

## Discrepancies between point process models and environmental envelopes identify the niche centroid – geography configuration



Gerardo Martín<sup>1,\*</sup>, Carlos Yáñez-Arenas<sup>2</sup>, Xavier Chiappa-Carrara<sup>1,2</sup>

<sup>1</sup> Departamento de Sistemas y Procesos Naturales, Escuela Nacional de Estudios Superiores unidad Mérida, Universidad Nacional Autónoma de México, Ucú, Yucatán 97357, México

<sup>2</sup> Laboratorio de Ecología Geográfica, Unidad de Conservación de la Biodiversidad, UMDI-Sisal, Facultad de Ciencias, Universidad Nacional Autónoma de México, Sierra Papacal, Yucatán 97302, México

### ARTICLE INFO

#### Keywords:

Point process models  
Minimum volume ellipsoids  
Niche centroid  
Multivariate skewness  
Presence only

### ABSTRACT

Contrasting outcomes of statistical modelling methods for the same data may represent critical pieces of information about the biological process. Ecological niche and species distribution modelling are notorious for their keenness on algorithm testing although differences between method outcomes are seldom used to gain biological insight. Here we use the differences and similarities between point process models (PPMs) and minimum volume ellipsoids (MVEs) to help identify the configuration of species' climatic niches in relation to geographic space and species' interactions. Poisson PPMs represent the abundance of points in an euclidean plane as a function of spatially defined covariates, while MVEs are used to estimate the (niche) centroid and represent similarity with its centroid via Mahalanobis distance. The niche centroid is receiving increasing attention due to its simplicity and ability to capture complex demographic processes such as species' abundance. MVEs' simplicity makes them sensitive to deviations from symmetry in the statistical distribution of environmental axes, or the lack of a defined centroid in geographic space. Using synthetic and real species we test the ability of PPMs and MVEs to characterise niche centroids in relation to each other and to the statistical properties of the environment. We also study a scenario in which non-linear responses and biological interactions define a species' fundamental niche and realised distribution respectively. MVEs were less precise but estimated useful centroids more frequently than PPMs. When centroids clearly existed in geographic and environmental spaces, PPMs' estimated centroids that were closer to the truth. MVEs' ability to estimate a similarity surface, unlike PPMs, depends on the correct estimation of the centroid. We suggest then that contrasting similarity surfaces estimated by both methods indicate the absence of the centroid in geographic space.

### Introduction

Ecological niche modelling (ENM) aims to estimate the climatic conditions under which species perform best (Peterson and Soberón, 2012). In the strictest sense, such aim is mostly attainable with tightly controlled laboratory experiments, which can be prohibitively expensive or unethical (Angilletta, 2009; Qiao et al., 2016). Given these limitations, the statistical analysis of the climatic conditions experienced by free ranging organisms were proposed as a suitable alternative, and are collectively known as correlative niche modelling (Soberón and Nakamura, 2009). Such an approach however has several limitations ranging from observation bias, migration, mismatch between the area of occurrence and physiologically tolerated climatic conditions (Soberón

and Peterson, 2005; Yáñez-Arenas et al., 2020), to name a few. Much has been learned since such analyses began to be implemented, and after much testing (e. g. Elith et al., 2006) now there are a plethora of analytical methods that partially address some of the formerly raised issues. Ecological niche models that predict geographical distributions on the basis of tolerated climatic conditions are now routinely used in ecology, conservation biology and public health risk mapping (Papes et al., 2012; Saupe et al., 2011; Tingley et al., 2009). Their use for predicting abundances or the impacts of climate change, however, remain controversial (Osorio-Olvera et al., 2020; Dallas et al., 2017).

Predicting species abundances via ENMs is inspired by Hutchinson (1957)'s conception of niche as a multidimensional space in whose centre occur the most favourable conditions for the species (the niche

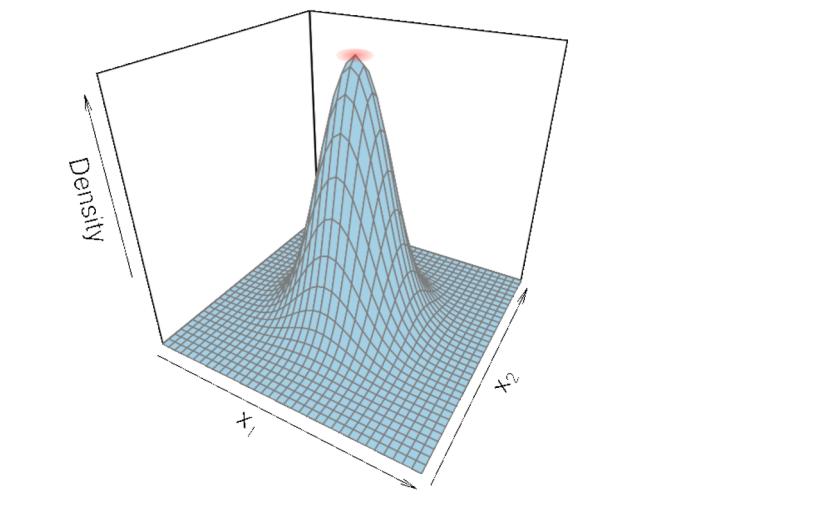
\* Corresponding author

E-mail address: [gerardo.mmc@enesmerida.unam.mx](mailto:gerardo.mmc@enesmerida.unam.mx) (G. Martín).

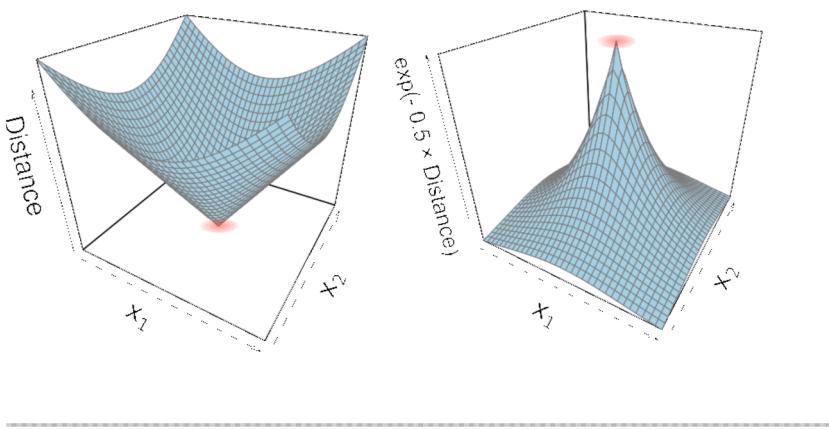
centroid). The centroid of the multidimensional environmental space represents the combination, for instance, of temperature, rainfall, solar radiation and humidity in which population growth rates reach their biological maximum. Hence, niches began to be modelled with minimum volume ellipsoids (MVEs; Qiao et al., 2016; Osorio-Olvera et al.,

2019) which are mathematically related to Hutchinson's rationale. The centroid then is estimated with the arithmetic mean of environmental conditions in which a species has been observed to be present, and represented in the geography with the Mahalanobis distance (Euclidean  $\times$  covariance matrix) to the centroid. This method assumes that species'

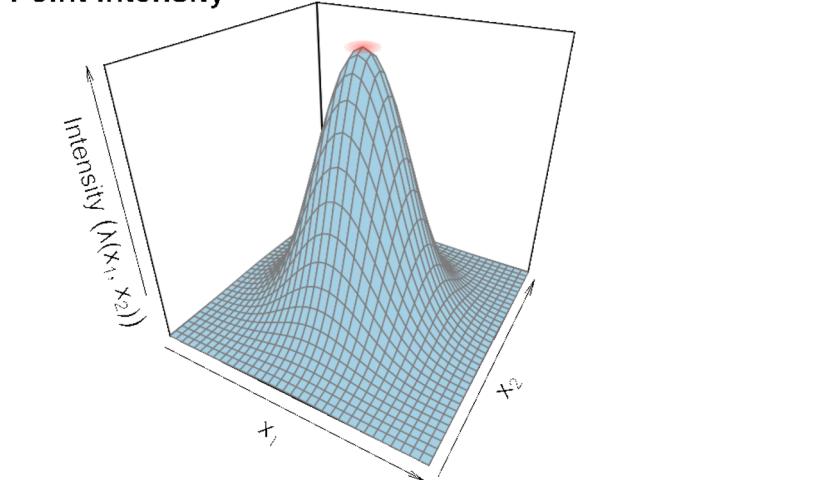
### Statistical density



### Distance to centroid



### Point intensity



**Fig. 1.** Similarities and differences between methods.

performance curves along each environmental dimension are unimodal and symmetric, which makes MVEs susceptible to characterise biased centroids.

In contrast to MVEs, point process models (PPMs) were developed as an ad-hoc tool for the statistical analysis of random variables consisting of geographical locations (Baddeley et al., 2016). Data for correlative ecological niche models are collections of geographical locations which, in most cases, represent only the places where species have been observed to be present. Less often, absences are used in a probabilistic, binomial approach. Therefore, PPMs are among the most suitable statistical tools for that purpose, although their adoption in ecology has been limited. The popular algorithm Maxent (Phillips et al., 2006) is an exception, as it is mathematically equivalent to PPMs (Renner and Warton, 2013), but lacks many of the diagnostic tools, statistical rigour and transparency of PPMs (Renner and Warton, 2013; Baddeley et al., 2016). Moreover, there are statistically sound and flexible approaches to account for data deficiencies such as observation bias, and spatial clustering and exclusion (Warton et al., 2013; Isaac et al., 2020), which exceed Maxent's built-in capabilities. In practice, PPMs work in the same way as generalised linear models by maximising a likelihood function. The response variable in PPMs is point intensity, representing the expected number of geographical points per spatial unit, hence the model fitted is a log-linear function of geographically referenced climatic covariates with a Poisson likelihood (Baddeley et al., 2016; Baddeley and Turner, 2005).

In regression analysis, the possible number of model formulas can be very large. However, in the context of the niche centrality hypothesis, models only need to be formulated to be able to identify relationships characteristic of performance curves for a minimal set of climatic axes. For most organisms, performance curves are smooth, concave and unimodal (Angilletta, 2009), which can be broadly reproduced by log-linear second-degree polynomials. The relationship between the niche centroid, statistical density, distance, performance curves and second-degree polynomials in this context is closer than it seems at first glance. Statistical density represents the frequency of observations, the basis for estimating centroids, while distance is a measure of similarity with the centroid defined by the coordinates in each axis, and point intensity in a second-degree polynomial with a maximum in each axis represents point intensity at all possible combinations of the independent variables (Fig. 1).

In this study, we characterise the similarities and differences between PPMs and MVEs as both methods represent important aspects of the data traditionally used in species distribution and ecological niche modelling. PPMs are the ad-hoc method for occurrence-based analyses, while MVEs explicitly represent the very definition of the biological phenomenon. We use the differences between methods to develop a series of recommendations for the model-building process, and show some of the extensions that PPMs incorporate in order to account for biological interactions.

## Methods

### Overview

To have complete control and knowledge of causality, our study has a section that is entirely synthetic, in which we simulated environmental conditions and occurrence records in squared grids of  $100 \times 100$  pixels. Then we analyse of two contrasting real world bird species. For the synthetic analyses, environmental covariates were 100 layers, 40 with a normal, 20 with a log-normal, 20 with beta and 20 with gamma distributions. The objective was to create variables with statistical properties that range from symmetric, left, or right skewed to bimodal, all of which occur among temperature and precipitation (Pedersen et al., 2010; Yaya and Vo, 2020). To simulate species occurrence records we created 2500 unique combinations of three environmental covariates. With each of the 2500 variable combinations we simulated four species, one in each

of the following scenarios based on the centroid's location in relation to environmental and geographic space (Fig. 2):

- 1 **Centre:** Species' centroid located at the highest density interval (mode) in each environmental dimension
- 2 **Random:** Centroid located randomly across the possible range of values of each environmental dimension
- 3 **Edge:** Centroid located at the top or bottom 5% value of the possible range of values of each environmental dimension
- 4 **Outer:** Centroid located below or above the possible range of values of each environmental dimension

After simulating environmental conditions and point processes we tried to retrieve the centroids used for simulating occurrence data sets. Then, we performed two similar analyses with increasing levels of realism: 1) simulating occurrences for two species using a set of four bioclimatic variables from the Chelsa dataset (Karger et al., 2017), and 2) analysing occurrence datasets of two North American bird species with well-known distributional ecology, *Callipepla californica* (with a truncated niche), and *Calamospiza melanocorys* (with an ecological niche defined in geographic space; Osorio-Olvera et al., 2020).

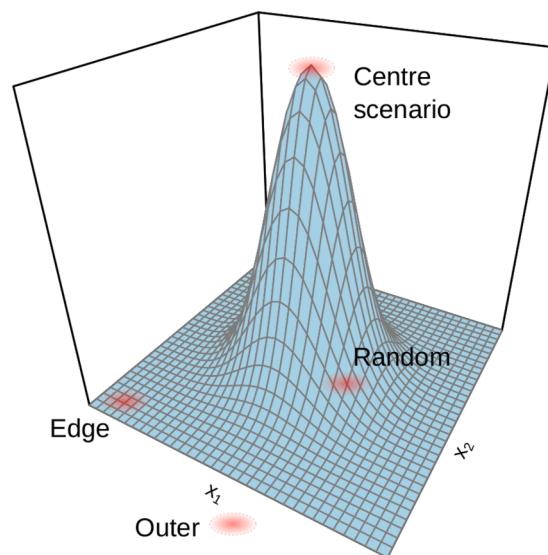
For the two simulated species with the Chelsa dataset, centroids along each climatic axis were generated using non-monotonic functions with a tipping point and then additively combined them in one single layer to use as average point intensity to simulate occurrences. Each simulated layer represents a species' fundamental niche, and we used them to simulate occurrences over a range of competition scenarios using pre-specified threshold distances within which species exclude each other. This method is called hardcore multi-type point interaction (Baddeley et al., 2016). We used R 4.0.2 for all the above analyses, full details below.

### Simulation of environmental covariates

All the covariates were simulated with a multivariate normal distribution to create a spatial structure dependent on the distance between pixel pairs. The degree of correlation between pixels was proportional to an exponential function of the distance between them:

$$\text{cov}(x_1, x_2) = \sigma e^{-\phi \cdot |s|}$$

$$|s| = \text{euclidean}(x_1, x_2)$$



**Fig. 2.** Scenarios simulated in environmental space.  $x_1$  and  $x_2$  axes show the values that occur in the geographic space of analysis and how centroids were located within or outside variable ranges.

The means of the normal distributions to simulate variables were chosen arbitrarily with values that resemble the averages of variables routinely used for modelling species' distributions (minimum and maximum temperatures and rainfall levels; Fick and Hijmans, 2017). To simulate spatially-structured variables with log-normal distributions, we simply exponentiated variables simulated with the multivariate normal distribution.

Variables with the gamma and beta distributions were simulated using log-normally distributed variables for each distribution's  $\alpha$  and  $\beta$  parameters:

- Gamma:

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

- Beta:

$$p(x) = \frac{\Gamma(\beta + \alpha)}{\Gamma(\beta)\Gamma(\alpha)} x^{\alpha-1} (1-x)^{\beta-1}$$

### Simulation of occurrence records

#### Species with simulated environmental covariates

To simulate points in the spatial domain of the simulated covariates we used the 2500 unique combinations described above. For each layer in the combination, we selected the values for the scenarios of the centroid's location in relation to environmental and geographical space. Then we projected the Mahalanobis distance to the selected centroids to create a layer that fully represents habitat suitability. The resulting layers were exponentially transformed:

$$p(DNC) = e^{-r \cdot DNC}$$

$$r = 10 / (\max(DNC) - \min(DNC))$$

in order to obtain values readable as probabilities by the randomPoints function of the dismo package used to simulate the collections of points. We acknowledge that the spatstat (Baddeley and Turner, 2005) functions for the same purpose are more methodologically sound than dismo's. However, randomPoints allows specification of the sample size and produces point patterns similar to a Poisson, while rpoispp determines sample size assuming that the probability layer is point intensity.

#### Interacting species based on historic climate

For these analyses we created two surfaces that represent each species' fundamental niche using four of Chelsa's bioclimatic variables: mean annual temperature (bio1), mean diurnal range (bio2), annual precipitation (bio12) and precipitation of driest month. For each variable, we created a response surface using bell-shaped (temperature optimum – Logan, 1989; double exponential) or asymptotic functions (negative exponential; formulas and parameters are given in supplementary materials). After generating two response surfaces for each variable, one for each species, we combined the responses linearly by drawing parameter values from a uniform distribution bounded by 0 and 2 and created a climatic suitability surface used as average point intensity to simulate occurrence datasets.

To simulate the datasets of interacting species' we used the ragsMultiHard function of spatstat using the 25 different possible combinations of radii from 0 to 0.5 degrees in 0.125 increments for each species. In the Hardcore model of inter-point interactions, radii refer to the distance in which one type of point excludes the occurrence of another, hence a radius of 0 is a simple Poisson point process (Baddeley

et al., 2016).

#### Analysis of the simulated point patterns

Analyses of both kinds of simulated species comprised: 1) retrieving the centroids used for generating the probability surface to simulate species using minimum volume ellipsoids (MVEs) as implemented in the ntbox R package (Osorio-Olvera et al., 2020), and with Poisson point process models (PPMs) using the spatstat package; 2) retrieving the surface used for simulating species using the same methods to retrieve centroids; 3) measuring the ability of both methods to retrieve the centroids via the Mahalanobis distance to the true centroid; 4) measuring the ability of both methods to retrieve the probability surface by measuring the spatial correlation between estimated distance to the centroid (for MVEs) and point intensity (for PPMs) with the probability surfaces; 5) measuring the influence of the statistical properties of environmental space (multivariate skewness) on the ability of both methods to retrieve the above quantities.

Retrieving centroids with MVEs is straightforward as it is part of the process, however, with PPMs it is only possible when the model formula can have maxima. Hence, in order to retrieve centroids we fitted all PPMs using a second-degree polynomial formula:

$$\log(\lambda) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta'_1 x_1^2 + \beta'_2 x_2^2 + \beta'_3 x_3^2$$

where  $\lambda$  is point intensity per grid cell,  $\beta$  are the regression coefficients, and  $x_i$  are the environmental covariates. Maxima, or centroids  $x_i$ , exist only if  $\beta'_i < 0$ , which we found differentiating  $d\log(\lambda)/dx_i$ , and then solving for  $x_i$ :

$$x'_i = \frac{-\beta_i}{2\beta'_i}$$

For the interacting species' scenarios, we tried to recover centroids even though it was not defined in one environmental dimension (rainfall) for one of the species. The interacting species' data were analysed as a marked Hardcore PPM with an interaction radius to be estimated, as a marked Poisson PPM (Baddeley et al., 2016) and with MVEs.

#### Analysis of retrieved centroids and generating surfaces

To compare the retrieved centroids with those used to simulate data, we measured the Mahalanobis distance between them. To compare the generated surfaces with those used for simulating data, we ran Pearson correlation tests. In order to characterise the causes of the differences obtained between modelling methods, we measured the multivariate skewness of each one of the 2500 sets of covariates using the mvn function of the MVN package with the Henze-Zirkler method (Henze and Zirkler, 1990).

#### Analysis of real-world species

Both analysed bird species are part of the North American Breeding Bird Survey, as it is the most comprehensive vertebrate spatially-referenced abundance dataset available. We chose *Callipepla californica* and *Calamospiza melanocorys* because these two species have contrasting distributional ecologies. The former has a truncated niche, poorly defined in geographical space, while the latter is well defined (Osorio-Olvera et al., 2020). To fit models we only used the occurrence records without abundance, and then tested the abundant niche centroid hypothesis with the records with abundance data.

To fit models we used the Chelsa bioclimatic variables and the first three components of a principal components analysis. To select a set of three climatic variables for each species we plotted point intensity in relation to each variable and identified those with a smooth bell-shaped response, and correlation pairs between variables, allowing no more than  $r = 0.7$ . All layers to fit models were cropped with 5° buffers around

occurrence records to exclude areas where there could be analogous climates.

With the subset of variables with the desired bell-shaped responses we identified sets of three to four variables that could be included in the same model due to low collinearity. Then we generated all the possible combinations of compatible variables using the same 2<sup>nd</sup> degree polynomial formula described above and fitted the PPMs. Once we had a full set of fitted models we excluded all those with convergence or statistical significance estimation problems, and then identified the model with the lowest Akaike Information Criterion (AIC). We then revised if the squared terms had statistically significant negative coefficients, and diagnosed the model with lurking plots of cumulative residuals along the  $x$  and  $y$  axes. Once we obtained a satisfactory PPM, we extracted the centroid and covariance matrix using the same set of climatic variables, and projected the Mahalanobis distance to the centroid.

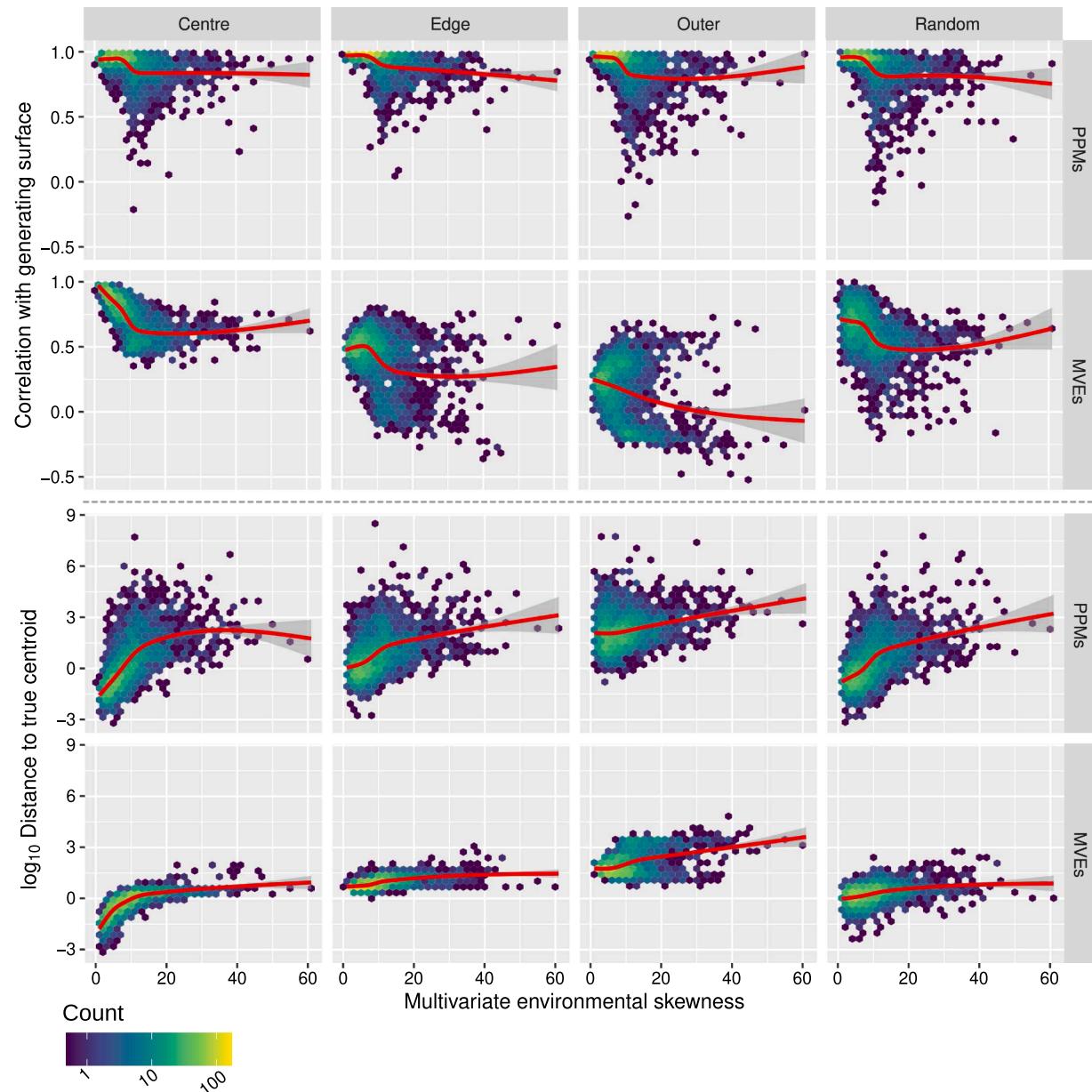
Finally we proceeded as before, comparing the surfaces generated by both methods and the estimated centroids using correlation and

Mahalanobis distance respectively. To test the abundant niche centroid hypothesis we used the abundance data, and measured its correlation with point intensity and distance to the centroid (exponentially transformed, Fig. 1.)

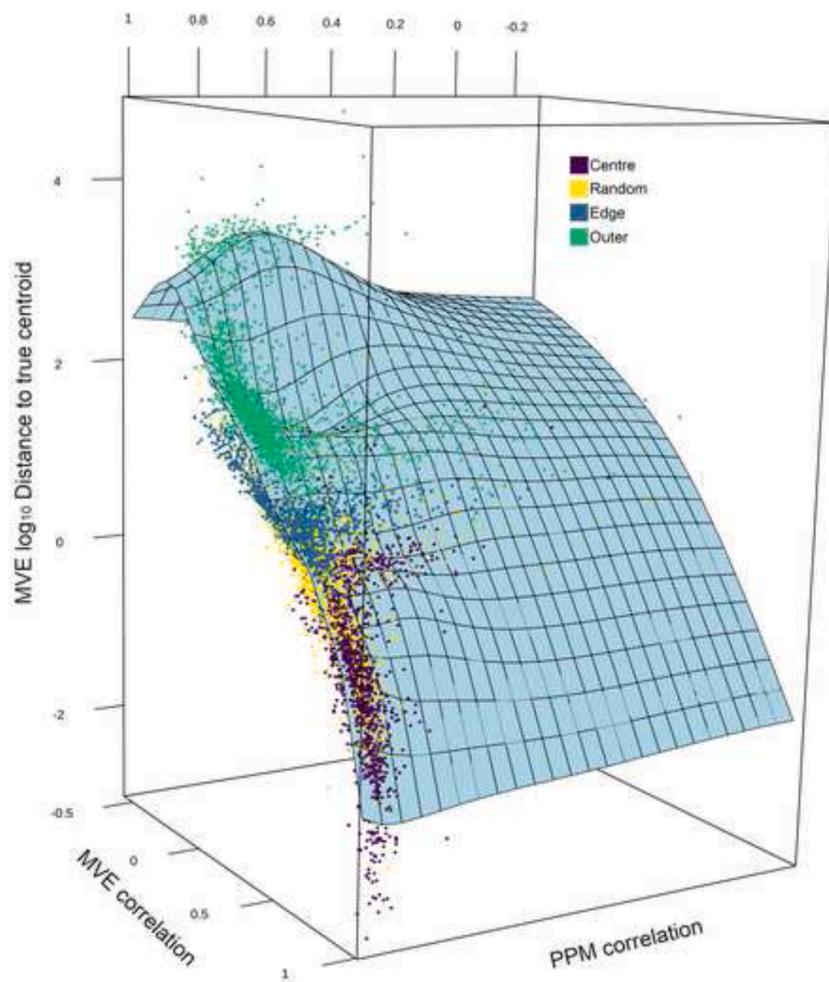
## Results

### Species based on simulated conditions

The scenario of the centroid's location in environmental and geographic space that allowed the best estimation was the centre, followed by random, edge, and outer allowed the least accurate estimates (Fig. 3). Between methods, we observed large differences. With PPMs we estimated both the most and least accurate centroids, because in many cases the optimisation routine did not estimate  $\beta_i < 0$ , which explains the very large distances to the true centroids (Fig. 4). Whereas, regardless of the non-existence of the centroid in the data, MVEs



**Fig. 3.** Correlation between the surface used to simulate each point process and the surface generated with PPMs and MVEs in each scenario as a function of multivariate environmental skewness. Red lines show the smoothed trend in the scatter plot. Environmental Mahalanobis distance to the true centroid in each simulated scenario between PPMs and MVEs as a function of multivariate environmental skewness. Red lines show the smoothed trend in the scatter plot.



**Fig. 4.** Cross comparison of the correlation between the surfaces generated by both methods for the same data sets and the distance to the true centroids obtained with MVEs (z-axis). Colours show the different scenarios for simulating point patterns. Distance to true centroid decreases in MVEs when the generated surface is highly correlated with the generating surface. However, distance to the true centroid does not change significantly when surfaces estimated by PPMs are more correlated with the generating surface (low inclination of the blue surface along its axis). The blue surface was generated with a generalised additive model with smoothing splines for PPM correlation and MVE correlation to represent the trend in the data points.

estimate a quantity that is more similar to the true centroid. The performance of both PPMs and MVEs was significantly affected by multivariate skewness. Greater departure from normality resulted in less accurate estimated centroids (Fig. 3).

Regarding the qualities of the surfaces generated by both methods, PPMs in most cases generated surfaces that were more positively correlated with the surface used to simulate occurrence data than MVEs. Environmental skewness also affected the ability of PPMs and MVEs to retrieve the surface that generated occurrence data.

The correlation with the generated surfaces were less affected by greater distance from the true centroid and vice versa among PPMs than MVEs. In fact, MVE results can be compared with PPMs, in that distance to the true centroid also changes very little with PPMs' correlation with the generating surface (Fig. 4). These results indicate that disparities in estimated centroids and generated surfaces between methods can be used to assess the environmental centroid - geography configuration being analysed. Part of the reason that PPMs' centroid estimations can be less accurate is due to the estimation of positive effects for the squared terms (Fig. S1).

#### Competing species based on climate databases

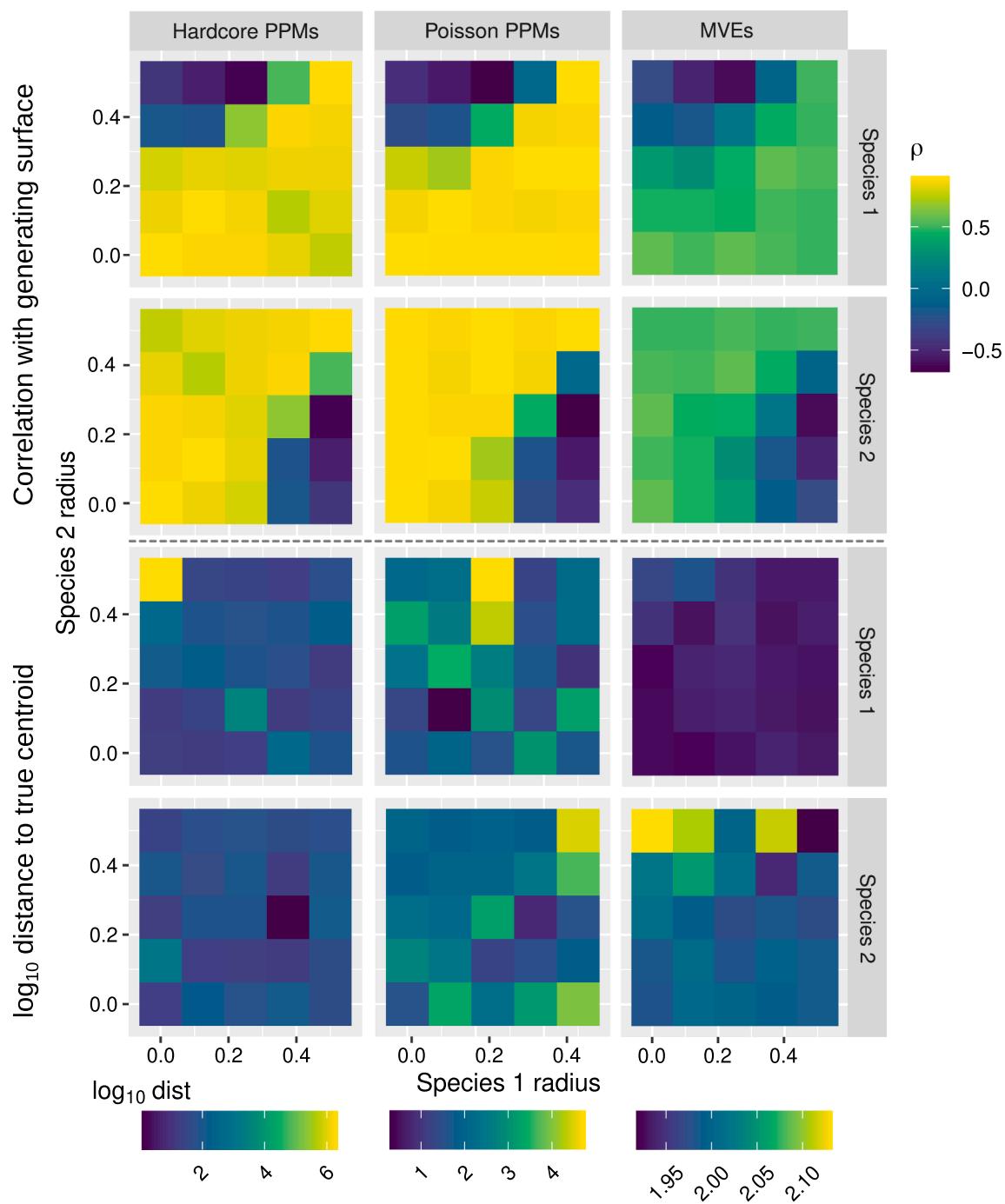
The radii of exclusion affected the ability of the different methods to retrieve true centroids. Centroids estimated with Poisson PPMs were further away in environmental space from true centroids compared to centroids estimated with Hardcore PPMs and MVEs. The method less affected by the size of the exclusion radius were MVEs, and Hardcore PPMs were relatively unaffected (Fig. 5).

The ability to retrieve the generating surfaces was also affected by the size of the exclusion radius (Supplementary materials – Interacting species). The method that produced the surfaces less similar to the original were MVEs, followed by Hardcore PPMs and Poisson. However the latter two only produced surfaces negatively correlated with the generating surface when the exclusion radii of the other species were greater than  $0.4^\circ$ , which was close to the highest values tested (Fig. 5).

#### Analysis of real-world species

We obtained four candidate model formulas for *Callipepla californica*, and 10 for *Calamospiza melanocorys'* PPMs. The fitted effects for the linear ( $\beta$ ) and quadratic terms ( $\beta''$ ) for each variable included in each model and their statistical significance appear in Table 1. The statistics comparing PPMs and MVEs are shown in Table 2. The selected models' diagnosis show cumulative residuals larger than tolerance in some areas, although generally close to zero (Figures S2 and S3 for *Callipepla californica* and *Calamospiza melanocorys* respectively).

The estimated surfaces (Fig. 6) and centroids and their similarities are consistent with the analyses of simulated species. For *Calamospiza melanocorys*, which has a centroid well defined in geographic space, the suitability surfaces generated by both methods were more similar than for *Calamospiza melanocorys*, albeit differences in this respect were more subtle than for simulated species with similar characteristics (Table 2). As previously found, *Calamospiza melanocorys'* abundance decreases significantly with distance from the niche centroid, and *Callipepla californica*'s does not. When comparing the correlation between abundance and point intensity, in both cases the correlation was higher than with



**Fig. 5.** Correlation between surfaces generated by methods and surfaces used to simulate occurrence records (top six panels). Distance between the centroids estimated by each method and true centroids in response to the radii in which species exclude each other (bottom six panels).

distance from the centroid. The centroid estimated with PPMs for *Calipepla californica* suggests that it could be defined outside the geographically occupied area at least on some climatic axes (bio8 for instance, see response plots in Fig. S4, and see Fig. S5 for the same responses for *Calamospiza melanocorys*).

## Discussion

Both PPMs and MVEs are capable of estimating niche centroids, and both are sensitive to the skewness of environmental conditions. MVEs are less sensitive to environmental skewness than PPMs in their estimates of the environmental centroid, although given certain centroid-

geographic space configurations PPMs can provide more precise centroid estimates than MVEs. PPMs are capable of estimating a suitability surface that resembles more the actual suitability surface than MVEs, regardless of the distance between estimated and true centroids. MVEs' performance in this department, however, depends on correct estimation of the multivariate centroid. Therefore, differences in environmental suitabilities produced by both methods likely indicate that the centroid does not exist in geographic space (Yáñez-Arenas et al., 2020). Similar estimates of the centroid and suitability surfaces between methods indicate that the centroid, the optimal climatic conditions, does exist in the geographic space occupied by the species. These insights from simulated species are confirmed with the analysis of real species

**Table 1**

Coefficients and fitted effects for each species' model. Asterisks indicate significance, \* indicates  $P \leq 0.05$ , \*\* is  $P \leq 0.01$  and \*\*\*  $P \ll 0.01$ . Grey background colour highlights estimates for *Calamospiza melanocorys*. For reference, bio5 = Max temperature of warmest month, bio7 = Temperature annual range, bio8 = Mean temperature of the wettest quarter, bio12 = Total (annual) precipitation, bio16 = Precipitation of wettest quarter.

Species	Variable	$\beta$	$\beta^*$
<i>Callipepla californica</i>	bio8	-0.014 **	$-4.5 \times 10^{-05}$
	bio11	0.028 ***	$-1.3 \times 10^{-04} ***$
	bio12	0.001 **	$-8.4 \times 10^{-07} ***$
<i>Calamospiza melanocorys</i>	bio5	0.13 ***	$-2.1 \times 10^{-04} ***$
	bio7	0.38 ***	$-4.4 \times 10^{-04} ***$
	bio16	0.046 ***	$-1.3 \times 10^{-04} ***$

with contrasting distributional ecologies.

The first assumption in modelling species with the methods used is that an optimum exists along each environmental dimension. In addition to the above, we also suggest that estimation of positive coefficients in the second degree polynomial indicate that such optima are weakly expressed among data, although this could be addressed with more complex model formulas, for which there are more sophisticated solutions within PPMs (Baddeley et al., 2016). Specific configurations of Maxent, for instance, are equivalent to our methods (Renner and Warton, 2013), but splines can also accommodate such complex non-linear responses (Renner et al., 2015). However, as shown with the second set of analyses, both MVEs and PPMs with a polynomial formula, can also estimate relatively close centroids in the presence of non-linear responses. Thus, we provide evidence that simple methods are still powerful enough to begin making ecologically sound decisions in the process of developing an ecological niche model.

Even though we analysed a comprehensive number of scenarios, covering the role of both geographic space and biotic interactions, more configurations and method synergies are possible. For instance, supraspecific units have been proposed for increasing the environmental space when the centroid does not exist within species' geographic ranges (Castaño-Quintero et al., 2020), and PPMs could prove useful for estimating it or assessing if it actually exists in the geography or environmental space of analysis. As for areas of application, PPMs may prove an important tool in quantifying the controversial abundant niche centroid hypothesis (Dallas et al., 2017; Osorio-Olvera et al., 2020), or may clarify as we have done, the circumstances in which the hypothesis holds on real species via PPMs' direct relationship with point abundance (Renner and Warton, 2013). Ultimately, improvements to the methods used, will improve the predictive capacity of niche and species distribution models.

In the majority of ecological niche modelling studies models are selected by random or spatial partitioning data into training and testing fractions (Raes and Ter Steege, 2007; Peterson et al., 2008). Currently, we ignore how such model selection methodologies affect the qualities of the suitability surfaces produced by models. It is likely that optimising performance with a partial ROC analysis for instance (Peterson et al., 2008) improves the accuracy of the generated surface. Although recent developments suggest that there are more robust ways of measuring spatial predictive performance (Jiménez and Soberón, 2020). If optimising any performance metric equates improving the estimation of

suitability surface, for MVEs this process implies improving the estimated centroids. Similarly, the customary approach when developing PPMs is optimising for the Akaike Information Criterion without data partitioning (Baddeley et al., 2016), which has not been tested in relation to any of the validation methods frequently used in ecological niche modelling.

Very few analyses in ecological niche modelling attempt to quantify biological interactions (e. g. González-Salazar et al., 2013). Our multi-species analyses show that even when suitable conditions co-occur, competition between species breaks down the positive relationship between abundance and climatic suitability. Such phenomena are expected in trophic webs, but are seldom attributed to failures to prove the abundant niche centroid hypothesis (Dallas et al., 2017; Osorio-Olvera et al., 2020). The biology of the phenomena that breaks down the suitability-abundance relationship may be related to Allee effects (Osorio-Olvera et al., 2019). Such ecological understanding, nevertheless is still difficult to implement in spatial analyses as we still lack straightforward ways to represent those dynamic processes in a simple statistical model. As a result, we found that quantifying interactions did not help to better estimate the centroid or the suitability surface when the effect of one species on another was very strong (very large exclusion radius). In such cases, discrepancies between MVEs and PPMs were less informative of the centroid-geography configuration than they were for single species cases. However, when species interactions were weaker, or bi-directional, there was still some inferential power to assess the niche centroid – geography configuration (Fig. 4).

To our knowledge the statistical properties of environmental dimensions have been widely neglected as a source of error in model development. Our results suggest that environmental skewness can complicate estimation by both methods used. Hence, we recommend choosing environmental covariates with lower skewness, or considering using methods that accommodate non-linearities and allow the user to estimate maxima or minima. As mentioned above, PPMs with splines represent one such alternative to more automated algorithms like Maxent (Phillips et al., 2006), when the number of models to be developed is small, when there are important ecological questions to be answered from distributional data, or when the question is one of forecasting and requires better estimation of the niche centroid (Soberón and Peterson, 2005).

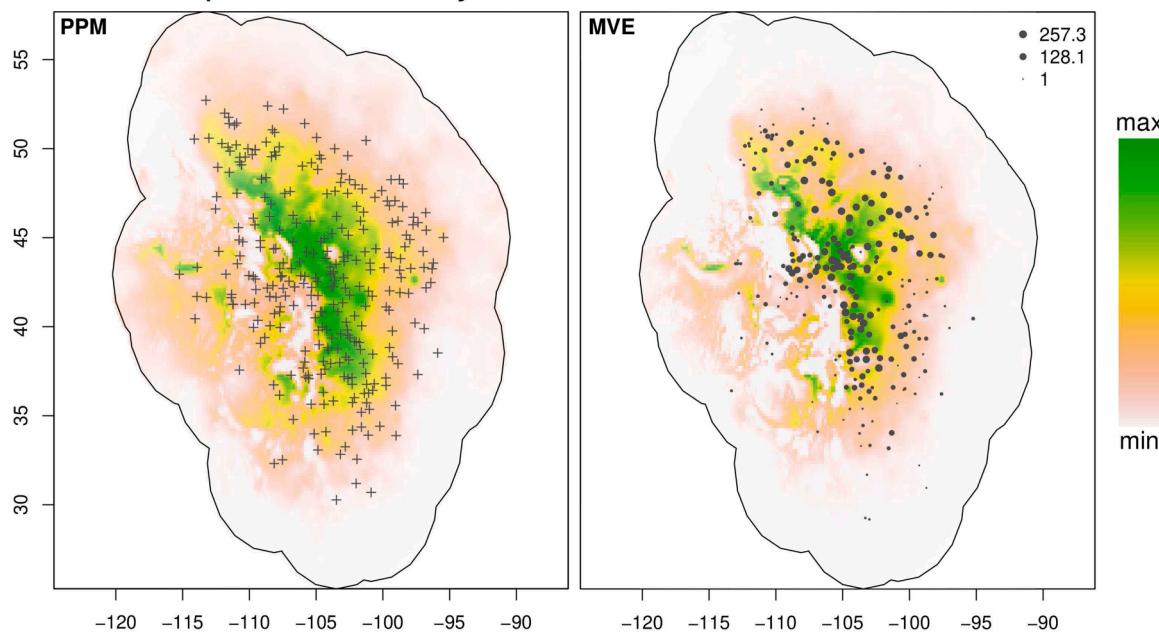
The imputative nature of MVEs represents both its strength (closer estimation of the centroid when it is absent or weakly defined) and its weakness (poor representation of the suitability surface when centroid is absent). However, their performance in relation to PPMs in both departments can be used to an advantage to identify the centroid-geography configuration (Yañez-Arenas et al., 2020), and make decisions on the steps to follow in the model-building process, such as including closely related species (Castaño-Quintero et al., 2020) to increase the environmental domain of the analyses. As we show with our analyses of real world species, the simulation findings are robust and we greatly encourage the synergistic use of modelling methods to improve estimates of optimal conditions and suitability – abundance relationships, to test important ecological questions and to develop more robust, ecologically sound, forecasting methodologies.

**Table 2**

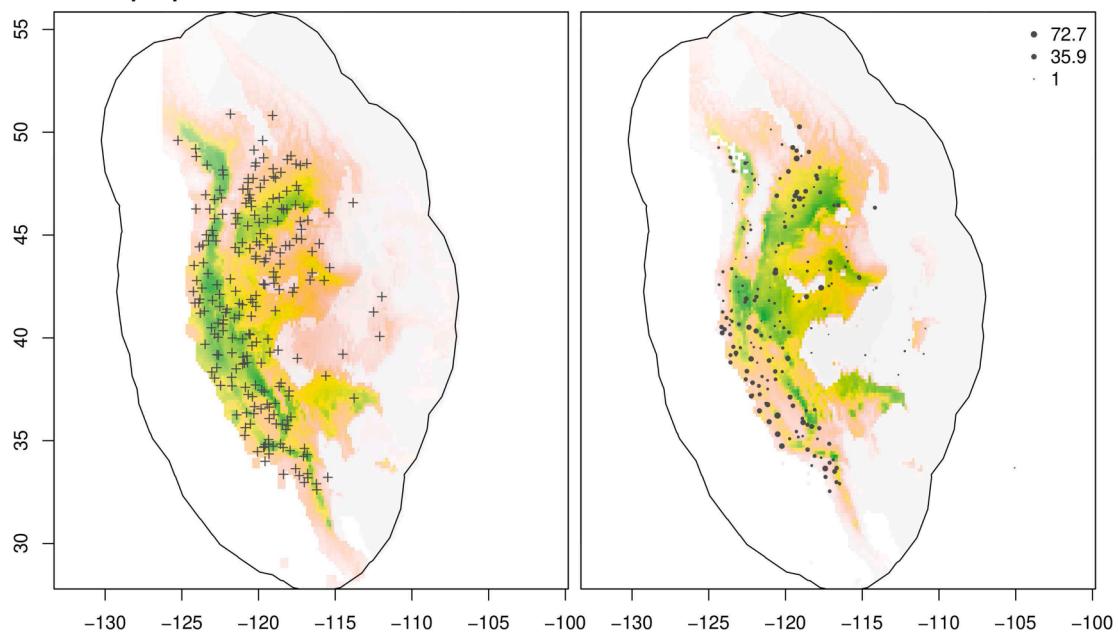
Results of the comparison of PPMs and MVEs to estimate niche centroids and suitability surfaces. Grey background colour highlights results for PPMs, blue for MVEs, and asterisks indicate statistical significance of the test. Centroid coordinates correspond to bioclimatic variable values given in the same order as they appear in Table 1 for each species. "Surfaces" refers to the correlation between the exponentiated distance to centroid and point intensity.

Species	Centroid coordinates		Distance between centroids	Correlation Surfaces ( $r$ )	Abundance ( $\rho$ )	
	PPM	MVE			PPM	MVE
<i>Callipepla californica</i>	-155.02, 104.9, 828.8	28.6, 20.6, 686.9	405.8	0.87***	0.04	0.006
<i>Calamospiza melanocorys</i>	309.9, 434.0, 176.6	297.2, 433.2, 181.0	0.28	0.94***	0.56***	0.52***

### *Calamospiza melanocorys*



### *Callipepla californica*



**Fig. 6.** Point intensity (left) and distance to centroid (right, exponentially transformed) for both bird species. Crosses are occurrence used to fit both PPMs and MVEs, and point sizes show average abundance estimates used to test the abundant niche centroid hypothesis.

#### Credit Author Statement

Gerardo Martín: conceptualisation, data curation, formal analysis, investigation, methodology, visualisation. Carlos Yáñez-Areas: conceptualisation, data curation, writing – review & editing. Xavier Chiappa-Carrara: writing – review & editing, acquiring funding, project administration.

#### Data accessibility statement

All data and code used in this study are publicly available in the repository <https://zenodo.org/badge/latestdoi/296682843>.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

We thank the Dirección General del Personal Académico (DGAPA) of the Universidad Nacional Autónoma de México for the postdoctoral scholarship granted to GM in September 2020. This research was funded by DGAPA-PAPIIT project number IA200822 awarded to GM.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ecolmodel.2022.109974](https://doi.org/10.1016/j.ecolmodel.2022.109974).

## References

- Angilletta, M.J., 2009. Thermal adaptation: a theoretical and empirical synthesis, Oxford biology. Oxford University Press, Oxford ; New York.
- Baddeley, A.J., Rubak, E., Turner, R., 2016. Spatial Point Patterns 830.
- Baddeley, A.J., Turner, R., 2005. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12, 1–42.
- Castaño-Quintero, S., Escobar-Luján, J., Osorio-Olvera, L., Peterson, A.T., Chiappa-Carrara, X., Martínez-Meyer, E., Yañez-Arenas, C., 2020. Supraspecific units in correlative niche modeling improves the prediction of geographic potential of biological invasions. *PeerJ* 8, e10454. <https://doi.org/10.7717/peerj.10454>.
- Dallas, T., Decker, R.R., Hastings, A., 2017. Species are not most abundant in the centre of their geographic range or climatic niche. *Ecol Lett* 20, 1526–1533. <https://doi.org/10.1111/ele.12860>.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Schachetti-Pereira, R., Schapire, R.E., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 2, 129–151.
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>.
- González-Salazar, C., Stephens, C.R., Marquet, P.A., 2013. Comparing the relative contributions of biotic and abiotic factors as mediators of species' distributions. *Ecological Modelling* 248, 57–70. <https://doi.org/10.1016/j.ecolmodel.2012.10.007>.
- Henze, N., Zirkler, B., 1990. A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods* 19, 3595–3617.
- Hutchinson, G.E., 1957. Concluding remarks. In: *Cold Spring Harbor Symposium on Quantitative Biology*, pp. 415–457.
- Isaac, N.J.B., Jarzyna, M.A., Keil, P., Dambly, L.I., Boersch-Supan, P.H., Browning, E., Freeman, S.N., Golding, N., Guillera-Arroita, G., Henrys, P.A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O.L., Schmucki, R., Simmonds, E.G., O'Hara, R.B., 2020. Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution* 35, 56–67. <https://doi.org/10.1016/j.tree.2019.08.006>.
- Jiménez, L., Soberón, J., 2020. Leaving the area under the receiving operating characteristic curve behind: An evaluation method for species distribution modelling applications based on presence-only data. *Methods in Ecology and Evolution* 11, 1571–1586. <https://doi.org/10.1111/2041-210X.13479>.
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4, 170122. <https://doi.org/10.1038/sdata.2017.122>.
- Logan, J.A., 1989. Derivation and Analysis of Composite Models for Insect Populations. McDonald, L.L., Manly, B.F.J., Lockwood, J.A., Logan, J.A. (Eds.). Estimation and Analysis of Insect Populations, Lecture Notes in Statistics. Springer, New York, NY, pp. 278–288. [https://doi.org/10.1007/978-1-4612-3664-1\\_19](https://doi.org/10.1007/978-1-4612-3664-1_19).
- Osorio-Olvera, L., Lira-Noriega, A., Soberón, J., Peterson, A.T., Falconi, M., Contreras-Díaz, R.G., Martínez-Meyer, E., Barve, V., Barve, N., 2020. ntbox: An r package with graphical user interface for modelling and evaluating multidimensional ecological niches. *Methods in Ecology and Evolution* 11, 1199–1206. <https://doi.org/10.1111/2041-210X.13452>.
- Osorio-Olvera, L., Soberón, J., Falconi, M., 2019. On population abundance and niche structure. *Ecography* 42, 1415–1425. <https://doi.org/10.1111/ecog.04442>.
- Osorio-Olvera, L., Yañez-Arenas, C., Martínez-Meyer, E., Peterson, A.T., 2020. Relationships between population densities and niche-centroid distances in North American birds. *Ecology Letters* 23, 555–564. <https://doi.org/10.1111/ele.13453>.
- Pápes, M., Peterson, A.T., Powell, G.V.N., 2012. Vegetation dynamics and avian seasonal migration: clues from remotely sensed vegetation indices and ecological niche modelling. *Journal of Biogeography* 39, 652–664. <https://doi.org/10.1111/jbi.12699.2011.02632.x>.
- Pedersen, L., Jensen, N.E., Christensen, L.E., Madsen, H., 2010. Quantification of the spatial variability of rainfall based on a dense network of rain gauges. *Atmospheric Research* 95, 441–454. <https://doi.org/10.1016/j.atmosres.2009.11.007>.
- Peterson, A.T., Pápes, M., Soberón, J., 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213, 63–72. <https://doi.org/10.1016/j.ecolmodel.2007.11.008>.
- Peterson, A.T., Soberón, J., 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *NatCon* 10, 102–107. <https://doi.org/10.4322/natcon.2012.019>.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Qiao, H., Peterson, A.T., Campbell, L.P., Soberón, J., Ji, L., Escobar, L.E., 2016. NicheA: creating virtual species and ecological niches in multivariate environmental scenarios. *Ecography* 39, 805–813. <https://doi.org/10.1111/ecog.01961>.
- Raes, N., Ter Steege, H., 2007. A null-model for significance testing of presence-only species distribution models. *Ecography* 30, 727–736. <https://doi.org/10.1111/j.2007.0906-7590.05041.x>.
- Renner, I.W., Elith, J., Baddeley, A.J., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I., 2015. Point process models for presence-only analysis. *Methods in Ecology and Evolution* 6, 366–379. <https://doi.org/10.1111/2041-210X.12352>.
- Renner, I.W., Warton, D.I., 2013. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* 69, 274–281. <https://doi.org/10.1111/j.1541-0420.2012.01824.x>.
- Saupe, E.E., Pápes, M., Selden, P.A., Vetter, R.S., 2011. Tracking a medically important spider: climate change, ecological niche modeling, and the brown recluse (*Loxosceles reclusa*). *PLoS one* 6, e17731. <https://doi.org/10.1371/journal.pone.0017731>.
- Soberón, J., Nakamura, M., 2009. Niches and distributional areas : Concepts, methods, and assumptions. *PNAS* 106, 19644–19650.
- Soberón, J., Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity informatics* 2, 1–10. <https://doi.org/10.1093/wber/lhm022>.
- Tingley, M.W., Monahan, W.B., Beissinger, S.R., Moritz, C., 2009. Birds track their Grinnellian niche through a century of climate change. *Proceedings of the National Academy of Sciences of the United States of America* 106 (Suppl), 19637–19643. <https://doi.org/10.1073/pnas.0901562106>.
- Warton, D.I., Renner, I.W., Ramp, D., 2013. Model-based control of observer bias for the analysis of presence-only data in ecology. *PLoS one* 8, e79168. <https://doi.org/10.1371/journal.pone.0079168>.
- Yañez-Arenas, C., Martín, G., Osorio-Olvera, L., Escobar-Luján, J., Castaño-Quintero, S., Chiappa-Carrara, X., Martínez-Meyer, E., 2020. The Abundant Niche-centroid Hypothesis: Key Points About Unfilled Niches and the Potential Use of Supraspecific Modeling Units. *Biodiversity Informatics* 15, 92–102. <https://doi.org/10.17161/bi.v15i2.13218>.
- Yaya, O.S., Vo, X.V., 2020. Statistical analysis of rainfall and temperature (1901–2016) in south-east Asian countries. *Theor Appl Climatol* 142, 287–303. <https://doi.org/10.1007/s00704-020-03307-z>.