

## RESEARCH ARTICLE

## Understanding spatial effects in species distribution models

Iosu Paradinas<sup>1,2\*</sup>, Janine Illian<sup>2</sup>, Sophie Smout<sup>2,3</sup>

**1** Scottish Ocean's Institute, University of St Andrews, East sands, St Andrews, United Kingdom, **2** AZTI, Txatxarramendi Ugarteia z/g, Sukarrieta, Bizkaia, Spain, **3** School of Mathematics and Statistics, University of Glasgow, Glasgow, United Kingdom

\* [ip30@st-andrews.ac.uk](mailto:ip30@st-andrews.ac.uk)

## Abstract

Species Distribution Models often include spatial effects which may improve prediction at unsampled locations and reduce Type I errors when identifying environmental drivers. In some cases ecologists try to ecologically interpret the spatial patterns displayed by the spatial effect. However, spatial autocorrelation may be driven by many different unaccounted drivers, which complicates the ecological interpretation of fitted spatial effects. This study aims to provide a practical demonstration that spatial effects are able to smooth the effect of multiple unaccounted drivers. To do so we use a simulation study that fit model-based spatial models using both geostatistics and 2D smoothing splines. Results show that fitted spatial effects resemble the sum of the unaccounted covariate surface(s) in each model.

## OPEN ACCESS

**Citation:** Paradinas I, Illian J, Smout S (2023) Understanding spatial effects in species distribution models. PLoS ONE 18(5): e0285463. <https://doi.org/10.1371/journal.pone.0285463>

**Editor:** Judi Hewitt, The University of Auckland - City Campus: University of Auckland, NEW ZEALAND

**Received:** August 1, 2022

**Accepted:** April 24, 2023

**Published:** May 30, 2023

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0285463>

**Copyright:** © 2023 Paradinas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** This paper is based on a simulation study. All R code is available in <https://github.com/iparperspective/Understanding-spatial-effects>.

## Introduction

Understanding and predicting species spatial patterns through Species Distribution Models (SDM) is pivotal for ecology, evolution and conservation [1]. SDMs quantify the relationship between species occurrence or abundance with biotic and abiotic factors in order to gain ecological and evolutionary understanding [2, 3]. This way SDMs allow us to predict distributions across landscapes and make future predictions based on identified drivers. Ideally, these factors would fully explain the distribution of the species under study, but this is practically unfeasible due to the large number of factors that drive the distribution of a species. In this regard, by including a spatial effect into a SDM, one can accommodate the spatial structure of the data that is unaccounted by our covariates.

Spatial autocorrelation refers to the dependence between pairs of observations in space. In SDMs, spatial effects allow us to predict better and reduce Type I errors in the presence of covariates [4, 5]. In species distribution, spatial autocorrelation arise due to unaccounted environmental or biotic drivers that are often hard to measure or estimate, for example a geographical range dispersion process or a highly dynamic processes such as wind and current [6–8]. These drivers can influence species distribution at all scales, from micrometres to continental and ocean-wide scales [9]. However, the size, spacing and extent of sampling units will constrain the scale of inferable drivers, and the scale of spatial autocorrelation [8, 10]. In other words, if we sample at a kilometer scale, we cannot infer processes at a smaller scale, and

**Funding:** IP was funded by a Marie Skłodowska-Curie Research Fellowship (GAP-847014). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

inversely, if our study area is one kilometer long, we cannot infer processes that affect at a larger scale.

The statistical interpretation of a spatial effect is related to the sign and link function of our linear predictor, but in general terms, positive values refer to areas where we expect more than that predicted by the rest of the linear predictor and vice versa. Ecologically, many SDM studies have linked spatial effects to biological features like home-range [6], hot-spot size [11] and unaccounted environmental drivers [12], providing reasonable arguments. For example, given a species that is driven by two environmental variables, one that drives the large-scale variation and another that drives the small-scale variation, the residual spatial pattern of a SDM that includes one of the two covariates will resemble the pattern of the unaccounted explanatory variable, either the large-scale or small-scale one. However, as we mentioned before, reality behind ecological processes is often high dimensional and variables that drive spatial correlation can occur at several different scales. In fact, SDMs are seldom able to identify more than a small portion of all the drivers that influence the distribution of the species under study. This results on spatial effects that are potentially driven by many different unaccounted drivers, diluting their interpretability in terms of an individual process. Although this interpretation issues have sporadically been addressed in the literature [7, 8, 13–17], many modellers fail to acknowledge this probably due to the lack of an explicit study that shows this.

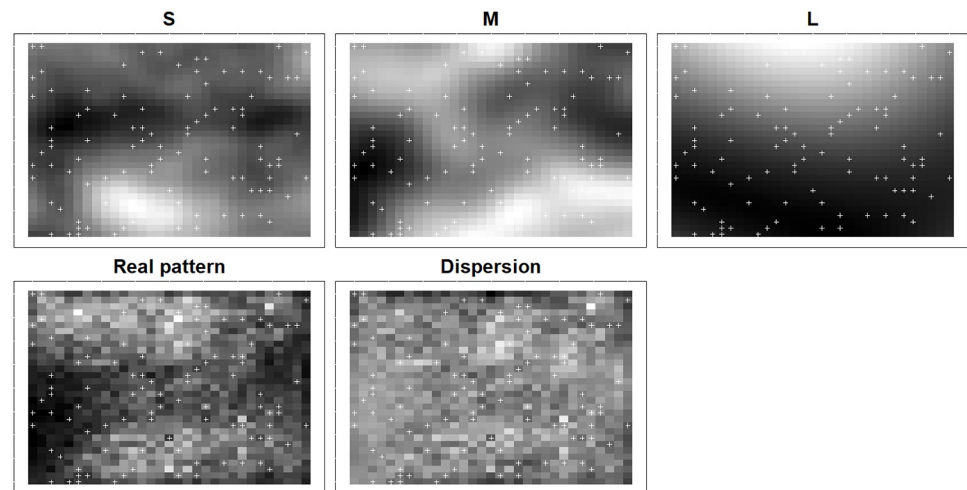
The aim of this article was to complement the MBGapp Shiny App, a user-friendly interface created by [18], used for teaching geostatistical analysis to scientists with only minimal statistical training. We provide a practical demonstration that spatial effects are able to smooth the effect of multiple unaccounted drivers, making the biological interpretation of spatial effects complicated. To do so, we used spatial models applied over simulated species distribution surfaces. Simulated fields were based on three spatially structured environmental covariates acting at different spatial scales, and a geographical range dispersion process.

## Methods

We used an iterative simulation approach to produce spatially aggregated distributions that include a dispersion effect. At each iteration we added a fixed number of new specimens to the study area based on a probability surface constituted by three spatially structured covariates, each operating at different scales (i.e., small, medium and large scale), and a spatial dispersion process driven by the abundance of the neighbouring areas, mimicking the colonization of a plant species for example. As a result, our simulated species abundance were driven by the sum of four different effects (Fig 1): the influence of three explanatory environmental variables operating at different spatial scales (S = small, M = medium and L = large) and a spatial dispersal effect that increase the spatial autocorrelation of the response variable.

We simulated fifty different scenarios, selected 100 random samples for each scenario and fitted all the possible combinations of Poisson spatial models that ranged from a purely spatial model to a full model that accounted for the three covariates (see Table 1). We used two spatial modelling approaches, geostatistics through the Integrated Nested Laplace Approximation approach (INLA) [19] and 2D smoothing splines through the MGCV package for R [20, 21] using their defaults settings.

Our aim was to assess the resemblance between fitted spatial effects and unaccounted covariate surface combinations. Resemblance was assessed through the similarity in pattern score (SIP) [22]. SIP scores are bound between zero and one, and high scores denote high similarity in pattern and vice versa. For each simulated scenario, we calculated the SIP score between the spatial effect of every fitted model (rows in Table 2) and all the possible different combinations of covariate surfaces (columns in Table 2), and recorded the absolute difference



**Fig 1. Visualization of the different autocorrelated drivers that influence the abundance pattern in a simulated scenario.** S, M and L refer to the small, medium and large scaled covariate fields, respectively. Dispersion refers to the geographical range dispersion. White crosses refer to the simulated 100 samples.

<https://doi.org/10.1371/journal.pone.0285463.g001>

between the best SIP score and the rest (i.e., SIP differences calculated per row in Table 2). This way, the spatial effect that best resembled a given combination of covariate surfaces scored a zero and that with the worst resemblance recorded the highest value (see Supporting information for a more detailed explanation of the procedure). As a result, we obtained fifty scores per model and combination of covariate surfaces. Finally, we summarised these scores by their mean and standard deviation. All the R script is available at <https://tinyurl.com/2p8n3e4r>.

## Results

Results show that fitted spatial effects resemble the sum of the unaccounted covariate surfaces in each model (see highlighted diagonal scores in Table 2). Fitted 2D splines using generalized additive models (GAM) seemed to perform a little worse than model based-geostatistics, probably due to the default selection of knots, but the overall pattern is very similar. This result suggests that spatial effects are able to smooth complex residual spatial patterns originated by a set of covariates that operate at very different scales. For example, model M\_M, which only accounts for the mid-scale covariate, estimates a spatial effect that resembles the sum of the

**Table 1. Summary of fitted models.** W refers to a geostatistical spatial correlation term, S, M and L refer to the small, medium and large scale covariates, respectively.

Model	Linear predictor	Missing covariates
M_0	$\beta_0 + W$	S, M & L
M_S	$\beta_0 + S + W$	M & L
M_M	$\beta_0 + M + W$	S & L
M_L	$\beta_0 + L + W$	S & M
M_ML	$\beta_0 + M + L + W$	S
M_SM	$\beta_0 + S + M + W$	L
M_SL	$\beta_0 + S + L + W$	M
M_SML	$\beta_0 + S + M + L + W$	–

<https://doi.org/10.1371/journal.pone.0285463.t001>

**Table 2. Resemblance between fitted spatial effects, using geostatistics and 2D smoothing splines, against all the possible combinations of covariate surfaces (per simulation).** Scores must be read by row, and reflect the difference between the best SIP score and all possible combinations of drivers for each simulation and model. Therefore, lower values represent higher resemblance and have been highlighted in bold. We present the mean difference and standard deviation (in parenthesis). See [Supporting information](#) for a more detailed explanation of the procedure that we followed.

	Model	Combination of drivers							
		Dispersion	S	M	L	S & M	S & L	M & L	S, M & L
Geostatistics (INLA)	M_0	0.62 (0.14)	0.30 (0.13)	0.27 (0.18)	0.35 (0.22)	0.11 (0.06)	0.17 (0.08)	0.12 (0.15)	<b>0.01 (0.02)</b>
	M_S	0.56 (0.18)	0.66 (0.22)	0.19 (0.16)	0.25 (0.19)	0.33 (0.17)	0.41 (0.16)	<b>0.01 (0.03)</b>	0.16 (0.12)
	M_M	0.47 (0.17)	0.19 (0.15)	0.71 (0.25)	0.26 (0.22)	0.26 (0.21)	<b>0.04 (0.08)</b>	0.37 (0.21)	0.11 (0.14)
	M_L	0.55 (0.17)	0.21 (0.14)	0.24 (0.21)	0.78 (0.35)	<b>0.04 (0.04)</b>	0.29 (0.20)	0.33 (0.24)	0.08 (0.12)
	M_SM	0.34 (0.17)	0.50 (0.24)	0.61 (0.28)	<b>0.07 (0.13)</b>	0.48 (0.26)	0.22 (0.15)	0.21 (0.18)	0.24 (0.20)
	M_SL	0.41 (0.23)	0.51 (0.21)	<b>0.08 (0.11)</b>	0.67 (0.35)	0.18 (0.11)	0.53 (0.25)	0.17 (0.20)	0.20 (0.16)
	M_ML	0.36 (0.18)	<b>0.06 (0.10)</b>	0.59 (0.24)	0.65 (0.26)	0.11 (0.11)	0.13 (0.14)	0.60 (0.24)	0.16 (0.17)
	M_SML	<b>0.09 (0.15)</b>	0.27 (0.16)	0.40 (0.22)	0.43 (0.24)	0.25 (0.16)	0.28 (0.18)	0.40 (0.24)	0.25 (0.22)
2D splines (GAM)	M_0	0.50 (0.09)	0.18 (0.10)	0.11 (0.08)	0.07 (0.07)	0.09 (0.08)	0.12 (0.08)	<b>0.04 (0.05)</b>	<b>0.04 (0.05)</b>
	M_S	0.50 (0.09)	0.38 (0.17)	0.10 (0.10)	0.08 (0.09)	0.22 (0.14)	0.28 (0.17)	<b>0.02 (0.04)</b>	0.14 (0.11)
	M_M	0.48 (0.10)	0.15 (0.11)	0.35 (0.19)	<b>0.03 (0.04)</b>	0.24 (0.16)	<b>0.07 (0.08)</b>	0.23 (0.17)	0.15 (0.14)
	M_L	0.33 (0.19)	0.13 (0.13)	<b>0.06 (0.08)</b>	0.16 (0.19)	<b>0.08 (0.10)</b>	0.16 (0.17)	0.11 (0.12)	0.11 (0.13)
	M_SM	0.49 (0.10)	0.38 (0.17)	0.36 (0.19)	<b>0.00 (0.02)</b>	0.42 (0.22)	0.23 (0.14)	0.18 (0.14)	0.28 (0.19)
	M_SL	0.35 (0.17)	0.25 (0.19)	<b>0.05 (0.08)</b>	0.16 (0.17)	0.16 (0.13)	0.26 (0.19)	0.10 (0.12)	0.18 (0.13)
	M_ML	0.33 (0.16)	<b>0.09 (0.10)</b>	0.20 (0.19)	0.14 (0.16)	0.16 (0.14)	0.13 (0.14)	0.23 (0.19)	0.18 (0.14)
	M_SML	0.34 (0.17)	0.23 (0.21)	0.20 (0.20)	0.14 (0.20)	0.27 (0.20)	0.26 (0.19)	0.22 (0.18)	0.28 (0.17)

<https://doi.org/10.1371/journal.pone.0285463.t002>

small-scale and large-scale covariate effects (S and L respectively). Similarly, the spatial effect of model M\_0, which is a purely spatial model (no covariates included), mirrors the combination of all three covariate surfaces (S, M and L). In the particular cases where we included two covariates (i.e., only one unaccounted covariate), spatial effects resembled the missing covariate.

## Discussion

Many studies have analysed the characteristics of spatial effects to describe the unaccounted ecological mechanisms that drive the distribution of species and try to associate spatial effect patterns to single unaccounted drivers such as home range or dispersion [4, 14, 23]. However, most species distributions are driven by a large number of factors and we are seldom able to identify most of these drivers in our statistical models. As a consequence, SDM spatial effects constitute a combination of many unaccounted factors [6–8].

This study used a simulation study to illustrate the difficulty in interpreting spatial effects with regards to unaccounted environmental drivers. Here, we did not attempt to account for all possible cases, instead, we aimed to illustrate our point using a simple and intuitive approach. Fitted spatial effects resembled the sum of the unaccounted covariate surfaces, including spatial patterns originated by covariates that operated at very different scales. Therefore the biological interpretation of spatial effects may only be valid when the unexplained spatial heterogeneity of the data is characterised by a single dominant driver. At this point, the question is: how many times do SDMs account for all but one driver? One can only speculate this answer but our guess would be: hardly ever. The environmental and ecological processes that drive the distribution of species are complex and diverse, and one could only arbitrarily assume that there is only one covariate missing in our SDM predictor to make biological interpretations over fitted spatial effects.

In this regard, one could use a multiresolution decomposition approach to identify dominant features within the residual spatial correlation of the data [16, 17]. This method essentially estimates the range of spatial correlation at different resolutions of the data, or in this case, residuals of the SDM to help us identify the scale-dependent features within the spatial effect of the residuals. Then, assuming that each scale is characterised by a single dominant driver [13], one could relate them to underlying process generating mechanisms.

## Conclusions

Spatial autocorrelation is a common feature in ecological data. As a consequence, spatial correlation models are important to correctly estimate covariate standard errors and therefore reduce Type I errors. Additionally, spatial correlation terms estimate the residual spatial structure of the data, improving the predictive capacity of our models at locations that are within the sampled area. In ecology, residual spatial patterns are potentially driven by complex multivariate and multi-scaled systems, which can be accommodated by a single spatial effect. Therefore, the biological interpretation of spatial effects is very difficult. A multiresolution decomposition of residual spatial patterns [17] could help us identify the scale-dependent features within the spatial correlation structure of the residuals assuming that each scale is characterised by a single dominant driver.

## Supporting information

**S1 File.**  
(ZIP)

## Author Contributions

**Conceptualization:** Iosu Paradinas, Janine Illian, Sophie Smout.

**Formal analysis:** Iosu Paradinas.

**Funding acquisition:** Iosu Paradinas, Janine Illian, Sophie Smout.

**Investigation:** Iosu Paradinas.

**Methodology:** Iosu Paradinas, Janine Illian, Sophie Smout.

**Writing – original draft:** Iosu Paradinas.

**Writing – review & editing:** Sophie Smout.

## References

1. Zurell D, Franklin J, König C, Bouchet PJ, Dormann CF, Elith J, et al. A standard protocol for reporting species distribution models. *Ecography*. 2020;. <https://doi.org/10.1111/ecog.04960>
2. Elith J, Leathwick JR. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*. 2009; 40:677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
3. Redding DW, Lucas TC, Blackburn TM, Jones KE. Evaluating Bayesian spatial methods for modelling species distributions with clumped and restricted occurrence data. *PloS one*. 2017; 12(11):e0187602. <https://doi.org/10.1371/journal.pone.0187602> PMID: 29190296
4. Lennon JJ. Red-shifts and red herrings in geographical ecology. *Ecography*. 2000; 23(1):101–113. <https://doi.org/10.1111/j.1600-0587.2000.tb00265.x>
5. Legendre P, Dale MR, Fortin MJ, Gurevitch J, Hohn M, Myers D. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*. 2002; 25(5):601–615. <https://doi.org/10.1034/j.1600-0587.2002.250508.x>

6. Keitt TH, Bjrnstad ON, Dixon PM, Citron-Pousty S. Accounting for spatial pattern when modeling organism-environment interactions. *Ecography*. 2002; 25(5):616–625. <https://doi.org/10.1034/j.1600-0587.2002.250509.x>
7. Dormann CF. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global ecology and biogeography*. 2007; 16(2):129–138. <https://doi.org/10.1111/j.1466-8238.2006.00279.x>
8. De Knecht H, van Langevelde Fv, Coughenour M, Skidmore A, De Boer W, Heitkonig I, et al. Spatial autocorrelation and the scaling of species-environment relationships. *Ecology*. 2010; 91(8):2455–2465. <https://doi.org/10.1890/09-1359.1> PMID: 20836467
9. Legendre P. Spatial autocorrelation: trouble or new paradigm? *Ecology*. 1993; 74(6):1659–1673.
10. Dungan JL, Perry J, Dale M, Legendre P, Citron-Pousty S, Fortin MJ, et al. A balanced view of scale in spatial statistical analysis. *Ecography*. 2002; 25(5):626–640. <https://doi.org/10.1034/j.1600-0587.2002.250510.x>
11. Ungaro F, Zasada I, Piorr A. Mapping landscape services, spatial synergies and trade-offs. A case study using variogram models and geostatistical simulations in an agrarian landscape in North-East Germany. *Ecological indicators*. 2014; 46:367–378. <https://doi.org/10.1016/j.ecolind.2014.06.039>
12. Borcard D, Legendre P. Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environmental and Ecological statistics*. 1994; 1(1):37–61. <https://doi.org/10.1007/BF00714196>
13. Perry J, Liebhold A, Rosenberg M, Dungan J, Miriti M, Jakomulka A, et al. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. *Ecography*. 2002; 25(5):578–600. <https://doi.org/10.1034/j.1600-0587.2002.250507.x>
14. Diniz-Filho JAF, Bini LM, Hawkins BA. Spatial autocorrelation and red herrings in geographical ecology. *Global ecology and Biogeography*. 2003; 12(1):53–64. <https://doi.org/10.1046/j.1466-822X.2003.00322.x>
15. Legendre P, Mi X, Ren H, Ma K, Yu M, Sun IF, et al. Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology*. 2009; 90(3):663–674. <https://doi.org/10.1890/07-1880.1> PMID: 19341137
16. Pasanen L, Aakala T, Holmstrom L. A scale space approach for estimating the characteristic feature sizes in hierarchical signals. *Stat*. 2018; 7(1):e195. <https://doi.org/10.1002/sta4.195>
17. Flury R, Gerber F, Schmid B, Furrer R. Identification of dominant features in spatial data. *Spatial Statistics*. 2021; 41:100483. <https://doi.org/10.1016/j.spasta.2020.100483>
18. Johnson O, Fronterre C, Diggle PJ, Amoah B, Giorgi E. MBGapp: A Shiny application for teaching model-based geostatistics to population health scientists. *PloS one*. 2021; 16(12):e0262145. <https://doi.org/10.1371/journal.pone.0262145> PMID: 34972193
19. Lindgren F, Rue H, et al. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*. 2015; 63(19):1–25. <https://doi.org/10.18637/jss.v063.i19>
20. Augustin NH, Trenkel VM, Wood SN, Lorange P. Space-time modelling of blue ling for fisheries stock management. *Environmetrics*. 2013; 24(2):109–119. <https://doi.org/10.1002/env.2196>
21. Wood SN. *Generalized additive models: an introduction with R*. CRC press; 2017.
22. Jones EL, Rendell L, Pirotta E, Long JA. Novel application of a quantitative spatial comparison tool to species distribution data. *Ecological Indicators*. 2016; 70:67–76. <https://doi.org/10.1016/j.ecolind.2016.05.051>
23. Rossi RE, Mulla DJ, Journel AG, Franz EH. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological monographs*. 1992; 62(2):277–314. <https://doi.org/10.2307/2937096>