MedMax Publishers

ISSN: 2769-2310

International Journal of Environmental & Earth Sciences

Review Article

Understanding Spatial Effects in Species Distribution Models

Iosu Paradinas[1,2], Janine Illian[3] and Sophie Smout[1]

[1]Scottish Ocean's Institute. University of St Andrews. East sands, St Andrews, UK.

[2]AZTI, Txatxarramendi Ugartea z/g, 48395, Sukarrieta, Bizkaia, Spain

[3]School of Mathematics and Statistics, University of Glasgow, Glasgow, G12 8QQ, UK

***Corresponding author**
Iosu Paradinas, Scottish Ocean's Institute. University of St Andrews. East sands, St Andrews, UK. E-mail: paradinas.iosu@gmail.com.

Received: 04 February 2022; **Published**: 18 February 2022

Abstract

Most Species Distribution Models include spatial effects to improve prediction and reduce Type I errors. Ecologists tend to try 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 wants 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.

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 and abundance with biotic and abiotic factors in order to gain ecological and evolutionary understanding [2]. This way SDMs allow us to predict distributions across landscapes and make future predictions based on identified drivers, as well as other latent variables such as spatial or spatiotemporal correlation effects. Generally, a SDM is composed by three types of predictors: non-spatial covariates; spatially structured covariates; and spatial or spatiotemporal autocorrelation effects that accommodate the spatial or spatiotemporal autocorrelation 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 [3,4]. In species distribution, spatial autocorrelation may arise as a combination of different factors such as: a geographical range dispersion process, e.g. colonization; unaccounted environmental or biotic drivers; and other highly dynamic processes such as wind and current [5-7]. These drivers can influence species distribution at all scales, from micrometers to continental and ocean-wide scales [8]. However, the size, spacing and extent of sampling units will constrain the scale of inferable drivers, and the scale of spatial autocorrelation [7,9]. In other words, if we sample at a kilometer scale, we cannot infer processes at a smaller scale, and 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, hot-spot size and unaccounted environmental drivers, providing reasonable arguments [5, 10,11]. 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, many modellers fail to acknowledge this probably due to the lack of an explicit study that shows this [7,12-16].

The aim of this article was to provide a practical demonstration that spatial effects are able to smooth the effect of multiple unaccounted drivers, making the biological interpretation of spatial effects rather complicated. To do so, we used model-based 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.

Simulation

We used an iterative simulation approach to produce spatially aggregated distributions (link to code in Annex A). 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), plus a spatial aggregation process driven by the abundance of the neighboring areas, mimicking the colonization of a plant species for example. As a result, our simulated species distributions were driven by the sum of four different effects (Figure 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.

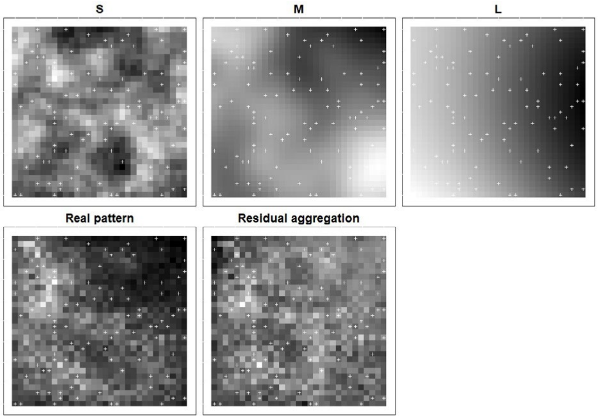


Figure 1: Visualization of the different autocorrelated drivers that influence the abundance pattern in a simulated scenario. S, M and L refer to the small, medium and large scaled covariate fields, respectively. Residual aggregation refers to the geographical range dispersion. White crosses refer to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

We simulated fifty different scenarios, selected 100 random samples for each scenario and fitted all the possible combinations of model-based 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) and 2D smoothing splines through the MGCV package for R [1719].

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

Model	Linear predictor	Missing covariates
M 0	0 + W	S, M&L
M S	0+S+W	M & L
M M	0+M+W	S & L
M L	0+L+W	S & M
M ML	0+M +L+W	S
M SM	0+S+M+W	L
M SL	0+S+L+W	M
M SML	0+S+M +L+W	{

Our aim was to assess the resemblance between fitted spatial effects and un- accounted covariate surface combinations. Resemblance was assessed through the similarity in pattern score (SIP) [20]. 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 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 Annex 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.

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 aggregation of the small-scale and large-scale covariates (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. 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.

Combination of drivers

	Model	Residual	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)	0.01 (0.02)
	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)	0.01 (0.03)	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)	0.04 (0.08)	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)	0.04 (0.04)	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)	0.07 (0.13)	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)	0.08 (0.11)	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)	0.06 (0.10)	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	0.09 (0.15	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)	0.04 (0.05)	0.04 (0.05)
	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)	0.02 (0.04)	0.14 (0.11)
	M M	0.48 (0.10)	0.15 (0.11)	0.35 (0.19)	0.03 (0.04)	0.24 (0.16)	0.07 (0.08)	0.23 (0.17)	0.15 (0.14)
	M L	0.33 (0.19)	0.13 (0.13)	0.06 (0.08)	0.16 (0.19)	0.08 (0.10)	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)	0.00 (0.02)	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)	0.05 (0.08)	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)	0.09 (0.10)	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)

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).

Discussion

Many studies have analyzed 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. 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 [5-7].

This study used a simulation study to illustrate the difficulty in interpreting spatial effects with regards to unaccounted environmental drivers. Readers must realize that did not attempt an exhaustive account of all possible cases, in- stead, we aimed at illustrating 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 characterized by a single dominant driver. However, 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 [15,21]. 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 characterized by a single dominant driver, one could relate them to underlying process generating mechanisms [12].

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 cor- relation terms estimate the residual spatial structure of the data, improving the predictive capacity of our models at locations that are within range. 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 could help us identify the scale-dependent features within the spatial correlation structure of the residuals assuming that each scale is characterized by a single dominant driver [21].

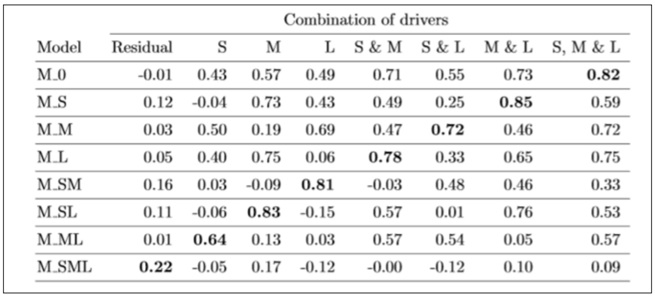
A Annex: script and further explanations

The R script that we used to do all the analysis is available at: https://github. com/iparperspective/Understanding-spatial-effects-/blob/main/Simulation_ script%20understanding%20 spatial%20effects.R

The aim of this annex is to explain the procedure that we followed to create Table 2. To do so we use a single simulated species distribution (as compared to 50 simulations in the study) that is also driven by three spatially structured environmental covariates acting at different spatial scales and a geographical range dispersion process.

We fitted all the models described in Table 1 and we computed SIP scores between each model's spatial effect and all the possible different combinations of covariate surfaces. By doing so, we get Table 3:

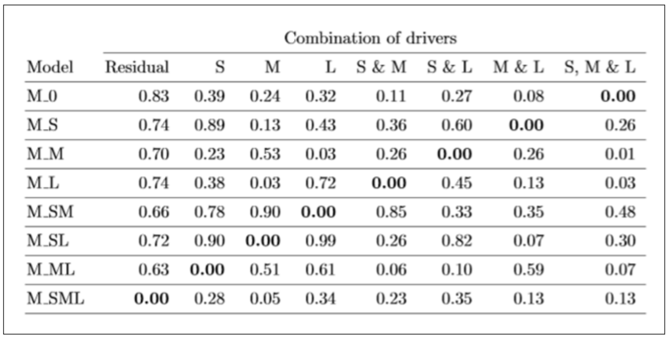
See Annex for a more detailed explanation of the procedure that we followed.



	Combination of drivers							
Model	Residual	S	M	L	S & M	S & L	M & L	S, M & L
M_0	-0.01	0.43	0.57	0.49	0.71	0.55	0.73	**0.82**
M_S	0.12	-0.04	0.73	0.43	0.49	0.25	**0.85**	0.59
M_M	0.03	0.50	0.19	0.69	0.47	**0.72**	0.46	0.72
M_L	0.05	0.40	0.75	0.06	**0.78**	0.33	0.65	0.75
M_SM	0.16	0.03	-0.09	**0.81**	-0.03	0.48	0.46	0.33
M_SL	0.11	-0.06	**0.83**	-0.15	0.57	0.01	0.76	0.53
M_ML	0.01	**0.64**	0.13	0.03	0.57	0.54	0.05	0.57
M_SML	**0.22**	-0.05	0.17	-0.12	-0.00	-0.12	0.10	0.09

Table 3: SIP scores between fitted spatial effects and all the combinations of covariate surfaces. Scores must be read by row. Values closer to one reflect bigger resemblance between spatial fields.

By replicating the simulation 50 times we would get 50 SIP scores for each position, which could be summarised by the mean and standard deviation of these 50 values. However, we decided to use the difference between the best SIP score for each model and combinations of covariate fields, i.e. differences by row in the Table 3, because results were clearer. This way Table 3 becomes Table 4: where zero values represents the best SIP score per model (by row) and the rest of the scores represent the SIP score difference with respect to the best score by row.



	Combination of drivers							
Model	Residual	S	M	L	S & M	S & L	M & L	S, M & L
M_0	0.83	0.39	0.24	0.32	0.11	0.27	0.08	**0.00**
M_S	0.74	0.89	0.13	0.43	0.36	0.60	**0.00**	0.26
M_M	0.70	0.23	0.53	0.03	0.26	**0.00**	0.26	0.01
M_L	0.74	0.38	0.03	0.72	**0.00**	0.45	0.13	0.03
M_SM	0.66	0.78	0.90	**0.00**	0.85	0.33	0.35	0.48
M_SL	0.72	0.90	**0.00**	0.99	0.26	0.82	0.07	0.30
M_ML	0.63	**0.00**	0.51	0.61	0.06	0.10	0.59	0.07
M_SML	**0.00**	0.28	0.05	0.34	0.23	0.35	0.13	0.13

Table 4: The difference in score between the best SIP score and the rest for each model (by row). Values closer to zero reflect bigger resemblance between spatial fields.

where zero values represent the best SIP score per model (by row) and the rest of the scores represent the SIP score difference with respect to the best score by row.

References

1. Damaris Zurell, Janet Franklin, Christian Konig, Phil J Bouchet, Carsten F Dormann, et al. (2020) A standard protocol for reporting species distribution models. Ecography 43: 1261-1277.
2. Jane Elith and John R Leathwick (2009) Species distribution models: ecological ex- planation and prediction across space and time. Annual review of ecology, evolution, and systematics 40: 677-697.
3. Jack J Lennon (2000) Red-shifts and red herrings in geographical ecology. Ecography 23: 101-113.
4. Pierre Legendre, Mark RT Dale, Marie-Jos´ee Fortin, Jessica Gurevitch, Michael Hohn, et al. (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25: 601-615.
5. Timothy H Keitt, Ottar N Bjørnstad, Philip M Dixon, and Steve Citron-Pousty (2002) Accounting for spatial pattern when modeling organism-environment interac- tions. Ecography 25: 616-625.
6. Carsten F Dormann (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global ecology and biogeography 16: 129-138.
7. HJ De Knegt, F van van Langevelde, MB Coughenour, AK Skidmore, WF De Boer, et al. (2010) Spatial autocorrelation and the scaling of species–environment relationships. Ecology 91: 2455-2465.
8. Pierre Legendre (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74: 1659-1673.
9. Jennifer L Dungan, JN Perry, MRT Dale, Pousty Legendre, S Citron-Pousty, et al. (2002) A balanced view of scale in spatial statistical analysis. Ecography 25: 626-640.
10. 10. Fabrizio Ungaro, Ingo Zasada, and Annette Piorr (2014) 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 46: 367-378.
11. 11. Daniel Borcard and Pierre Legendre (1994) Environmental control and spatial structure in ecological communities: an example using oribatid mites (acari, oribatei). Environmental and Ecological statistics 1: 37-61.
12. 12. JN Perry, AM Liebhold, MS Rosenberg, J Dungan, M Miriti, et al. (2002) Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography 25: 578-600.
13. 13. Jos´e Alexandre Felizola Diniz-Filho, Luis Mauricio Bini, and Bradford A Hawkins (2003) Spatial autocorrelation and red herrings in geographical ecology. Global ecology and Biogeography 12: 53-64.
14. 14. Pierre Legendre, Xiangcheng Mi, Haibao Ren, Keping Ma, Mingjian Yu, et al. (2009) Partitioning beta diversity in a subtropical broad- leaved forest of china. Ecology, 90: 663-674.
15. 15. Leena Pasanen, Tuomas Aakala, and Lasse Holmstr¨om (2018) A scale space approach for estimating the characteristic feature sizes in hierarchical signals. Stat 7:e195.
16. 16. Roman Flury, Florian Gerber, Bernhard Schmid, and Reinhard Furrer (2021) Identification of dominant features in spatial data. Spatial Statistics 41:100483.
17. 17. Finn Lindgren, H˚avard Rue, et al. (2015) Bayesian spatial modelling with r-inla. Jour- nal of Statistical Software 63: 1-25.
18. 18. Nicole H Augustin, Verena M Trenkel, Simon N Wood, and Pascal Lorance (2013) Space-time modelling of blue ling for fisheries stock management. Environ- metrics 24:109-119.
19. 19. Simon N Wood (2017) Generalized additive models: an introduction with R. CRC press.
20. 20. Esther L Jones, Luke Rendell, Enrico Pirotta, and Jed A Long (2016) Novel application of a quantitative spatial comparison tool to species distribution data. Ecological Indicators 70: 67-76.
21. 21. Roman Flury, Florian Gerber, Bernhard Schmid, and

Citation: Iosu Paradinas, Janine Illian, Sophie Smout. Understanding Spatial Effects in Species Distribution Models. Int J Envi & Eart Scie 2022, 3: 1-4