APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 7, 2023, 49-62

Inhomogeneous Spatial Point Process Models for Species Distribution Analysis: A Systematic Review

Judie Armel Bourobou Bourobou[1,2*], Adandé Belarmain Fandohan[1,3] and Roman Lucas Glèlè Kakaï[1]

[1]Laboratoire de Biomathématiques et d'Estimations Forestières - LABEF, Université d'Abomey-Calavi – UAC, Cotonou, Benin
[2]Institut de Recherches Agronomique et Forestière - IRAF, Centre National de Recherche Scientifique et Technologique - CENAREST, Libreville, Gabon
[3]Unité de Recherche en Foresterie et Conservation des BioRessources - URFCBio, Ecole de Foresterie Tropicale - EForT, Université Nationale d'Agriculture - UNA, Kétou, Benin

[*]Corresponding author: judiarmelbourobou@gmail.com

Submitted 06 March 2023, Revised 01 May 2023, Accepted 03 May 2023, Available online 05 May 2023.

Abstract: This study aims to systematically review the application of inhomogeneous spatial point process models (ISPPMs) for species distribution analysis. The review focused on (i) the trend in the use of ISPPMs, (ii) the general characteristics of the studies reviewed, and (iii) the practice of inhomogeneous spatial point process modeling. Based on specific criteria, a search using Publish or Perish (PoP) software and Google Scholar databases was performed for published papers on ISPPMs from 2006 to 2020. The study revealed a significant evolution in the use of ISPPMs. Most of the studies were conducted at regional, national, and continental scales. More than 60% of the papers used presence-only data. The linear model was the most used (47.12%). Maximum likelihood (21%) and minimum contrast estimation (19%) were the primary methods for estimating the fitted model parameters. The goodness of fit, performance analysis and model comparison guided fitting model validation. Moreover, many of these studies (56.91%) did not explicitly address the issues of model specification and spatial dependence. Furthermore, 47% of the articles considered did not clarify the estimation method used. New challenges and perspectives are to be explored.

Keywords: Cox point process; Inhomogeneous point process; Inhomogeneous Poisson process; Markov point process; Spatial point process models.

1. INTRODUCTION

In practice, knowledge of the spatial distribution of plants or animals is of great value in planning landscape management actions, biology conservation, and predicting the impacts of climate change on species [1]. The most typical method for estimating a species's actual or potential geographic distribution is first to characterize the environmental conditions suited to the species and then map out the spatial spread of those settings [2]. The suitable ecological conditions for a species may be characterized using a species distribution model (SDM). Therefore, over the past decades, ecologists have devoted significant attention to using SDMs to estimate those conditions by associating known species' occurrence records with suites of environmental variables that could reasonably be assumed to affect the species' physiology and probability of persistence [2-4].

The increase in the use of SDMs by ecologists in the last years has permitted the development of new estimation methods such as Ecological Niche Factor (ENFA; [5]), Genetic Algorithm for Rule Set Production (GARP; [6]), Maximum Entropy (Maxent; [7]) and more recently Bayesian methods [8,9], for presence-only data, site-occupancy models [10] or zero-inflated binomial (ZIB) for presence-absence data, N-mixture models [11] or zero-inflated Poisson (ZIP) for abundances data.

Recent studies have indicated inhomogeneous Poisson point process models (IPPMs; [12]) as a natural framework to conceptualize all these data types [13–16]. The IPPMs assume that (i) the locations of the n point events $(y_i, \ldots, y_n)$ are independent, and (ii) the intensity $\lambda(y_i)$, at point y_i, can be modeled as a function of the k explanatory variables. Then, a species occurrence could be modeled as arising from an inhomogeneous Poisson point process (IPP). We assume the following log-linear specification:

$$log\lambda_i = \beta_0 + \sum_{j=1}^{k} x_{ij}\,\beta_j \qquad (1)$$

Let Z be a random set of occurrence records falling in some geographical study area $\mathcal{D}$ $(here\ \mathcal{D} \subseteq \mathbb{R}^2)$. The IPPM is a probabilistic model for the random set $Z = \{z_i\} \subseteq \mathcal{D}$. It can be defined by an intensity function λ, which assigns a positive real-valued intensity $\lambda(z)$ to each point $z \in \mathcal{D}$. The density of IPP over the domain $\mathcal{D}$ is defined by:

$$P_\lambda = \frac{\lambda(z)}{\int_{\mathcal{D}}^{\cdot} \lambda(z)\, dz} \tag{2}$$

This density function represents the species distribution, while the intensity function $\lambda(z)$ informally quantifies how many z_i are likely to occur near z. The number of points N_z, equal to the number of presences in region $\mathcal{D}$, follows itself a Poisson distribution with mean:

$$\Lambda(\mathcal{D}) = \int_{\mathcal{D}}^{\cdot} \lambda(z)\, dz \tag{3}$$

The assumption of linearity can be relaxed, e.g., using quadratic terms or splines. The interest of these models is that they can be extended to the hierarchical framework, which can consider spatial autocorrelation [16]. The Cox process relaxes the independence assumption and deals with clustering and the effects of unmeasured covariates [17]. The most commonly used example of the Cox process model is the spatial log-Gaussian Cox process model (LGCP). Its intensity $\Lambda(z)$ can be expressed as follows:

$$\Lambda(z) = \boldsymbol{x}(z)'\boldsymbol{\beta} + \xi(z) \tag{4}$$

where $\boldsymbol{x}$ is a vector of environmental variables and $\xi(z)$ is a spatial Gaussian process with zero as the mean, its covariance function depends on the distance between observations. However, the Gibbs process relaxes this assumption by assuming interactions between points. An example of this process is area interaction (see [18]).

$$\Lambda(z) = \boldsymbol{x}(z)'\boldsymbol{\beta} + t_z(Z_p)\theta \tag{5}$$

t_z is the area of a disc of radius r centered at the location z that does not intersect with similar discs centered around each of the presence points Z_p. θ is an interaction parameter. $If\ \theta = 1$, then the model is reduced to the Poisson process with intensity $\boldsymbol{x}(z)'\boldsymbol{\beta}$; $if\ \theta < 1$, then the process is regular; $if\ \theta > 1$, then the process is clustered.

The three models described above are examples of inhomogeneous spatial point process models (ISPPMs). In the modeling framework, ISPPMs are powerful tools for analyzing spatial data in various applications such as ecology, astronomy, and epidemiology [19]. For example, point patterns with nonhomogeneous intensities are frequently seen in both nature and technology. And maps showing plant locations in more significant regions with changing conditions typically look inhomogeneous [20]. Thus, the ISPPMs are particularly relevant in ecology, where there is a strong interest in understanding the spatial distribution and abundance of individuals or groups in space [21].

To our knowledge, there is no synthesis on the use of ISPPMs. A review of these methods should help SDM users to choose which methods to use for their analyses. Hence, the primary purpose of this study is to review the papers involving ISPPMs in species distribution analysis. For this study, we will examine (i) the trend in the use of ISPPMs, (ii) the general characteristics of the studies conducted, and (iii) the practice of inhomogeneous spatial point process modeling. Based on this assessment, critical confounding factors that need to be addressed concerning the performance of ISPPMs will be identified.

2. RESEARCH METHODOLOGY

This study followed the Preferred Reporting Items for Systematic Reviews and Metanalyses (PRISMA) statement [22,23]. A strategic search of records in Inhomogeneous Spatial Point Process Modeling between 2006 and 2020 was conducted using Publish or Perish (PoP) software and Google scholar databases. This search included the following three equations: (i) ("inhomogeneous Poisson process" OR" spatial point process" AND" species distribution"), (ii) (" Cox point process" OR" Gibbs process" OR" Markov process" AND" species distribution"), and (iii) (" inhomogeneous spatial point process" AND" species distribution").

The inclusion criteria were determined based on the original research papers written in English and reporting an application of ISPPM. The documents related to statistical methodology development were excluded. The search records have been downloaded and processed using Zotero software. Figure 1 summarizes all steps in the paper selection process. In total, 699 records were identified, including 37 duplicates excluded from the analysis. We also excluded all records considering non-original articles (books, reviews, reports, preprinted articles) and articles published in a language other than English. Finally, a detailed review of 88 published articles was conducted (see Appendix A).

Based on these 88 papers, the number of papers published by the journal and by year of publication was first examined. The general features of these reviewed studies were then examined. For that purpose, the following aspects were examined: field of study (plant or animal ecology, conservation, evolution), area of study (range, type of environment), sampling design, and sample size. Finally, inferential and model validation issues were examined. To achieve this, we were interested in the type and source of data, the specification of models (linear, non-linear, smoothing, spline, and quadratic, the appropriateness of the number of covariates included in the model provided, and the number of observations), the fitted model, the estimation methods used, the model evaluation methods, the statistical software, and packages.

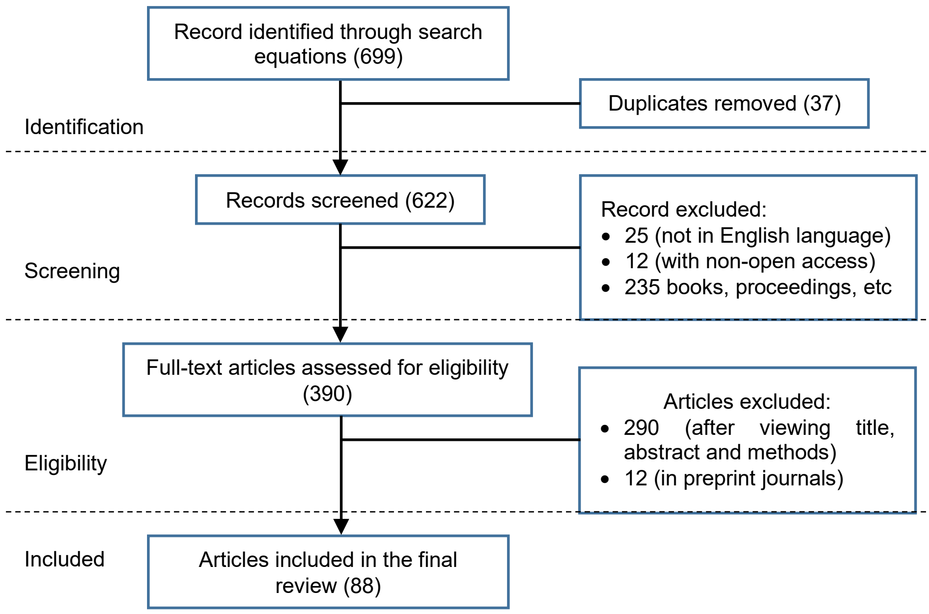


Figure 1. Flowchart of the selection of papers reviewed

3. RESULTS

Up to 31 December 2020, 699 records associated with the search equations were identified. After carefully examining these documents, duplicates, grey literature records, and non-relevant papers were excluded. Finally, only 88 published articles dealing with inhomogeneous spatial point process modeling were retained for an eventual examination.

3.1 Trend in the Use of Inhomogeneous Spatial Process Model

The reviewed studies were published between 2006 and 2020 (Figure 2). Despite the peaks observed in 2015 and 2019, the evolution of ISPPM for species distribution analysis showed a general exponential trend. (Figure 2a). As shown in Figure 2b, the top journals in which the published papers were published were the following: Plos One (8.14%), Methods in Ecology and Evolution (6.98%), Ecology (6.98%), Landscape Ecology (3.49%), Journal of Vegetation (6.38%) and Ecography (3.49%).

3.2 Characteristics of Reviewed Studies

Around 22% of these studies used opportunistic sampling, 68.48% used systematic design, and 9.78% used both (opportunistic and systematic designs). Published studies were carried out mainly in Plant Ecology (51.14%) and Animal Ecology (31.82%), but also in Epidemiology (11.36%), microbial Ecology (2.27%), Fungal ecology (1.14%), Paleoecology (1.14%), and Animal conservation (1.14%) (Figure 2c). These studies were conducted in dryland (92%), Seas (2%), and Antarctica (1%). They were carried out worldwide, especially in Australia, the USA, China, France, and Africa (Figure 2d).

Approximately 43% of the studies included in this review were conducted on areas smaller than one km², 15.12% on [800; + [km², 10.47% on national regions, 18.60% on regional areas, and 12.79% on [1; 800 [km² (Figure 3a). In these studies, the sample size ranged from 3 to over 1000 (Figure 3b).

3.3 Practice of ISPPMs

3.3.1 Types of Data and Their Source

Only 31.91% analyzed count data, 3.19% presence-absence data, and 4.26% papers considered presence-only and presence-absence. Among the papers, 60.64% used presence-only data. Moreover, 50.85% of presence-only data came from systematic studies (census, survey). In comparison, 37.29% came from various sources such as GBIF, citizen science database, Atlas, PubMed database, eBird database, and Museum collections, and collected through opportunistic sightings, reporting, and historical samples. Only 11.86% of Presence-only data were from both Systematic and opportunistic studies.

3.3.2 ISPPMs Used in the Consulted Papers

The inhomogeneous Poisson process model was the predominant model in the papers collected, followed by the Log Gaussian Cox process model and the inhomogeneous Thomas process model (Figure 3c). Gibbs class models such as Strauss point process model, Area-interaction model, hard-core point process model, and Geyer point process model were poorly used.

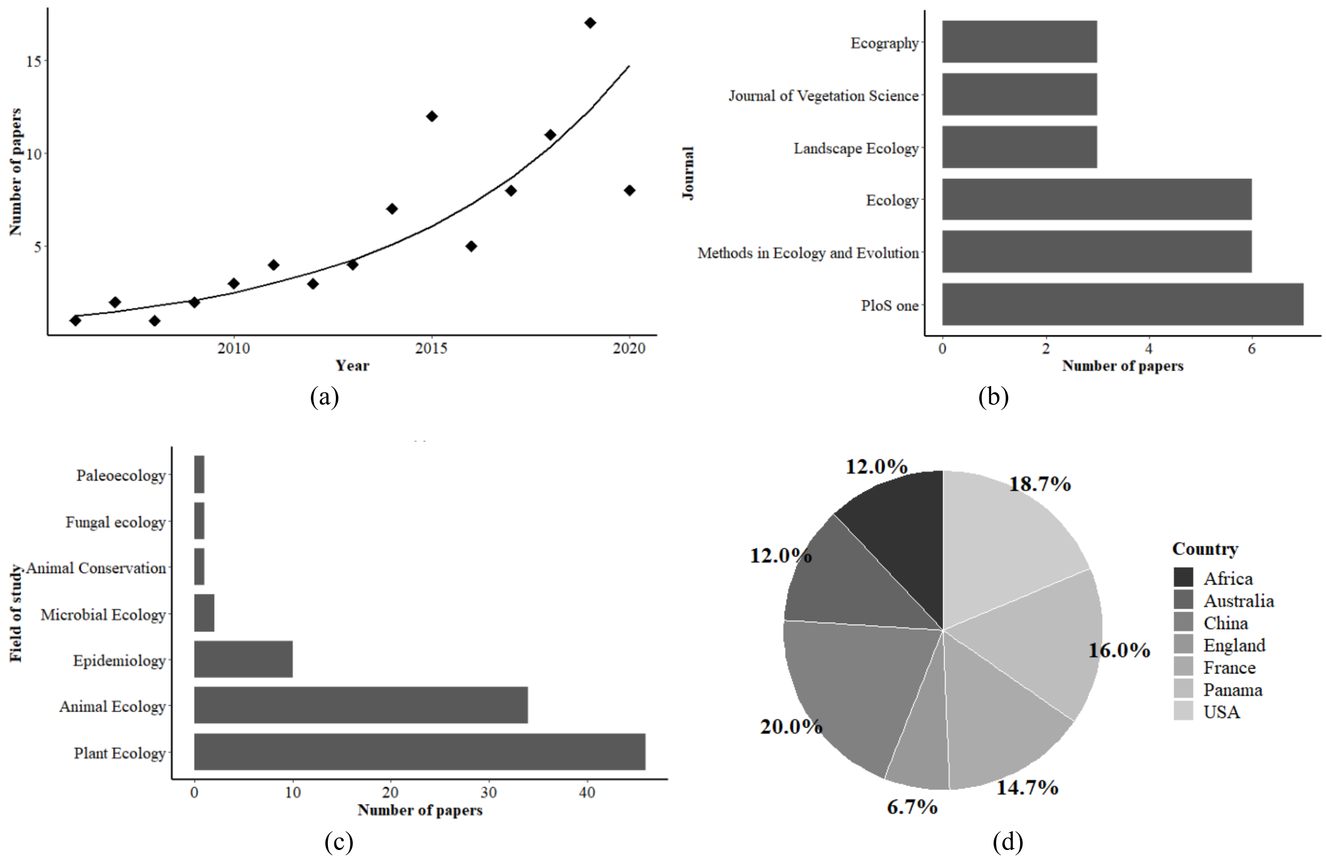
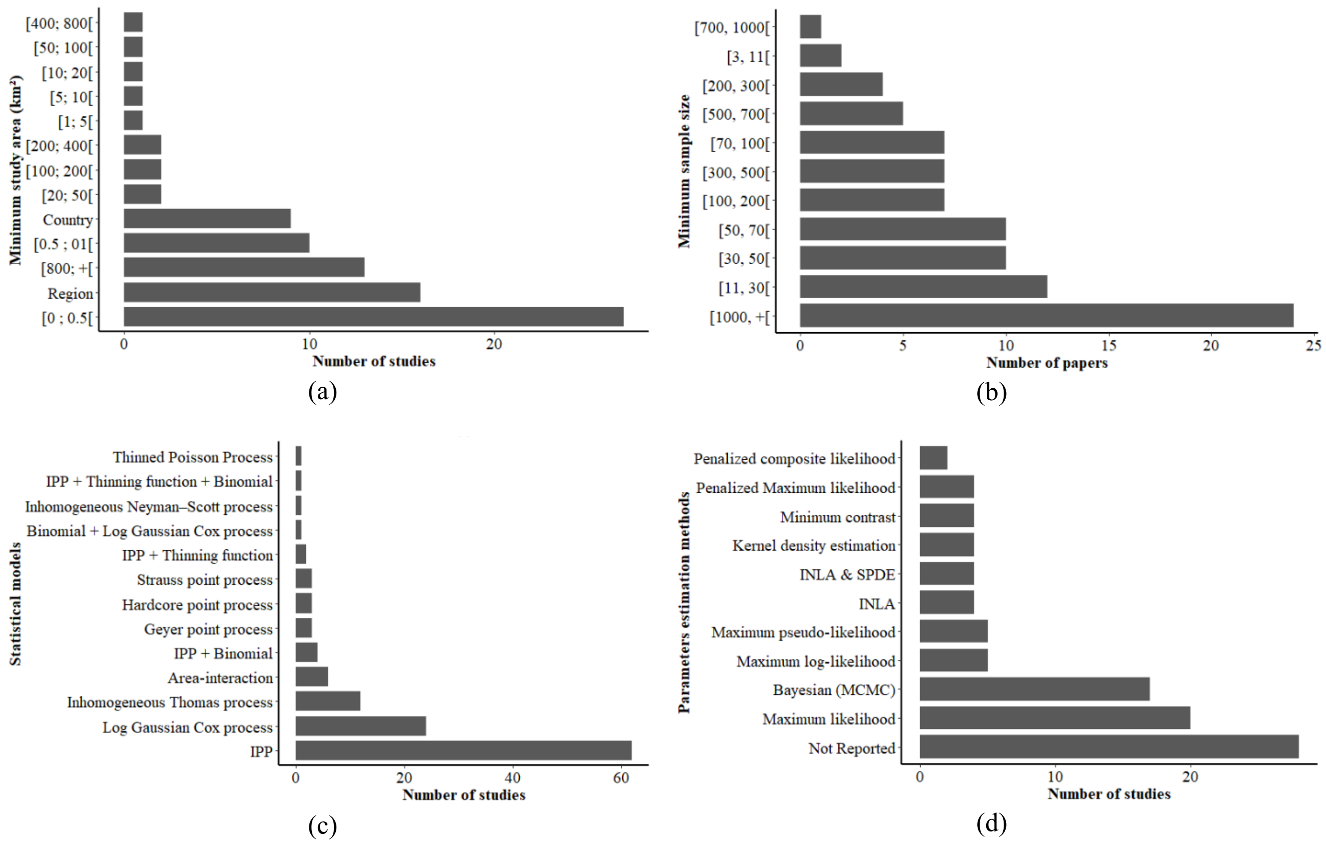


Figure 2. Number of published papers on ISPPMs: (a) by year; (b) by journal; (c) Area of study; (d) Major countries (with more than one study)



IPP: Inhomogeneous Poisson Process, INLA: Integrated Nested Laplace Approximation, SPDE: Stochastic Partial Differential Equation, MCMC: Markov chain Monte Carlo

Figure 3. (a) Minimum number of locations used in each study, (b) Extends of study areas, (c) Statistical models used in the reviewed studies, (d) Main parameters estimation methods used in the reviewed studies

Table 1. Model specifications

Form of model	Frequency in the reviewed studies (%)
Cubic term	0.96
Interaction term	1.92
Linear	47.12
Non-linear (unspecified form)	14.42
Quadratic term	7.69
Smooth term	3.85
Not reported	24.04

3.3.3 Model Specifications and Spatial Dependence

The linear model remained the most used (47.2%) compared to the non-linear models (28.14%) (Table 1). Nevertheless, many papers did not provide information on model specification (24.04%). Furthermore, even when mentioned, the non-linear specification was frequently unclear on which specific effect was included (i.e., cubic, quadratic, or smoothing terms). However, models including quadratic terms are more commonly used than other non-linear effect models. The spatial dependence was not considered in about 57% of the reviewed studies.

3.3.4 Parameter Estimation Methods

More than ten parameter estimation methods were identified in the ISPPMs literature (Figure 3d), with the main methods being the maximum likelihood (21%) and the minimum contrast estimation (19%). However, 47% of the papers did not mention the estimation method.

3.3.5 Model Validations

Appendix B provides a comprehensive list of statistical models, parameter estimation methods, model validation approaches, and software packages used in the reviewed papers. It was noted that statistical modeling principles and steps were often followed. A wide range of model formulations can be used to implement the inhomogeneous Poisson point or log-Gaussian Cox process models (see Appendix B). Furthermore, almost all statistical analyses were performed in R software. Three validation approaches for the fitted models were identified: goodness of fit, performance analysis, and model comparison.

4. DISCUSSION

This study aimed to examine the use of inhomogeneous spatial point process models in species distribution analysis. It examined how these ISPPMs were used over the last fifteen years and identified some critical confounding factors regarding their performance that need to be discussed.

While the number of studies using ISPPMs has followed an exponential trend over the past 15 years, these statistical methods remain unpopular and poorly explored in depth, especially in Africa. It also indicated that ISPPMs remain undervalued in the SDM literature; fewer than 100 studies used ISPPMs in this review. It may be imputable to the lack of ISPPM techniques in standard training curricula for ecologists, epidemiologists, and conservationists [24].

ISPPMs have gained increasing interest in animal and plant ecology over the past fifteen years, especially with presence-only data as anticipated by previous works [16,25–28]. Therefore, ISPPMs offer a better framework for analyzing presence-only data [26]. The linear form of ISPPM was used in most of the reviewed papers without any appropriate validation of its suitability to the used data set. Indeed, when models include continuous predictors, linear models are often preferred [29]. However, inadequate model specification and suboptimal parametrization often result in poor performance [29,30]! Thus, checking, detecting, and accounting for non-linear effects in the model specification is crucial to improve the model's performance. In this regard, some papers advocate using interaction terms or smooth functions such as splines or wavelets [13,26] as in generalized additive models [31].

Several methods were used in the reviewed papers to estimate the parameters of the fitted models. The maximum likelihood (21% for the inhomogeneous Poisson point process) and the minimum contrast estimation (19% for Log Gaussian Cox and inhomogeneous Thomas processes) were used most. Although weakly used in the published studies, the Integrated Nested Laplace Approximation (INLA) implemented in the R-INLA package [23, 29] is helpful for Cox process estimation. The phenomenon of spatial dependence or spatial autocorrelation (SAC) is well documented in spatial point analysis (see [32-34]). However, most of the reviewed papers did not account for it. This implies a risk of overestimating the relationship between environmental covariates and species presence, potentially resulting in a poor predictive ability [32] extrapolating to unknown geographical areas since the process generating the spatial structure varies with space and time [35].

The growing use of species distribution modeling (SDM) has sparked new concerns about inaccuracies and misuse of this critical tool. Collinearity between environmental predictors is an example of the source of these issues. It is hypothesized to be a significant source of model uncertainty; collinearity is expected to increase the uncertainty of model parameters and decrease statistical power [36]. Many methods have been designed to deal with collinearity problems, but they seem to be neglected by the current statistical machines applied to SDM [36]. Dormann et el. [37] reviewed these methods, showing considerable variation in performance among them. Nevertheless, latent variable methods have shown promising results, such as using scores from a principal component analysis (PCA) in regression. Variables selection is another alternative to address the collinearity problem. Recent work has been done on implementing variable selection for inhomogeneous spatial point models to reduce variance inflation due to overfitting and bias due to underfitting. The authors examined using the LASSO, adaptive LASSO, and elastic net methods in a simulation study. Their simulations found that the adaptive LASSO method performed best for the Poisson process (see [19]).

4.1 Challenges and Prospects

Inhomogeneous spatial point process models are a relatively new class of methods in species distribution modeling, with a variety of models today. How statistical models can be implemented permits substantial flexibility in SDM complexity, critical for robust inference [38].

The inhomogeneous spatial point process models would be a natural approach for modeling presence-only data (see, for example, [13, 27]). However, the PO data generally suffer from various key issues [39], such as:

- Sample selection bias [40, 41], where opportunistic sampling occurs in areas that are easily accessible (e.g., near roads);
- The imperfect detection [42];
- Moreover, because absences are unavailable, only relative suitability/probability measures can typically be modeled [43, 44].

Several authors advocate using multiple data sources in modeling species distributions to reduce uncertainties and increase model performance (see [45-49]). However, appropriately combining various data is a challenge since each data source has different assumptions and biases, and each may provide additional information on species distribution [39]. Thus, recent advances in species distribution modeling have focused on better integrating these different data sources to achieve better predictions and statistical inference [46, 47, 50, 51].

However, using PO data without combining other data sources remains an alternative [39]. For example, [52] demonstrated that citizen science datasets could reproduce forecasts from models based on systematically collected colonization–extinction data and lead to the same forest management conclusions. Second, researchers can use only PO data for species distribution modeling due to systematic studies' logistical challenges and financial costs [53].

Therefore, evaluating the empirical performance of ISPPMs under specific conditions and developing case-specific model specifications would lead to more promising results. In this regard, potential research avenues could consider evaluating the empirical performance of ISPPMs by varying the combination of some factors such as (i) statistical approach (frequentist or Bayesian), (ii) parameter estimation methods, (iii) sample size, (iv) sampling bias, (v) pixel resolution, (vi) model specification, (vii) type of data to be analyzed, (viii) collinearity between predictor variables.

5. CONCLUSION

This paper is intended to conduct a systematic review of the use of ISPPMs to analyze species distributions. From 2006-2020, the use of ISPPMs in ecology, conservation, and epidemiology grew to account for various aspects of statistical modeling, such as model specification, spatial dependence, model validation, data type, and source. However, ISPPMs have shown some limitations when applied to PO data, which is generally subject to sampling bias or imperfect detection. Several authors advocated combining multiple data sources to reduce uncertainties and improve model performance when modeling species distributions with PO data. Nevertheless, for African researchers facing logistical challenges and financial costs associated with systematic studies, using only opportunistic data, without combining other data sources, remains an option. Thus, knowing the empirical performance of ISPPMs trained on PO data would benefit those researchers in choosing statistical models for the species spatial distribution analysis.

ACKNOWLEDGMENT

We acknowledge the German Academic Exchange Service (DAAD) support under the "Ph.D. In-Country/In-Region Scholarship Programme FSA/UAC 20".

REFERENCES

[1] C. Botella, A. Joly, P. Bonnet, P. Monestiez and F. Munoz, Species distribution modeling based on the automated identification of citizen observations, *Applications in Plant Sciences*, 6(2), 2018, e1029.
[2] R. G. Pearson, Species' distribution modeling for conservation educators and practitioners, *Synthesis. American Museum of Natural History*, 50, 2019, 54-89.
[3] A. Guisan and W. Thuiller, Predicting species distribution: Offering more than simple habitat models, *Ecology Letters*, 8, 2005, 993-1009.
[4] A. T. Peterson, J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura and M. B. Araújo, *Ecological Niches and Geographic Distributions* (MPB-49), Princeton University Press, 2011.
[5] A. H. Hirzel, J. Hausser, D. Chessel and N. Perrin, Ecological-niche factor analysis: How to compute habit suitability maps without absence data?, *Ecology*, 83(7), 2002, 2027-2036.
[6] D. Stockwell, The GARP modelling system: problems and solutions to automated spatial prediction, *International Journal of Geographical Information Science*, 13, 1999, 143-158.

[7] M. Dudík, S. Phillips and R. E. Schapire, Correcting sample selection bias in maximum entropy density estimation, In *Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS'05)*, 2005, 323-330.
[8] F. Tonini, F. Divino, G. J. Lasinio, H. H. Hochmair and R. H. Scheffrahn, Predicting the geographical distribution of two invasive termite species from occurrence data, *Environmental Entomology*, 43, 2014, 1135-1144.
[9] F. Divino, N. Golini, G. Jona Lasinio and A. Penttinen, Bayesian logistic regression for presence-only data, *Stochastic Environmental Research and Risk Assessment*, 29, 2015, 1721-1736.
[10] D. I. Mackenzie, J. D. Nichols, J. E. Hines, M. G. Knutson and A. B. Franklin, Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly, *Ecology*, 84, 2003, 2200-2207.
[11] J. A. Royle, N-Mixture models for estimating population size from spatially replicated counts, *Biometrics*, 60, 2004, 108-115.
[12] N. A. C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York, 1993.
[13] W. Fithian and T. Hastie, Finite-sample equivalence in statistical models for presence-only data, *Annals of Applied Statistics*, 7, 2013, 1917.
[14] S. J. Phillips, R. P. Anderson, M. Dudík, R. E. Schapire and M. E. Blair, Opening the black box: An open-source release of Maxent, *Ecography*, 40, 2017, 887-893.
[15] I. W. Renner, J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic and D. I. Warton, Point process models for presence-only analysis, *Methods in Ecology and Evolution*, 6, 2015, 366-379.
[16] D. I. Warton and L. C. Shepherd, Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology, *Annals of Applied Statistics*, 4, 2010, 1383-1402.
[17] J. Møller, A. R. Syversveen and R. P. Waagepetersen, Log Gaussian cox processes, *Scandinavian Journal of Statistics*, 25, 1998, 451-482.
[18] A. J. Baddeley and M. N. M. Van Lieshout, Area-interaction point processes, *Annals of the Institute of Statistical Mathematics*, 47, 1995, 601-619.
[19] Y. R. Yue and J. M. Loh, Variable selection for inhomogeneous spatial point process models, *Scandinavian Journal of Statistics*, 43, 2015, 288-305.
[20] U. Hahn, E. B. V. Jensen, M. -C. Van Lieshout and L. S. Nielsen, Inhomogeneous spatial point processes by location-dependent scaling, *Advances in Applied Probability*, 35(2), 2003, 319-336.
[21] C. Liu, G. Newell and M. White, The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites, *Ecography*, 42, 2019, 535-548.
[22] B. Hutton, G. Salanti, A. Chaimani, D. M. Caldwell, C. Schmid, K. Thorlund, E. Mills, F. Catalá-López, L. Turner, D. G. Altman and D. Moher, The quality of reporting methods and results in network meta-analyses: An overview of reviews and suggestions for improvement, *PLOS ONE*, 9, 2014, e92508.
[23] D. Moher, A. Liberati, J. Tetzlaff and D. G. Altman, The PRISMA Group, preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, *PLOS Medicine*, 6, 2009, e1000097.
[24] J. Velásquez-Tibatá, C. H. Graham and S. B. Munch, Using measurement error models to account for georeferencing error in species distribution models, *Ecography*, 39, 2016, 305-316.
[25] N. E. Zimmermann, T. C. Edwards Jr, C. H. Graham, P. B. Pearman and J.-C. Svenning, New trends in species distribution modelling, *Ecography*, 33, 2010, 985-989.
[26] I. W. Renner, J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic and D. I. Warton, Point process models for presence-only analysis, *Methods in Ecology and Evolution*, 6, 2015, 366-379.
[27] A. Guisan, R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. Tulloch, T. J. Regan, L. Brotons, E. McDonald-Madden and C. Mantyka-Pringle, Predicting species distributions for conservation decisions, *Ecology Letters*, 16, 2013, 1424-1435.
[28] J. Elith and J. R. Leathwick, Species distribution models: ecological explanation and prediction across space and time, *Annual Review of Ecology, Evolution, And Systematics*, 40, 2009, 677-697.
[29] M. Perera and C. Tsokos, A statistical model with non-linear effects and non-proportional hazards for breast cancer survival analysis, *Advances in Breast Cancer Research*, 7, 2018, 65.
[30] R. Halvorsen, A gradient analytic perspective on distribution modelling, *Sommerfeltia*, 35, 2012, 1-165.
[31] T. Hastie and R. Tibshirani, Generalized additive models: some applications, *Journal of the American Statistical Association*, 82, 1987, 371-386.
[32] C. F. Dormann, J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz and W. Daniel Kissling, Methods to account for spatial autocorrelation in the analysis of species distributional data: A review, *Ecography*, 30, 2007, 609-628.
[33] M. -J. Fortin, P. Drapeau and P. Legendre, Spatial autocorrelation and sampling design in plant ecology, *Vegetatio*, 83(1:2), 1990, 209-222.
[34] R. Halvorsen, S. Mazzoni, J. W. Dirksen, E. Næsset, T. Gobakken and M. Ohlson, How important are choice of model selection method and spatial autocorrelation of presence data for distribution modelling by MaxEnt?, *Ecological Modelling*, 328, 2016, 108-118.
[35] J. Elith, M. Kearney and S. Phillips, The art of modelling range-shifting species, *Methods in Ecology and Evolution*, 1, 2010, 330-342.
[36] P. De Marco and C. C. Nóbrega, Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation, *PLOS ONE*, 13, 2018, e0202403.
[37] C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell and S.

Lautenbach, Collinearity: A review of methods to deal with it and a simulation study evaluating their performance, *Ecography*, 36, 2013, 27-46.
[38] C. Merow, M. J. Smith, T. C. Edwards, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wüest, N. E. Zimmermann and J. Elith, What do we gain from simplicity versus complexity in species distribution models?, *Ecography*, 37, 2014, 1267-1281.
[39] R. J. Fletcher Jr, T. J. Hefley, E. P. Robertson, B. Zuckerberg, R. A. McCleery and R. M. Dorazio, A practical guide for combining data to model species distributions, *Ecology*, 100, 2019, e02710.
[40] Y. Fourcade, J. O. Engler, D. Rödder and J. Secondi, Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias, *PLOS ONE*, 9, 2014, e97122.
[41] S. J. Phillips, M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick and S. Ferrier, Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data, *Ecological Applications*, 19, 2009, 181-197.
[42] T. J. Hefley, A. J. Tyre, D. M. Baasch and E. E. Blankenship, Nondetection sampling bias in marked presence-only data, *Ecology and Evolution*, 3, 2013, 5225-5236.
[43] J. A. Royle, R. B. Chandler, C. Yackulic and J. D. Nichols, Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions, *Methods in Ecology and Evolution*, 3, 2012, 545-554.
[44] R. M. Dorazio, Predicting the geographic distribution of a species from presence-only data subject to detection errors, *Biometrics*, 68, 2012, 1303-1312.
[45] L. Bedriñana-Romano, R. Hucke-Gaete, F. A. Viddi, J. Morales, R. Williams, E. Ashe, J. Garcés-Vargas, J. P. Torres-Florez and J. Ruiz, Integrating multiple data sources for assessing blue whale abundance and distribution in Chilean Northern Patagonia, *Diversity and Distributions*, 24, 2018, 991-1004.
[46] W. Fithian, J. Elith, T. Hastie and D.A. Keith, Bias correction in species distribution models: pooling survey and collection data for multiple species, *Methods in Ecology and Evolution*, 6, 2015, 424-438.
[47] K. Pacifici, B. J. Reich, D. A. Miller, B. Gardner, G. Stauffer, S. Singh, A. McKerrow and J. A. Collazo, Integrating multiple data sources in species distribution modeling: A framework for data fusion, *Ecology*, 98, 2017, 840-850.
[48] I. W. Renner, J. Louvrier and O. Gimenez, Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization, *Methods in Ecology and Evolution*, 10, 2019, 2118-2128.
[49] J. P. Rose, B. J. Halstead and R. N. Fisher, Integrating multiple data sources and multi-scale land-cover data to model the distribution of a declining amphibian, *Biological Conservation*, 241, 2020, 108374.
[50] S. L. Peel, N. A. Hill, S. D. Foster, S. J. Wotherspoon, C. Ghiglione and S. Schiaparelli, Reliable species distributions are obtainable with sparse, patchy and biased data by leveraging over species and data types, *Methods in Ecology and Evolution*, 10, 2019, 1002-1014.
[51] M.V. Talluto, I. Boulangeat, A. Ameztegui, I. Aubin, D. Berteaux, A. Butler, F. Doyon, C.R. Drever, M.-J. Fortin, T. Franceschini, J. Liénard, D. McKenney, K.A. Solarik, N. Strigul, W. Thuiller and D. Gravel, Cross-scale integration of knowledge for predicting species ranges: A metamodelling framework, *Global Ecology and Biogeography*, 25, 2016, 238-249.
[52] L. Mair, P. J. Harrison, M. Jönsson, S. Löbel, J. Nordén, J. Siitonen, T. Lämaas, A. Lundström and T. Snäll, Evaluating citizen science data for forecasting species responses to national forest management, *Ecology and Evolution*, 7, 2017, 368-378.
[53] A. Srivathsa, M. Puri, N. S. Kumar, D. Jathanna and K. U. Karanth, Substituting space for time: Empirical evaluation of spatial replication as a surrogate for temporal replication in occupancy modelling, *Journal of Applied Ecology*, 55, 2018, 754-765.

APPENDIX A

Different Articles Reviewed as Part of the Study

No	ID	Title	Journal	Authors and Year
1	Obs1	Trees outside forests in agricultural landscapes: spatial distribution and impact on habitat connectivity for forest organisms	Landscape Ecology	Rossi et al., 2016
2	Obs10	Spatial ecology of bacteria at the microscale in soil	PLOS ONE	Raynaud & Nunan, 2014
3	Obs11	Spatial Association and Diversity of Dominant Tree Species in Tropical Rainforest, Vietnam	Forests	Nguyen et al., 2018
4	Obs12	Size-class effect contributes to tree species assembly through influencing dispersal in tropical forests	PLOS ONE	Hu et al., 2014
5	Obs13	Saltcedar (Tamarix mascatensis inhibits growth and spatial distribution of eshnan (Seidlitzia rosmarinus by enrichment of soil salinity in a semi-arid desert	Plant and Soil	Erfanifard & Khosravi, 2019

6	Obs14	Quantifying effects of habitat heterogeneity and other clustering processes on spatial distributions of tree species	Ecology	Shen et al., 2013
7	Obs15	Point patterns of tree distribution determined by habitat heterogeneity and dispersal limitation	Oecologia	Lin et al., 2011
8	Obs16	Opportunistic records reveal Mediterranean reptiles' scale-dependent responses to anthropogenic land use	Ecography	De Solan et al., 2019
9	Obs17	On the existence of maximum likelihood estimates for presence-only data	Methods in Ecology and Evolution	Hefley & Hooten, 2015
10	Obs18	On extrapolating past the range of observed data when making statistical predictions in ecology	PLOS ONE	Conn et al., 2015
11	Obs19	Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization	Methods in Ecology and Evolution	Renner et al., 2019
12	Obs2	The landscape configuration of zoonotic transmission of Ebola virus disease in West and Central Africa: interaction between population density and vegetation cover	PeerJ	Walsh & Haseeb, 2015
13	Obs20	Mapping the risk of Nipah virus spillover into human populations in South and Southeast Asia	Transactions of The Royal Society of Tropical Medicine and Hygiene	Walsh, 2015
14	Obs22	Landscape-scale distribution of tree roosts of the northern long-eared bat in Mammoth Cave National Park, USA	Landscape Ecology	Thalken et al., 2018
15	Obs23	Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection	Methods in Ecology and Evolution	Koshkina et al., 2017
16	Obs24	Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions	Landscape Ecology	Fletcher et al., 2016
17	Obs25	Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology	The Annals of Applied Statistics	Warton & Shepherd, 2010
18	Obs26	Environmental heterogeneity blurs the signature of dispersal syndromes on spatial patterns of woody species in a moist tropical forest	PLOS ONE	Ramón et al., 2018
19	Obs27	Ecological drivers of tree assemblage in tropical, subtropical and subalpine forests	Journal of Vegetation Science	Asefa et al., 2020
20	Obs28	Ecological drivers of spatial community dissimilarity, species replacement and species nestedness across temperate forests	Global Ecology and Biogeography	Wang et al., 2018
21	Obs29	Drivers of forest regeneration patterns in drought prone mixed-species forests in the Northern Calcareous Alps	Forest Ecology and Management	Simon et al., 2019
22	Obs3	The impact of human population pressure on flying fox niches and the potential consequences for Hendra virus spillover	Scientific reports	Walsh et al., 2017
23	Obs30	Disentangling the functional trait correlates of spatial aggregation in tropical forest trees	Ecology	McFadden et al., 2019
24	Obs31	Determinants of spatial patterns of canopy tree species in a tropical evergreen forest in Gabon	Journal of Vegetation Science	Engone Obiang et al., 2019
25	Obs32	Detangling the effects of environmental filtering and dispersal limitation on aggregated distributions of tree and shrub species: life stage matters	PLOS ONE	Yang et al., 2016
26	Obs33	Comparative interpretation of count, presence–absence and point methods for species distribution models	Methods in Ecology and Evolution	Aarts et al., 2012
27	Obs34	Can the Spatial Point Patterns of Animal Distributions Be Detected Using Sparse Samples? A Case Study of Four Soricomorpha (Mammalia Species in Poland)	Zoologica Poloniae	Chen, 2014
28	Obs35	Bias correction in species distribution models: pooling survey and collection data for multiple species	Methods in ecology and evolution	Fithian et al., 2015
29	Obs37	Assessing the influence of environmental heterogeneity on bird spacing patterns: a case study with two raptors	Ecography	Cornulier & Bretagnolle, 2006

30	Obs38	Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway	Spatial Statistics	Sicacha-Parada et al., 2021
31	Obs4	The effects of habitat filtering and non-habitat processes on species spatial distribution vary across life stages	American journal of botany	Shi et al., 2018
32	Obs40	Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts	Ecological Monographs	Conn et al., 2015
33	Obs41	Ultra-fine scale spatially-integrated mapping of habitat and occupancy using structure-from-motion	PLOS ONE	McDowall & Lynch, 2017
34	Obs42	The relative influence of niche versus neutral processes on Ediacaran communities	bioRxiv	Mitchell et al., 2018
35	Obs43	The cost of postponing biodiversity conservation in Mexico	Biological Conservation	Fuller et al., 2007
36	Obs44	Tales from the underground: soil heterogeneity and not only aboveground plant interactions explain fine-scale species patterns in a Mediterranean dwarf-shrubland	Journal of Vegetation Science	Pescador et al., 2020
37	Obs48	Spatial associations of tree species in a subtropical evergreen broad-leaved forest	Journal of Plant Ecology	Luo et al., 2012
38	Obs49	Spatial associations among major tree species in a cool-temperate forest community under heterogeneous topography and canopy conditions	Population ecology	Torimaru et al., 2013
39	Obs5	The assembly and interactions of tree species in tropical forests based on spatial analysis	Ecosphere	Zhang et al., 2017
40	Obs50	Spatial analysis of a cat-borne disease reveals that soil pH and clay content are risk factors for sarcocystosis in sheep	Frontiers in veterinary science	Taggart et al., 2019
41	Obs51	Seasonal occurrence and abundance of dabbling ducks across the continental United States: Joint spatio-temporal modelling for the Genus Anas	Diversity and Distributions	Humphreys et al., 2019
42	Obs53	Reliable species distributions are obtainable with sparse, patchy and biased data by leveraging over species and data types	Methods in Ecology and Evolution	Peel et al., 2019
43	Obs57	Point pattern modelling for degraded presence-only data over large regions	Journal of the Royal Statistical Society: Series C (Applied Statistics	Chakraborty et al., 2011
44	Obs58	A Bayesian semiparametric GLMM for historical and newly collected presence-only data: An application to species richness of Ross Sea Mollusca	Environmetrics	Carota et al., 2017
45	Obs59	A hierarchical model for estimating the spatial distribution and abundance of animals detected by continuous-time recorders	PLOS ONE	Dorazio & Karanth, 2017
46	Obs6	Species–area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity	Ecology	Shen et al., 2009
47	Obs60	A practical guide for combining data to model species distributions	Ecology	Fletcher Jr et al., 2019
48	Obs61	A space-time point process model for analyzing and predicting case patterns of diarrheal disease in northwestern Ecuador	Spatial and spatio-temporal epidemiology	Ahn et al., 2014
49	Obs62	Analysis of Minnesota colon and rectum cancer point patterns with spatial and non-spatial covariate information	The Annals of Applied Statistics	Liang et al., 2009
50	Obs63	Analyzing spatial point patterns subject to measurement error	Bayesian Analysis	Chakraborty & Gelfand, 2010
51	Obs64	Careful prior specification avoids incautious inference for log-Gaussian Cox point processes	Journal of the Royal Statistical Society: Series C (Applied Statistics	Sørbye et al., 2019
52	Obs65	Climate change could increase the geographic extent of Hendra virus spillover risk	EcoHealth	Martin et al., 2018

53	Obs66	Discrete versus continuous domain models for disease mapping	Spatial and Spatio-temporal Epidemiology	Konstantinoudis et al., 2020
54	Obs67	Efficient Modelling of Presence-Only Species Data via Local Background Sampling	Journal of Agricultural, Biological and Environmental Statistics	Daniel et al., 2019
55	Obs68	Estimating Individual-Level Risk in Spatial Epidemiology Using Spatially Aggregated Information on the Population at Risk	Journal of the American Statistical Association	Diggle et al., 2010
56	Obs69	Estimating occupancy and abundance using aerial images with imperfect detection	Methods in Ecology and Evolution	Williams et al., 2017
57	Obs7	Spatial Statistics	NA	Bar-Hen et al., 2015
58	Obs70	Estimating seal pup production in the Greenland Sea by using Bayesian hierarchical modelling	Journal of the Royal Statistical Society: Series C (Applied Statistics	Jullum et al., 2020
59	Obs71	Evaluating citizen science data for forecasting species responses to national forest management	Ecology and evolution	Mair et al., 2017
60	Obs72	Going off grid: Computationally efficient inference for log-Gaussian Cox processes	Biometrika	Simpson et al., 2016
61	Obs73	Habitat use of toothed whales in a marine protected area based on point process models	Marine Ecology Progress Series	Silva et al., 2019
62	Obs74	Hen harrier Circus cyaneus nest sites on the Isle of Mull are associated with habitat mosaics and constrained by topography	Bird Study	Geary et al., 2018
63	Obs75	Hierarchical spatial modeling for estimation of population size	Environmental and Ecological Statistics	Barber & Gelfand, 2007
64	Obs76	Identifying priority conservation areas for a recovering brown bear population in Greece using citizen science data	Animal Conservation	Bonnet-Lebrun et al., 2019
65	Obs77	Importance of habitat heterogeneity and biotic processes in the spatial distribution of a riparian herb (Carex remota L.: a point process approach	Stochastic Environmental Research and Risk Assessment	Uria-Diez et al., 2013
66	Obs78	Improved spatial ecological sampling using open data and standardization: an example from malaria mosquito surveillance	Journal of the Royal Society Interface	Sedda et al., 2019
67	Obs79	Integrating distance sampling and presence-only data to estimate species abundance	Ecology	Farr et al., 2020
68	Obs8	Spatial patterns reveal negative density dependence and habitat associations in tropical trees	Ecology	Bagchi et al., 2011
69	Obs80	Key Community Assembly Processes Switch between Scales in Shaping Beta Diversity in Two Primary Forests, Southwest China	Forests	Asefa et al., 2020
70	Obs81	Spatial modelling of lupus incidence over 40 years with changes in census areas	Journal of the Royal Statistical Society: Series C (Applied Statistics	Li et al., 2012
71	Obs82	Log Gaussian Cox processes and spatially aggregated disease incidence data	Statistical methods in medical research	Li et al., 2012
72	Obs83	Mapping species richness using opportunistic samples: a case study on ground-floor bryophyte species richness in the Belgian province of Limburg	Scientific reports	Neyens et al., 2019
73	Obs84	Modelling and classification of species abundance: a case study in the Barro Colorado Island plot	Journal of Applied Statistics	Jalilian, 2017
74	Obs85	Modelling the abundance of an endangered medicinal species, Phellodendron amurense: generalised linear model vs. generalised additive model.	Botanica Serbica	Naiqi & Zhang, 2018
75	Obs86	A spatial pattern analysis of the halophytic species distribution in an arid coastal environment	Environmental Monitoring and Assessment	Bana & Goossens, 2015

76	Obs87	On Consistent Nonparametric Intensity Estimation for Inhomogeneous Spatial Point Processes	Journal of the American Statistical Association	Guan, 2008
77	Obs88	Pairwise interaction point processes for modelling bivariate spatial point patterns in the presence of interaction uncertainty	Journal of Environmental Statistics	Nightingale et al., 2015
78	Obs89	Penalized composite likelihoods for inhomogeneous Gibbs point process models	Computational Statistics & Data Analysis	Daniel et al., 2018
79	Obs9	Spatial patterns of tree species richness in two temperate forests	Journal of Ecology	Wang et al., 2011
80	Obs90	Phylogenetic and functional diversity area relationships in two temperate forests	Ecography	Wang et al., 2013
81	Obs91	Predicting the geographical distribution of two invasive termite species from occurrence data	Environmental Entomology	Tonini et al., 2014
82	Obs92	Regularized estimating equations for model selection of clustered spatial point processes	Statistica Sinica	Thurman et al., 2015
83	Obs93	Response of individual sizes and spatial patterns of Deyeuxia angustifolia to increasing water level gradient in a freshwater wetland	Environmental Science and Pollution Research	Ren et al., 2020
84	Obs94	Seal encounters at sea: A contemporary spatial approach using R-INLA	Ecological modelling	Carson & Flemming, 2014
85	Obs95	Use of opportunistic sightings and expert knowledge to predict and compare Whooping Crane stopover habitat	Conservation Biology	Hefley et al., 2015
86	Obs96	Using INLA to fit a complex point process model with temporally varying effects-a case study	Journal of Environmental Statistics	Illian et al., 2012
87	Obs97	Variable selection for spatial Poisson point processes via a regularization method	Statistical Methodology	Thurman & Zhu, 2014
88	Obs98	Is my sdm good enough? insights from a citizen science dataset in a point process modeling framework	Ecological Modelling	Leandro et al., 2020

APPENDIX B

Key Elements Used in the Practice of the Inhomogeneous Poisson Process Modeling

Statistical model	**Inhomogeneous point process models or R functions used**	**Parameters estimation method**	**Model validation**	**Package**
Area interaction	Area-interaction model	Likelihood maximization, Combined penalized, Maximum pseudolikelihood, MCMC, Penalized Composite likelihood, Regularized Composite likelihood	AUC, TSS, Residual Analysis, Envelopes from 99 Monte Carlo Simulations	Ppmlasso, Spatstat, Maxnet
Geyer point process	Geyer point process	Maximum likelihood, Maximum profile pseudo-likelihood, Maximum pseudo-likelihood	Smoothed partial residual diagnostics	Spatstat
Hard-core point process	Ppm function, Hybrid model	Maximum likelihood	Smoothed partial residual diagnostics	Spatstat
Neyman–Scott process	Log-intensity Inhomogeneous Thomas process			Spatstat

Inhomogeneous Poisson process	Log-linear regression (ppm function, GAM (gam function, mgcv function, weighted logistic regression, Inhomogeneous Poisson process model, Integrated model, GLM, joint log-linear IPP, Pooled data model, Point process model, Semiparametric Bayesian GLMM, Spatial Capture-recapture (SCR model, Inhomogeneous Poisson cess model, Hierarchical Bayesian model, Regularized IPP model, PLCL-IPP model, Marked IPP model boosted	Poisson maximum pseudo-likelihood, Penalized Composite likelihood, Regularized Composite Likelihood, MCMC, Regularized (penalized quasi-log-likelihood, Boosted regression trees, Maximum log-likelihood)	Methods proposed by Jara-Guerrero et al. (2015), Residual Analysis, AIC, BIC, AUC, simulation study, Kappa, BIC, Sensitivity, AIC	Spatstat, Sce, Ade4, Bayesglm, Glm, Logistf, Spat-Pred, Mgcv, MultispeciesPP
IPP + Binomial	Integrated Model, N-mixture model, IPP model (ppm function, pooled data model, multi species PP)	Joint likelihood, MCMC	Simulation study	R codes
IPP + Thinning function	Integrated model, IPP model	Maximum joint log-likelihood, Likelihood maximization	AUC, TSS, kappa statistic, AIC, block validated log-likelihood	MultispeciesPP
IPP +Tf+ Bin	Integrated SDM	Weighted joint likelihood	AUC statistic and block validated log-likelihoods	
Inhomogeneous Thomas process	Log-linear regression, Integrated model, Logintensity Inhomogeneous Thomas process, inhomogeneous Thomas process model	Maximum likelihood, Minimum contrast	Methods proposed by Jara-Guerrero et al. (2015), AIC, 99 simulations envelopes, Confidence interval, AIC	Spatstat, Sce, Selectspm, Scripts from Shen et al. (2009)
LGCP	Log-linear model, Hierarchical GLMs and Spatio-temporal regression model, Naive model, VSE model, point process model, Log Gaussian Cox process model, GAMpo model, Bayesian hierarchical spatio-temporal model	The two-step approach of Waagepetersen and Guan (2009), MCMC, INLA & SPDE, INLA	Pair-Correlation function, Loosmore's goodness of fit, diagnostic plots, cross-validation scheme, AIC, DIC, Simulation study, Average Relative Squared Prediction Error (RSPE, DIC, Partial ROC test, Sensitivity, Specificity, AUC, WAIC, PML	Spatstat, Randomfield, SpatPred, RINLA library, Stabundance, Lgcp, PrevMap
LGCP + Binomial	Joint model		Spearman rank correlation, sensitivity, specificity, AUC, TSS	INLA-package
Strauss point process				Spatstat

	Ppm function, Hybrid model	Maximum profile pseudolikelihood, Maximum likelihood	Smoothened partial residual diagnostics, AIC	
Thinned Poisson process	Joint log-linear inhomogeneous Poisson process model		Fold block cross-validation, AUC, predictive log-likelihood	MultispeciesPP