

## Understanding species distribution in dynamic populations: a new approach using spatio-temporal point process models

Andrea Soriano-Redondo<sup>1,2,3</sup>, Charlotte M. Jones-Todd<sup>4,5</sup>, Stuart Bearhop<sup>1</sup>, Geoff M. Hilton<sup>2</sup>, Leigh Lock<sup>3</sup>, Andrew Stanbury<sup>3</sup>, Stephen C. Votier<sup>6</sup>, Janine B. Illian<sup>4</sup>

<sup>1</sup>Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Cornwall Campus TR10 9EZ, UK

<sup>2</sup>Wildfowl & Wetlands Trust, Slimbridge, Gloucester, GL2 7BT, UK

<sup>3</sup>Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire, SG19 2DL, UK

<sup>4</sup>CREEM, School of Mathematics and Statistics, University of St Andrews, St Andrews, KY16 9SS, UK

<sup>5</sup>National Institute for Water and Atmospheric, Hamilton, 3216, NZ

<sup>6</sup>Environment and Sustainability Institute, University of Exeter, Cornwall Campus TR10 9EZ, UK

**Corresponding author:** Andrea Soriano-Redondo, Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Cornwall Campus TR10 9EZ, UK. Email:

as746@exeter.ac.uk

**Decision date:** 06-Feb-2019

---

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/ecog.03771].

**ABSTRACT**

Understanding and predicting a species' distribution across a landscape is of central importance in ecology, biogeography, and conservation biology. However, it presents daunting challenges when populations are highly dynamic (i.e. increasing or decreasing their ranges), particularly for small populations where information about ecology and life history traits is lacking. Currently, many modelling approaches fail to distinguish whether a site is unoccupied because the available habitat is unsuitable or because a species expanding its range has not arrived at the site yet. As a result, habitat that is indeed suitable may appear unsuitable. To overcome some of these limitations, we use a statistical modelling approach that applies spatio-temporal log-Gaussian Cox processes. These model the spatial distribution of the species across available habitat and how this distribution changes over time, relative to covariates. In addition, the model explicitly accounts for spatio-temporal dynamics that are unaccounted for by covariates through a spatio-temporal stochastic process. We illustrate the approach by predicting the distribution of a recently established population of Eurasian cranes (*Grus grus*) in England and estimate the effect of a reintroduction in the range expansion of the population. Our models show that wetland extent and perimeter-to-area ratio have a positive and negative effect, respectively, in crane colonisation probability. Moreover, we find that cranes are more likely to colonise areas near already occupied wetlands and that the colonisation process is progressing at a low rate. Finally, the reintroduction of cranes in SW England can be considered a human-assisted long-distance dispersal event that has increased the dispersal potential of the species along a longitudinal axis in S England. Spatio-temporal log-Gaussian Cox process models offer an excellent opportunity for the study of species where information on life history traits is lacking, since these are represented through the spatio-temporal dynamics reflected in the model.

**Keywords:** Species distribution models, point process models, spatio-temporal log-Gaussian Cox process, latent field, dispersal, colonisation, reintroduction.

## INTRODUCTION

Understanding and predicting a species' distribution across a landscape is of central importance in ecology, biogeography, and conservation biology (Elith and Leathwick 2009). The distribution of species is an outcome of dynamic processes, as population ranges change in response to both extrinsic (i.e. environmental conditions) and intrinsic (i.e. feedback processes within the populations) factors. Climate change, habitat loss, human disturbance or the introduction of non-native species are well-known causes of shifts and reductions in population distribution ranges (Mooney and Cleland 2001, Lenoir et al. 2008, Moritz et al. 2008, Mantyka- pringle et al. 2012, Haddad et al. 2015). Likewise, efforts to restore original habitats, enhance natural resources and reduce human threats have resulted in the range expansion of several species of vertebrates (Deinet et al. 2013, Tauler et al. 2015, Sun et al. 2016). Nevertheless, the expansion of these populations is strongly influenced by intrinsic factors; life history traits, such as intrinsic growth rate or dispersal ability, play a major role in shaping species occurrence (Hastings et al. 2005). Predicting the distribution of declining or increasing populations across space and time is extremely challenging (Elith and Leathwick 2009); especially, for small populations where data are usually scarce and potentially missing.

The most common approach to estimating the distributions of species is to use phenomenological species distribution models (SDMs) that correlate observations of species occurrences with environmental variables. Many phenomenological SDMs rely on the implicit assumption that species are spread throughout all the suitable environmental space (Guisan and Thuiller 2005, Dormann et al. 2012). However, when species are still expanding their range, as for example in the context of a recolonisation, this assumption is unrealistic. In such a scenario, favourable habitat yet unoccupied might be available just beyond the range of the species. Standard models draw conclusions based on how frequently species have been observed in specific environmental conditions. However, these cannot distinguish between habitat that has rarely been visited as it is unsuitable and habitat that has rarely been visited as it is outside a species' current range. This leads to wrong conclusions as to a species' habitat preferences and can make predictions unreliable.

To overcome some of these limitations several approaches have been developed to estimate species distribution. For example, generalized functional responses account for the availability of all environmental units (Matthiopoulos et al. 2011), while hierarchical models can account for imperfect

detection of the species' occurrences (Mang et al. 2017) and allow predicting species distribution in new areas. In addition, some recent models combine phenomenological and mechanistic approaches, which explicitly capture population dynamics (Gallien et al. 2010). In some cases to parameterize the mechanistic component of these models a detailed understanding of either key life history traits (including fecundity, mortality, dispersal ability), carrying capacity or density-dependence processes is required (Gallien et al. 2010, Fitzpatrick et al. 2012, Zurell et al. 2012, Sun et al. 2016). On the other hand, some methods have been able to link habitat suitability with population dynamics using georeferenced species counts (or presence- only data) (Matthiopoulos et al. 2015).

In this study, we propose an alternative approach which models the dynamics of the species distribution through time conditional on the distribution of potentially suitable habitat in space. Our approach is based on a marked spatio-temporal point process model. The relevance of spatial and spatio-temporal point process models in ecological contexts is becoming increasingly recognised. For example, Law et al. 2009, Illian et al. 2013, Brown et al. 2016 determine underlying processes that support species coexistence in plant communities and Yuan et al. 2017 model distance sampling data as a thinned log-Gaussian Cox process (LGCP).

We further develop the use of point process models in ecological contexts by considering a marked spatio-temporal point process model. This allows us to model the distribution of species in space and time conditional on both potential and suitable habitat. These dynamics are reflected by stochastic spatio-temporal terms in the model that i) represent the underlying processes governing habitat location and ii) conditional on potential habitat represent the spatial processes governing species distribution.

### **A marked spatio-temporal point process**

More specifically, spatio-temporal point processes model the pattern formed by the x- and y-locations of any objects (e.g. individuals, cells, potentially suitable habitat) or events (e.g. earthquakes) in space and time. In addition to the locations of the objects represented by the points, qualitative or quantitative information on the objects may be available (e.g. health status of individuals, size of the habitat or whether it is occupied or not, magnitude of an earthquake). These additional variables are commonly referred to as *marks*. To improve the understanding of spatio-temporal dynamics these marks can be considered alongside the pattern and jointly fitted as part of a complex marked point process model (Illian

et al. 2012b, Illian et al. 2013, Jones- Todd et al. 2018). In ecological studies these have been used to model the spatial distribution of above ground biomass (Ledo et al. 2016), to assess the processes that regulate spatial locations of muskoxen (*Ovibos moschatus*) herds in Greenland (Illian et al. 2012a) and to estimate factors influencing the frequency of koala (*Phascolarctos cinereus*) visits to eucalyptus trees (Illian et al. 2012b). Typically marks are not the main focus in a marked point process model and are interpreted as additional data to inform the spatial structure of the pattern. However, in our study they are the main focus. The marks here relate to the presence (or absence) of cranes within a suitable habitat (i.e. wetland). The spatial pattern of wetlands is assumed to be a realisation of a log-Gaussian Cox process, and conditional on this we infer the spatial distribution of crane presence.

In other words, we jointly fit a model to i) the spatial pattern formed by the locations of wetlands in England and ii) the associated mark information on the presence or absence of cranes in each of these wetlands over time. The two components in the model share a common stochastic element, a spatio-temporal Gaussian random field that reflects spatio-temporal dynamics in the system. Clearly, in an ecological study, modelling the location of suitable habitat as such is not of as much interest in itself as is the spatial distribution of the target species. However, modelling the spatial pattern formed by the locations of suitable habitat alongside the marks accounts for the spatial autocorrelation in the variable of interest, and most importantly so, allows for prediction in time, given the spatial structure - and hence availability - of suitable habitat in space.

The main objective of this study is to illustrate the opportunities provided by joint spatio-temporal log-Gaussian Cox process (STLGCP) models to study and predict population range expansion for species where knowledge about key life history traits, such as rates and patterns of dispersal, is limited. For illustration, we use data on the recolonisation process of Eurasian cranes (*Grus grus*, cranes hereafter) in England. Here, we try to (i) determine the environmental and anthropogenic variables affecting crane breeding distribution; (ii) establish the effect of the colonisation dynamics in the species distribution, using the spatial autocorrelation structure (latent field) as a proxy; (iii) assess how the colonisation dynamics might change over time, through the temporal variation of the shape of the spatial autocorrelation structure; (iv) understand the effect of the reintroduction on the species range expansion; and (v) identify the most suitable wetlands for cranes and the wetlands more likely to be colonised in the near future. To fit the complex spatio-temporal model, we use a recently developed statistical model

fitting method, integrated nested Laplace approximation (INLA), designed to provide a fast and accurate approach to fitting complex spatio-temporal models.

## METHODS

### Study area and species data

Historically, cranes were resident birds in the UK, with presence all year around, but they were extirpated in the 16th century after hunting completely depleted the population (Gurney 1921, Stanbury 2011).

After they had disappeared, cranes only visited the UK as vagrants (Stanbury 2011). In 1979, more than 400 years after the local extirpation of the breeding population a pair established in a wetland area in the Norfolk Broads, eastern England (Stanbury 2011). Over the following years the population grew slowly, to 13-14 breeding pairs in 2010 (Stanbury 2011). The current distribution in the UK comprises areas in the east of the UK, in Norfolk, Cambridgeshire, Oxfordshire, Yorkshire, and Aberdeenshire (Stanbury 2011). To boost the species population in the UK, between 2010 and 2014, 90 juvenile birds were released in the Somerset Levels ( $51^{\circ} 2' N$ ,  $2^{\circ} 55' W$ ; 56,650 ha), a wetland area situated in SW England, ca. 250 km away from the naturally established population (Soriano-Redondo et al. 2016).

Cranes use a wide variety of wetland habitats for breeding (BirdLife-International 2017). For this reason, we aim to determine the wetland features influencing the species colonisation in England. To do so, we mapped the wetlands across England combining information from two sources: the UK Land Cover Map 2007 (Morton et al. 2011) and the Wetland Vision map of current wetlands (Hume 2008). We overlaid both maps to define the wetland areas because the UK Land Cover Map 2007 has a similar resolution across the country, while the Wetland Vision map focuses on major wetland areas. We limited our study to England because the Wetland Vision map of current wetlands does not cover Scotland and Wales (Hume 2008). We only considered wetland areas over 5 ha since the highest nesting densities recorded for the species are around 8 ha/pair (Johnsgard 1983). In addition, since in the UK cranes have been exclusively nesting in lowland wetlands we excluded wetlands situated in the moorland areas (Stanbury and Sills 2012) based on information from Natural England data products (<http://www.natureonthemap.naturalengland.org.uk>).

For each wetland we calculated four different metrics that could affect habitat suitability: (i) wetland extent, (ii) perimeter-to-area ratio, (iii) proportion of wetland in the surrounding landscape (10km

terrestrial buffer around the wetland), (iv) proportion of surrounding landscape (10km terrestrial buffer) that is urbanized, using urban and suburban areas in England mapped in the UK Land Cover Map 2007. We scaled the four metrics by subtracting the mean and dividing the resulting value by the standard deviation. To model the temporal dynamics of cranes we used an auto-regressive process of order one, AR(1), since our data indicate that a crane breeding location is highly dependent on the breeding location of the previous year.

To determine whether a wetland had been colonised by cranes, we recorded crane nesting locations (from 2000 to 2015) using a range of sources across England: landowners, site-managers, the Rare Breeding Birds Panel (<http://www.rbbp.org.uk/>), and the Scarce Migrant Reports in the journal *British Birds*. We considered that a wetland was occupied by breeding cranes when (i) an incubating bird or a chick was seen, (ii) there was post-breeding evidence of eggs being laid or young being present, (iii) a crane family was seen post-fledging or (iv) adults were acting in a manner that strongly suggested the presence of a nest or young. From 2013 onwards (the year in which released cranes first bred) we also recorded the nesting locations of the reintroduced birds. In the few cases where there was uncertainty about the exact location of the nesting site, we assigned it to the closest wetland area. Imperfect detection is unlikely to have had an impact because cranes are very recognisable birds and it is improbable that any nesting location was missed. In addition, we used cranes presence/absence as the mark which is a more conservative approach than using the number of pairs in the wetland, which can be subject to bigger errors.

### **Marked log-Gaussian Cox process**

As mentioned above, to estimate the drivers of crane distribution in England and predict crane occurrence over time, we used a spatio-temporal point process model, specifically a marked STLGCPC to jointly model the wetland locations as a point pattern and the presence of cranes as a mark. A STLGCPC is a mathematically tractable type of Cox process (Cox and Isham 1980) where the log-intensity of the process is a spatial Gaussian random field with some mean, variance, and correlation function (Møller et al. 1998). In our model we use the Matérn covariance function to describe the spatial-autocorrelation between points. Given a realization of the Gaussian random field, in year  $t$  the STLGCPC has a Poisson number of points with mean intensity  $\lambda(\mathbf{s}, \mathbf{t})$ . This construction implies that given the intensity function,

the objects represented by the points, are scattered independently (Rathbun 1996; Møller et al. 1998).

Specifically, the intensity  $\Lambda(\mathbf{s}, \mathbf{t})$  we use here is given by,

$$\Lambda(\mathbf{s}, \mathbf{t}) = \exp(\alpha_0 + \mathbf{G}(\mathbf{s}, \mathbf{t}))$$

Here  $\alpha_0$  is an intercept term that we set to zero.  $\mathbf{G}(\cdot)$  is a Gaussian random field, accounting for spatial structures in the point pattern and the exponential ensures that the intensity is always positive. This intensity, in the context of the application presented here represents the density of wetlands across the UK. Understanding the spatial distribution of wetlands is in itself not of interest here. What is of interest, however, is the spatial distribution of crane breeding pairs given the spatial distribution of the wetlands. As mentioned above we treat the information whether a breeding pair can be found in a particular wetland or not as a mark and model it alongside the spatial pattern, rather than in isolation.

The joint modelling framework of locations of wetlands and the presence of cranes (marks) utilizes multiple latent Gaussian fields and multiple likelihoods to represent the spatial processes inherent in the data through space and time. In addition, each wetland has an associated  $m(\mathbf{s}, \mathbf{t})$ , which is assumed to follow a probability distribution and changes in space and over time. In the application discussed herein, this mark is a binary variable indicating the presence or absence of crane breeding pairs. This mark is assumed to follow a Bernoulli distribution with the spatio-temporal probability of crane presence conditional on the locations of wetlands being given by  $p(\mathbf{s}, \mathbf{t})$  (i.e.  $m(\mathbf{s}, \mathbf{t}) \sim \text{Bernoulli}(p(\mathbf{s}, \mathbf{t}))$ ). Hence, in addition to the intensity we also model the marks.

$$\mathbf{p}(\mathbf{s}, \mathbf{t}) = \text{logit}^{-1}(\beta_0 + \sum_{q=1}^m \alpha_q x_q(\mathbf{s}, \mathbf{t}) + \beta \mathbf{G}(\mathbf{s}, \mathbf{t}) + \mathbf{g}(\mathbf{s}, \mathbf{t})).$$

Here  $\beta_0$  is an intercept term that we set to zero and  $x_q(\mathbf{s}, \mathbf{t})$  refers to the value of the  $q$ th covariate at location  $\mathbf{s}$  at time  $\mathbf{t}$ . Both  $\mathbf{g}(\mathbf{s}, \mathbf{t})$  and  $\mathbf{G}(\mathbf{s}, \mathbf{t})$  are zero-mean spatio-temporal Gaussian random fields, which are approximated by an SPDE model in space that follows an AR(1) process over time (2000-2015) with dependence parameter  $\rho$  (Lindgren et al. 2011, Cameletti et al. 2013) The latent field  $\mathbf{G}(\mathbf{s}, \mathbf{t})$  is assumed to reflect both the log spatio-temporal intensity of the LGCP describing the location of wetlands as well as their availability to the crane population as they are expanding their range through time. The strength and direction of the interaction between the process governing the probability of observing a

breeding pair and that governing the spatial distribution of wetlands is reflected in the parameter  $\beta$ . For example a positive  $\beta$  would indicate that in areas of higher wetland density the probability of observing a pair of breeding cranes is greater. The random field  $g(s, t)$  is the spatially varying process referring to the probability of presence of a breeding pair of cranes, which is conditionally independent of wetland density. In combination with the intercept term,  $\beta_0$ , which is constant in space, it reflects those processes that are not captured by environmental covariates contained in the model. We also use the four scaled environmental variables mentioned above, as covariates  $x_q(s, t)$  for the model of the mark. We assess their importance in inferring the spatio-temporal process describing the probability of observing a breeding pair conditional on wetland location through estimating the respective regression coefficients  $\alpha_q$  (further information in Appendix 1).

As mentioned, this joint model is fitted using a computationally efficient Bayesian model fitting procedure, INLA, along with the SPDE approximation to the shared latent spatio-temporal process. The section below briefly summarises how these approaches are used in the application discussed in this article (Rue et al. 2009, Lindgren et al. 2011). In addition, we provide an online tutorial and a simulated dataset to illustrate fitting a marked STLGCP as discussed above ([https://github.com/cmjt/examples/blob/master/species\\_distribution.md](https://github.com/cmjt/examples/blob/master/species_distribution.md)).

### **Fitting latent Gaussian models using the INLA-SPDE approach**

The INLA approach is designed to fit complex spatio-temporal models (formally the class of latent Gaussian models) in a computationally efficient way. The joint model structure we use here that contains a Gaussian random field in space and an AR(1) process in time is a case where we strongly benefit from the computational efficiency of INLA. In addition we use an SPDE as an approximation to a Gaussian random field as introduced by Lindgren et al. 2011. Here, weighted sums of basis functions are used to approximate the spatial random functions arising from the solution to the SPDE. Thus, the continuous interpretation of space is preserved, whilst benefiting from the computational advantages of the approximation and the INLA approach.

### **Model implementation**

The model was fitted using the R-library R-INLA, which has functionality for applying the SPDE approach (Rue et al. 2009; Lindgren et al. 2011). In practice, a Delaunay triangulation (called a “mesh”) is constructed on the spatial domain. This mesh is the structure on which the representation of the field is based (Fig 1). To avoid edge effects (i.e. increased variance at borders) in the latent field prediction we extended the mesh beyond the English border. Since we are modelling in a Bayesian context, we need to specify priors for all parameters; we used normal uninformative priors (with mean set to 0 and variance set to 10) for the regression parameters.

To understand the effect of the reintroduction on the species range expansion we run two main models: the first only comprising the natural population (NAT model hereafter) and the second one adding the nesting locations of the reintroduced birds (NAT-RE model hereafter).

Measures to test the robustness for log-Gaussian Cox models are not well established and remain under development. Therefore, to test the strength of the results we run three additional models: the first including the nesting pairs from the natural population from 2000 to 2010 (note that during that period both datasets, including and excluding the reintroduced birds, are identical), the second one including the nesting pairs from the natural population from 2011 to 2015 and the third one including the nesting pairs from the natural population and the reintroduced population from 2011 to 2015. In addition, to illustrate the diagnostic ability of our approach, we plotted a receiver operating characteristic (ROC) curve of the NAT model, which compares the fitted probability of presence at each wetland and the observed presence/absence (Appendix 2 Fig S2) (Mason and Graham 2002).

Finally, we determined the 20 most likely areas to be colonised using the two models (NAT and NAT-RE). For each model we also calculated two scenarios for the predicted values, the first using only the estimates of the significant covariates and the second one adding the spatial autocorrelation structure in the predicted values.

## RESULTS

### Changes in wetland occupancy by breeding pairs

We mapped a total of 2,526 wetlands across England ranging from 5 to 24,707 ha with perimeter-to-area ratios ranging from 0.002 to 0.077. The proportion of surrounding wetland and the proportion of

urbanized areas ranged from 0 to 0.57 and from 0 to 0.81, respectively. When considering the natural population only (NAT dataset), the number of occupied wetlands increased steadily in the 15-year time period, from one wetland in 2000 in the Norfolk Broads to seven in 2015 spread across Norfolk, Cambridgeshire and Yorkshire (Fig 1 and Fig S1). The NAT-RE dataset, which included the reintroduced birds only differed from the NAT dataset from 2013 onwards, i.e. from when the reintroduced birds started nesting in several areas in Somerset, Gloucestershire and Oxfordshire (Fig 1). The total number of breeding pairs reached 30 in 2015 spread over 12 wetland areas across England (Fig 1).

### **Drivers of crane colonisation**

In the NAT model, wetland extent positively affected crane colonisation probability, i.e. cranes were more likely to settle in larger wetlands (Table 1). This is in contrast to the perimeter-to-area ratio, which had a negative effect, i.e. cranes preferred wetlands where edges were minimised (Table 1). There was no evidence that the two landscape features, proportion of wetland and urban areas in the surroundings affected the cranes' establishment probability (Table 1). The results from the NAT-RE model were similar: wetland extent and the perimeter-to-area ratio showed the same pattern (Table 2). However, in this instance the proportion of surrounding wetland area had a small negative effect on colonisation probability (Table 2). The results from the three additional models were similar: wetland extent had a significant positive effect in the two models including only the nesting pairs from the natural population from 2000 to 2010 and from 2011 to 2015 (Appendix 2 Table S1 and S2), and perimeter-to-area ratio had a significant negative effect in the model including the nesting pairs from the natural population from 2000 to 2010 and the model including the nesting pairs from the natural population and the reintroduced population from 2011 to 2015 (Appendix 2 Table S1 and S3). In none of the three models the proportion of surrounding wetland and urban areas had significant effects. Therefore, we focused on the effect of the wetland extent and the perimeter-to-area ratio in the discussion section. To assess model fit we use a ROC curve (Appendix 2 Fig S2) which indicates that the NAT-model produces a higher proportion of true-positives than false-positives, i.e. it predicts the presence of nesting cranes more often in wetlands where cranes are observed than in areas where they are not.

The spatial autocorrelation structure as reflected in the random field  $g(\mathbf{s}, t)$  had a strong influence on the species' distribution, especially when considering only the natural population (Fig 2a). Nevertheless, the spatial autocorrelation structure had a substantial uncertainty, which increased with the distance from the

colonised wetlands (Fig 2b). However, when the reintroduced birds were accounted for, the intensity of the spatial autocorrelation structure decreased (Fig 2c). Likewise, the uncertainty was lower, although there was increased uncertainty further away from the colonised wetlands (Fig 2d). Thus, the weight of the spatial autocorrelation structure relative to the covariates was higher in NAT dataset models than in NAT-RE models. We also found an extremely high temporal autocorrelation ( $\rho > 0.99$  in both models) in the spatial autocorrelation structure from one year to the next, implying that the spatial structure is unlikely to change in the near future, which also means that the colonisation process is progressing at an extremely low rate (Tables 1 and 2).

With respect to the NAT model, when accounting for the spatial autocorrelation and the significant covariates, we found that the 20 wetlands most likely to be colonised were in close proximity to the currently occupied wetlands in eastern England (Fig 3a). However, when we ignored the spatial autocorrelation and simply determined the more suitable wetlands for cranes based on the covariates, we found that they were spread across much of England (Fig 3b). The NAT-RE model followed a similar pattern: when considering the spatial autocorrelation structure and the significant variables, the most likely areas to be colonised were near to the already occupied wetlands (Fig 3c). Likewise, the more suitable wetlands in the NAT-RE model were spread across England (Fig 3d). Moreover, the site where cranes were reintroduced in the Somerset Levels obtained the third highest score using the NAT model estimates. It is also encouraging to note that the reintroduced cranes have nested in two of the top 20 wetlands predicted by the NAT model (Fig 3b). Of the top 20 wetlands that each model highlighted, seven were selected by both NAT and NAT-RE models.

## DISCUSSION

In this study, for the first time, point process models were applied to gain an understanding of and to predict the recolonisation process of a bird species in a dynamic population. Specifically, we used marked log-Gaussian Cox process models to model the spatio-temporal distribution of cranes in England, where the species is recolonising several wetland areas, through immigration from mainland Europe and a subsequent reintroduction in SW England, after the extirpation of its breeding population in the 16<sup>th</sup> century. We found that crane colonisation is progressing at a very slow rate and that they favour areas near already colonised wetlands. However, within those areas, cranes are selecting large wetlands with

low perimeter-to-area ratios. In addition, the crane reintroduction has increased the probability of recolonisation in the west of England.

The main advantage of log-Gaussian Cox process models is that they account for spatial structure that remains unexplained by the covariates by using a latent field that can be modelled over time and captures the dynamism of ecological systems. In our case, the spatial autocorrelation structure was a proxy for the colonisation process; since cranes started recolonising England from the east, the occupied wetland tend to be restricted to the eastern part of the country, even though there are apparently suitable wetlands further afield. In ecological terms the factors affecting the spatial structure of the occupied wetlands are likely to be natal philopatry and conspecific attraction (Timoney 1999) which would explain why occupied wetlands aggregate in clusters in some areas, such as in the Norfolk Broads and Cambridgeshire Fens, with cranes tending to settle adjacent to areas that are already occupied (Timoney 1999). These phenomena are well captured in the spatial autocorrelation structure, which show a gradient through the east-west axis, with higher values on the eastern side. The high temporal autocorrelation obtained in our models show that the spatial autocorrelation structure should remain constant over time, which indicates that turnover of occupied wetlands between years is low and as such colonisation will progress slowly (Table 1 and 2). Nevertheless, it is important to note that the spatial autocorrelation structure could also be capturing spatio-temporal aspects other than the colonisation process that we are unaware of.

Overall, our models suggest that cranes prefer large wetlands with low perimeter-to-area ratios (i.e. low amounts of edge habitat) (Tables 1 and 2). Nevertheless, the four selected covariates only captured general wetland features; therefore, there could be a number of site-level characteristics relevant for crane colonisation that remained unaccounted for. We observe that at a country-level, the spatial autocorrelation structure of the marked point process is more relevant in determining the potential colonisation areas than the spatial characteristics of the wetlands, with wetland near to occupied sites having a higher probability of occupancy than wetlands further away (Figs 2a and 2c). However, at a regional-level, differences in the spatial autocorrelation structure among wetlands are reduced and the importance of the environmental covariates increases significantly. Therefore, assuming constant environmental conditions and the same

rate of spread as the previous 15 years, our study indicates that cranes will colonise large wetlands with low perimeter-to-area ratios near to the currently occupied wetlands.

Interestingly, the two covariates that positively affected crane colonisation (in both the NAT and the NAT-RE model), large wetland extent and low perimeter-to area ratio, are indicators of high quality habitat and low fragmentation (Laurance and Yensen 1991). These variables are linked to high species richness, abundance and nest success in several grassland and wetland bird species, such as upland sandpiper (*Bartramia longicauda*), purple heron (*Ardea purpurea*) and little bittern (*Ixobrychus minutus*) (Burger et al. 1994, Helzer and Jelinski 1999; Winter et al. 2006, Benassi et al. 2007). Several mechanisms may explain birds' preference for larger wetlands: higher habitat heterogeneity (Golet et al. 2001), higher conspecific attraction (Benoit and Askins 2002), longer hydroperiods (Snodgrass et al. 2000), adequate territory size requirements (Worrall et al. 1997) or higher invertebrate food availability (Euliss and Mushet 1999, Schulz 2004, Tarr et al. 2005). In addition, in England, most of the remaining large wetlands are under protection being commonly designated as Sites of Special Scientific Interest (SSSI, Natural England, <https://designatedsites.naturalengland.org.uk/>) and there is evidence that protected wetlands have been used by colonising birds as stepping stones for the permanent population establishment (Hiley et al. 2013).

The comparison between the NAT and the NAT-RE model produced some interesting information about the effect of the reintroduction on potential population spread. Our results indicate that, prior to the reintroduction, even though the model predicted a westward spread, the probability of the species reaching the west of England was extremely low, since the spatial autocorrelation structure should remain constant over time (Fig 2a). Thus, even large wetlands with low perimeter-to-area ratios scored substantially low colonisation probabilities compared to less suitable sites in eastern England. Our ability to predict the likelihood of colonisation (Fig 2b) in wetlands further away from the core colonised areas is hindered by the uncertainty associated to the spatial autocorrelation structure. However, after the reintroduction, the potential colonisation areas spread along an east-west axis in the south of England (Fig 2c). Remarkably, the spatial autocorrelation decreased dramatically in the RE-NAT model (Figs 2a and 2c), implying that the importance of relevant covariates (the habitat) in the colonisation process increased, which has relevant ecological implications: reintroduced cranes are tending to colonise wetlands more as a function of habitat suitability than proximity to already occupied wetlands. This increase in dispersal

might be due to the particular aspects of the reintroduction. For example, large numbers of birds reaching maturity the same year, in contrast to the natural population where the process is gradual, could have increased competition for suitable breeding territories locally which forced the pairs to disperse further away; and/or the effect of the natal philopatry could have been reduced because the reintroduced birds were hand-reared in enclosed facilities. Thus, even though our results suggest areas far away from the current core range should remain unoccupied, the reintroduction could be seen as human-assisted long-distance dispersal event that would promote the population spread westward. Moreover, there is already evidence of connectivity between the natural and the reintroduced population, with a pair formed by one individual from each population nesting in 2015.

To date, log-Gaussian Cox process models have been applied in several areas of research, such as meteorology or policy (Illian et al. 2012b, Illian et al. 2013, Cosandey-Godin et al. 2014, Gómez-Rubio et al. 2015). However, these modelling methods do not use a joint approach as suggested here. Thus, this study provides a new approach to modelling spatial change in dynamic populations. Although we focus on a population that is expanding its range, this methodology can be used when population ranges are contracting or shifting and it could be particularly relevant for the study and conservation of rare species under climate change, where information on life history traits is lacking and management strategies are time sensitive. Another field that could benefit from this approach is the study of biological invasion, since the study of particular invasion events is usually challenging due to small sample sizes and the influence of local factors (Guisan et al. 2014).

STLGCP models could help predict the spatial change of population distribution through the modelling of the latent field even when the underlying causes for that change are only partially understood. In our case we only modelled potentially suitable habitat (wetlands), but this approach could be extended by adding non-suitable habitat in the model when source-sink dynamics are suspected. Moreover, density dependence processes could be added by modelling population counts rather than presence/absence data. Furthermore, another potential application of this method would be to jointly model changes in population distribution and habitat over time, which would be particularly relevant in light of climate change and habitat fragmentation. Finally, we want to make a call to encourage cross-disciplinary collaborations between ecologists and statisticians to shed light on complex processes that require cutting-edge methodologies.

## ACKNOWLEDGMENTS

We thank the members of the UK Crane Working Group for recording cranes nest locations and breeding outcomes for over 30 years. We also thank the two anonymous reviewers for their comments to improve the manuscript. We are grateful to the Great Crane Project, a partnership between the Wildfowl and Wetlands Trust, the Royal Society for the Protection of Birds and Pensthorpe Conservation Trust, with major funding from Viridor Credits Environmental Company, for their financial and logistic support. SB is funded by an EU consolidator's grant: STATEMIG 310820. ASR is supported by a joint pre-doctoral grant from the University of Exeter, the Wildfowl and Wetlands Trust and the Royal Society for the Protection of Birds.

**Author contributions:** A.S.R, C.J.T, S.B., G.M.H, S.C.V and J.B.I designed the study, A.S collected the data, A.S.R and C.J.T analysed the data and A.S.R drafted the paper, all authors contributed substantially to revising and preparing the paper; J.B.I supervised all stages of the research.

**Data accessibility:** Computer code is available in the supplementary material and GitHub ([https://github.com/cmjt/examples/blob/master/species\\_distribution.md](https://github.com/cmjt/examples/blob/master/species_distribution.md)).

## REFERENCES

Benassi, G. et al. 2007. Area effect on bird species richness of an archipelago of wetland fragments in Central Italy. – *Community Ecol.* 8: 229-237.

Benoit, L. K. and Askins, R. A. 2002. Relationship between habitat area and the distribution of tidal marsh birds. – *Wilson Bull.* 114: 314-323.

BirdLife-International 2017. Species factsheet: *Grus grus*. – Downloaded from <http://www.birdlife.org> on 16/01/2017.

Brown, C. et al. 2016. Success of spatial statistics in determining underlying process in simulated plant communities. – *J. Ecol.* 104: 160-172.

Burger, L. D. et al. 1994. Effects of prairie fragmentation on predation on artificial nests. – *J. Wildl. Manag.*: 249-254.

- Cameletti, M. et al. 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. – *AStA Adv. Stat. Anal.* 97: 109-131.
- Cosandey-Godin, A. et al. 2014. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. – *Can. J. Fish. Aquat. Sci.* 72: 186-197.
- Cox, D. R. and Isham, V. 1980. Point processes (Vol. 12). – CRC Press.
- Deinet, S. et al. 2013. Wildlife comeback in Europe: The recovery of selected mammal and bird species. – Zoological Society of London.
- Dormann, C. F. et al. 2012. Correlation and process in species distribution models: bridging a dichotomy. – *J. Biogeogr.* 39: 2119-2131.
- Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – *Annu. Rev. Ecol. Evol. Syst.* 40: 677-697.
- Euliss, N. H. and Mushet, D. M. 1999. Influence of agriculture on aquatic invertebrate communities of temporary wetlands in the prairie pothole region of North Dakota, USA. – *Wetlands* 19: 578-583.
- Fitzpatrick, M. C. et al. 2012. Modeling range dynamics in heterogeneous landscapes: invasion of the hemlock woolly adelgid in eastern North America. – *Ecol. Appl.* 22: 472-486.
- Gallien, L. et al. 2010. Predicting potential distributions of invasive species: where to go from here? – *Divers. Distributions* 16: 331-342.
- Golet, F. C. et al. 2001. Relationship between habitat and landscape features and the avian community of red maple swamps in southern Rhode Island. – *Wilson Bull.* 113: 217-227.
- Gómez-Rubio, V. et al. 2015. Analysis of massive marked point patterns with stochastic partial differential equations. – *Spat. Stat.* 14: 179-196.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – *Ecol. Lett.* 8: 993-1009.

Guisan, A. et al. 2014. Unifying niche shift studies: insights from biological invasions. – *Trends Ecol. Evol.* 29: 260-269.

Gurney, J.H. 1921. *Early Annals of Ornithology*. – Witherby.

Haddad, N. M. et al. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. – *Sci. Adv.* 1: e1500052.

Hastings, A. et al. 2005. The spatial spread of invasions: new developments in theory and evidence. – *Ecol. Lett.* 8: 91-101.

Helzer, C. J. and Jelinski, D. E. 1999. The relative importance of patch area and perimeter–area ratio to grassland breeding birds. – *Ecol. Appl.* 9: 1448-1458.

Hiley, J. R. et al. 2013. Protected areas act as establishment centres for species colonizing the UK. – *Proc. R. Soc. Lond. B Biol. Sci.* 280: 20122310.

Hume, C. 2008. *Wetland Vision Technical Document: overview and reporting of project philosophy and technical approach*. – The Wetland Vision Partnership: 80.

Illian, J. B. et al. 2012a. Using INLA to fit a complex point process model with temporally varying effects—a case study. – *J. Env. Stat.* 3: 7.

Illian, J. B. et al. 2012b. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). – *Ann. Appl. Stat.*: 1499-1530.

Illian, J. B. et al. 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. – *Methods Ecol. Evol.* 4: 305-315.

Johnsgard, P. A. 1983. *Cranes of the World*. – Cambridge Univ. Press.

Jones-Todd, C. M. et al. 2018. A spatiotemporal multispecies model of a semicontinuous response. – *J. R. Stat. Soc. Ser. C Appl. Stat.* 67: 705-722.

- Laurance, W. F. and Yensen, E. 1991. Predicting the impacts of edge effects in fragmented habitats. – *Biol. Conserv.* 55: 77-92.
- Law, R. et al. 2009. Ecological information from spatial patterns of plants: insights from point process theory. – *J. Ecol.* 97: 616-628.
- Ledo, A. et al. 2016. Lianas and soil nutrients predict fine-scale distribution of above-ground biomass in a tropical moist forest. – *J. Ecol.* 104: 1819-1828.
- Lenoir, J. et al. 2008. A significant upward shift in plant species optimum elevation during the 20th century. – *Science* 320: 1768-1771.
- Lindgren, F. et al. 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: 423-498.
- Mang, T. et al. 2017. Accounting for imperfect observation and estimating true species distributions in modelling biological invasions. – *Ecography* 40: 1187-1197.
- Mantyka-pringle, C. S. et al. 2012. Interactions between climate and habitat loss effects on biodiversity: a systematic review and meta-analysis. – *Glob. Change Biol.* 18: 1239-1252.
- Mason, S. J. and Graham, N. E. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. – *Quarterly Journal of the Royal Meteorological Society* 128: 2145-2166.
- Matthiopoulos, J. et al. 2011.. Generalized functional responses for species distributions. – *Ecology* 92: 583-589.
- Matthiopoulos, J. et al. 2015. Establishing the link between habitat selection and animal population dynamics. – *Ecol. Monogr.* 85: 413-436.
- Møller, J. et al. 1998. Log Gaussian Cox processes. – *Scand. J. Stat.* 25: 451-482.

- Mooney, H. A. and Cleland, E. E. 2001. The evolutionary impact of invasive species. – Proc. Natl. Acad. Sci. U.S.A. 98: 5446-5451.
- Moritz, C. et al. 2008. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. – Science 322: 261-264.
- Morton, D. et al. 2011. Final Report for LCM2007-the new UK land cover map. – Countryside Survey Technical Report No 11/07.
- Rathbun, S. L. 1996. Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. J. Stat. Plan. Inference 51: 55-74.
- Rue, H. et al. 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: 319-392.
- Schulz, R. 2004. Field studies on exposure, effects, and risk mitigation of aquatic nonpoint-source insecticide pollution. – J. Environ. Qual. 33: 419-448.
- Snodgrass, J. W. et al. 2000. Relationships among isolated wetland size, hydroperiod, and amphibian species richness: implications for wetland regulations. – Conserv. Biol. 14: 414-419.
- Soriano-Redondo, A. et al. 2016. Ecological Responses to Extreme Flooding Events: A Case Study with a Reintroduced Bird. – Sci. Rep. 6: 28595.
- Stanbury, A. 2011. The changing status of the Common Crane in the UK. British Birds, 104, 432-447.
- Stanbury, A. and Sills, N. 2012. Common Crane habitats in Britain. – British Wildlife 23: 381.
- Sun, Y. et al. 2016. Predicting and understanding spatio-temporal dynamics of species recovery: implications for Asian crested ibis *Nipponia nippon* conservation in China. – Divers. Distributions 22: 893-904.

- Tarr, T. et al. 2005. Macroinvertebrate community structure across a wetland hydroperiod gradient in southern New Hampshire, USA. – *Wetl. Ecol. Manag.* 13: 321-334.
- Tauler, H. et al. 2015. Identifying key demographic parameters for the viability of a growing population of the endangered Egyptian Vulture *Neophron percnopterus*. – *Bird Conserv. Int.* 25: 426-439.
- Timoney, K. 1999. The habitat of nesting whooping cranes. – *Biol. Conserv.* 89: 189-197.
- Winter, M. et al. 2006. Patch size and landscape effects on density and nesting success of grassland birds. – *J. Wildl. Manag.* 70: 158-172.
- Worrall, P. et al. 1997. Constructed wetlands and nature conservation. – *Wat. Sci. Tech.* 35: 205-213.
- Yuan, Y. et al. 2017. Point process models for spatio-temporal distance sampling data. – arXiv preprint arXiv:1604.06013.
- Zurell, D. et al. 2012. Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps. – *Ecography* 35: 590-603.

**Figure Legends**

Figure 1: Delaunay triangulation (mesh) over the study area. Green dots represent the 2,526 wetlands in England, black dots the wetlands occupied by nesting pairs from the natural population and grey dots the wetlands occupied by nesting pairs from the reintroduced population in 2015.

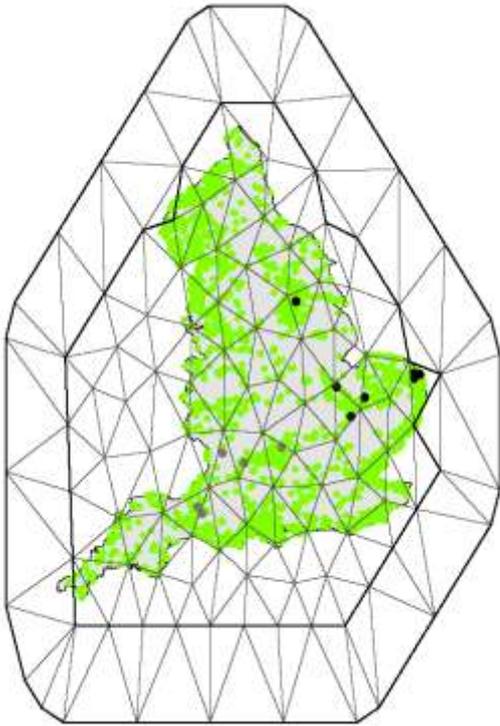


Figure 2: The (a) posterior mean and (b) the posterior standard deviation of the random field  $g(\mathbf{s}, t)$  (including the intercept  $\beta_0$ ) in 2015 estimated in the NAT model which, only includes the nesting pairs from the natural population; and the (c) posterior mean and (d) the posterior standard deviation of the random field  $g(\mathbf{s}, t)$  (including the intercept  $\beta_0$ ) in 2015 estimated in the NAT-RE model, which includes the nesting pairs from the natural population and the reintroduced population. Once accounted for the effect of the covariates, given the spatial autocorrelation structure, the observations of cranes occurrence are considered to be independent. The spatial autocorrelation structure accounts for the colonisation dynamics. The higher values in the posterior mean indicate a stronger autocorrelation structure.

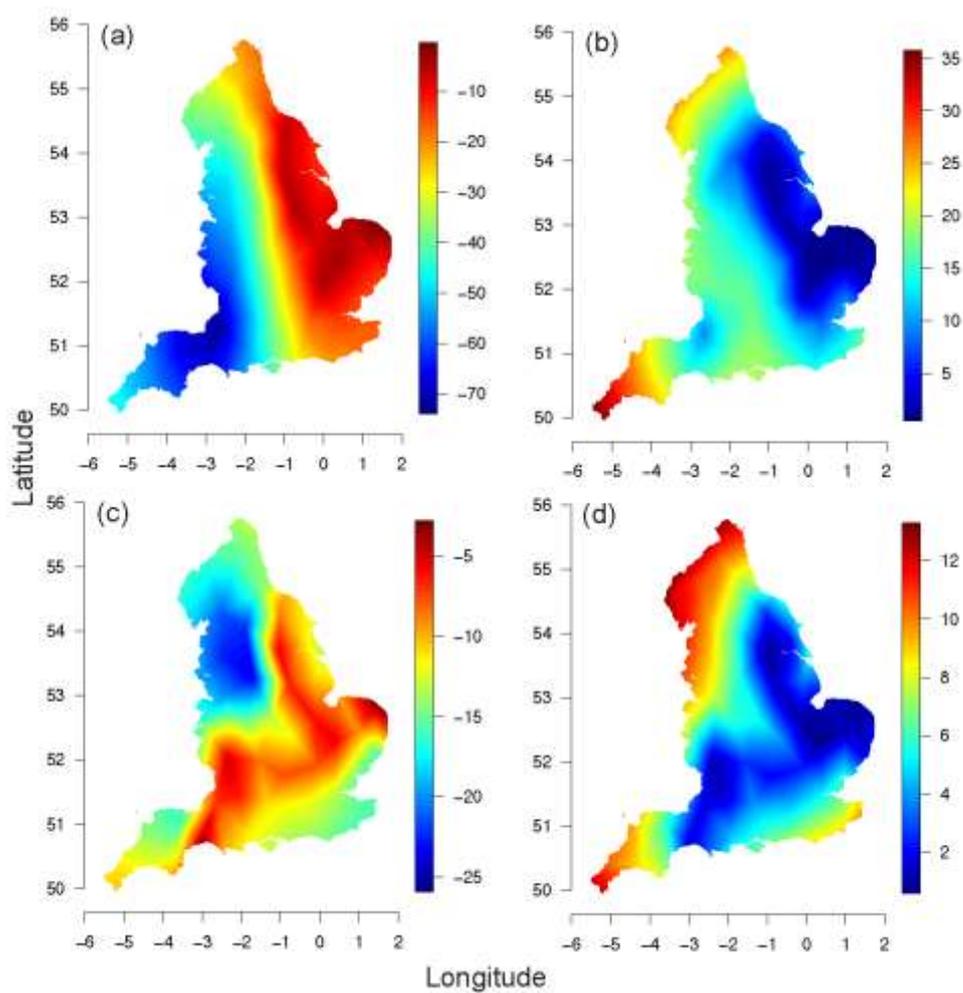
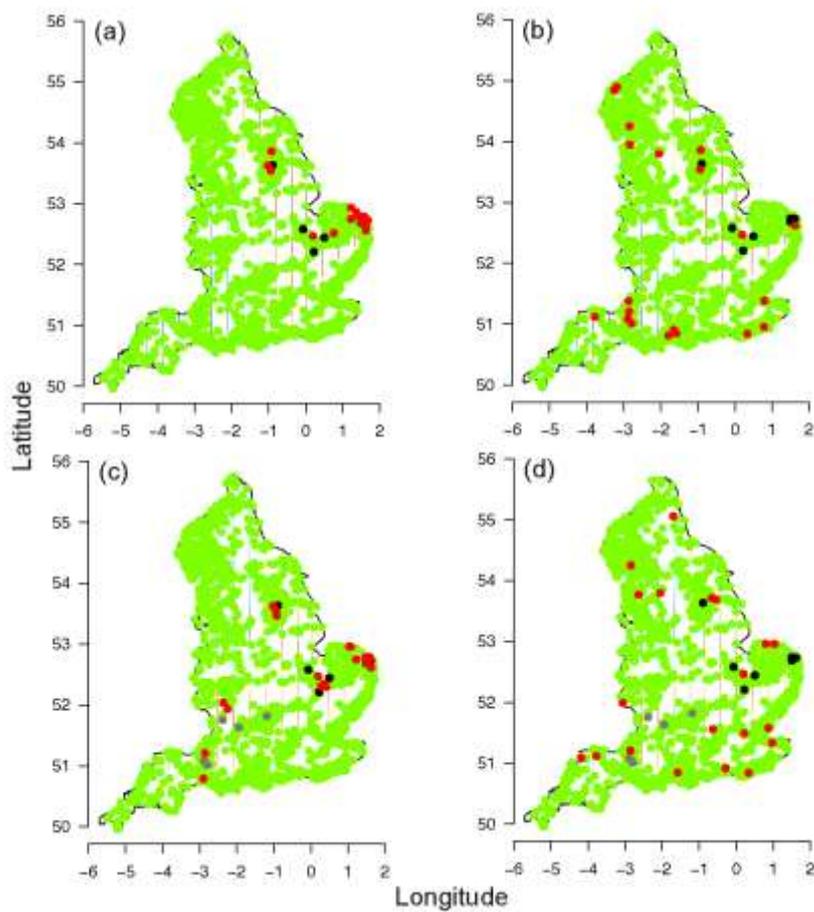


Figure 3: Wetlands most likely to be colonised in the near future. Green dots represent the 2,526 wetlands in England, black dots the wetlands occupied by nesting pairs from the natural population and grey dots the wetlands occupied by nesting pairs from the reintroduced population. The red dots represent the 20 most likely wetlands to be colonised, when (a) considering the posterior mean of the covariates and the spatial autocorrelation structure and (b) considering only the posterior mean of the covariates estimated in the NAT model which only includes the nesting pairs from the natural population; and when (c) considering the posterior mean of the covariates and the spatial autocorrelation structure and (d) considering only the posterior mean of the covariates estimated in the NAT-RE model which includes the nesting pairs from the natural population and the reintroduced population.



**Table Legends**

Table 1: Posterior means, standard deviations and 95% credible intervals for the covariates and random field  $g(\mathbf{s}, t)$  (spatial smoothness, scale and temporal autocorrelation of the random field and strength and direction of the response, respectively) estimated in the NAT model which only includes the nesting pairs from the natural population.

	Mean	SD	2.5% quant.	97.5% quant.
Wetland extent	1.7793	0.2042	<b>1.3784</b>	<b>2.1798</b>
Perimeter-to-area ratio	-0.9678	0.3575	<b>-1.6697</b>	<b>-0.2664</b>
Proportion of surrounding wetland	-0.6140	0.5329	-1.6604	0.4314
Proportion of surrounding urbanized areas	0.5525	0.5334	-0.4948	1.5990
Log( $\tau$ )	-3.9672	0.0335	-4.0386	-3.8480
Log( $\kappa$ )	-0.5706	0.0213	-0.6164	-0.4842
$\rho$	0.9998	0.0001	0.9995	1.0000
$\beta$	0.0458	0.0267	-0.0108	0.1402

Table 2: Posterior means, standard deviations and 95% credible intervals for the covariates and random field  $g(\mathbf{s}, t)$  (spatial smoothness, scale and temporal autocorrelation of the random field and strength and direction of the response, respectively) estimated in the NAT-RE model which includes the nesting pairs from the natural population and the reintroduced population.

	Mean	SD	2.5% quant.	97.5% quant.
Wetland extent	0.1234	0.0494	<b>0.0263</b>	<b>0.2203</b>
Perimeter-to-area ratio	-3.1559	0.3486	<b>-3.8403</b>	<b>-2.4721</b>
Proportion of surrounding wetland	-0.6920	0.3377	<b>-1.3550</b>	<b>-0.0296</b>
Proportion of surrounding urbanized areas	0.0402	0.3966	-0.7385	0.8183
Log( $\tau$ )	-3.5466	0.4247	-4.4008	-2.7320
Log( $\kappa$ )	-0.4607	0.2433	-0.9746	-0.0249
$\rho$	0.9992	0.0007	0.9973	0.9999
$\beta$	0.0795	0.3169	-0.5212	0.7213