# Predicting Population Change from Models Based on Habitat Availability and Utilisation

**Authors:**

Jason Matthiopoulos (University of Glasgow, Jason.matthiopoulos@glasgow.ac.uk),

Christopher Field (University of Glasgow, chris.field93@gmail.com),

Ross MacLeod (Liverpool John Moores University, r.c.macleod@ljmu.ac.uk and University of Glasgow)

**Running title:** Habitat composition and population change

**Corresponding author***:* Room 312, Graham Kerr Building, Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, G12 8QQ, Tel.: 0141 330 5969 Fax: 0141 330 5971 E-mail: Jason.Matthiopoulos@glasgow.ac.uk, Orchid ID: 0000-0003-3639-8172

## Abstract

The need to understand the impacts of land management for conservation, agriculture and disease prevention are driving demand for new predictive ecology approaches that can reliably forecast future changes in population size. Currently, although the link between habitat composition and animal population dynamics is undisputed, its function hasn't been quantified in a way that enables accurate prediction of population change in nature. Here, using 12 house sparrow colonies as proof-of-concept, we apply recent theoretical advances to predict population growth or decline from detailed data on habitat composition and habitat selection. We show, for the first time, that statistical population models using derived covariates constructed from parametric descriptions of habitat composition and habitat selection can explain an impressive 92% of observed population variation. More importantly, they provide excellent predictive power under cross-validation, anticipating 81% of variability in population change. These models may be embedded in readily available Generalised Linear Modelling frameworks, allowing their rapid application to field systems. Furthermore, we use optimisation on our sample of sparrow colonies to demonstrate how such models, linking populations to their habitats, permit the design of practical and environmentally sound habitat manipulations for managing populations.

***Keywords:*** habitat engineering, habitat selection, habitats-to-populations (HTP), house sparrow (*Passer domesticus*), population declines, population modelling

# Introduction

Accelerating environmental change is driving widespread demand for predictive ecology approaches that are able to make reliable forecasts about our impacts on biological populations (Evans et al. 2011, Mouquet et al. 2015, Dietze et al. 2018). Currently, ecology is particularly successful at explaining the population impacts of measured environmental variables (Mouquet et al. 2015). However, much more needs to be achieved if we are to produce accurate forecasts of how ecological systems will respond to future change (Evans et al 2011, Dietze et al. 2018). The ability to make reliable anticipatory predictions will have widespread applications in species and land management, whether for conservation, environmentally sustainable agriculture, the management of wildlife for food stocks or defence against disease vectors and alien species.

A priority focus in such new predictive ecology approaches is the crucial relationship between habitat quality and population growth, a link whose existence is undisputed but whose form and strength are harder to quantify (Morales et al. 2010, Matthiopoulos et al. 2011, Matthiopoulos et al. 2015). The proliferation of spatial modelling in ecology in recent decades is beginning to see data on the spatial distribution of populations being connected to individual aspects of population demography (Johnson et al. 2004, McLoughlin et al. 2006, 2007, Aldridge & Boyce 2007, 2008, Blouin-Demers & Weatherhead 2008, Erickson et al. 2014). By harnessing the predictive power of the link between habitats and populations, such efforts can help build the sort of anticipatory ecological science we urgently need.

The growth rate of a population is ultimately driven by a mosaic of spatially heterogeneous habitat covariates representing conditions, resources and risks (Guisan & Zimmermann 2000, Matthiopoulos et al. 2015). To understand and manage population change via habitat composition, we need to identify which habitat variables are likely to be influential. This requires two types of methods, used in tandem. First, transferrable models of habitat selection (Yates et al. 2018) that can successfully capture a species' distribution at both fine and coarse scales and, second, frameworks for connecting habitat use to concurrent population change data. In this paper, we exploit recent developments in both of these directions integrated in a statistical approach recently developed by Matthiopoulos et al. (2015). This framework, henceforth called Habitats-To-Populations (HTP), was derived from first principles, and is thus well supported theoretically, but its utility and power have yet to be tested in a real system. Here, we make the first test in nature of the ability of the HTP framework to predict population change from habitat composition. We do so using a set of house sparrow (*Passer domesticus*) colonies, a species that has experienced rapid, but poorly understood, population declines in parts of its global range, while becoming a highly invasive species elsewhere (Hole et al 2002). Further, we use this quantitative link between habitat composition and change in population size to examine how nascent habitat engineering approaches could be developed to achieve – often elusive – ecological objectives of conservation or pest management.

## Materials and Methods

Population change was monitored over a two-year period in a total of 12 sparrow colonies where the colony's nesting area had been identified at the start of monitoring. In these, and a further 20 colonies, added to provide a broader representation of habitat scenarios, the composition of all the gardens within each colony's spatial range was documented and observations of habitat use by sparrows were made. Following the paradigm illustrated in Fig. 1 (adapted from Matthiopoulos et al. 2015), our analysis approaches the combined data on habitat use and population change by linking a flexible modelling framework of habitat preferences with a habitat-sensitive model of population growth. In order to test the performance and utility of the HTP theoretical framework for achieving the goals of predictive ecology, we investigated the model’s explanatory power, its predictive power under cross-validation and its ability to generate specific, testable recommendations regarding habitat improvements.

### *Study species*

Although increasingly considered as an invasive species in many parts of the world, the house sparrow has been suffering heavy declines in much of its native range in the last decades. Declines have been particularly severe in North-Western Europe and in the UK, where the breeding population has dropped by 71% since the 1970s to 5.1 million today, placing the species on the UK conservation red-list (Robinson et al. 2016). The causes of sparrow declines in urban/suburban environments, where the majority of the population are found, remain uncertain but are probably connected to changing habitat composition of gardens brought on by changes in socioeconomic status and home-improvement trends (Shaw, Chamberlain & Evans 2008). It has been suggested that garden paving reduces suitable foraging area, that the use of introduced plants reduces native vegetation and leads to a reduction in invertebrate prey (Chace & Walsh 2004, Shaw et al. 2008, Seress et al 2012), while tidier gardens and modern roofing techniques reduce the availability of cover from predation and the number of nesting sites (Daniels & Kirkpatrick 2006). However, none of these variables have been definitively linked to population change at a landscape scale using population replicates. This requirement, as well as the ease with which sparrows can be observed and their urban and suburban environment can be manipulated, make the species particularly suitable as a test-bed for new modelling ideas and subsequent validation experiments.

### *Data collection*

The set of 32 colonies for our study lay within an area of 200km$^2$ in and around the city of Glasgow in Scotland and were chosen to maximise the diversity of structural habitat characteristics and observed colony size.

To achieve the fine spatial scales needed for modelling habitat selection, the habitat usage and availability study was carried out by surveying all the individual gardens (150 in total) within the range of each sparrow colony. High-resolution (2x2m grid square size) habitat and space use survey data were collected covering the gardens in each colony. During the habitat usage survey, all the cells in each colony were continually observed until a total of 20 separate sparrow visits to grid cell locations were recorded. We aimed to distribute our observation equally for each unit of area within the complex of gardens belonging to each colony's domain. An additional 20 locations were then randomly selected from all the grid squares within the colony. This provided us with a use-availability data set comprising an equal number of ones and zeroes.  For our 32 colonies, this led to 1280 distinct spatial cell observations of habitat usage. Using Google Earth imagery (GOOGLE EARTH, 2016), for each of the 1280 cells we estimated the percentage covered by hedge, grass, bush, roof, artificial surfaces and trees to the nearest 10%. We also calculated the distance of each cell from the nearest hedge and roof. For 12 focal sparrow colonies, baseline population surveys were available from a citizen science programme, the Glasgow House Sparrow Project, part of a partnership between the Royal Society for the Protection of Birds (RSPB) and the University of Glasgow. This provided data on house sparrow colony size as measured by counting all males present in each colony during the main part of the breeding season (April to July) in 2014.  The population surveys were undertaken by trained sparrow surveyors who repeatedly walked every street and back lane in their survey area and counted the total number of adult sparrows in each colony. House sparrows are confident around humans, very vocal and easily identifiable, making exact determination of colony population size in the breeding season straightforward in urban areas. During the breeding season, house sparrows are extremely site faithful to their nest area, with radio tracking showing maximum travel distances of <100m (Shaw 2009). However, even this distance is rare given that 95% of the foraging locations during breeding occur in an area that can be represented by a 16m radius circle (calculated from data in Shaw 2009), allowing precise population counts for each survey area. House sparrows breed in pairs and males are particularly easy to detect and count because they sing for long periods from highly visible song posts next to their nests. We therefore used the maximum number of males recorded in each colony as the measure of colony size. Population growth rates were established as the difference in colony sizes recorded between the years 2014 and 2016 using follow-up colony surveys in the 2016 breeding season based on the same methodology as the first citizen science surveys. The increased sample size (N=32) for the habitat usage survey part of the modelling was chosen to help ensure that the habitat model was provided with adequate data (our habitat use model was particularly data hungry because of the multiple pairwise interactions it contained - see next sub-section).

### *Statistical analysis of habitat use*

Within the broad field of Habitat Selection Functions (HSFs, a term that we use here to describe closely related approaches such as Resource Selection Functions – Manly et al. 2004 – and Maximum Entropy models – Philips, Anderson & Schapire 2006), it is widely recognised that animals respond not merely to the habitat characteristics of their immediate location but also to the overall habitat composition of their surrounding area (e.g., the availability of all habitats within their entire home range). Matthiopoulos et al. (2011) proposed the framework of Generalised Functional Responses (GFR), which capture nonlinearities in the habitat preferences of animals by modelling the regression coefficients of an HSF as functions of local habitat availability. By using data from multiple scenarios of availability (e.g., the different sparrow colonies in our study) a GFR can interpolate (in environmental space) the response of the species under as-yet unobserved scenarios of habitat availability. For *n* explanatory variables $\mathbf{X} = (X_1, \ldots, X_n)$ with $n+1$ regression coefficients $\beta = (\beta_0, \ldots, \beta_n)$, the general HSF takes the form. $h$

$$h(\mathbf{X}) = \exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i) \quad (1)$$

These functions can be obtained from statistical models that are fitted equivalently to count, use-availability or point-process data (Aarts et al. 2012), either via likelihood or maximum entropy criteria. To extend the HSF framework to account for the regional availability of all habitats influencing local usage, the Generalised Functional Response writes each beta coefficient as a function of habitat availability

$$\beta_i = \sum_{i=1}^{n} \sum_{j=1}^{k} b_{ij} E(X_i^j) \quad (2)$$

Where $E(X_i^j)$ is the *j*th order expectation of the *i*th covariate.  These expectation terms provide statistical summaries for the distribution of explanatory variables in the neighbourhood of each spatial point of interest. Their inclusion allows the model to quantify the environmental context within which utilisation decisions are made by animals (Paton & Matthiopoulos 2016). Including expectations up to the maximum *k*th order allows the GFR to account for the average, variance, skewness, kurtosis  etc. of the distribution of variables in surrounding space. However, in the simplest case (requiring the least amount of usage data to be estimated), only the averages (i.e. 1st order expectations) are used. When the coefficients in eq. (2) are placed into the model of eq. (1), they result in interaction terms between local covariate values and their regional expectations.

Eq. (2) implies that each sampling scenario (in our study, data within the home range of each sparrow colony) is characterised by a unique response (i.e. a unique regression slope) to each of the covariates, based on characteristics of regional habitat availability. It is, however possible to calculate this unique numerical value of each coefficient for each sampling scenario and to compare the response of animals to covariates between different sampling scenarios. In practice, the GFR takes the form of a generalised linear (or additive) model that comprises all the pairwise interactions between the values of environmental variables at any one sampled location and their (1st, 2nd, etc.) expectations in the vicinity of a location across the sampling scenario (Matthiopoulos et al. 2011, 2015, Aarts et al. 2013, Paton & Matthiopoulos 2016). The specific form of the GLM depends on the data to be analysed (Aarts et al. 2008, Aarts et al. 2012). In our case, the value of use or availability (1 or 0) was used as the response variable in a logit GLM with binomial likelihood, modelling the probability that a cell with particular habitat characteristics contained one of our sparrow detections. To create the GFR, we distinguished between spatially referenced variables (i.e. the characteristics of each 2x2m cell in a garden) and colony/survey-specific variables (e.g. the known size of a colony, or prevailing weather information). Colony size was used as a colony-specific covariate in the habitat model to account for the effect of crowding on apparent habitat suitability (i.e. the possibility that suboptimal habitats are used by birds when they are displaced from optimal habitats by conspecifics).

Analysis was performed using R v3.3.0 (R Core Team 2016) with the aid of the R library HaToPo (HAbitats-TO-POpulations, available from https://github.com/JasonMat/HATOPO.git).

*Linking habitat to population growth*

The second stage of the HTP analysis links population change data to habitat composition and was recently derived by Matthiopoulos et al. (2015). Here, and in Fig. 1, we briefly outline the components of HTP, but the mathematical details can be found in that earlier paper. HTP utilises estimated parameters, generated by two distinct statistical models, namely, a detailed approximation of habitat availability and a species-habitat association model (the HSF/GFR, see previous subsection). Both of these constituent models need to be fitted to multiple *sampling scenarios*, to give them the opportunity to learn how the same organisms respond to a diversity of habitat compositions. A sampling scenario corresponds to a definable population or sub-population and its spatial range during a short time window. In this study, it refers to individual house-sparrow colonies. The parameter estimates from these two models are then used in pre-derived mathematical formulae (Matthiopoulos et al. 2015) to construct candidate covariates to be regressed against observations of population change. These new, constructed covariates, although not explicitly spatial, nevertheless contain information about the combinations of habitats experienced by the individuals living in different sampling scenarios. The theoretical motivation behind these new covariates (and consequently, the hypothesis put to the test in the present paper, using real data) is that, by more precisely encompassing the portfolio of habitats exploited by the study individuals, a population model will be better able to retrieve predictive signal from data of population change. This theoretical approach currently entails a set of simplifying assumptions (see also Matthiopoulos et al. 2015), whose relevance to the sparrow system we examine below:

1) Accessibility: The entire home range of each colony is assumed equally accessible to each sparrow, such that points are not likely to receive differing degrees of usage due to their distance from the nest. Given the very strong site fidelity of sparrows (see earlier discussion) and resulting short range of movements, this assumption is likely to hold.

2) Spatial-pseudoequilibrium: We assume (in a way similar to most habitat analyses - Guisan and Thuiller 2005) that our space-use data come from a population whose distribution is not currently undergoing radical adjustments.

3) Negligible individual variation: We assume that habitat use by any one individual is representative of the use of space by the entire colony and that individuals are similar to each other in their behaviour and reproductive potential.

4) Non-depletable covariates: Although the approach does take into account secondary effects in distribution and colony growth resulting from crowding and density dependence, we are not explicitly modelling resource depletion. Consequently, our sparrow case study has used habitat characteristics, rather than measurements of prey abundance, as the candidate covariates.

5) Exclusion of more sophisticated features: Finally, the model does not currently examine genetic change during the window of observations, formulations with non-additive habitat effects on population fitness, Allee effects in population growth, or saturating responses to superabundant resources.

The constructed covariates are employed in a generalised linear model of population change to investigate their impact on population change data in a set of sampling scenarios. This stage of the analysis (corresponding to boxes 3-6 in Fig. 1) relates population growth rates to habitat composition. The deterministic population model used as a platform by Matthiopoulos et al.

(2015) has extensive roots in ecology and was most recently proposed by Turchin (2003) as a convenient starting point for population inference

$$\frac{N_{t+1}}{N_t} = \exp(F(\mathbf{f},\mathbf{N})) \tag{3}$$

where $F$ is some function of habitat composition $\mathbf{f}$ and past population densities $\mathbf{N} = \{N_t, N_{t-1}, ...\}$ . The stochastic version of this model takes the form

$$\begin{aligned} N_{t+1} &\sim Poisson(\lambda_t) \\ \lambda_t &= N_t \exp(F(\mathbf{f},\mathbf{N})) \end{aligned} \tag{4}$$

The Poisson distribution is used here purely for illustrative purposes. Alternative distributions such as the negative binomial may be used to capture over-dispersion in population count data. Eqs (4) can be conveniently fitted to population count data as a generalised linear model (GLM) with linear predictor given by

$$\ell_t = F(\mathbf{f},\mathbf{N}) + \ln(N_t) \tag{5}$$

In regression terminology, $\ln(N_t)$ is the GLM's offset. Two types of statistical population models were fitted to the colony growth data. The first, used here as a baseline, we call the "mean-field" model. It relates population growth to the average values of the habitat variables across the range of each colony (based on samples from gardens around the nesting colony). This model ignores the nuances of habitat availability and habitat usage. The second, called here the "spatial" model was based on the HTP approach and relates population change to the detailed profile of habitat composition in the domain of each colony. As an example of implementation of eq. (5) for the two cases, our selected mean-field model in the sparrow analysis (see results below) was a GLM with linear predictor

$$\ell_t = \alpha_0 + \alpha_1 \left\langle Bush \right\rangle + \alpha_2 N_t + \ln(N_t) \tag{6}$$

Here, log-linear population growth is predicted by the average availability of bushes across the domain of each colony, with the added effect of the colony's size (representing density dependence). Such a model, employing average conditions across the domain of a colony, may be able to capture coarser aspects of habitat dependence but it ignores the fact that members of the population are exposed to combinations of covariates spatially, and they can modulate that exposure (whether beneficial or otherwise) by differentiating their use of different habitat types within their range. This is where information from the habitat model can extend the predictive reach of mean-field models such as eq. (6). The effect of habitat availability and habitat use on population growth was estimated via the HTP model, implemented in the HaToPo library for R (R-Core 2016). HaToPo yields a set of constructed candidate covariates of population change, each of which is a function of previously estimated parameters from the models of habitat availability and habitat use. Furthermore, it captures the spatial features of conspecific interference to create a constructed covariate for density dependence. Our final spatial model was, once again, a GLM of the form of eq. (5) but this time, the linear predictor comprised constructed covariates in bush, grass, roof and density dependence (see results, below).

$$\ell_t = \alpha_0 + \alpha_1 Bush + \alpha_2 Grass + \alpha_3 Roof + \alpha_4 f(N_t) + \ln(N_t) \tag{7}$$

*Model fitting and selection*

It was possible to subject our habitat model (the GFR) to autonomous selection via the Akaike Information Criterion (AIC – e.g. Burnham and Anderson, 2002). This would lead to the most parsimonious model of habitat preference and use. However, we found that this approach led to habitat models that were over-parameterised for the later purposes of population prediction. Thus far, this is a unique problem in the literature, because previous studies on habitat preferences did not historically need to consider predictions of population growth. We decided to perform a search for parsimonious habitat models evaluating each candidate GFR on the basis of how well it ultimately enabled the population model to predict population change. This was done by means of leave-one-out cross-validation on our 12 colonies for which population change data were available.

Our model fitting and model selection protocol was as follows (the full R-code is provided in the Appendix, subdivided in sections): We first specified a candidate model (Section 2 of R-code) and fitted it to the aggregate space-use data from all 32 colonies (Section 3 of R-code). The gfr() function in the HaToPo library automatically extends the dataframe to include habitat availability expectations and specifies the necessary interaction terms for the fully expanded GFR model formula. Using the functions favail() and ga.gfr() in the HaToPo library, we obtained two objects containing 1) parametric approximations of habitat availability for each colony and 2) the colony-specific habitat selection coefficients. These two objects were then passed to the pop.covariates() function in HaToPo (Section 5 in R-code) which constructed a set of candidate covariates for population growth. Population growth was regressed against these constructed covariates (Section 5 in R-code). Since we only had population data for 12 colonies, we subjected this population GLM to model selection by AIC. The ability of the combined habitat (HSF/GFR) and population (HTP) model to explain population data was evaluated by the percentage deviance explained in the final population model. The ability of the combined models to predict new data was evaluated by the Sum of Squared Residuals (SSR) under leave-one-out cross validation applied to the entire modelling workflow (Section 6 in R-code). We selected the combined habitat and population model based on predictive ability (all the habitat models examined are listed at the final section of the R-code).

*Optimising population growth via habitat manipulation*

The resulting fully fitted spatial population model (in eq. (7)) connects habitat composition to population growth. To evaluate how responsive the species might be to conservation actions through changes in gardening practices, we carried out the following optimisation protocol (Section 8 in R-code). We perturbed the non-structural aspects of different garden cells in our data set (i.e. the percentages of grass, tree, hedge and bush), until we obtained the maximum improvement in predicted growth rates in all 12 of our colonies. However, we constrained our optimisation so that we only accepted proposed landscaping improvements if the confidence intervals associated with the predicted growth rates did not also include a wide range of possible deteriorations (specifically, no proposed improvement was accepted if its lower confidence interval was less than 70% of the previously accepted mean improvement). There is considerable precedent in the reserve design literature for such constrained optimisation that incorporates objective functions constructed from combinations of point and interval model estimates (Moilanen & Wintle 2006). As part of our optimisation, we sought to understand 1) whether all colonies could benefit from such manipulations, 2) how much improvement we might expect and 3) how drastic a manipulation would be required to generate these improvements.

## *Results*

### *Habitat usage*

Following model selection, the preferred model of habitat use contained three main effects (percentage of grass, bush and roof in each cell) and 16 other terms representing average habitat values across the colony and interactions between the model's main terms and colony-specific averages. The combination of main effects and interactions in the GFR allows the construction of colony-specific coefficients of habitat selection arising from the response of the animals to the particular habitat configuration in the vicinity of each colony, a phenomenon, known in the literature as a functional response in habitat selection (Arthur et al. 1996, Mysterud & Ims 1998, Mauritzen et al. 2003, Matthiopoulos et al. 2011, Aarts et al 2013). We examined the histograms and pairwise correlations between different habitat-selection coefficients, estimated for different colonies (Fig. 2). Looking at the single-coefficient histograms (on the diagonal of Fig. 2) we see that sparrows predominantly avoid areas with grass (histogram centred at negative values in Fig. 2aa) and prefer areas with bush cover (histogram centred at positive values in Fig. 2bb). The histogram of coefficients for roof (Fig. 2cc) does not present a predominantly positive or negative bias, indicating that in different colonies (depending on habitat composition) sparrows may use roofs more, or less than expected by chance. However, the variable has been retained by model selection because it lends the habitat model explanatory power. The off-diagonal plots in Fig. 2 also provide evidence of strong pairwise correlation between modelled responses to all three environmental variables. The specific coefficient values estimated are determined by the existence of trade-offs and synergies between coefficients for different variables. Trade-offs (negative correlations) would indicate that these covariates are substitutable. Synergies, shown by a positive relationship between coefficients, indicate that the associated covariates operate in a complementary way, increasing each-other's apparent preference.

The habitat usage model explained a modest percentage (33%) of the null deviance, suggesting either an inherently high noise-to-signal ratio in habitat use by sparrows (e.g. individual variation), or that not all relevant features (e.g. direct measures on availability of food) were included in the analysis. The parameters of the GFR model were next embedded in the HTP approach (see next section), to represent the expected frequency of interactions between sparrows and different features of their environment.

### *Linking habitat to population growth*

The 12 focal house sparrow colonies at which we measured population growth varied in size from 2 to 24 pairs, with a mean colony size in the 2014 breeding season of 6.5+1.7 pairs increasing to a mean colony size of 8.2±1.7 pairs in the breeding season of 2016.

Cross-validation served as the ultimate arbiter for the best spatial and mean-field models (Table S1). Despite using such a numerically greedy approach to model selection, computational requirements were very low (in the order of seconds) because HTP is based on pre-derived mathematical expressions (Matthiopoulos et al. 2015) and generalised linear models. Both the mean-field and spatial models found relationships between population change and their respective bush covariates, as well as evidence for density dependence. The spatial model additionally found links with grass and roof. The best-performing spatial model explained 92% of the deviance in the observed population growth rates ($r^2 = 0.9$ - Fig. 3a) and was able to retain high performance under prediction, where it achieved an $r^2$ of 0.81 (Fig. 3b). The spatial model (Table S1) was characterised by a large, positive intercept, expressing a tendency for population growth, which was moderated strongly by too much exposure to grass and roof structures, but also (to a lesser extent) exposure to bushes. This was somewhat unexpected since the habitat model indicated an overall preference for bushes. The mean-field model explained a comparatively lower but still respectable percentage of deviance (72% with $r^2 = 0.68$ - Fig. 3c) but suffered a greater drop when tested for predictive ability under cross-validation ($r^2 = 0.5$ -

Fig. 3d). The mean-field model was also poorer at detecting factors linked to population growth with the overall availability of bushes and density dependence the only main explanatory variables identified as important.

Therefore, the HTP approach was able to extract considerably more explanatory and predictive power by looking at the exact composition of habitats and whether they were over- or under-used by sparrows, rather than the average of single environmental variables on which most existing modelling approaches would be based. We also observed a difference in specificity between mean-field and spatial models. In particular, whereas several of the mean-field candidate models had similar cross-validation scores to the best model of that category, the spatial model category comprised highly variable cross-validation scores (therefore, it strongly differentiated between the high-performance model selected, and several others that were similar in structure but poorer in performance than the mean-field equivalents).

### *Optimising population growth via habitat manipulation*

Our optimisation exercise aimed at using the HTP approach to predict how colony growth rates might be increased by judiciously manipulating habitat composition. We arrived at recommendations for improving growth for 10 out of 12 colonies (Fig. 4). Our results suggested that six out of seven colonies with currently zero or negative growth rates could be pushed to positive growth. Notably, the recommended changes (increases or decreases in any one type of habitat) were not proportional across colonies, underlining the point that joint availability of all habitats is important for determining habitat use and population impacts. For the specific set of colonies we focused on, a reduction in the amount of grass was the most frequently recommended habitat manipulation (suggested as a way to increase population growth rate in 11 out of the 12 colonies, colonies in Fig. 4). The other habitat manipulations were more colony-specific but involved offsetting the reduction in grass with recommended increases to the various habitat variables that represent potential cover from predators (bushes, hedges or trees depending on the specific colony). All these increases were achievable with low potential for unpleasant surprises (i.e. the confidence intervals shown as rhombuses in the proposed growth rates of Fig. 4 did not encompass the possibility of population declines).

## Discussion

Here, we have exploited new theoretical developments in statistical ecology to demonstrate, for the first time, that predictive modelling of population growth in nature is possible based on habitat composition. Further, we have demonstrated an early example of an optimisation process for habitat composition that shows how a predictive ecology modelling approach can be used to derive tailored land management recommendations that could be used to promote, or control population growth for conservation, sustainable agriculture, wildlife stock management and defence against disease vectors.

In our study species, habitat selection was noisy. Our habitat model was able to account for only 33% of the variability in the space use data. Its modest explanatory power could be because important covariates of usage were absent. Examples might include the actual (Barnard 1980, Gotmark & Post 1996), or perceived (Brown & Kotler 2004, Zanette et al. 2011) risk of predation. Alternatively, sparrow behaviour in response to habitat could be genuinely very noisy due to unmodelled individual variation or due to the time of year at which habitat usage was being measured (the survey was carried out during a period when resources are less constrained than they would be in winter (MacLeod et al. 2006)). Despite this, inclusion of habitat use and availability into our statistical model of colony growth rates pushed explanatory power from

72% to 92% (compare Figs. 3c and 3a) with even greater gains (50% to 81%) in predictive power (compare Figs 3d with 3c). To our knowledge this is the first modelling approach to successfully predict the majority of population growth variation across replicate populations based on habitat composition and usage.

It should be noted that this is a proof-of-concept study on a relatively small sample of 12 populations, and the above metrics of predictive performance are based on the same cross-validation procedure used to carry out model selection. More work will be needed to examine if the high percentages of predictive power can be sustained for other species and out-of-data predictions for larger samples of independent populations. It is nevertheless possible to evaluate our results based on a comparison between the spatial and mean-field approaches. Our constructed covariates, based on habitat availability and expected use, had much greater predictive power than the average availability of habitat variables. The conclusion therefore is that, even if the animals' usage of the different components of their habitat has elements of randomness (as the noisy GFR seems to suggest), the signal of habitat composition may remain strong within population growth data.

The HTP framework in its current form is particularly suited to the life history of house sparrows. However, future research could be directed to make the HTP approach applicable to populations with variable accessibility constraints, non-equilibrium distributions, population age- or stage-structure, depletable covariates, genetic change, non-additive habitat effects on population fitness, Allee effects in population growth, or saturating responses to superabundant resources. From the point of view of statistical inference, in future implementations, the two model components of habitat selection and population growth would ideally be implemented as simultaneous (or, at least connected) inference, so that uncertainty from both models is propagated through the full workflow to the final results. Such joint inference is, as yet unavailable in the literature. Our cross-validation approach in the present paper (applied to the full workflow) was a step in this direction.

Our garden habitat composition manipulation exercise (Fig. 4) is mainly a sensitivity analysis used to illustrate the extent of garden modifications required to improve colony viability and the improvements that might be possible. Even though we chose a simplistic algorithm that altered garden cells independently of each other (an approach that is too fragmented from a garden landscaping point of view), our analysis does demonstrate that habitat manipulation can have large impacts on expected population growth, in several cases even reversing population declines. Such emerging exemplars of the possibilities for habitat engineering for the purposes of management have considerable future potential, and deserve further algorithm development and experimental testing. Since the habitat modifications suggested here were achieved by focusing just on changing vegetation types in gardens, rather than changes to more fixed structural elements of the landscape, the predictions from these types of models should be practical to test. Ultimately, this type of testing could enable conservation organisations, such as the Royal Society for the Protection of Birds in the UK, to provide bespoke guidance to individual homeowners and other land owners who wish to enhance the wildlife potential of their properties.

Quite apart from its applied importance, our study contributes several conceptual developments relating to habitat. For example, it casts a critical light on the implicit assumption (pervading the broad class of applied habitat models) that hotspots in spatial usage are indicative of fitness-conferring habitats. Although it is understood that, biologically, the relationship between space use and population growth may not be straightforward (Pulliam 2000, Morales et al. 2010, Peterson et al. 2011), high-usage and high-fitness are often conflated in contemporary literature (e.g. Stockwell, 2006), in order to provide solutions to urgent applied problems. However, the fact that usage is observed and modelled directly, but fitness is not, leaves open the possibility of mismanaging wildlife whose distribution does not directly reflect habitat suitability. The predictive ecology modelling approach we demonstrate here shows how such problems could be overcome by the generation of fully estimable population models, able to predict how populations would change under specific management and environmental change scenarios.

Packaging habitat selection and population fitness in simple qualitative rules of thumb will continue to be a challenge long into the future, because such rules cannot apply indiscriminately across different species or even subpopulations. This challenge means that models, such as the one presented here, could prove valuable in the design of conservation and landscape management strategies, particularly when it is imperative to exploit efficiencies in management to navigate human-wildlife conflicts, where only narrow strips of common ground can reconcile conservation with human exploitation (Redpath et al. 2013). Much work remains to be done before habitat manipulations can be designed with regularity and without the close supervision of ecological experts. However, such efficiencies now appear more achievable, given our finding that, at least in this study system, moderate manipulations in habitat are predicted to have large effects on population growth, even reversing ongoing declines.

## Acknowledgments

We thank the all the RSPB staff and volunteers involved in the Glasgow House Sparrow Project without whose immense support, and enthusiasm over several years the data collection for this study would have been impossible. In particular Dr Kat Jones, Toby Wilson and Sarah-Jayne Foster were instrumental in developing the ideas and providing the support that lead to the long-term population survey programme. Rebecca Dickson collected much of the habitat use data as part of her BSc project. RCM was supported by a Royal Society of Edinburgh Scottish Government Research Fellowship. Three anonymous reviewers provided valuable comments.

## References

Aarts, G., MacKenzie, M., McConnell, B., Fedak, M. & Matthiopoulos, J. (2008) Estimating space use and environmental preference from telemetry data. Ecography. 31, 140-160.

Aarts, G., Matthiopoulos, J. & Fieberg, J. (2012) Comparative interpretation of count and point methods for species distribution models. Methods in Ecology and Evolution. 3: 177-187.

Aarts, G., Brasseur, S., Fieberg, J., Matthiopoulos, J. (2013) Quantifying the effect of habitat availability on species distributions. Journal of Animal Ecology. 82: 1135-1145.

Aldridge, C. L. & Boyce, M. S. (2007) Linking occurrence and fitness to persistence: a habitat-based approach for endangered greater sage-grouse. Ecol. Appl. 17:508–526.

Aldridge, C. L. & Boyce, M. S. (2008) Assessing occurrence and fitness for persistence: a case study of greater sage-grouse brood habitat. Israel J. Ecol. Evol. 54: 389–419.

Barnard, C.J. (1980) Flock feeding and time budgets in the house sparrow (*Passer-domesticus l*). Animal Behaviour. 28: 295-309.

Blouin-Demers, G. & Weatherhead, P.J. (2008) Habitat use is linked to components of fitness through the temperature-dependence of performance in rat snakes (*Elaphe obsoleta*). Israel J. Ecol. Evol. 54: 361–372.

Boyce, M.S. & McDonald, L.L. (1999) Relating populations to habitats using resource selection functions. TREE, 14:268-272.

Boyce, M.S., McDonald, L.L. & Manly, B.F.J. (1999) Relating populations to habitats: Reply. TREE, 14: 490.

Brown, J.S., Kotler, B.P. (2004) Hazardous duty pay and the foraging cost of predation. Ecology Letters, 7: 999-1014

BTO, 2016, *House Sparrow Research* [online] Availible at: http://www.bto.org/volunteer-surveys/gbw/about/background/projects/sparrows [Accessed 22/08/16]

Burnham, K. P. and Anderson, D.R. (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach. 2d ed. New York: Springer-Verlag.

Chace, F.J. & Walsh, J.J. (2004) Urban effects on native avifauna: a review, *Landscape and Urban Planning*, 74: 46-79.

Chamberlain, D.E., Cannon, A.R., Toms, M.P., Leech, D.I., Hatchwell, B.J., & Gaston, K.J., (2009) Avian production in urban landscapes: a review and metanalysis, *Ibis*, 151: 1-18.

Daniels, G.D. & Kirkpatrick, J.B. (2006) Does variation in garden characteristics influence the conservation of birds in suburbia? *Biological Conservation*, 133: 326– 335.

DeLaet, J. & Summers-Smith, J.D. (2007) The status of the urban house sparrow Passer domesticus in north-western Europe: a review. *Journal of Ornithology*. 148: 275–278.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen, L.G., Loescher, H.W., Lunch, C.K., Pijanowski, B.C., Randerson, J.T., Read, E.K., Tredennick, A.T., Vargas, R., Weathers, K.C. & White, E.P. (2018) Iterative near-term ecological forecasting. *Proceedings of the National Academy of Sciences* DOI: 10.1073/pnas.1710231115

Evans, K.L., Newson, S.E. & Gaston, K.J. (2009) Habitat influences on urban avian assemblages, *Ibis*, 151: 19–39.

Evans, M.R., Norris, K.J. and Benton, T.G. (2012) Predictive ecology: systems approaches. Phil. Trans. R. Soc. B 367: 163 -169.

GOOGLE EARTH version 6.1, 2016 URL: https://www.google.com/earth/ [accessed 22/08/16]

Gotmark, F., Post, P. (1996) Prey selection by sparrowhawks, Accipiter nisus: Relative predation risk for breeding passerine birds in relation to their size, ecology and behaviour. *Philosophical Transactions of the Royal Soc. B.* 351: 1559-1577.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. Ecol. Model., 135:147-186.

Hole, D.G., Whittingham, M.J., Bradbury, R.B., Anderson, G.Q.A., Lee, P.L.M., Wilson, J.D. & Krebs, J.R. (2002b) Widespread local house-sparrow extinctions - Agricultural intensification is blamed for the plummeting populations of these birds. *Nature (London),* **418**, 931–932.

Johnson, C. J., M. S. Boyce, C. C. Schwartz, and M. A. Haroldson. (2004) Modeling survival: application of the multiplicative hazards model to Yellowstone grizzly bear. *J. Wildl. Manage.* 68: 966–978.

MacLeod, R., Barnett, P., Clark, J. & Cresswell, W. (2006) Mass-dependent predation risk as a mechanism for house sparrow declines? *Biology letters*, 2: 43–6.

Manly, B., McDonald, L., Thomas, D., McDonald, T. & Erickson, W.P. (2004) *Resource selection by Animals - Statistical design and analysis for field studies*. Chapman & Hall, London, 177 pp.

Matthiopoulos, J., Hebblewhite, M., Aarts, G. & Fieberg, J. (2011) Generalized functional responses for species distributions. *Ecology*. 92: 583-589.

Matthiopoulos, J., Cordes, L., Mackey, B. et al. (2014) State-space modelling reveals proximate causes of harbour seal population declines. *Oecologia*. 174: 151.

Matthiopoulos, J., Fieberg, J., Aarts, G., Beyer, H. L., Morales, J. M. and Haydon, D. T. (2015) Establishing the link between habitat selection and animal population dynamics. *Ecological Monographs*. 85: 413–436.

Mauritzen, M., Belikov, S.E., Boltunov, A.N., Derocher, A.E., Hansen, E., Ims, R.A., Wiig, O. & Yoccoz, N. (2003) Functional responses in polar bear habitat selection. *Oikos*. 100, 112–124.

McLoughlin, P., Boyce, M. S., Coulson, T. & Clutton-Brock. T. (2006) Lifetime reproductive success and density-dependent, multi-variable resource selection. *Proc. Royal Soc. B*. 273: 1449–1454.

McLoughlin, P.D., Gaillard, J.-M., Boyce, M. S., Bonenfant, C., Messier, F., Duncan, P., Delorme, D., Van Moorter, B., Saïd, S. & Klein. F. (2007) Lifetime reproductive success and composition of the home range in a large herbivore. *Ecology* 88: 3192–3201.

Moilanen, A. & Wintle B.A. (2006) Uncertainty analysis favours selection of spatially aggregated reserve networks. *Biological Conservation*.129: 427–434

Morales, J.M., Moorcroft, P.R., Matthiopoulos, J., Frair, J., Kie, J.G., Powell, R.A., Merill, E.H. & Haydon, D.T. (2010) Building the bridge between animal movements and population dynamics. *Philosophical Transactions of the Royal Society B.* 365: 2289-2301.

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., Faure, D., Garnier, E., Gimenez, O., Huneman, P., Jabot, F., Jarne, P., Joly, D., Julliard, R., Kéfi, S., Kergoat, G. J., Lavorel, S., Le Gall, L., Meslin, L., Morand, S., Morin, X., Morlon, H., Pinay, G., Pradel, R., Schurr, F. M., Thuiller, W. and Loreau, M. (2015), Predictive ecology in a changing world. *J. Appl. Ecol.* 52: 1293–1310.

Mysterud, A. & Ims, R.A. (1998) Functional responses in habitat use: availability influences relative use in trade-off situations. *Ecolog*y. 79, 1435–1441.

Paton, R. S. and Matthiopoulos, J. (2016) Defining the scale of habitat availability for models of habitat selection. *Ecology*. 97: 1113–1122.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions.* Princeton University Press. Princeton. 314pp.

Phillips, S.J., Anderson, R.P. & Scapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*. 190: 231-259.

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*. 3: 349-361.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Redpath, S.M., Young, J., Evely, A., Adams, W.M., Sutherland, W.J, Whitehouse, A., Amar, A., Lambert, R.A., Linnell, J.D.C., Watt, A. & Gutierrez, R.J. (2013) Understanding and managing conservation conflicts. *TREE*. 28: 100-109.

Robinson, R.A., Leech, D.I., Massimino, D., Woodward, I., Eglington, S.M., Marchant, J.H., Sullivan, M.J.P., Barimore, C., Dadam, D., Hammond, M.J., Harris, S.J., Noble, D.G., Walker, R.H. & Baillie, S.R. (2016) BirdTrends 2016: *trends in numbers, breeding success* *and survival for UK breeding birds. Research Report* 691. BTO, Thetford. www.bto.org/birdtrends.

Seress, G., Bkony, V., Pipoly, I., Sz, T. Nagy, K. & Liker, A. (2012) Urbanization, nestling growth and reproductive success in a moderately declining house sparrow population. *Journal of Avian Biology*, 43: 403–414.

Shaw, L.M. & Chamberlain, D. & Evans. M. (2008) The House Sparrow Passer domesticus in urban areas: reviewing a possible link between post-decline distribution and human socioeconomic status. *Journal of Ornithology*. 149: 293-299.

Shaw, L.M. 2009. *Investigating the role of socioeconomic status in determining urban habitat quality for the House Sparrow, Passer domesticus.* PhD thesis University of Exeter.

Stockwell, D. (2006) *Niche modelling: predictions from statistical distributions.* CRC Press, Boca Raton, Florida. 201pp.

Turchin, P. (2003) *Complex Population Dynamics A Theoretical/Empirical Synthesis*. Princeton University Press. 472pp.

Yates, K.L. et al. (2018) Outstanding Challenges in the Transferability of Ecological Models, *TREE*. 33: 790-802.

Zanette, L.Y., White, A.F., Allen, M.C. and Clinchy, M. (2011) Perceived predation risk reduces the number of offspring songbirds  produce per year. *Science*. 334: 1398-1401.

## Figure legends

**Figure 1:** The six building blocks of the Habitats-To-Populations framework (HTP, modified from Matthiopoulos et al. 2015) can be divided into categories: Three steps (in blue background) modelling habitat use, and three steps (in white background) accounting for the effects of habitat on population growth. The environment can be characterised by the availability of different habitat types (Step 1). Animals have habitat preferences (Step 2) which may lead them to use habitats disproportionately to their availability. Together, the combination of habitat availability and habitat preferences are assumed to give rise to the observed spatial distribution of a population (Step 3). These first three steps comprise the Habitat Selection Model and they determine the exposure of individuals to different habitat types that influence their fitness (Step 4), which, in turn, determines the collective capability of a population to grow (Step 5). Processes of population growth determine current population density (Step 6), which has the opportunity to mould habitat availability but also the potential to feed back directly into population growth and habitat preferences (via density dependent and spatial crowding). The key advantage of this approach is that all six processes can be estimated from field data. Statistical inference for the steps in dark blue requires spatial data on habitat availability and use. The steps shown in white require additional temporal data on population change.

**Figure 2:** Analysis of the colony-specific habitat selection coefficients obtained via the GFR model, incorporating main effects and relevant interactions for each covariate. The histograms on the diagonal summarise the coefficient values obtained for each environmental covariate (green

curves are smoothed versions of the same information). The pairwise scatter-plots, off the diagonal show the correlations (in this example, all synergistic) between regression coefficients.

**Figure 3:** Goodness-of-fit and predictive ability for the spatial model (a and b, respectively) and for the mean-field model (c and d, respectively). The spatial model uses information extracted from models of habitat availability and use in each garden cell around the colonies. The mean field model only uses average availability. For the goodness of fit plots (a & c), the regression line and corresponding prediction intervals come from a linear model fitted to the observations against the population model fit. For the prediction plots, each dot represents an iteration of our leave-one-out cross-validation algorithm.

**Figure 4:** Results of the garden improvement optimisation. The black line segments show the improvements possible in each of our 12 colonies (sorted in order of low-to-high observed growth rate). The pie charts show the alterations in composition of the gardens around each colony in terms of six habitat categories (explained in the key), two of which (artificial surface and roof) were constrained to starting conditions and four of which (bush, grass, hedge and tree) were allowed to be varied to optimise population growth. The confidence intervals of the predicted growth rates are represented as the top and bottom vertices of the grey rhombuses surrounding each line segment.