Check for updates

Received: 1 October 2021 | Accepted: 23 June 2022

DOI: 10.1111/2041-210X.13957

Methods in Ecology and Evolution

RESEARCH ARTICLE

Realising the Promise of Large Data and Complex Models

Balancing structural complexity with ecological insight in Spatio-temporal species distribution models

Megan R. Laxton[1] | Óscar Rodríguez de Rivera[2] | Andrea Soriano-Redondo[3,4] | Janine B. Illian[1]

[1]School of Mathematics and Statistics, University of Glasgow, Glasgow, UK

[2]Statistical Ecology at Kent (SE@K), School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK

[3]Helsinki Lab of Interdisciplinary Conservation Science, Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

[4]Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland

Correspondence
Megan R. Laxton
Email: megan.laxton@glasgow.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: 00713939

Handling Editor: Ruth King

Abstract

1. The potential for statistical complexity in species distribution models (SDMs) has greatly increased with advances in computational power. Structurally complex models provide the flexibility to analyse intricate ecological systems and realistically messy data, but can be difficult to interpret, reducing their practical impact. Founding model complexity in ecological theory can improve insight gained from SDMs.
2. Here, we evaluate a marked point process approach, which uses multiple Gaussian random fields to represent population dynamics of the Eurasian crane *Grus grus* in a spatio-temporal species distribution model. We discuss the role of model components and their impacts on predictions, in comparison with a simpler binomial presence/absence approach. Inference is carried out using Integrated Nested Laplace Approximation (INLA) with inlabru, an accessible and computationally efficient approach for Bayesian hierarchical modelling, which is not yet widely used in SDMs.
3. Using the marked point process approach, crane distribution was predicted to be dependent on the density of suitable habitat patches, as well as close to observations of the existing population. This demonstrates the advantage of complex model components in accounting for spatio-temporal population dynamics (such as habitat preferences and dispersal limitations) that are not explained by environmental variables. However, including an AR1 temporal correlation structure in the models resulted in unrealistic predictions of species distribution; highlighting the need for careful consideration when determining the level of model complexity.
4. Increasing model complexity, with careful evaluation of the effects of additional model components, can provide a more realistic representation of a system, which is of particular importance for a practical and impact-focused discipline such as ecology (though these methods extend to applications for a wide range of systems). Founding complexity in contextual theory is not only fundamental to maintaining model interpretability but can be a useful approach to improving insight gained from model outputs.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

KEYWORDS
INLA, marked point process, spatio-temporal model, species distribution model

1 | INTRODUCTION

The continuing increase and the improvement both of the availability and detail of ecological information, and of computational resources allows realistically complex and flexible statistical models to be fitted to ecological data. However, increasing structural complexity through the inclusion of additional model components without ecological justification will not improve our ability to understand fundamental ecological responses and predictions. Hence increased model complexity is only relevant if the more complex models provide us with additional insight into ecological processes than simpler and often computationally cheaper models (Bolker, 2009). For instance, more complex models may improve the accuracy of predictions or abundance estimates: they may have less restrictive assumptions, thus avoiding incorrect inference and model interpretation, and they may have the ability to explicitly reflect complex ecological processes.

In the specific context of species distribution models, which seek to provide insight into the distribution of species in geographic and environmental space, complex spatio-temporal models are often utilised. Increased model complexity is supported by the fact that those complex models may reflect the complex ecological processes driving the species distribution, such as demography (Pagel & Schurr, 2012), dispersal (Elith & Leathwick, 2009; Iverson et al., 2004), and physiology (Buckley et al., 2011) and hence provide additional ecological insight. Increasing model complexity through including model structures that reflect spatio-temporal dynamics not only accounts for autocorrelation not explained by covariates, but may also reflect ecological processes such as dispersal limitation or site fidelity. However, caution is needed to avoid such random effects being overfitted and concealing the impact of model covariates (Sørbye et al., 2019).

Model complexity is occasionally influenced by the maximisation of predictive performance, not ecological theory, as has been noted for correlative species distribution modelling (Austin, 2002, 2007). In many studies of species distribution, the research aims include interpretation of effects or mechanisms, or extrapolation to spatial areas or time periods outside of those studied. In these instances, model interpretability and ecological realism should be considered alongside predictive performance during model selection.

In this paper, we highlight that the level of complexity in a spatio-temporal ecological model has to be carefully chosen and tailored to the system under investigation. To this end, we carefully inspect the complexity and limitations of a marked spatio-temporal point process model related to a model that has previously been fitted (Soriano-Redondo et al., 2019) with the aim of understanding and predicting the spatial distribution of a reintroduced species. The modelling in Soriano-Redondo et al. (2019) is facilitated through a Bayesian hierarchical modelling approach where spatio-temporal dynamics in behaviour are represented by a flexible stochastic partial differential equation (SPDE) (Lindgren et al., 2011). This allows the model to be fitted in continuous space (Simpson et al., 2017) and within a realistic timeframe through the computational efficiency of integrated nested laplace approximation (INLA) (Rue et al., 2009). We revisit this approach here, fitting a similar model using the user-friendly software package `inlabru`, with the specific aim of reflecting on the role of complex model components and how they represent ecological processes in species distribution models. We show that model complexity has to be carefully aligned with the complexity of the ecological processes that can be observed in the data, to provide adequate model interpretability and hence relevant ecological insight.

1.1 | Approaches to species distribution modelling

Species distribution models (SDMs) link information on the presence/absence or abundance of a species to environmental variables to predict where (and how much of) a species is likely to be present in unsampled locations or time periods (Martinez-Minaya et al., 2018).

Existing methods for species distribution modelling include: approaches developed to deal with presence-only datasets (such as maximum entropy algorithm, distance sampling, similarity, and envelope methods such as MAXENT, Gower metric, Mahalanobis distance, and ecological niche factor analysis); machine-learning algorithms that are iterative in ecological systems (such as artificial neural networks and deep learning); and classic regression models (such as generalised linear models (GLM), their semiparametric extension, generalised additive models (GAM), and a related method, multivariate adaptive regression splines (MARS)).

Several of the above methods are based on the assumption that the observations being modelled are independent. This is not always the case because ecological (and particularly species distribution) data usually present residual spatial or spatio-temporal correlation structures (Kneib et al., 2008). These structures can be the effects of processes that are unaccounted for in simple habitat models, such as social interaction with conspecifics, dispersal limitations or interspecific competition (Lichstein et al., 2002; Miller, 2012; Storch et al., 2003). Spatial data are also subject to spatial autocorrelation—the phenomenon that two points are more similar the closer they are in space, (Tobler, 1970).

Failing to account for unexplained spatial or temporal correlation structures in species distribution modelling can lead to spurious significance and an increase in the rate of Type I errors (false positives) in the interpretation of explanatory variables (Dormann, 2007; Dormann et al., 2007; Miller, 2012; Segurado et al., 2006). In some

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cases, not including these effects can lead to inference of an incorrect direction in the relationship between environmental covariates and response (Kühn, 2006). However, the inclusion of model components to account for unexplained spatial or temporal structures can rectify the above issues, as well as allowing insight into otherwise overlooked effects (Lichstein et al., 2002) and improving model fit (Dormann, 2007). Although, careful prior choice is advised to reduce the risk of correlation structures incorrectly lowering the estimated significance of fixed effects (Hodges & Reich, 2010; Sørbye et al., 2019).

GAM and MARS can account for spatio-temporal autocorrelation by taking a semiparametric approach through including smoothing spline terms. A very powerful and flexible alternative is to use a parametric approach in which the spatial and temporal correlation structures are represented by a (Gaussian) random field and approximated by an SPDE (Lindgren et al., 2011; Miller et al., 2020). This allows us to flexibly model species distribution data in continuous space as georeferenced data (for count or presence-absence data) or as point patterns (for presence only or effort-corrected presence data; Yuan et al., 2017). This approach is taken in Soriano-Redondo et al. (2019) as well as in this paper.

1.2 | Point process methodology and INLA

As model complexity increases—for example, when spatio-temporal correlation structures are included as complex model components—an advantageous approach can be found in Bayesian hierarchical modelling (Cressie et al., 2009). Hierarchical models allow parameters to vary at more than one level via the introduction of random effects. Constructing such models in a Bayesian framework enables full inference through the quantification of parameters and uncertainty; a feature of great practical benefit in applied conservation contexts (Wade, 2000; Wintle et al., 2003). Bayesian modelling also enables the inclusion of informative prior information, which can help to avoid overfitting of random effects (Sørbye et al., 2019). Several authors have applied a Bayesian approach to analyse different aspects of species distribution, such as comparing Gaussian processes against other popular approaches to modelling (Golding & Purse, 2016), visualising spatial patterns with Bayesian hierarchical logistic regression (Gelfand et al., 2006), analysing species distribution using presence/absence data (Latimer et al., 2006), working with occupancy models (MacKenzie et al., 2002), or analysing distance sampling data using a Bayesian GAM (Sigourney et al., 2020).

Hierarchical Bayesian models have traditionally relied on Markov chain Monte Carlo (MCMC) simulation techniques, which are computationally expensive and technically challenging, consequently limiting their application. However, INLA methodology and its powerful application to modelling complex datasets has opened hierarchical Bayesian modelling to a wider range of applications (Illian et al., 2013). As opposed to MCMC simulations, INLA uses an approximation for inference and hence avoids the intense computational needs, convergence, and mixing problems sometimes encountered by MCMC algorithms. Moreover, INLA can be used in combination with the SPDE approach, which approximates Gaussian random fields with a Matérn covariance structure, allowing for computationally efficient modelling in continuous space (Lindgren et al., 2011; Simpson et al., 2012). Implementation of these methods in `R` (R Core Team, 2021) is made accessible through the package `inlabru` (Bachl et al., 2019), which was designed as a user-friendly wrapper and extension to `R-INLA` (Rue et al., 2009), providing easier syntax for point process model fitting as well as reflecting ecological observation processes. These statistical methods and their packages for implementation constitute an efficient and reliable framework for complex spatio-temporal modelling, with a range of applications in ecology and beyond (Williamson et al., 2021; Yuan et al., 2017).

Here, we focus on the application of spatio-temporal marked point process models to analyse complex ecological systems, made feasible through the use of `R-INLA` and `inlabru`. Spatio-temporal point process methods model the locations of objects or events in time and space to provide an understanding of the overall spatio-temporal pattern formed by the locations, which might reflect interindividual interactions or habitat preferences. A simple example of a point process is the inhomogeneous Poisson process, where points are distributed following a fixed intensity function that varies through space. A particularly flexible class of point process models are Cox processes, specifically the log Gaussian Cox process. Cox processes are generalisations of the inhomogeneous Poisson process, where the intensity is represented by a random field. This implies that both the intensity function and the distribution are random ('doubly-stochastic'). In the case of the log Gaussian Cox process, the intensity $\Lambda(.)$ has the following form: $\Lambda(s) = \exp(G(s))$, where $G(.)$ is a Gaussian random field. Given this intensity, the point locations are assumed to be independent, that is, to follow a Poisson process. As a result, these are a special case of latent Gaussian models, the class of models that may be fitted with INLA (Cox & Isham, 1980; Illian et al., 2012; Møller et al., 1998).

An extension of the simple point process model is the marked point process model, in which information on the location of objects or events is modelled along with the qualitative or quantitative properties (commonly referred to as 'marks') of these events or objects. If a dependency between the values of a mark and the point distribution is assumed, these can be jointly modelled with two or more dependent likelihoods. Simultaneously modelling the distribution of points (as a point process) and some data feature collected at the point locations greatly increases the model complexity and computational cost. The computational efficiency of INLA, however, makes fitting these models feasible (Illian et al., 2012) and inlabru facilitates the fitting of both simple as well as marked point processes, which can be rather cumbersome in plain `R-INLA`.

When fitting a marked point process model in inlabru, the spatio-temporal structure of the marks (independent of the point distribution) and the spatio-temporal structure of the points can be represented with different Gaussian random fields in a shared representation of continuous space and discrete time. In this way, multiple

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

data features and different sources of spatial clustering can be incorporated into a single model of species distribution.

In this paper, we reflect on a modelling structure first developed by Soriano-Redondo et al. (2019), which follows the population spread of the Eurasian crane *Grus grus* across England. Here, we construct several models based on this work, to assess the role of complex model components in capturing spatial and spatiotemporal clustering dynamics in the species distribution. We take advantage of the new and advanced analytical approaches developed in inlabru, to explore the spatiotemporal structures and AR1 temporal correlation component in Soriano-Redondo et al. (2019). Here, the point pattern of habitat patches also reflects the observation process, as data-collectors used prior information on species habitat preference to influence the sampling design. Accounting for spatially varying detection probability is a particular strength of `inlabru`, which was developed specifically for (ecological) datasets with complex observation processes.

2 | MATERIALS AND METHODS

2.1 | Data

We investigate the spatial distribution of a resident breeding population of Eurasian crane in England following the return of the species to the UK in 1979 (Stanbury, 2011), with the aim of predicting the distribution of the population in future years. Breeding pairs of cranes are only found in wetland habitats, which creates a spatial restriction on the suitable habitat available to the species, although little is known about its dispersal ability. The distribution of wetland locations in England is heterogeneous, with various underlying factors creating a clustered structure of habitat patches. Thus, the distribution of the crane population throughout the UK and hence its future spread is dependent on the availability of suitable wetland habitat. Due to this known species habitat preference, data on crane presence were collected only at wetland locations, and not in interstitial areas. We demonstrate how this type of data can be modelled as a marked point process in `inlabru`, in order to infer habitat preferences and predict species distribution, while accounting for known species habitat requirements and observation processes. Specifically, we model the observation process as reflected in the wetland locations as a point pattern, jointly modelled with the presence/absence of nesting cranes as a mark.

The dataset includes information on a small breeding population that naturally settled in the UK in 1979 after earlier extirpation of the species, as well as animals used in a reintroduction project in 2010–2014. It also includes records for 2,526 wetlands in England derived from the UK Land Cover Map 2007 (Morton et al., 2011) and the Wetland Vision map of current wetlands (Hume, 2008). While the UK Land Cover Map 2007 assigns the dominant habitat to a 25 m raster grid, the Wetland Vision map determines the boundaries of the major wetland areas in England. Combining both sources provides an approximation of most available wetland patches in England. This selection was reduced by eliminating wetlands in moorland areas since cranes in the UK are only nesting in lowland wetland areas (Stanbury & Sills, 2012). Wetland patches smaller than 5 ha were also removed, as estimates show that cranes required at least 8 ha to nest (Johnsgard, 1983). The fine grain of this classification allows us to treat the wetland areas as points, since extensive wetland regions are fragmented by other types of habitats, roads or other human infrastructure, and are indeed composed by multiple smaller wetland areas. Each wetland area has associated information: its coordinates, data on the presence or absence of breeding pairs of cranes, wetland extent, perimeter-to-area ratio, the proportion of urbanised areas in a 10 km surrounding terrestrial buffer.

To evaluate the structure of the model in Soriano-Redondo et al. (2019), we present two modelling approaches with the aim of estimating the spatio-temporal distribution of cranes in England. The first approach makes use of a single Gaussian random field to represent spatio-temporal correlation in the distribution of the observed population of cranes. The second is a marked point process approach, which incorporates a second Gaussian random field, accounting for the distribution of wetlands. Within these modelling frameworks, we explore differences arising from the inclusion of a temporal correlation structure via a first-order autoregressive (AR1) process, where estimates for each year are dependent on the immediately previous year, compared to treating each year of data as independent and identically distributed (IID).

2.2 | Single-field models

In order to explore what level of model complexity is needed to answer relevant ecological questions based on the crane data, we start with a relatively simple spatio-temporal model in continuous space. To improve our understanding of the spatio-temporal distribution of cranes across England, we initially ignore the fact that cranes will only nest in wetland habitats, and we first construct a model containing a single spatio-temporal Gaussian random field. This is essentially a model of geo-referenced (binomial) data (often referred to as a geostatistical model) with a single likelihood, which does not take the observation process into account and assumes that crane presence is equally likely across England. Here, crane presence is modelled as a binomial distribution:

$$O_{s,t} \sim \text{Bernoulli}(P(s,t)). \tag{1}$$

The occurrence of cranes $O_{s,t}$ at each location s, in each year t can be considered an individual Bernoulli trial with probability $P(s,t)$. Here, the location s is a vector representing coordinates s_1 and s_2 in two-dimensional space. Note that for larger-scale models, location could also be represented on the sphere (Simpson et al., 2016).

The probability of crane presence P in a location s at time t is dependent on the observed distribution of the existing population in space and time, as well as the effect of environmental covariates:

$$P(s,t) = \text{logit}^{-1}\left(\beta_0 + \sum_{i=1}^{I} \alpha_i x_i(s,t) + M(s,t)\right). \tag{2}$$

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Here, β_0 represents an intercept term and $x_i(s,t)$ is the value of each covariate i at location s and time t. The environmental covariates included in the model were: the density of surrounding urbanised areas (within a 10 km terrestrial buffer), wetland perimeter-to-area ratio, and wetland extent. The values of these covariates were standardised using the Z-score formula prior to inclusion in the model, in order to improve the stability of numerical computation. The regression coefficient α_i of each covariate is estimated in the model. The spatio-temporal Gaussian random field $M(s,t)$ represents the spatial structure in distribution of the observed crane population in space and time unexplained by the intercept and covariates. The Gaussian random field $M(s,t)$ is approximated using an SPDE. Penalised complexity (PC) priors defining a likely minimum spatial range of 200 km and maximum standard deviation of 1.5 were used to inform this covariance structure, according to ecological understanding of the distances across which values may be correlated, and the extent to which values may vary. PC priors are interpretable default priors that operate under the principle of Occam's Razor, penalising complexity away from a simpler base model (Blangiardo et al., 2013; Simpson et al., 2017). Further information on the priors used can be found in Appendix B. We explore two options for the temporal element of this field, (a) incorporating a random effect in which each year of data is considered independent and identically distributed (IID) and (b) modelling temporal correlation between consecutive years with a first-order autoregressive (AR1) process.

Model fitting and inference were carried out in R version 4.1.1 (R Core Team, 2021) using the packages `inlabru` version 2.3.1.9000 (Bachl et al., 2019) and R-INLA version 21.02.23 (Rue et al., 2009). The code used in this analysis is available on Zenodo (Laxton et al., 2022a). A transformed version of the data is available on Dryad (Laxton et al., 2022b). These data have been randomly transformed to prevent an exact identification of nest locations and avoid potential disturbance to cranes during the breeding period.

2.3 | Marked point process models

The binomial presence/absence model in Equations (1) and (2) captures the spatial correlation structure in the occurrence of breeding pairs of cranes, but ignores the distribution of suitable habitat patches (wetlands). However, sampling only took place in locations where cranes would be expected to nest, this observation process can also be interpreted as preferential sampling (Diggle et al., 2010; Pennino et al., 2019). In this example, crane detectability is assumed to be known, as due to the large size of the animals, they are easy to detect. However, with rare or small organisms that are more difficult to detect, it may be of interest to use models to predict and estimate abundance across space. In these cases, such a model would likely predict presence where the species of interest does not occur, and overestimate abundance as a result. Despite the certainty of detection of cranes here, the distribution of wetlands is still an important factor in determining species distribution, as it represents habitat preference. This can be accounted for through modelling the data using a marked point process, by jointly modelling the distribution of wetlands as a point process and the presence/absence of breeding pairs of cranes as a binomial mark. The presence/absence of cranes can be viewed as a mark on the wetland point process; a feature which only occurs at the point locations, and so is dependent on the underlying spatial structure of the points. In this second approach, we extend the model presented in Equation (2) to include the density of available wetland habitat. This model structure is taken from Soriano-Redondo et al. (2019), and the models developed in this manuscript are based on those created by Soriano-Redondo et al. (2019), in order to evaluate the advantages and disadvantages of including complex model components.

The point pattern of wetland locations is modelled as a log Gaussian Cox process in `inlabru`. The distribution of wetlands is independent conditional on the point process intensity $\Lambda(s)$:

$$\Lambda(s) = \exp\big(\alpha_0 + G(s)\big). \tag{3}$$

The log-intensity of the spatial point process model is given by an intercept term α_0 and a spatial Gaussian random field $G(s)$. Since we are not interested in gaining an understanding of the underlying drivers of wetland distribution in this context, we do not include covariates in the model that would help explain their spatial distribution; however, this could be easily done.

The probability of crane presence $P(s,t)$ is dependent on both the distribution of wetlands in space, and the distribution of the existing crane population in space and time:

$$P(s,t) = \text{logit}^{-1}\left(\beta_0 + \sum_{i=1}^{I} \alpha_i x_i(s,t) + \beta G(s) + M(s,t)\right). \tag{4}$$

Here, wetland density is incorporated into the model through the inclusion of the spatial Gaussian random field $G(s)$ multiplied by a scaling parameter β, which determines the strength and direction of the interaction between wetland density and crane presence. A prior is placed on β (see the Data Availability Statement for code) to provide stability due to convergence issues. This ensures a simple interpretation of the interaction between $G(s)$ and crane presence (i.e., positive effect of high density as opposed to negative effect of low density). The spatio-temporal Gaussian random field $M(s,t)$ represents the distribution of the observed crane population in space and time, independent of wetland distribution. As in the binomial presence/absence model, both an AR1 and IID temporal correlation structure are considered for $M(s,t)$. However, the point pattern of wetland locations does not change between years, and so represents a constant restriction on species distribution. Therefore, $G(s)$ is a purely spatial Gaussian random field, and does not incorporate a temporal element. This represents an improvement in computational efficiency compared to the model structure in Soriano-Redondo et al. (2019), which includes a spatiotemporal field instead of a spatial one.

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3 | RESULTS

In both the binomial presence/absence model (Equation 2) and marked point process model (Equation 4), which incorporated an AR1 temporal correlation structure, the correlation parameter for the AR1 process, ρ_t, (which is bounded between −1 and 1) was estimated to be extremely high (Table 1). This indicates a very strong temporal correlation across years. Due to this strong temporal correlation, the model estimated the same spatial structure in the mark random field $M(s,t)$ and predictions of crane probability of presence across all 5 years considered in the study (see Appendix A, Figure A2). These estimates constituted a spatial structure reflecting an average of all years of data, as opposed to accurately representing changes in spatial structure each year. Such averaged estimates may be inaccurate when interpreting within-year spatial effects, and so graphical representations of results are provided here for the models which treat each year of data as IID only.

In the marked point process and binomial presence/absence models, the mark random field $M(s,t)$ represents the spatial distribution of the crane population each year (Figure 1, centre plot). This constrains predictions to near where nesting pairs of cranes have been observed, and without it, the species could be predicted to occur in areas of suitable environmental conditions, but that are an unrealistic distance from the established population.

The marked point process models also contain a point random field, $G(s)$, which models the spatial distribution of wetlands (Figure 1, left plot; Figure A4). The strength and direction of the interaction between the wetland density and probability of crane presence is estimated through the scaling parameter β. In the models containing the point random field, β is estimated as a significant positive effect (Table 1). This means that probability of crane presence is estimated to be higher in areas of high wetland density, as opposed to low density areas. Without the inclusion of the point random field (as in the binomial presence/absence model), predictions of species distribution could be correlated across wide expanses of interstitial habitat and are not assumed to depend on the density of suitable habitat patches.

Estimated regression coefficients were similar across all models for both wetland perimeter-to-area ratio and wetland extent. For all models, wetland perimeter-to-area ratio had a significant negative effect on crane presence, and wetland extent had a significant positive effect (Table 1). However, there was a slight difference in estimated effect of density of surrounding urbanised areas between the IID and AR1 models. Although density of surrounding urbanised areas was found to have a negative effect in all models, this was estimated as not being significant in the binomial presence/absence and marked point process models which incorporated an AR1 temporal correlation structure, but was estimated as being significant in those models with an IID temporal structure (Table 1).

Probability of presence of breeding pairs of cranes across space in 2015 was predicted using the binomial presence/absence and marked point process models with IID temporal structure. Predictions from the binomial presence/absence model (Figure 2, left plot) show a wider spatial range in areas of high probability of presence, compared to predictions from the marked point process model (Figure 2, right plot).

TABLE 1 Posterior mean and 95% credible intervals for: Regression coefficients of environmental covariates; scaling parameter (β) representing the interaction between $G(s)$ and the probability of crane presence; temporal correlation parameter from the AR1 process; parameters of the spatial, $G(s)$, and spatiotemporal, $M(s,t)$, fields. The covariate coefficients and scaling parameter values with credible intervals, which do not cross zero are in bold. Model run times and Watanabe–Akaike information criterion (WAIC) scores are also given. The covariates included in the model were: Density of surrounding urbanised areas (within a 10 km terrestrial buffer), wetland perimeter-to-area ratio and wetland extent. The values of these covariates were standardised using the *Z*-score formula prior to inclusion in the model. All values given are rounded to 2 decimal places

	Binomial presence/absence		Marked point process	
	IID	AR1	IID	AR1
Urban density coefficient	**−0.58 [−1.19, −0.04]**	−0.14 [−1.15, 0.72]	**−0.74 [−1.48, −0.1]**	−0.09 [−1.19, 0.85]
Wetland extent coefficient	**0.08 [0.02, 0.14]**	**0.12 [0.02, 0.24]**	**0.1 [0.03, 0.16]**	**0.1 [0, 0.21]**
Perimeter:Area ratio coefficient	**−3.02 [−3.92, −2.2]**	**−3.49 [−4.57, −2.53]**	**−3.02 [−3.94, −2.19]**	**−3.47 [−4.55, −2.51]**
Scaling (β)	NA	NA	**0.62 [0.11, 1.16]**	**0.99 [0.04, 2.05]**
Temporal correlation (ρ_t)	NA	0.99 [0.98, 1]	NA	0.99 [0.98, 1]
$M(s,t)$ spatial range	346.7 [183.45, 612.59]	119.95 [70.38, 197.16]	238.66 [142.06, 315.93]	145.15 [67.5, 282.16]
$M(s,t)$ standard deviation	1.22 [0.66, 2.01]	3.03 [1.94, 4.25]	0.98 [0.56, 1.29]	3.34 [2.24, 4.86]
$G(s)$ spatial range	NA	NA	105.97 [88.57, 185.44]	124.19 [83.17, 188.41]
$G(s)$ standard deviation	NA	NA	1.22 [1.07, 1.32]	1.34 [1.01, 1.83]
Run time	63.26	123.64	108.62	174.94
WAIC	464.38	425.38	28,045.01	28,024.94

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

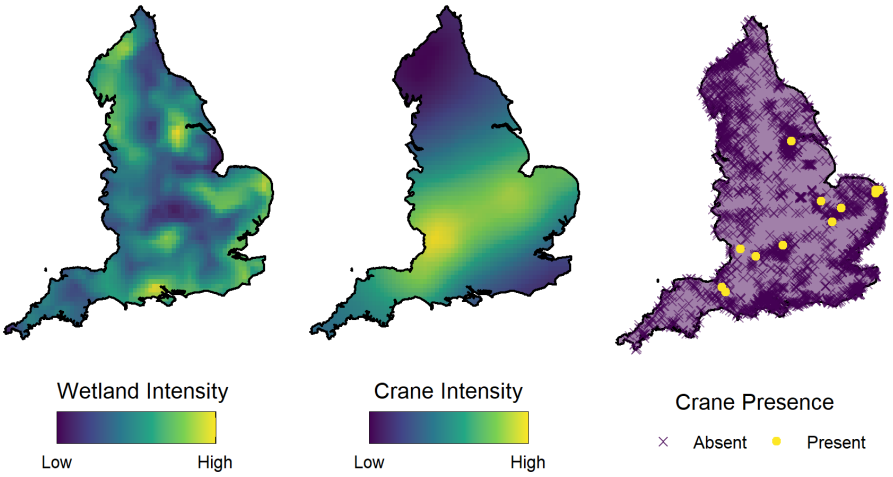

FIGURE 1 Estimated point random field ($G(s)$, left) and mark random field ($M(s, t)$, centre) for 2015 from the marked point process model with IID temporal structure. Colour scale is given in low-high intensity as we are interested in relative differences across space and not absolute values. Distribution of wetland locations and observed crane presences in England in 2015 (right). Purple background colour is used for visual clarity and does not represent a value on the colour scale.

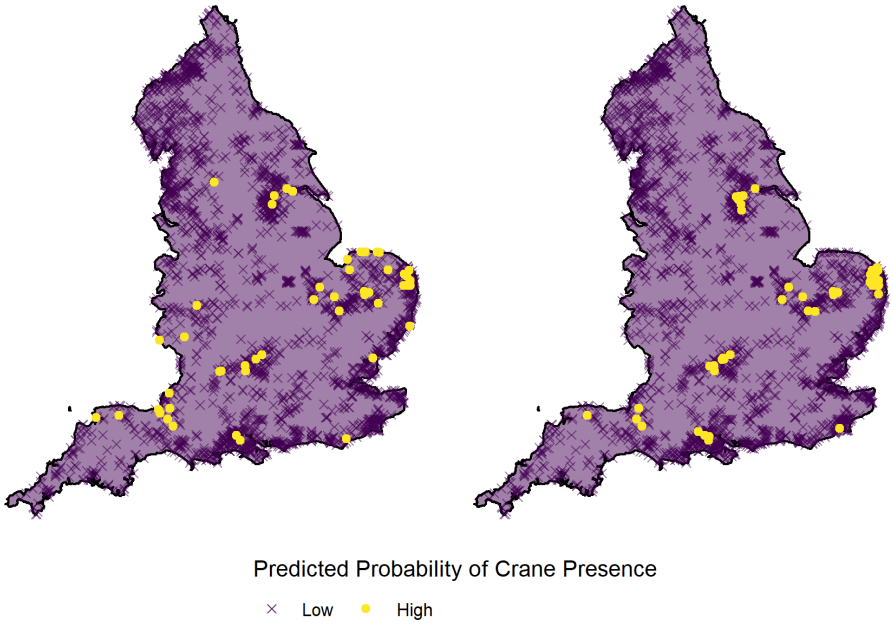

FIGURE 2 Mean predicted probability of the presence of a breeding pair of cranes at each wetland location in 2015. Predictions were made using the binomial presence/absence model (left) and the marked point process model (right) with IID temporal structure. Colour scale is given in low-high probability as we are interested in relative differences across space and not absolute values. 'Low' represents a value of mean predicted probability of crane presence less than or equal to the 99th quantile of mean predicted probabilities across all four models analysed. 'High' represents a value of mean predicted probability of crane presence greater than this cutoff point. Purple background colour is used for visual clarity, and does not represent a value on the colour scale.

4 | DISCUSSION

In this paper, we have fitted four different models of varying complexity. The simplest model is a spatio-temporal model with a single likelihood with an IID assumption between years. We will now compare the relative benefits of the different models with regard to the ecological aim of predicting species distribution and representing relevant ecological processes in space and time. Predictions of high probability of crane presence from the marked point process model (Figure 2, right plot) correspond to areas of high wetland density (Figure 1, left plot), provided these high density areas overlap with areas of high species intensity (Figure 1, centre plot). Differences can be observed in the high probability of presence along the Norfolk Coast (East) predicted by the binomial presence/absence model (Figure 2, left plot), likely due to the presence of cranes to the South East of this area (Figure 1, right plot). However, this is not the case in the marked point process prediction, where it is taken into account that this area has a lower density of wetland habitat relative to the area where the existing population is observed. Additionally, the binomial presence/absence model predicts a high probability of presence at a single wetland in the North-West of England, likely due to the large extent of this wetland. Below this, several wetlands spanning the West and South-West of England are predicted to have mid-to-high probability of crane presence. This does not occur in the marked point process prediction, where spatial structure of the predictions is more constrained and dependent on wetland density.

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The scaling parameter β represents the strength and direction of the interaction between the point field $G(s)$ and the probability of crane presence. In the marked point process model with an AR1 temporal structure, the significant positive effect of wetland density on the probability of crane presence is of greater magnitude than that of the model with IID temporal structure (Table 1). This is due to the fact that the AR1 model borrows strength across years, meaning that the repeated presence of cranes in high-wetland-density areas over time indicates a stronger effect than that observed in the IID model, which considers each year of data independently.

The spatiotemporal field $M(s,t)$ in the AR1 models has a smaller posterior spatial range and larger standard deviation than that of the IID models (Table 1). This is due to the fact that the AR1 models borrow strength across years, meaning that the spatial structure is influenced by the repeated occurrence of relatively isolated crane presences over time. Similarly, the standard deviation is high, as this allows sufficient variation in the field for the estimated species intensity to range from low to very high values. On the other hand, the IID models consider each year of data independently, and so are more strongly influenced by the priors, which suggest a larger spatial range and lower standard deviation, with the intensity of the species correlated across larger intervals and lowering smoothly away from areas of higher intensity. While the priors are biologically realistic, they may contrast with the trends observed in the data due to the early stage of reintroduction at which the data was collected.

In the models with AR1 temporal structure, the temporal correlation, ρ_t, is estimated to be extremely high (Table 1) and the spatial structure of $M(s,t)$ is predicted to be the same for each year of data (Appendix A, Figure A2). This indicates that the spatiotemporal field in these models captures a spatial structure in the data that is constant over time. This strong trend of a static spatial structure may be due to the early stage of reintroduction when observations were made. At this early stage, the population is small and so may not be subject to strong dispersal pressures such as habitat density dependence, meaning that movement between habitat patches may be less common than it would be in a densely populated area. The population is also only observed in a small proportion of its potential range, meaning that a large amount of the study area is taken up by consistent zero counts. Limited movement, large numbers of consistent absences, and the binary nature of presence/absence data all lead to few observed changes in the dataset between years (Appendix A, Figure A1). Therefore, the static spatial structure picked up by the model may not reflect the behaviour of the study species, for example, strong site fidelity, and instead could be a feature of the stage of the reintroduction process the population was in when the data were collected. A longer period of observation may be required to accurately capture and predict the dispersal dynamics of the species.

Since the AR1 models detect the strong trend of a static spatial structure over time, they estimate no change in the spatial representation of the spatio-temporal field $M(s,t)$ between years. This means that the spatio-temporal field must represent the most likely spatial correlation structure for all years, and so produces an average from the full dataset. This averaging can produce misleading estimates of species distribution, particularly when the species is observed to 'appear' in a new area of the study region part-way through the study, as it can be predicted to have a high probability of presence in this area before the appearance event. For example, breeding pairs of cranes were first observed in Somerset (South West) in 2013 as the result of a successful reintroduction project (Appendix A, Figure A1). However, the spatio-temporal field from the AR1 models shows a high intensity of cranes in this area in the years prior to this event, when no breeding pairs of cranes had been observed in the area (Appendix A, Figure A2). In this example, the covariates included in the model are also stationary across the interval of observation. As such, predictions from the models with AR1 temporal structure are very similar across all years in the study period (Appendix A, Figure A5). This not only conceals fine scale interannual differences in spatial distribution but also removes novel insight from future predictions. Such an effect may be avoided by including an extra, purely spatial, field in the mark likelihood and setting priors to manually restrict this to capturing the static spatial effect, separating this strong signal from the changes in spatial structure over time which could then be picked up by the spatiotemporal field. As mentioned above, the effect observed here is a result of the data structure, and so a larger dataset collected over a longer duration may provide sufficient information to avoid these issues.

The issues created by the strong, static spatial trend picked up by the spatiotemporal field in this example demonstrate that inclusion of complexity (here, an effect accounting for temporal correlation) may not necessarily improve ecological insight gained from a model. The averaged predictions created by the most complex model we consider here (the marked point process with AR1 temporal correlation structure) mean that it was not the most suitable model with regard to prediction of the species distribution within the observation period. It is likely that this effect was present in the INLA model of this data (Soriano-Redondo et al., 2019) but was interpreted as a strong temporal trend in direction of species spread, as opposed to a static effect, due to a lack of accessible plotting methods in `INLA` at the time of analysis. However, due to the accessible `predict()` function and connection with visualisation tool `ggplot2` in `inlabru`, the true nature of this effect was more easily identified. An even more complex model, containing another purely spatial field and strong priors, may improve output quality when aiming to predict future distribution of the population.

Of the four models considered in this paper, the most complex (the marked point process model with AR1 temporal structure) had the longest running time, and the simplest (the binomial presence/absence model with IID temporal structure) had the shortest (Table 1). While running time is an important methodological consideration, each of these models is made extremely computationally efficient through the use of the INLA approach, allowing model choice to be more strongly determined by ecological insight gained, as opposed to computational efficiency.

Watanabe–Akaike information criterion (WAIC) scores were computed for each of the four models (Table 1), although it is

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

important to note that the binomial presence/absence and marked point process models cannot be compared using this criterion. However, within each model type, it is interesting to note that the AR1 models have a lower score than the IID models, indicating a better fit to the data. This is unsurprising when considering that the AR1 models are more strongly influenced by the data, as information is borrowed across years, whereas the IID models are more reliant on the priors. The true meaning of such criteria in the area of point process modelling is not well understood, due to the historically theoretical nature of the topic, and as demonstrated in this manuscript, these criteria do not necessarily provide a definitive answer as to selecting the 'best' model for interpretation and ecological insight.

Species distribution data is often a combination of the true underlying distribution of the species, and the observation process used to collect data, creating a need to disentangle the true drivers of species distribution from sampling effects. In the early stages of a species invasion or reintroduction, there is a time lag, which means that the observed distribution may not match the true potential distribution of the species once the population has been established (De Marco et al., 2008). Making observations at this initial stage could bias inference of the effects of environmental covariates, as the distribution of the species is restricted by dispersal limitations, which remain unaccounted for. Including complex model components to represent spatial correlation structures, as is demonstrated here via the inclusion of $G(s)$ in the marked point process model, can account for the fact that the species distribution is limited by the early stage of its introduction to the environment, giving a more accurate prediction of the range being constrained close to where the species has been observed. Accounting for spatial correlation can also aid in avoiding Type I errors when inferring the significance of covariate effects (Dormann et al., 2007). However, caution is advised when incorporating additional structural complexity into models of this type of data, due to its zero-inflated and relatively static nature (observed here through the issues associated with inclusion of an AR1 temporal correlation structure). Accounting for the observation process, and spatially varying detection probability is a particular strength of the `inlabru` package, which can also be used to model subsampled and distance sampling data (Jullum et al., 2020; Yuan et al., 2017).

Complex model components such as spatial and spatio-temporal random fields can be used to represent population dynamics in species distribution models. In this example, density of wetland habitat patches has a positive effect on the probability of crane presence in the marked point process model, representing the habitat preferences of this species. Through explicit understanding of the dynamics represented by each field, the benefits of increased complexity can be achieved while maintaining interpretability of results. However, when increasing model complexity, it is important to maintain an understanding of the effects and role of each model component, in order to determine whether it benefits the outcomes of the modelling aims.

AUTHOR CONTRIBUTIONS

Megan Laxton, Óscar Rodríguez de Rivera and Janine Illian conceived the ideas and designed the methodology; Andrea Soriano-Redondo provided access to the data and advised on ecological background and model interpretation; Megan Laxton analysed the data; Megan Laxton, Óscar Rodríguez de Rivera and Janine Illian led the writing of the manuscript; Janine Illian supervised all stages of research. All authors contributed critically to the drafts and gave the final approval for publication.

ACKNOWLEDGEMENTS

The authors thank Andrew Seaton (University of Glasgow) for assistance in the development of Appendix B on PC priors.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13957.

DATA AVAILABILITY STATEMENT

A transformed version of the data used in this analysis is available on Dryad Digital Repository https://doi.org/10.5061/dryad.2z34tmpps (Laxton et al., 2022b). These data have been randomly transformed to prevent an exact identification of nest locations and avoid potential disturbance to cranes during the breeding period. R code which can be used to fit the models developed in this manuscript to the supplied data is available on Zenodo https://doi.org/10.5281/zenodo.6907553 (Laxton et al., 2022a).

ORCID

Megan R. Laxton https://orcid.org/0000-0002-5017-7395
Óscar Rodríguez de Rivera https://orcid.org/0000-0002-2754-4265
Andrea Soriano-Redondo https://orcid.org/0000-0003-0728-730X
Janine B. Illian https://orcid.org/0000-0002-6130-2796

REFERENCES

Austin, M. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling*, *157*(2–3), 101–118.

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, *200*(1–2), 1–19. https://doi.org/10.1016/j.ecolmodel.2006.07.005

Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). Inlabru: An r package for bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, *10*(6), 760–766. https://doi.org/10.1111/2041-210x.13168

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with r-inla. *Spatial and Spatiotemporal Epidemiology*, *4*, 33–49. https://doi.org/10.1016/j.sste.2012.12.001

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bolker, B. (2009). Learning hierarchical models: Advice for the rest of us. *Ecological Applications*, *19*(3), 588–592. https://doi.org/10.1890/08-0639.1

Buckley, L. B., Waaser, S. A., MacLean, H. J., & Fox, R. (2011). Does including physiology improve species distribution model predictions of responses to recent climate change? *Ecology*, *92*(12), 2214–2221.

Cox, D. R., & Isham, V. (1980). *Point processes* (Vol. *12*). CRC Press.

Cressie, N., Calder, C. A., Clark, J. S., ver Hoef, J. M., & Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, *19*(3), 553–570.

De Marco, P., Diniz-Filho, J. A. F., & Bini, L. M. (2008). Spatial analysis improves species distribution modelling during range expansion. *Biology Letters*, *4*(5), 577–580. https://doi.org/10.1098/rsbl.2008.0210

Diggle, P. J., Menezes, R., & Su, T.-L. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(2), 191–232. https://doi.org/10.1111/j.1467-9876.2009.00701.x

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, *16*(2), 129–138. https://doi.org/10.1111/j.1466-8238.2006.00279.x

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., & Wilso, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, *30*(5), 609–628.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677–697.

Gelfand, A. E., Holder, M., Latimer, A., Lewis, P. O., Rebelo, A. G., Silander, J. A., Jr., & Wu, S. (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, *1*(1), 41–92.

Golding, N., & Purse, B. V. (2016). Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Methods in Ecology and Evolution*, *7*(5), 598–608.

Hodges, J. S., & Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, *64*(4), 325–334. https://doi.org/10.1198/tast.2010.10052

Hume, C. (2008). *Wetland vision technical document: Overview and reporting of project philosophy and technical approach* (Technical Report No. 80). The Wetland Vision Partnership.

Illian, J. B., Martino, S., Sørbye, S. H., Gallego-Fernández, J. B., Zunzunegui, M., Esquivias, M. P., & Travis, J. M. (2013). Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution*, *4*(4), 305–315.

Illian, J. B., Sørbye, S. H., & 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. https://doi.org/10.1214/11-aoas530

Iverson, L. R., Schwartz, M. W., & Prasad, A. M. (2004). How fast and far might tree species migrate in the eastern United States due to climate change? *Global Ecology and Biogeography*, *13*(3), 209–219.

Johnsgard, P. A. (1983). *Cranes of the world*. Cambridge University Press.

Jullum, M., Thorarinsdottir, T., & Bachl, F. E. (2020). Estimating seal pup production in the Greenland Sea by using Bayesian hierarchical modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(2), 327–352. https://doi.org/10.1111/rssc.12397

Kneib, T., Müller, J., & Hothorn, T. (2008). Spatial smoothing techniques for the assessment of habitat suitability. *Environmental and Ecological Statistics*, *15*(3), 343–364.

Kühn, I. (2006). Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, *13*(1), 66–69. https://doi.org/10.1111/j.1472-4642.2006.00293.x

Latimer, A. M., Wu, S., Gelfand, A. E., & Silander, J. A., Jr. (2006). Building statistical models to analyze species distributions. *Ecological Applications*, *16*(1), 33–50.

Laxton, M. R., Rodríguez de Rivera, O., Soriano-Redondo, A., & Illian, J. B. (2022a). Code from: Balancing structural complexity with ecological insight in spatio-temporal species distribution models. *Zenodo*, https://doi.org/10.5281/zenodo.6907553

Laxton, M. R., Rodríguez de Rivera, O., Soriano-Redondo, A., & Illian, J. B. (2022b). Transformed crane data from: Balancing structural complexity with ecological insight in spatio-temporal species distribution models. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.2z34tmpps

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, *72*(3), 445–463.

Lindgren, F., Rue, H., & 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. https://doi.org/10.1111/j.1467-9868.2011.00777.x

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255.

Martinez-Minaya, J., Cameletti, M., Conesa, D., & Pennino, M. G. (2018). Species distribution modeling: A statistical review with focus in spatio-temporal issues. *Stochastic Environmental Research and Risk Assessment*, *32*(11), 3227–3244.

Miller, D. L., Glennie, R., & Seaton, A. E. (2020). Understanding the stochastic partial differential equation approach to smoothing. *Journal of Agricultural, Biological and Environmental Statistics*, *25*(1), 1–16. https://doi.org/10.1007/s13253-019-00377-z

Miller, J. A. (2012). Species distribution models. *Progress in Physical Geography: Earth and Environment*, *36*(5), 681–692. https://doi.org/10.1177/0309133312442522

Møller, J., Syversveen, A. R., & Waagepetersen, R. P. (1998). Log Gaussian cox processes. *Scandinavian Journal of Statistics*, *25*(3), 451–482. https://doi.org/10.1111/1467-9469.00115

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R., & Simpson, I. C. (2011). *Final report for LCM2007—The new UK land cover map* (Technical Report No. 11/07). Countryside Survey Technical Report.

Pagel, J., & Schurr, F. M. (2012). Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, *21*(2), 293–304.

Pennino, M. G., Paradinas, I., Illian, J. B., Muñoz, F., Bellido, J. M., López-Quílez, A., & Conesa, D. (2019). Accounting for preferential sampling in species distribution models. *Ecology and Evolution*, *9*(1), 653–663. https://doi.org/10.1002/ece3.4789

R Core Team. (2021). *R: A language and environment for statistical computing* [computer software manual]. Retrieved from https://www.R-project.org/

Rue, H., Martino, S., & 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. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Segurado, P., Araujo, M. B., & Kunin, W. E. (2006). Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, *43*(3), 433–444. https://doi.org/10.1111/j.1365-2664.2006.01162.x

Sigourney, D. B., Chavez-Rosales, S., Conn, P. B., Garrison, L., Josephson, E., & Palka, D. (2020). Developing and assessing a density surface model in a Bayesian hierarchical framework with a focus on uncertainty: Insights from simulations and an application to fin whales (*Balaenoptera physalus*). *PeerJ*, *8*, e8226. https://doi.org/10.7717/peerj.8226

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., & Rue, H. (2016). Going off grid: Computationally efficient inference for log-Gaussian cox processes. *Biometrika*, *103*(1), 49–70. https://doi.org/10.1093/biomet/asv064

Simpson, D., Lindgren, F., & Rue, H. (2012). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, *23*(1), 65–74. https://doi.org/10.1002/env.1137

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: Aprincipled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. https://doi.org/10.1214/16-sts576

Sørbye, S. H., Illian, J. B., Simpson, D. P., Burslem, D., & Rue, H. (2019). Careful prior specification avoids incautious inference for log-Gaussian cox point processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *68*(3), 543–564. https://doi.org/10.1111/rssc.12321

Soriano-Redondo, A., Jones-Todd, C. M., Bearhop, S., Hilton, G. M., Lock, L., Stanbury, A., Votier, S. C., & Illian, J. B. (2019). Understanding species distribution in dynamic populations: A new approach using spatio-temporal point process models. *Ecography*, *42*(6), 1092–1102.

Stanbury, A. (2011). The changing status of the common crane in the UK. *British Birds*, *104*, 432–447.

Stanbury, A., & Sills, N. (2012). *Common crane habitats in Britain* (Technical Report No. 23: 381). British Wildlife.

Storch, D., Konvicka, M., Benes, J., Martinková, J., & Gaston, K. J. (2003). Distribution patterns in butterflies and birds of The Czech Republic: Separating effects of habitat and geographical position. *Journal of Biogeography*, *30*(8), 1195–1205. https://doi.org/10.1046/j.1365-2699.2003.00917.x

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234. https://doi.org/10.2307/143141

Wade, P. R. (2000). Bayesian methods in conservation biology. *Conservation Biology*, *14*(5), 1308–1316.

Williamson, L. D., Scott, B. E., Laxton, M. R., Bachl, F. E., Illian, J. B., Brookes, K. L., & Thompson, P. M. (2021). Spatiotemporal variation in harbor porpoise distribution and foraging across a landscape of fear. *Marine Mammal Science*, *38*, 42–57. https://doi.org/10.1111/mms.12839

Wintle, B. A., McCarthy, M. A., Volinsky, C. T., & Kavanagh, R. P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, *17*(6), 1579–1590.

Yuan, Y., Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., Buckland, S. T., Rue, H., & Gerrodette, T. (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics*, *11*(4), 2270–2297. https://doi.org/10.1214/17-aoas1078

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Laxton, M. R., Rodríguez de Rivera, Ó., Soriano-Redondo, A., & Illian, J. B. (2023). Balancing structural complexity with ecological insight in Spatio-temporal species distribution models. *Methods in Ecology and Evolution*, *14*, 162–172. https://doi.org/10.1111/2041-210X.13957

2041210x, 2023, 1, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13957 by Test, Wiley Online Library on [17/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License