breeding birds. Trends are published annually for up to 111 bird species for the UK, England, Scotland, Wales, Northern Ireland and nine regions within England. These trends are useful at showing large-scale population changes but we know that some species show spatially varying trends that are masked by national trends, and cannot be captured by regional trends due to sample size constraints. Previous work has developed maps of abundance for different periods of the BBS time series but these are now out of date and lack the temporal resolution of standard population trends. Here, we update previous work, aiming to produce annual predictions of relative abundance, which can vary spatially and temporally, for the 111 species featured in the published BBS trends from 1994 to 2016. We evaluated two methods, Generalised Additive Models (GAMs) and Geographically Weighted Regression (GWR) to model spatial and temporal variation in species abundance relative to habitat and elevation.

The GAM method worked well, yielding models with acceptable fit metrics and producing annual maps of relative abundance which agreed well with known distribution and abundance patterns from *Bird Atlas 2007–11*. From these maps we extracted trend estimates which were highly correlated with the published long-term and 10-year trends. These maps could be useful for setting local conservation priorities and may be better than the published BBS trends at estimating abundance for areas with poor coverage or rare and localised species. However, further work is needed to validate small area trends derived in this way against independent data.

Jennifer A. Border & Simon Gillings (2019). Fine-scale mapping of relative abundance and trends, and extraction of small-area population trends for breeding birds. BTO Research Report **720**, BTO, Thetford, UK.

ISBN 978-1-912642-08-3



9 781912 642083 >