

# Inhomogeneous Poisson Point Process for Species Distribution Models: Relative performance of methods accounting for sampling bias and imperfect detection

Yannick MUGUMAARHAHAMA <sup>\*1,2</sup>, Adandé Belarmain FANDOHAN<sup>1,3</sup>, Arsène Ciza MUSHAGALUSA<sup>1,2</sup>, Idelphonse Akoeugnigan SODE<sup>1</sup> and Romain GLELE KAKAÏ<sup>1</sup>

<sup>1</sup>Laboratoire de Biomathématiques et d'Estimations Forestières, Faculty of Agronomic Sciences, University of Abomey-Calavi, 01 PO. Box: 526 Cotonou, Benin

<sup>2</sup>Unité de Biométrie, Faculté des Sciences Agronomiques et Environnement, Université Evangélique en Afrique, PO. Box: 3323 Bukavu, Democratic Republic of Congo

<sup>3</sup>Unité de Recherche en Foresterie, Agroforesterie et Biogéographie, Ecole de Foresterie Tropicale, Université Nationale d'Agriculture, PO. Box : 45, Kétou, Benin

April 12, 2021

## Abstract

Species distribution models have become tools of great importance in ecology since the advanced knowledge of suitable habitat of species is needed in the process of the world's biodiversity conservation. Models that use presence-only data are of great interests and are widely used in ecology due to their easy access. However, these models do not estimate accurately the true spatial species distribution based solely on presence-only data since they do not account for biases induced by the sampling techniques used and imperfect detection. To address this gap, Hierarchical integrated models have been recently introduced. Through this study, we assessed the relative performance of these new SDMs models using simulated data. The performance of the models was tested by comparing the estimates of parameters of the distribution models they provide with parameters used to simulate the distribution of the virtual species. The best model was the one whose estimates were close to the true distribution parameters of the virtual species. Results showed that analyzing Presence-only data in conjunction with Point-counts data through the Dorazio's Hierarchical model produced estimates of the coefficients of the species intensity models with high precision and less bias while the Koshkina integrated model showed poor performance. Site-occupancy data, being not informative of species abundance, did not allow reducing biases in Presence-only data. The Dorazio's Hierarchical model produced estimates with high precision even with low detection probability. We have also found that the species rarity tends to inflate the variability of the models' estimates making modelling abundant species to be more accurate than modelling less abundant species. Hence, to model the species distribution with high precision based on Presence-only data, additional Point-counts data are required to account for sampling bias and imperfect detection.

**Key words :** Integrated Species distribution models, Maximum likelihood, Point-counts, Presence-only data, Site-Occupancy, species abundance

\*Corresponding author : [lesmas2020@gmail.com](mailto:lesmas2020@gmail.com)

# 1 Introduction

Species distribution models (SDMs) have become important tools in ecology to inform conservation actions in the context of climate change. They are used to identify suitable habitats from threatened, harmful invasive or endangered species or those of special concern or to assess the impact of climate change or land use on their future distribution (Crall et al., 2013; De Siqueira et al., 2009; Fei et al., 2012; Fuller et al., 2008; Guillera-Arroita et al., 2014; Hefley et al., 2017; Kearney et al., 2010; Koshkina et al., 2017; Kremen et al., 2008; Li and Wang, 2013). The main purpose of SDMs is to plan conservation activities such as habitat restoration, bio-monitoring or creation of reserve networks. Accurate predictive SDMs are of particular importance for effective and adaptive management and conservation activities or for ecological research in general. Several modeling approaches have been developed including presence only SDMs.

Presence-only SDMs were developed to use data where knowledge of absences is inadequate or unavailable (Brotons et al., 2004; Guisan and Thuiller, 2005a,b; Peterson et al., 2011). The data collection techniques are unknown for most of Presence-only data resulting in an uncertainty in SDMs predictions (Barry and Elith, 2006; Elith et al., 2002). Furthermore, despite the numerous attempts made by ecologists, it is not possible to accurately predict the true spatial species distribution of the species of interest based solely on presence-only data (Fithian and Hastie, 2013; Hastie and Fithian, 2013). The uncertainty associated with ecological data (mostly collected from websites) is a great challenge in species distribution modeling. It must be accounted for when results are to be appropriately interpreted or if they are the basis of a decision-making process (Barry and Elith, 2006; Elith et al., 2002).

Among factors affecting SDMs accuracy, sampling bias and imperfect detection are receiving increasing attention. Sampling bias involves collecting occurrence records of the species of interest in such a way that some locations have a lower probability of being visited than others. Imperfect detection is the failure to detect an individual of that species in a given location, even if it is present. Correcting sampling bias effects can increase accuracy of Presence-only SDMs otherwise the model can reflect sampling effort rather than the true distribution of a species (Phillips et al., 2009). In addition, imperfect detection can substantially prevaricate the maximum likelihood estimates of Presence-only SDMs' parameters (Dorazio, 2012). The assumption of perfect detection prevaricates parameters estimation and weakens statistical inference. Therefore, this situation could lead to a misunderstanding of the issue of interest and hence conclusions for policy making could be misleading (Kellner and Swihart, 2014). Yet, it is unclear to which extent, accounting for sampling bias and imperfect detection in the specification of these models would improve accuracy of presence only SDMs (Koshkina et al., 2017).

To address this challenge, a hierarchical species distribution model with presence only data together with replicated point counts was firstly proposed (Dorazio, 2014). Recently Koshkina et al. (2017) introduced an integrated species distribution model that combines Presence-only data with repeated survey site-occupancy records. These models account for sampling bias and imperfect detection simultaneously by using a thinned Poisson point-process model for presence only data after Dorazio (2014). They also incorporate either an N-mixture model for replicated point counts (Royle, 2004) or a revised version of the conventional site-occupancy (SO) model for repeated surveys SO data (MacKenzie et al., 2002). Yet, SDMs performances may also be affected by the type of species, considering its spatial patterning in geographic and environmental spaces (i.e. rare versus abundant; large distribution versus short range; different combination of these two aspects). For instance, how will SDMs performance be affected depending of the following situations? The

species is abundant and largely distributed vs abundant and short range; it is rare and largely distributed vs rare and short range.

Effect of species rarity and range on models' performances has become an issue of interest only in the recent SDM literature. In this study, we examine effect of the type of species (rare versus abundant) on the performance of presence only SDMs when accounting for sampling bias and imperfect detection. To our knowledge, no previous study has already compared the predictive performance of these recently proposed models in the situation of rare vs abundant species.

## 2 Methods

### 2.1 Data generation process

For the data generation process, the simulation design was similar to that described in Dorazio (2014) and Koshkina et al. (2017). For this purpose, let us denote  $B$ , a square area divided into  $1000 \times 1000$  grid cells representing the study area where occurs our virtual species. Let us also denote  $x(s)$  and  $w(s)$ , two environmental predictors (covariates) generated using bivariate distributions that vary spatially and assumed to be independent of each other. These two predictors  $x(s)$  and  $w(s)$  were generated so that they were defined at every point  $s$  on the 2D grid of  $B$  (see Dorazio, 2014; Koshkina et al., 2017).

As in Dorazio (2014) and Koshkina et al. (2017), we simulated the species intensity  $\lambda(s)$  using the log-linear function that depends on the single covariate  $x(s)$ :

$$\log(\lambda(s)) = \beta_0 + \beta_1 x(s)$$

The detection probability  $b(s)$  was simulated using the logit function that depends on the single covariate  $w(s)$ :

$$\text{logit}(b(s)) = \alpha_0 + \alpha_1 w(s)$$

#### *Simulation of the abundant species*

We considered  $\beta_1 = 0.5$  and  $\beta_0 = \log(8000) \approx 8.9872$  for what we considered as an abundant species under varying detection probability  $b(s)$ . We considered  $\alpha_1 = -1$  and  $\alpha_0$  was assigned values ranging from -5 to 5 so that detection probabilities at the average value of  $w(s)$  ranged from very low values ( $b(s) \approx 0$ ) to very high values ( $b(s) \approx 1$ ). The simulation of the abundant species was done in such a way as to have presence-only data that could go beyond 10000 occurrence records per cell when assuming perfect detection and decreasing as the probability of detection decreases without going below 100 records per cell when the detection probability is close to 0.

#### *Simulation of the less abundant species*

We considered  $\beta_1 = 0.5$  and  $\beta_0 = \log(200) \approx 5.2983$  for what we considered as a less abundant species under varying detection probability  $b(s)$  as we did for the abundant species. For this scenario, the simulations was done in such a way as to have presence-only data that could not go beyond 300 occurrence records per cell when assuming perfect detection and decreasing as the probability of detection decreases with the possibility of going below 5 records per cell when the detection probability is close to 0.

For both species abundance scenarios, the point patterns that represent the true presences were simulated with the intensity function  $\lambda(s)$  but thinned by the detection probability  $b(s)$ . Also,

the data were assumed to be subject to an extra sampling bias that cannot be accounted for by covariate  $w(s)$ . To achieve this, we did like Koshkina et al. (2017), by choosing randomly 30 % of the total number of true detected presences and considered them as the observed presences. This ensures that we were working with true presence-background data instead of a presence-absence type of data.

To simulate the Point-counts data and site-occupancy data, we divided the study area B into 100 x 100 square quadrats of equal size, each quadrat being constituted by 100 grid cells of B. Different samples of sizes  $Z = 50, 100, 200, 400$  or 800 quadrats were randomly selected throughout the whole region B. For each quadrat, the corresponding values of intensity  $\lambda(s)$  or true number of individuals present in it was calculated by summing the intensities corresponding to grid cells that fall in the considered quadrat. For each quadrat, the corresponding values of covariates  $x(s)$  and  $w(s)$  were calculated using the mean values of  $x(s)$  and  $w(s)$  from the aggregated areas. Simulated point-counts and presence-absence (site-occupancy status) were obtained by conducting J independent binomial draws from individuals of each quadrat (for more details, see Dorazio (2014) for point-counts data and Koshkina et al. (2017) for site-occupancy data).

## 2.2 Simulation design and data analysis

In all experiments, 2000 replications of the simulated data, each one containing presence-only, site-occupancy and point-counts data were generated. Maximum likelihood estimates of the predictor's weight  $\beta_1$  and the intercept  $\beta_0$  were computed for each simulated dataset by fitting the model of presence-only data, the Koshkina Integrated model (the model with combined data of presence only observations and Site-Occupancy data), and the Dorazio model (the model with combined data of presence only observations and point counts). All models were fitted using the maximum likelihood method, in R (version 3.5.1) (R Development Core Team, 2018).

We estimated parameters ( $\beta_0$  and  $\beta_1$ ) from the virtual species distribution (intensity) based on the data sets containing records of points selected randomly over the study area B and the associated covariate. Hence, we obtained parameters' estimates ( $\beta_0$  and  $\beta_1$ ) for each model that were compared to the simulated values of ( $\beta_0$  and  $\beta_1$ ) that were used to simulate the virtual species distribution (intensity).

The three models were compared by testing the effects of accounting for sampling bias and imperfect detection, the number of sites in planned surveys and the detection probability on model's performance based on  $\beta$  estimates. To assess the performance of models, operating characteristics of the estimators (bias and variance) were used.

## 3 Results

### 3.1 Comparison of $\beta_0$ and $\beta_1$ estimates

The student t test was used to test the conformity of the  $\beta$  estimates while the Levene test was used to test the effect of species abundance on their variance. The results obtained are presented in table 1.

Table 1 illustrates how the bias evolves in estimating parameters for the different models under study. The results showed that the estimates of parameters  $\beta_0$  and  $\beta_1$  have high variance in case of less abundant species while they become low in case of abundant species. The increase in species abundance decreases the variance of the estimates. Estimates of parameter  $\beta_0$  are less precise

Table 1: Maximum-likelihood estimates of parameters  $\beta_0$  and  $\beta_1$  obtained by fitting different models.

Parameter	Model	Species abundance	Real	Estimate	SD	t test p-value
$\beta_0$	PO	Low	5.298	4.842	1.866	0.000***
		High	8.987	7.718	1.193	0.000***
		Levene test p-value			0.000***	
	KIM	Low	5.298	0.434	3.13	0.000***
		High	8.987	7.471	1.66	0.000***
		Levene test p-value			0.000***	
	DHM	Low	5.298	5.286	0.153	0.111
		High	8.987	8.987	0.027	0.93
		Levene test p-value			0.000***	
$\beta_1$	PO	Low	0.5	0.521	0.222	0.025*
		High	0.5	0.498	0.051	0.06
		Levene test p-value			0.000***	
	KIM	Low	0.5	0.508	0.126	0.361
		High	0.5	0.552	0.217	0.347
		Levene test p-value			0.09368	
	DHM	Low	0.5	0.504	0.106	0.342
		High	0.5	0.5	0.017	0.401
		Levene test p-value			0.000***	

for Presence-only and Koshkina integrated model. These models underestimate  $\beta_0$  while Dorazio model has shown to be superior in estimating  $\beta_0$ . All models, except Presence-only model, appear to estimate parameter  $\beta_1$  accurately. Based on these results, we conclude that the combination of Presence-only data and Site-occupancy data does not significantly increase the accuracy of species distribution models under any species abundance scenario. Dorazio's Hierarchical SDM has shown the best performance in the estimation of parameters  $\beta_0$  and  $\beta_1$  for both, rare and abundant species when compared to other models. Accounting for sampling bias and imperfect detection by using Koshkina Integrated model appears hence to not be meaningful at all. Site-occupancy surveys are not hence good ways to improve SDM accuracy. Point-count data appear to be helpful for that purpose.

### 3.2 Effects of detection probability on operating characteristics of maximum likelihood estimates

As mentioned above, the decrease in relative abundance leads to inflation of the variance of  $\beta_0$  estimates. The  $\beta_0$  estimates of the Presence-only model and Koshkina Integrated model are less precise and are also found to be sensitive to the variation in the detection probability (Fig.1 and Fig.2). As the probability of detection decreases, the bias and variance of the estimators in  $\beta_0$  increases. Bias of  $\beta_0$  due to low probabilities of detection is thus higher for less abundant species than for abundant species. In both scenarios of species abundance (Less abundant species versus abundant species), the Dorazio Hierarchical model differs from the two models in that it is very insensitive to changes in the probability of detection. For this model, the bias of the  $\beta_0$  estimates is almost zero but only its variance is slightly altered if the species is rare, regardless of the detection probability (Fig.1c, Fig.1f, Fig.2c and Fig.2f). As for the other models,  $\beta_1$  is not sensitive either to variation in detection probabilities or to changes in the abundance of the species of interest.

For the three models, we noted that the variation in detection probability has no effect on the

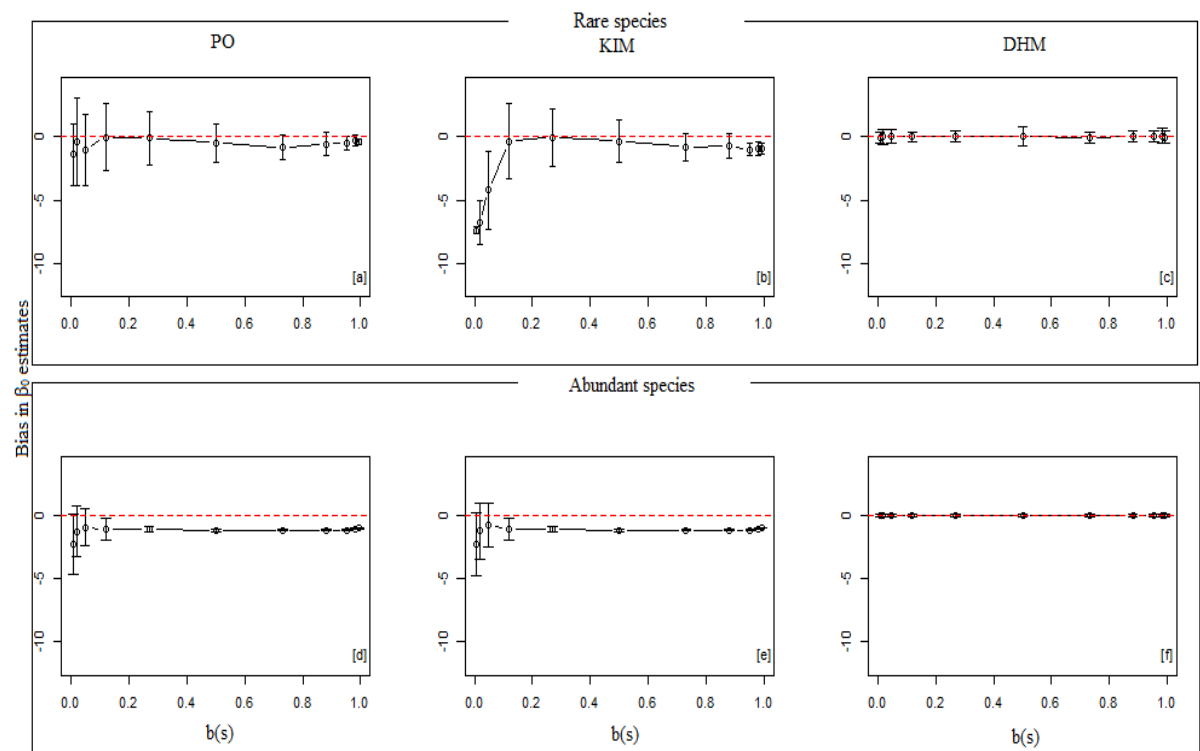


Figure 1: **Bias of maximum-likelihood estimates of the model parameters  $\beta_0$  with varying detection probability.** The top row shows results for the less abundant species while the bottom row shows results for the abundant species. The bar lines represent the standard deviation of the bias. The red line indicates bias equal zero

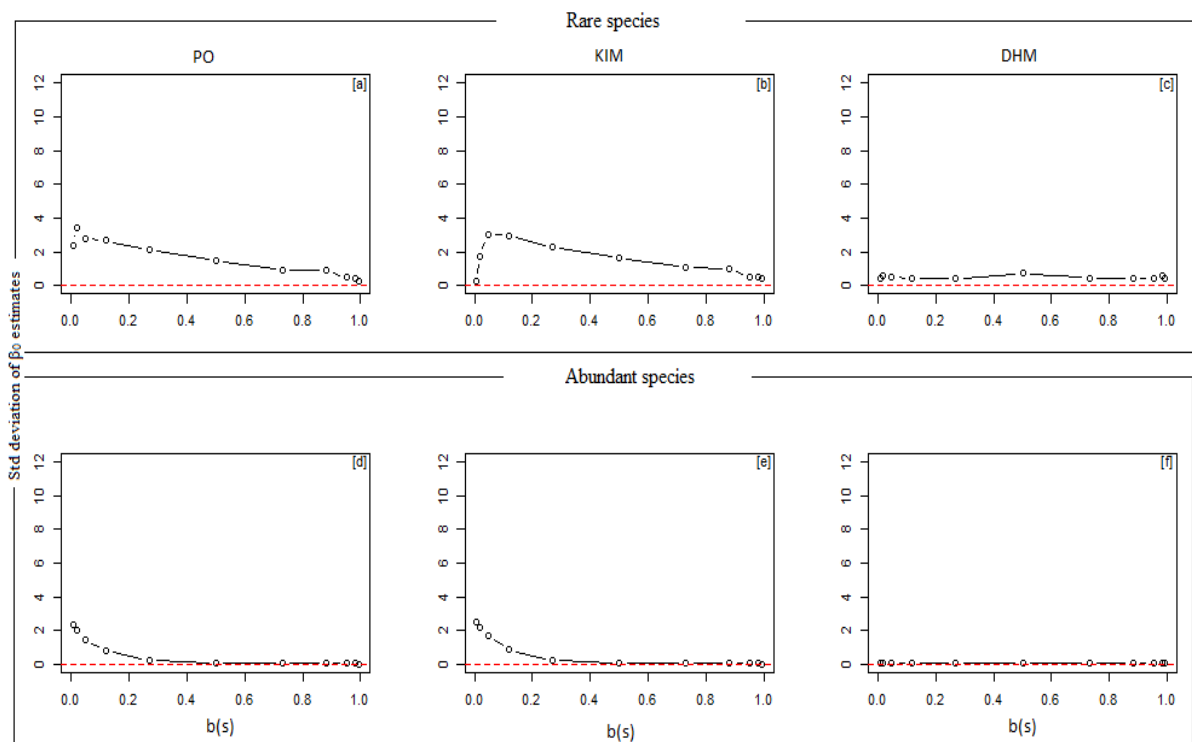


Figure 2: **Standard deviation of maximum-likelihood estimates of the model parameters  $\beta_0$  with varying detection probability.** The top row shows results for the less abundant species while the bottom row shows results for the abundant species. The red line indicates standard deviation equal zero

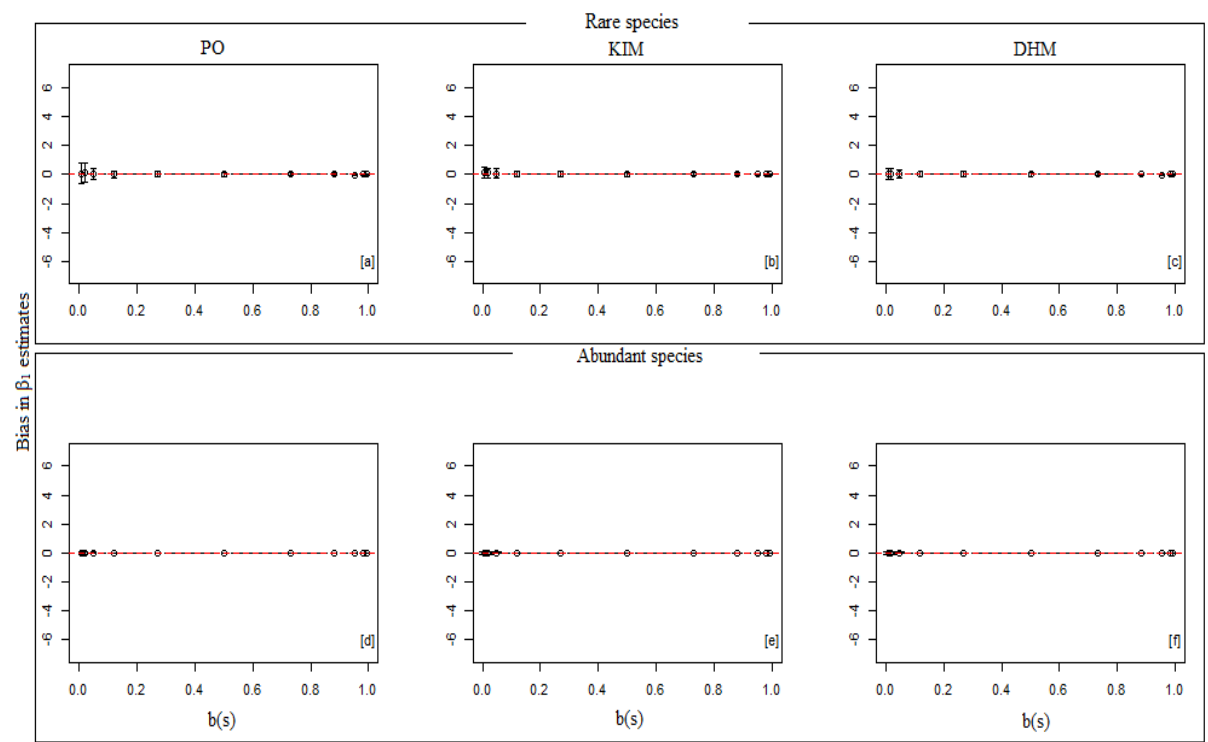


Figure 3: **Bias of maximum-likelihood estimates of the model parameters  $\beta_1$  with varying detection probability.** The top row shows results for the less abundant species while the bottom row shows results for the abundant species. The bar lines represent the standard deviation of the bias. The red line indicates bias equal zero.

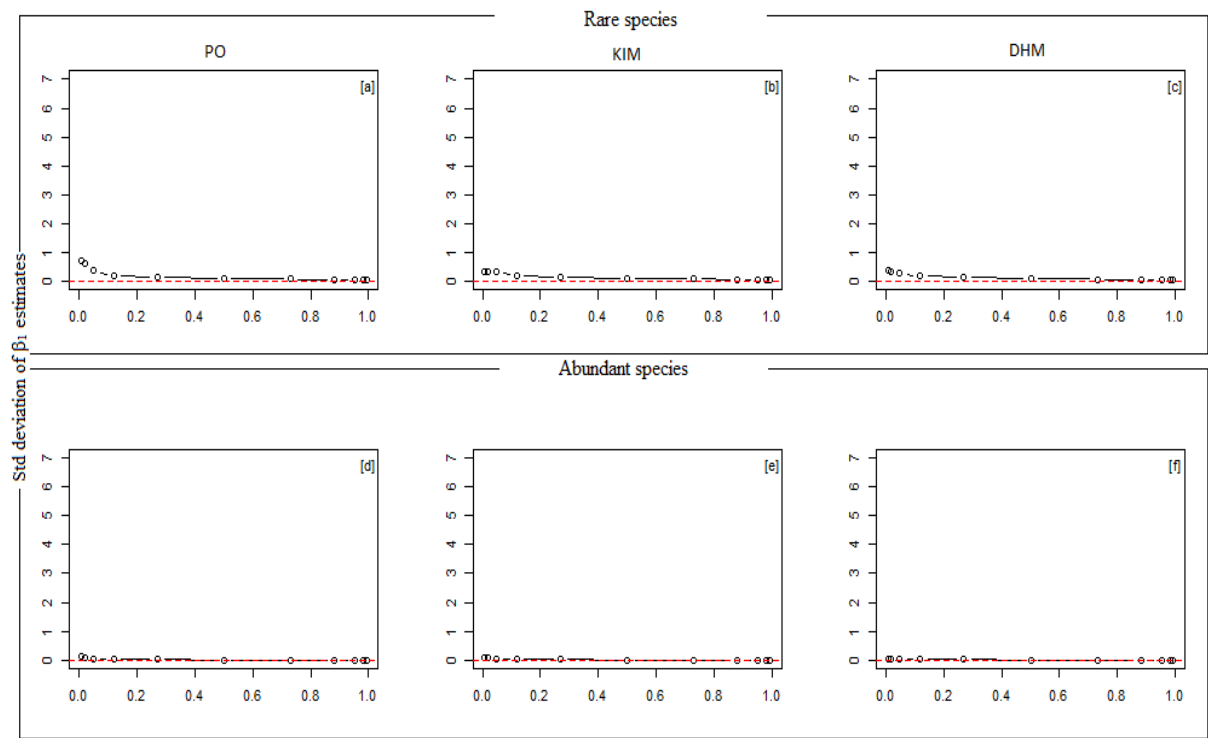


Figure 4: **Standard deviation of maximum-likelihood estimates of the model parameters  $\beta_1$  with varying detection probability.** The top row shows results for the less abundant species while the bottom row shows results for the abundant species. The red line indicates standard deviation equal zero



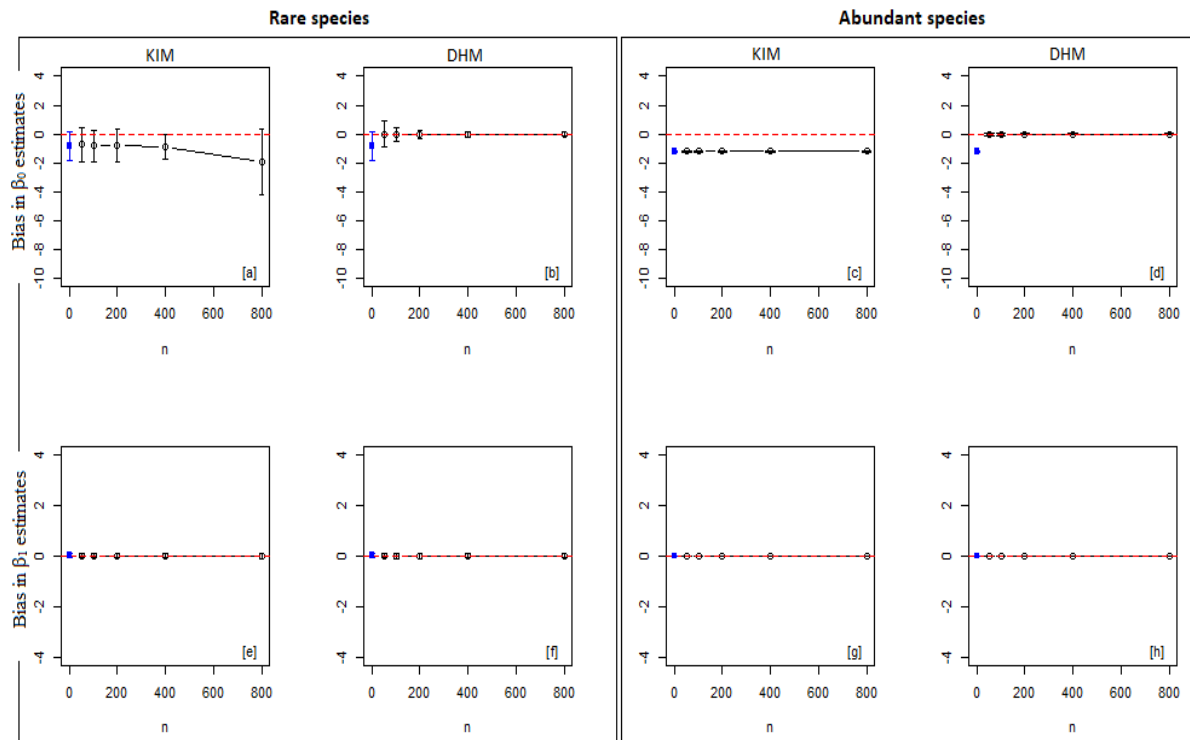


Figure 5: Mean bias of maximum-likelihood estimates of the model parameters  $\beta_0$  and  $\beta_1$  with varying numbers of Site-Occupancy or Point-counts sites. The top row shows results for the  $\beta_0$  while the bottom row shows results for  $\beta_1$ . The red line indicates bias equal zero. The bar lines represent the standard deviation of the bias. The blue square represents the mean bias of maximum-likelihood estimates obtained with the Presence-only model.

precision (variance) of the  $\beta_1$  estimate when modelling spatial distribution of an abundant species (Fig.3d-3f and Figd-4f). For all the three models, regardless of the detection probability, the variance of the  $\beta_1$  estimate is almost zero. On the other hand, when there is a less abundant species, the precision decreases when the detection probability becomes less than 0.2 (Fig.3a-3c and Fig.4a-4c).

### 3.3 Effects of number of sampled sites on operating characteristics of maximum likelihood estimates of $\beta_0$ and $\beta_1$

Apart from the effect of species abundance on the variance of  $\beta_0$  estimates, the Koshkina Integrated model is less precise regardless of the number of surveyed sites (Fig.5c). On the other hand, the Dorazio Hierarchical model has shown to be superior with a minimum number of sites (50 sites) sampled in case of an abundant species. For a less abundant species, to achieve good accuracy (near-zero variance) of the  $\beta_0$  estimates, 200 or more surveyed sites are required (Fig.6b).

In case of abundant species, all models estimated  $\beta_0$  with almost zero variance. This is not the case for the less abundant species for which it is necessary to increase the number of surveyed sites during planned surveys to reduce the variance of the  $\beta_0$  estimated by Dorazio model (Fig.6b). For the Koshkina Integrated Model, the variance of  $\beta_0$  estimates tends to be slightly erratic with varying numbers of Site-occupancy sites for the less abundant species while it appears constant for the abundant species. Furthermore, it appears to decrease with the increase of the species abundance (see Fig.6a and Fig.6c). Indeed, the rarer the species, the greater the variance of  $\beta_0$  estimates and



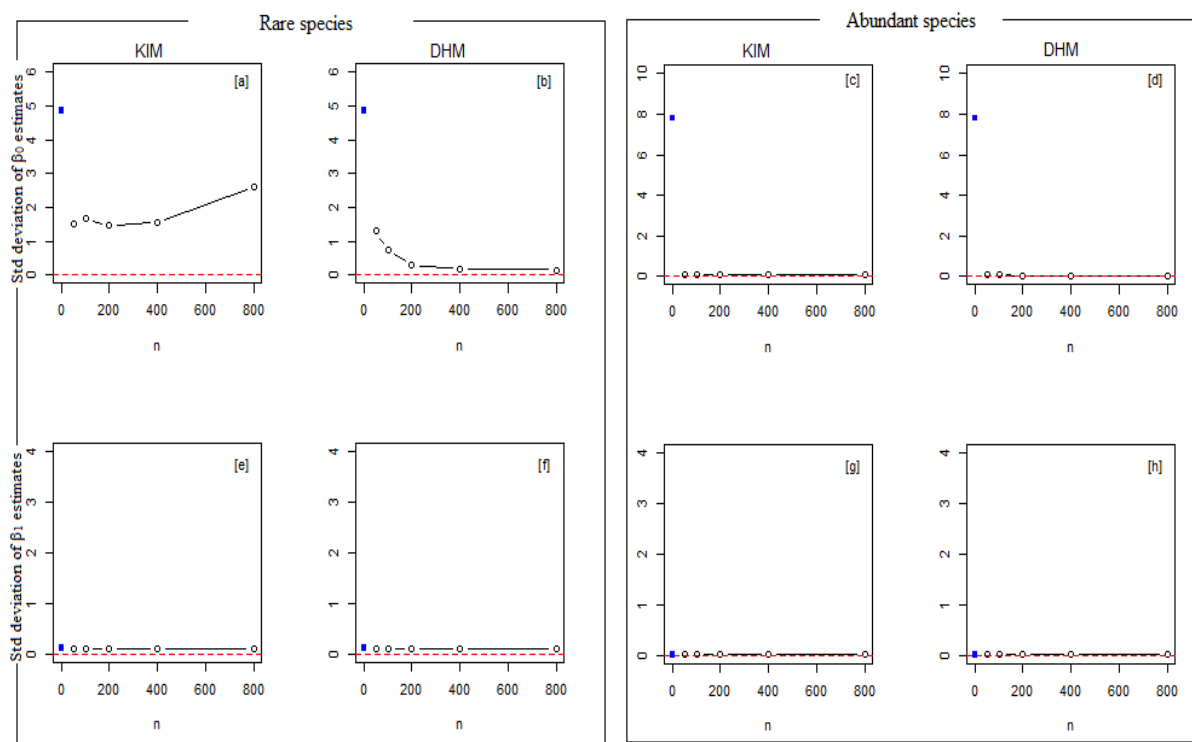


Figure 6: **Standard deviation of maximum-likelihood estimates of the model parameters  $\beta_0$  and  $\beta_1$  with varying numbers of Site-Occupancy or Point-counts sites.** The top row shows results for the  $\beta_0$  while the bottom row shows results for  $\beta_1$ . The red line indicates bias equal zero. The blue square represents the standard deviation of maximum-likelihood estimates obtained with the Presence-only model.

vice versa.

As for  $\beta_0$ , the rarity of the species of interest seems to have a leverage effect on the variance of  $\beta_1$  estimates. The Presence-only model, the Koshkina Integrated model and the Dorazio Hierarchical model estimate  $\beta_1$  with almost zero variance and almost constant as for the estimation of  $\beta_0$  (Fig.6h). For the Koshkina Integrated model and the Dorazio Hierarchical model, the variance of the  $\beta_1$  estimates for the two scenarios of abundance of the species of interest is negligible (almost zero) and does not change regardless of the number of surveyed sites (Fig.6e-6h).

## 4 Discussion

### 4.1 Performance in accounting for imperfect detection and sampling bias

Modelling species occurrence without accounting for bias induced by imperfect detection and/or sampling bias could lead to capture only sampling effort. In this circumstance, it is difficult to distinguish between predictions that reliably reflect ecological processes that shape the spatial distribution of the species and those that are related to detectability effects or sampling effect (Guillera-Arroita, 2017; Lahoz-Monfort et al., 2014).

Presence-only model and the Koshkina Integrated model were unable to estimate accurately  $\beta_0$  (the intercept). The two models seem to behave in the same way. With poor precision in  $\beta_0$  estimates, it is then clear that analyzing Presence-only data in conjunction with the Site-occupancy data does not correctly account for sampling bias and imperfect detection. On the other hand, the Dorazio Hierarchical model has shown good performance by producing highly precise estimates of  $\beta_0$  with less bias. All models have shown good performance in estimating  $\beta_1$  parameter (slope). It appears then that analyzing Presence-only data in conjunction with Points-count data improve significantly the precision of Presence-only models and appears to be the best way to account for sampling bias and imperfect detection to which these data are subject. The performance of Dorazio Hierarchical model in estimating  $\beta_0$  is inherited from Point-count model while its performance in estimating  $\beta_1$  is inherited from the Presence-only model. The performance of Dorazio Hierarchical model we have observed can be explained by the fact that Point-counts model is known to produce accurate estimates of abundance and detectability in circumstances where populations are assumed truly closed (Royle, 2004). Another reason can be that the Point-counts model are known to be more versatile in the sense that it can be used even in populations where site-specific abundances may be high and where estimates of site occupancy equal one (Dodd and Dorazio, 2004). The Point-counts model has several advantages among which its ability to be extended to allow heterogeneity in the detection probability per individual between sample locations independently of the variation in site specific abundances (Royle and Dorazio, 2006).

Surprisingly, our results do not corroborate those of Koshkina et al. (2017) while the simulation plan was almost identical to ours. Also, they do not corroborate those of MacKenzie et al. (2002) and Mackenzie and Royle (2005) who had shown that the Site-Occupancy model can correct imperfect detection due to false absences (that is, failure to detect a species present on a given site) provided that the planned surveys are conducted in such a way that it is possible to estimate the detectability. One factor could explain the bad performance of the Koshkina integrated model: The lack of information on the local abundance of the species of interest. Dorazio (2007) showed that the estimates of population site occupancy and abundance based on the Point-counts and Site-occupancy models are technically equivalent, but since point-counts data potentially contain more information (relative to species abundance) than Site-occupancy data, the Dorazio Hierarchical

model may produce estimates that are more accurate than those produced by Koshkina Integrated model. Hence, the major weakness of the Site-Occupancy data is that they are not informative of the abundance of the species of interest. That could be why the Koshkina Integrated models did not perform well. In addition, Jiménez-Valverde et al. (2009), Nielsen et al. (2005), and Pearce et al. (2001) and other empirical studies have demonstrated a lack of correlation between occupancy and local abundance. In these studies, it is demonstrated that imposing a presence-absence design to point-counts design leads to loss of information except when each spatial unit contains only one observation. In the case where each spatial unit contains at least one observation, the use of Site-occupancy data leads to a loss of the ability to investigate the influence of environmental conditions on the species distribution (Aarts et al., 2012).

Dorazio (2014) showed that if abundance within sample units is not too high, Site-Occupancy survey can be used to reduce or eliminate the bias induced by imperfect detection and sampling bias. This statement supposes that Koshkina integrated model could produce relatively accurate estimates if the species of interest is rare or rare. Furthermore, Comte and Grenouillet (2013) found that Site-occupancy model provided better accuracy than models that do not account for imperfect detectability and demonstrated that correcting bias induced by imperfect detection improved species distributions models performance. Their results demonstrated that Site-Occupancy model is an attractive way to predict the distribution of poorly detectable species but it does not always lead to a substantial improvement over conventional model predictions.

Conduct planned surveys is expensive and biologist interested in modelling species distribution are not always fortunate enough to collect Point-counts data in planned surveys. Then, the Dorazio Hierarchical model offers an alternative to model spatial distribution of species with high precision by combining opportunistic data (Presence-only data) and Point-counts data obtained during planned surveys. The advantage offered by Dorazio Hierarchical model is that it does not necessarily require a large number of point-count observations to collect during repeated planned surveys.

## 4.2 Effect of variation in detection probability and species rarity on the models' precision

Detection probability is intrinsically heterogeneous among individuals of a species of interest. Such intrinsic heterogeneity could be manifest as inherent differences among individuals in levels of activity or movement, crypsis (the ability of an animal to avoid observation or detection) and perhaps even body size (Cutrera et al., 2006; Karlsson et al., 2008; Karpestam et al., 2014; Storfer et al., 1999). The Presence-only model and Koshkina Integrated model have shown to be relatively sensitive to the variation in detection probability during opportunistic surveys while the Dorazio hierarchical model has shown to be less sensitive. When doing simulations, we assumed that the detection probability at each surveyed site was constant during repeated planned surveys and we did not then assess the effect of variation in detection probability during planned surveys to know how the Dorazio Hierarchical model and Koshkina Integrated model perform when varying detection probability in that circumstance. Indeed, the concern of this study was the imperfect detection to which Presence-only data are subject. Nevertheless, Dorazio Hierarchical model and Koshkina Integrated model are assumed to explicitly account for species detectability (MacKenzie et al., 2002; Tyre et al., 2003). The two models have two components, one of which is the spatial point process model and the other is the point-counts model or the site-occupancy model. It is the second component of these models that is intended to correct the bias in the  $\beta_0$  estimates, the relative abundance of the species of interest. By comparing the performance of the PO model and the Dorazio Hierarchical model as probability detection varies, we can see that the Point-counts model component makes the Dorazio Hierarchical model very insensitive to the variation in prob-

ability detection. The component relating to Point-counts data can thus be considered as accurate in estimating the abundance of the species of interest.

Kery et al. (2005) and Zylstra et al. (2010) found that the Point-counts models provided greater abundance estimates than those obtained by other methods, that is, conventional territory mapping for different bird species (Kery et al., 2005) and distance sampling for desert tortoise (Zylstra et al., 2010). Furthermore, Veech et al. (2016) found that in case of homogeneity in individual detection, Point-counts model generally perform well with relatively little bias. Royle (2004) obtained similar result with constant detection probabilities of individuals. He found that if detection probability was high, then the Point-counts models tended to be very accurate; if detection probability was low, the average difference between estimated and known abundance could be as great as 20 % for small sample sizes or as low as 5 % for large sample sizes. Furthermore, Veech et al. (2016) demonstrated that when there is heterogeneity in detection probability, the models still perform well as long as detection probability is moderately high on average. Martin et al. (2011) found Point-counts models to exhibit substantial bias towards overestimation of actual abundance when there was bimodal heterogeneity in individual detection probability represented by correlated detection probabilities at the sampling sites. Warren et al. (2013) found both underestimation and overestimation of abundance without any strong bias. Couturier et al. (2013), Joseph et al. (2009), and Kery et al. (2005) and Other studies that assessed performance of Point-counts models on real data have also suggested that model performance can be affected by heterogeneity in individual detection probability. Indeed, heterogeneity in detection probability among individuals may be a relevant concern in other abundance-estimation techniques such a Site-Occupancy models.

Our results have shown that the species being rare tends to increase the bias of the  $\beta_0$  estimates (intercept of the intensity model) only. When a species or individual is poorly detected, detectability models may not converge to a solution even for a species occurring across 50 % of the sites (Welsh et al., 2013). In this circumstance models cannot be accurate in estimating parameters. According to Welsh et al. (2013), another issue arising with the analyses of less abundant species with lower detection probabilities is that their fitted probabilities are often found to be equal to one, a result that would generally suggest that the species is in effect widespread. But what is a less abundant species? The definition of rarity often varies with study system and taxa (Yu and Dobson, 2000). A given species is considered as rare if it is undeniable that it was captured once or twice in such a large sample effort and if it failed to occur in at least eight out of 11 sites. That is one of the criteria used by Ferraz et al. (2007). Based on this, we suspect that the abundance for the less abundant species scenario (200 individuals at average values of the intensity covariate) could not be considered to be really the abundance corresponding to a less abundant species. It might have been better to consider an average abundance far below that considered for the less abundant species scenario in our simulations and this could possibly lead to different conclusions.

### 4.3 Models limitations and suggestions for further works

Presence-only data suffer from the disadvantage of containing the problem of spatial auto-correlation (non-independence). We can not afford to ignore this bias in the modeling of species distribution. Species distribution models implicitly assume that the geographical data points for species records are independent and the environmental layers used as hypothetical predictive variables and associated to the geographical records of species also show problems of spatial auto-correlation (Cruz-Cárdenas et al., 2014; Segurado et al., 2006). Disregard and not avoid spatial auto-correlation has consequences such as incurring in poor precision in model coefficients and hence inflate type I errors (Cruz-Cárdenas et al., 2014; Dormann, 2007). Therefore, where biological or population processes induce substantial auto-correlation in the species distribution, and this is not modeled,

model predictions will be inaccurate. The models we have studied should therefore also be evaluated in relation to their predictive performance when the data are subject to spatial auto-correlation. Another unexplored fact of this study is the benefit of adding interaction terms in the models. Additiannal study are needed to explore it.

## 5 Conclusion

In this study, we have shown that the best way to correct the bias induced by sampling bias and imperfect detection is to analyze Presence-only data in conjunction with Point-counts data through Hierarchical species distribution model proposed by Dorazio (2014) which is the combination of Spatial Point Process Model and N-mixture model (also called Point-counts model). Indeed, analyzing Presence-only data in conjunction with Site-occupancy data resulted in poor precision in estimates of species intensity whereas analyzing Presence-only data in conjunction with Point-counts estimated parameters of species intensity without bias. Whatever the species abundance, the Dorazio Hierarchical model has shown to be less sensitive to the variation in detection probability during opportunistic surveys. It has shown to outperform the other models by estimating model's coefficients with almost zero bias and very low variance of estimates. In the case of less abundant species, the advantage offered by the Dorazio Hierarchical model is that, if the repeated planned surveys are conducted in such a way to accurately estimate the detection probability, they need only Point-counts data of small number of sampled sites to correct such a bias. Furthermore, in the case of the abundant species, the number of sites to be sampled is much smaller.

### 5.1 Acknowledgements

The authors sincerely thank students from the Faculty of Agriculture and Environmental studies of the Université Evangélique en Afrique for their valuable contributions in data collection. Authors acknowledge the Université Evangélique en Afrique for support to this work through the University project on improvement of research and teaching quality funded by Pain pour le Monde (Project A-COD-2018-0383). We greatly appreciated technical support of the team of the Laboratoire de Biomathématiques et d'Estimations Forestières (LABEF). A. Belarmain Fandohan was supported by Georg Forster-Hermes Research Fellowship programs of the Alexander von Humboldt Foundation: postdoctoral fellowship number 3.4-BEN/1155509 STP, return fellowship no 3.4 - RKS - BEN/1155509 and equipment subsidy no 3.6 – BEN/1155509.

### 5.2 Conflicts of Interest

The authors declare no conflict of interest.

## References

- Aarts, Geert, John Fieberg, and Jason Matthiopoulos (2012). "Comparative interpretation of count, presence-absence and point methods for species distribution models". In: *Methods in Ecology and Evolution* 3.1, pp. 177–187. ISSN: 2041210X. DOI: [10.1111/j.2041-210X.2011.00141.x](https://doi.org/10.1111/j.2041-210X.2011.00141.x).
- Barry, Simon and Jane Elith (2006). "Error and uncertainty in habitat models". In: *Journal of Applied Ecology* 43.3, pp. 413–423. ISSN: 00218901. DOI: [10.1111/j.1365-2664.2006.01136.x](https://doi.org/10.1111/j.1365-2664.2006.01136.x).
- Brotons, Lluís, Wilfried Thuiller, Miguel B. Araújo, and Alexandre H. Hirzel (2004). "Presence-absence versus presence-only modelling methods for predicting bird habitat suitability". In: *Ecography* 27.4, pp. 437–448. ISSN: 09067590. DOI: [10.1111/j.0906-7590.2004.03764.x](https://doi.org/10.1111/j.0906-7590.2004.03764.x).

- Comte, Lise and Gaël Grenouillet (2013). “Species distribution modelling and imperfect detection: Comparing occupancy versus consensus methods”. In: *Diversity and Distributions* 19.8, pp. 996–1007. ISSN: 13669516. DOI: [10.1111/ddi.12078](https://doi.org/10.1111/ddi.12078).
- Couturier, Thibaut, Marc Cheylan, Albert Bertolero, Guillelme Astruc, and Aurelien Besnard (2013). “Estimating abundance and population trends when detection is low and highly variable: A comparison of three methods for the Hermann’s tortoise”. In: *Journal of Wildlife Management*. ISSN: 0022541X. DOI: [10.1002/jwmg.499](https://doi.org/10.1002/jwmg.499).
- Crall, Alycia W., Catherine S. Jarnevich, Brendon Panke, Nick Young, Mark Renz, and Jeffrey Morisette (2013). “Using habitat suitability models to target invasive plant species surveys”. In: *Ecological Applications* 23.1, pp. 60–72. ISSN: 10510761. DOI: [10.1890/12-0465.1](https://doi.org/10.1890/12-0465.1).
- Cruz-Cárdenas, Gustavo, Lauro López-Mata, José Luis Villaseñor, and Enrique Ortiz (2014). “Potential species distribution modeling and the use of principal component analysis as predictor variables.” In: *Revista Mexicana de Biodiversidad* 85.1, pp. 189–199. ISSN: 18703453. DOI: [10.7550/rmb.36723](https://doi.org/10.7550/rmb.36723).
- Cutrer, A. P., C. Daniel Antinuchi, M. S. Mora, and A. I. Vassallo (2006). “Home-Range and Activity Patterns of the South American Subterranean Rodent *Ctenomys Talarum*”. In: *Journal of Mammalogy*. ISSN: 0022-2372. DOI: [10.1644/05-MAMM-A-386R1.1](https://doi.org/10.1644/05-MAMM-A-386R1.1). arXiv: [NIHMS150003](https://arxiv.org/abs/NIHMS150003).
- De Siqueira, Marinez Ferreira, Giselda Durigan, Paulo de Marco Júnior, and A. Townsend Peterson (2009). “Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species”. In: *Journal for Nature Conservation* 17.1, pp. 25–32. ISSN: 16171381. DOI: [10.1016/j.jnc.2008.11.001](https://doi.org/10.1016/j.jnc.2008.11.001).
- Dodd, C. Kenneth and Robert M. Dorazio (2004). “Using counts to simultaneously estimate abundance and detection probabilities in a Salamander community”. In: *Herpetologica*. ISSN: 0018-0831. DOI: [10.1655/03-60](https://doi.org/10.1655/03-60).
- Dorazio, Robert M. (2007). “On the choice of statistical models for estimating occurrence and extinction from animal surveys”. In: *Ecology*. ISSN: 00129658. DOI: [10.1890/07-0006.1](https://doi.org/10.1890/07-0006.1).
- Dorazio, Robert M. (2012). “Predicting the Geographic Distribution of a Species from Presence-Only Data Subject to Detection Errors”. In: *Biometrics* 68.4, pp. 1303–1312. ISSN: 0006341X. DOI: [10.1111/j.1541-0420.2012.01779.x](https://doi.org/10.1111/j.1541-0420.2012.01779.x).
- Dorazio, Robert M. (2014). “Accounting for imperfect detection and survey bias in statistical analysis of presence-only data”. In: *Global Ecology and Biogeography* 23.12, pp. 1472–1484. ISSN: 1466822X. DOI: [10.1111/geb.12216](https://doi.org/10.1111/geb.12216).
- Dormann, Carsten F. (2007). “Effects of incorporating spatial autocorrelation into the analysis of species distribution data”. In: *Global Ecology and Biogeography* 16.2, pp. 129–138. ISSN: 1466822X. DOI: [10.1111/j.1466-8238.2006.00279.x](https://doi.org/10.1111/j.1466-8238.2006.00279.x).
- Elith, Jane, Mark A. Burgman, and Helen M. Regan (2002). “Mapping epistemic uncertainties and vague concepts in predictions of species distribution”. In: *Ecological Modelling* 157.2-3, pp. 313–329. ISSN: 03043800. DOI: [10.1016/S0304-3800\(02\)00202-8](https://doi.org/10.1016/S0304-3800(02)00202-8). arXiv: [1124](https://arxiv.org/abs/1124).
- Fei, Songlin, Liang Liang, Frederick L. Paillet, Kim C. Steiner, Jingyun Fang, Zehao Shen, Zhiheng Wang, and Frederick V. Hebard (2012). “Modelling chestnut biogeography for American chestnut restoration”. In: *Diversity and Distributions* 18.8, pp. 754–768. ISSN: 13669516. DOI: [10.1111/j.1472-4642.2012.00886.x](https://doi.org/10.1111/j.1472-4642.2012.00886.x).
- Ferraz, Gonçalo, James D. Nichols, James E. Hines, Philip C. Stouffer, Richard O. Bierregaard, and Thomas E. Lovejoy (2007). “A large-scale deforestation experiment: Effects of patch area and isolation on Amazon birds”. In: *Science* 315.5809, pp. 238–241. ISSN: 00368075. DOI: [10.1126/science.1133097](https://doi.org/10.1126/science.1133097).
- Fithian, William and Trevor Hastie (2013). “Finite-sample equivalence in statistical models for presence-only data”. In: *Annals of Applied Statistics* 7.4, pp. 1917–1939. ISSN: 19326157. DOI: [10.1214/13-AOAS667](https://doi.org/10.1214/13-AOAS667). arXiv: [1207.6950](https://arxiv.org/abs/1207.6950).



- Fuller, Trevon, David P. Morton, and Sahotra Sarkar (2008). “Incorporating uncertainty about species’ potential distributions under climate change into the selection of conservation areas with a case study from the Arctic Coastal Plain of Alaska”. In: *Biological Conservation* 141.6, pp. 1547–1559. ISSN: 00063207. DOI: [10.1016/j.biocon.2008.03.021](https://doi.org/10.1016/j.biocon.2008.03.021).
- Guillera-Arroita, Gurutzeta (2017). “Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities”. In: *Ecography* 40.2, pp. 281–295. ISSN: 16000587. DOI: [10.1111/ecog.02445](https://doi.org/10.1111/ecog.02445). arXiv: [ecog.02097](https://arxiv.org/abs/ecog.02097) [[10.1111](https://doi.org/10.1111)].
- Guillera-Arroita, Gurutzeta, José J. Lahoz-Monfort, Darryl I. MacKenzie, Brendan A. Wintle, and Michael A. McCarthy (2014). “Ignoring imperfect detection in biological surveys is dangerous: A response to ‘fitting and interpreting occupancy models’”. In: *PLoS ONE* 9.7. ISSN: 19326203. DOI: [10.1371/journal.pone.0099571](https://doi.org/10.1371/journal.pone.0099571).
- Guisan, Antoine and Wilfried Thuiller (2005a). “Predicting species distribution: Offering more than simple habitat models”. In: *Ecology Letters* 8.9, pp. 993–1009. ISSN: 1461023X. DOI: [10.1111/j.1461-0248.2005.00792.x](https://doi.org/10.1111/j.1461-0248.2005.00792.x). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Guisan, Antoine and Wilfried Thuiller (2005b). “Predicting species distribution: Offering more than simple habitat models”. In: *Ecology Letters* 8.9, pp. 993–1009. ISSN: 1461023X. DOI: [10.1111/j.1461-0248.2005.00792.x](https://doi.org/10.1111/j.1461-0248.2005.00792.x). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Hastie, Trevor and Will Fithian (2013). “Inference from presence-only data; the ongoing controversy”. In: *Ecography* 36.8, pp. 864–867. ISSN: 09067590. DOI: [10.1111/j.1600-0587.2013.00321.x](https://doi.org/10.1111/j.1600-0587.2013.00321.x). arXiv: [NIHMS150003](https://arxiv.org/abs/NIHMS150003).
- Hefley, Trevor J., Brian M. Brost, and Mevin B. Hooten (2017). “Bias correction of bounded location errors in presence-only data”. In: *Methods in Ecology and Evolution* 8.11, pp. 1566–1573. ISSN: 2041210X. DOI: [10.1111/2041-210X.12793](https://doi.org/10.1111/2041-210X.12793). arXiv: [0608246v3](https://arxiv.org/abs/0608246v3) [[arXiv:physics](https://arxiv.org/abs/0608246v3)].
- Jiménez-Valverde, Alberto, Jaf Diniz Filho, Eduardo B. De Azevedo, and Paulo a. V. Borges (2009). “Species Distribution Models Do Not Account for Abundance: The Case of Arthropods on Terceira Island”. In: *Annales Zoologici Fennici* 46.6, pp. 451–464. ISSN: 0003-455X. DOI: [10.5735/086.046.0606](https://doi.org/10.5735/086.046.0606).
- Joseph, Liana N., Ché Elkin, Tara G. Martin, and Hugh P. Possingham (2009). “Modeling abundance using N-mixture models: The importance of considering ecological mechanisms”. In: *Ecological Applications* 19.3, pp. 631–642. ISSN: 10510761. DOI: [10.1890/07-2107.1](https://doi.org/10.1890/07-2107.1).
- Karlsson, Magnus, Sofia Caesar, Jonas Ahnesjö, and Anders Forsman (2008). “Dynamics of colour polymorphism in a changing environment: Fire melanism and then what?” In: *Oecologia* 154, pp. 715–724. ISSN: 00298549. DOI: [10.1007/s00442-007-0876-y](https://doi.org/10.1007/s00442-007-0876-y).
- Karpestam, Einat, Sami Merilaita, and Anders Forsman (2014). “Body size influences differently the detectabilities of colour morphs of cryptic prey”. In: *Biological Journal of the Linnean Society* 113.1, pp. 112–122. ISSN: 10958312. DOI: [10.1111/bij.12291](https://doi.org/10.1111/bij.12291).
- Kearney, Michael R., Brendan A. Wintle, and Warren P. Porter (2010). “Correlative and mechanistic models of species distribution provide congruent forecasts under climate change”. In: *Conservation Letters* 3.3, pp. 203–213. ISSN: 1755263X. DOI: [10.1111/j.1755-263X.2010.00097.x](https://doi.org/10.1111/j.1755-263X.2010.00097.x).
- Kellner, Kenneth F. and Robert K. Swihart (2014). “Accounting for imperfect detection in ecology: A quantitative review”. In: *PLoS ONE* 9.10, e111436. ISSN: 19326203. DOI: [10.1371/journal.pone.0111436](https://doi.org/10.1371/journal.pone.0111436).
- Kery, Marc, J. Andrew Royle, and Hans Schmid (2005). “Modeling avian abundance from replicated counts using binomial mixture models”. In: *Ecological Applications*. ISSN: 10510761. DOI: [10.1890/04-1120](https://doi.org/10.1890/04-1120).
- Koshkina, Vira, Yan Wang, Ascelin Gordon, Robert M. Dorazio, Matt White, and Lewi Stone (2017). “Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection”. In: *Methods in Ecology and Evolution* 8.4, pp. 420–430. ISSN: 2041210X. DOI: [10.1111/2041-210X.12738](https://doi.org/10.1111/2041-210X.12738).



- Kremen, C., A. Cameron, A. Moilanen, S. J. Phillips, C. D. Thomas, H. Beentje, J. Dransfield, B. L. Fisher, F. Glaw, T. C. Good, G. J. Harper, R. J. Hijmans, D. C. Lees, E. Louis, R. A. Nussbaum, C. J. Raxworthy, A. Razafimpahanana, G. E. Schatz, M. Vences, D. R. Vieites, P. C. Wright, and M. L. Zjhra (2008). “Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools”. In: *Science* 320.5873, pp. 222–226. ISSN: 00368075. DOI: [10.1126/science.1155193](https://doi.org/10.1126/science.1155193).
- Lahoz-Monfort, José J., Gurutzeta Guillera-Aroita, and Brendan A. Wintle (2014). “Imperfect detection impacts the performance of species distribution models”. In: *Global Ecology and Biogeography* 23.4, pp. 504–515. ISSN: 1466822X. DOI: [10.1111/geb.12138](https://doi.org/10.1111/geb.12138). arXiv: [arXiv:1101.0479v1](https://arxiv.org/abs/1101.0479v1).
- Li, Xinhai and Yuan Wang (2013). “Applying various algorithms for species distribution modelling”. In: *Integrative Zoology* 8.2, pp. 124–135. ISSN: 17494877. DOI: [10.1111/1749-4877.12000](https://doi.org/10.1111/1749-4877.12000).
- MacKenzie, Darryl I., James D. Nichols, Gideon B. Lachman, Sam Droege, Andrew A. Royle, and Catherine A. Langtimm (2002). “Estimating site occupancy rates when detection probabilities are less than one”. In: *Ecology* 83.8, pp. 2248–2255. ISSN: 00129658. DOI: [10.1890/0012-9658\(2002\)083\[2248:ESORWD\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Mackenzie, Darryl I. and J. Andrew Royle (2005). “Designing occupancy studies: General advice and allocating survey effort”. In: *Journal of Applied Ecology* 42.6, pp. 1105–1114. ISSN: 00218901. DOI: [10.1111/j.1365-2664.2005.01098.x](https://doi.org/10.1111/j.1365-2664.2005.01098.x).
- Martin, Julien, J. Andrew Royle, Darryl I. Mackenzie, Holly H. Edwards, Marc Kéry, and Beth Gardner (2011). “Accounting for non-independent detection when estimating abundance of organisms with a Bayesian approach”. In: *Methods in Ecology and Evolution* 2.6, pp. 595–601. ISSN: 2041210X. DOI: [10.1111/j.2041-210X.2011.00113.x](https://doi.org/10.1111/j.2041-210X.2011.00113.x).
- Nielsen, Scott E., Chris J. Johnson, Douglas C. Heard, and Mark S. Boyce (2005). “Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history”. In: *Ecography* 28.2, pp. 197–208. ISSN: 09067590. DOI: [10.1111/j.0906-7590.2005.04002.x](https://doi.org/10.1111/j.0906-7590.2005.04002.x).
- Pearce, J., S. Ferrier, and D. Scotts (2001). “An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales”. In: *Journal of Environmental Management* 62.2, pp. 171–184. ISSN: 03014797. DOI: [10.1006/jema.2001.0425](https://doi.org/10.1006/jema.2001.0425).
- Peterson, A Townsend, Jorge Soberón, Richard G Pearson, Robert P Anderson, Enrique Martínez-Meyer, Miguel Nakamura, and Miguel Bastos Araújo (2011). *Ecological Niches and Geographic Distributions*. Vol. 2011. Princeton University Press, p. 336. ISBN: 0691136882. DOI: [10.5860/CHOICE.49-6266](https://doi.org/10.5860/CHOICE.49-6266).
- Phillips, Steven J., Miroslav Dudík, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier (2009). “Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data”. In: *Ecological Applications* 19.1, pp. 181–197. ISSN: 10510761. DOI: [10.1890/07-2153.1](https://doi.org/10.1890/07-2153.1). arXiv: [1132](https://arxiv.org/abs/1132).
- R Development Core Team (2018). *R statistical software*. Vienna, Austria.
- Royle, J. Andrew (2004). “N-Mixture Models for Estimating Population Size from Spatially Replicated Counts”. In: *Biometrics* 60.1, pp. 108–115. ISSN: 0006341X. DOI: [10.1111/j.0006-341X.2004.00142.x](https://doi.org/10.1111/j.0006-341X.2004.00142.x).
- Royle, J. Andrew and Robert M. Dorazio (2006). “Hierarchical models of animal abundance and occurrence”. In: *Journal of Agricultural, Biological, and Environmental Statistics* 11, pp. 249–263. ISSN: 10857117. DOI: [10.1198/108571106X129153](https://doi.org/10.1198/108571106X129153).
- Segurado, P., M. B. Araújo, and W. E. Kunin (2006). “Consequences of spatial autocorrelation for niche-based models”. In: *Journal of Applied Ecology* 43.3, pp. 433–444. ISSN: 00218901. DOI: [10.1111/j.1365-2664.2006.01162.x](https://doi.org/10.1111/j.1365-2664.2006.01162.x). arXiv: [1132](https://arxiv.org/abs/1132).

- Storfer, Andrew, Jonra Cross, Victor Rush, and Joseph Caruso (1999). “Adaptive coloration and gene flow as a constraint to local adaptation in the streamside salamander”. In: *Evolution* 53.3, pp. 889–898. ISSN: 00143820. DOI: [10.1111/j.1558-5646.1999.tb05383.x](https://doi.org/10.1111/j.1558-5646.1999.tb05383.x).
- Tyre, Andrew J., Brigitte Tenhumberg, Scott A. Field, Darren Nijalke, Kirsten Parris, and Hugh P. Possingham (2003). *Improving precision and reducing bias in biological surveys: Estimating false-negative error rates*. DOI: [10.1890/02-5078](https://doi.org/10.1890/02-5078).
- Veech, Joseph A., James R. Ott, and Jeff R. Troy (2016). “Intrinsic heterogeneity in detection probability and its effect on N-mixture models”. In: *Methods in Ecology and Evolution* 7.9, pp. 1019–1028. ISSN: 2041210X. DOI: [10.1111/2041-210X.12566](https://doi.org/10.1111/2041-210X.12566).
- Warren, Christopher C., Joseph A. Veech, Floyd W. Weckerly, Lisa O'Donnell, and James R. Ott (2013). “Detection heterogeneity and abundance estimation in populations of Golden-cheeked Warblers”. In: *The Auk* 130.4, pp. 677–688. ISSN: 00048038. DOI: [10.1525/auk.2013.13022](https://doi.org/10.1525/auk.2013.13022).
- Welsh, Alan H., David B. Lindenmayer, and Christine F. Donnelly (2013). “Fitting and Interpreting Occupancy Models”. In: *PLoS ONE* 8.1, e52015. ISSN: 19326203. DOI: [10.1371/journal.pone.0052015](https://doi.org/10.1371/journal.pone.0052015).
- Yu, Jinping and F. Stephen Dobson (2000). “Seven forms of rarity in mammals”. In: *Journal of Biogeography* 27.1, pp. 131–139. ISSN: 03050270. DOI: [10.1046/j.1365-2699.2000.00366.x](https://doi.org/10.1046/j.1365-2699.2000.00366.x).
- Zylstra, Erin R., Robert J. Steidl, and Don E. Swann (2010). “Evaluating Survey Methods for Monitoring a Rare Vertebrate, the Sonoran Desert Tortoise”. In: *Journal of Wildlife Management* 74.6, pp. 1311–1318. ISSN: 0022-541X. DOI: [10.2193/2009-331](https://doi.org/10.2193/2009-331).