Stockholms universitet

Spatial and Temporal point process models for estimation of brown bear population size

An attempt to separate the measures of density and search effort

Daria Medvedeva

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2021:15
Matematisk statistik
Juni 2021

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2021:15**
http://www.math.su.se

Spatial and Temporal point process models for estimation of brown bear population size

An attempt to separate the measures of density and search effort

Daria Medvedeva*

June 2021

Abstract

Monitoring of the population of brown bears is important for its management and preservation. The primary aim is to estimate the size of the population, distribution and trends within the population. The size is currently estimated by using Capture-Recapture (CR) models for closed populations.

In this thesis we propose a different model which has its roots in the Spatially Explicit Capture-Recapture models which make use of spatial information in data. The model consists of a spatial point process model for the activity centres of bears and a temporal point process model for the detection times of each bear. A measure of search effort is obtained through the temporal point process. The search effort is used to ensure the identifiability of the measure of density. The model is fit with a Bayesian approach using Markov Chain Monte Carlo sampling.

Results of a simulation study show that the proposed model works well for estimation of the parameters of density and search effort. An application to real data collected in the county of Västerbotten during autumn 2019 yields a more narrow 95% credible interval of the population size than the 95% confidence interval obtained by using the current CR models. In addition, spatially varying estimates of density, which are corrected for the spatial variation in search effort, are produced.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: dasha.medv@gmail.com. Supervisor: Martin Sköld.

Acknowledgements

I would like to thank my supervisor Martin Sköld for the opportunity to write this thesis at the Swedish Museum of Natural History, for introducing me to this interesting field and for his support, patience and guidance. I would also like to thank Claes Hällström for his valuable inputs and enormous support during these extraordinary times.

Contents

1 Introduction

Monitoring of low-density and elusive large carnivores, such as the brown bear, is important for management and preservation of these species. The primary aim of the monitoring is to estimate the population size, distribution and trends within the population in order to detect change. These estimates may also be used to determine yearly hunting quota and in order to preserve the desired population size the estimates are required to be as precise as possible.

Non-invasive Genetic Sampling (NGS) is a class of methods where DNA of individuals is collected through sources such as hair, droppings and saliva. The collection is made in a way which does not disturb the individuals by e.g. fixed detectors or area searches. In Sweden, monitoring and population estimates of brown bear are mainly based on NGS of bear droppings made by volunteers, since the method was introduced in 2001 (Bellemain et al. 2005). Management in Sweden takes place at the county level and population estimates are calculated each year in a different county, or counties, with established bear population. Every five years an estimate of the national population size is calculated using estimates of size and population trend in each county, see Jonas Kindberg et al. (2011) for calculations in 2008. The trend estimates are based on the Large Carnivore Observation Index which uses effort-corrected observations of bear made by moose hunters (Jonas Kindberg et al. 2009).

The Swedish Museum of Natural History (SMNH) were given the task to estimate the population size of brown bear in 2020, at county level, by using the currently established method which is based on Capture-Recapture (CR) models for closed populations. In conjunction with this, SMNH were interested in evaluating whether a new method and model could be obtained and lead to improved usage of data and analysis. The model was desired to make use of the spatial distribution of brown bears as well as be able to obtain and possibly separate measures of density and search effort in order to correct for uneven sampling. The aim of this masters' thesis is thus to evaluate if a model with the desired properties can be obtained, whether this model produces an estimate of the population size with good precision and whether a correction for uneven sampling is meaningful.

1.1 Background

CR methods are commonly used when the interest lies in estimating the total number of individuals present in a population. The basic idea of CR models is to use detections of individuals to model and estimate a capture probability p and the total population size N. The earliest class of these models developed are closed population models which assume that the population under observation is closed during the period of inventory, i.e. no individual dies, is born or moves over the boundary of the region of inventory (Amstrup et al. 2005, Ch. 1.2). Other classes such as open population models, where the population may change during the period of inventory, have been developed but will not be discussed further in this thesis (consult Amstrup et al. (2005) for more information). Detections of individuals, more commonly named *captures*, can be made on a number k of sampling occasions during a period of inventory which is limited in time. The *captures* of an individual i, consisting of a first *capture* and possible additional *recaptures*, are assumed to occur independently in time and space. On a sampling occasion j, $j = 1, \ldots, k$, the number of captures n_j is counted out of which u_j are a first capture and m_j are recaptures, where $m_1 = 0$. In the simplest case with $k = 2$, which is the origin of CR methodology (Amstrup et al. 2005, Ch. 2.3), $n_1 = u_1$ individuals are captured on the first occasion. These n_1 individuals constitute the unknown proportion

$$\frac{n_1}{N} \tag{1}$$

of the whole population N. On the second occasion n_2 individuals are captured out of which m_2 are recaptured. The m_2 individuals constitute the known proportion

$$\frac{m_2}{n_2} \tag{2}$$

of the n_2 captured individuals on the second occasion. It is assumed that the two proportions in Eq. 1 and Eq. 2 are approximately equal, i.e.

$$\frac{n_1}{N} \approx \frac{m_2}{n_2} \tag{3}$$

which gives the so called Petersen-Lincoln estimator, (Amstrup et al. 2005, Eq. 2.1, p. 27), of the population size N

$$\widehat{N}_P = \frac{n_1 n_2}{m_2}. \tag{4}$$

To extend the field of CR methods, beyond the estimator in Eq. 4, a variety of models have been developed for the case when $k > 2$. These can be divided into two groups, discrete-time models and continuous-time models.

Discrete-time models, as described in Ch. 4.2 of Amstrup et al. (2005), assume that captures can be made on a distinct number of occasions where the exact time of a capture is unknown and at most one capture per individual during an occasion is allowed. These models include the additional parameter P_{ij}, i.e. the capture probability for individual i on sampling occasion j. Variations in capture probability are considered and can originate from three sources: (1) time effects, e.g. uneven sampling effort between occasions; (2) behavioural responses to captures, e.g. individuals getting affected by their first capture; and (3) individual heterogeneity from either observable factors, e.g. sex or age, or unobservable characteristics. There are a total of eight models consisting of the so called null model with no variation and seven models reflecting all possible combinations of the sources for variation. The simplest model $\mathbf{M}_0$, assumes that the capture probability P_{ij} is equal for all individuals on each occasion and between occasions, i.e. $P_{ij} = p$, and contains the two parameters (p, N) to be estimated. Model $\mathbf{M}_t$, t stands for time effect, assumes that the capture probability P_{ij} is equal for all individuals on each occasion but varies between occasions, i.e. $P_{ij} = p_j$, and contains the $k + 1$ parameters $(p_1, \ldots, p_k, N)$. Model $\mathbf{M}_b$, b for behavioural response, assumes that P_{ij} changes for individuals after their first capture, i.e. for any individual $P_{ij} = p$ until its first capture and $P_{ij} = \tilde{p}$ after its first capture, and contains the three parameters $(p, \tilde{p}, N)$. Model $\mathbf{M}_h$, h for heterogeneity, assumes that P_{ij} is unique for each individual but does not change between occasions, i.e. $P_{ij} = p_i$, and thus contains the $N + 1$ parameters $(p_1, \ldots, p_N, N)$. The remaining four models are $\mathbf{M}_{tb}$ assuming $P_{ij} = p_j$ until first capture and $P_{ij} = \tilde{p}_j$ after first capture, $\mathbf{M}_{th}$ assuming $P_{ij} = p_i e_j$ where e_j is a time effect, $\mathbf{M}_{bh}$ assuming $P_{ij} = p_i$ until first capture and $P_{ij} = \tilde{p}_i$ after first capture and the most general model $\mathbf{M}_{tbh}$ assuming $P_{ij} = p_{ij}$ until first capture and $P_{ij} = \tilde{p}_{ij}$ after first capture.

Continuous-time models, as described in Ch. 4.3 of Amstrup et al. (2005), assume that captures can be made at any time during a fixed period of length T where the exact time t of a capture is recorded with one capture per occasion. These models include the additional parameter $\lambda_i^*(t)$, i.e. the capture rate of individual i in a small time interval around time t. As for the discrete-time models, there are a total of eight models which reflect the three different sources for variation of the capture rate. For example, the simplest model $\mathbf{M}_0$ assumes $\lambda_i^*(t) = \lambda$ and the most general model $\mathbf{M}_{tbh}$ assumes $\lambda_i^*(t) = \lambda_i \alpha(t)$ until first capture and $\lambda_i^*(t) = \phi \lambda_i \alpha(t)$ after first capture where $\alpha(t)$ represents time effects as a function in $(0, T)$, λ_i represents heterogeneity as a capture rate of individual i and ϕ represents the behavioural response to the first capture.

A limitation of the closed population CR models mentioned above is that variation due to spatial effects is not accounted for and density estimates reflecting variation in space cannot be obtained. Due to this limitation David L Borchers and Murray G Efford (2008) presented the Spatially Explicit Capture-Recapture (SECR) methods where a spatial component is introduced into the, likelihood based, CR models above. This is accomplished by the concept of activity centres of individuals, whose geographical locations are modelled by a spatial point process. The capture probability of an individual is then modelled as a function of the distance between a detector, i.e. a fixed location where animals can be captured/detected, and its activity centre. This function is such that the probability of capture decreases with increasing distance between a detector and an individuals activity center. Captures of individuals at different detectors are thus modelled conditionally on their activity centres. Since sampling processes for SECR methods do not always include fixed detectors, Murray G. Efford (2011) modified the general SECR model and presented likelihood based SECR methods for data collected through area searches. The region of inventory is divided into a number of polygons and the capture probability of an individual is modelled as a function of the quantitative overlap of a polygon and its home range. This modification was done after Royle and Young (2008) introduced Bayesian SECR implementations for the Binomial model *complete-data* likelihood, as described in D. Borchers and Fewster (2016), for the same type of data. *Complete-data* likelihood refers to a likelihood as if all activity centres in the population of interest were observed. The Bayesian SECR approach for Binomial

model *complete-data* likelihood thus uses data augmentation where an unknown number of unobserved individuals are included and a vector of binary variables is added to indicate if an individual was observed or not. The number of unobserved individuals can be large, which makes data augmentation costly in term of computational efficiency. Due to this, King et al. (2016) developed a more efficient Bayesian implementation by using *semi-complete-data* likelihood approach, i.e. likelihood as if activity centres of only captured individuals were observed, for the Binomial model without data augmentation.

The NGS method used in Sweden, when collecting brown bear droppings for individual identification, is a type of sampling process without fixed detectors. The collection of droppings is mostly done by volunteers resulting in a type of data called opportunistic data where search effort is often unknown. It is desirable that a large number of volunteers participate, to get a good representation of the reality and large amount of samples for better precision of estimates, and too demanding sampling regulations can reduce the number of participants. Thus, opportunistic data are often less structured and come with sampling bias. The bias can arise from different sources such as sampling method, effort, species and habitats sampled (Dickinson et al. 2010), where variation in effort can arise from uneven sampling intensity over time, uneven spatial coverage, uneven sampling effort per occasion and uneven detectability (Isaac et al. 2014). Monitoring programs allowing any amount of effort in sampling can result in over-reporting, under-reporting or failure to report all observations. This can lead to conclusions of analysis which do not reflect the actual biological patterns correctly and reflect the variation in effort instead (Dickinson et al. 2010). Further, uneven spatial coverage can results in conclusions which reflect the spatial sampling bias instead of the actual spatial differences (Dickinson et al. 2010). In the case of the collection of brown bear droppings in Sweden, the problem with spatial sampling bias arises e.g. in mountain areas which few volunteers visit. The result is that these areas are being subject to sampling with less intensity than other areas. Thus, to avoid biased estimates there is a need to account for variation in search effort and Bischof et al. (2019), (see also Bischof et al. (2020)), developed such a model which provides an assessment of spatial variation in detection.

The CR models which are currently used in Sweden to estimate the population size of brown bear are discrete-time models although the data collection with NGS methods is made continuously during the period of inventory and not on a distinct number of sampling occasions. In order for the collected data to suit the discrete-time models the period of inventory is divided into smaller time intervals equivalent to distinct sampling occasions. Multiple captures of a bear during one occasion, i.e. time interval, are registered as one capture. This adaptation of data thus leads to information loss and possible sensitivity to how the period of inventory is divided. The CR models which are used can account for variations in sampling effort between occasions but not for spatial variations in sampling effort.

The model we propose can be though of as a modification of the Poisson model *semi-complete-data* likelihood approach described in D. Borchers and Fewster (2016). Due to the lack of fixed traps/detectors as well as fixed occasions in the process of data collection, this approach is not applied directly. The activity centres can be seen as detectors which can detect an individual several times, and in light of this the distance between activity centre and detector is zero and will not be modelled. Locations of activity centres are modelled with an inhomogeneous spatial Poisson process. Captures of individuals are modelled with an inhomogeneous temporal Poisson process, through which a measure of search effort is obtained. In this way, the spatial information in data is used as well as all capture information, the information loss of the current methods is avoided and spatial variation in search effort is accounted for. During the work on this thesis, a related approach by Zhang et al. (2020) has appeared, where the detection probability is modelled as a function of the distance between the activity centre and the detection locations. As in Zhang et al. (2020), the proposed model is fit with a Bayesian approach and Markov Chain Monte Carlo (MCMC) methods for sampling. A simulation study is conducted to evaluate if the proposed model works well for parameter estimation after which the model is applied to real data of brown bears collected in the county of Västerbotten in 2019.

2 Materials and methods

2.1 Data collection and structure

The collection of brown bear droppings for the inventory of the population is made yearly in different areas of the 2/3 northernmost part of Sweden. These areas consist of one or several counties with an established bear population and each county is responsible for the inventory of the brown bear approximately every fifth year, resulting in a gap between inventories in each county. The collection of droppings is made during the period between the 21st of August and the 31st of October and is a voluntary contribution of people who go out in the woods. This period is preferable for two reasons: (1) during this period of time a large number of hunters and berry pickers are out in the woods; (2) bears are less mobile during this period and eat a lot in the state of hyperphagia, before going into hibernation (Sahlén et al. 2006). This results in a larger density of droppings and a larger amount of visitors in the woods, which combined gives good conditions for a large amount of collections needed for better precision of estimates. The estimation of population size thus depends strongly on the amount of people participating in the collection. A more detailed description of the method of inventory can be found in e.g. Naturvårdsverket & Rovdata (2014).

Those who go into the woods receive a sampling kit for collection of droppings. When a dropping is found it is collected and the geographical location and coordinates as well as time of collection are noted manually on a paper. After collection, the sample with the note attached is sent directly to the SMNH for analysis. Each sample goes through an extraction of DNA from which it can determined if the DNA comes from a bear or not. If bear-DNA is confirmed the individual identity and possibly sex may or may not be determined. A likely reason for not being able to determine individual identity and sex is that the DNA is too decomposed, where the speed of decomposition depends on the diet of a bear and the temperature to which the dropping has been exposed. Further, too decomposed DNA can be a reason for not confirming bear-DNA in a sample. In order for a sample to be valid and used for estimation of population size it thus needs to be collected in the right way with all required information written down and needs to contain DNA intact enough to be confirmed

as bear-DNA and for individual identity to be determined. Thus, errors can occur already in the collection and DNA-analysis phases.

After collection and analysis the individual identity, location and time of collection and possibly sex for each collected dropping is known. The different locations (x_l, y_l) at which a particular bears' droppings have been found can be seen as being inside a circle with centre at the mean of those scattered locations. For each observed bear j, an unique spatial coordinate (x_j, y_j) is thus computed as the coordinate-wise arithmetic mean of the locations of its droppings, i.e. $(x_j, y_j) = \frac{1}{k_j} \sum_{l=1}^{k_j} (x_{lj}, y_{lj})$, where k_j is the number of locations at which droppings of bear j have been found. This corresponds to our notion of an activity centre (Royle et al. (2013), Ch. 1.5.3), the location where the bear is assumed to have its home. The activity centre of a bear can replace the locations (x_l, y_l) as a detector which can observe this bear several times. Each observed bear has one activity centre resulting in number of observed activity centres equal to the number of uniquely observed bears m. As a result, k_j also corresponds to the number of times bear j has been observed during the whole search period. The following information is thus available from the collected data:

- m - the number of bears observed at least once during the whole search period.
- k_j - number of times bear j has been observed during the whole search period, $j = 1, \ldots, m$.
- (x_j, y_j) - geographical location of the activity centre of an observed bear j, $j = 1, \ldots, m$.
- t_{lj} - recorded time of observation l of bear j, $l = 1, \ldots, k_j$, $j = 1, \ldots, m$.

2.2 Model

From the previous section we know that the data has spatio-temporal dimensions. The geographical locations (x, y) of activity centres are limited by the region of inventory B, i.e. the spatial dimension is limited to $B \in \mathbb{R}^2$. The recorded times of observations t of bears are limited by the period of inventory $[0, T]$ of length T, i.e. the temporal dimension is limited to $[0, T] \in \mathbb{R}$. In the following sub-sections we will describe the model of (x, y) and t in more detail. The notations which will be used in the description are summarized below:

- B - whole region of inventory.
- T - length of inventory period.
- B_i - sub-region i of B, $i = 1, \ldots, r$.
- a_i - area of sub-region i.
- N - the total number of bears in B.
- M - the number of bears observed at least once in B.
- M_0 - the number of non-observed bears in B.
- N_i - the total number of bears, in sub-region i.
- M_i - the number of bears observed at least once in sub-region i.
- $M_{0,i}$ - the number of non-observed bears in sub-region i.
- K_j - number of observations of bear j, $j = 1, \ldots M$.
- $\mu(x, y)$ - intensity of the point process model for locations (x, y) of activity centres.
- $q(x, y)$ - probability of observing a bear at lest once at location (x, y) of its activity centre.
- μ_i - intensity of the point process for locations of activity centres in sub-region i.
- q_i - probability of observing a bear at least once at location of its activity centre in sub-region i.
- $\lambda_i^*(t)$ - intensity of the point process for observation times of a bear at its activity centre in sub-region i.

2.2.1 Spatial Poisson process

We assume that the locations of activity centres of bears, observed and unobserved, in the whole region of interest B constitute an inhomogeneous spatial Poisson process with coordinate dependent intensity $\mu(x, y)$, corresponding to the average number of activity centres per km^2. If a bear with activity centre at (x, y) is observed at least once during the search period T with probability $q(x, y)$ and not observed at all with probability $1 - q(x, y)$ then, following standard results of thinning of a Poisson process (Illian et al. (2008), Ch. 6.2.1), the locations of activity centres of observed bears and the non-observed bears constitute two independent inhomogeneous spatial Poisson processes with intensities $\mu(x, y)q(x, y)$ and

$\mu(x,y)(1-q(x,y))$ respectively.

Further, splitting the whole region of interest B into r disjoint sub-regions $B_1,\ldots,B_r$ and letting $\mu(x,y)$ and $q(x,y)$ be constant within regions, i.e. letting $\mu(x,y)=\mu_i$ and $q(x,y)=q_i$ for all $(x,y)\in B_i$, $i=1,...,r$, yields that locations of activity centres in sub-region i constitute a homogeneous spatial Poisson processes with intensity μ_i. It follows that locations of activity centres of observed and non-observed bears in sub-region i constitute homogeneous spatial Poisson processes with intensities $\mu_i q_i$ and $\mu_i(1-q_i)$ respectively. Note that all activity centres in the whole region B constitute an inhomogeneous spatial Poisson process with region dependent intensity. Since the sub-regions are disjoint the sub-regional processes are independent.

In this sub-regional setting, the number of bears N_i in sub-region i follows a Poisson distribution (Illian et al. (2008), Ch. 2.3.1), $N_i \sim \text{Pois}(\mu_i a_i)$, where

$$\mu_i a_i = \int_{B_i} \mu_i \, dxdy. \tag{5}$$

The number a_i in Eq. 5 is the area of region i and μ_i is the average number of bears per km^2 in region i. Following the same reasoning yields that the number of observed bears M_i in sub-region i follows a Poisson distribution with parameter $\mu_i a_i q_i$, i.e. $M_i \sim \text{Pois}(\mu_i a_i q_i)$, and the number of non-observed bears $M_{0,i}$ in sub-region i follows a Poisson distribution with parameter $\mu_i a_i(1-q_i)$, i.e. $M_{0,i} \sim \text{Pois}(\mu_i a_i(1-q_i))$, where q_i can be seen as a measure of search effort.

2.2.2 Temporal Poisson Process

The number of bears in a sub-region is unknown. What is known is the number of observed bears, but without further information it can't be deduced if the magnitude of this number is primarily a consequence of the parameter μ or q, i.e. we do not know if the number of bears observed primarily depends on the average number of bears per km^2 or the search effort. An attempt to resolve this issue is to make use of the times at which each bear has been observed by modelling a temporal Poisson Process.

We assume that the recorded times t of collection of droppings from a bear, with activity centre (x, y) in sub-region i, constitute an inhomogeneous temporal Poisson process with intensity $\lambda_i^*(t)$, and that these processes of bears are independent. During the whole search period T, the number of observations K_j of bear j, given that its activity centre (x, y) is in sub-region i, follows a Poisson distribution (Ross (2014), Ch. 5.4.1), $K_j \sim \text{Pois}(\lambda_i)$, where

$$\lambda_i = \int_T \lambda_i^*(s)\, ds\,. \tag{6}$$

In this setting, the number of observations of a bear is allowed to equal zero. Since we do not observe bears for which the number of observations is zero, we have to modify the Poisson distribution above by conditioning on the event $K_j > 0$. The number of observations K_j of bear j, conditionally on the event that it has been observed at least once, follows a Zero-Truncated Poisson distribution with parameter given by Eq. 6, $K_j \sim \text{TruncPois}(\lambda_i)$, whose probability mass function is given by

$$p(k_j|i) = \frac{\tilde{p}(k_j|i)}{1 - \tilde{p}(0|i)}\,, k_j = 1, 2, \ldots. \tag{7}$$

The probability mass function $\tilde{p}(k_j|i)$ in Eq. 7 is of a $\text{Pois}(\lambda_i)$ distribution. The expected number of observations of a bear in region i, conditionally on the event that it has been observed at least once, is thus $\lambda_i/(1 - e^{-\lambda_i})$.

2.2.3 Combining the processes

From the spatial Poisson process above we have that the probability of observing a bear, with activity centre in sub-region i, at least once is q_i which is the same as the probability of that bear having at least one observation during the whole search period, that is $1 - \tilde{p}(0|i) = 1 - \exp(-\lambda_i)$. The spatial and temporal Poisson processes are thus combined through the equality

$$q_i = 1 - \exp(-\lambda_i) \tag{8}$$

where $1 - q_i = \exp(-\lambda_i)$ therefore is the probability of not observing a particular bear in sub-region i which is the probability of that bear having zero observations during the whole search period. Thus, the temporal process contains information on parameter q_i through the equality in Eq. 8, which together with the spatial process ensures the identifiability of the parameter of interest μ_i.

2.2.4 Spatial dependencies

In some cases, the number of observed bears and the number of their captures can be small in some sub-region(s) and it can be meaningful to introduce dependencies between sub-regions. This is done in order to reflect the assumption that neighbouring sub-regions tend to have more similar values of bear density and search effort than sub-regions far apart from each other.

This can be done by first noting that μ_i and λ_i, $i = 1, \ldots, r$, can be represented as linear models with linear predictor consisting of smooth functions of spatial covariates s_i, $i = 1, \ldots, r$, which in our case are the sub-regions so that $s_i \in \{1, \ldots, r\}$. For simplicity we will only show the concept for parameters μ_i. In light of Generalised Additive Models (GAMs) (S. N. Wood 2017), the smooth function $f(s_i)$ in region i is related to μ_i through a log-link function, i.e.

$$\log(\mu_i) = f(s_i) + \epsilon_i, \tag{9}$$

where $\epsilon_i \sim N(0, \sigma^2)$. The function $f(s_i)$ is often denoted simply as a regression coefficient, say α_i, so that $f(s_i) = \alpha_i$ is the spatial coefficient of region i (Ludwig Fahrmeir et al. 2013, Ch. 8.2.4). In this way we get the matrix formulation of Eq. 9 as follows

$$\log(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{10}$$

where $\boldsymbol{X}$ is the model matrix with $X_{ij} = 1$ if $i = j$ and 0 otherwise.

Further, in order for neighbouring regions to have similar estimates of μ_i and thus similar

estimates of α_i we introduce the penalty

$$J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \boldsymbol{S} \boldsymbol{\alpha}, \tag{11}$$

where $S_{ii} = n_i$ where n_i is the number of neighbouring regions of region i, $S_{ij} = S_{ji} = -1$ if region i and j are neighbours and $S_{ij} = S_{ji} = 0$ if they are not. The penalty in Eq. 11 is used to form a penalized least squares criterion as described in Ch. 4.2.2 of S. N. Wood (2017) and Ch. 8.2.4, p.522-523 of Ludwig Fahrmeir et al. (2013). This penalty is a sum of squared differences between parameters α of neighbouring regions, and minimizing the sum is the same as minimizing the differences between parameters of neighbouring regions.

The penalty in Eq. 11 is employed to reflect a prior belief, and can also be expressed in a Bayesian way by setting an (improper) Gaussian distribution prior, i.e. a prior which does not integrate to one, on the vector of spatial parameters $\boldsymbol{\alpha}$ as

$$\boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, \tau \boldsymbol{S}^{-1}) \tag{12}$$

where the matrix $\boldsymbol{S}$ is sparse and does not have full rank, and τ is a precision parameter (S. N. Wood (2017), Ch. 5.8.1). This type of Gaussian distribution is called an (intrinsic) Gaussian Markov Random Field (GMRF), with precision matrix $\boldsymbol{S}$ containing the conditional independence structure which reflects the neighbourhood information in a vector. The precision parameter τ controls the degree of smoothness in the field. For more general theory, definitions and construction of GMRFs consult Rue and Held (2005) as well as Ludwig Fahrmeir et al. (2013) (Ch. 8.2.4).

2.2.5 Prior distributions

In the Bayesian approach the parameters μ_i and λ_i are assumed to be random variables and need to be assigned probability distributions to reflect prior beliefs about them. These distributions are called prior distributions, or simply priors. Priors can have different forms depending on convenience and prior information available (Carlin and Louis 2008, Ch. 2.2).

Conjugate priors, where the prior is conjugate with the data likelihood, lead the posteriors to belong to the same family of distributions as themselves and are often more computationally convenient. In our case with Poisson and Zero-Truncated Poisson likelihoods, we choose the gamma priors $\mathrm{G}(a, b)$ and $\mathrm{G}(c, d)$, with mean a/b and c/d as well as variance a/b^2 and c/d^2, for the parameters μ_i and λ_i.The gamma prior for the parameter μ_i of the Poisson likelihood is conjugate but not for the parameter λ_i since its posterior is not a gamma distribution. The values of a, b, c and d are chosen to reflect prior beliefs about the possible values of μ_i and λ_i.

Prior belief that parameters of neighbouring regions have more similar values than non-neighbouring ones are incorporated by using GMRF priors instead of the gamma priors. In that case, we set the GMRF prior $\mathrm{N}(\mathbf{0}, \tau \boldsymbol{S}^{-1})$ on the vector of spatial parameters $\boldsymbol{\alpha}$ from the linear model of $\log(\boldsymbol{\mu})$ in Eq.10. Following the same reasoning as for $\log(\boldsymbol{\mu})$ in section 2.2.4, we let $\log(\boldsymbol{\lambda}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, and set the GMRF prior $\mathrm{N}(\mathbf{0}, \rho \boldsymbol{S}^{-1})$ on the vector of spatial parameters $\boldsymbol{\beta}$, where ρ is a precisions parameter in the same way as τ. Further, priors are set on the precision parameters τ and ρ of the GMRFs, which are called hyperpriors and are most commonly gamma distributions. For more details about Bayesian implementation and applications of GMRFs consult L. Fahrmeir and Kneib (2011) (Ch. 5).

2.3 Computation

The model with the sub-regional setting was fit in R (R Core Team 2020) using the package 'rjags' (Plummer 2019), an interface from R to JAGS, which uses Bayesian MCMC. The Poisson distribution was written using the built in distribution available in the JAGS language and the Zero-Truncated Poisson distribution, which is not a standard distribution in JAGS, was written using the zero-trick (Lunn et al. (2012), Ch. 9.5.1).

The GMFR was constructed using the function `jagam` (S. N. Wood 2016) in the package 'mgcv' (S. Wood 2019), for GAMs, in a similar way as the package 'secrgam' (D. L. Borchers and D. Kidney 2014) does to fit regression spline models using the package 'secr' (M. Efford 2020), for SECR models. The function `jagam` returns JAGS model code with appropriate priors as well as the model and precision matrix which are parametrized in accordance with

S. N. Wood (2017) (Ch. 1.8.1, 5.4.1).

Inference about model parameters was made from a sequence of dependent samples generated from the posterior distribution. Inference about the number of non-observed bears was made from a sequence of dependent samples generated from the posterior predictive distribution (Carlin and Louis (2008), p.26).

3 Results

3.1 Simulation study

In order to evaluate if the suggested model works well for parameter estimation and if the proposed separation of μ and q works we conduct a simulation study. The value of μ is chosen from $\{1, 4\}$ and the value of λ is chosen from $\{2, 4\}$. This creates 4 different sub-regions with $(\mu_1, \lambda_1) = (1, 2)$, $(\mu_2, \lambda_2) = (1, 4)$, $(\mu_3, \lambda_3) = (4, 2)$ and $(\mu_4, \lambda_4) = (4, 4)$. The area of each region is set to 100 km^2. This corresponds to an expected number of individuals equal to 100 and 400 for regions with $\mu = 1$ and $\mu = 4$ respectively and an expected number of observations of one individual approximately equal to 2.31 and 4.07 for regions with $\lambda = 2$ and $\lambda = 4$. Thus, the expected number of observed individuals is approximately equal to 86.47 and 98.17 in the regions with $\mu = 1$ as well as 345.87 and 392.67 in the regions with $\mu = 4$. A larger value of search effort increases the expected number of observed individuals, even though the underlying expected number of individuals is the same.

For each region we simulate $m_1, \ldots, m_{100}$ realisations of M from the corresponding Poisson distribution. For each of these realisations of M we simulate $k_1, \ldots, k_{m_i}$, $i = 1, \ldots, 100$, realizations of K from the corresponding Zero-Truncated Poisson distribution. How values from a Zero-Truncated Poisson distribution were simulated can be seen in Dalgaard (2005). This amounts to 100 sets for each sub-region consisting of a value m for the number of bears observed and m values of k for the number of observations of each of the m bears observed.

The proposed model was fit to the simulated data with a G$(1, 1)$ prior set on all parameters. For each of the simulated data sets we ran three parallel chains of the MCMC sampling for 11,000 iterations and removed the 1000 first as burn in, resulting in 10,000 samples from the posterior distributions in each chain. The posterior means and medians, of the 100 obtained posterior distributions for each parameter in the 4 sub-regional setting, as estimators of the model parameters were evaluated using the frequentist approach of RMSE, relative bias and coverage of the 95% credible intervals (Morris et al. 2019). The results are displayed in the Table 1 where it can be seen that the posterior means and medians estimates are practically

unbiased with low values of RMSE. The coverage of credible intervals is overall roughly 95%, with not substantial smaller/larger values. It can thus be concluded that the proposed model works well for estimation of the proposed separated parameters of bear density and search effort.

Table 1: Simulation results with RMSE and relative bias of the posterior means and medians of the parameters. Proportion of 95% Credible intervals which include the data generating parameter values is displayed under Coverage.

	Means			Medians			
	Mean	RMSE	Rel. Bias	Mean	RMSE	Rel. Bias	Coverage
Region 1							
$\lambda_1 = 2$	1.985	0.1837	-0.0076	1.981	0.1843	-0.0096	0.92
$\mu_1 = 1$	1.019	0.0958	0.0187	1.014	0.0948	0.0141	0.99
Region 2							
$\lambda_2 = 4$	3.981	0.1785	-0.0048	3.978	0.1787	-0.0055	0.97
$\mu_2 = 1$	1.006	0.1026	0.0058	1.002	0.1025	0.0019	0.93
Region 3							
$\lambda_3 = 2$	2.003	0.0837	0.0013	2.002	0.0837	0.0008	0.98
$\mu_3 = 4$	3.980	0.2347	-0.0050	3.975	0.2350	-0.0062	0.91
Region 4							
$\lambda_4 = 4$	4.005	0.0934	0.0012	4.004	0.0934	0.0010	0.98
$\mu_4 = 4$	3.976	0.1843	-0.0059	3.972	0.1850	-0.0069	0.97

3.2 Application

The current method, used by SMNH 2020 for estimation of the population size of brown bear in the county of Västerbotten, is based on CR models for closed populations. This is the method that was used by the Scandinavian Brown Bear Research Project the previous times the inventory of brown bear was conducted in this county in 2014 (J. Kindberg and Swenson 2015) and 2009 (J. Kindberg and Swenson 2010). Inventory was also made in this county 2004 (J. Kindberg and Swenson 2006), where the population estimate was calculated using open population models. This estimate was further re-calculated using closed population models during the inventory in 2009. There is a variety of different closed population models which can incorporate time effects, behavioural responses and heterogeneity of individuals as sources for variation in the capture probability P_{ij} (Amstrup et al. (2005), Ch. 4). The models which are currently considered are discrete-time models, as described in Section 1.1, which incorporate time effects t as well as possibly heterogeneity h and gender-effects sex. Heterogeneity is incorporated through separating individuals into two classes, the

easy captured and the elusive, and is denoted as $h2$. The different models are thus: $\mathbf{M}_t$, with temporal variation in P_{ij}; $\mathbf{M}_{t*sex}$, with temporal variation in P_{ij} for each sex; $\mathbf{M}_{th2}$, with temporal and heterogeneity variation in P_{ij}; $\mathbf{M}_{th2*sex}$, with temporal and heterogeneity variation in P_{ij} for each sex. The considered models are fitted and ranked according to their Akaike's Information Criterion (AIC) values, where the model with smallest AIC, i.e. highest rank, is chosen. The model with highest rank for the inventory in Västerbotten 2019 was $\mathbf{M}_{th2}$ and resulted in an estimate of 516 brown bears with a 95% confidence interval of 416-640 bears, (Åsbrink et al. 2020).

In order to demonstrate how the proposed model would perform on a real data set, we apply it to the data set of brown bear droppings collected in the county of Västerbotten during autumn 2019. This is the data set which was used to estimate the size of the brown bear population in Västerbotten during the spring of 2020 at SMNH, see Åsbrink et al. (2020) for full report. The data is available in the database Rovbase (www.rovbase.se), a joint Swedish-Norwegian database on large carnivore findings, and consists of 792 samples in which the individual identity of bears could be determined. This amounts to 73% of the total number of samples in which bear-DNA was found (1080 samples) and 64% of the total number of samples that were analysed for DNA (1240 samples). Out of the 792 samples 359 unique individuals were identified.

As stated in the background, Section 1.1, the collection of brown bear droppings in Sweden is made continuously during the inventory period and not during pre-specified occasions. Therefore, the data has to be adapted by dividing the inventory period into smaller time intervals equivalent to occasions. The present way is to divide the inventory period in calendar weeks, which resulted in a loss of approximately 20% (Åsbrink et al. 2020) of the 792 samples in the data set from Västerbotten in 2019. The proposed model avoids this loss of information and uses all samples of an individual as well as the spatial information given by the coordinates of samples. The data is modified by aggregating samples for each unique bear and calculating the arithmetic mean of the coordinates, resulting in a total number of observations and an activity centre. In this way, the time of collection of each sample is discarded and the interest lies in the number of observations during the whole search period.

The proposed model uses a division of the region of inventory into a number of smaller disjoint regions. A simple and straightforward choice of sub-regions in the county of Västerbotten are its municipalities, resulting in 15 sub-regions of different sizes. The coordinates of borders and areas of municipalities are taken from Statistiska Centralbyrån. Using the border coordinates of municipalities, it can be determined to which of them the activity centres belong. The municipalities and the distribution of activity centres from the inventory in Västerbotten 2019 are displayed in Figure 1. Two of the activity centres are located just outside the southern border of Västerbotten, and these are manually registered as belonging to the municipality nearest to them. As can be seen, the number of activity centres tends to be smaller in the northern and coastal parts of Västerbotten and larger in the southern and inland parts. In addition, Table 2 displays the number of activity centres in each municipality, where the index is set manually and does not correspond to the national index, and Figure 2 displays the frequency of the number of observations of each observed bear in the different municipalities. As can be seen, municipalities Malå and Umeå only have one observed bear with only one capture each. Overall, number of observation of each bear tends to be small, with median of 2 and mean of 2.2 observation per bear, and a higher value is desired for better precision of estimates.

Table 2: Number of unique individuals observed during autumn 2019 in each of the municipalities of Västerbotten together with the index of each munucipality, area and number of observed individuals per km^2.

Municipality	Id.	Number of observed ind.	Area (km^2)	Number/km^2
Nordmaling	1	13	1230.67	0.01056
Bjurholm	2	24	1306.73	0.01837
Vindeln	3	13	2630.14	0.00494
Robertsfors	4	3	1292.32	0.00232
Norsjö	5	8	1739.13	0.00460
Malå	6	1	1598.21	0.00063
Storuman	7	41	7298.88	0.00562
Sorsele	8	16	7366.62	0.00217
Dorotea	9	43	2764.44	0.01555
Vännäs	10	2	529.51	0.00378
Vilhelmina	11	49	8047.19	0.00609
Åsele	12	79	4223.45	0.01871
Umeå	13	1	2316.68	0.00043
Lycksele	14	62	5518.12	0.01124
Skellefteå	15	4	6802.34	0.00059

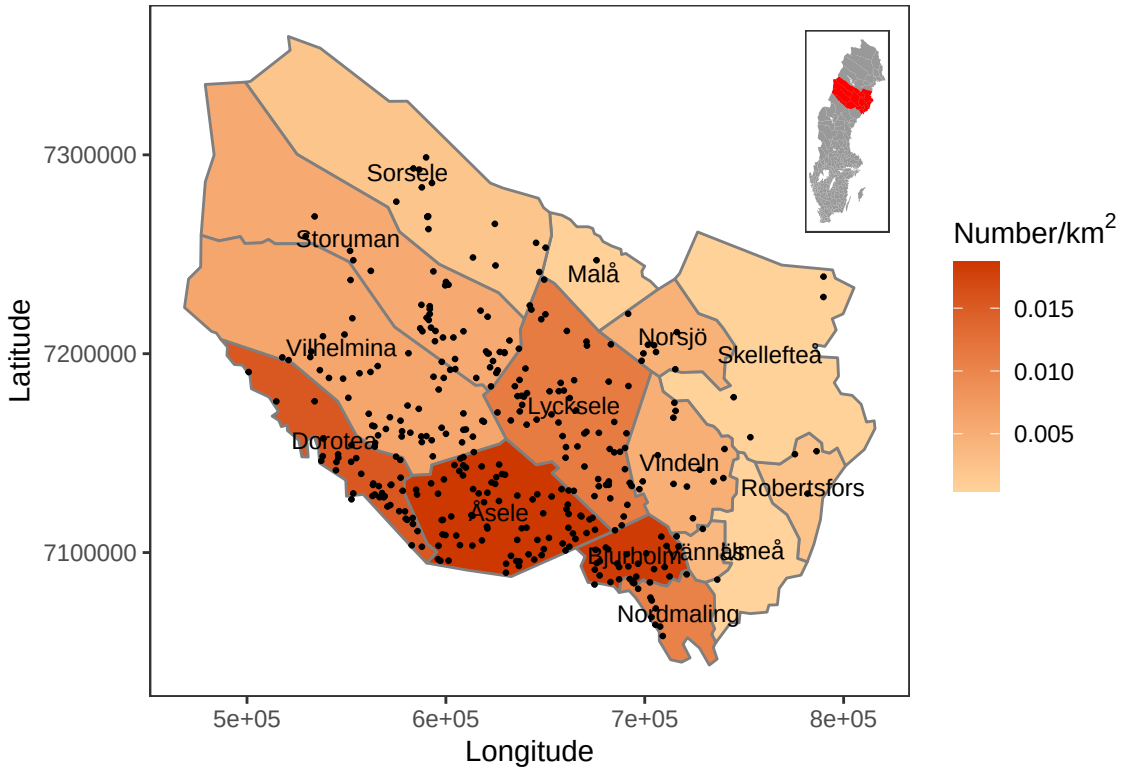

Figure 1: The county of Västerbotten with its municipalities coloured according to the number of observed individuals per km^2. Black dots respresent the activity centers of bears observed during the collection period in autumn 2019.

The proposed model was fit to the data with a G$(1, 1)$ prior, where the mean and variance is 1, set on all 15 parameters μ and a G$(2, 1)$ prior, where the mean and variance is 2, set on all 15 parameters λ. These prior distributions are relatively informative especially for the parameters λ, since the probability of these parameters being > 4 is apriori very small, but can be seen as reasonable first assumptions considering that the number of observations of a bear and the number of observed bears per km^2 tends to be small. Three parallel chains

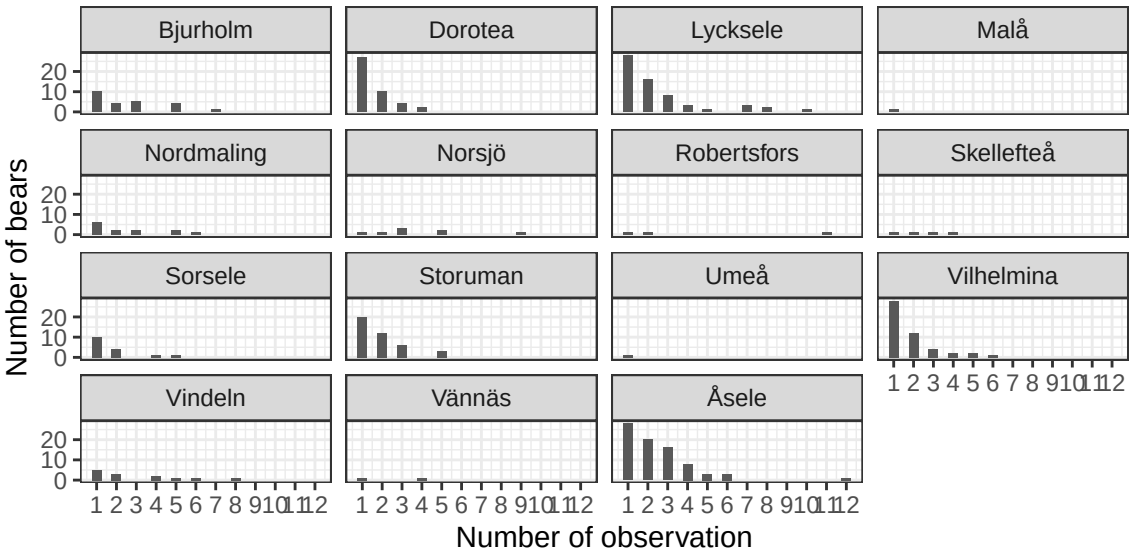

Figure 2: Frequency of the number of observations of each observed bear in the different municipalities of Västerbotten.

of the MCMC sampling were run for 50,000 iterations and the 25,000 first were removed as burn in, resulting in 25,000 samples from the posterior distributions in each chain. Visual investigations of trace plots and potential scale reduction factor $\hat{R}$ (Brooks and Gelman 1998; Gelman and Rubin 1992) were made as convergence checks for all parameters that were monitored, i.e. for N and μ_i, λ_i, $M_{0,i}$, $i = 1, \ldots, 15$. All $\hat{R}$ values were < 1.1 as desired for convergence and all trace plots showed good mixing with low autocorrelation within chains. Summary of the posterior statistics, including mean, standard deviation, median and 95% credible intervals, of the monitored parameters is presented in Table 3 and trace as well as density plot for N can be seen in Figure 6 in Appendix B. The posterior distribution of N is skewed and has a long right tail with small mass, where values as large as 5000 bears appear. This results in a moderate but skewed 95% credible interval with slightly high upper bound. This skewed distribution is arising from the posterior distributions of parameters μ in the two municipalities Malå and Umeå, which can be seen in Figures 7, 8 in Appendix B. The uncertainty in these two municipalities is due to the small number of observed bears. Also, posterior correlation between N and μ in municipalities Malå and Umeå are large compared with correlations with the remaining parameters. Thus, the uncertainty in the posterior distribution of N is mostly due to the small number of observed bears in Malå and Umeå.

Table 3: Summary of the estimation results when fitting the proposed sub-regional model to the bear data collected in the county of Västerbotten during autumn 2019.

				Credible Interval	
	Mean	SD	Median	2.5%	97.5%
N	481.612	95.741	468.000	434.000	607.000
μ_1	0.013	0.004	0.013	0.007	0.021
$M_{0,1}$	2.153	1.934	2.000	0.000	7.000
λ_1	2.128	0.431	2.103	1.361	3.046
μ_2	0.022	0.004	0.021	0.014	0.031
$M_{0,2}$	3.355	2.335	3.000	0.000	9.000
λ_2	2.199	0.328	2.185	1.599	2.883
μ_3	0.006	0.002	0.006	0.003	0.009
$M_{0,3}$	1.218	1.320	1.000	0.000	5.000
λ_3	2.649	0.469	2.624	1.800	3.640
μ_4	0.003	0.002	0.003	0.001	0.007
$M_{0,4}$	0.135	0.409	0.000	0.000	1.000
λ_4	3.903	1.025	3.820	2.149	6.130
μ_5	0.005	0.002	0.005	0.002	0.009
$M_{0,5}$	0.333	0.636	0.000	0.000	2.000
λ_5	3.550	0.652	3.512	2.382	4.943
μ_6	0.008	0.034	0.003	0.000	0.046
$M_{0,6}$	10.585	53.946	2.000	0.000	71.000
λ_6	0.812	0.742	0.598	0.027	2.751
μ_7	0.008	0.001	0.008	0.005	0.010
$M_{0,7}$	13.952	5.960	13.000	5.000	28.000
λ_7	1.428	0.217	1.419	1.028	1.880
μ_8	0.004	0.001	0.003	0.002	0.006
$M_{0,8}$	8.766	5.670	8.000	1.000	23.000
λ_8	1.172	0.315	1.148	0.629	1.857
μ_9	0.026	0.005	0.026	0.018	0.037
$M_{0,9}$	28.250	11.063	27.000	11.000	54.000
λ_9	0.976	0.185	0.965	0.646	1.367
μ_{10}	0.007	0.005	0.006	0.001	0.019
$M_{0,10}$	0.885	1.637	0.000	0.000	5.000
λ_{10}	1.980	0.907	1.855	0.591	4.087
μ_{11}	0.009	0.001	0.008	0.006	0.012
$M_{0,11}$	19.175	7.299	18.000	8.000	36.000
λ_{11}	1.315	0.193	1.306	0.962	1.718
μ_{12}	0.021	0.002	0.021	0.017	0.027
$M_{0,12}$	10.751	4.144	10.000	4.000	20.000

Table 3: Summary of the estimation results when fitting the proposed sub-regional model to the bear data collected in the county of Västerbotten during autumn 2019. *(continued)*

	Mean	SD	Median	2.5%	97.5%
λ_{12}	2.153	0.182	2.149	1.808	2.522
μ_{13}	0.006	0.033	0.002	0.000	0.037
$M_{0,13}$	12.768	77.468	2.000	0.000	82.000
λ_{13}	0.813	0.758	0.595	0.023	2.801
μ_{14}	0.013	0.002	0.013	0.010	0.017
$M_{0,14}$	9.255	3.944	9.000	3.000	18.000
λ_{14}	2.082	0.204	2.076	1.698	2.497
μ_{15}	0.001	0.000	0.001	0.000	0.002
$M_{0,15}$	1.032	1.519	1.000	0.000	5.000
λ_{15}	2.082	0.721	2.008	0.894	3.692

As an attempt to resolve this issue, spatial dependencies between neighbouring municipalities are introduced by replacing the gamma priors with GMRF priors. The function `jagam` returns model and precision matrices from a parametrized model, with an intercept which is independent of the remaining spatial parameters. The proposed model is fit to the data with a $\mathrm{N}(0, (0.00096)^{-1})$ and $\mathrm{N}(0, (0.017)^{-1})$ prior set on the intercept parameters denoted by α_1 and β_1 in the parametrizations of the linear models $\log(\boldsymbol{\mu})$ and $\log(\boldsymbol{\lambda})$ respectively. A $\mathrm{G}(0.05, 0.005)$ prior is set on the two precision parameters τ and ρ. All these priors are returned by the function `jagam` and not chosen manually. Three parallel chains of the MCMC sampling are run for 50,000 iterations, after an adaptation period of 10,000 iterations, and the 25,000 first are removed as burn in, resulting in 25,000 samples from the posterior distributions in each chain. Trace plots show good mixing and $\hat{R}$ values were below 1.1 and close to 1 for all monitored parameters, as desired for convergence. The summary of posterior statistics is presented in Table 4. The issue with the slightly high upper bound of the credible intervals is resolved and the posterior distribution of N is close to having zero skewness with difference between mean and median being less tha)n 1. The two posteriors of N, one obtained with gamma priors and one with GMRF priors respectively, are displayed in Figure 3. The estimated number of brown bears in the whole county of Västerbotten is 446 (median) with a 95% credible interval of 420-480 bears.

Table 4: Summary of the estimation results when fitting the proposed sub-regional model with GMRF-spline to the bear data collected in the county of Västerbotten during autumn 2019.

				Credible Interval	
	Mean	SD	Median	2.5%	97.5%
N	446.945	15.283	446.000	420.000	480.000
μ_1	0.011	0.003	0.011	0.006	0.018
$M_{0,1}$	1.593	1.514	1.000	0.000	5.000
λ_1	2.229	0.384	2.210	1.541	3.037
β_1	0.700	0.068	0.701	0.560	0.828
α_1	-5.255	0.107	-5.252	-5.470	-5.054
μ_2	0.019	0.004	0.019	0.013	0.028
$M_{0,2}$	2.832	2.022	3.000	0.000	8.000
λ_2	2.245	0.293	2.230	1.711	2.863
β_2	0.114	0.240	0.117	-0.373	0.575
α_2	-0.309	0.508	-0.292	-1.363	0.619
μ_3	0.005	0.001	0.005	0.003	0.008
$M_{0,3}$	1.120	1.197	1.000	0.000	4.000
λ_3	2.590	0.382	2.570	1.890	3.403
β_3	-0.419	0.143	-0.416	-0.701	-0.148
α_3	0.315	0.184	0.313	-0.039	0.676
μ_4	0.002	0.001	0.002	0.001	0.005
$M_{0,4}$	0.097	0.331	0.000	0.000	1.000
λ_4	3.703	0.956	3.605	2.119	5.821
β_4	0.021	0.110	0.022	-0.193	0.239
α_4	1.255	0.153	1.254	0.950	1.552
μ_5	0.005	0.002	0.004	0.002	0.008
$M_{0,5}$	0.409	0.703	0.000	0.000	2.000
λ_5	3.156	0.590	3.107	2.152	4.448
β_5	0.096	0.261	0.111	-0.463	0.583
α_5	-1.626	0.523	-1.583	-2.767	-0.739
μ_6	0.002	0.001	0.002	0.000	0.004
$M_{0,6}$	0.609	1.026	0.000	0.000	3.000
λ_6	1.817	0.632	1.763	0.734	3.249
β_6	-0.001	0.106	-0.003	-0.203	0.215
α_6	0.721	0.168	0.722	0.388	1.048
μ_7	0.007	0.001	0.007	0.005	0.010
$M_{0,7}$	12.933	5.351	12.000	4.000	25.000
λ_7	1.458	0.199	1.450	1.087	1.867
β_7	0.189	0.219	0.197	-0.271	0.598
α_7	-1.979	0.395	-1.960	-2.814	-1.259
μ_8	0.003	0.001	0.003	0.002	0.005

Table 4: Summary of the estimation results when fitting the proposed sub-regional model with GMRF-spline to the bear data collected in the county of Västerbotten during autumn 2019. *(continued)*

	Mean	SD	Median	2.5%	97.5%
$M_{0,8}$	6.276	3.935	6.000	1.000	16.000
λ_8	1.362	0.284	1.347	0.846	1.960
β_8	0.083	0.135	0.080	-0.175	0.354
α_8	1.145	0.226	1.148	0.701	1.579
μ_9	0.024	0.004	0.023	0.016	0.033
$M_{0,9}$	22.695	8.607	22.000	9.000	43.000
λ_9	1.085	0.181	1.075	0.758	1.463
β_9	0.223	0.147	0.223	-0.072	0.515
α_9	-0.190	0.266	-0.186	-0.725	0.324
μ_{10}	0.005	0.003	0.005	0.001	0.012
$M_{0,10}$	0.306	0.617	0.000	0.000	2.000
λ_{10}	2.375	0.620	2.310	1.340	3.768
β_{10}	0.558	0.252	0.560	0.052	1.042
α_{10}	-1.231	0.556	-1.193	-2.416	-0.257
μ_{11}	0.008	0.001	0.008	0.006	0.011
$M_{0,11}$	17.779	6.537	17.000	7.000	33.000
λ_{11}	1.359	0.178	1.353	1.027	1.725
β_{11}	0.414	0.189	0.413	0.051	0.791
α_{11}	-0.330	0.345	-0.322	-1.039	0.307
μ_{12}	0.021	0.002	0.021	0.017	0.026
$M_{0,12}$	11.281	4.224	11.000	4.000	21.000
λ_{12}	2.101	0.176	2.097	1.770	2.462
β_{12}	-0.184	0.350	-0.156	-0.971	0.429
α_{12}	-1.359	0.586	-1.304	-2.666	-0.348
μ_{13}	0.001	0.001	0.001	0.000	0.003
$M_{0,13}$	0.382	0.714	0.000	0.000	2.000
λ_{13}	2.348	0.648	2.298	1.216	3.804
β_{13}	-0.350	0.145	-0.348	-0.640	-0.077
α_{13}	0.186	0.188	0.185	-0.190	0.557
μ_{14}	0.013	0.002	0.012	0.009	0.016
$M_{0,14}$	9.043	3.756	9.000	3.000	17.000
λ_{14}	2.056	0.182	2.052	1.714	2.428
β_{14}	-0.431	0.205	-0.425	-0.857	-0.048
α_{14}	-0.697	0.262	-0.692	-1.219	-0.188
μ_{15}	0.001	0.000	0.001	0.000	0.002
$M_{0,15}$	0.591	0.902	0.000	0.000	3.000
λ_{15}	2.545	0.590	2.503	1.500	3.832
β_{15}	-0.650	0.174	-0.649	-1.004	-0.317
α_{15}	1.345	0.202	1.347	0.943	1.738

Table 4: Summary of the estimation results when fitting the proposed sub-regional model with GMRF-spline to the bear data collected in the county of Västerbotten during autumn 2019. *(continued)*

	Mean	SD	Median	2.5%	97.5%
τ	5.989	2.892	5.449	1.975	13.087
ρ	60.699	36.673	52.507	14.956	153.788

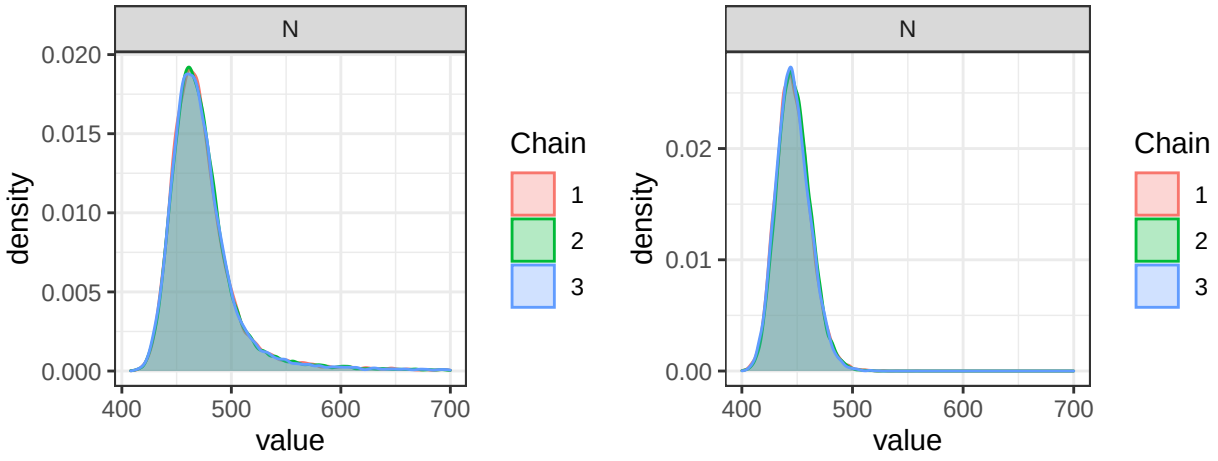

Figure 3: Posterior distributions of N. Left plot with gamma priors. Right plot with GMRF priors.

Concerning density and search effort, in Figure 4 we see the municipalities coloured according to the posterior means of μ and λ. The separation of density and search effort is meaningful, municipality Åsele (12) which had largest number of collected droppings in total and per km^2 does not have the largest estimate of density which is corrected for the search effort. Municipality Dorotea (9) has largest estimate of density when corrected for the low search effort. Overall, density is estimated to be higher in the southern and more inland parts of Västerbotten and search effort is estimated to be higher in the parts closer to the coast. In addition, the posterior standard deviations of μ and λ are displayed in Figure 5, where we can see that the posterior is widest for μ in Dorotea (9) where the mean is largest and widest for λ in Robertsfors (4) where the mean is largest, i.e. posterior means of these parameters are less precise than other.

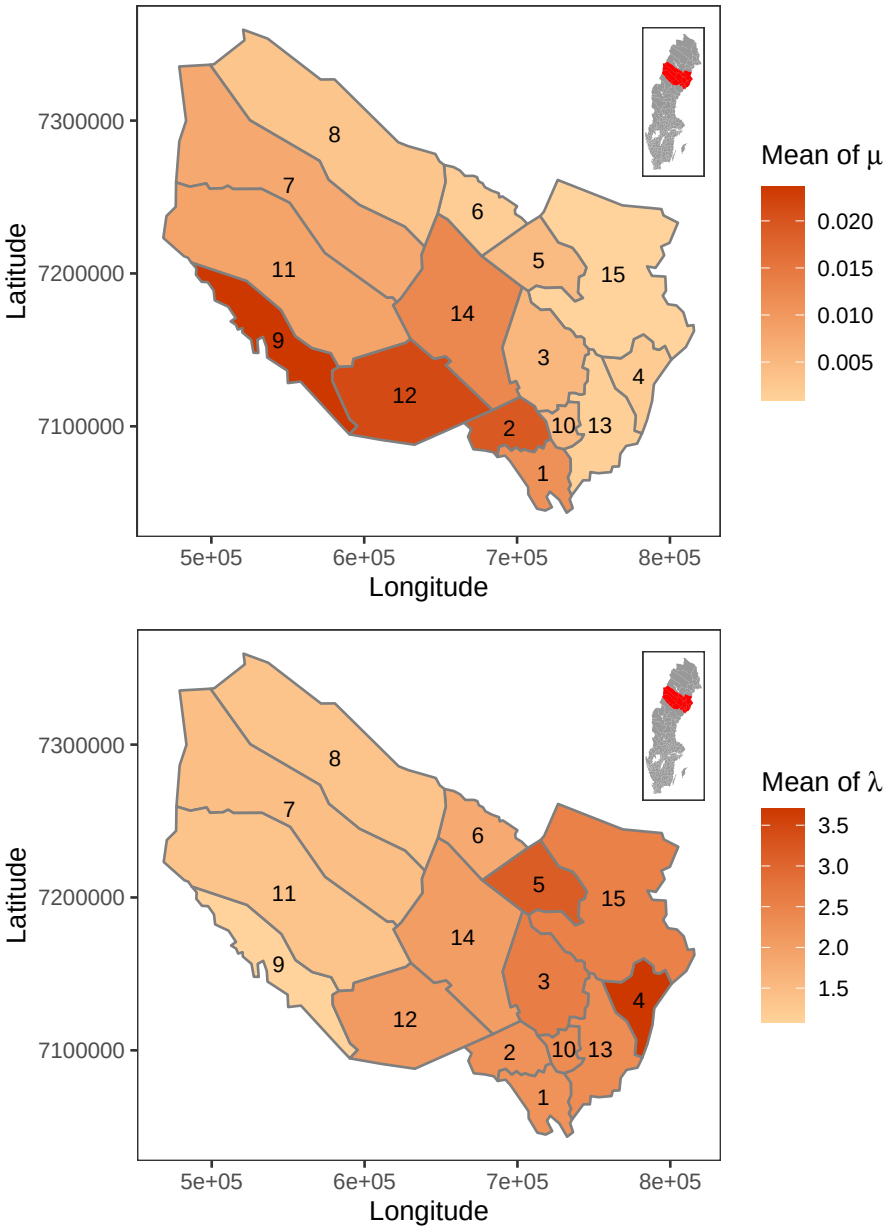

Figure 4: Municipalities in Västerbotten coloured according to the values of posterior means of μ (top) and λ (bottom).

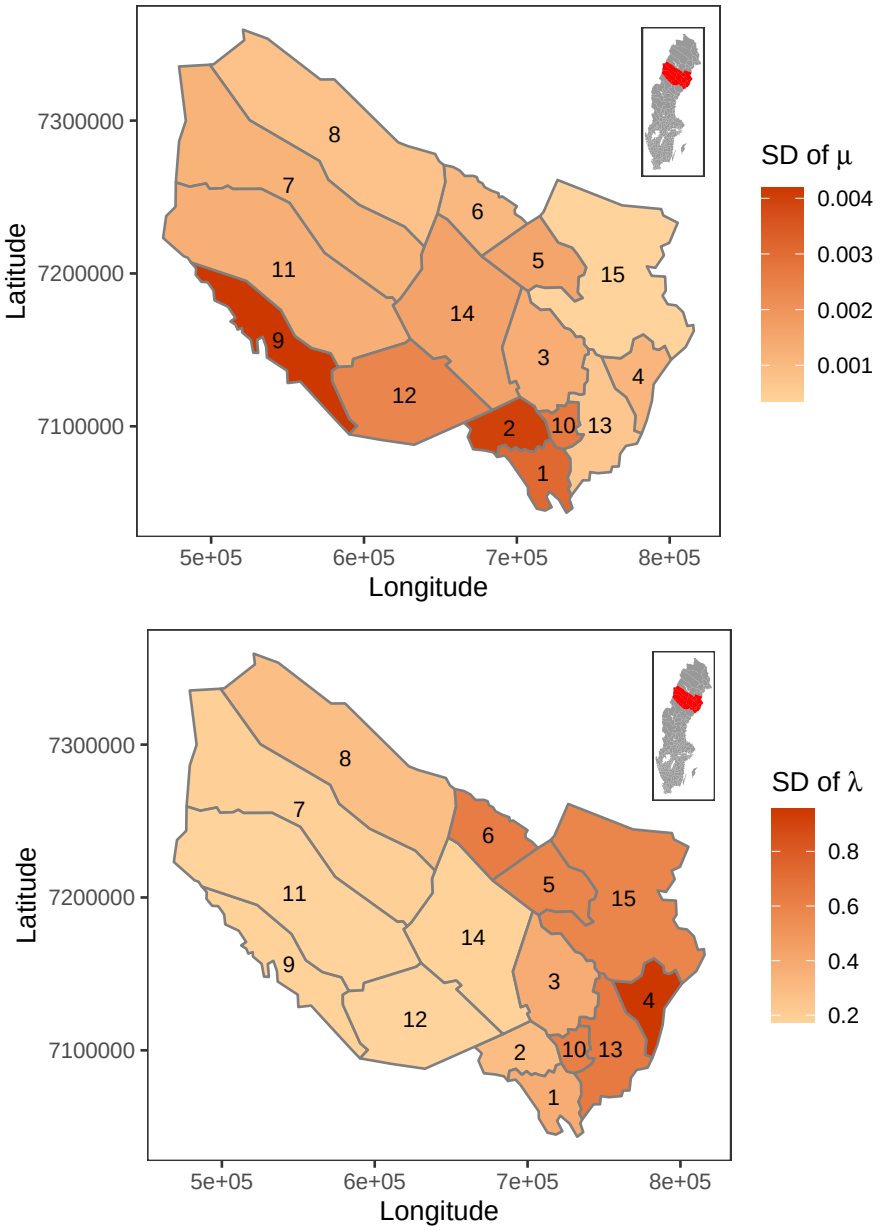

Figure 5: Municipalities in Västerbotten coloured according to the values of posterior standard deviations of μ (top) and λ (bottom).

4 Conclusions and discussion

In this thesis, we have developed and evaluated a spatial and temporal point process model, for Swedish NGS data of brown bear droppings collected in a fixed large region. The current method used in Sweden does not utilize the spatial information in data and leads to information loss in samples when dividing the search period into calendar weeks (Åsbrink et al. 2020). Existing SECR models are not directly applicable to the data in Sweden since the collection process does not comprise of fixed sampling occasions and fixed traps or detectors.

The proposed model takes the spatial information into account and does not require an arbitrary division of the search period into weeks. The loss of information which is present when using the current method is thus avoided. In addition, a measure of search effort can be obtained and separated from the density. This enables a correction of density by the spatial variation in search effort when estimating the population size. The model accounts for the uneven search effort in space in order to better reflect the true collection process and distribution of brown bears. By dividing the search area into sub-regions the need to calculate continuous coordinate dependent intensities of the point processes was removed. The model was formulated with an approach similar to the Poisson model *semi-complete-data* likelihood, as described in D. Borchers and Fewster (2016), which allows Bayesian implementations without data augmentation and without the need to specify prior assumptions on the total population size parameter N. Different prior distributions were set on density and detection rate through which assumptions about spatial dependencies between sub-regions could be incorporated as spatial covariates in a linear function.

A simulation study showed that the model works well for parameter estimation and that the measures of density and search effort can be separated. Application of the model on the data of brown bear from the county Västerbotten in 2019, with prior assumptions about spatial dependencies between municipalities, yielded a narrow 95% credible interval of 420-480 bears compared with the 95% confidence interval of 416-640 bears which was obtained when using the current method (Åsbrink et al. 2020). The separation of density and search effort as well as effort corrected density estimates showed to be meaningful. In addition, spatial

dependencies were useful when dealing with sub-regions in which counts were low.

Even though the proposed model yielded a more precise estimate of the total population size N than the one obtained by using the current models, there still remains a problem with the bears which move close to the border of the region of inventory. When a bear moves outside the county of inventory as well as in the county of inventory, it is not obvious to which county inventory this bear should be counted. The current way is to count all captured individuals as belonging to the county in which inventory is made. This can result in overestimation of the number of bears in the areas close to the border. Also, possibly not all droppings are collected from bears which cross county borders. This affects the model since these bears seem to have a smaller probability of recapture. A possible solution to this problem is to impose weights such that each bear counts as a fraction depending on where in the county the bear has its activity centre, e.g. a bear with activity centre close to the boundary is counted as half a bear. The fractions can be estimated based on e.g. a bivariate normal distribution fit to each bears capture locations. This is however beyond the scope of this thesis, but is an interesting problem for further research.

The approach with modelling density and detection rate as linear functions gives the ability to incorporate a variety of different covariates, such as sex in the case where it is interesting to keep track of the ratio between number of female and male bears. Other covariates that possibly can explain the variation in capture probability, such as distance from detection location to nearest road (Bischof et al. 2019) or other geographical parameters, can be of interest. This flexibility may have a potential influence on what is recorded during collection of droppings in future brown bear inventories.

The model which we have proposed and evaluated provides a relatively simple and flexible improvement to the current models used to estimate population size. It uses all capture information as well as spatial information in the data and produces effort corrected spatially varying estimates of density from which the number of unobserved bears is estimated. We see no apparent obstacle for this model to be applied to data from previous inventories in other counties with established brown bear population, and we recommend further explorations of this model and approach.

A Model code

```
################################################
# JAGS model with Gamma priors
################################################

BearModel <- "model{
  ## Poisson
  for (i in 1:length(m)) {
    # Link between q and lambda
    q[i] <- 1-exp(-lambda[i])
    # Distribution of number of observed ind. in region i
    m[i] ~ dpois(mu[i]*q[i]*a[i])
  }

  ## Trunc-Poisson
  C <- 10000
  for (j in 1:length(k)) {
    L[j] <- lambda[r[j]]^k[j]/((exp(lambda[r[j]])-1)*exp(logfact(k[j])))
    tau[j] <- -log(L[j]) + C
    zeros[j] ~ dpois(tau[j])
  }

  ## Posterior predictive
  for (i in 1:length(m)) {
    # Posterior predictive of number of ind. in region i
    n[i] <- m0[i]+m[i]
    # Posterior predictive of number of non-observed ind. in region i
    m0[i] ~ dpois(mu[i]*a[i]*(1-q[i]))
  }

  # Posterior predictive of number of ind. in total
  N <- sum(n)

  ## Priors
  for (i in 1:length(m)) {
    mu[i] ~ dgamma(s1,r1)
    lambda[i] ~ dgamma(s2,r2)
  }
}"
```

```
################################################
# JAGS model with GMRF
################################################

BearModelGMRF <- "model{
  ## Linear predictors
  mu.log <- X_mu %*% alpha
  lambda.log <- X_lambda %*% beta

  ## Poisson
  for (i in 1:length(m)) {
    # Link functions to linear predictors
    lambda[i] <- exp(lambda.log[i])
    mu[i] <- exp(mu.log[i])
    # Link between q and lambda
    q[i] <- 1-exp(-lambda[i])
    # Distribution of number of observed ind. in region i
    m[i] ~ dpois(mu[i]*a[i]*q[i])
  }

  ## Trunc-Poisson
  C <- 10000
  for (j in 1:length(k)) {
    L[j] <- lambda[r[j]]^k[j]/((exp(lambda[r[j]])-1)*exp(logfact(k[j])))
    theta[j] <- -log(L[j]) + C
    zeros[j] ~ dpois(theta[j])
  }

  ## Posterior predictive
  for (i in 1:length(m)) {
    # Posterior predictive of number of ind. in region i
    n[i] <- m0[i]+m[i]
    # Posterior predictive of number of non-observed ind. in region i
    m0[i] ~ dpois(mu[i]*a[i]*(1-q[i]))
  }

  # Posterior predictive of number of ind. in total
  N <- sum(n)

  ## Priors
  # Parametric effect prior: 1/32^2
  for (i in 1:1) { alpha[i] ~ dnorm(0,0.00096) }
  # Parametric effect prior: 1/7.7^2
  for (i in 1:1) { beta[i] ~ dnorm(0,0.017) }
```

```
  # priors for s(kod):
  K_alpha <- S_alpha[1:14,1:14] * tau[1]
  alpha[2:15] ~ dmnorm(zero_alpha[2:15],K_alpha)
  K_beta <- S_beta[1:14,1:14] * rho[1]
  beta[2:15] ~ dmnorm(zero_beta[2:15],K_beta)

  # smoothing parameter priors:
  for (i in 1:1) {
    tau[i] ~ dgamma(.05,.005)
    eta[i] <- log(tau[i])
    rho[i] ~ dgamma(.05,.005)
    kappa[i] <- log(rho[i])
  }
}"
```

B Complementary plots

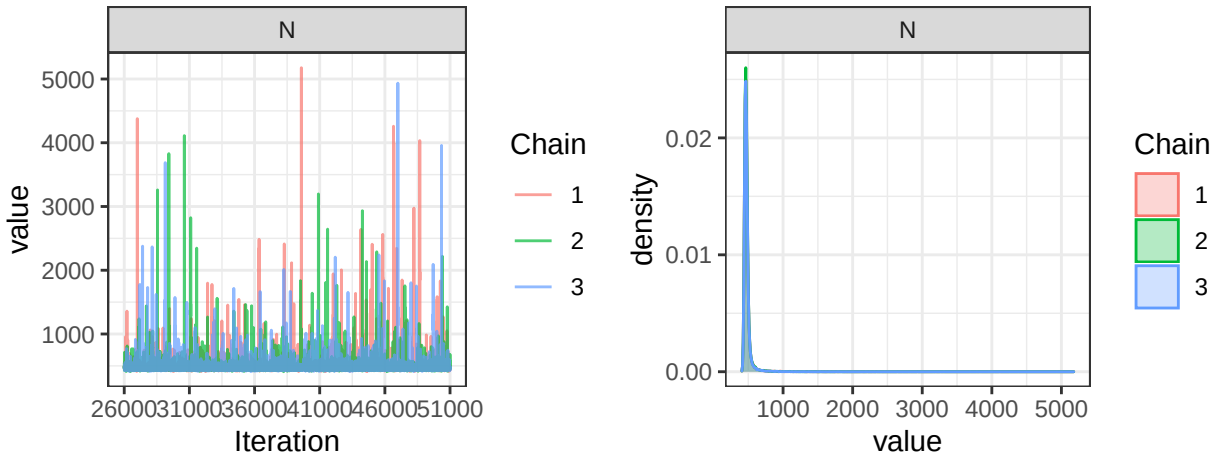

Figure 6: Trace and density plots of posterior samples from the proposed model for parameter N.

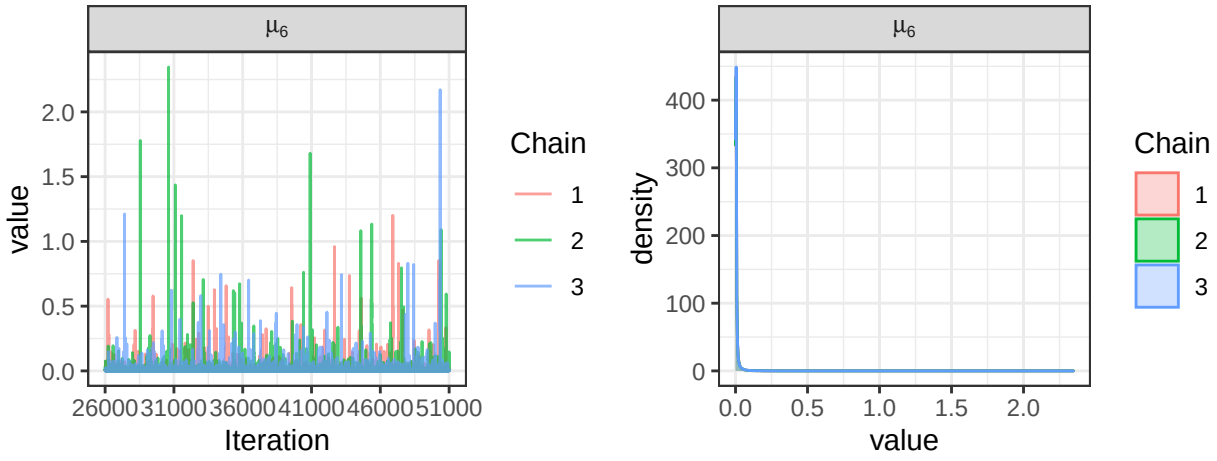

Figure 7: Trace and density plots of posterior samples from the proposed model for parameter μ in municipality Malå.

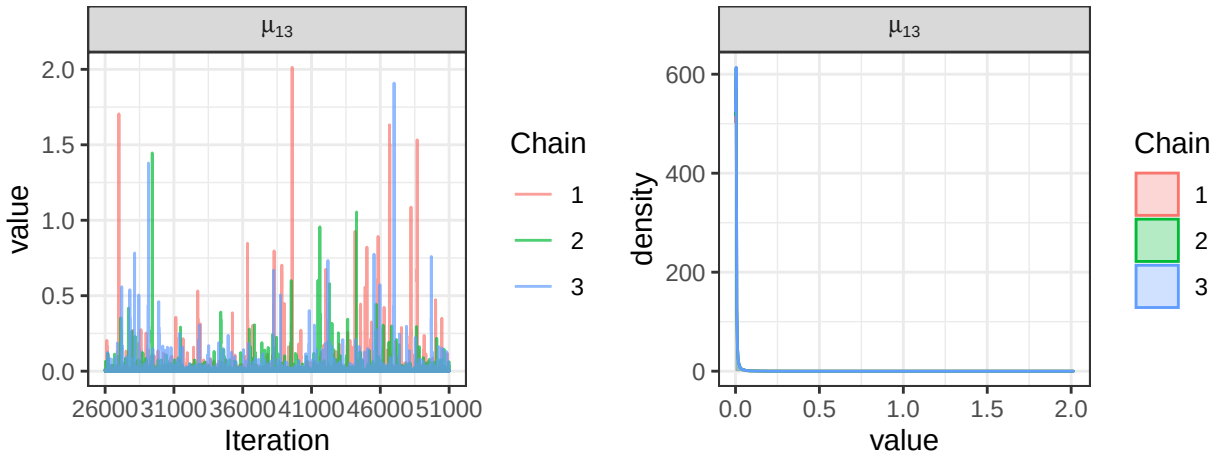

Figure 8: Trace and density plots of posterior samples from the proposed model for parameter μ in municipality Umeå.

References

Amstrup, Steven C., Trent L. McDonald, and Bryan F. J. Manly (2005). *Handbook of capture-recapture analysis*. Princeton: Princeton University Press. ISBN: 0691089671.

Åsbrink, Jessica, Martin Sköld, Thomas Källman, and Niclas Gyllenstrand (2020). *Resultat från inventeringen av brunbjörn i Västerbottens län 2019*.

Bellemain, Eva, Jon E. Swenson, David Tallmon, Sven Brunberg, and Pierre Taberlet (2005). "Estimating population size of elusive animals with DNA from hunter-collected feces: four methods for brown bears". In: *Conservation biology* 19.1, pp. 150–161.

Bischof, Richard, Cyril Milleret, Pierre Dupont, Joseph Chipperfield, Henrik Brøseth, and Jonas Kindberg (2019). "RovQuant: estimating density, abundance and population dynamics of bears, wolverines and wolves in Scandinavia". In:

Bischof, Richard, Cyril Milleret, Pierre Dupont, Joseph Chipperfield, Mahdieh Tourani, Andrés Ordiz, Perry de Valpine, Daniel Turek, J Andrew Royle, Olivier Gimenez, et al. (2020). "Estimating and forecasting spatial population dynamics of apex predators using transnational genetic monitoring". In: *Proceedings of the National Academy of Sciences* 117.48, pp. 30531–30538.

Borchers, D. L. and D. Kidney (2014). *Flexible density surface estimation for spatially explicit capture-recapture surveys*. Tech. rep. CREEM Technical Report, no. 2014-1, University of St Andrews.

Borchers, David L and Murray G Efford (2008). "Spatially explicit maximum likelihood methods for capture–recapture studies". In: *Biometrics* 64.2, pp. 377–385.

Borchers, David and Rachel Fewster (2016). "Spatial Capture–Recapture Models". In: *Statistical Science* 31.2, pp. 219–232. DOI: 10.1214/16-STS557.

Brooks, S. and A. Gelman (Dec. 1998). "General Methods for Monitoring Convergence of Iterative Simulations". In: *J. Comput. Graphi. Stat.* 7, pp. 434–455. DOI: 10.1080/10618600.1998.10474787.

Carlin, B.P. and T.A. Louis (2008). *Bayesian Methods for Data Analysis*. 3rd ed. Chapman & Hall/CRC Texts in Statistical Science. CRC Press. ISBN: 9781584886983.

Dalgaard, P. (May 1, 2005). *Simulate Zero-Truncated Poisson distribution. r-help mailing list.* URL: https://stat.ethz.ch/pipermail/r-help/2005-May/070680.html.

Dickinson, Janis L., Benjamin Zuckerberg, and David N. Bonter (2010). "Citizen science as an ecological research tool: challenges and benefits". In: *Annual review of ecology, evolution, and systematics* 41, pp. 149–172.

Efford, Murray (2020). *'secr': Spatially explicit capture-recapture models*. R package version 4.3.3. URL: https://CRAN.R-project.org/package=secr.

Efford, Murray G. (2011). "Estimation of population density by spatially explicit capture–recapture analysis of data from area searches". In: *Ecology* 92.12, pp. 2202–2207.

Fahrmeir, L. and Thomas Kneib (2011). *Bayesian smoothing and regression for longitudinal, spatial and event history data*. Oxford: Oxford University Press. ISBN: 9780191728501.

Fahrmeir, Ludwig, Thomas Kneib, Stefan Lang, and Brian Marx (2013). *Regression. Models, Methods and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 9783642343339.

Gelman, A. and D.B. Rubin (1992). "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4, pp. 457–472.

Illian, Janine, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan (2008). *Statistical analysis and modelling of spatial point patterns*. Vol. 70. John Wiley & Sons.

Isaac, Nick JB, Arco J. van Strien, Tom A. August, Marnix P. de Zeeuw, and David B. Roy (2014). "Statistics for citizen science: extracting signals of change from noisy ecological data". In: *Methods in Ecology and Evolution* 5.10, pp. 1052–1060.

Kindberg, J. and J. E. Swenson (Nov. 27, 2006). *Results from the genetic analyses performed on feces samples from bears in Västerbotten County, Sweden. Part II Population estimate.* Tech. rep. Report No. 2006:3 from the Scandinavian Brown Bear Research Project. URL: http://bearproject.info/wp-content/uploads/2015/11/2006-E68-3-Result-genetic-analyses-Vasterbotten-pop-estimate.pdf. published.

– (May 3, 2010). *Beräkning av björnstammens storlek i Västerbotten 2009*. Tech. rep. Report No. 2010:4 from the Scandinavian Brown Bear Research Project. URL: http://bearproject.info/wp-content/uploads/2015/11/2010-E89-4-Bj%C3%B6rnstammen-i-V%C3%A4sterbotten.pdf. published.

Kindberg, J. and J. E. Swenson (Nov. 9, 2015). *Björnstammens storlek i Västerbotten 2014*. Tech. rep. Rapport 2015:6 from the Scandinavian Brown Bear Research Project. URL: http://bearproject.info/wp-content/uploads/2016/01/2015-E131-6-Bj%C3%B6rnstammens-storlek-i-V%C3%A4sterbotten-2014.pdf. published.

Kindberg, Jonas, Göran Ericsson, and Jon E. Swenson (2009). "Monitoring rare or elusive large mammals using effort-corrected voluntary observers". In: *Biological conservation* 142.1, pp. 159–165.

Kindberg, Jonas, Jon E. Swenson, Göran Ericsson, Eva Bellemain, Christian Miquel, and Pierre Taberlet (2011). "Estimating population size and trends of the Swedish brown bear Ursus arctos population". In: *Wildlife Biology* 17.2, pp. 114–123. DOI: 10.2981/10-100.

King, Ruth, Brett T. McClintock, Darren Kidney, David Borchers, et al. (2016). "Capture–recapture abundance estimation using a semi-complete data likelihood approach". In: *The Annals of Applied Statistics* 10.1, pp. 264–285.

Lunn, David, Chris Jackson, Nicky Best, Andrew Thomas, and David Spiegelhalter (2012). *The BUGS book: A practical introduction to Bayesian analysis*. CRC press.

Morris, Tim P., Ian R. White, and Michael J. Crowther (2019). "Using simulation studies to evaluate statistical methods". In: *Statistics in medicine* 38.11, pp. 2074–2102.

Naturvårdsverket & Rovdata (2014). *BJÖRN: Övervakning i Skandinavien*. URL: http://naturvardsverket.se/Documents/publikationer6400/978-91-620-8710-4.pdf?pid=14094.

Plummer, Martyn (2019). *'rjags': Bayesian Graphical Models using MCMC*. R package version 4-10. URL: https://CRAN.R-project.org/package=rjags.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ross, Sheldon M. (2014). *Introduction to Probability Models*. Academic press.

Royle, J Andrew, Richard B. Chandler, Rahel Sollmann, and Beth Gardner (2013). *Spatial capture-recapture*. Academic Press.

Royle, J Andrew and Kevin V Young (2008). "A hierarchical model for spatial capture–recapture data". In: *Ecology* 89.8, pp. 2281–2289.

Rue, Håvard and Leonhard Held (2005). *Gaussian Markov random fields: theory and applications*. Boca Raton: Chapman & Hall/CRC Press. ISBN: 1584884320.

Sahlén, V., J. Swenson, S. Brunberg, and J. Kindberg (Dec. 25, 2006). *Björnen i Sverige. En rapport från Skandinaviska Björnprojektet till den Svenska Rovdjursutredningen. (In Swedish; The Brown Bear in Sweden. A report from SBBRP to Svenska Rovdjursutredningen.)* Tech. rep. Report No. 2006:4 from the Scandinavian Brown Bear Research Project. URL: http://bearproject.info/wp-content/uploads/2015/11/2006-E69-4-Bjornen-i-Sverige_Rapport-till-rovdjursfor-2006.pdf. published.

Wood, S. (2019). *'mgcv': Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. R package version 1.8-31.

Wood, Simon N. (2016). "Just another gibbs additive modeller: interfacing JAGS and mgcv". In: *arXiv preprint arXiv:1602.02539*.

– (2017). *Generalized additive models: an introduction with R*. 2nd ed. Boca Raton: CRC Press. ISBN: 9781498728331.

Zhang, Wei, Joseph D. Chipperfield, Janine B. Illian, Pierre Dupont, Cyril Milleret, Perry de Valpine, and Richard Bischof (2020). "A hierarchical point process model for spatial capture-recapture data". In: *bioRxiv*.