POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Matematica

Tesi di Laurea Magistrale

Bayesian hierarchical models for presence-availability data, with application to GPS locations of bears and wolves.

Relatore
Prof. Gianluca Mastrantonio
firma del relatore

. . . . . .

Candidato
Valeria Schellino
firma del candidato

. . . . . .

Anno Accademico 2020-2021

A mia sorella

Summary

Resource Selection Functions allow knowing which habitat characteristics are preferred by a species. Two types of locations are considered: a set of locations in which the animal has been recorded (usually through GPS), which are called "used locations", and a set of "available locations" where the individual has not been observed. The latter are usually sampled uniformly over the animal home range.

The thesis aims to construct Bayesian hierarchical models for use-availability data of two animals that are modeled jointly. We formalized four models and, through a simulation study, we evaluated the identifiability of the parameters and the convergence of the model with respect to the number of available locations. The models are then estimated on a real dataset of GPS locations of bears and wolves, collected in the PNALM area (National Park of Abruzzo, Lazio, and Molise). We fit a different model for each of the four time-windows where the data are recorded (spring, early summer, late summer, autumn), using environmental data as model covariates. Results obtained are interpretable and give insights on how the animals interact and choose the resources.

Acknowledgements

Ringrazio innanzitutto il mio relatore Gianluca Mastrantonio per la sua disponibilità e il suo aiuto durante la stesura della tesi; le sue risposte rapide e chiare mi hanno permesso di risolvere i miei dubbi.

Ringrazio il Professore Paolo Ciucci dell'Università La Sapienza di Roma e la sua alunna Cecilia Parracciani per averci gentilmente fornito i dati utilizzati per la mia tesi.

Ringrazio la mia professoressa delle superiori Caterina Mellano per avermi fatto apprezzare la matematica e avermi aiutata a capire che questa era la mia strada, anche se arrivavo dal Liceo Classico.

Ringrazio, poi, le persone importanti della mia vita. Grazie mamma e papà per avermi supportata e sopportata durante questi cinque anni di università.

Grazie Paola per essere una costante e una certezza nella mia vita, so che posso sempre contare su di te e che il nostro legame è unico e forte.

Grazie ai miei nonni Renza e Valerio per i loro insegnamenti di vita. Porterò sempre nel cuore il ricordo della mano grande di nonno che mi accompagnava ogni giorno a scuola e le gare di operazioni che facevo il pomeriggio con nonna. Grazie anche a nonna Lucia, che non capiva perché studiassi sempre.

Grazie a mia zia Ornella, che mi ha insegnato che *le salite più dure sono quelle che ci conducono più in alto* e ha tenuto la sua mano sulla mia spalla quando queste salite mi sembravano un po' troppo faticose.

Grazie alle mie amiche del liceo, per aver condiviso con me momenti speciali anche durante l'università. Grazie in particolare a Camilla, per esserci sempre per me, per gioire genuinamente dei miei successi e per ascoltarmi senza giudizio.

Grazie ai miei compagni di università, per aver condiviso con me dubbi, progetti, videochiamate, viaggi in treno, pause caffè, pranzi, pizzate e gelati. Con ognuno di voi ho un ricordo felice. Grazie in particolare a Giulia, Lidia e Giulia, per essere diventate amiche preziose.

Grazie a Elena, Andrea, Marta, Giovanni e nonna Rita per avermi accolta nella loro famiglia e per aver condiviso con me tempo prezioso durante la quarantena.

Grazie, infine, a Lorenzo, per aver creduto in me anche quando io non ci riuscivo e avermi sempre aiutata e spronata a dare il meglio di me. Senza di te non ce l'avrei fatta.

Contents

Le salite più dure
sono quelle che ci portano
sulle cime più alte

[zia Ornella]

Chapter 1

Introduction

Studying where large animals move within an area allows seeing which habitat characteristics are preferred by a species and detecting interactions between different species. In this thesis, species of interest are the Marsican bear (*Ursus arctos marsicanus*) and the wolf (*Canis lupus*) in the National Park of Abruzzo, Lazio, and Molise. The latter constitutes the only area in the mountain range of the Apennines where these two species live in conditions of sympatry. There are several studies on the interaction between different animals that have a dominant/subordinate relationship, but relatively few on the relationship between large dominant species [Milleret et al., 2018, Darnell et al., 2014, Belant et al., 2010], as bears and wolves. Milleret et al. [2018] studied habitat segregation between bears and wolves in a human-dominated landscape in Sweden; also in the National Park of Abruzzo, Lazio, and Molise, there is a human presence. The coexistence between bears and wolves is also characterized by common prey. This could lead to kleptoparasitism: one species captures prey and the other species takes advantage of that booty, either by stealing it or by eating its leftover carcasses. Milleret et al. [2018] divide the interactions between wolves and bears in two types: a consumer-resource interaction, where the bear benefits from the presence of the wolf, for instance through kleptoparasitism and the wolf has a disadvantage, or an exploitative competition, where both the species are disadvantaged by their interaction.

This thesis aims to create four Bayesian hierarchical models for spatial locations of bears and wolves, taking into account environmental covariates (to see which habitat components are preferred by the species) and possible interaction between the two species. In Chapter 2, Resource selection functions are presented, focusing on a logistic regression approach, in which two types of animal locations are considered: locations of presence, where the individual has been observed (for example through GPS), and locations of availability, that are sampled in the individual home range. The different orders of selection of Johnson [1980] are described, and a definition and a mathematical way to estimate the home range are presented. Then, theoretical insights about Bayesian inference, hierarchical models and Markov Chain Monte Carlo are given. In Chapter 3 data is described, divided into four time-windows. In Chapter 4 four different hierarchical models are presented. The first one considers that the probability of finding the bear in a location depends only on the environment in that location, while the probability of finding the wolf depends on the

presence of the bear beside environmental covariates; the second one vice versa. The third one considers that if a species is present in a location, the other has a constant probability of being present, that is equal in each location where the other species is present; otherwise, the probability of finding it depends only on the environmental characteristics of the considered location. The fourth model takes into consideration also a possible attractive or repulsive effect between the two species, with different strengths. In Chapter 5 a simulation study is presented, in which two different aspects are considered: the identifiability of the models, and a check on the number of available locations for each model and each dataset taken into consideration. Finally, in Chapter 6 results obtained on real datasets and a comparison between the different models are shown.

1.1 Notations

In this section, some useful notations that are used in the thesis are presented.

- $\mathbb{E}[X]$: expected value of the random variable X.
- $X \perp\!\!\!\perp Y$: the random variables X and Y are independent.
- $\mathcal{B}(\mathcal{X})$: Borel sigma-algebra of the space $\mathcal{X}$.
- $\int f(x)P(dx)$: Lebesgue integral with respect to the measure P.
- Bold symbols (e.g. $\boldsymbol{x}$) denote vectors, which are columns by default. The transposed version of $\boldsymbol{x}$ is $\boldsymbol{x}'$, so that $\boldsymbol{x}' \cdot \boldsymbol{y}$ denotes scalar product.
- $X \sim P$: the random variable X is distributed according to P.
- Subscripts denote vector element, while superscripts denote sample index.

Chapter 2

Methods

2.1 Resource selection functions

Resource Selection Functions (**RSFs**), consist in a class of functions that allow knowing which habitat characteristics are preferred by a species. A RSF is usually proportional to the probability that a unit of habitat is used by an animal [Mark S. Boyce, 1999, Manly et al., 2002]. In particular, RSFs link species choices to the resources available in a given unit of habitat [Hooten et al., 2017].

A possibility is to write the RSF as the product of two non-negative functions representing the selection process (with which the animal chooses the location) and the availability of resources [Hooten et al., 2017]. This product is normalized, in order to find a probability density function. Let us denote a spatial location by $\boldsymbol{\mu}^i$, the resources at a given spatial location as $\boldsymbol{x}(\boldsymbol{\mu}^i)$. If you consider the locations to be independent, the RSF can be written as:

$$p\left(\boldsymbol{\mu^i}|\boldsymbol{\beta},\boldsymbol{\theta}\right) = \frac{g\left(\mathbf{x}\left(\boldsymbol{\mu^i}\right),\boldsymbol{\beta}\right) f\left(\boldsymbol{\mu}^i,\boldsymbol{\theta}\right)}{\int g(\mathbf{x}(\boldsymbol{\mu}),\boldsymbol{\beta}) f(\boldsymbol{\mu},\boldsymbol{\theta}) d\boldsymbol{\mu}}, \tag{2.1}$$

where g represents the selection process function which depends on the selection coefficients $\boldsymbol{\beta}$ and the resources $\boldsymbol{x}(\boldsymbol{\mu}^i)$, while the availability function f depends on the availability coefficients $\boldsymbol{\theta}$. In theory, availability and selection functions can assume any possible form; however, if we suppose that the individual can go anywhere in the considered region with equal probability, the availability function is uniform: in this case, availability coefficients disappear and only the selection coefficients $\boldsymbol{\beta}$ are present [Hooten et al., 2017].

Eq. (2.1) represents a Point process model, that describes a RSF considering only locations $\boldsymbol{\mu}^i$ chosen by an animal. However, there are also other alternative methods to fit RSF models (for more details see Hooten et al. 2017); here, we will focus on a logistic regression approach. In the latter, two types of locations are considered: a set of locations in which the animal has been recorded (usually through GPS, Global Positioning System, or similar methods), that are called **used locations**, and a set of **available locations** where the individual has not been observed but is free to go. The latter are not locations where the individual has never passed through, but simply it has never been recorded there. These are usually sampled in a **domain of availability** [Manly et al.,

2002, Forester et al., 2009] where the animal moves. The available locations are used to understand the distribution of the resources in the domain of availability and to compare those with the resource distribution in the used locations; in this way, it is possible to understand in which measure each resource impacts where the animal goes.

A target variable y_i is assigned to each location $\boldsymbol{\mu}^i$ and it is equal to 1 if the considered location is used, 0 if it is available. A logistic regression model is the following:

$$\begin{aligned} y_i &\sim Bernoulli(\pi_i) \\ logit(\pi_i) &= \beta_1 + \boldsymbol{x}(\boldsymbol{\mu}^i)' \cdot \boldsymbol{\beta} \quad \text{for} \quad i = 1, ..., N_{tot}, \end{aligned} \tag{2.2}$$

where π_i is the probability that the individual has been present in location $\boldsymbol{\mu}^i$, N_{tot} is the total number of observations, and $\boldsymbol{\beta} = (\beta_2, ..., \beta_{n+1})$ are the selection coefficients. n is equal to the number of resources taken into consideration, and β_1 is the intercept. This is the simplest logistic regression model for RSFs and considers only the dependence on resources.

An RSF estimates the effect of a resource on a species [Michelot et al., 2019]. In Eq (2.2), the coefficient β_j represents the effect of the j^{th} covariate x_j on the species. $\beta_j > 0$ represents a *preference* of positive values of the covariate by the animal, while if $\beta_j = 0$ there is *indifference*, and if $\beta_j < 0$ there is *avoidance* of positive values of the covariate [Michelot et al., 2019].

It is now important to describe how available locations can be obtained. If used locations are the ones recorded through GPS or similar methods, available locations are usually sampled from the domain of availability, that can be different based on the granularity of the study. Johnson [1980] individuates different scales at which individuals select resources:

1. *First-order selection*: the species selects a geographical area;
2. *Second-order selection*: represents the home range of an individual or a social group;
3. *Third-order selection*: the species selects different components of habitat within the home range;
4. *Fourth-order selection*: the species selects food.

Understanding in which scale of selection the analysis happens is fundamental in order to obtain a good interpretation of the model [Johnson, 1980]. In our case, the region of interest is the **home range** (*Third-order selection*). It is important to clarify what is the home range and how it can be mathematically estimated.

2.1.1 Estimating the home range

There are different ways to define the home range for an animal. The most common definition of home range is the one in Burt [1943], that defines the home range as the area occupied by an individual in its normal activities of food gathering, mating, and taking care of offspring; the individual can sometimes go outside this area for explorations or other reasons, but these territories should not be considered part of the home range.

Estimating the home range is difficult. In fact, as Powell and Mitchell [2012] argue, most works use heuristic estimates of the home range rather than information and data on habitat and resources and rather than understanding what really is home range for the considered species.

The choice of the home range for an animal can be seen as *"a product of decision-making processes shaped by natural selection"* in order to increase its fitness [Powell and Mitchell, 2012]. Mitchell et al. [2012] propose different economical models in which animals, in choosing their home ranges, try to maximize their benefits over the costs: more specifically, they maximize resources while minimizing their home range area (in order to avoid dangers). Finally, Powell and Mitchell [2012] argue that home ranges are different among animals of different species and possibly also among individuals of the same species and for a single individual over time.

In Hooten et al. [2017], we can find another definition for the home range: *"non linear feature in the multidimensional space that serves as a semi-permeable boundary to movement"*. There are two common techniques used for mathematically estimating the home range:

- **isopleth of the Kernel Density Estimate**; a probability density in the space is obtained from the observed animal locations via Kernel Density Estimate, and then a contour line or isopleth (a line drawn through all the points on the surface with same density value) is used. For instance, if we consider the 95% isopleth, the home range will contain 95% of the total density in the considered region.

- **convex hull**: the smallest convex polygon containing all the observed locations is used as home range. This method is less subjective than the previous one, even if also in this case it is possible to choose a percentage of points to be used for the creation of the convex polygon.

Once the home range is estimated, available locations are sampled within it and used to fit the RSFs. The aim is to find the selection coefficients $\boldsymbol{\beta}$ in order to estimate, for each possible location $\boldsymbol{\mu}^i$, if it is an used location or an available one. Here, we adopt a Bayesian approach, which implies choosing a prior for $\boldsymbol{\beta}$ and finding a posterior distribution for $[\boldsymbol{\beta} \mid \boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^{N_{tot}}]$.

2.2 The Bayesian approach

In this section, an historical description of the different definitions of probability is presented. Then, the focus will be on Bayesian statistics.

2.2.1 Different definitions of probability

Different definitions of **probability** are possible (for example, see Regis 2020): **classical**, **frequentist** and **Bayesian**.

For the **classical** definition, let us consider a space of events Σ and let us suppose that these events are equiprobable. Then, the probability of an event is given by *the ratio*

between the number of favorable cases and the number of possible cases [Stigler, 2005]. Mathematically, if N is the number of possible cases and N_E is the number of favorable cases for an event $E \subseteq \Sigma$, the probability of the event E is:

$$P_E = \frac{N_E}{N}.$$

This definition is based on a finite number of discrete events, so it is not easily extendible to the case of continuous variables. Moreover, another weakness of this definition is that equiprobability between events is supposed before defining the notion of probability itself, resulting in a problem of circularity in the definition [Regis, 2020].

The **frequentist** definition is given by Von Mises [1939]. If you perform an experiment N times and the event E occurs N_E times, the probability of E is the limit of the relative frequency $N_E(N)/N$ for $N \to \infty$ [Regis, 2020]:

$$P_E = \lim_{N\to\infty} \frac{N_E(N)}{N}.$$

In this case it is not necessary to know the space of the events before giving the definition and it is not necessary to have equiprobability of events, but we have to assume that the experiment is ideally repeatable an infinite number of times under the same conditions [Regis, 2020]. This can be a problem since it could be not possible to ensure that all the experiments are repeatable many times without changing some conditions.

The **Bayesian definition** considers the probability as *a measure in [0,1] of the degree of plausibility of a proposition* [Regis, 2020]. This definition can be applied to every possible proposition, as complex propositions can be seen as combination of simpler propositions through logic operators (like *and*, *or*, *not*). Personal considerations are used to assign a probability to a given event before conducting an experiment [Regis, 2020]. This probability is called **prior** probability. Then, once finished the experiment, a new probability is computed based on results of the experiment: this is called **posterior** probability and is not absolute, since always conditioned to previous knowledge. To summarize, we start from a given prior belief and then we update that belief using new observations when they arrive.

2.2.2 Bayesian inference

In Bayesian inference, a model described by some parameters $\boldsymbol{\theta}$ is assumed true and a prior distribution is assigned to $\boldsymbol{\theta}$. Then, a posterior distribution of $\boldsymbol{\theta}$ is obtained after the observation of data $\boldsymbol{x}$ [Regis, 2020] via **Bayes theorem**:

Theorem 1 (Bayes theorem)**.** *The posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{x})$ can be obtained by:*

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{x})},$$

where $f(\boldsymbol{x}|\boldsymbol{\theta})$ is called likelihood, $f(\boldsymbol{x})$ is called marginal probability and $f(\boldsymbol{\theta})$ is the prior probability.

Bayesian statistics differs from frequentist statistics, due to putting a prior distribution on $\boldsymbol{\theta}$ and considering $\boldsymbol{\theta}$ to be a random variable; in contrast, frequentist statistics considers $\boldsymbol{\theta}$ to be a fixed quantity.

Example 1. *Let us now show an example that clarifies these concepts. Let us consider a linear model of the type:*

$$x_i = \beta_1 + \beta_2 t_i + \epsilon_i \quad \forall\ i \in I,$$

where $\epsilon_i \sim Normal(0, \sigma^2)$ *and* $I = (1,2,...,N_{tot})$ *with* N_{tot} *being the total number of observations, and* $t_i \in \mathbb{R}$*. Moreover, let us suppose that* $\forall i \neq j \in I \quad x_i \perp\!\!\!\perp x_j$*. Parameters of interest are in this case* $\boldsymbol{\theta} = (\sigma^2, \beta_1, \beta_2)$*.*

The frequentist approach considers the set of parameters as fixed and uses the maximum likelihood estimator

$$\arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} f(x_i|\boldsymbol{\theta})$$

to find them (where $f(x_i|\boldsymbol{\theta})$ *is the conditional probability distribution of* x_i *given* $\boldsymbol{\theta}$*).*

In the Bayesian approach, a prior distribution $f(\boldsymbol{\theta})$ *is initially chosen, then a sample* $(\boldsymbol{t}, \boldsymbol{x})$ *is observed and finally a posterior distribution for* $\boldsymbol{\theta}$ *is computed through Bayes theorem.*

The arbitrariness in the choice of the prior is the most criticized aspect of Bayesian statistics. There are two possibilities for the choice of the prior distribution:

- to choose a **non-informative prior**; this means to consider a distribution that does not give information. This strategy of *complete ignorance* is usually pursued when it is not possible to have any information on the parameters. For instance, in the case of a parameter representing a probability, a non-informative prior could be an $Uniform(0,1)$ (this is called flat prior);
- to choose an **informative prior**, that means to choose a prior distribution based on past experiments or theoretical assumptions.

Note that with particular choices of the prior distribution, it is possible to analytically compute posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{x})$. If prior and posterior belong to the same family of distributions, they are said to be *conjugate*.

Point estimators of parameters values can be obtained from the posterior; it is common to use the mean of the posterior distribution [Regis, 2020]:

$$\mathbb{E}[\boldsymbol{\theta}] = \int \boldsymbol{\theta}\ f(\boldsymbol{\theta}|\boldsymbol{x})\ \mathrm{d}\boldsymbol{\theta}.$$

Alternatively, the maximum of the posterior can be used. Moreover, it is possible to compute the *credible intervals* (also called *intervals of credibility*) of the parameters: a credible interval of amplitude α is indicated with C_α and defined as:

$$\int_{C_\alpha} f(\boldsymbol{\theta}|\boldsymbol{x})\ \mathrm{d}\boldsymbol{\theta} = \alpha.$$

Bayesian statistics can be applied to particular types of models called hierarchical models, which we discuss next.

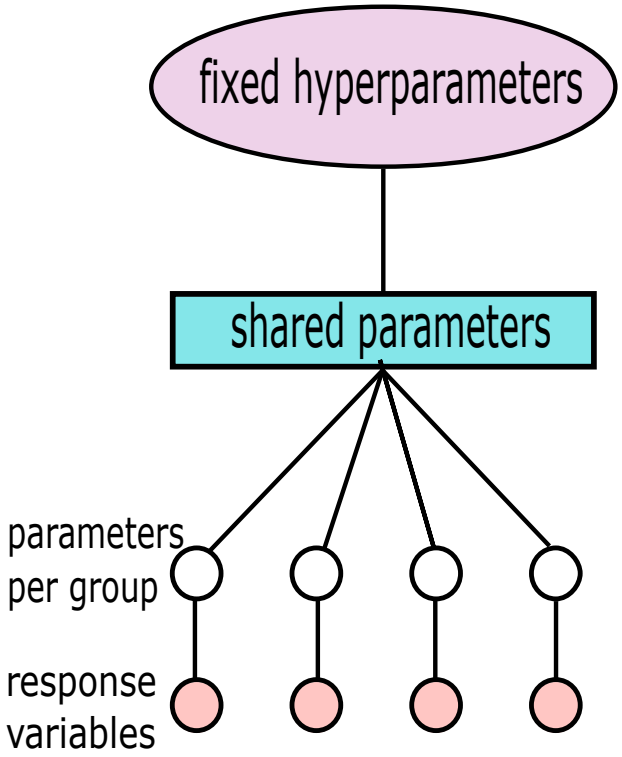

Figure 2.1. Structure of a generic hierarchical model.

2.3 Hierarchical models

Hierarchical models are used when the data presents a multi-level structure. A generic hierarchical model is represented in Figure 2.1, where the following components are present:

- *fixed hyperparameters*: are chosen a priori and are fixed. Fixed hyperparameters are for instance the parameters of the prior distributions. In Figure 2.1, this is represented by violet oval blocks.
- *shared parameters*: parameters common to different groups, represented by a light blue rectangular block.
- *parameters per group*: they are different for each group taken into consideration and are represented by white circles.
- *response variables*: each group has the correspondent response variable, represented into pink circles.

A hierarchical model has a conditional structure: lines between blocks represents conditionality between random variables, while blocks not connected by lines are independent conditioned on the value of the other blocks. Thus, it is natural to apply Bayes theorem

in hierarchical models: by observing the response variables, you can recursively infer the posterior distribution for the other levels.

Note that Figure 2.1 is only an explicative example, but, in practise, hierarchical models can have more complex structures.

2.4 Sampling from the posterior distribution

Let us consider the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{x})$ and let us suppose we want to compute the expected value of a function $g(\boldsymbol{\theta})$ with respect to it:

$$\mathbb{E}[g(\boldsymbol{\theta})] = \int g(\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}) d\boldsymbol{\theta}.$$

If it is not possible to compute the integral, we can approximate its value with a **Monte Carlo** estimate, which requires to sample from the posterior distribution and use these samples to build an empirical estimate of the integral. Say that we have b different samples $\boldsymbol{\theta}^1, ..., \boldsymbol{\theta}^b$ from $f(\boldsymbol{\theta}|\boldsymbol{x})$. Thus,

$$\frac{1}{b} \sum_{j=1}^{b} g(\boldsymbol{\theta}^j) \approx \mathbb{E}[g(\boldsymbol{\theta})] = \int g(\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}) d\boldsymbol{\theta}.$$

If it is not possible to directly sample from $f(\boldsymbol{\theta}|\boldsymbol{x})$, iterative methods can be used to obtain approximate samples. To this aim, we will now introduce **Markov chains** over a general *measurable state space*.

2.4.1 Markov chains

First, let us consider a state space $\boldsymbol{\Sigma}$, and let us denote by $\mathcal{B}(\boldsymbol{\Sigma})$ the Borel sigma-algebra on $\boldsymbol{\Sigma}$.

Definition 1. A **Markov kernel** or **transition kernel** is a function $\mathcal{K}$ defined on $\boldsymbol{\Sigma} \times \mathcal{B}(\boldsymbol{\Sigma})$ such that [Robert, 1999]:

- $\forall s \in \boldsymbol{\Sigma}$, $\mathcal{K}(s, \cdot)$ is a probability measure;
- $\forall A \in \mathcal{B}(\boldsymbol{\Sigma})$, $\mathcal{K}(\cdot, A)$ is measurable.

We are now ready to define a Markov chain.

Definition 2. Given a transition kernel $\mathcal{K}$, a **discrete time Markov chain in measurable space** is a sequence of random variables $X^0, X^1, ..., X^n$ denoted as $(X^n)_n$ such that $\forall k$ the conditional distribution of X^{k+1} given $x^k, x^{k-1}, ..., x^0$ is the same as the distribution of X^{k+1} given x^k [Robert, 1999]:

$$P\left(X^{k+1} \in A \mid x^0, x^1, ..., x^k\right) = P\left(X^{k+1} \in A \mid x^k\right) = \int_A \mathcal{K}(x^k, dx).$$

When the state space $\mathbf{\Sigma}$ is discrete (namely finite or countably infinite), the transition kernel $\mathcal{K}$ coincides with the matrix of transition probabilities:

$$p_{(i,j)} = P\left(X^{k+1} = j \mid X^k = i\right) \ \forall i, j \in \mathbf{\Sigma}.$$

In the continuous case, the transition probability can be obtained via integration of the transition kernel [Robert, 1999]:

$$P\left(X \in A \mid x\right) = \int_A \mathcal{K}(x, d\widetilde{x}).$$

By denoting $\mathcal{K}^1(x, A) = \mathcal{K}(x, A)$, if we take into consideration $n > 1$ transitions, the n-steps kernel can be obtained as[1]:

$$\mathcal{K}^n(x, A) = \int_{\mathbf{\Sigma}} \mathcal{K}^{n-1}(y, A)\ \mathcal{K}(x, dy).$$

This allows to define the **Chapman-Kolmogorov equations**:

$$\forall (m, n) \in \mathbb{N}^2,\ x \in \mathbf{\Sigma},\ A \in \mathcal{B}(\mathbf{\Sigma}),$$

$$\mathcal{K}^{m+n}(x, A) = \int_{\mathbf{\Sigma}} \mathcal{K}^n(y, A)\ \mathcal{K}^m(x, dy).$$

This means that to get from x to A in $m + n$ steps it is necessary to pass through some y on the n^{th} step [Robert, 1999].

Definition 3. Let $A \in \mathcal{B}(\mathbf{\Sigma})$. The **stopping time** at A is

$$\tau_A = \inf_{n \geq 1}\{X^n \in A\}$$

and represents the first step n at which the chain enters in A. If $X^n \notin A\ \forall n$, by convention $\tau_A = \infty$ [Robert, 1999].

Now, we define two useful notions for characterizing Markov Chains, those of irreducibility and recurrence.

Definition 4. In the case of *discrete state space*, a chain is **irreducible** if from each state of the chain you can go to any other state, that means that all the states communicate. Mathematically, this can be written as:

$$P_x(\tau_y < \infty) > 0\ \forall x, y \in \mathbf{\Sigma},$$

where P_x indicates the probability starting from the point x [Robert, 1999].

In the case of *continuous state space*, in order to define the irreducibility of a chain, it is necessary to consider a measure ϕ and a transition kernel $\mathcal{K}(\cdot, \cdot)$.

[1] For $\mathcal{K}$ superscripts indicate power.

Definition 5. A Markov chain $(X^n)_n$ with transition kernel $\mathcal{K}$ is ϕ-**irreducible** if $\forall A \in \mathcal{B}(\mathbf{\Sigma})$ such that $\phi(A) > 0$, there exists $n > 0$ such that $\mathcal{K}^n(x, A) > 0 \ \forall x \in \mathbf{\Sigma}$. This is equivalent to $P_x(\tau_A < \infty) > 0$ [Robert, 1999].

Definition 6. A set C is **small** if there exists $m \in \mathbb{N}$ and there exists a non-zero measure ν_m such that [Robert, 1999]:

$$\mathcal{K}^m(x, A) \geq \nu_m(A) \quad \forall x \in C, \ \forall A \in \mathcal{B}(\mathbf{\Sigma}).$$

Definition 7. A chain $(X^n)_n$ that is ϕ-irreducible is **recurrent** if there exists a small set C with $\phi(C) > 0$ such that $P_x(\tau_C < \infty) = 1 \ \forall x \in C$ [Robert, 1999].

Some Markov Chains can be associated to an *invariant* probability measure:

Definition 8. A probability measure π is invariant for the transitional kernel $\mathcal{K}(\cdot, \cdot)$ [Robert, 1999] if

$$\pi(B) = \int_{\mathbf{\Sigma}} \mathcal{K}(x, B) \ \pi(dx) \ \forall B \in \mathcal{B}(\mathbf{\Sigma}).$$

The invariant distribution is also called **stationary** distribution. In fact, if π is an invariant probability measure, $X^0 \sim \pi \implies X^n \sim \pi \ \forall n$ that means that the chain is stationary in distribution [Robert, 1999]. For a recurrent chain, there exists a unique invariant probability measure π [Robert, 1999].

Definition 9. A stationary Markov chain is **reversible** if the distribution of X^{n+1} given $X^{n+2} = x$ is the same as the distribution of X^{n+1} given $X^n = x$

In other words, the direction of time does not change the dynamics of a reversible chain [Robert, 1999].

Definition 10. A Markov chain with a kernel $\mathcal{K}$ satisfies the **detailed balance condition** if there exists a function g satisfying:

$$\mathcal{K}(y, x) \ g(y) = \mathcal{K}(x, y) \ g(x) \quad \forall x, y.$$

Theorem 2. *If a Markov chain with transition kernel $\mathcal{K}$ satisfies the detailed balance condition for $g = \pi$ being a probability density function, then the following two results hold [Robert, 1999]:*

1. *π is the invariant distribution of the chain;*
2. *the chain is reversible.*

Theorem 2 is an essential building block to develop Markov Chains with a specific invariant distribution, by specifying a suitable transition kernel. However, Markov Chains with invariant distribution without detailed balance condition also exist.

2.4.2 Markov Chain Monte Carlo

Coming back to Bayesian inference, it is possible to consider a **discrete time Markov chain in continuous space** having as invariant (or target) distribution the posterior $f(\boldsymbol{\theta}|\boldsymbol{x})$. The realizations of the chain can be considered as approximate posterior samples from $f(\boldsymbol{\theta}|\boldsymbol{x})$. Those can be used to compute, for example, the expected value of some function through Monte Carlo approximation, or to obtain estimates of the posterior distribution. This procedure is called Markov Chain Monte Carlo (MCMC).

The most common MCMC algorithms use a kernel satisfying the detailed balance condition in order to ensure the correct invariant distribution. Here two of them are presented: **Metropolis-Hastings algorithm** and **Gibbs sampling**.

In **Metropolis-Hastings algorithm**, a kernel satisfying the balance detailed condition is defined by the product of a proposal distribution and an acceptance rate (representing the probability of accepting a new proposal). Metropolis-Hastings algorithm (Algorithm 1) requires us to fix an initial value for the parameters $\boldsymbol{\theta}^0$ and to choose a proposal distribution $Q(\cdot)$ with density $q(\cdot)$.

Algorithm 1 Metropolis-Hastings

1: **for** $j = 1, ..., b$ **do**
2: Choose a value $\boldsymbol{\theta}^*$ near to $\boldsymbol{\theta}^{j-1}$ with density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{j-1})$
3: Compute the **acceptance rate** $\alpha = \min\left[1, \frac{f(\boldsymbol{\theta}^*|\boldsymbol{x})q(\boldsymbol{\theta}^{j-1}|\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}^{j-1}|\boldsymbol{x})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{j-1})}\right]$
4: Generate $u \sim Uniform(0,1)$
5: **if** $u \leq \alpha$ **then**
6: $\boldsymbol{\theta}^j = \boldsymbol{\theta}^*$ (acceptance step)
7: **else**
8: $\boldsymbol{\theta}^j = \boldsymbol{\theta}^{j-1}$ (refusal step)
9: **end if**
10: **end for**

Gibbs sampling is used in the case of multivariate distributions, when one or more of the marginal distributions for $\theta_i,\ i = 1, \ldots, N$ is unknown, when they are known but sampling from them is difficult or when sampling from the joint distribution $f(\theta_1, \theta_2, ..., \theta_N)$ is not possible. In these cases, sampling from conditional distributions can be easier. Through Gibbs sampling, the i^{th} sample at step j is sampled from:

$$f\left(\theta_i^j \mid \theta_1^j, \ldots, \theta_{i-1}^j, \theta_{i+1}^{j-1}, \ldots, \theta_N^{j-1}, \boldsymbol{x}\right).$$

An example of Gibbs sampling can be found in Algorithm 2 for the case $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$; Gibbs sampling requires to choose an initial value for the parameters $\boldsymbol{\theta}^0$.

Gibbs sampling is particularly useful for *hierarchical models* [Robert, 1999], where a conditional structure is present.

Note that with MCMC there is no need to write the full posterior explicitly: in the Metropolis acceptance rate, the marginal $f(\boldsymbol{x})$ cancels, such that you do not need to normalize the posterior with respect to $\boldsymbol{\theta}$; with Gibbs, you only need the full conditionals.

The obtained sequence $\boldsymbol{\theta}^1, ..., \boldsymbol{\theta}^b$ contains b dependent samples, since each sample depends in some ways on the precedent. Therefore, it is important to adopt some procedures

Algorithm 2 Gibbs

1: **for** $j = 1, ..., b$ **do**
2: Sample θ_1^j from $f\left(\theta_1 |\ \theta_2^{j-1}, \theta_3^{j-1}, \boldsymbol{x}\right)$
3: Sample θ_2^j from $f\left(\theta_2 |\ \theta_1^{j}, \theta_3^{j-1}, \boldsymbol{x}\right)$
4: Sample θ_3^j from $f\left(\theta_3 |\ \theta_1^{j}, \theta_2^{j}, \boldsymbol{x}\right)$
5: **end for**

in order to obtain posterior samples that are less correlated and more independent from the initial choice of $\boldsymbol{\theta}^0$. When steps of the chain increase, the distribution is better approximated. However, at the beginning the distribution can be very different from the invariant distribution. The period necessary to converge to the invariant distribution is called **burn-in**. Thus, in order to avoid high dependence on the choice of the initial value $\boldsymbol{\theta}^0$, the initial samples are not considered, in a number that can vary. This procedure is called **burn-in**. Moreover, the procedure of **thinning** can be adopted: it consists in taking into consideration only 1 sample every K samples, where K represents the parameter of thinning. It is also useful to consider more than one chain for each parameter, in order to see if the different chains have similar behaviours and converge to the same target.

2.4.3 Measures of MCMC performance

Statistics can be computed after obtaining the posterior samples, and some measures can be considered in order to evaluate the MCMC performance.

The first one is the **autocorrelation (AC) of samples**, that can be computed at a chosen distance (called *lag*) and is defined as follows:

Definition 11. The autocorrelation of a sample X^n at lag k (**AC.k**) is given by [Venables and Ripley, 2002]:

$$\rho(k) = \frac{c(k)}{c(0)},$$

where

$$c(k) = \frac{1}{N} \sum_{n=\max(1,-k)}^{\min(N-k,N)} \left[X^{n+k} - \mathbb{E}[X]\right] \left[X^n - \mathbb{E}[X]\right].$$

For instance, **AC.10** represents the autocorrelation at lag 10.

Another measure of interest is the effective sample size. To this aim, let us consider a sample of N independent scalar observations $\theta^1, \ldots, \theta^N$ that are identically distributed with mean μ and variance σ^2; in this case, the mean of the distribution is estimated by the mean of the sample:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \theta^i$$

and the variance of $\hat{\mu}$ is the following:

$$\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{N}.$$

On the contrary, if the observations are correlated to each other, $\mathrm{Var}(\hat{\mu})$ is higher:

$$\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{n_{\text{eff}}},$$

where $n_{\text{eff}} < N$ represents the **effective sample size** (**SSeff**) [Freeman, 1966] and indicates how many independent samples correspond to the dependent samples set.

Definition 12. In the case of MCMC samples, SSeff can be defined as:

$$\textbf{SSeff} = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho(k)},$$

where $\rho(k)$ is the autocorrelation of samples at distance k [Kass et al., 1998, Ripley, 1987].

If the observations are independent, they are uncorrelated, so SSeff is equal to the number of observations. If the correlation decreases slowly with respect to the lag k, SSeff is smaller. SSeff is usually between 0 and N [Kass et al., 1998]. Let us consider a number of samples equal to N after burn-in and thinning and let n_0 be the minimum distance at which θ^t and θ^{t+n_0} are approximately independent $\forall t$ (where θ is one of the parameters to be estimated). Then, SSeff can be heuristically approximated as $\frac{N}{n_0}$. If SSeff is low, the posterior estimated mean of parameters is not good, because samples are correlated. Thus, higher values of SSeff are preferred. If we fix the number of samples N, the slower the decay of autocorrelation is, the larger n_0, and the smaller SSeff is.

Other two parameters of interest are **Naive SE** and **Time-series SE**. These are two ways to compute posterior samples standard error: **Naive SE** assumes independent samples, while **Time-series SE** assumes that samples can be dependent. If the observations are effectively independent, these two error measures should be equal.

Chapter 3

Data

Data has been collected in the National Park of Abruzzo, Lazio, and Molise (PNALM), in Italy, from 2005 to 2010 and has been analyzed previously (for instance, by Parracciani 2020). The study area is 1800 km^2 and is covered for about 60% by deciduous forests, followed by pastures and grasslands and agricultural areas. Primary and secondary roads are present in density of 1.1/ km. More details about the configuration of the park and species that populate it can be found in Parracciani [2020]. We only concentrate on the two species of interest for the thesis: **bear** and **wolf**. The latter populate the PNALM area with high densities: for the wolves 5 individuals/100 km^2, for the bears 40 individuals/1000 km^2 are present. Biologists estimate the presence of 8 different packs of wolves that are distributed in different zones of the park, while bears are concentrated in a restricted area [Parracciani, 2020, Ciucci et al., 2015, Mancinelli et al., 2018].

In our work, we use individual locations of bears and wolves, recorded through GPS (Global Positioning System) collars. In particular, for the bear 19 individuals have been recorded, whose 11 were females and 8 males; for the wolf 7 individuals, whose 3 were females and 4 males and attributable to 6 different packs. Different bears and different packs of wolves are identified by an identification code; for the bears, this contains the letter F if the individual is a female and the letter M if it is a male. If an individual has been recorded for two years, it is considered as two different individuals. Moreover, bears and wolves locations are divided into four time-windows: spring (containing locations from March to May), early summer (from June to July), late summer (from August to September), and autumn (from October to December).

Since GPS locations have been originally recorded in different temporal intervals for the two species, they have been filtered in order to obtain a location every 3 hours for both the species, namely 8 locations per day. For each location, the GPS coordinates (latitude and longitude) and the values of some environmental covariates have been recorded. Moreover, the environmental covariates are divided into different categories by Parracciani [2020], as shown in Table 3.1. The *distance to the forest edge* can assume positive or negative values, where positive ones indicate that the animal is out of the forest, negative ones that it is inside the forest.

Besides locations recorded with GPS collars, in our analysis we consider also available locations. These are obtained by sampling in the home range; thus, we conduct our

ENVIRONMENTAL COVARIATES	
CATEGORY	**VARIABLE**
Soil cover	*Percentage of agriculture areas* *Percentage of non-vegetated rocky areas* *Percentage of shrubland* *Percentage of pastures and grasslands*
Forest structure	*Distance to forest edge* *Average density of trees in beech forest* *Average density of trees in oak and hop-hornbeam forest*
Anthropogenic variables	*Distance in metres from urban centres and primary roads* *Distance in metres from urban centres and secondary roads*
Orographic variables	*Terrain ruggedness index* *Hillshade*

Table 3.1. Environmental covariates present in the dataset, divided in different categories by Parracciani [2020].

analysis at the *Third-order selection* of Johnson [1980]. In our case, the home range for each bear and each pack of wolves has been computed by taking the **convex hull** with 100% of the used points. The average dimension of home ranges is 150 km^2 [Mancinelli et al., 2018].

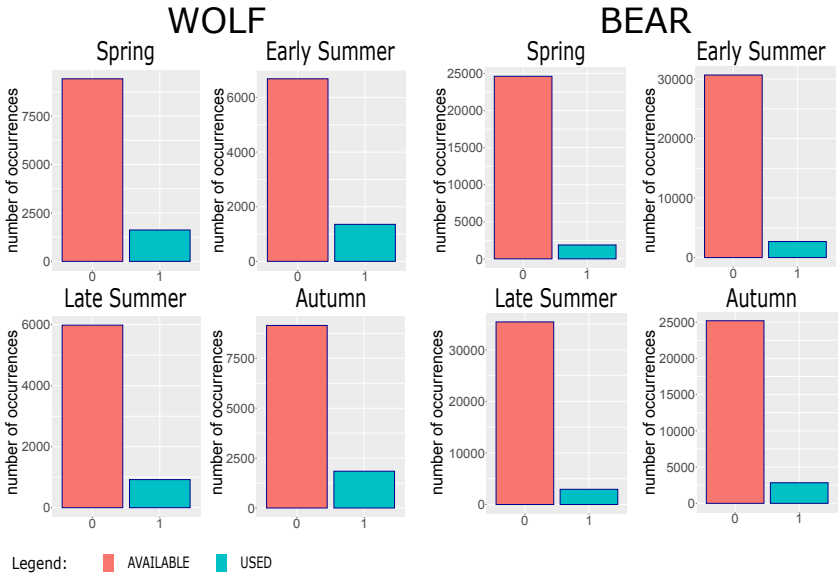

Figure 3.1. Bears and wolves: histograms of number of occurrences; presence locations (used locations) in blue and available locations in red.

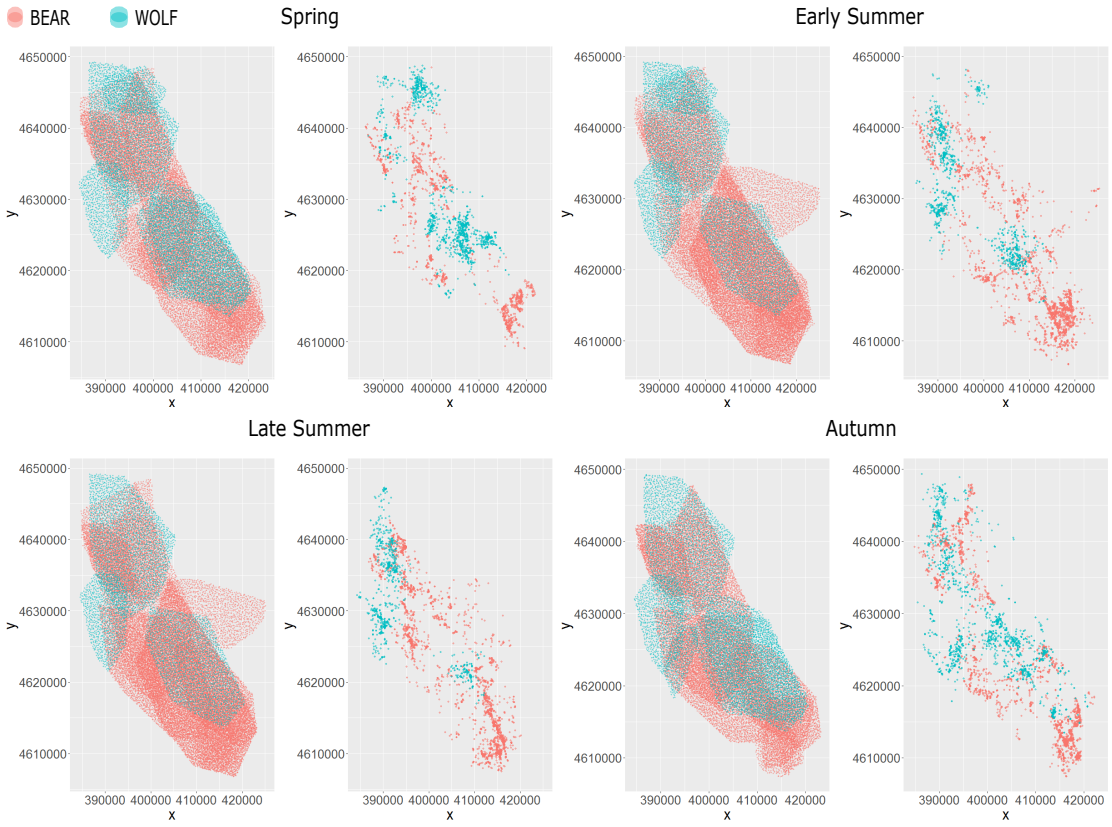

Figure 3.2. Home ranges representation for bears (red) and wolves (blue) in the four time-windows: used locations (right) and sampled available locations (left).

We represent in Figure 3.2 home ranges of bears and wolves, focusing on used and available locations in the four time-windows, where it can be seen that home ranges of bears and wolves sometimes overlap. From the pictures, it is evident that the number of used locations is less than the number of available locations, for both bears and wolves in all the time-windows. We represent in Figure 3.1 some histograms showing the distribution of locations with response variable equal to 1 (used locations, in blue) and equal to 0 (available locations, in red) for both the species and for each time-window. Other histograms representing the number of occurrences of each bear and each pack of wolves can be found in Appendix A.

Chapter 4

Models

In this Chapter, we present the implementation of four different models of resource selection functions for bears and wolves considering not only resources but also interaction between the two species.

First, we consider two uni-lateral models: in the first one, we model the presence of the bear considering only resources, and the presence of the wolf considering both resources and the presence of the bear; on the contrary, in the second one, we model the presence of the wolf considering only resources, and the presence of the bear considering the presence of the wolf beside the resources.

Then, we model the interaction between bear and wolf in the two directions (bear $\leftrightarrow$ wolf) considering two other models. The first of them models the presence of a species in a location considering only resources if the other species is not present, or with constant probability otherwise. The second one takes into consideration attractiveness or repulsiveness with different intensities on a species if the other species is present.

These models are all hierarchical since they include interactions between species and a multi-level structure is present. Resources are in this case represented by the environmental covariates presented in Section 3; their values are standardized, in order to obtain comparable results.

We use a binary response variable (called *Presence*), that assumes value 1 if it is referred to a used point (that is one of the locations recorded through GPS, called also *real location*, or *location of presence*) and value 0 if it is referred to an available point (sampled in the home range). These models remember logistic regression models, where the response variable is binary; however, in this case, the binary response variable does not represent a presence/absence variable: $Presence = 0$ does not mean that the animal is not present in the considered location, but it indicates an available position, namely the animal has never been observed in that location, but we can not exclude that it has passed through that location, since this belongs to its home range. For this reason, locations with response variable equal to 0 are called *pseudo-zeros*. Data involving response variables of this type are called **use-availability data**.

In general, in all models a parameter is associated to each environmental covariate for both the wolf and the bear: sets of parameters associated to environmental covariates are indicated with $\boldsymbol{\beta}^{wolf}$ and $\boldsymbol{\beta}^{bear}$ respectively. Table 4.1 shows the association between an

PARAMETER	ENVIRONMENTAL COVARIATE
β_2	Percentage of shrubland
β_3	Percentage of pastures and grasslands
β_4	Percentage of agricultural areas
β_5	Distance to forest edges
β_6	Average density of trees in beech forest
β_7	Average density of trees in oak and hop-hornbeam forest
β_8	Distance from urban centres and primary roads
β_9	Distance from secondary roads
β_{10}	Terrain ruggedness index
β_{11}	Hillshade
β_{12}	Percentage of non-vegetated rocky areas

Table 4.1. Association of a parameter to each environmental covariate.

environmental covariate and a parameter. Note that $\boldsymbol{\beta}$ can indicate both $\boldsymbol{\beta}^{wolf}$ and $\boldsymbol{\beta}^{bear}$. Moreover, the first component β_1 is associated to the intercept of the models.

The particular response variable for use-availability data has led to difficulties in the interpretation of the parameters. Fieberg et al. [2021] give a guide for the interpretation of parameters in resource selection analysis when we have use-availability data. In particular, Fieberg et al. [2021] argue that the confusion in the interpretation of the parameters is due to using logistic regression in a non-standard way. Binary response variables obtained in the logistic regression approach are usually associated with something that happens (when the response variable is equal to 1) or does not happen (if it is equal to 0), so it is natural to associate a response variable equal to 1 to a presence location and a response variable equal to 0 to an absence location. In resource selection analysis with a logistic regression approach, however, a response variable equal to 0 indicates a pseudo-zero location, namely an available location, as explained above. According to Fieberg et al. [2021], applying logistic regression to use-availability data can be seen as a simple way to estimate the coefficients $\boldsymbol{\beta}$ of a resource selection function. In fact, if the number of pseudo-zero locations is high, the logistic regression parameters converge to ones obtained with a Point Process model [Warton and Shepherd, 2010], which is an exact model, since it is constructed considering only presence data, namely only locations in which the animal has been observed.

Models are run in **JAGS** (Just Another Gibbs Sampler), an **R** package that allows analyzing Bayesian hierarchical models through MCMC simulation (in particular, by using Gibbs sampling). In JAGS, different settings are possible; first, we always consider more

than a single chain to see if the different chains have similar behaviors and converge to the same target. Then, we set a burn-in phase and thinning. Finally, it is usually preferable to allow JAGS with an **adaptive phase**, during which samplers are optimized. Samples obtained in this phase are not considered, since they do not form a Markov chain. The number of adaptation represents the length of the adaptive phase.

4.1 First model: presence of the wolf dependent on presence of the bear

In the First model, we start by modeling the probability of finding the bear in a location considering only environmental covariates, then we model the probability of finding the wolf given the presence of the bear beside the environmental covariates. To this aim, we consider two matrices X^{bear} and X^{wolf} containing the standardized values of the environmental covariates, for the bear and the wolf respectively. Components of vectors $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ represent a measure of the effect of the correspondent environmental covariate on the probability of finding the animal. Moreover, we consider a parameter γ representing the effect of the presence of the bear on the presence of the wolf, two vectors $\boldsymbol{\pi}^{bear}$ and $\boldsymbol{\pi}^{wolf}$ (representing the probability of finding the bear and the wolf in each location), and the vectors $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ (containing response variables for the bear and the wolf in each location).

The first model is defined by, for $i = 1, ..., N_{bear}$:

$$y_i^{bear} \sim Bernoulli(\pi_i^{bear}),$$

$$logit(\pi_i^{bear}) = X_i^{bear}\boldsymbol{\beta}^{bear}$$

and, for $j = 1, ..., N_{wolf}$:

$$y_j^{wolf} \sim Bernoulli(\pi_j^{wolf}),$$

$$logit(\pi_j^{wolf}) = X_j^{wolf}\boldsymbol{\beta}^{wolf} + \gamma\ logit^{-1}(X_j^{wolf}\boldsymbol{\beta}^{bear}).$$

In particular, the matrices X^{bear} and X^{wolf} have a number of columns equal to the number of covariates taken into consideration plus 1 since we add a column of ones used for the intercept in the logistic regression, and have a number of rows equal to the number of observations in the considered dataset. $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are vectors of dimension equal to the number of covariates taken into consideration for the bear and the wolf plus 1, since β_1^{bear} and β_1^{wolf} are the respective intercepts. Together with with γ, $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are the parameters of the model to be estimated. $\boldsymbol{\pi}^{bear}$ and $\boldsymbol{\pi}^{wolf}$ are vectors of dimension equal to the number of locations of the bear and the wolf respectively. π_i^{bear} is the probability of finding the bear in location i. $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ are binary vectors of dimension equal to $\boldsymbol{\pi}^{bear}$ and $\boldsymbol{\pi}^{wolf}$. y_i^{bear} is equal to 1 if the i^{th} location is a presence location (used location); it is equal to 0 if it is an available location. The same happens for y_j^{wolf}. The above model is thus very similar to a standard logistic model, but contains in addition an interaction term between the two species. Its hierarchical structure is represented in Figure 4.1.

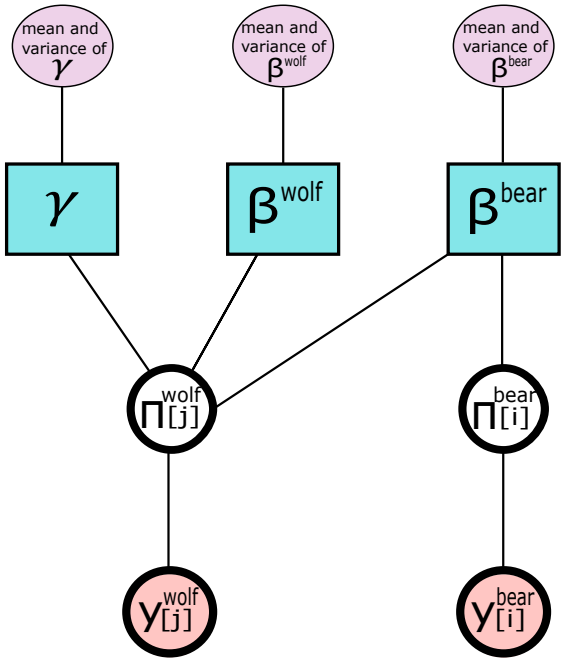

Figure 4.1. First model: representation of the hierarchical structure. i indicates a bear location ($i = 1, ..., N_{bear}$), j indicates a wolf location ($j = 1, ..., N_{wolf}$).

4.2 Second model: presence of the bear dependent on presence of the wolf

In the Second model, the probability of finding the wolf in a location depends only on environmental covariates, while the probability of finding the bear depends on the presence of the wolf besides the environmental covariates. As for the First model, we consider matrices X^{wolf} and X^{bear}, vectors $\boldsymbol{\pi}^{bear}$, $\boldsymbol{\pi}^{wolf}$, $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$, that have the same role assumed in the First model. $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are the parameters associated to the environmental covariates as in the previous model, while in this case the parameter γ represents the effect of the presence of the wolf on the presence of the bear. $\boldsymbol{\beta}^{bear}$, $\boldsymbol{\beta}^{wolf}$ and γ are the parameters to be estimated through the Bayesian model.

The model can be written as, for $i = 1, ..., N_{wolf}$:

$$y_i^{wolf} \sim Bernoulli(\pi_i^{wolf}),$$

$$logit(\pi_i^{wolf}) = X_i^{wolf}\boldsymbol{\beta}^{wolf}$$

and, for $j = 1, ..., N_{bear}$:

$$y_j^{bear} \sim Bernoulli(\pi_j^{bear}),$$

$$logit(\pi_j^{bear}) = X_j^{bear}\boldsymbol{\beta}^{bear} + \gamma\ logit^{-1}(X_j^{bear}\boldsymbol{\beta}^{wolf}).$$

Its hierarchical structure is represented in Figure 4.2.

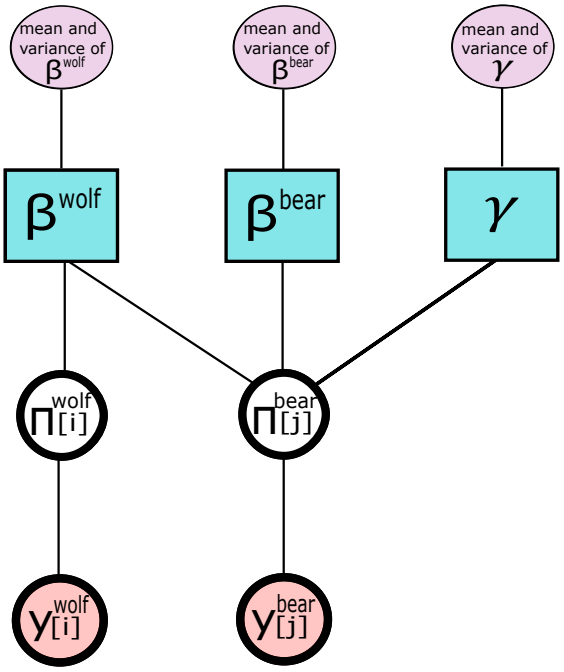

Figure 4.2. Second model: representation of the hierarchical structure. i indicates a wolf location ($i = 1, ..., N_{wolf}$), j indicates a bear location ($j = 1, ..., N_{bear}$).

4.3 Third model: mixture model with constant probability of presence of a species if the other species is present

In the Third model, the probability of finding the bear in a location depends on the presence of the wolf, and vice versa the probability of finding the wolf in a location depends on the presence of the bear. In particular, if the wolf is not present in a considered location, the probability of finding the bear depends only on the environmental covariates in that location, while if the wolf is present, the probability of finding the bear does not depend on the environmental covariates and is a fixed value common to all bear locations in which the wolf is present. The wolf is modelled in a similar fashion.

As for the previous models, matrices X^{bear} and X^{wolf}, vectors $\boldsymbol{\pi}^{bear}$, $\boldsymbol{\pi}^{wolf}$, $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ are considered. The parameters to be estimated are, besides $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ defined as before, γ^{bear}, that represents the probability of finding the bear in any location in which the wolf is present, and γ^{wolf}, that represents the probability of finding the wolf in any location in which the bear is present.

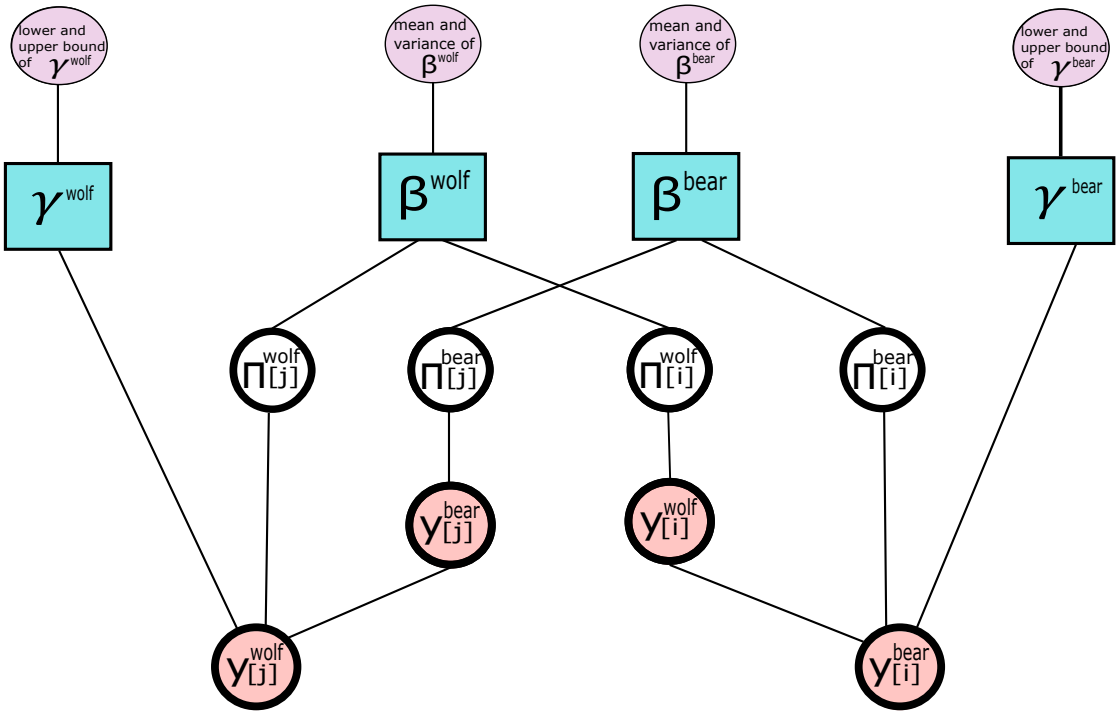

Figure 4.3. Third model: representation of the hierarchical structure. i indicates a bear location ($i = 1, ..., N_{bear}$), j indicates a wolf location ($j = N_{bear} + 1, ..., N_{bear} + N_{wolf}$).

The model is, for $i = 1, ..., N_{bear}$:

$$y_i^{bear} \sim Bernoulli((1 - y_i^{wolf})\pi_i^{bear} + y_i^{wolf}\gamma^{bear}),$$
$$logit(\pi_i^{bear}) = X_i^{bear}\boldsymbol{\beta}^{bear},$$
$$y_i^{wolf} \sim Bernoulli(\pi_i^{wolf}),$$
$$logit(\pi_i^{wolf}) = X_i^{bear}\boldsymbol{\beta}^{wolf}.$$

For $j = N_{bear} + 1, ..., N_{bear} + N_{wolf}$, instead:

$$y_j^{wolf} \sim Bernoulli((1 - y_j^{bear})\pi_j^{wolf} + y_j^{bear}\gamma^{wolf}),$$
$$logit(\pi_j^{wolf}) = X_j^{wolf}\boldsymbol{\beta}^{wolf},$$
$$y_j^{bear} \sim Bernoulli(\pi_j^{bear}),$$
$$logit(\pi_j^{bear}) = X_j^{wolf}\boldsymbol{\beta}^{bear}.$$

In this case, $\boldsymbol{y}^{bear}$, $\boldsymbol{y}^{wolf}$, $\boldsymbol{\pi}^{bear}$ and $\boldsymbol{\pi}^{wolf}$ are vectors of dimension $N_{wolf} + N_{bear}$; their first N_{bear} entries are referred to locations of the bear, the others to locations of the wolf. As for the previous ones, the Third model is hierarchical, as Figure 4.3 shows.

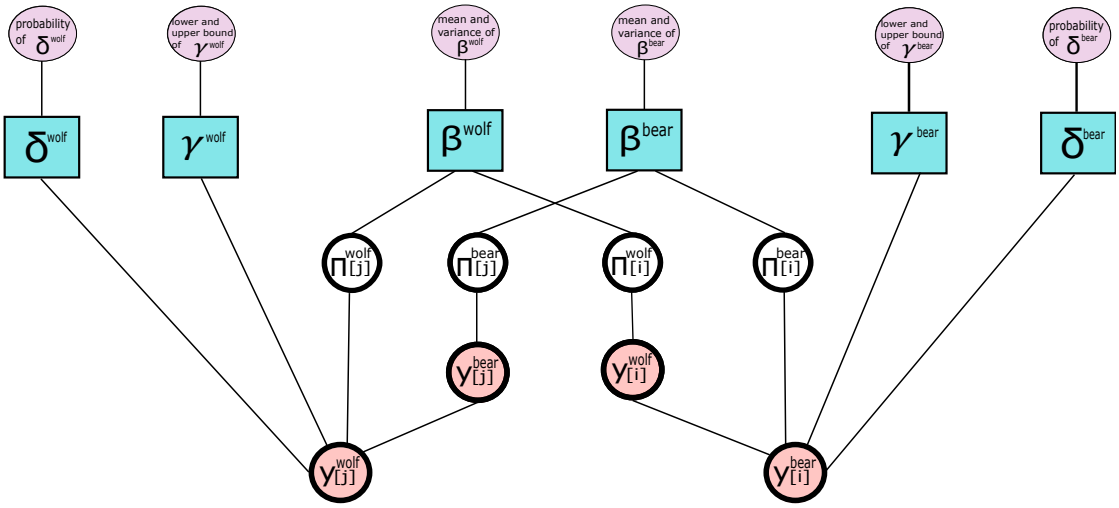

Figure 4.4. Fourth model: representation of the hierarchical structure. i indicates a bear location ($i = 1, ..., N_{bear}$), j indicates a wolf location ($j = N_{bear} + 1, ..., N_{bear} + N_{wolf}$).

4.4 Fourth model: mixture model with attractiveness or repulsiveness between the two species

Finally, we model the probability of finding the bear in a location considering also the presence of the wolf besides the environmental covariates, and vice-versa, as follows: if the wolf is not present in a location, the probability of finding the bear depends only on the environmental covariates; on the contrary, if the wolf is present, there are two possible cases represented by a binary parameter δ^{bear}:

1. $\delta^{bear} = 1$: the presence of the wolf increases the probability of finding the bear. To this aim, we introduce a parameter $\gamma^{bear} \in [0,1]$ indicating the strength of the relationship between wolf and bear: in the extreme $\gamma^{bear} = 0$, there is no effect of the presence of the wolf; on the other hand, $\gamma^{bear} = 1$ means that if the wolf is present, the bear is present (attractiveness).

2. $\delta^{bear} = 0$: the presence of the wolf decreases the probability of finding the bear. This results in the following situation: in the extreme $\gamma^{bear} = 0$, there is no effect of the presence of the wolf; on the other hand, $\gamma^{bear} = 1$ means that if the wolf is present, the bear is not present (repulsiveness).

An analogous construction is done for the wolf according to whether the bear is present or not.

As for the previous models, the different environmental covariates are contained in matrices X^{bear} and X^{wolf}, $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are the parameters associated to them.

Besides $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$, the parameters of the model to be estimated are:

- $\delta^{bear} \in \{0,1\}$ and indicates if there is attractiveness (1) or repulsiveness (0) for the bear with respect to the presence of the wolf;
- $\delta^{wolf} \in \{0,1\}$ and indicates if there is attractiveness (1) or repulsiveness (0) for the wolf with respect to the presence of the bear;
- γ^{bear} represents the strength of the effect of the presence of the wolf on the presence of the bear. If it is equal to 1, there is total attractiveness (if $\delta^{bear} = 1$) or total repulsiveness (if $\delta^{bear} = 0$);
- γ^{wolf} represents the strength of the effect of the presence of the bear on the presence of the wolf. If it is equal to 1, there is total attractiveness (if $\delta^{wolf} = 1$) or total repulsiveness (if $\delta^{wolf} = 0$).

The model is thus, for $i = 1, ..., N_{bear}$:

$$y_i^{bear} \sim Bernoulli(\pi_i^{bear} - y_i^{wolf}\gamma^{bear}\pi_i^{bear} + y_i^{wolf}\gamma^{bear}\delta^{bear}),$$

$$logit(\pi_i^{bear}) = X_i^{bear}\boldsymbol{\beta}^{bear},$$

$$y_i^{wolf} \sim Bernoulli(\pi_i^{wolf}),$$

$$logit(\pi_i^{wolf}) = X_i^{bear}\boldsymbol{\beta}^{wolf};$$

for $j = N_{bear} + 1, ..., N_{bear} + N_{wolf}$, instead:

$$y_j^{wolf} \sim Bernoulli(\pi_j^{wolf} - y_j^{bear}\gamma^{wolf}\pi_j^{wolf} + y_j^{bear}\gamma^{wolf}\delta^{wolf}),$$

$$logit(\pi_j^{wolf}) = X_j^{wolf}\boldsymbol{\beta}^{wolf},$$

$$y_j^{bear} \sim Bernoulli(\pi_j^{bear}),$$

$$logit(\pi_j^{bear}) = X_j^{wolf}\boldsymbol{\beta}^{bear}.$$

In Figure 4.4 the hierarchical structure of the model is represented.

As for the Third model, $\boldsymbol{y}^{bear}$, $\boldsymbol{y}^{wolf}$, $\boldsymbol{\pi}^{bear}$ and $\boldsymbol{\pi}^{wolf}$ are vectors of dimension $N_{wolf} + N_{bear}$; their first N_{bear} entries are referred to locations of the bear, the others to locations of the wolf.

Chapter 5

Simulation study

Before proceeding with applying the models to the real dataset, it is important to evaluate the goodness of the models, by considering two aspects: their identifiability and the impact of the number of pseudo-zeros on the parameter estimates.

5.1 Identifiability of the models

Definition 13. A model is **identifiable** if, letting f be its likelihood:

$$f(\boldsymbol{x}|\boldsymbol{\theta_1}) = f(\boldsymbol{x}|\boldsymbol{\theta_2}) \; \forall \; \boldsymbol{x} \implies \boldsymbol{\theta_2} = \boldsymbol{\theta_1}.$$

It is important to check that the models presented in Sections 4.1, 4.2, 4.3 and 4.4 are identifiable. In fact, the aim is to obtain a set of parameters that express the effect of each covariate on the probabilities of the presence of the wolf and the bear and their interaction, thus these parameters have to be univocal.

We check the identifiability of the models by considering a synthetic dataset with lower dimensions than the real one, in order to reduce computational time. In particular, for each model, the values for the environmental covariates are sampled from a Normal distribution (with mean equal to 0 and variance equal to 1, in order to have values of similar size), and are inserted in the matrices X^{bear} and X^{wolf}. Moreover, we arbitrarily choose a set of values for the parameters. Then, the correspondent values of response variables $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ are generated via the model. The resulting X^{bear}, X^{wolf}, $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ are used as dataset for the model in JAGS and posterior MCMCs for the parameters are obtained. Finally, statistics on the posteriors are considered in order to check if the arbitrarily chosen parameters values (thereafter, the *true* parameter values) are well estimated or not: in particular, we check whether the 95% posterior credible intervals contain the true parameter value.

For all models we choose relatively small values of $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$, usually in the range $[-2,2]$. In Figure 5.1, we can see that if we take too large values for the logit, the probabilities will be near 1 and will assume very similar values. Numerically, this makes distinguishing between close parameter values hard. In our case, models are logistic with some modifications and the values of environmental covariates are low, so choosing

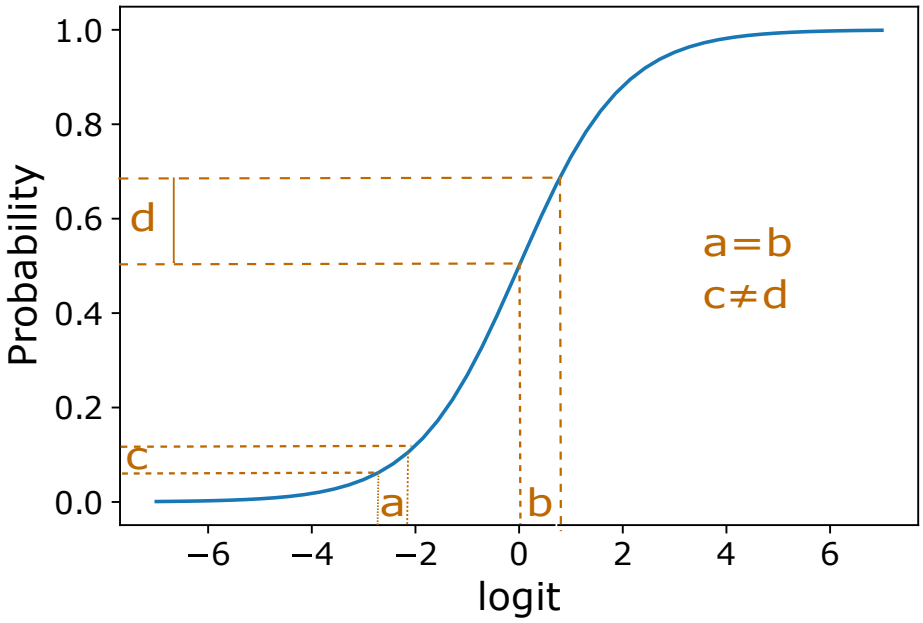

Figure 5.1. Probabilities values with respect to logit values.

relatively small values for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ is better in order to have a large range of probabilities.

For the Third and Fourth model, we also find that allowing a large prior range leads to non-identifiability of the parameters, likely due to the presence of a different set of parameters with large numerical value which have same likelihood as the true ones. We have tried to analytically prove the non-identifiability of the Third and the Fourth models, without succeeding in general; however, a proof of the non-identifiability of these models in particular cases ($\gamma^{bear} = \gamma^{wolf} = 1$, and $\gamma^{bear} = \gamma^{wolf} = \delta^{bear} = \delta^{wolf} = 1$ respectively) can be found in Appendix B.

5.1.1 Identifiability of the First model

In order to check the identifiability of the First model, only 3 environmental covariates are taken into consideration, even if the number of effective covariates is larger in the real dataset. Nevertheless, if the model is identifiable, the identifiability should not change by taking a larger number of covariates. The choice of using a small number of covariates in the check of identifiability is made for computational reasons. Matrices X^{wolf} and X^{bear} have 4 columns in this case. The arbitrarily chosen values of the parameters are $\gamma = -1$, $\boldsymbol{\beta}^{bear} = (2, 1, -1.5, 1.7)$, $\boldsymbol{\beta}^{wolf} = (1.3, -1.5, 0.3, 2)$. A number of observations equal to 500 is fixed for both bears and wolves. In order to be sure that the chains reach convergence, we considered different JAGS settings; in particular, we fixed a number of chains equal to 3, we took 7000 samples for chain, we set an adaptation phase of 2000 samples and a

AC.10			
Parameter	**thinning=1**	**thinning=2**	**thinning=20**
γ	0.874	0.746	0.097
β_1^{bear}	0.048	0.013	0.008
β_2^{bear}	0.048	0.033	0.002
β_3^{bear}	0.069	0.04	0.01
β_4^{bear}	0.05	0.019	-0.003
β_1^{wolf}	0.871	0.746	0.097
β_2^{wolf}	0.19	0.153	0.032
β_3^{wolf}	0.516	0.432	0.05
β_4^{wolf}	0.566	0.486	0.068

Table 5.1. Autocorrelation between samples at distance 10 for the parameters, running the model with different values of thinning.

burn-in equal to 1000 (in order to reduce dependence on initial values of the chain) and we tried different values for thinning: $thinning = 1$ (which corresponds to no thinning), $thinning = 2$ and $thinning = 20$. Prior distributions of the parameters (γ, $\boldsymbol{\beta}^{bear}$, $\boldsymbol{\beta}^{wolf}$) are $Normal(0,10^7)$.

We compute the measures of MCMC performance presented in Section 2.4.3. Autocorrelation between samples at distance 10 is very high for some parameters (as for instance γ and β_1^{wolf}) when thinning is low, but if we run again the model increasing the value of thinning, the autocorrelation between samples at distance 10 decreases, as Table 5.1 shows. In Figure 5.2, we represent the running mean and the autocorrelation decay with respect to the lag for γ; the behaviours of the other problematic parameters are similar. In general, it is clear that by taking a larger value of thinning the autocorrelation decays faster. Nevertheless, taking a large value of thinning is possible in this case because the dataset is not too large, which results in a relatively small computational time, but it could be not always possible with a larger dataset.

Another measure of interest is the Effective Sample Size (*SSeff*). If we fix the number of samples, the slower the autocorrelation decays, the smaller *SSeff* is. In Table 5.2 we represent how *SSeff* changes for the parameters by changing the value of thinning. γ, β_1^{wolf}, β_3^{wolf} and β_4^{wolf} are the parameters that, in general, have the lowest values of *SSeff* for each value of thinning considered. This confirms the more problematic behaviours of their chains with respect to other parameters chains.

Two other parameters of interest are Naive SE and Time-series SE, where the first one computes posterior samples standard error assuming independent samples, while the

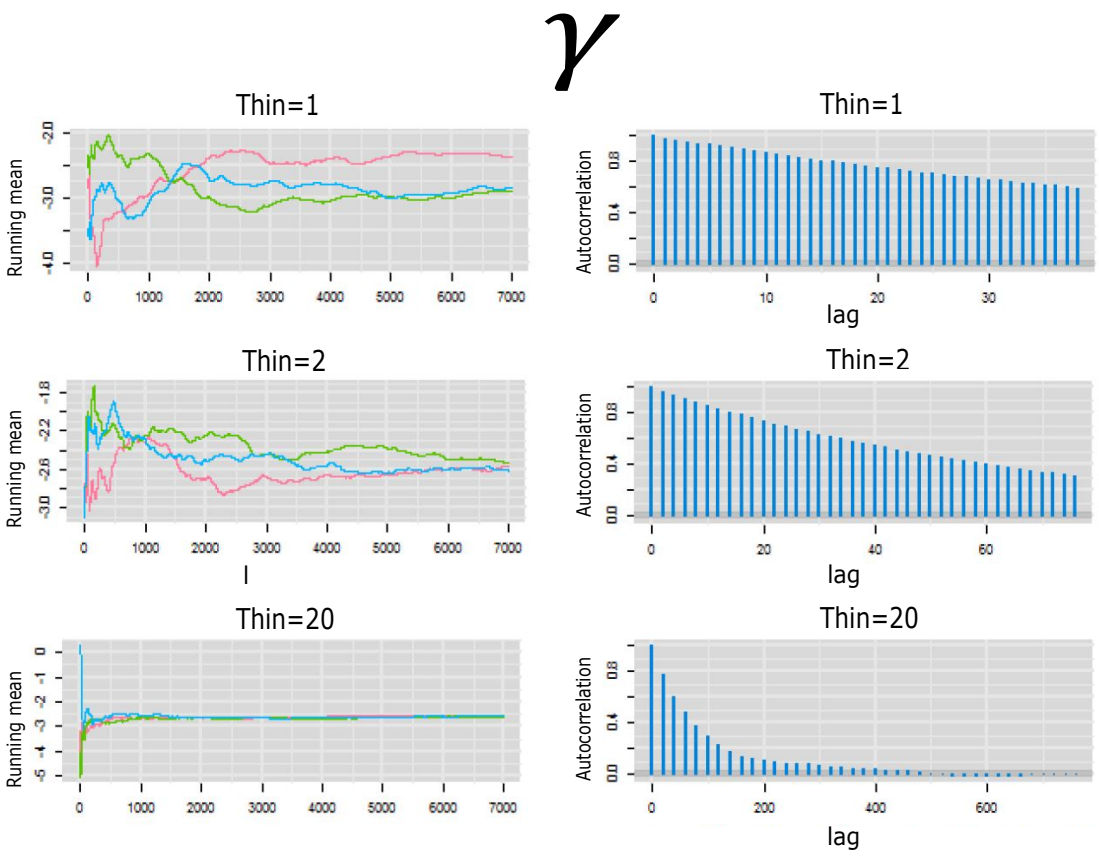

Figure 5.2. *Left*: Posterior running mean of the 3 chains for γ with different values of thinning. *Right*: Autocorrelation decay with respect to lag for γ by taking different values of thinning.

second one considers that samples can be dependent. We look for equal (or at least very similar) measures for these two errors. In Table 5.3 the values of the two errors measures are shown for different values of thinning. In particular, looking at these parameters for the case of no thinning ($thinning = 1$), we can confirm that thinning is necessary, as the two measures are different, mostly for γ, β_1^{wolf}, β_3^{wolf} and β_4^{wolf}, for which they have different orders of magnitude. By taking $thinning = 2$, the two errors become more similar, but improvements can still be done. Finally, with $thinning = 20$ the two errors become very similar for all the parameters. In Table 5.4 the absolute value of the difference between Naive SE and Time-series SE for the parameters with the different values of thinning is shown: it decreases for all the parameters if we run the model with a larger value of thinning. For γ, β_1^{wolf}, β_3^{wolf} and β_4^{wolf} the order of magnitude remains larger than for the other parameters, but Naive SE and Time-series SE are more similar with $thinning = 20$ than with the other values of thinning.

EFFECTIVE SAMPLE SIZE			
Parameter	**thinning=1**	**thinning=2**	**thinning=20**
γ	140	312	2602
β_1^{bear}	3882	7165	19731
β_2^{bear}	4942	8737	17797
β_3^{bear}	3929	5568	16710
β_4^{bear}	3862	7787	20248
β_1^{wolf}	137	301	2557
β_2^{wolf}	1192	1814	9023
β_3^{wolf}	284	626	4082
β_4^{wolf}	275	499	3525

Table 5.2. Effective Sample Size value for the parameters, running the model with different values of thinning.

	thinning=1		**thinning=2**		**thinning=20**	
Parameter	**Naive SE**	**Time-series SE**	**Naive SE**	**Time-series SE**	**Naive SE**	**Time-series SE**
γ	0.0083	0.1011	0.0079	0.0651	0.0084	0.0238
β_1^{bear}	0.0012	0.0028	0.0012	0.002	0.0012	0.0012
β_2^{bear}	0.001	0.0021	0.001	0.0016	0.001	0.0011
β_3^{bear}	0.0012	0.0029	0.0012	0.0024	0.0012	0.0014
β_4^{bear}	0.0012	0.0029	0.0012	0.002	0.0012	0.0013
β_1^{wolf}	0.0062	0.0756	0.0058	0.0489	0.0062	0.0177
β_2^{wolf}	0.0016	0.0067	0.0016	0.0054	0.0016	0.0024
β_3^{wolf}	0.0015	0.0131	0.0015	0.0086	0.0015	0.0035
β_4^{wolf}	0.0022	0.0199	0.0021	0.014	0.0022	0.0054

Table 5.3. Comparison between Naive SE and Time-series SE for different values of thinning.

Since we have seen that it is better to set a large value of thinning, we consider *thinning* = 20 and we look at the parameters posterior densities. In order to see if parameters are well estimated, we check if the true values of parameters (namely the ones arbitrarily chosen at the beginning) are included in the 95% credible intervals of the posteriors. Figure 5.3 shows that parameters are well estimated: in each plot, the red line represents the true value of the correspondent parameter, the blue lines are the two extremes of the 95% credible intervals (lower and upper bound). The model can be

Parameter	$Thin = 1$	$Thin = 2$	$Thin = 20$
γ	$9.2753 \cdot 10^{-2}$	$5.7252 \cdot 10^{-2}$	$1.5418 \cdot 10^{-2}$
β_1^{bear}	$1.617 \cdot 10^{-3}$	$8.52 \cdot 10^{-4}$	$3.8 \cdot 10^{-5}$
β_2^{bear}	$1.116 \cdot 10^{-3}$	$5.92 \cdot 10^{-4}$	$9.3 \cdot 10^{-5}$
β_3^{bear}	$1.644 \cdot 10^{-3}$	$1.164 \cdot 10^{-3}$	$1.48 \cdot 10^{-4}$
β_4^{bear}	$1.66 \cdot 10^{-3}$	$7.89 \cdot 10^{-4}$	$2.4 \cdot 10^{-5}$
β_1^{wolf}	$6.9439 \cdot 10^{-2}$	$4.3125 \cdot 10^{-2}$	$1.1513 \cdot 10^{-2}$
β_2^{wolf}	$5.121 \cdot 10^{-3}$	$3.843 \cdot 10^{-3}$	$8.42 \cdot 10^{-4}$
β_3^{wolf}	$1.1623 \cdot 10^{-2}$	$7.144 \cdot 10^{-3}$	$1.937 \cdot 10^{-3}$
β_4^{wolf}	$1.77 \cdot 10^{-2}$	$1.19 \cdot 10^{-2}$	$3.2 \cdot 10^{-3}$

Table 5.4. Absolute value of the difference between Naive SE and Time-series SE for different values of thinning.

considered identifiable with this dataset.

5.1.2 Identifiability of the Second model

Checking the identifiability of the Second model on a synthetic dataset is not necessary, since it is symmetric with respect to the First model. Positions of the bear and the wolf are simply inverted in the Second model with respect to the First; thus, since the dataset has been sampled and parameters have been arbitrarily chosen, also the identifiability of the Second model is ensured.

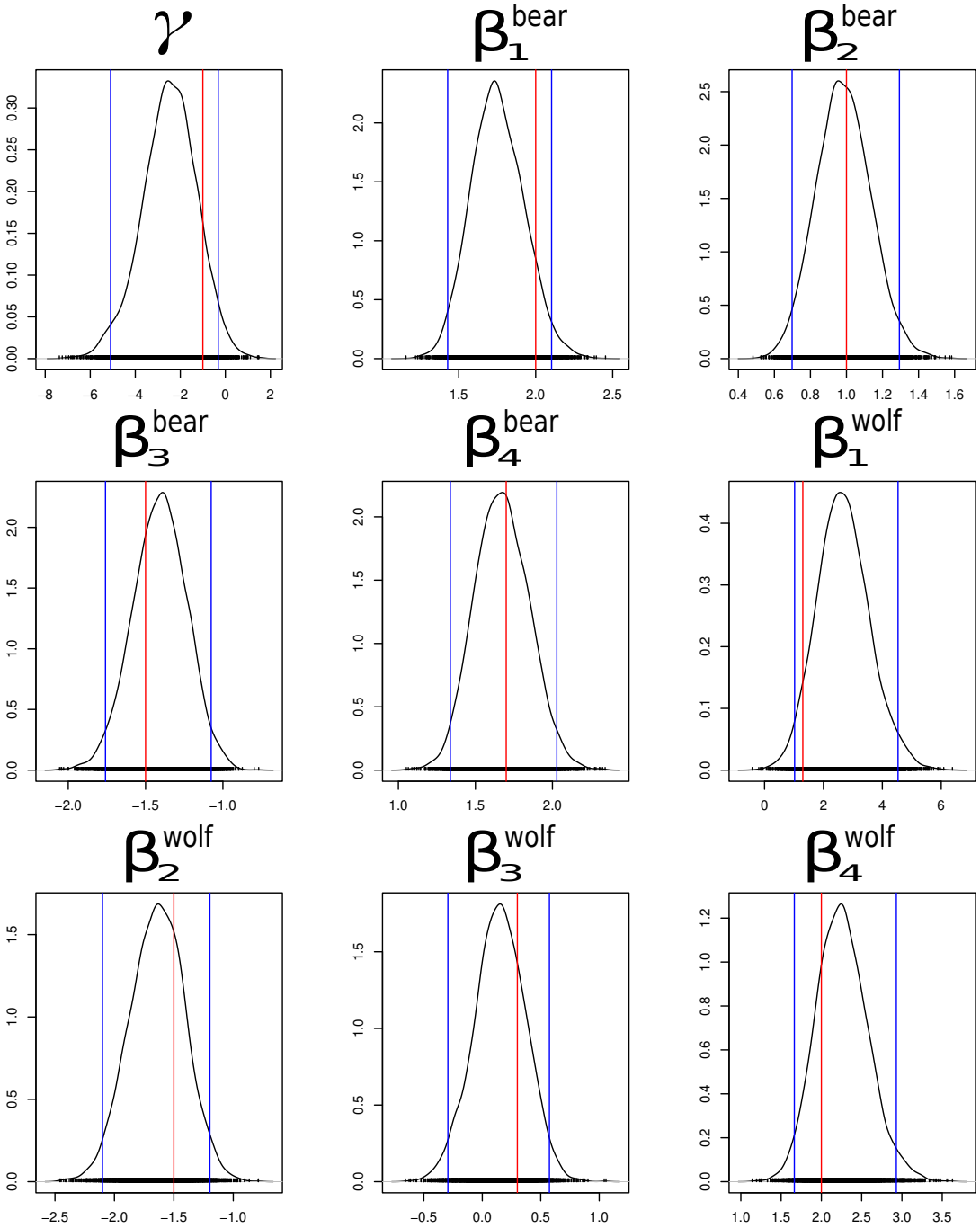

Figure 5.3. **Identifiability** of the First model on a synthetic dataset: density plots for the parameters. The red line represents the true value of the parameter, the blue lines represent the 95% credible intervals extremes. Case of *thinning* = 20.

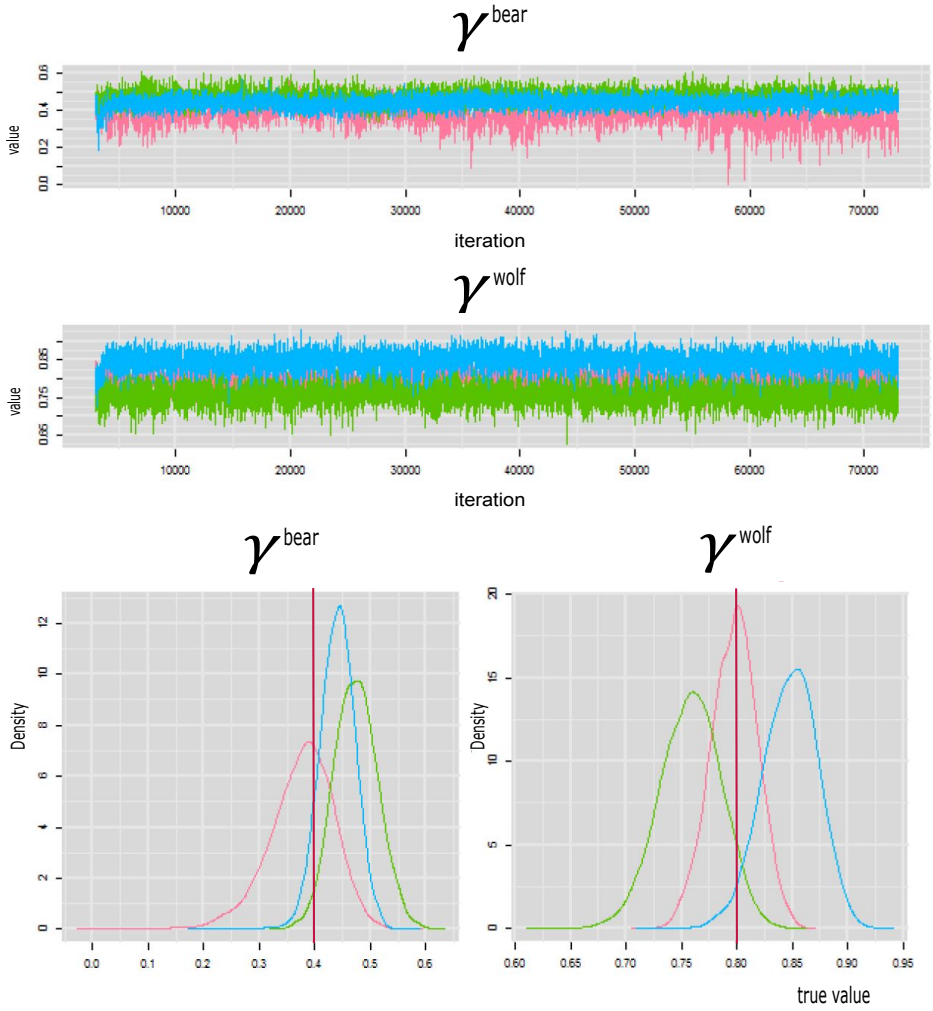

Figure 5.4. Third model: values and densities of the 3 posterior chains for γ^{bear} and γ^{wolf}. Priors for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ were $Normal(0,10^7)$.

5.1.3 Identifiability of the Third model

As for the First model, identifiability of this model is initially checked considering a synthetic dataset. In particular, 3 environmental covariates are considered and their values are sampled from a Normal distribution with mean equal to 0 and variance equal to 1. Thus, X^{wolf} and X^{bear} have 4 columns, where the first one has all the values equal to 1 and is used for the intercept, the other three are referred to the environmental covariates. 400 observations are sampled for the wolf and 500 for the bear: X^{wolf} and X^{bear} have a number of rows equal to 400 and 500, respectively. Regarding the parameters, we arbitrarily chose the following values: $\gamma^{bear} = 0.4$, $\gamma^{wolf} = 0.8$, $\boldsymbol{\beta}^{bear} = (2.3, -0.1, -0.5, 1.4)$ and $\boldsymbol{\beta}^{wolf} = (1, -1, 1.9, 0.2)$. Also in this case, the values of $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ are simulated from

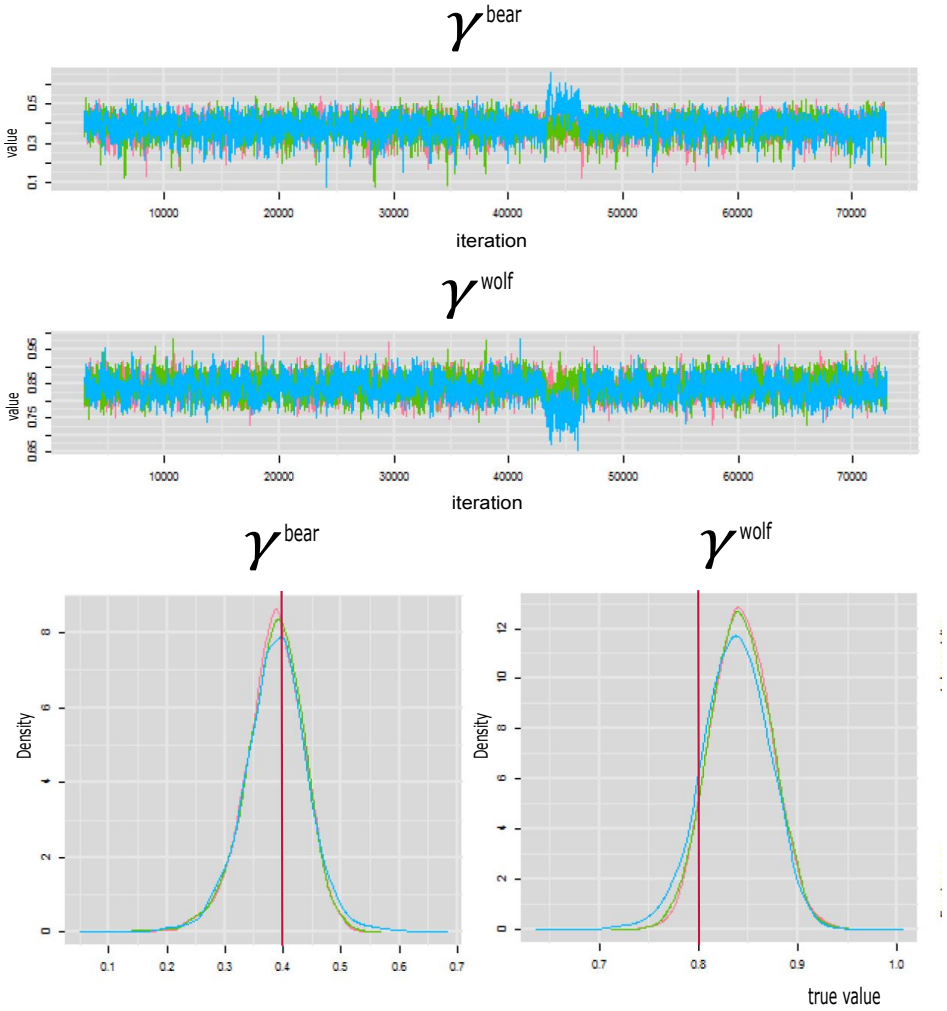

Figure 5.5. Third model: values and densities of the 3 posterior chains for γ^{bear} and γ^{wolf}. Priors for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ were $Normal(0,10)$.

the model considering X^{bear} and X^{wolf} sampled before and the arbitrarily chosen values of the parameters. The model is run in JAGS and posterior estimates of the parameters are obtained and used to check if the parameters are well estimated or not. The sampling is done with JAGS using 3 chains with 7000 samples each chain; additionally, we use a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and $thinning = 10$. For γ^{bear} and γ^{wolf} the prior distributions are taken $Uniform(0,1)$, since these are probabilities and we do not have any previous information on them. For $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$, we considered different prior distributions, less or more informative.

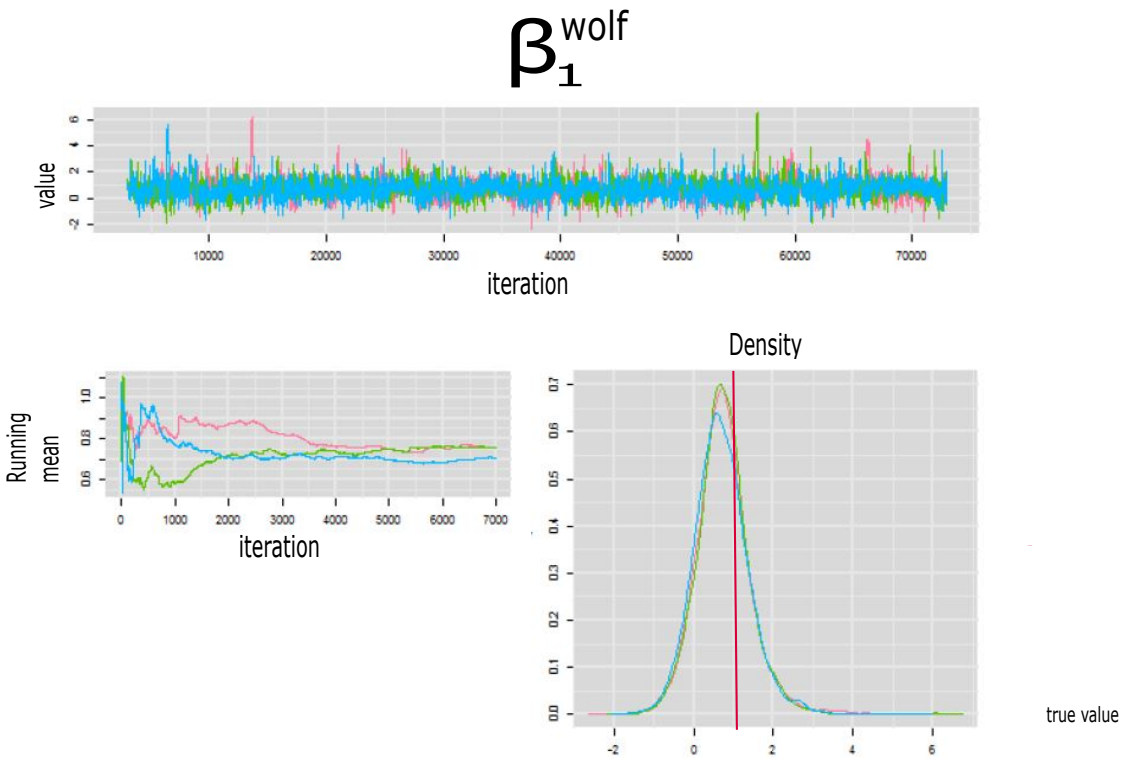

Figure 5.6. Third model: representation of posteriors for β_1^{wolf}. Priors for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ were $Uniform(-10,10)$.

If we take very large priors like $Normal(0,10^7)$ for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$, for some parameters the posterior chains do not converge, as Figure 5.4 shows for γ^{bear} and γ^{wolf}[1]. If we reduce the variance of the Normal distribution used as prior for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ and we take, for instance, a $Normal(0,100)$ for $\boldsymbol{\beta}^{bear}$, $\boldsymbol{\beta}^{wolf}$, posterior chains do not stabilize, even if the situation is slightly better than the previous case. Finally, if we reduce again the variance of the prior distributions, for instance taking $Normal(0,10)$ as prior for $\boldsymbol{\beta}^{bear}$, $\boldsymbol{\beta}^{wolf}$, chains are more stabilized, as Figure 5.5 shows, and the parameters are well estimated (their true values are included in the 95% credible intervals of the posteriors). Moreover, looking at the densities of the posterior chains represented in Figures 5.4 and 5.5, it is clear that if we decrease the value of the prior variance, the densities of the 3 posterior chains are closer.

We now consider Uniform priors for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$. In theory, uniform distributions on large intervals are preferable, as they are less informative. A good alternative to them

[1]A choice for the priors on $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ has a consequence on the posterior estimates of other parameters.

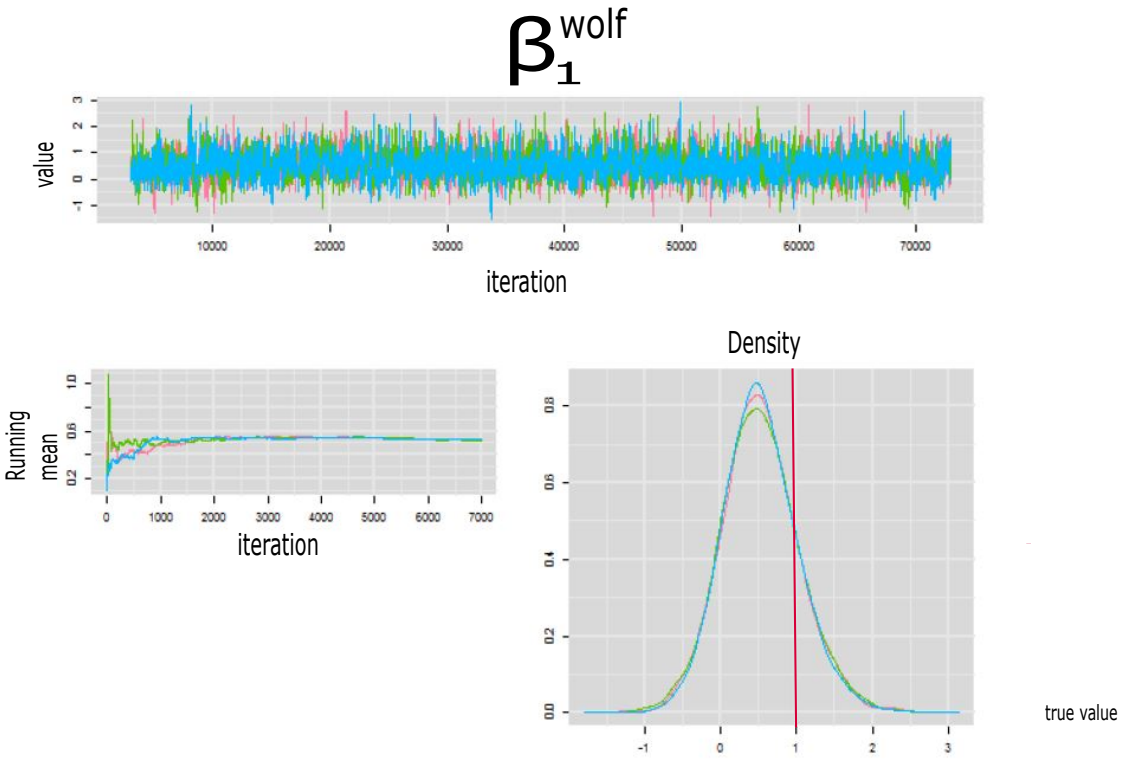

Figure 5.7. Third model: representation of posteriors for β_1^{wolf}. Priors for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ were $Uniform(-3,3)$.

could be a normal distribution with a very large variance, but, as described above, for this model taking a large prior variance results in a non convergence of the chains. Setting priors for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ to be $Uniform(-10,10)$ leads to non-convergent posterior chains, as Figure 5.6 shows for β_1^{wolf}. In particular, the running means of the 3 chains take several iterations to converge to a single value and the blue chain tends to a slightly different value from the others. Figure 5.6 shows the behaviour of β_1^{wolf}, but the situation is similar for the other parameters. If we reduce the support of the $Uniform$ distribution, for instance by taking the prior for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ to be $Uniform(-5,5)$, the situation improves for most of the parameters. If we reduce again the support considering the prior $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ to be $Uniform(-3,3)$, chains are even better and converge around a single value for all the parameters. Moreover, all the true values of the parameters are included in the correspondent 95% credible intervals of the posteriors. In order to make a comparison with the case of priors $Uniform(-10,10)$ (Figure 5.6), we represent in Figure 5.7 posterior chains for β_1^{wolf}: this parameter is identifiable, the 3 chains converge and their running means tend to a single value.

To summarize, we represent in Table 5.5 the results obtained with the different prior distributions for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$. In general, this shows that the chains reach convergence

Priors	Convergence	Priors	Convergence
$Normal(0,10^7)$	no	$Uniform(-10,10)$	no
$Normal(0,10^2)$	no	$Uniform(-5,5)$	for some parameters only
$Normal(0,10)$	yes	$Uniform(-3,3)$	yes

Table 5.5. Third model. Convergence of the chains for the parameters with different prior distributions for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$.

if we take as priors uniform distributions centred in 0 and with a not too large support, or normal distributions with low variance. In these cases, the model seams identifiable. Likely, for this model there exists another set of parameters with similar likelihood as the true one, but this has very large values, thus reducing the support of the prior distribution avoids this issue.

5.1.4 Identifiability of the Fourth model

As for the previous models, the dataset used to check the identifiability of the Fourth model is composed by covariates sampled from a *Normal* distribution (with mean equal to 0 and variance equal to 1). The latter are inserted in matrices X^{bear} and X^{wolf}. In this case, the arbitrarily chosen values of the parameters are: $\gamma^{bear} = 0.5$, $\gamma^{wolf} = 0.8$, $\delta^{bear} = 1$, $\delta^{wolf} = 0$, $\boldsymbol{\beta}^{bear} = (2.3, -0.1, -0.5, 1.4)$, $\boldsymbol{\beta}^{wolf} = (1, 0.5, 1.9, 0.2)$. In order to check the identifiability of the Fourth model, we follow the same procedure of the previous models, sampling values of the response variables and using them with X^{bear} and X^{wolf} to retrieve posteriors estimates of the parameters through JAGS. Also in this case, the synthetic dataset is small for computational reasons: only 3 covariates, 500 observations for the bear and 400 for the wolf are taken into consideration. Prior distributions for δ^{bear} and δ^{wolf} are $Bernoulli(0.5)$, for γ^{bear} and γ^{wolf} are $Uniform(0,1)$, while for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ different priors are considered.

As for the Third model, we first consider a normal distribution with large variance: the prior for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ is taken to be $Normal(0,10^7)$. In this case, chains do not converge. If we reduce the variance of the normal distribution to 100, we obtain both convergent chains and good estimates of the parameters: the model is identifiable with this choice of prior distributions. Similarly, with $Normal(0,10)$ priors for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$, parameters are well estimated and the chains converge into a single target value.

On the contrary, taking $Uniform(-10,10)$ priors leads to non convergent chains for $\boldsymbol{\beta}^{wolf}$. If we reduce the support to $(-5,5)$ or to $(-3,3)$, chains converge for all the parameters, which are also well estimated. Table 5.6 shows that the chains reach convergence and the model seems identifiable if we take as prior distributions Uniform distributions centred in 0 and with a not too large support, or Normal distributions with low variance. Likely, as for the Third model, there exists another set of parameters with similar likelihood as the true one, but this has very large values, thus reducing the support of the prior distribution avoids this issue.

Priors	Convergence	Priors	Convergence
$Normal(0,10^7)$	no	$Uniform(-10,10)$	no
$Normal(0,10^2)$	yes	$Uniform(-5,5)$	yes
$Normal(0,10)$	yes	$Uniform(-3,3)$	yes

Table 5.6. Fourth model. Convergence of the chains for the parameters with different prior distributions for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$.

5.2 Effect of the number of pseudo-zeros

Another important aspect to consider is the number of pseudo-zeros in the dataset. In fact, the number of available locations used in an RSF model is arbitrarily decided by the researcher: in our case, available locations have been sampled in the home range of the animal. If the number of available locations is low with respect to the number of observed locations (locations with response variable equal to 1, used locations), problems in the estimates of the parameters may arise (for instance, parameters estimates could change). It is necessary to check if the estimates of the parameters through the considered models are dependent on the number of the sampled available points compared to the number of used locations. In particular, if the number of pseudo-zeros is large enough, the estimates of the parameters obtained with the logistic regression approach converge to the estimates of the parameters obtained with a Poisson Point Process method, which is an exact method based only on presence locations [Warton and Shepherd, 2010]. Thus it is necessary to show that the considered number of available points is enough to guarantee the stability of the parameter estimates: more precisely, if we slightly decrease the number of available points, the estimates do not change.

For instance, Fieberg et al. [2021] show an example by using standard logistic regression and varying the number of available points from 1 available point for each used point to 100 available points for each used point: results show that by taking a number of pseudo-zeros larger than a threshold (that in his case was 10 available locations per used location) the parameters estimates stabilize. On the contrary, if the total number of pseudo-zeros is lower than the threshold, the estimates of the parameters change by taking different proportion of available points with respect to used points. Moreover, Fieberg et al. [2021] show that in the standard logistic regression the intercept has a different behaviour with respect to the other parameters: its estimate changes by increasing the number of available points.

For each model and for each dataset (corresponding to the different time-windows), it is important then to check if the number of pseudo-zeros is enough to guarantee the convergence of the posterior chains and, if this is verified, the goodness of the parameters estimates. Thus, for each considered model and for each dataset we eliminate some percentages of pseudo-zeros and we check for any changes in the convergence of the posterior

SEASON	**SPRING**		**EARLY SUMMER**	
SPECIES	**Bear**	**Wolf**	**Bear**	**Wolf**
Total number of observations	26509	11047	33374	8039
Number of observations equal to 1	1876	1619	2698	1355
Number of observations equal to 0	24633	9428	30676	6684
Percentage of observations equal to 1	7.08%	14.66%	8.08%	16.86%
Percentage of observations equal to 0	92.92%	85.34%	91.92%	83.14%
SEASON	**LATE SUMMER**		**AUTUMN**	
SPECIES	**Bear**	**Wolf**	**Bear**	**Wolf**
Total number of observations	38381	6906	28037	10989
Number of observations equal to 1	2949	921	2840	1843
Number of observations equal to 0	35432	5985	25197	9146
Percentage of observations equal to 1	7.68%	13.34%	10.13%	16.77%
Percentage of observations equal to 0	92.32%	86.66%	89.87%	83.23%

Table 5.7. Statistics on the number of available/used locations in the different time-windows. Note that *equal to* has to be intended as *with response variable* $\boldsymbol{y}$ *equal to*.

chains and in the estimates of the parameters. In this way, we find a threshold above which the percentage of discarded pseudo-zeros leads to bad estimates of the parameters or non-convergent chains. Given a model, this procedure is implemented for each dataset (both the simulated one and those related to each different time-window), since in each dataset the percentage of pseudo-zeros and their distribution can change.

5.2.1 Number of pseudo-zeros for the First model

First, we consider the synthetic dataset used in Section 4.1 and we randomly eliminate the following percentages of pseudo-zeros: 10%, 25%, 40%, 50%, 75%. For each percentage of discarded pseudo-zeros, we run the model in JAGS in order to check if the estimates of the parameters change. JAGS settings are in this case: 3 chains, 7000 samples for chain, a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and $thinning = 20$. Priors for all the parameters are $Normal(0,10^7)$. If we eliminate 40% or less of the pseudo-zeros, parameters continue to be well estimated, while if we eliminate 50% or 75% of the pseudo-zeros the true value of the intercept β_1^{bear} is not included in the 95% credible interval of the posterior. In the case of 50% of discarded pseudo-zeros, the distance between true value of β_1^{bear} and the lower bound of the interval is equal to 0.15575, in the case of 75% of discarded pseudo-zeros, this becomes equal to 0.8306. Moreover, with

	Bear		Wolf	
SEASON	Original dataset	Reduced dataset	Original dataset	Reduced dataset
SPRING	7.08%	7.13%	14.66%	14.66%
EARLY SUMMER	8.08%	8.09%	16.86%	16.92%
LATE SUMMER	7.68%	7.73%	13.34%	13.46%
AUTUMN	10.13%	10.16%	16.77%	16.83%

Table 5.8. Percentages of used locations in the original dataset and in the reduced dataset (1 observation every 30 for the bear and 1 observation every 10 for the wolf).

75% of discarded pseudo-zeros β_1^{wolf} as well is not well estimated. Nevertheless, the true values of the other parameters are included in the respective 95% credible intervals also with a low percentage of original pseudo-zeros. Thus, the theory of Fieberg et al. [2021] is confirmed for the First model on the synthetic dataset, even if this involves interactions between the two species: if the number of pseudo-zeros is larger than a threshold, the parameter estimates are not affected by the number of pseudo-zeros. Note that in our case the initial percentage of pseudo-zeros was 42.6% for the wolf and 24% for the bear: these percentages are not obtained for a particular reason, but simply because $\boldsymbol{y}^{wolf}$ and $\boldsymbol{y}^{bear}$ have been sampled through the model starting by values of the environmental covariates sampled by $Normal(0,1)$ and arbitrary chosen parameters values.

We now apply the same procedure to the real datasets with all the covariates given in Table 4.1, since, as already mentioned, the number and the distribution of available locations change in each dataset. In Table 5.7 some statistics on the number of pseudo-zeros in the different time-windows are shown. For computational reasons, the original dataset is "reduced" considering an observation every 30 for the bear and an observation every 10 for the wolf. Taking a smaller number of observations should not change the situation, since the percentages of available and used locations are very similar, as Table 5.8 shows. Priors for all the parameters are $Normal(0,10^7)$.

Let us start considering the dataset for **SPRING**; we randomly eliminate different percentages of pseudo-zeros (10%, 25%, 50% and 75%) and run the model in JAGS with the following settings: 3 chains, 7000 samples for chain, a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and $thinning = 5$. For each percentage of randomly discarded pseudo-zeros, posterior chains converge. Thus, we look at the differences in the mean and in the extremes of 95% credible intervals of the posterior estimates varying the proportion of randomly discarded pseudo-zeros: in general, if we eliminate 75% or more of the pseudo-zeros, posterior estimates of all the parameters slightly change. Moreover, posterior estimates of the intercept β_1^{bear} are different if we eliminate 50% of the available locations. This confirms again what found by Fieberg et al. [2021] in standard logistic regression. We can conclude that for the First model applied to Spring dataset the initial percentage of pseudo-zeros is enough to guarantee the convergence of the methodology and it is possible to obtain good estimates of the parameters.

For **EARLY SUMMER** as well, JAGS settings are the following: 3 chains, 7000 samples for chain, a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and *thinning* = 10. For this time-window as well, we randomly eliminate several percentages of pseudo-zeros (5%, 10%, 25%, 50% and 75%) and we compare the posterior estimates (mean and 95% credible intervals) with those obtained without the elimination of available locations: the estimates do not change too much, except for β_1^{bear} (when the percentage of discarded pseudo-zeros is 25% or more) and β_1^{wolf} (when the percentage of discarded pseudo-zeros is 50% or more). Thus, as for Spring, we can argue that the number of pseudo-zeros in the dataset is enough to guarantee the convergence of the methodology.

In the case of **LATE SUMMER**, we randomly eliminate different percentages of pseudo-zeros (10%, 25%, 50% and 75%). JAGS settings are: 3 chains, 7000 samples for chain, a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and *thinning* = 5. If we randomly eliminate 50% or less of the initial pseudo-zeros, mean, lower bound and upper bound of the 95% credible intervals do not change too much for the posterior estimates of the parameters, while eliminating 75% of the initial pseudo-zeros leads to different estimates of the parameters, mostly for γ and $\boldsymbol{\beta}^{wolf}$. As already mentioned, we are considering a "reduced" dataset (obtained by taken 1 observation every 30 for the bear and 1 observation every 10 for the wolf) for computational reasons: in the case of non-elimination of pseudo-zeros there are 1280 observations for the bear and 691 for the wolf, where the available locations are 1181 for the bear and 598 for the wolf. If we randomly eliminate 75% of the pseudo-zeros, we obtain 295 pseudo-zeros and 99 ones for the bear, 150 pseudo-zeros and 93 ones for the wolf. Thus, problems arise when we have around one used location every 3 available locations for the bear and one used location every 2 available locations for the wolf. Likely, these are not enough for the model we are considering: it is necessary to take a larger number of pseudo-zeros in order to obtain good estimates of the parameters. To conclude, for the First model the number of pseudo-zeros in Late Summer dataset is enough to guarantee the convergence of the methodology and parameters are well estimated.

Finally, for **AUTUMN**, the JAGS settings are the same as Spring. The percentages of randomly eliminated pseudo-zeros are 10%, 25%, 50% and 75%. Posterior estimates of the parameters do not change too much if the percentage of discarded pseudo-zeros is less than or equal to 50%. For the intercept the estimates change with lower percentages of discarded pseudo-zeros. Nevertheless, we can argue that also in Autumn dataset the original number of pseudo-zeros is enough to guarantee the convergence of the methodology.

As an example, Figure 5.8 represents the effect of eliminating a percentage of pseudo-zeros on lower bound, mean and upper bound of the 95% credible interval, with respect to the proportion of discarded pseudo-zeros in Early Summer for the parameter γ. For the other parameters and the other time-windows, similar Figures are shown in Appendix A.

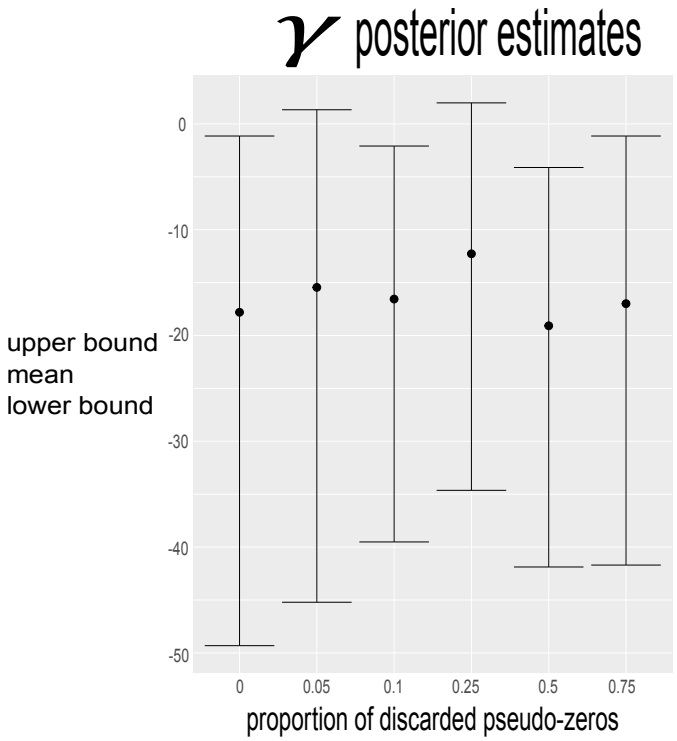

Figure 5.8. First model: effect of eliminating a percentage of pseudo-zeros on posterior estimates (lower bound, mean, upper bound) for γ, with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Early Summer.

5.2.2 Number of pseudo-zeros for the Second model

For the Second model, we do not perform the study over the synthetic dataset as that would be identical to the First model. We consider the real dataset. Prior distributions for the parameters are $Normal(0,10^7)$. We check for each time-window if the posterior estimates of the parameters change by randomly eliminating different percentages of pseudo-zeros. As for the First model, we consider only one observation every 10 for the wolf and one every 30 for the bear.

For **SPRING** dataset, we randomly eliminate the following percentages of pseudo-zeros: 10%, 25%, 50% and 75%. JAGS settings are in this case: 3 chains, 7000 samples for chain, a burn-in phase of 2000 samples, an adaptation phase of 2000 samples and $thinning = 7$. For the intercepts β_1^{bear} and β_1^{wolf} the estimates change if we eliminate 50% of the pseudo-zeros or more, while for the other parameters the estimates do not change until we eliminate 75% of the pseudo-zeros.

For **EARLY SUMMER** dataset, the randomly discarded percentages of pseudo-zeros and JAGS settings are the same as for Spring. Posterior estimates do not change too much if we eliminate less than 75% of the pseudo-zeros for all the parameters, except for β_1^{wolf}, for which the elimination of 50% of the pseudo-zeros or more leads to very different values

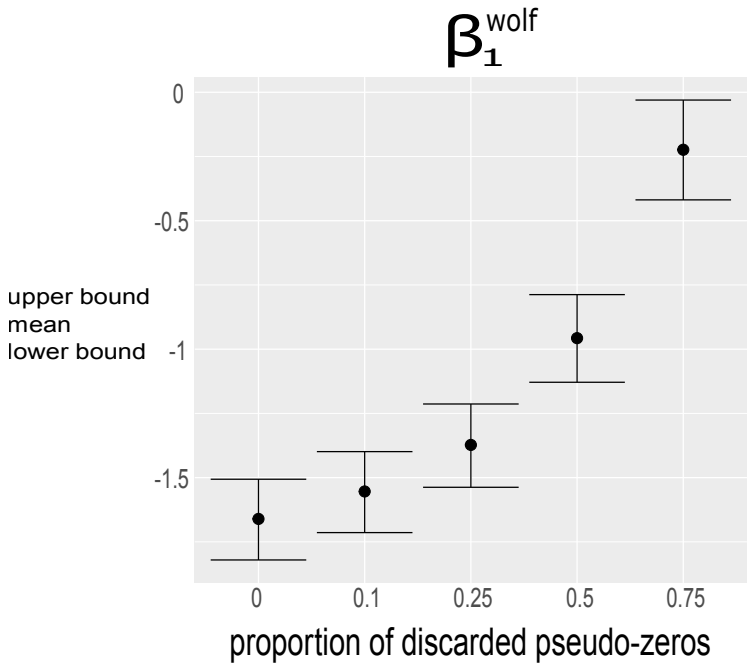

Figure 5.9. Second model: effect of eliminating a percentage of pseudo-zeros on posterior estimates (lower bound, mean, upper bound) of β_1^{wolf}, with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Autumn.

of the posterior mean and the 95% credible intervals.

For **LATE SUMMER** dataset, the randomly discarded percentages of pseudo-zeros are the same as for the previous time-windows, while JAGS settings are in this case: 3 chains, 7000 samples for chain, a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and $thinning = 5$. Results are very similar to what was found for Early Summer.

Finally, for **AUTUMN** dataset, the randomly discarded percentages of pseudo-zeros are the same as for the previous time-windows and JAGS settings are the following: 3 chains, 7000 samples for chain, a burn-in phase of 2000 samples, an adaptation phase of 2000 samples and $thinning = 15$. The posterior estimates do not change too much until the proportion of discarded pseudo-zeros is equal to 75% for $\boldsymbol{\beta}^{bear}$ (except the intercept β_1^{bear}) and γ, and until we eliminate 50% of the pseudo-zeros for $\boldsymbol{\beta}^{wolf}$ (except the intercept β_1^{wolf}) and β_1^{bear}. However, the posterior estimates of β_1^{wolf} change if we eliminate any percentage of pseudo-zeros, as Figure 5.9 shows. This is not surprising, since β_1^{wolf} is an intercept.

As for the First model, the number of pseudo-zeros is enough to have good parameter estimates for the Second model.

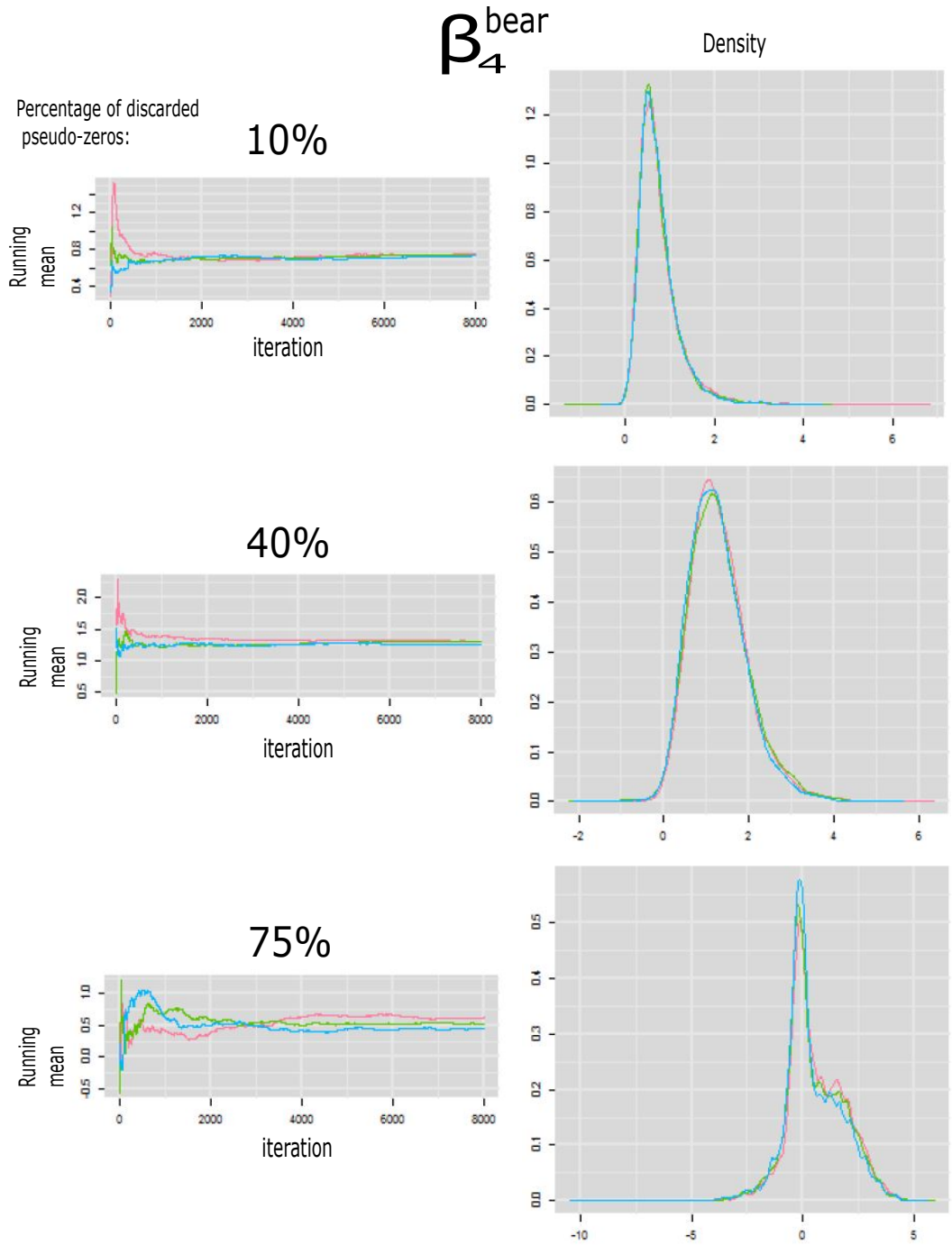

Figure 5.10. Third model: effect of randomly eliminating different percentages of pseudo-zeros on density and running mean of β_4^{bear}, considering the synthetic dataset.

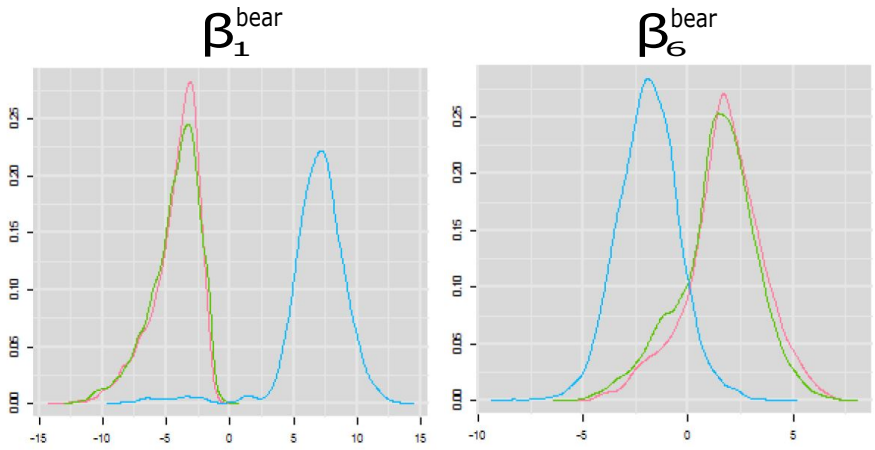

Figure 5.11. Third model: posterior densities for β_1^{bear} and β_6^{bear} with 60% of the pseudo-zeros eliminated. Autumn.

5.2.3 Number of pseudo-zeros for the Third model

We first look at the number of pseudo-zeros for the synthetic dataset used to check the identifiability. We run the model in JAGS (with 3 chains, 8000 samples for chain, a burn-in phase of 2000 samples, an adaptation phase of 2000 samples and $thinning = 15$) and we randomly eliminate the following percentages of pseudo-zeros: 10%, 40% and 75%. As prior distributions, we set $Uniform(0,1)$ for γ^{bear} and γ^{wolf}, and $Normal(0,10)$ for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$. In the case of 10% and 40% of discarded pseudo-zeros, posterior chains reach convergence, while eliminating 75% of the pseudo-zeros leads to chains that converge more slowly. Figure 5.10 shows the posterior running mean and the density in the three cases described above. Moreover, if the percentage of discarded pseudo-zeros is too high, some parameters (like γ^{bear} and γ^{wolf}) are not well estimated.

Then, we consider the four datasets associated to the different time-windows: we take as environmental covariates for each time-window the same as the First model and we reduce the dataset size in the same way. The discarded percentages of pseudo-zeros are also in this case 10%, 40% and 75%.

For **AUTUMN** dataset, JAGS settings are the following: 3 chains, 8000 samples for chain, a burn-in phase of 2000 samples, an adaptation phase of 2000 samples and $thinning = 15$. If we do not eliminate any pseudo-zeros, the chains converge, but eliminating a percentage of pseudo-zeros (like 10% or 40%) leads to non-convergent chains. In particular, with 40% of discarded pseudo-zeros, for some parameters two of the three posterior chains have very similar densities, that approximate a normal distribution with a certain mean, while the density of the other chain approximates a normal distribution symmetric to the others with respect to the value 0, as Figure 5.11 shows for two of the parameters. This could be an empirical proof of the non-identifiability of the Third model on that dataset. We tried to analytically prove the non identifiability of the Third model,

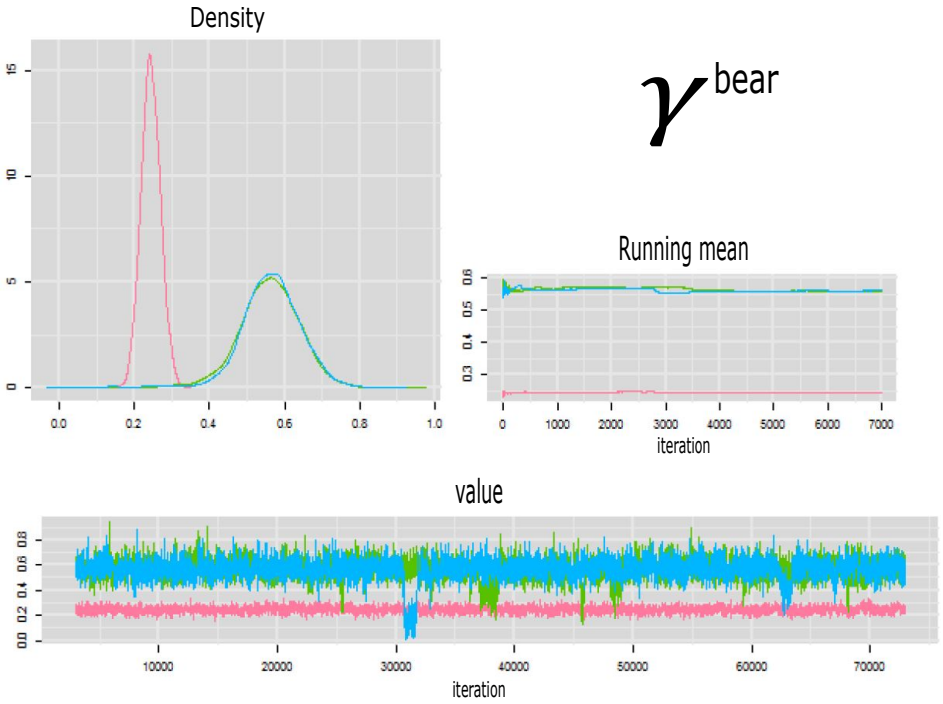

Figure 5.12. Third model: posterior densities for γ^{bear} with elimination of 75% of the pseudo-zeros. Late summer.

without succeeding in general. However, we have proven that if $\gamma^{bear} = \gamma^{wolf} = 1$ swapping $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ gives the same likelihood for the data. See Appendix B for more details.

For **EARLY SUMMER** dataset, JAGS settings are the following: 3 chains, 7000 samples for chain, a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and *thinning* = 10. If we do not eliminate any pseudo-zeros, the chains converge. Also with the elimination of 10% of the pseudo-zeros, chains converge, but eliminating an higher percentage of pseudo-zeros (for instance 40% or 75%) leads to non-convergent chains. The situation is similar to what shown in Figure 5.11 for the Autumn case.

For **LATE SUMMER** dataset, JAGS settings are the same as the ones used for Early summer. The chains reach convergence with the following percentages of randomly discarded pseudo-zeros: 10% and 40%. However, with the random elimination of 75% of the pseudo-zeros, we obtain non-convergent chains. In Figure 5.12 we represent the posterior density of the three chains for γ^{bear}: likely, the model is not identifiable on this reduced dataset, since it seams that there are two possible values for γ^{bear}. The situation is similar for the other parameters.

For **SPRING** dataset, JAGS settings are the following: 3 chains, 9000 samples for chain, a burn-in phase of 3000 samples, an adaptation phase of 2500 samples and *thinning* = 20. In this case, if we eliminate 10% of the pseudo-zeros, chains converge, but if we eliminate more than 40% of the pseudo-zeros, this no longer happens. The obtained

chains are similar to those shown for the other time-windows.

For this model, we find that if we eliminate a small percentage of pseudo-zeros in the Autumn and Early Summer datasets, the MCMC do not reach convergence, meaning that they either have a very bad behaviour or different chains converge to different parameter values. For the remaining time-windows, chains converge if we discard less than 40% of pseudo-zeros.

5.2.4 Number of pseudo-zeros for the Fourth model

For the Fourth model as well, we initially take into consideration the synthetic dataset used in Section 5.1.4 and we randomly eliminate the following percentages of pseudo-zeros: 10%, 25%, 50%, 75%. For each percentage of discarded pseudo-zeros, we run the model in JAGS in order to check if the estimates of the parameters change, with the following settings: 3 chains, 7000 samples for chain, a burn-in phase of 3000 samples, an adaptation phase of 2000 samples and $thinning = 15$. Prior distributions for δ^{bear} and δ^{wolf} are $Bernoulli(0.5)$, for γ^{bear} and γ^{wolf} are $Uniform(0,1)$, and for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are $Normal(0,10)$. For all the percentages of randomly eliminated pseudo-zeros, posterior chains of the parameters converge. Moreover, for γ^{bear} the posterior estimates do not change too much if we randomly eliminate less than 50% of the initial pseudo-zeros, for γ^{wolf} if we randomly eliminate less than 75% of the initial pseudo-zeros. For δ^{bear} the posterior estimates are the same for each percentage of discarded pseudo-zeros, while for δ^{wolf} the posterior estimates change if we randomly eliminate 50% of the pseudo-zeros. For $\boldsymbol{\beta}^{bear}$, the posterior estimates slightly change if we randomly eliminate 50% of the pseudo-zeros, and significantly change if the percentage of discarded pseudo-zeros reaches 75%. Finally, for $\boldsymbol{\beta}^{wolf}$, the posterior estimates do not change too much until we eliminate 75% of the initial pseudo-zeros. On the synthetic dataset, the number of pseudo-zeros is therefore enough to guarantee good parameter estimates.

Then, we study the behaviour of the chains considering real data in the different time-windows. For computational reasons, also in this case we take into consideration 1 observation every 30 for the bear and 1 observation every 10 for the wolf. Statistics on the number of used and available locations are the same as for the First model, as shown in Table 5.7. For all time-windows, prior distributions are set as following: δ^{bear} and δ^{wolf} are $Bernoulli(0.5)$, γ^{bear} and γ^{wolf} are $Uniform(0,1)$, $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are $Normal(0,10)$. Moreover, JAGS settings are the following: 3 chains, 7000 samples for chain, a burn-in phase of 1000 samples, an adaptation phase of 2000 samples and $thinning = 10$. The percentages of discarded pseudo-zeros are 10%, 40%, 75%.

For **SPRING** dataset, we run the model in JAGS, and we eliminate the different percentages of pseudo-zeros in order to see if the posterior estimates of the parameters change or not. If we eliminate 40% or 75% of the pseudo-zeros, the posterior chains do not reach convergence.

In **EARLY SUMMER** case, the posterior chains obtained considering the different percentages of pseudo-zeros described above reach convergence. However, for most of the parameters (γ^{bear}, γ^{wolf}, δ^{bear}, δ^{wolf}, $\boldsymbol{\beta}^{wolf}$ and some components of $\boldsymbol{\beta}^{bear}$) the posterior estimates change if we eliminate more than 40% of the pseudo-zeros. This is evident mostly for the intercepts β_1^{bear} and β_1^{wolf}, as Figure 5.13 shows. When we consider all the

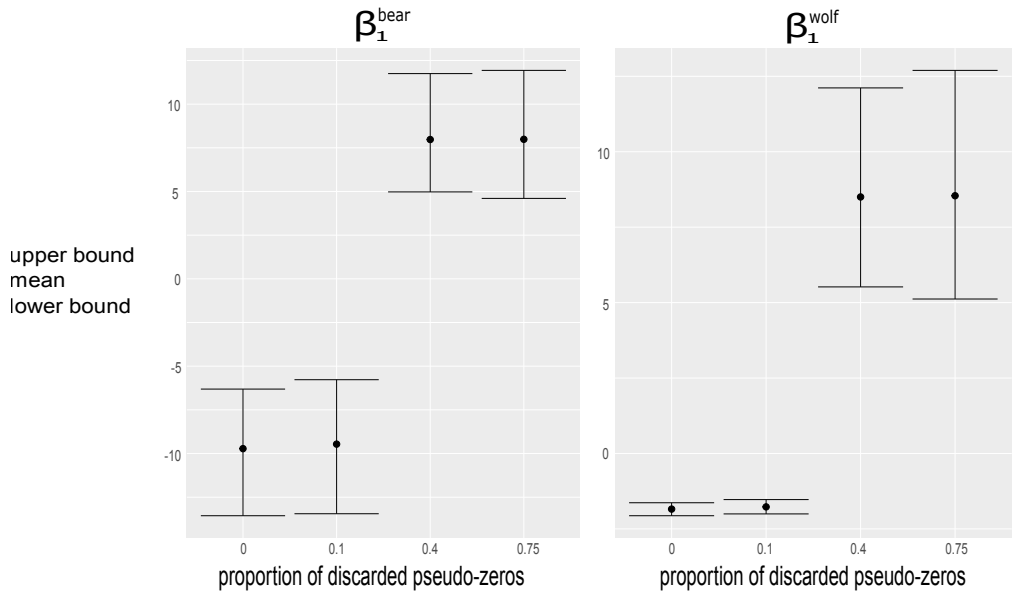

Figure 5.13. Fourth model: Effect of eliminating a percentage of pseudo-zeros on the posterior estimates (lower bound, mean, upper bound) of the intercepts, with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Early Summer.

pseudo-zeros or 90% of the pseudo-zeros, posterior estimates are concentrated on a certain value, while if we eliminate an higher percentage of pseudo-zeros, they are concentrated on a different value.

In **LATE SUMMER** case, the chains reach convergence with all the percentages of discarded pseudo-zeros we considered. In this case, posterior estimates of the parameters do not change until we eliminate 75% of the pseudo-zeros.

For these first three time-windows, then, the number of pseudo-zeros is enough to have good parameter estimates for the Fourth model. However, reducing them by more than 10% changes the parameter estimates for Spring and Early Summer, while the threshold is larger for Late Summer.

In **AUTUMN** case, if we take into consideration the original reduced dataset, chains do not reach convergence; also if we eliminate small percentages of pseudo-zeros (like 10% or 40%), they take a lot of steps to reach convergence, while randomly eliminating 75% of the original pseudo-zeros leads to convergent chains. This behaviour is different from what we expected and we are unable to give an explanation.

Chapter 6

Final results on real datasets and comparison between models

We now run the four models on the entire datasets corresponding to the different time-windows, in order to obtain posterior estimates of the parameters and to compare results obtained through the different models. JAGS settings used for all the models in the different time-windows are the following: 3 chains, 7000 samples for chain, a burn-in phase of 2000 samples, an adaptation phase of 2000 samples and $thinning = 10$. Posterior chains for the parameters of the different models converge, except for the Second model in Autumn. For each model and time-window in which convergence is reached, the mean and the extremes of the 95% credible intervals are computed. In Table 6.1 we represent the posterior mean of the parameters of the First model in the different time-windows, in order to compare the behaviours of the animals and their interaction in different parts of the year. Similarly, we represent the same in Tables 6.2, 6.3 and 6.4 for the Second, Third and Fourth models, respectively.

From the values of $\boldsymbol{\beta}^{bear}$ in Table 6.1, we can see the effect of each environmental covariate on the presence of the bear. The larger β_j^{bear} is for a given j, the larger is the effect of the corresponding covariate on the probability of finding the bear. If we compare the absolute values of two components of $\boldsymbol{\beta}^{bear}$, a greater effect on the probability of finding the bear is given by the component with larger absolute value. For instance, in Spring, β_3^{bear} (correspondent to the covariate representing the percentage of pastures and grasslands) is the largest component of $\boldsymbol{\beta}^{bear}$ in absolute value; in Early Summer the largest component is by β_7^{bear}, which corresponds to the average density of trees in oak and hop-hornbeam forest. The same reasoning can be done for Late Summer and Autumn, for which the greater effect is given by β_6^{bear} (associated to average density of trees in beech forest) and β_7^{bear}, respectively. Another distinction can be made in terms of signs: if a parameter is positive, the probability of presence of the bear is increased by positive

	Spring	**Early Summer**	**Late Summer**	**Autumn**
γ	2.88	-27.44	11.69	-8
β_1^{bear}	-2.96	-2.68	-2.82	-2.46
β_2^{bear}	-0.08	0.04	0.26	-0.09
β_3^{bear}	-0.65	0.18	0.45	-0.25
β_4^{bear}	-0.19	0.24	-0.16	0.12
β_5^{bear}	0.28	-0.06	-0.08	-0.24
β_6^{bear}	0.3	0.57	0.76	-0.1
β_7^{bear}	0.3	0.79	0.35	0.37
β_8^{bear}	-0.03	0.16	0.19	-0.05
β_9^{bear}	0.32	0.18	0.3	0.31
β_{10}^{bear}	0.32	0.17	0.01	0.16
β_{11}^{bear}	-0.13	-0.19	-0.06	-0.22
β_{12}^{bear}	-0.45	0.36	0.53	0.03
β_1^{wolf}	-2.28	-0.09	-3.44	-0.9
β_2^{wolf}	-0.03	-0.2	-0.34	-0.01
β_3^{wolf}	0.25	0.25	-0.55	-0.26
β_4^{wolf}	0.04	0.4	0.009	0.19
β_5^{wolf}	-0.6	-0.77	-0.18	-0.31
β_6^{wolf}	0.13	1.55	-0.24	-0.16
β_7^{wolf}	-0.03	1.78	0.04	0.34
β_8^{wolf}	-0.68	0.94	0.79	-0.03
β_9^{wolf}	0.27	0.8	0.22	0.27
β_{10}^{wolf}	-0.07	-0.36	-0.58	0.1
β_{11}^{wolf}	0.14	-0.22	0.52	-0.24
β_{12}^{wolf}	-0.42	0.8	-0.63	-0.15

Table 6.1. First Model: posterior mean of the parameters in the different time-windows.

values of the corresponding covariate; vice versa if it is negative[1]. Thus, we can look at the signs of the parameters considered above and notice that β_3^{bear} is negative for the Spring dataset, β_7^{bear} is positive both in Early Summer and Autumn, and β_6^{bear} is positive in Late Summer. Moreover, β_7^{bear} is positive in all the time-windows, thus if the average density of trees in oak and hop-hornbeam forest in a location is large, it increases the probability of finding the bear in that location; in particular, this happens mostly in Early Summer, since the value of β_7^{bear} is bigger in this time-window than in the others. On the contrary, a covariate that has a negative impact on the presence of the bear in all time-windows is the hillshade, represented by β_{11}^{bear}: this happens more strongly in Autumn, followed by Early Summer and Spring. β_9^{bear} (associated to the distance from secondary roads) and β_{10}^{bear} (associated to the terrain ruggedness index) have similar posterior means in Spring and Early Summer, while quite different posterior means in Late Summer and Autumn. β_{12}^{bear}, that is associated to the percentage of non-vegetated rocky areas, has a negative posterior mean only in Spring. The percentage of shrubland (β_2^{bear}) and the percentage of pastures and grasslands (β_3^{bear}) have a positive impact in Late Summer and Early Summer, and a negative impact in the other time-windows. β_4^{bear} (which is associated to the percentage of agricultural areas) has a positive posterior mean in Early Summer and Autumn, and negative in Spring and Late Summer. Similar comparisons can be made for the other components of $\boldsymbol{\beta}^{bear}$, but in general we can see in Table 6.1 that the posterior means of the parameters usually have same sign in Early Summer and Late Summer, except for β_4^{bear} (associated to the percentage of agricultural areas); the same happens for Spring and Autumn, except for β_4^{bear}, β_5^{bear} (associated to the distance to forest edges) and β_{12}^{bear}. If we fix the values of $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ and of the environmental covariates, we can look at the behaviour of γ, which represents the effect of the presence of the bear on the presence of the wolf. From the Table, we can see that the absolute value of the posterior mean of γ is largest in Early Summer, for which the impact is negative, followed by Late Summer, for which the impact is positive. Also in Spring the impact is positive, but the value of the posterior mean is smaller than in Late Summer. On the contrary, in Autumn the impact is again negative. As for the bear, we can look at the posterior estimates of $\boldsymbol{\beta}^{wolf}$. In Spring, β_8^{wolf} (associated to the distance from urban centres and primary roads) has the highest negative impact on the probability of finding the wolf, followed by β_{12}^{wolf}, while the highest positive impact is given by β_9^{wolf}, followed by β_3^{wolf}. In Early Summer the highest positive impact is given by β_7^{wolf}, that has a very large posterior mean in absolute value compared to the other parameters (1.78); on the contrary, β_5^{wolf} has the biggest negative impact (-0.77). In the same way, we can see from the Table that β_8^{wolf} has the highest positive impact in Late Summer, β_7^{wolf} in Autumn.

For the Second model, we do not report posterior means for the Autumn case in Table 6.2 since, as already mentioned before, chains did not converge in that case. Interpretation of the parameters is analogous to the First model, since they are symmetric. In this case, an high probability of presence of the wolf has a negative impact on the probability of presence of the bear in all considered time-windows. In particular, this impact is very big

[1]Recall here that we are considering standardized covariates.

	Spring	**Early Summer**	**Late Summer**
γ	-11.8	-1.66	-12.83
β_1^{bear}	-1.34	-2.44	-1.34
β_2^{bear}	-0.18	-0.05	0.98
β_3^{bear}	-0.42	-0.05	0.56
β_4^{bear}	-0.18	0.21	0.46
β_5^{bear}	-0.6	-0.12	-0.82
β_6^{bear}	0.6	0.51	1.56
β_7^{bear}	0.29	0.71	0.82
β_8^{bear}	-1.36	0.35	1.3
β_9^{bear}	0.77	0.3	0.63
β_{10}^{bear}	0.37	0.06	-0.66
β_{11}^{bear}	0.04	-0.16	0.14
β_{12}^{bear}	-0.54	0.22	0.58
β_1^{wolf}	-2.07	-2.16	-2.34
β_2^{wolf}	-0.14	-0.26	0.45
β_3^{wolf}	0.11	-0.19	0.06
β_4^{wolf}	0.02	-0.18	0.5
β_5^{wolf}	-0.48	-0.65	-0.5
β_6^{wolf}	0.17	0.3	0.56
β_7^{wolf}	0.01	0.13	0.35
β_8^{wolf}	-0.71	0.65	0.98
β_9^{wolf}	0.28	0.47	0.23
β_{10}^{wolf}	0.05	-0.68	-0.42
β_{11}^{wolf}	0.1	0.16	0.18
β_{12}^{wolf}	-0.45	-0.14	-0.008

Table 6.2. Second Model: posterior mean of the parameters in the different time-windows.

in Late Summer (for which the posterior mean of γ is -12.83) and in Spring (for which the posterior mean of γ is -11.8), while it is lower in Early Summer (for which the posterior mean of γ is -1.66). In Spring, the component of $\boldsymbol{\beta}^{bear}$ that has the biggest impact in absolute value on the presence of the bear is β_8^{bear}, in Early Summer β_7^{bear} (according to what found with the First model), and in Late Summer β_6^{bear}. Similar reasoning can be done for $\boldsymbol{\beta}^{wolf}$, for which in general the values of the posterior means are lower than $\boldsymbol{\beta}^{bear}$. In particular, β_{11}^{bear} has posterior mean equal to 0: this means that the covariate associated to this parameter (hillshade) likely has a negligible impact on the probability of finding the wolf.

For the Third model, In Table 6.3, parameters γ^{bear} and γ^{wolf} are very interesting, since the first one represents the probability of finding the bear in a location where the wolf is observed, the second one vice versa. In Spring the probability of finding the wolf in a location where the bear is observed is, in average, almost equal to 0 (more precisely 0.006), and the probability of finding the bear in a location where the wolf is observed is 0.12. The latter increases in Early Summer, Late Summer and Autumn, where, in particular, it reaches the value 0.5. The probability of finding the wolf in a position in which the bear is present is small in Autumn (0.02), while in Late Summer it is equal to 0.18. However, in Early Summer the probability of finding the wolf in a location in which the bear is observed is, in average, higher than in the other time-windows (0.39), even if it remains lower than 0.5. In the case in which a species is not observed in a location, the probability of finding the other species depends only on the environmental covariates. In that case, each component of $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ gives an idea of the effect of the correspondent covariate on the presence of the considered animal. Values of the parameters associated to the different covariates can be compared: if a parameter is in absolute value larger than another parameter, this means that the first has a greater impact on the probability of finding the animal. The biggest impact on the probability of finding the bear in the different time-windows is given by: β_9^{bear} in Spring (with a posterior mean equal -0.32, negative impact), β_6^{bear} in Early Summer (0.35, positive impact) and Late Summer (-0.32, negative impact) and β_4^{bear} in Autumn (0.33, positive impact). Regarding the probability of finding the wolf, the highest positive impact is given by β_8^{wolf} in Early Summer (0.27) and Autumn (0.29), by β_{10}^{wolf} in Spring (0.12), and by β_6^{wolf} in Late Summer (0.21).

Finally for the Fourth model, Table 6.4 shows that δ^{bear} has posterior mean equal to 1 in Early Summer, Late Summer and Autumn and almost equal to 1 in Spring. This means that the presence of the wolf attracts the bear in all the time windows considered. The strength of this attractiveness is represented by γ^{bear}: we can see in Table 6.4 that this is very low in Spring (since $\gamma^{bear} = 0.075$ in average), then it increases in Early Summer and Late Summer ($\gamma^{bear} = 0.18$ and 0.26 respectively) and reaches its maximum value in Autumn (0.47). However, this attractiveness is never total, as the posterior mean of γ^{bear} is not equal to 1. On the contrary, the posterior mean of δ^{wolf} is equal to 0 in Spring and Autumn, and assumes quite low value in Early Summer and Late Summer. In Autumn the posterior mean of γ^{wolf} is high (0.83), so we can argue that the wolf tends to avoid

locations in which the bear is observed, while the bear is attracted by the wolf. This is in accordance with what we found in the Third model. Regarding Spring, the posterior mean of γ^{wolf} is very high (0.95): this means that there is almost total repulsiveness of the wolf with respect to the bear. In Early Summer and Late Summer, the posterior means of δ^{wolf} are 0.37 and 0.31 respectively, so there is a larger number of times in which there has been a repulsive effect on the wolf with respect to the bear than an attractive effect. Likely, in these time-windows, the bear is attracted by the wolf with a lower strength with respect, for instance, to the Autumn case; thus, the wolf tends less to avoid locations in which the bear is observed. This confirms what we found with the Third model, that is the probability of finding the wolf in a location of the bear is slightly higher than in Autumn.

	Spring	**Early Summer**	**Late Summer**	**Autumn**
γ^{bear}	0.12	0.23	0.3	0.5
γ^{wolf}	0.006	0.39	0.18	0.02
β_1^{bear}	-2.76	-2.9	-3.04	-3.48
β_2^{bear}	0.08	0.04	-0.04	-0.28
β_3^{bear}	-0.2	0.02	-0.17	-0.19
β_4^{bear}	-0.23	-0.04	0.08	0.33
β_5^{bear}	0.14	0.05	0.02	0.1
β_6^{bear}	0.08	0.35	-0.32	-0.06
β_7^{bear}	-0.04	-0.06	-0.09	-0.05
β_8^{bear}	-0.09	0.06	-0.002	0.25
β_9^{bear}	-0.32	-0.09	-0.13	-0.15
β_{10}^{bear}	-0.11	-0.11	0.15	0.02
β_{11}^{bear}	0.1	0.06	-0.002	0.1
β_{12}^{bear}	0.03	-0.04	-0.08	0.04
β_1^{wolf}	-1.75	-1.66	-2.23	-1.6
β_2^{wolf}	-0.08	-0.02	-0.07	-0.11
β_3^{wolf}	0.05	0.04	0.11	0.04
β_4^{wolf}	0.06	-0.09	-1.06	-0.02
β_5^{wolf}	-0.13	-0.02	-0.06	0.05
β_6^{wolf}	0.04	0.03	0.21	-0.04
β_7^{wolf}	-0.02	0.005	-0.18	-0.03
β_8^{wolf}	-0.11	0.27	0.18	0.29
β_9^{wolf}	0.06	0.002	-0.004	0.002
β_{10}^{wolf}	0.12	0.04	-0.04	0.06
β_{11}^{wolf}	-0.05	0	-0.12	-0.06
β_{12}^{wolf}	0.11	-0.01	0.09	-0.06

Table 6.3. Third Model: posterior mean of the parameters in the different time-windows.

	Spring	**Early Summer**	**Late Summer**	**Autumn**
γ^{bear}	0.075	0.18	0.26	0.47
γ^{wolf}	0.95	0.24	0.22	0.83
δ^{bear}	0.99	1	1	1
δ^{wolf}	0	0.37	0.31	0
β_1^{bear}	-2.78	-2.88	-3.05	-3.46
β_2^{bear}	0.08	0.05	-0.04	-0.29
β_3^{bear}	-0.18	0.03	-0.18	-0.2
β_4^{bear}	-0.21	-0.01	0.08	0.32
β_5^{bear}	0.14	0.04	0.05	0.1
β_6^{bear}	0.07	0.31	-0.29	-0.07
β_7^{bear}	-0.02	-0.01	-0.07	-0.06
β_8^{bear}	-0.08	0.07	0.006	0.23
β_9^{bear}	-0.3	-0.07	-0.12	-0.14
β_{10}^{bear}	-0.1	-0.09	0.14	0.03
β_{11}^{bear}	0.09	0.07	-0.008	0.09
β_{12}^{bear}	0.02	-0.02	-0.08	0.05
β_1^{wolf}	-1.75	-1.62	-2.15	-1.6
β_2^{wolf}	-0.08	-0.03	-0.06	-0.09
β_3^{wolf}	0.06	0.03	0.12	0.05
β_4^{wolf}	0.06	-0.12	-0.79	-0.02
β_5^{wolf}	-0.13	0.007	-0.08	0.04
β_6^{wolf}	0.04	0.05	0.2	-0.04
β_7^{wolf}	-0.02	-0.005	-0.2	0.03
β_8^{wolf}	-0.1	0.25	0.16	0.28
β_9^{wolf}	0.06	-0.01	-0.003	0.005
β_{10}^{wolf}	0.12	0.02	-0.03	0.06
β_{11}^{wolf}	-0.05	-0.006	-0.11	-0.06
β_{12}^{wolf}	0.11	-0.02	0.09	-0.07

Table 6.4. Fourth Model: posterior mean of the parameters in the different time-windows.

Chapter 7

Conclusion

In this thesis, four different models have been presented. We studied the identifiability of the parameters and checked if the number of pseudo-zeros of each considered dataset was enough to guarantee the convergence of the methodology for each model. The First and the Second models are identifiable, while we analytically proved that the Third and the Fourth models are not identifiable in particular cases ($\gamma^{bear} = \gamma^{wolf} = 1$, and $\gamma^{bear} = \gamma^{wolf} = \delta^{bear} = \delta^{wolf} = 1$ respectively), and we empirically proved that if we took prior distributions with large supports for $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$, posterior chains converge to different parameter values. Regarding the effect of the number of pseudo-zeros, the considered datasets (the synthetic one and the reduced datasets corresponding to the different time-windows) contain a number of pseudo-zeros that is enough to guarantee the convergence of the posterior chains and good estimates of the parameters for the First and the Second model, while if we discard a too high percentage of pseudo-zeros from some of the considered datasets, some problems in the convergence of the chains and in the estimates of the parameters arise for the Third and the Fourth models.

We then applied these models to the real datasets and compare results obtained in the different time-windows with the different models (except the Second model in Autumn, for which the posterior chains did not converge). In particular, the Third and the Fourth models give similar conclusions regarding the interaction between the two species, while the Second model gives a different conclusion. However, it is important to notice that the temporal dependency has not been modelled, thus if we estimate a high probability of both species being present in a given location, this does not mean that they encountered, but likely they have been passed through that location in different moments. This leads to difficulties in the interpretation of the probability to find a species given the presence of the other species, and can also explain different conclusions from different models.

Finally, performing inference on the real datasets with all models was computationally expensive.

Future works could involve the temporal dimension in order to better understand animal movement taking into consideration time as well.

Appendix A

Additional figures

A.1 Data representation

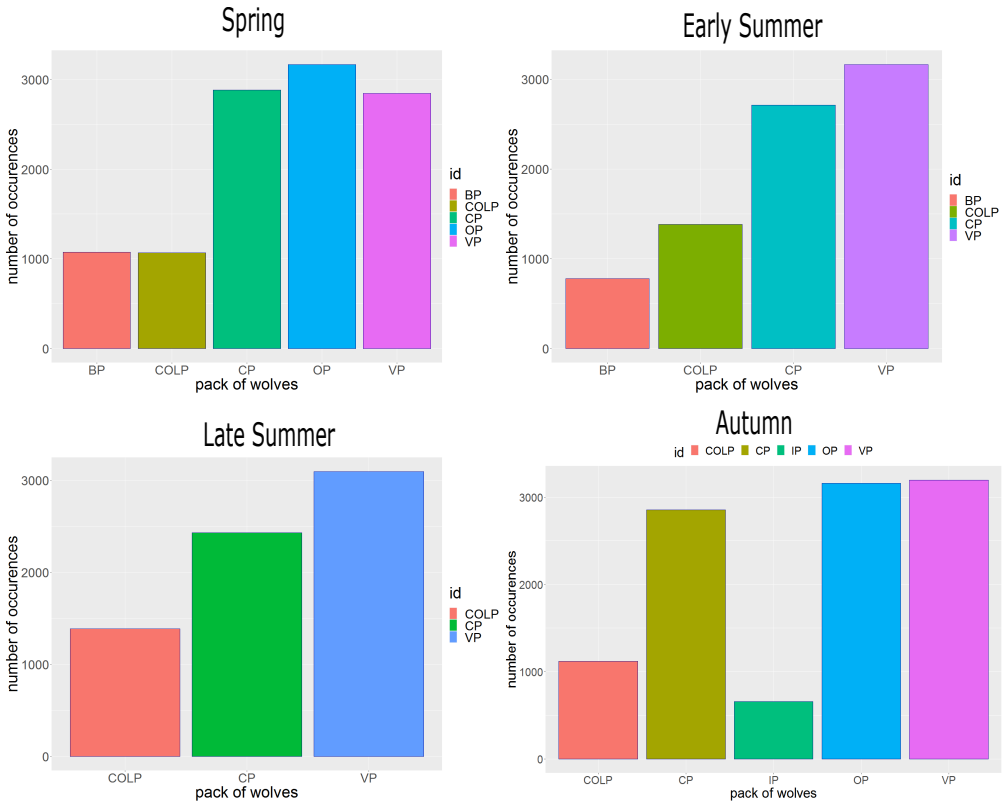

Figure A.1. Wolves: number of occurrences for each pack in the different time-windows.

Figures A.1 and A.2 represent, for each time-window, the number of occurrences for each pack of wolves (in the case of wolves) or for each individual (in the case of bears).

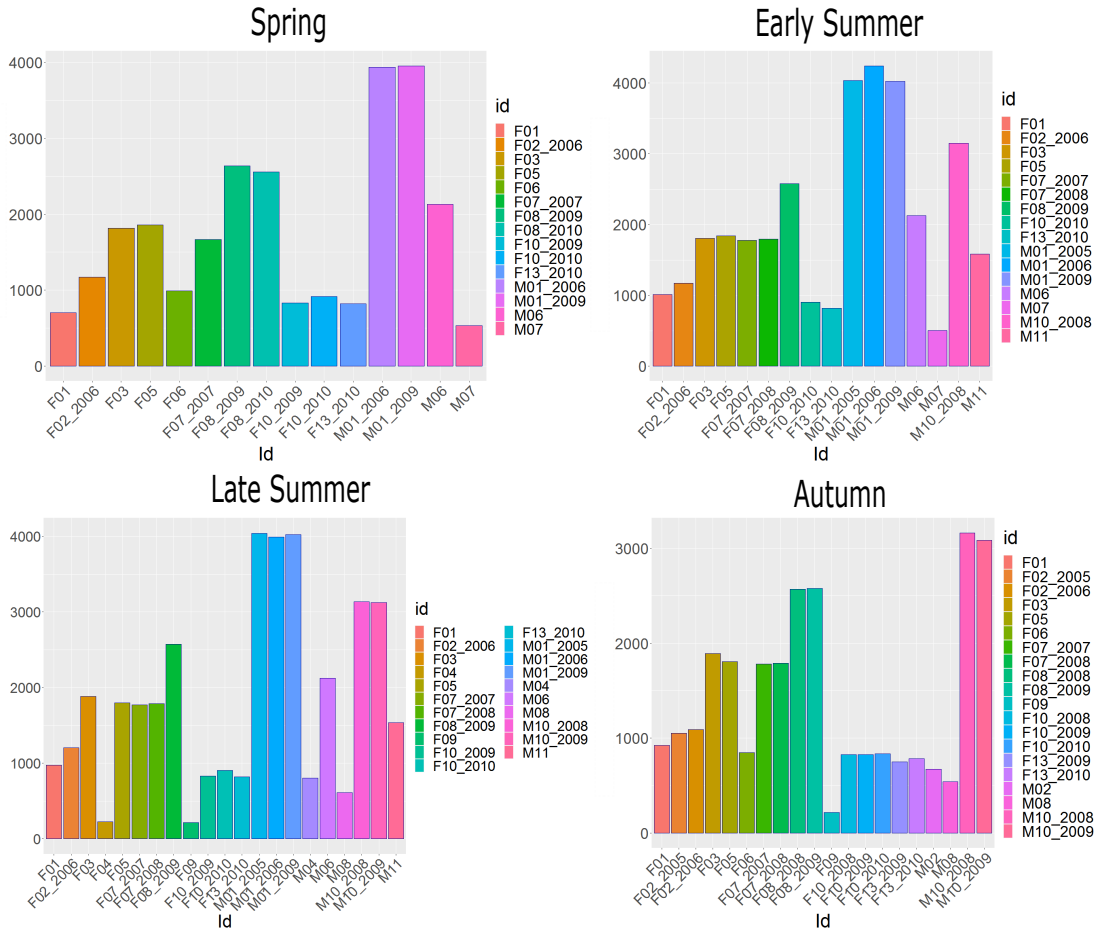

Figure A.2. Bears: number of occurrences for each individual in the different time-windows.

A.2 Number of pseudo-zeros for the First model

The following Figures show how the posterior estimates change if we eliminate different percentages of available locations, in different time-windows.

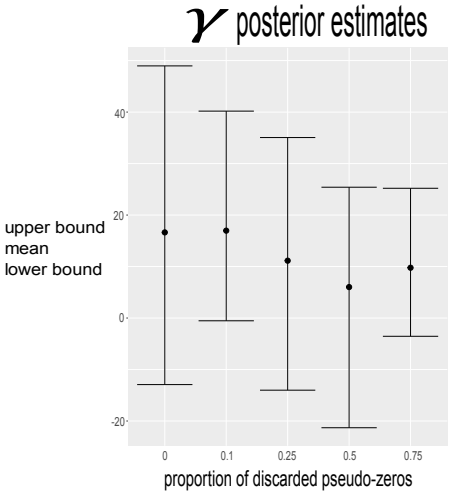

Figure A.3. First model: effect of eliminating a percentage of pseudo-zeros on γ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Spring.

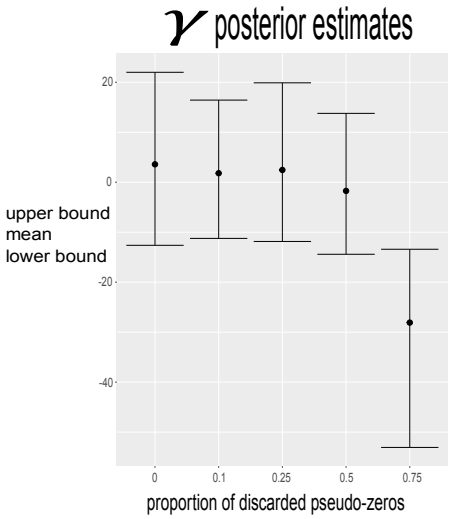

Figure A.4. First model: effect of eliminating a percentage of pseudo-zeros on γ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Late Summer.

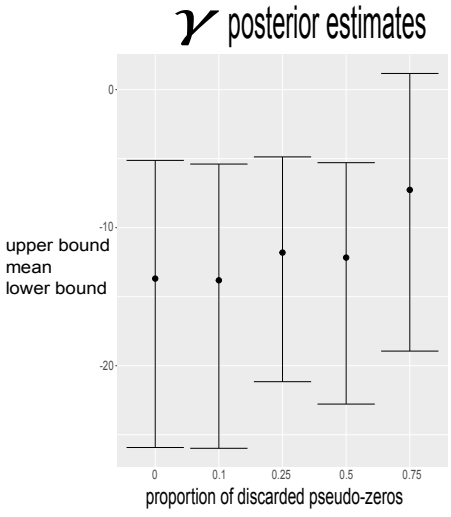

Figure A.5. First model: effect of eliminating a percentage of pseudo-zeros on γ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Autumn.

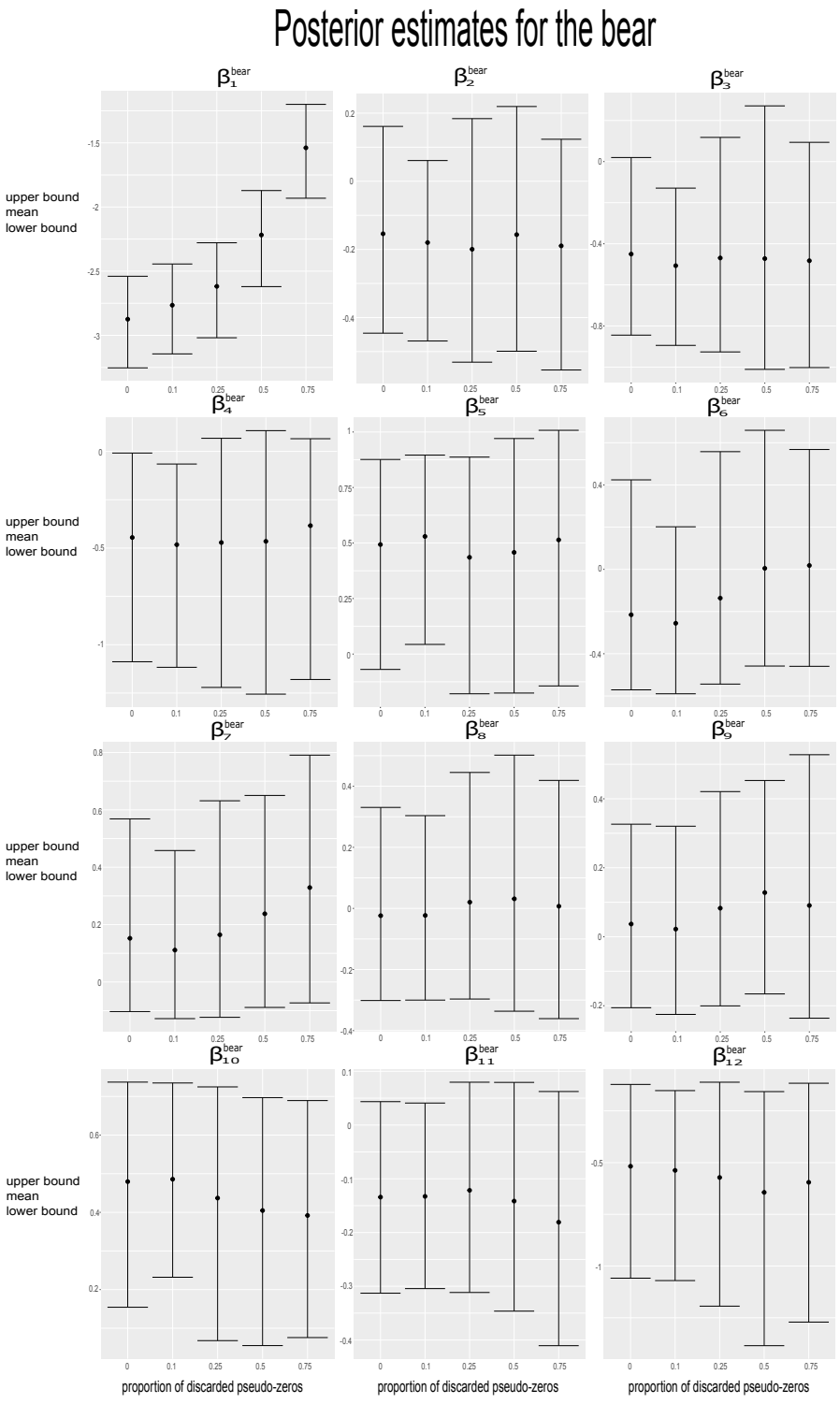

Figure A.6. First model: effect of eliminating a percentage of pseudo-zeros on β^{bear} posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Spring.

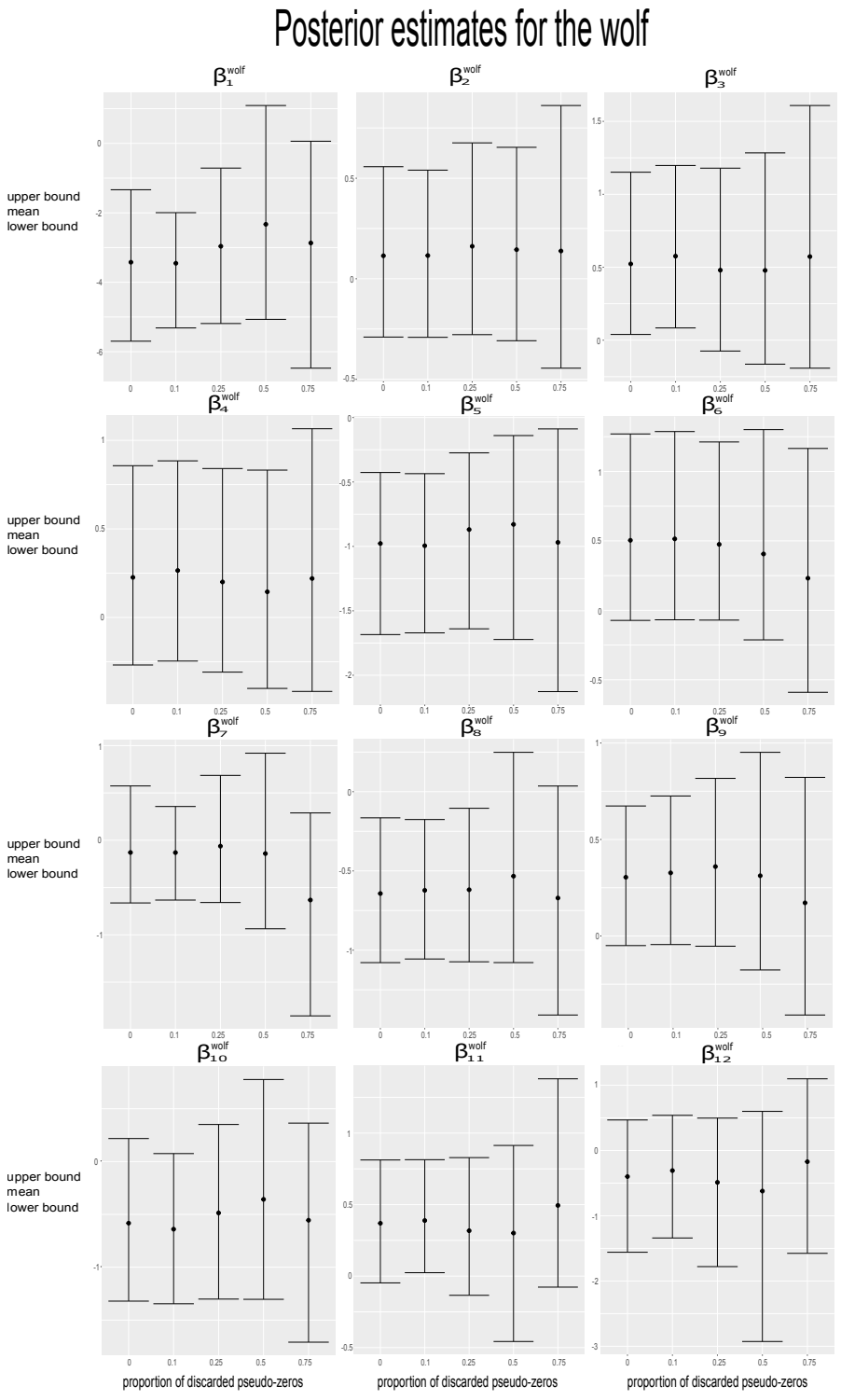

Figure A.7. First model: effect of eliminating a percentage of pseudo-zeros on $\boldsymbol{\beta}^{wolf}$ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Spring.

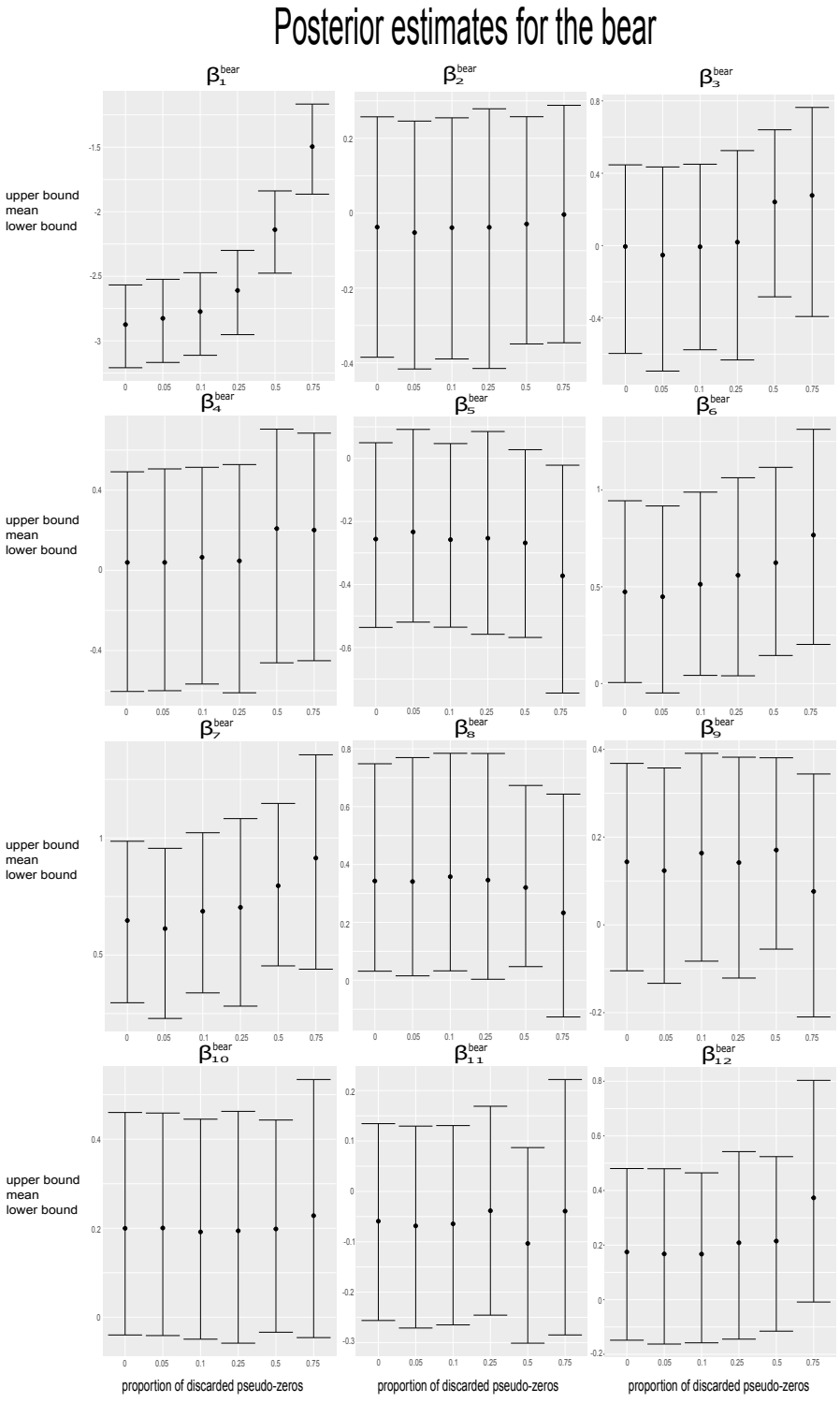

Figure A.8. First model: effect of eliminating a percentage of pseudo-zeros on $\boldsymbol{\beta}^{bear}$ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Early Summer.

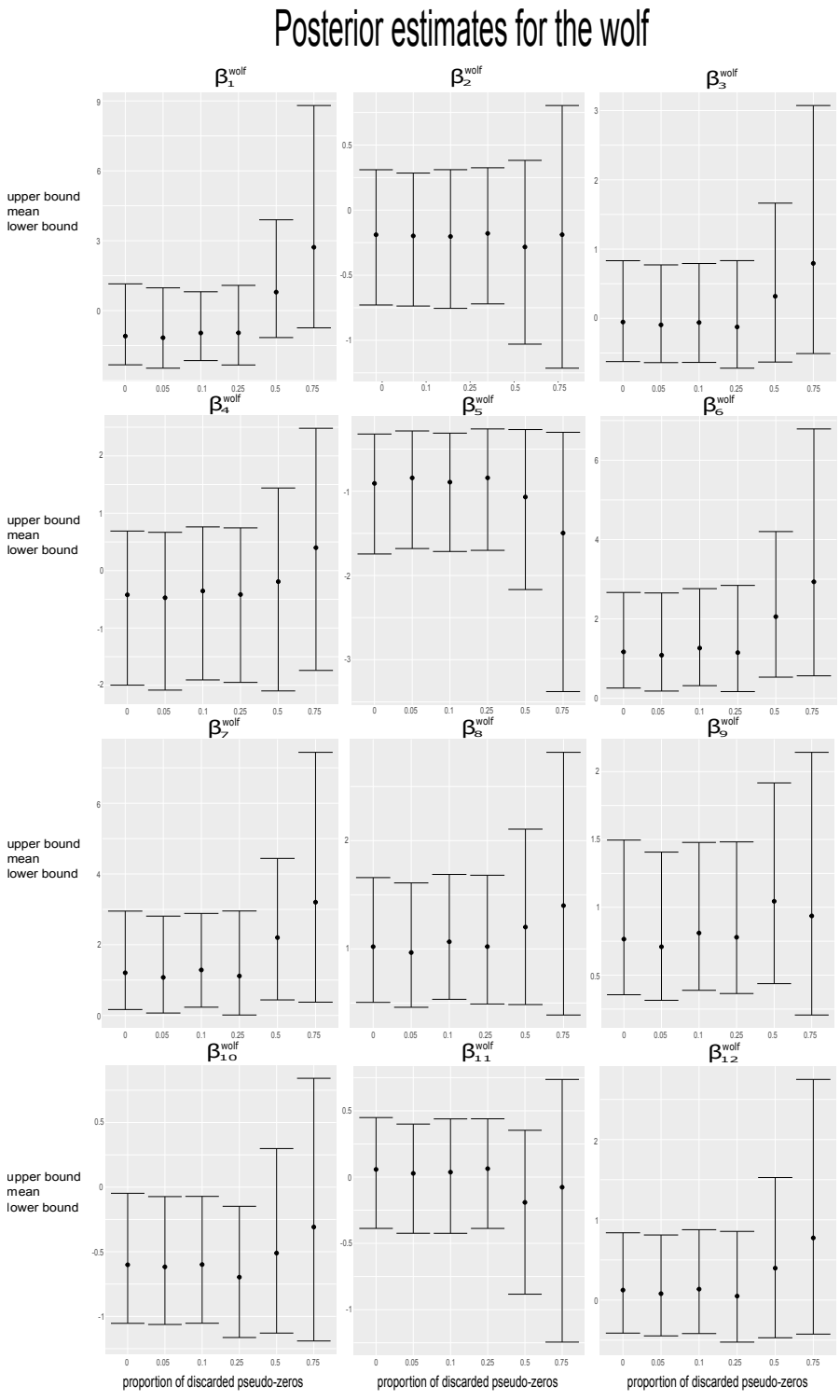

Figure A.9. First model: effect of eliminating a percentage of pseudo-zeros on $\boldsymbol{\beta}^{wolf}$ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Early Summer.

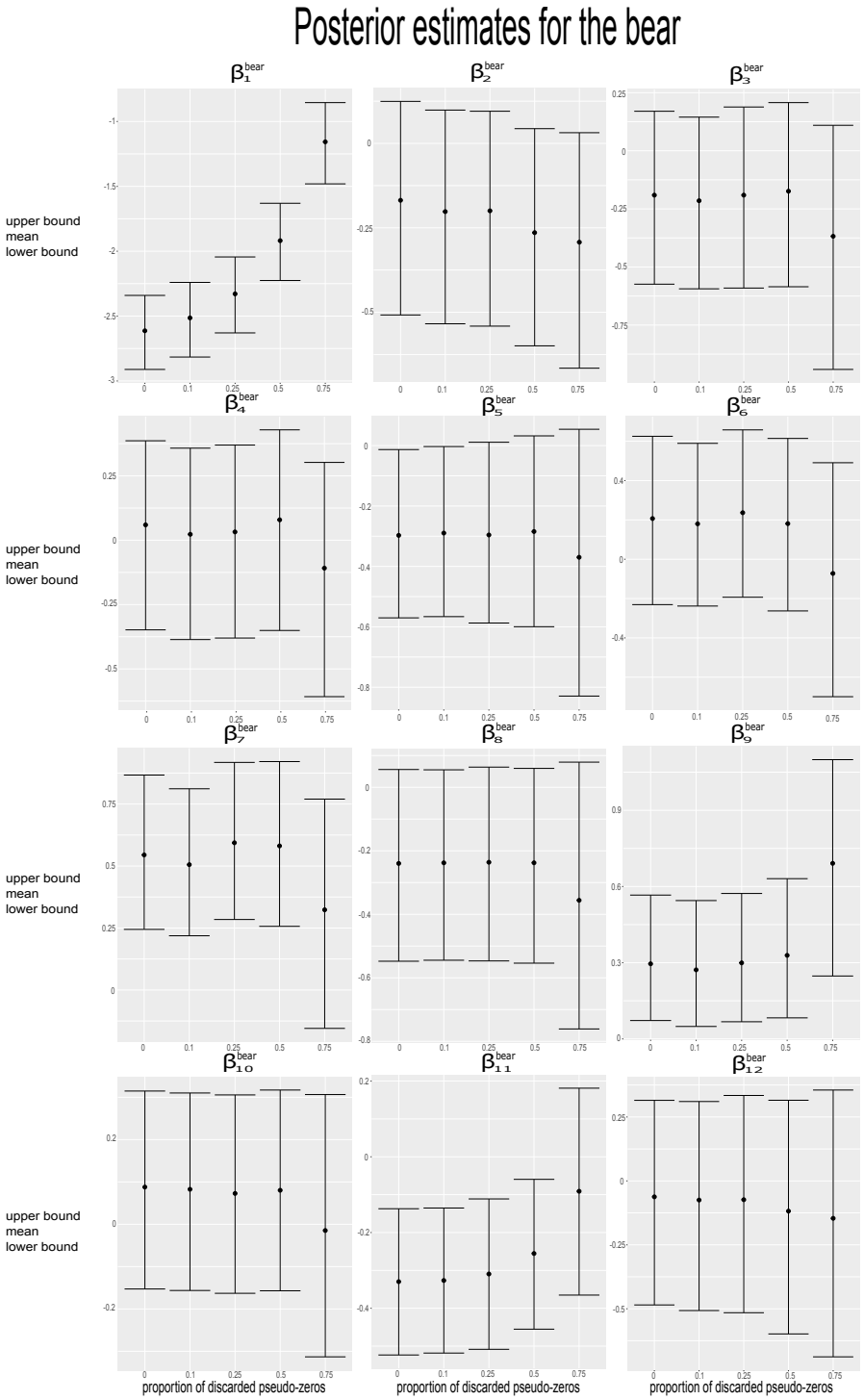

Figure A.10. First model: effect of eliminating a percentage of pseudo-zeros on $\boldsymbol{\beta}^{bear}$ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Autumn.

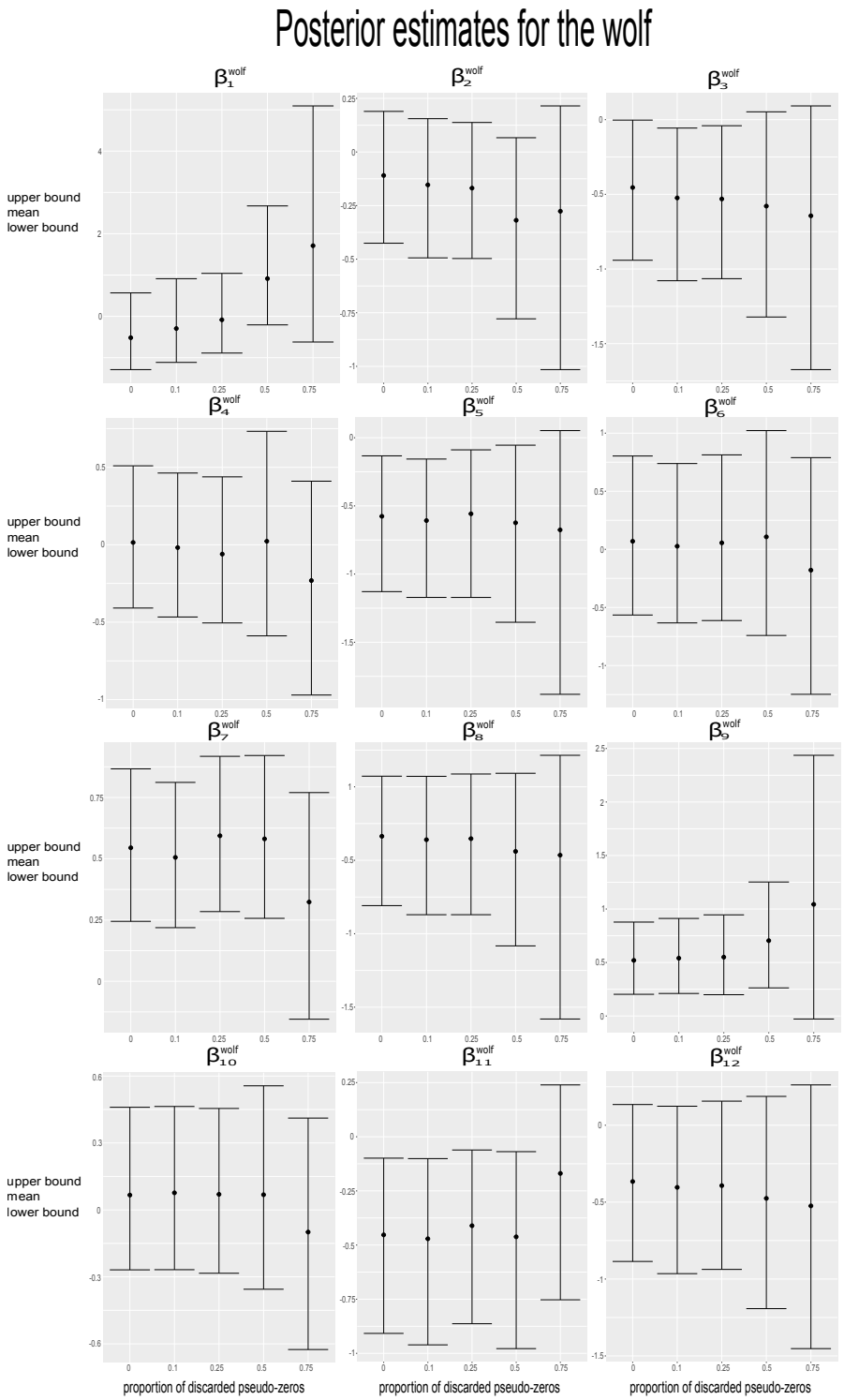

Figure A.11. First model: effect of eliminating a percentage of pseudo-zeros on $\boldsymbol{\beta}^{wolf}$ posterior estimates (lower bound, mean, upper bound), with respect to the proportion of discarded pseudo-zeros. 1 observation every 30 for the bears, 1 observation every 10 for the wolves. Autumn.

Appendix B

Proofs

B.1 Identifiability of the Third model

We now analytically compute the likelihoods for $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ given a set of parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^{bear}, \boldsymbol{\beta}^{wolf}, \gamma^{bear}, \gamma^{wolf}\right)$. For $i = 1, ..., N_{bear}$:

$$P\left(y_i^{bear} = 1 \mid \boldsymbol{\theta}, \boldsymbol{y}^{wolf}\right) = \left(1 - y_i^{wolf}\right)\pi_i^{bear} + y_i^{wolf}\gamma^{bear},$$

$$P\left(y_i^{bear} = 1 \mid \boldsymbol{\theta}\right) = \mathbb{E}_{\boldsymbol{y}^{wolf}}\left[\left(1 - y_i^{wolf}\right)\pi_i^{bear} + y_i^{wolf}\gamma^{bear}\right] = \left(1 - \pi_i^{wolf}\right)\pi_i^{bear} + \pi_i^{wolf}\gamma^{bear}.$$

Since

$$\pi_i^{bear} = logit^{-1}\left(X_i^{bear}\boldsymbol{\beta}^{bear}\right) = \frac{e^{X_i^{bear}\cdot\boldsymbol{\beta}^{bear}}}{1 + e^{X_i^{bear}\boldsymbol{\beta}^{bear}}},$$

$$\pi_i^{wolf} = logit^{-1}\left(X_i^{bear}\boldsymbol{\beta}^{wolf}\right) = \frac{e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}{1 + e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}},$$

we obtain:

$$P\left(y_i^{bear} = 1 \mid \boldsymbol{\theta}\right) = \frac{1}{1 + e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}\frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}}}{1 + e^{X_i^{bear}\boldsymbol{\beta}^{bear}}} + \gamma^{bear}\frac{e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}{1 + e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}} =$$

$$= \frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}} + \gamma^{bear}e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\left(1 + e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)}{\left(1 + e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\right)\left(1 + e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)}.$$

For $j = N_{bear}, ..., N_{bear} + N_{wolf}$:

$$P\left(y_j^{wolf} = 1 \mid \boldsymbol{\theta}, \boldsymbol{y}^{bear}\right) = \left(1 - y_j^{bear}\right)\pi_j^{wolf} + y_j^{bear}\gamma^{wolf},$$

$$P\left(y_j^{wolf} = 1 \mid \boldsymbol{\theta}\right) = \mathbb{E}_{\boldsymbol{y}^{bear}}\left[\left(1 - y_j^{bear}\right)\pi_j^{wolf} + y_j^{bear}\gamma^{wolf}\right] = \left(1 - \pi_j^{bear}\right)\pi_j^{wolf} + \pi_j^{bear}\gamma^{wolf}.$$

Since

$$\pi_j^{bear} = logit^{-1}\left(X_j^{wolf}\boldsymbol{\beta}^{bear}\right) = \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}}{1 + e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}},$$

$$\pi_j^{wolf} = logit^{-1}\left(X_j^{wolf}\cdot\boldsymbol{\beta}^{wolf}\right) = \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}}{1 + e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}},$$

we obtain:

$$P\left(y_j^{wolf}=1 \mid \boldsymbol{\theta}\right)=\frac{1}{1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}}\frac{e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}}{1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}}+\gamma^{wolf}\frac{e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}}{1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}}=$$
$$=\frac{e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}+\gamma^{wolf}e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\right)}{\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\right)\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)}.$$

If $\gamma^{bear}=\gamma^{wolf}=1$, in the two likelihoods above $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are interchangeable. Thus, in that case, the model is not identifiable.

If $\gamma^{bear}$and$\gamma^{wolf}\neq 1$, nothing can be said from an analytical point of view.

B.2 Identifiability of the Fourth model

As for the Third model, we analytically compute the likelihoods for $\boldsymbol{y}^{bear}$ and $\boldsymbol{y}^{wolf}$ given a set of parameters $\boldsymbol{\theta}=\left(\boldsymbol{\beta}^{bear},\boldsymbol{\beta}^{wolf},\gamma^{bear},\gamma^{wolf},\delta^{bear},\delta^{wolf}\right)$. For $i=1,...,N_{bear}$:

$$P\left(y_i^{bear}=1 \mid \boldsymbol{\theta},\boldsymbol{y}^{wolf}\right)=\pi_i^{bear}-y_i^{wolf}\gamma^{bear}\pi_i^{bear}+y_i^{wolf}\gamma^{bear}\delta^{bear},$$
$$P\left(y_i^{bear}=1 \mid \boldsymbol{\theta}\right)=\mathbb{E}_{\boldsymbol{y}^{wolf}}\left[\pi_i^{bear}-y_i^{wolf}\gamma^{bear}\pi_i^{bear}+y_i^{wolf}\gamma^{bear}\delta^{bear}\right]=$$
$$=\pi_i^{bear}-\pi_i^{wolf}\gamma^{bear}\pi_i^{bear}+\pi_i^{wolf}\gamma^{bear}\delta^{bear}.$$

Since

$$\pi_i^{bear}=logit^{-1}\left(X_i^{bear}\boldsymbol{\beta}^{bear}\right)=\frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}}}{1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}},$$
$$\pi_i^{wolf}=logit^{-1}\left(X_i^{bear}\cdot\boldsymbol{\beta}^{wolf}\right)=\frac{e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}{1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}},$$

we obtain:

$$P\left(y_i^{bear}=1 \mid \boldsymbol{\theta}\right)=\frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}}}{1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}}-\frac{e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}{1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}\gamma^{bear}\frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}}}{1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}}+\frac{e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}{1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}}\gamma^{bear}\delta^{bear}=$$
$$=\frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\right)-e^{X_i^{bear}\boldsymbol{\beta}^{bear}}e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\gamma^{bear}+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\gamma^{bear}\delta^{bear}\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)}{\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\right)\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)}.$$

For $j=N_{bear},...,N_{bear}+N_{wolf}$:

$$P\left(y_j^{wolf}=1 \mid \boldsymbol{\theta},\boldsymbol{y}^{bear}\right)=\pi_j^{wolf}-y_j^{bear}\gamma^{wolf}\pi_j^{wolf}+y_j^{bear}\gamma^{wolf}\delta^{wolf},$$
$$P\left(y_j^{wolf}=1 \mid \boldsymbol{\theta}\right)=\mathbb{E}_{\boldsymbol{y}^{wolf}}\left[\pi_j^{wolf}-y_j^{bear}\gamma^{wolf}\pi_j^{wolf}+y_j^{bear}\gamma^{wolf}\delta^{wolf}\right]=$$
$$=\pi_j^{wolf}-\pi_j^{bear}\gamma^{wolf}\pi_j^{wolf}+\pi_j^{bear}\gamma^{wolf}\delta^{wolf}.$$

Since

$$\pi_j^{bear} = logit^{-1}\left(X_j^{wolf} \cdot \boldsymbol{\beta}^{bear}\right) = \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}}{1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}},$$
$$\pi_j^{wolf} = logit^{-1}\left(X_j^{wolf} \cdot \boldsymbol{\beta}^{wolf}\right) = \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}}{1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}},$$

we obtain:

$$P\left(y_j^{wolf}=1 \mid \boldsymbol{\theta}\right) = \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\right) - e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\gamma^{wolf}}{\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\right)\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\right)} + \\ + \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\gamma^{wolf}\delta^{wolf}\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\right)}{\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\right)\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\right)}.$$

We have that: δ^{bear} and $\delta^{wolf} \in \{0,1\}$. If they are both equal to 1 and also $\gamma^{bear} = \gamma^{wolf} = 1$, then

$$P\left(y_i^{bear}=1 \mid \boldsymbol{\theta}\right) = \frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\right) - e^{X_i^{bear}\boldsymbol{\beta}^{bear}}e^{X_i^{bear}\boldsymbol{\beta}^{wolf}} + e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)}{\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\right)\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)} = \\ = \frac{e^{X_i^{bear}\boldsymbol{\beta}^{bear}} + e^{X_i^{bear}\boldsymbol{\beta}^{wolf}} + e^{X_i^{bear}\boldsymbol{\beta}^{bear}+X_i^{bear}\boldsymbol{\beta}^{wolf}}}{\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{wolf}}\right)\left(1+e^{X_i^{bear}\boldsymbol{\beta}^{bear}}\right)},$$

and

$$P\left(y_j^{wolf}=1 \mid \boldsymbol{\theta}\right) = \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\right) - e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}e^{X_j^{wolf}\boldsymbol{\beta}^{bear}} + e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\right)}{\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\right)\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\right)} = \\ = \frac{e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}} + e^{X_j^{wolf}\boldsymbol{\beta}^{bear}} + e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}+X_j^{wolf}\boldsymbol{\beta}^{bear}}}{\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{bear}}\right)\left(1+e^{X_j^{wolf}\boldsymbol{\beta}^{wolf}}\right)}.$$

Also in this case, $\boldsymbol{\beta}^{bear}$ and $\boldsymbol{\beta}^{wolf}$ are interchangeable. Thus, the Fourth model is not identifiable if $\gamma^{bear} = \gamma^{wolf} = \delta^{bear} = \delta^{wolf} = 1$. However, if this particular condition is not verified, nothing can be said on the identifiability of the Fourth model from an analytical point of view.

Bibliography

Jerrold L Belant, Brad Griffith, Yingte Zhang, Erich H Follmann, and Layne G Adams. Population-level resource selection by sympatric brown and American black bears in Alaska. *Polar Biology*, 33(1):31–40, 2010.

William Henry Burt. Territoriality and Home Range Concepts as Applied to Mammals. *Journal of Mammalogy*, 24(3):346–352, 08 1943. ISSN 0022-2372. doi: 10.2307/1374834. URL https://doi.org/10.2307/1374834.

Paolo Ciucci, Vincenzo Gervasi, Luigi Boitani, John Boulanger, D Paetkau, R Prive, and E Tosoni. Estimating abundance of the remnant Apennine brown bear population using multiple noninvasive genetic data sources. *Journal of Mammalogy*, 96:206–220, 02 2015. doi: 10.1093/jmamma/gyu029.

Angela M Darnell, Jan A Graf, Michael J Somers, Rob Slotow, and Micaela Szykman Gunther. Space use of African wild dogs in relation to other large carnivores. *PloS one*, 9 (6):e98846, 2014.

John Fieberg, Johannes Signer, Brian Smith, and Tal Avgar. A 'how to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, 2021.

James D. Forester, H. Im, and P. Rathouz. Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, 90 12:3554–65, 2009.

Linton C. Freeman. Survey sampling. *Social Forces*, 45(1):132–133, 09 1966. ISSN 0037-7732. doi: 10.1093/sf/45.1.132-a. URL https://doi.org/10.1093/sf/45.1.132-a.

Mevin Hooten, Devin Johnson, Brett Mcclintock, and Juan Morales. *Animal Movement: Statistical Models for Telemetry Data*. 03 2017. ISBN 9781315117744. doi: 10.1201/9781315117744.

Douglas H. Johnson. The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, 61(1):65–71, 1980.

Robert E. Kass, Bradley P. Carlin, Andrew Gelman, and Radford M. Neal. Markov Chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, 52(2):93–100, 1998. doi: 10.1080/00031305.1998.10480547. URL https://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480547.

Sara Mancinelli, Luigi Boitani, and Paolo Ciucci. Determinants of home range size and space use patterns in a protected wolf (canis lupus) population in the central Apennines, Italy. *Canadian Journal of Zoology*, 96:828–838, 2018.

Bryan Manly, Lyman Mcdonald, Dana Thomas, Trent Mcdonald, and Wallace Erickson. *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*, volume 63. 01 2002. ISBN 1402006772. doi: 10.2307/5247.

Lyman L. McDonald Mark S. Boyce. Relating populations to habitats using resource selection functions. *Trends in Ecology & Evolution*, 14(7):268–272, 1999. ISSN 0169-5347. doi: https://doi.org/10.1016/S0169-5347(99)01593-1. URL https://www.sciencedirect.com/science/article/pii/S0169534799015931.

Théo Michelot, Paul G. Blackwell, and Jason Matthiopoulos. Linking resource selection and step selection models for habitat preferences in animals. *Ecology*, 100(1): e02452, 2019. doi: https://doi.org/10.1002/ecy.2452. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2452.

Cyril Milleret, Andrés Ordiz, Guillaume Chapron, Harry Peter Andreassen, Jonas Kindberg, Johan Månsson, Aimee Tallian, Petter Wabakken, Camilla Wikenros, Barbara Zimmermann, Jon E. Swenson, and Håkan Sand. Habitat segregation between brown bears and gray wolves in a human-dominated landscape. *Ecology and Evolution*, 8(23): 11450–11466, 2018. doi: https://doi.org/10.1002/ece3.4572.

Michael Mitchell, Roger Powell, and Janet Rachlow. Foraging optimally for home ranges. *Journal of Mammalogy*, 93:917–928, 08 2012. doi: 10.2307/23259838.

Cecilia Parracciani. Coesistenza tra grandi carnivori: analisi di segregazione dell'habitat tra orso e lupo nell'Appennino centrale, 2020.

Roger A. Powell and Michael S. Mitchell. What is a home range? *Journal of Mammalogy*, 93(4):948–958, 09 2012. ISSN 0022-2372. doi: 10.1644/11-MAMM-S-177.1. URL https://doi.org/10.1644/11-MAMM-S-177.1.

Marco Regis. Introduzione alla statistica Bayesiana, January 2020.

Brian D. Ripley. *Stochastic Simulation*. Wiley, 1987.

George Robert, Christian. Casella. *Monte Carlo Statistical Methods*. Springer, New York, NY, 1999.

Stephen M. Stigler. Chapter 24 - P.S. Laplace, Théorie analytique des probabilités, first edition (1812); Essai philosophique sur les probabilités, first edition (1814). In I. Grattan-Guinness, Roger Cooke, Leo Corry, Pierre Crépel, and Niccolo Guicciardini, editors, *Landmark Writings in Western Mathematics 1640-1940*, pages 329–340. Elsevier Science, Amsterdam, 2005. ISBN 978-0-444-50871-3. doi: https://doi.org/10.1016/B978-044450871-3/50105-4. URL https://www.sciencedirect.com/science/article/pii/B9780444508713501054.

Bill Venables and B Ripley. *Modern Applied Statistics With S*. 01 2002. doi: 10.1007/b97626.

Richard Von Mises. Probability, statistics and truth. *Journal of the Institute of Actuaries*, 70(2):266–267, 1939. doi: 10.1017/S0020268100011781.

David I. Warton and Leah C. Shepherd. Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *The Annals of Applied Statistics*, 4(3):1383 – 1402, 2010. doi: 10.1214/10-AOAS331. URL https://doi.org/10.1214/10-AOAS331.