

# **A spatio-temporal multi-species model of a semi-continuous response**

Charlotte M. Jones-Todd\*, Ben Swallow\*, Janine B. Illian\*, and Mike Toms\*\*

*\*Centre for Research into Ecological and Environmental Modelling, The Observatory, University of St Andrews, Scotland, KY16 9LZ*

*\*\*British Trust for Ornithology, The Nunnery, Thetford, IP24 2PU, UK.*

## **Summary.**

As accessible and potentially vulnerable species high up in the food chain, birds are often used as indicator species to highlight changes in ecosystems. This study focuses on multiple spatially dependent relationships between a raptor (sparrowhawk), a potential prey species (house sparrow) and a sympatric species (collared doves) in space and time.

We construct a complex spatio-temporal latent Gaussian model to incorporate both predator-prey and sympatric relationships, which is novel in two ways. First, different types of species interactions are represented by a shared spatio-temporal random effect, which extends existing approaches to multivariate spatial modelling through the use of a joint latent modelling approach. Second, we use a delta-gamma model to capture the semi-continuous nature of the data to jointly model the binary and continuous sections of the response.

The results indicate that sparrowhawks have a localised effect on the presence of house sparrows, which could indicate that house sparrows avoid sites where sparrowhawks are present.

*Keywords:* INLA, joint model, multivariate spatio-temporal modelling, SPDE approach.

## **1. Background—modelling multi-species coexistence**

The co-occurrence of different species in the same spatial location at the same time is referred to as coexistence. Many ecologists are interested in understanding the mechanisms that facilitate such coexistence. Recent decades have seen an increasing push to conserve and monitor biodiversity, both in Europe and world-wide (Pereira and Cooper, 2006). Birds are often used as an indicator species as they are more accessible for monitoring than other species, and are particularly vulnerable to changes in ecosystems due to their position

high up in the food chain (Bibby et al., 2000; Gregory and van Strien, 2010). In order to investigate long-term trends in bird species a number of monitoring schemes have been set up (Robinson et al., 2014). Studies that analyse data resulting from these schemes show a rather mixed picture for many songbirds in a UK context (e.g. Baillie et al., 2014). It is not well understood as to why some studies show a decline in certain bird species and others do not. In order to ascertain why this is the case it is vital to gain an improved understanding of the multivariate mechanisms that govern long-term population dynamics. This will help in the understanding of the ecological processes taking place and will have the potential to aid conservation efforts.

Songbirds are a group of species that have attracted particular interest from conservationists as well as from the general public. Specifically, the effect of increases in the abundance and distribution of some species of raptor on declines in songbirds has been the focus of much research interest (e.g. Newson et al., 2010; Bell et al., 2010). The interactions between predator and prey species are often much more complicated than generally perceived, and hence simple explanations (e.g., blaming raptors exclusively for songbird declines) are not sufficient (Newson et al., 2010). Thus, in order to gain a better understanding of the species dynamics of songbirds it is necessary to account for different types of interactions operating among several species in space and over time.

The majority of multi-species approaches usually explore either singularly, (i) a predator-prey relationship (Garneau et al., 2007), or (ii) co-occurrence of sympatric species (Schweiger et al., 2012). However, in nature these different types of relationships rarely occur in isolation, but instead operate simultaneously. Thus, to more realistically account for species dynamics we propose modelling different types of species interactions within the same spatio-temporal model. The framework we propose utilises shared spatio-temporal random effects that represent the interaction amongst species, which enable the relationships amongst species to be described.

### *Multi-type latent Gaussian models*

Bird species sharing similar habitats typically compete for similar resources, as do many other species. Thus, modelling individual species whilst ignoring any effects of this competition, and other relevant biotic interactions, oversimplifies the underlying dynamics of the process (Boulangéat et al., 2012). In some cases it may be appropriate to model the dynamics of one species whilst treating another as a fixed effect. However, in many instances this approach fails to propagate the uncertainty relating to each species correctly, making

parameter estimates and predictions unrealistic (Clark, 2003). In addition, this approach assumes some specific directionality, which implies that it is possible to identify a focal species. This may be sensible in some contexts where background knowledge may be used to justify this simplification, such as in Illian et al. (2009); Högmander and Särkkä (1999). However, in complex ecosystems the dependence relationships are rarely this simple (Dray et al., 2012). Hence, a multivariate approach should be employed in order that the relevant dependence structures may be correctly inferred. As such the framework we propose both accounts for species dependencies in space and over time, and uses latent structures to facilitate simultaneous modelling of several species.

Through employing such shared latent structures, we are able to capture the relevant dependencies inherent in the multi-species data detailed in Section 2.1 below. In particular, our approach utilises additional random fields that are infinitesimally close to the target model enabling us to assess the interaction amongst multiple species. This approach is essentially an extension of the ‘linked model’ proposed by Diggle and Milne (1983), which sets two latent fields proportional to one another. This corresponds to a class of multivariate GRFs similar to those termed Linear Models of Coregionalization (LMCs) (Gelfand et al., 2002). In addition to the assumed shared stochastic structures we consider the observation locations (sites) directly, thus avoiding the loss of information resulting from transforming spatially continuous data into lattice data.

Our motivating data are not only spatially but also temporally indexed. As such, the interest is not only in each species’ spatial distribution, but also in assessing how the spatial distribution changes over time. This allows for the possibility that the interaction between the species within and between trophic levels may vary both in space and time. Understanding these inter- and intra-specific interactions is imperative in understanding the complex ecological dynamics present that drive the coexistence or otherwise of species.

To incorporate all such dependencies into one modelling framework we construct a spatio-temporal multivariate Gaussian Random Field (GRF) with a Matérn covariance for the spatial domain, and an autoregressive process of order 1 (AR(1)) to describe the temporal dependence. We take an integrated nested Laplace approximation (INLA) approach (Rue et al., 2009) for Bayesian inference, coupled with a stochastic partial differential equation (SPDE) model (Lindgren et al., 2011) to account for the spatial autocorrelation.

The remainder of this article is structured as follows: Section 2 describes the approach taken, and its application to GBFS data discussed in 2.1; Section 2.2 defines the model specifics; Section 3 illustrates the inference to be drawn from such a model, and its ade-

quacy. Finally, Section 4 outlines other possible applications and extensions of the flexible spatio-temporal model class proposed in this article.

## 2. Methods & materials

The modelling framework we propose is motivated by long-term geostatistical garden bird data. We focus on three species of bird: a species of garden bird that has shown large declines in population recently, one of its potential predators, and a sympatric species. We use a complex hierarchical Bayesian model for this purpose, where spatio-temporal structures are modelled with an SPDE approach and fit the model with INLA. This enables us to simultaneously model and estimate the spatial correlation between, and amongst, the species. The INLA approach is designed to fit latent Gaussian models where the values at each location are conditionally independent, given the covariance structure. The conditional independence of the latent Gaussian field translates into a sparse precision matrix. The sparsity of this precision matrix leads to a Gaussian Markov Random Field (GMRF); this aids the computational efficiency of the INLA approach. However, as typically approximations of mechanisms operating on continuous domains are required in this context, it is beneficial to model the joint behaviour of this process for all locations. We therefore use a SPDE model for the latent field. The use of a SPDE as an approximation to a GMRF was introduced by Lindgren et al. (2011), whereby weighted sums of basis functions are used to approximate the spatial random functions arising from the solution to the considered SPDE. Thus, the continuous interpretation of space is preserved, whilst benefiting from the computational advantages arising from the discrete Markovian structures of GMRFs. Further details relating to the SPDE model and the model fitting procedure are given in the appendix; the authors also recommend that readers refer to Rue et al. (2009); Lindgren et al. (2011) for more in-depth details.

### 2.1. *The motivating garden bird feeding survey data*

The modelling framework we propose is illustrated through the use of data collected by the Garden Bird Feeding Survey (GBFS) undertaken by the British Trust for Ornithology (BTO)†. We consider count data collected at approximately 200 sites per year across the UK over a 36 year period, 1970–2005. The data collected are the maximum count of each species seen feeding on provisioned food in the surveyed gardens. These counts are noted in each of up to 26 weeks every winter spanning the months October–March. We then

†<http://www.bto.org/volunteer-surveys/gbfs>

calculate the annual averages across weeks giving a mean of weekly maxima for each site-year combination. It should be noted therefore that the data relate to what is assumed to be a continuous variable, with a non-zero probability of obtaining exact zeros. Such data are sometimes referred to as a semi-continuous variable (Aitchison, 1955). Previous modelling approaches for such semi-continuous data introduce the use of a delta-gamma model. This jointly models the binary and continuous sections of the response (Foster and Bravington, 2013). Due to the nature of our data we propose utilising such a delta-gamma model in our multi-species modelling framework discussed in the following sections. In the following we use the term ‘density’ to refer to the assumed continuous variables mentioned above.

There is severe concern as to the causes of declines observed in many wild bird species over the past 40 years. Amongst the possible causes is an increase in the abundance and distribution of the Eurasian sparrowhawk *Accipiter wisps*, an avian predator that has increased in both abundance and distribution over a similar time period (Newton, 1986). The house sparrow *Passer domesticus* is of particular concern as it has decreased by approximately 60% over the past 40 years (Robinson et al., 2005). Previous work by Götmark and Andersson (2005) suggests that house sparrows may be at particular risk from sparrowhawk predation due to their decreasing population (Chamberlain et al., 2009), and hence we concentrate particularly on these species in this paper.

In addition, we include a third species in our analysis, the collared dove *Streptopelia decaocto*, to assess the shared effect between their spatial distributions over time and that of house sparrows. Collared doves have colonised the UK in a similar time frame but in the reverse direction to sparrowhawks, so act as a natural control species for testing potential effects of predators (Thomson et al., 1998; Newson et al., 2010; Swallow et al., 2016a). Adding collared dove to the joint model can ensure that spurious negative correlations are not occurring, whilst also testing for positive associations between species that have similar ecological requirements.

We are interested in assessing if there is spatial and temporal interaction between sparrowhawk increase and house sparrow decrease, which would be consistent with the idea of a causal relationship between the two. For example, a negative relationship between the occurrence of raptors and a prey species may suggest that the occurrence of the former is causing a reduction in the occurrence of the latter: either directly through predation, or as a behavioural response of the later to avoid areas where predators are abundant. In either case it can highlight relationships that may warrant further attention. A positive

correlation between the latent variables of two sympatric species would similarly suggest that the species are attracted to the same areas due to some unmeasured factor.

## 2.2. A multi-species spatio-temporal model

A delta gamma model is used to represent the biomass of each avian species. Let  $z_{ik}$  be a binary indicator of the  $k^{\text{th}}$  species' presence ( $k = 1, 2, 3 =$  sparrowhawk, collared dove, house sparrow) at site  $i$ . Then  $z_{ik} \sim \text{Bernoulli}(p_{ik})$ , where  $p_{ik}$  is the probability of presence of the  $k^{\text{th}}$  species at site  $i$ . Letting  $d_{ik}$  be the density of the  $k^{\text{th}}$  species at location  $i$ , then  $d_{ik}$  is given by,

$$d_{ik} = \begin{cases} \text{Gamma}(a_{ik}, b_{ik}) & \text{with probability } p_{ik} \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

with shape and scale parameters  $(a_{ik}, b_{ik})$  respectively, so that  $E[d_k] = a_k b_k = \mu_k$ .

The binary components of the response reveal at which site each species is most likely to appear, whereas the non-zero component reveal the spatial distribution of the abundance for each species. This facilitates accounting for both the remaining spatial auto-correlation and dependence among the species. Utilising multiple latent stochastic structures in the modelling procedure leads to the linear predictors for each species being constructed as,

$$\begin{aligned} \text{sparrowhawk} & \begin{cases} \text{logit}(p_{i1}) = \mathbf{x}_1\{\mathbf{s}_i, t\} \\ \log(\mu_{i1}) = \beta_1 \mathbf{x}_1\{\mathbf{s}_i, t\}, \end{cases} \\ \text{collared dove} & \begin{cases} \text{logit}(p_{i2}) = \mathbf{x}_2\{\mathbf{s}_i, t\} \\ \log(\mu_{i2}) = \beta_2 \mathbf{x}_2\{\mathbf{s}_i, t\}, \end{cases} \\ \text{house sparrow} & \begin{cases} \text{logit}(p_{i3}) = \alpha + \gamma_1 \mathbf{x}_1\{\mathbf{s}_i, t\} + \gamma_2 \mathbf{x}_2\{\mathbf{s}_i, t\} + \mathbf{x}_3\{\mathbf{s}_i, t\} \\ \log(\mu_{i3}) = \alpha_y + \gamma_3 \mathbf{x}_1\{\mathbf{s}_i, t\} + \gamma_4 \mathbf{x}_2\{\mathbf{s}_i, t\} + \beta_3 \mathbf{x}_3\{\mathbf{s}_i, t\}. \end{cases} \end{aligned} \quad (2)$$

Here each  $\mathbf{x}_j\{\mathbf{s}_i, t\}$  ( $j = 1, 2, 3$ ) is a spatio-temporal random effect modelled by a SPDE model (Lindgren et al., 2011), which follows an AR(1) process over time with parameter  $\rho_i$  (see appendix). Each  $\alpha$  is an intercept term for each component of the model referring to the house sparrows. The parameters  $\beta$  and  $\gamma$  are scaling (Blangiardo et al., 2013 Chapter 8) or 'interaction' parameters to the spatio-temporal random fields of which they are coefficients. That is, each shared random field (i.e., a random field that appears in more than one linear predictor) represents the shared inter- or intra-species spatial auto-correlation over time. Although each  $\beta$  and  $\gamma$  have the same mathematical interpretation we term each  $\beta$  a scaling parameter and each  $\gamma$  an interaction parameter. This is done in order to differentiate between species and amongst species interactions. Each  $\beta$  or  $\gamma$  parameter

represents the magnitude and direction of this spatial similarity. For example, Equation (2) assumes that  $\beta_1$  multiplied by  $\text{logit}(p_{i1})$  is equal to  $\log(\mu_{i1})$  etc..

Following the construction of the joint spatio-temporal model mentioned above below we summarise the interpretation of each stochastic structure. Each field may be thought of as follows:

- $\mathbf{x}_1\{\mathbf{s}_i, t\}$ , the spatially varying process referring to the probability of presence of sparrowhawk, which is also assumed to be a contributory factor to the spatially varying density of sparrowhawk over the time period. As such, one would expect  $\beta_1$  to be positive;
- $\mathbf{x}_2\{\mathbf{s}_i, t\}$ , the spatially varying process referring to the probability of presence of collared dove, which is also assumed to be a contributory factor to the spatially varying density of collared dove over the time period. As such, one would expect  $\beta_2$  to be positive;
- $\mathbf{x}_3\{\mathbf{s}_i, t\}$ , the spatially varying process referring to the probability of presence of house sparrow—conditional on the spatial distribution of both sparrowhawk and collared dove—which is also assumed to be a contributory factor to the spatially varying density of house sparrow over the time period. One would expect  $\beta_3$  to be positive. The field  $\mathbf{x}_3\{\mathbf{s}_i, t\}$  accounts for any remaining structure in the spatial distribution of house sparrows once it has been ascertained whether either the prey species—sparrowhawk—or sympatric species—collared dove—are considered to have an effect on the spatial distribution of house sparrow. Thus, if the spatial distribution of house sparrow were independent of either that of sparrowhawk or collared dove  $\mathbf{x}_3\{\mathbf{s}_i, t\}$  would account for the variation specific to the spatial distribution of house sparrow.

### 3. Results and inference

A wrapper function to fit the model given in Equation (2), along with the variants discussed below, can be found at <https://github.com/jt/cmjt/lgcp>. Further details regarding the modelling procedure are given in the appendix along with the prior specifications for the parameters of the spatio-temporal fields.

Table 1 gives the posterior means, standard errors, and 95% credible intervals, for the parameters of the joint spatio-temporal model given by Equation 2. Each AR(1) parameter,  $\rho_i$  ( $i = 1, 2, 3$ ), is close to one indicating a strong short-term dependence between the latent fields across years. Recall that each  $\beta_i$  parameter represents, on the link scale, the

relationship between the probability of presence and density of the sparrowhawks, collared doves, and house sparrows respectively. In the case of each species this scaling parameter is estimated to be positive, and the corresponding 95% credible intervals do not contain zero. This indicates a positive relationship between species presence and density. This parameter is essentially the ratio between the log density and logit probability of presence, which indicates that where there is a high probability of observing a species there are also more birds of that species. That is, the occurrence and density of birds of the same species share the same spatial patterning across the UK. In the case of sparrowhawks  $\hat{\beta}_1 = 0.5396$ , which indicates that the log density of sparrowhawks at any spatial location is approximately equal to twice the logit of the probability of observing them. Thus, if the probability of observing a sparrow hawk were 0.5 at any particular location, then the estimated density (at that location) of sparrow hawks would be  $\exp(0.5396) \sim 1.7$  birds. With respect to the collared dove  $\hat{\beta}_2 = 0.1697$  indicating that if at a particular location one was equally likely to observe or not a collared dove, the density of collared doves would be  $\exp(0.1697) \sim 1.18$ . In the case of the house sparrows this parameter has the same interpretation as above given the relationship to the other two species.

The intercept parameters  $\alpha$  and  $\alpha_y$  relate to the density and probability of presence of the house sparrow respectively. In each case if there were no spatio-temporal variation these parameters represent the fixed average value, on the link scale, of density and probability of presence.

Of most interest in our modelling framework are the interaction parameters represented by each  $\gamma_i$  ( $i = 1, 2, 3, 4$ ). These parameters represent the inter-species relationships. Both  $\gamma_2$  and  $\gamma_4$  are estimated to be positive, with 95% credible intervals that do not contain zero. This leads us to infer that the presence of collared doves is positively related to both the presence and density of house sparrows (i.e., areas of high collared dove occurrence relate to areas of high housesparrow occurrence and abundance). This is not surprising as they are known to be sympatric species. In particular, if at any particular location the probability of observing a collared dove is 0.5 then the density of house sparrows would be  $\exp(1.1227) \sim 3$ , ignoring the other effects in the model. Along the same vein at locations where collared doves are as equally likely to be observed as not then the probability of observing a house sparrow would be  $\text{logit}^{-1}(0.6047) \sim 0.65$ , given the other components of the linear predictor.

Interaction parameters  $\gamma_1$  and  $\gamma_3$  are estimated to be negative, with 95% credible intervals that do not contain zero. In contrast to the relationship between house sparrows



and collared doves there is a negative relationship between the presence of sparrowhawks and house sparrows. The occurrence of sparrow hawks does not share the same spatial structure as either the occurrence or density of housesparrows. This is not unsurprising as these species are predators and prey respectively. These parameter values indicate that if at any particular location the probability of observing a sparrow hawk is 0.5 then the density of house sparrows would be  $\exp(-0.1451) \sim 0.86$ , ignoring the other effects in the model. Along the same vein at locations where sparrow hawks are as equally likely to be observed as not then the probability of observing a house sparrow would be less than chance (i.e.,  $\text{logit}^{-1}(-0.8781) \sim 0.29$ ), given the other components of the linear predictor.

Figures 1 and 2 show the estimated response of each component of the proposed spatio-temporal delta-gamma model along with the inter-species interaction over the time period. Figure 1 shows the estimated probability maps for each species in 1970 (top row) and 2005 (bottom row). The maps, from left to right in each case, show the estimated probability of observing sparrowhawks, collared doves, and house sparrows respectively. In each case, areas of high probability of observing house sparrows correspond to areas of low probability of observing sparrowhawks, and high probability of observing collared doves. These spatial differences correspond to the predator prey and sympatric species relationships respectively. The central plot shows the mean probability of house sparrow presence from 1970–2005 in relation to each species effect. We can see that on average the probability of observing house sparrows declines over the time period (see solid line in Figure 1). In addition, given the other model components, the probability of house sparrow presence declines with respect to the presence of sparrowhawks (dot-dash line Figure 1) and remains pretty much constant with respect to collared dove presence (dotted line Figure 1). The remainder of the spatial and temporally varying effect of house sparrows (i.e., the species specific effect) is shown by the dashed line. This seemingly remains constant over the time period. It should be noted that the values plotted are average values of the random effects on the response scale in each year, the fixed effect has been omitted so that these effects can be seen in isolation.

In a similar vein to Figure 1 this Figure 2 shows the density component of the model for each of the species considered. The density maps relate again to 1970 (top row) and 2005 (bottom row), and show on the response scale the estimated density of sparrowhawks, collared dove, and house sparrows—from left to right respectively. The central plot shows the average values of the random effects in each year thought to contribute to the spatial distribution of house sparrows—note the fixed effect has been omitted. The solid line

illustrates the combination of the other three (i.e., it is the overall estimated house sparrow mean density in each year). This shows a decline over the period and mirrors most closely the decline in the number of collared doves (dotted line). Despite the positive value of  $\hat{\gamma}_4$ , the contribution of the collared dove effect to the density of house sparrows is seemingly negative. This is due to the estimated decline in the numbers of collared doves over that time suggesting there may be similar drivers of population change in the two species; it should be noted that the positive  $\hat{\gamma}_4$  indicated concordance between these two sympatric species. The decline in house sparrow numbers is illustrated by the downward slope of the dot-dash line; however, perhaps this is best illustrated by the maps where areas of high house sparrow density correspond to areas of low sparrowhawk density. Further details regarding model construction are given in the appendix.

The results presented from the modelling approach applied here detected a negative relationship between the probability of sparrowhawk presence and the probability of house sparrow presence, as well as between the probability of sparrowhawk presence and the density of house sparrows. This is consistent with the idea of a causal relationship, either directly on the populations of house sparrows due to predation, or a behavioural response of house sparrows avoiding sites that are frequented by sparrowhawks.

### *3.1. Model checking*

It is imperative that the model variables be assessed as to whether they contribute to the inference. However, assessing the validity of the assumptions of our proposed model is not a simple task and model assessment methods for complex spatio-temporal models are still being developed (Martins et al., 2014; Simpson et al., 2014; Yuan et al., 2017). One inbuilt method for model checking offered by `inla` is an approach based on leave one out cross-validation, see Blangiardo and Cameletti (2015, Chapter 5) and Rue et al. (2009); Held et al. (2010). In summary this approach aims to check the plausibility of the model assumptions through splitting the data into two groups; the assumed model is then fitted to one group of data, and the second group is used to calculate goodness of fit indices. The index used, for the assumed continuous response, in this article is the probability integral transform (PIT) (Dawid, 1984; Czado et al., 2009). This is an informal way to assess model fit, and can only be used as an indication of model fit.

The focus of our model is on the relationship between three species. In particular, we focus on the relationship between (i) sparrowhawks and house sparrows (a predator-prey scenario), and (ii) collared doves and house sparrows (two sympatric species). To assess

the suitability of this joint model we compute PIT values for the density component of the model, these are seen in histogram in Figure 3. This histogram should roughly follow a uniform distribution if the predictive distribution is in line with the data (Gneiting et al., 2007). The histogram in Figure 3 indicates a reasonable fit for the density component of the model. To assess the fit of the Bernoulli component of the model we may compare the estimated values of the joint model to those of a single model where no inter-species relationships are assumed. The percentage difference between the estimated values of the joint model and the 'null' model and the 95% quantiles for each year are shown in the right hand plot of Figure 3. One may consider this to indicate a reasonable fit, however it is evident that towards the latter end of the time period the difference between models increases. This may suggest that the effect of sparrowhawks was initially high but then reached an equilibrium.

#### 4. Discussion

The estimated negative relationship between the probability of sparrowhawk presence and both the density and probability of observing house sparrows detected in this article may equate to sparrowhawks having a localised effect on the presence of house sparrows, or causing house sparrows to avoid sites where sparrowhawks are present. However, the recolonisation of sparrowhawks across the UK has failed to cause extinction of their prey.

Perrins and Geer (1980) and Newton (1986) studied the effects of sparrowhawks on breeding densities of blue tits *Cyanistes caeruleus* and great tits *Parus major* and found no difference between years when sparrowhawks were present and those where they were absent. Their results suggested that sparrowhawks merely reduced the magnitude of the peak in post-breeding numbers, the seasonal pattern of mortality and the means by which the prey species were removed from the population (Newton, 1998). The results from this analysis are consistent with the idea that sparrowhawks are reducing post-breeding peaks in prey numbers as the data come from a survey that monitors these peaks. However, they also suggest that there is no evidence of a discernible effect on overall site occupation. Most previous analyses of sparrowhawks on breeding density of songbirds have failed to find any significant widespread effect (Newson et al., 2010) but Bell et al. (2010) and Swallow et al. (2016a) both found negative effects of sparrowhawks on house sparrows when using the same GBFS data. The sparrowhawk effects detected and outlined above, are not necessarily indicative of a causal effect of sparrowhawks on house sparrow presence. We do note that it is possible that the negative correlation between the two species could

be driven by independent and concurrent factors that have not been explicitly modelled here. Yet, an additional random effect is included relating to house sparrows to allow the associated variables to vary on their own (not tied to processes governing the other species). Our results do suggest, however, that the spatial structure of house sparrows has evolved over the last 36 years inversely to that of sparrowhawks, although it is impossible to confirm a causal relationship with observational data such as these. In contrast, positive correlation was found between house sparrow and collared dove. These two species have similar ecological requirements and the positive relationship reflects that the two species might respond similarly to the same environmental stimuli. We would therefore expect the spatial distribution of the two species through time to reflect this. From an ecological point of view our work exhibits the ability to effectively capture the complexity often inherent in multi-species data, and infer about the relationships and processes fundamental to the data structure.

The model discussed here has enabled us to simultaneously assess the spatio-temporal structure inherent in zero-inflated data relating to counts of house sparrows obtained from the GBFS over a 36 year period. Both the relationship between sympatric and predator-prey species are accounted for in a joint model of several species. It is now widely accepted that multi-species interactions are highly relevant in ecology, in particular in the context of changes in the spatial distribution of species over time. Previous analyses have tended to concentrate on single species independently (e.g. Newson et al., 2010) and small spatial regions (e.g. Lahoz-Monfort et al., 2011). The method outlined here accounts for spatial correlation across the whole of the UK, with uncertainty directly accounted for depending on the density of monitored sites locally. Our methodology allows us to infer different types of inter-specific interactions and processes inherent in such data. The results presented here may suggest positive future directions for further research into these multi-species dynamics. The use of shared random effects for understanding synchrony in multi-species data sets has been explored previously by Lahoz-Monfort et al. (2011) and Swallow et al. (2016b). The method outlined here further extends this synchrony across both space and time, allowing much greater understanding of the the complex correlation structures underpinning both intra- and inter-specific dynamics.

From a broader perspective, we combine spatio-temporal work done by Cameletti et al. (2013) and joint modelling work carried out by Illian et al. (2012) in the context of point process modelling. This enables us to infer the spatio-temporal effect, as well as estimate multiple response variables and assess the relationships among the response variables. We

used the computationally efficient method INLA for model fitting, suitable for latent Gaussian models (Rue et al., 2009), exploiting its computational efficiency to avoid prohibitive running times in the context of realistically complex spatio-temporal models (Cameletti et al., 2013; Blangiardo and Cameletti, 2015). The computational advantages of both INLA (Simpson et al., 2011) and the SPDE approach facilitate the incorporation of the complexity intrinsic in many practically relevant data sets. Using the methodology discussed here allows one to infer about processes which operate over both space and time as well as the dependency among processes operating simultaneously in one combined model. The flexibility and generality of model fitting with the INLA-SPDE approach and the associated software *R-INLA* allows us to view the model discussed here as just one example of a whole host of similar spatio-temporal joint models. For instance, different types of response variables may be considered, such as spatio-temporal log-Gaussian Cox processes as considered in Yuan et al. (2017). Similarly, models for data on a larger spatial scale such a global data may be fitted directly on the surface of the earth without the need for a projection into two-dimensional space as applied in Python et al. (2016) to model global terrorism in space and time. Using the joint modelling approach these models may be extended to a multi-species or to a multi-event situation.

In addition to being applicable elsewhere, the approach taken here can also be extended to provide more flexible models. For instance, in the current paper we assume that a stationary latent Gaussian model is appropriate for our data. Future extensions of our methodology may generalise our approach to non-stationary SPDE models which allow the covariance structure to vary spatially. (Bolin and Lindgren, 2011; Ingebrigtsen et al., 2014; Fuglstad et al., 2013). This might be particularly relevant in strongly patchy environments, and in the presence of physical barriers in space.

In summary we have successfully demonstrated that the type of model we propose and the methodology we use to fit such a model enabled us to simultaneously account for the latent spatio-temporal process inherent in many real-world data-sets. We concentrate here on fitting such models to complex multi-species data illustrating an extremely flexible class of spatio-temporal models, which can easily be extended to offer insight into processes operating in a broad range of fields. Not only can our methodology be applied to geostatistical data (i.e., where we fit a spatially continuous model to measurements taken at a finite number of locations), but where the interest may be to analyse the spatial pattern formed by individuals or events in space (i.e., point pattern data). In conclusion, such methodology is applicable to many complicated spatial/spatio-temporal data structures and due to

the generality of the INLA and SPDE framework can be flexibly extended.

## Tables &amp; Figures

Table 1: Posterior means, standard errors, and 95% credible intervals, for the parameters of the joint spatio-temporal model in Equation 2.

	Parameter	Mean	SE	Quantile of Order	
				2.5%	97.5%
<b>sparrowhawk</b>	$\rho_1$	0.9800	0.0001	0.9798	0.9801
	$\beta_1$	0.5396	0.0019	0.5363	0.5438
<b>collared dove</b>	$\rho_2$	0.9612	0.0004	0.9604	0.9620
	$\beta_2$	0.1697	0.0018	0.1670	0.1738
<b>house sparrow</b>	$\alpha$	1.2629	0.0574	1.1503	1.3755
	$\alpha_y$	1.2657	0.0289	1.2090	1.3224
	$\rho_3$	0.9683	0.0002	0.9680	0.9687
	$\gamma_1$	-0.8781	0.0059	-0.8872	-0.8647
	$\gamma_2$	0.6047	0.0036	0.5974	0.6114
	$\gamma_3$	-0.1451	0.0049	-0.1531	-0.1342
	$\gamma_4$	1.1227	0.0028	1.1170	1.1281
	$\beta_3$	0.5006	0.0036	0.4942	0.5082

## References

- Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, 50(271):901–908.
- Baillie, S. R., Marchant, J. H., Leech, D. I., Massimino, D., Sullivan, M. J. P., Eglington, S. M., Barimore, C., Dadam, D., Downie, I. S., Harris, S. J., Kew, A. J., Newson, S. E., Noble, D. G., Risely, K., and Robinson, R. (2014). Birdtrends 2014: trends in numbers, breeding success and survival for UK breeding birds. Technical report, BTO Research Report No 662, BTO, Thetford. (<http://www.bto.org/birdtrends>).
- Bell, C. P., Baker, S. W., Parkes, N. G., Brooke, M. D. L., and Chamberlain, D. E. (2010). The role of the eurasian sparrowhawk (*Accipiter nisus*) in the decline of the house sparrow (*Passer domesticus*) in Britain. *The Auk*, 127(2):411–420.
- Bibby, C., Burgess, N., Hill, D., and Mustoe, S. (2000). *Bird Census Techniques*. Academic Press, London, 2nd edition.
- Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian Models with R-INLA*. John Wiley & Sons, Chichester.
- Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and spatio-temporal epidemiology*, 7:39–55.
- Bolin, D. and Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, 5(1):523–550.
- Boulangéat, I., Gravel, D., and Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, 15(6):584–593.
- Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97(2):109–131.
- Chamberlain, D., Glue, D., and Toms, M. (2009). Sparrowhawk accipiter nisus presence and winter bird abundance. *Journal of Ornithology*, 150(1):247–254.
- Clark, J. S. (2003). Uncertainty and variability in demography and population growth: a hierarchical approach. *Ecology*, 84(6):1370–1381.
- Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261.





Fig. 1: Top row from left to right shows the estimated probability of presence of sparrowhawks, collared doves, and house sparrows in 1970; the bottom row shows these probabilities for 2005. The central panel shows the mean contribution of each species specific random effect to the probability of observing house sparrows over the time period. The solid line is the average value in each year of the probability of observing a house sparrow. The dotted line indicates the average contribution of the process governing the probability of presence of collared doves to the probability of presence of house sparrows. The dot-dash line indicates the average contribution of the process governing the probability of presence of sparrowhawks to the probability of presence of house sparrows. The dashed line indicates the average of the house sparrow specific random effect in each year.

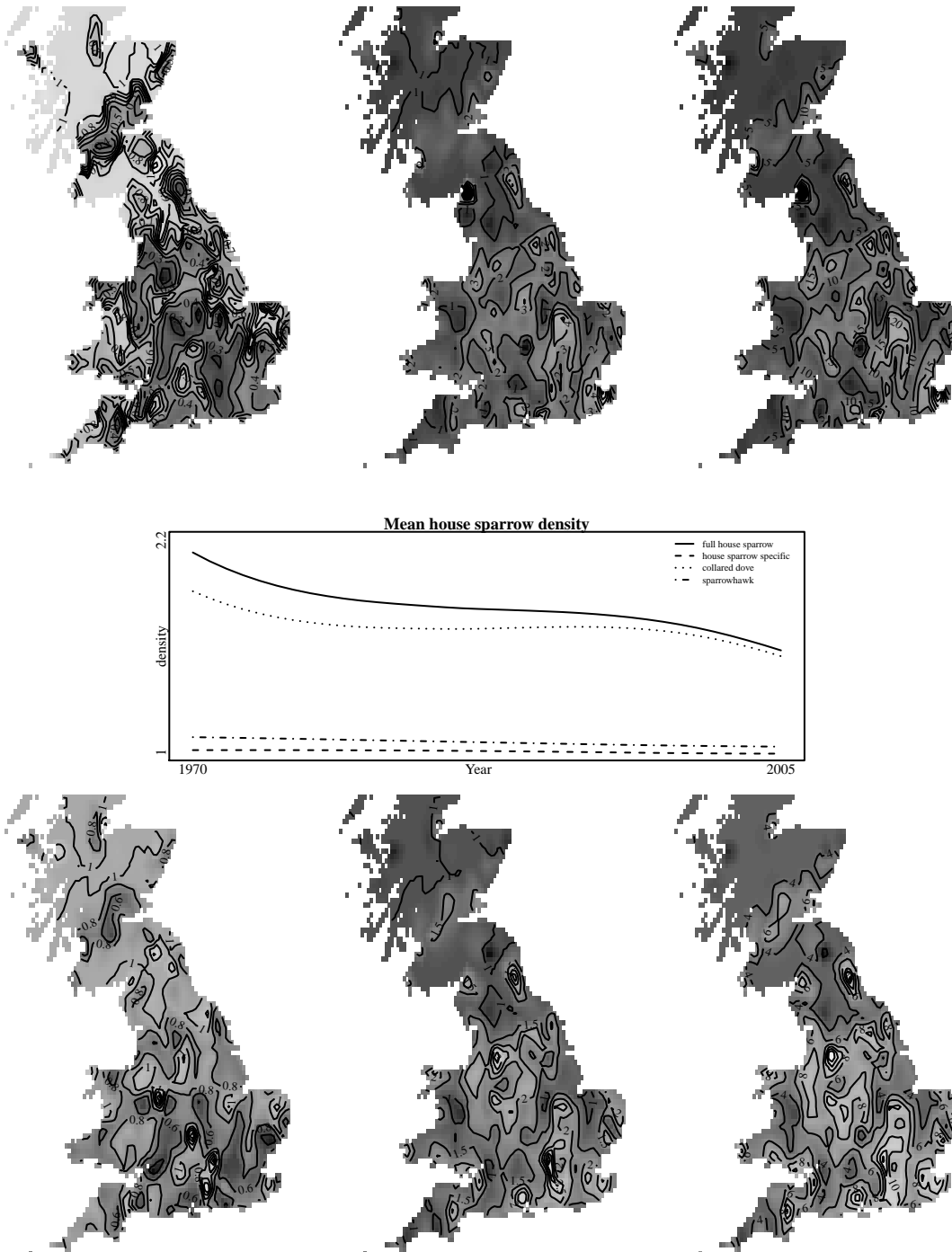


Fig. 2: Top row from left to right shows the estimated density of sparrowhawks, collared doves, and house sparrows in 1970; the bottom row shows these densities for 2005. The central panel shows the mean contribution of each species specific random effect to the density of house sparrows over the time period. The solid line is the average value in each year of house sparrow density. The dotted line indicates the average contribution of the process governing the probability of presence of collared doves to the spatial distribution of house sparrows. The dot-dash line indicates the average contribution of the process governing the probability of presence of sparrowhawks to the density of house sparrows. The dashed line indicates the average of the house sparrow specific random effect in each year.

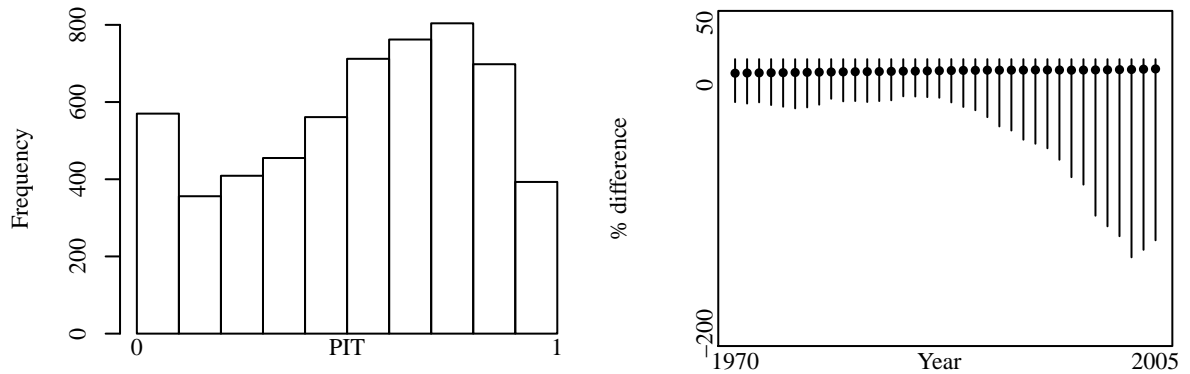


Fig. 3: Left: histogram of the cross-validated PIT values for the non-zero density house sparrow component of the joint model. Right: plot showing the percentage difference between the joint model component relating to the binary presence of house sparrows to the single likelihood model fitted. Lines indicate the 95% quantiles of the percentage difference.

Dawid, A. P. (1984). Statistical theory: the prequential approach. *Journal of the Royal Statistical Society. Series A*, 14(2):278–292.

Diggle, P. J. and Milne, R. K. (1983). Bivariate Cox processes: some models for bivariate spatial point patterns. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 45(1):11–21.

Dray, S., Péliissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P., Bellier, E., Bivand, R., Blanchet, F. G., De Cáceres, M., Dunfou, A., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J., and Wager, H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, 82(3):257–275.

Foster, S. D. and Bravington, M. V. (2013). A Poisson–gamma model for analysis of ecological non-negative continuous data. *Environmental and Ecological Statistics*, 20(4):533–552.

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2013). Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. *arXiv preprint arXiv:1304.6949*.

Garneau, D. E., Post, E., Boudreau, T., Keech, M., and Valkenburg, P. (2007). Spatio-temporal patterns of predation among three sympatric predators in a single-prey system. *Wildlife Biology*, 13(2):186–194.

Gelfand, A. E., Schmidt, A. M., and Sirmans, C. (2002). Multivariate spatial process models:

- conditional and unconditional Bayesian approaches using coregionalization. *Center for Real Estate and Urban Economic Studies, University of Connecticut*.
- Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268.
- Götmark, F. and Andersson, M. (2005). Predation by sparrowhawks decreases with increased breeding density in a songbird, the great tit. *Oecologia*, 142(2):177–183.
- Gregory, R. D. and van Strien, A. (2010). Wild bird indicators: using composite population trends of birds as measures of environmental health. *Ornithological Science*, 9(1):3–22.
- Held, L., Schrödle, B., and Rue, H. (2010). Posterior and cross-validated predictive checks: a comparison of MCMC and INLA. In *Statistical modelling and regression structures*, pages 91–110. Springer, New York.
- Högmander, H. and Särkkä, A. (1999). Multitype spatial point patterns with hierarchical interactions. *Biometrics*, 55(4):1051–1058.
- Illian, J. B., Møller, J., and Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, 16(3):389–405.
- Illian, J. B., Sørbye, S. H., and Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, 6(4):1499–1530.
- Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8(1):20–38.
- Lahoz-Monfort, J. J., Morgan, B. J., Harris, M. P., Wanless, S., and Freeman, S. N. (2011). A capture–recapture model for exploring multi-species synchrony in survival. *Methods in Ecology and Evolution*, 2(1):116–124.
- Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.
- Martins, T. G., Simpson, D. P., Riebler, A., Rue, H., and Sørbye, S. H. (2014). Penalising model component complexity: a principled, practical approach to constructing priors. *arXiv preprint arXiv:1403.4630*.

- Newson, S. E., Røst, E. A., Baillie, S. R., Buckland, S. T., and Aebischer, N. J. (2010). Population change of avian predators and grey squirrels in England: is there evidence for an impact on avian prey populations? *Journal of Applied Ecology*, 47(2):244–252.
- Newton, I. (1986). *The Sparrowhawk*. Poyser, London.
- Newton, I. (1998). *Population Limitation in Birds*. Academic Press, London.
- Pereira, H. M. and Cooper, H. D. (2006). Towards the global monitoring of biodiversity change. *Trends in Ecology & Evolution*, 21(3):123 – 129.
- Perrins, C. M. and Geer, T. A. (1980). The effect of sparrowhawks on tit populations. *Ardea*, 68:133–142.
- Python, A., Illian, J., Jones-Todd, C. M., and Blángiardo, M. (2016). A Bayesian approach to modelling fine-scale spatial dynamics of non-state terrorism: world study, 2002-2013.
- Robinson, R. A., Morrison, C. A., and Baillie, S. R. (2014). Integrating demographic data: towards a framework for monitoring wildlife populations at large spatial scales. *Methods in Ecology and Evolution*, 5(12):1361–1372.
- Robinson, R. A., Siriwardena, G. M., and Crick, H. Q. P. (2005). Size and trends of the house sparrow *Passer domesticus* population in Great Britain. *Ibis*, 147(3):552–562.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- Schweiger, O., Heikkinen, R. K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J., and Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, 21(1):88–99.
- Simpson, D., Lindgren, F., and Rue, H. (2011). Fast approximate inference with INLA: the past, the present and the future. *arXiv preprint arXiv:1105.2982*.
- Simpson, D. P., Martins, T. G., Riebler, A., Fuglstad, G.-A., Rue, H., and Sørbye, S. H. (2014). Penalising model component complexity: a principled, practical approach to constructing priors. *arXiv preprint arXiv:1403.4630*.
- Swallow, B., Buckland, S. T., King, R., and Toms, M. P. (2016a). Bayesian hierarchical modelling of continuous non-negative longitudinal data with a spike at zero: an application to a study of birds visiting gardens in winter. *Biometrical Journal*, 58(2):357.
- Swallow, B., King, R., Buckland, S. T., and Toms, M. P. (2016b). Identifying multispecies synchrony in response to environmental covariates. *Ecology and Evolution*, 6(23):8515–8525.

Thomson, D. L., Green, R. E., Gregory, R. D., and Baillie, S. R. (1998). The widespread declines of songbirds in rural Britain do not correlate with the spread of their avian predators. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1410):2057–2062.

Yuan, Y., Bachl, F., Lindgren, F., Brochers, D., Illian, J., Buckland, S., Rue, H., and Gerrodette, T. (2017). Point process models for spatio-temporal distance sampling data. *arXiv preprint arXiv:1604.06013*. to appear in *Annals of Applied Statistics*.