

Norwegian University of Life Sciences
Faculty of Environmental Sciences and Natural Resource Management

2024

ISSN 2535-2806

MINA fagrapport 96

Simultaneous range-wide genetic sampling of brown bears in Sweden: a pilot study

Cyril Milleret
Pierre Dupont
Jessica Åsbrink
Richard Bischof



Milleret, C., Dupont, P., Åsbrink, J., and Bischof, R., 2024. **Simultaneous range-wide genetic sampling of brown bears in Sweden: a pilot study** - MINA fagrapport 96. 23 pp.

Ås, April 2024

ISSN: 2535-2806

COPYRIGHT

©Norwegian University of Life Sciences (NMBU)

The publication may be freely cited where the source is acknowledged

AVAILABILITY

Open

PUBLICATION TYPE

Digital document (pdf)

QUALITY CONTROLLED BY

The Research committee (FU), MINA, NMBU

PRINCIPAL

Naturvårdsverket, Ref: NV-04419-21, Contact person: Robert Ekblom

COVER PICTURE

Female brown bear, Staffan Widstrand.

NØKKEORD

Ursus arctos, brunbjørn, bestandstetthet, oppdagbarhetssannsynlighet, ikke-invasiv innsamling av genetisk materiale, romlig fangst-gjenfangst, rovdyrforvaltning

KEY WORDS

Ursus arctos, brown bear, population density, detection probability, non-invasive genetic sampling, spatial capture-recapture, carnivore management

Cyril Milleret, Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway

Pierre Dupont, Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway

Jessica Åsbrink, Naturhistoriska riksmuseet, Enheten för populationsanalys och –övervakning, Box 50007, 104 05, Stockholm, Sweden

Richard Bischof (richard.bischof@nmbu.no), Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway

Summary

Background The Swedish brown bear (*Ursus arctos*) population is being monitored using non-invasive genetic sampling (NGS). Given the relatively large number of bears and the wide range they occupy in Sweden, NGS is logistically and financially costly. To spread effort and cost, bear monitoring in Sweden follows a five-year schedule where monitoring occurs successively in four distinct regions, followed by a year without sampling. This creates large spatial and temporal gaps in sampling and poses a challenge to the estimation of range-wide population sizes (Dupont et al., 2024). Estimation challenges are compounded by the fact that NGS is conducted opportunistically primarily by hunters without a direct measure of search effort.

Approach In this pilot study, we estimated the range-wide sampling effort required to obtain reliable population size estimates using spatial capture recapture (SCR) models. We did so using simulations based on current Swedish bear population and sampling characteristics. We considered two different scenarios with synchronous sampling in all four regions. The first scenario assumed that sampling was opportunistic (e.g., conducted by members of the public) and that only a proxy of spatio-temporal variation in search effort was available. The second scenario assumed that sampling was conducted in a structured fashion (e.g., by authorities) with known search effort. We simulated several levels of search effort intensity for both scenarios.

Results We found that an average of 1.5-2 spatial detections per individual detected throughout the entire Swedish bear population range should be sufficient to obtain robust range-wide and regional population size estimates. This would amount to approximately 5000-6000 DNA samples collected and analyzed each year. This estimated number of required samples accounts for genotyping failures. Our analysis also highlights that opportunistic sampling with inaccurate proxies of search effort can lead to an underestimation of population size at the regional and national levels. The severity of the underestimation increases as sampling intensity decreases.

Discussion We show that it is possible to obtain precise and accurate spatially-explicit estimates of the Swedish bear population with a reasonably low number of samples. This could be achieved by spreading the equivalent of the number of samples currently being collected across region C (Jämtland and Västernorrland) over the entire bear range in Sweden. Implementation of such synchronous range-wide sampling would however require solving logistic issues, including challenges arising from the prominent opportunistic component of bear monitoring in Sweden. Ours is a pilot study, with an overly simplistic model of sampling design. Additional analyses could adjust the spatio-temporal configuration of sampling to further improve estimation and cost efficiency. Nonetheless, our findings are promising and investigations into range-wide monitoring are worth pursuing further. Without them, reliable and complete population size estimates of the Swedish bear population will likely remain elusive.

Sammanfattning

Bakgrund Den svenska brunbjörnspopulationen (*Ursus arctos*) inventeras med hjälp av icke-invasiv genetisk provtagning (NGS). Med tanke på det relativt stora antalet björnar och det stora utbredningsområdet i Sverige är NGS både logistiskt och ekonomiskt kostsamt. För att sprida arbetsinsatser och kostnader följer övervakningen i Sverige ett femårigt schema där inventeringen sker i tur och ordning i fyra olika områden, följt av ett år där ingen inventering sker. Detta skapar stora rumsliga och tidsmässiga luckor i inventeringen vilket blir en stor utmaning när populationsuppskattningar som omfattar hela utbredningsområdet ska göras (Dupont et al., 2024). Skattningen försvåras dessutom av att NGS i första hand genomförs av jägare utan något direkt mått på sökinsatsen.

Metod I den här pilotstudien uppskattade vi den söksinsats som behövs över hela utbredningsområdet för att få tillförlitliga populationsuppskattningar med hjälp av SCR-modeller (spatial capture recapture). Det gjordes med hjälp av simuleringar baserade på den nuvarande svenska björnpopulationen och hur prover samlas in. Två olika alternativ med provinsamling i alla fyra områden samtidigt övervägdes. I det första alternativet antogs att provinsamlingen var opportunistisk (t.ex. utförd av allmänheten) och att endast en uppskattning av rumslig och tidsmässig sökinsats var möjlig. I det andra alternativet antogs att provinsamlingen genomfördes på ett strukturerat sätt (t.ex. av myndigheter) med känd sökinsats. Flera olika nivåer av sökinsats simulerades för båda alternativen.

Resultat Vi bedömer att 1,5-2 identifieringar i genomsnitt per identifierad björnindivid i hela utbredningsområdet för den svenska björnpopulationen bör vara tillräckligt för att få tillförlitliga populationsuppskattningar på både regional och nationell nivå. Detta skulle motsvara en insamling och analys av 5000-6000 dna-prov varje år och antalet tar även hänsyn till prover som inte ger någon genotyp. Analysen visar också att en opportunistisk provtagning med felaktig uppskattning av sökinsatsen kan resultera i en underskattning av populationsstorleken på regional och nationell nivå. Underskattningens omfattning ökar i takt med att insamlingsintensiteten minskar.

Diskussion Vi ser att det är möjligt att få mer exakta och noggranna geografiska uppskattningar av den svenska björnpopulationen med ett rimligt antal prover. Detta skulle kunna uppnås genom att sprida motsvarande antal som för närvarande samlas in över region C (Jämtland och Västernorrland) över hela björnens utbredningsområde i Sverige. Att genomföra en sådan samordnad områdesövergripande provinsamling kräver att man löser logistiska problem, däribland utmaningar som uppstår på grund av den stora opportunistiska delen av björnövervakningen i Sverige. Vår studie är en pilotstudie, med en alltför förenklad modell för provtagningens utformning. Ytterligare analyser kan justera den rumsliga och tidsmässiga utformningen av provinsamlingen för att förbättra uppskattning och kostnadseffektivitet. Trots detta är resultaten lovande och det är värt att fortsätta utreda provinsamling inom hela björnens utbredningsområde. Utan en sådan kommer det troligen även i fortsättningen att vara svårt att göra säkra uppskattningar av den svenska björnpopulationen.

Contents

1	Introduction	6
2	Methods	9
2.1	Bear population size and distribution	10
2.2	NGS simulation	10
2.3	Model fitting	12
2.4	Evaluation of model performance	13
2.5	Deriving the total number of DNA samples required	13
3	Results	14
3.1	Simulated detection data	14
3.2	Abundance estimates	15
3.3	Total number of DNA samples required	18
4	Discussion	19
5	Acknowledgements	20
6	Data availability	21
	References	23

1 Introduction

The Swedish brown bear (*Ursus arctos*) population occupies a large spatial range ($> 500\,000\text{ km}^2$) and is monitored using opportunistic non-invasive genetic sampling (NGS) implemented primarily by hunters (Bischof et al., 2019). A recent report by the Swedish Museum of Natural History estimated the population size to be between 2 587 and 3 080 bears in Sweden in 2022 (Sköld and Åsbrink, 2023). However, the latest analysis by RovQuant (Dupont et al., 2024) revealed that Swedish bear population size estimates can vary substantially between models used to analyze the data. The primary reason for the divergent results seems to be the non-synchronous sampling of 4 regions through a 5-year period (Dupont et al., 2024). This approach leaves large spatial and temporal gaps in the data, that models are unable to bridge without making a series of unrealistic assumptions (Dupont et al., 2024).

Despite the opportunistic nature of the sampling, the overall number of samples collected by hunters is impressive. To encourage future participation in sampling, all putative brown bear samples submitted are analyzed genetically. For example, during the sampling conducted in 2015 in region C (Figure 1), more than 5800 putative bear samples were collected and submitted to the authorities (Kindberg and Swenson, 2017). Of these, 4300 samples were successfully genotyped and revealed the sex and identity of 1015 different bears. This corresponds to >4.2 detections for each detected individual, on average. The drawback of this opportunistic data collection is that effort is neither directed nor documented (Bischof et al., 2019). Lack of direct measures of search effort and spatial variation therein can lead to erroneous population size inferences (Moqanaki et al., 2021).

The Swedish Environmental Protection Agency must produce an estimate of the national number of bears and their distribution once every five years or more frequently (Förordning, 2009:1263). Several methods have been used to obtain range-wide population size estimates of bears in Sweden from non-simultaneous sampling (Bischof et al., 2019; Sköld and Åsbrink, 2023; Kindberg et al., 2009). All of these methods face the challenge of having to predict population size in years and areas without available NGS data. Regardless of the method chosen, statistical and biological assumptions are required to estimate population size with spatio-temporal gaps in sampling. However, violation of those assumptions can lead to unreliable abundance estimates and impact our ability to detect changes in the population (Milleret et al., 2020; Dupont et al., 2024). This makes the estimation of population size in years and areas not sampled challenging, whether it is based on proxies that correlate with population size (e.g., bear observations during moose hunts, Kindberg et al. 2009) or using data that gives information on a part of the population only (e.g., dead recoveries, Bischof et al. 2019; Sköld and Åsbrink 2023).

In this pilot study, we used simulations grounded in empirical data to evaluate and compare different scenarios of sampling configuration and intensity for range-wide simultaneous sampling of brown bears across the species range in Sweden. We considered two main scenarios. In the first scenario, we assumed that NGS was conducted opportunistically by hunters and that a coarse proxy for spatial variation in search effort was available. In the second scenario, we assumed that NGS was conducted during "structured" searches (e.g., such as performed by authorities for wolves and wolverines, Milleret et al. 2022, 2023a) and that an accurate record of search effort was available. We then simulated NGS data using different search effort intensities and evaluated the robustness of the range-wide population size estimates arising from the different sampling scenarios. Although the scenarios explored are simplistic, they constitute a first attempt to inform a potential move towards range-wide bear monitoring in Sweden. Range-wide sampling of bears in Sweden presents its own set of logistic challenges. We nonetheless hope that the present analysis can provide initial indication of the theoretical feasibility from a population estimation perspective and stimulate further exploration of more diverse and realistic scenarios.

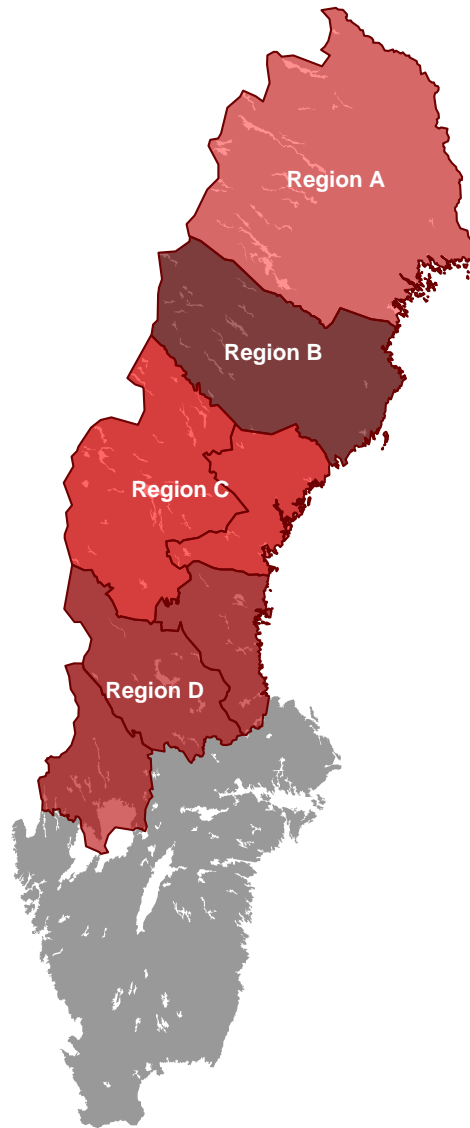


Figure 1: Bear NGS sampling regions in Sweden. Region A is composed of the Norrbotten county last sampled in 2021. Region B is composed of the Västerbotten county last sampled in 2019. Region C is composed of counties Jämtland and Västernorrland, last sampled in 2020, and Region D is composed of counties Värmland, Dalarna and Gävleborg, last sampled in 2022.

Box 1: Terms and acronyms used

AC: Activity center. Model-based equivalent of the center of an individual's home range during the monitoring period. "AC location" refers to the spatial coordinates of an individual AC in a given year and "AC movement" to the movement of an individual AC between consecutive years.

CMR: Capture mark-recapture.

CrI: 95% credible interval associated with a posterior sample distribution.

Detectors: Potential detection locations in the spatial capture-recapture framework. These can refer to fixed locations (e.g., camera-trap locations) or in this report to areas searched (e.g., habitat grid cells in counties where searches for genetic samples were conducted).

Habitat buffer: Buffer surrounding the searched area that is considered potentially suitable habitat but was not searched.

Länsstyrelserna: Swedish County Administrative Boards, in charge of the monitoring of large carnivores at the county level.

Legal culling: Lethal removal of individuals by legal means, including licensed recreational hunting, management removals, and defense of life and property.

MCMC: Markov chain Monte Carlo.

NGS: Non-invasive genetic sampling.

OPSCR: Open-population spatial. capture-recapture

p_0 : Baseline detection probability; probability of detecting an individual at a given detector, if the individual's AC is located exactly at the detector location.

σ : Scale parameter of the detection function; related to the size of the circular home-range.

SCR: Spatial capture-recapture.

SNO: Statens naturoppsyn (Norwegian Nature Inspectorate) is the operative field branch of the Norwegian Environment Directorate (Miljødirektoratet).

Statsforvalteren: Norwegian state's representative in the county, responsible for following up decisions, goals, and guidelines from the legislature and the government.

RovQuant: Research project at the Norwegian University of Life Sciences (NMBU, Ås, Norway) that develops and applies SCR models.

2 Methods

The methodological approach taken in this study consisted in 1) simulating the abundance and distribution of the Swedish bear population using the most recent estimation results (Dupont et al., 2024); 2) simulating opportunistic and systematic NGS of varying intensity across Sweden; 3) fitting single-season spatial capture recapture (SCR) models to the simulated datasets to estimate abundance; 4) evaluate the robustness of inferences.

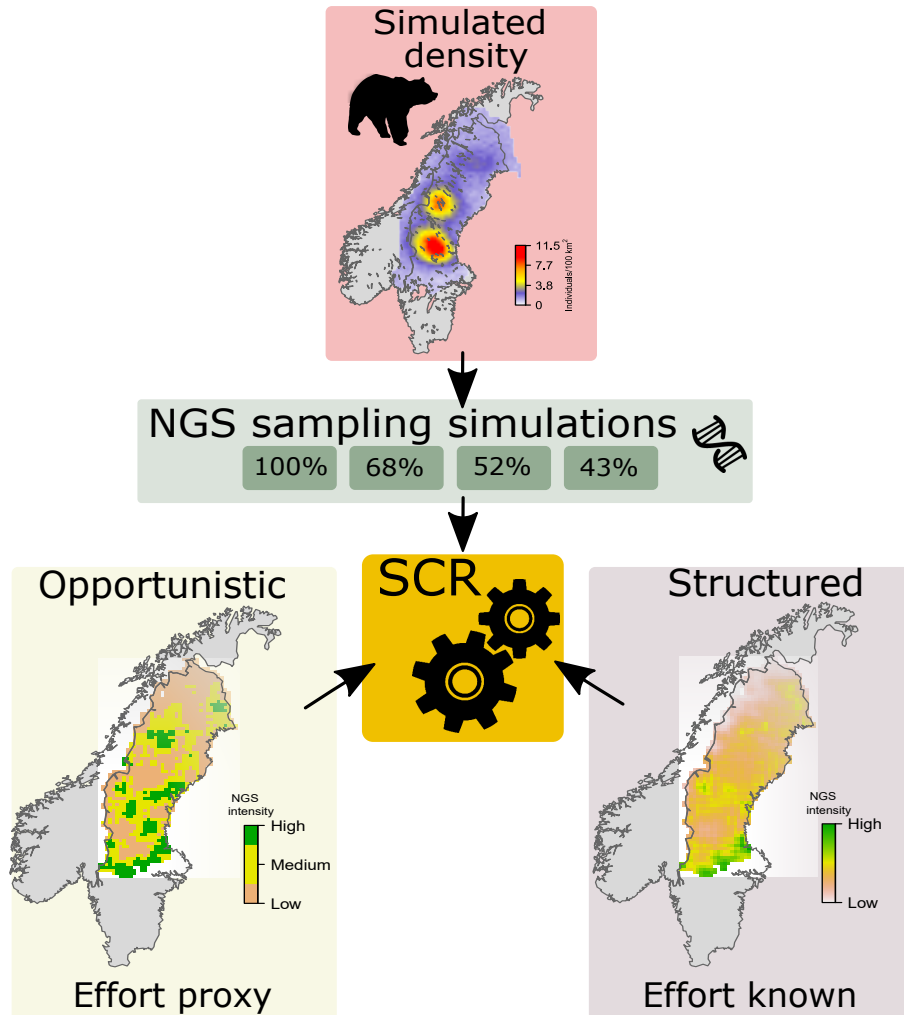


Figure 2: Diagram summarising the main steps of the study: 1) Spatially-explicit simulation of the bear population. 2) Simulation of genetic sampling with different levels of intensity. 3) Model fitting (estimation) with either a known effort (structured sampling) or a proxy for effort (opportunistic sampling). 4) Evaluation and comparison of the resulting estimates.

SCR models require multiple detections of unique individuals in different locations. We chose to perform the study using the female part of the population because females have smaller home ranges than males. If the NGS scenario chosen allows the multiple detections of females, multiple spatial detections of males are guaranteed. In addition, the current NGS conducted in Sweden revealed a higher detection probability for males than for females, suggesting that if the minimum number of samples to fit SCR models is reached for females, the number of samples will also be sufficient to fit SCR models for males.

With this simulation study, our goal was not to capture the full complexity of the biology and monitoring of bears in Sweden, but to provide a first assessment of the minimum number of genetic samples necessary to obtain reliable estimates of range-wide abundance of bears in

Sweden using SCR models. We used the simulation capabilities of R package NIMBLE (de Valpine et al., 2022) and nimbleSCR (Bischof et al., 2021) to perform the simulations.

2.1 Bear population size and distribution

In SCR models, population density is represented by the number and distribution of individual activity centers (ACs, s_i). Here, we used the spatial distribution of bear ACs in Sweden as predicted by a sex-specific Bayesian open population spatial capture recapture (OPSCR) model. Full details of the model can be found in Dupont et al. (2024). The point process formulation of the OPSCR model estimates the association between spatial covariates and density (Zhang et al., 2022). Dupont et al. (2024) found that female bear density was associated positively with the spatial distribution of dead recoveries (95%CI: $\beta_{dead} = 0.44;0.65$) and negatively ($\beta_{obs} = -0.18;-0.11$) with the distribution of bear observations (available in the roibase and skandobs datasets). Using these covariates and the associated coefficients estimated by the OPSCR model (Figure 3), we simulated 25 bear populations, represented by the number and spatial distribution of individual activity centers. Each simulation was obtained using a randomly chosen iteration in the posterior distribution for the total abundance (N), β_{dead} , and β_{obs} . This allowed us to propagate uncertainty in the number and distribution of bears.

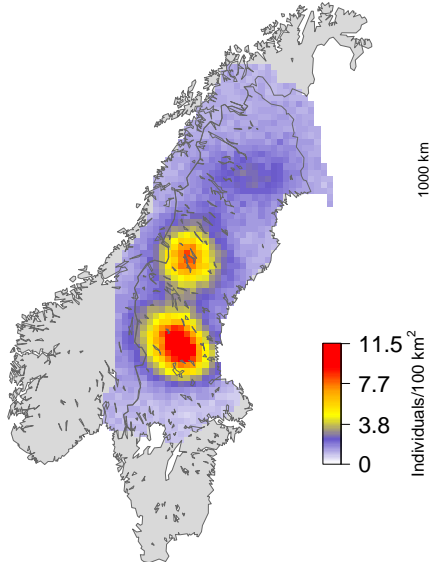


Figure 3: Average predicted female bear density in Sweden over 25 realized values taken randomly from the posterior distributions of an OPSCR model (Dupont et al., 2024)

2.2 NGS simulation

In order to simulate NGS data for each of the simulated bear populations, we assumed that the sampling process followed the detection process of a SCR model. This means that 1) the simulated number of detections only reflects bear samples successfully genotyped, 2) detections occur at discrete spatial locations (i.e., detectors), and that 3) only one detection per individual is retained within a 1×1 km grid cell (Milleret et al., 2018).

SCR models account for individual heterogeneity in detectability associated with the location of the individual and its home range size. Detection probability p of a given individual i at a given detector j can be defined as a decreasing function of the distance from the individual activity center d_{ij} :

$$p_{ij} = p0_j \exp(-d_{ij}^2/2\sigma^2)$$

where σ is the spatial scale parameter of the detection function and controls for the size of individual home range, and $p0$ describes the baseline detection probability, i.e. the detection probability at the activity center location (s_i). Here, we considered that $p0$ at a given detector j was a function of the county c in which it was located (Figure 1), distance from roads (*Road*), and a covariate representing spatial heterogeneity in search effort (*Effort*):

$$\text{logit}(p0_j) = pZero_c + \beta_{road} * Road_j + \beta_{effort} * Effort_j$$

Because of the opportunistic nature of NGS in Sweden, there is no direct measure of spatial variation in effort. In an attempt to obtain a representative picture of the spatial heterogeneity in detection probability, we created a kernel distribution of all carnivore observations reported in skandobs and from all genotyped carnivore samples designated to be bears that were available in rovbase. We then weighted the kernel distribution by the density of bears in Sweden to account for a potential correlation between bear density and number of samples collected. Values of this *Effort* covariate were scaled to obtain a map of relative search effort (Figure 2). We also modelled the effect of the distance to roads (β_{road}) as this variable can explain spatial heterogeneity in sampling intensity (Dupont et al., 2024).

Because we used the partially aggregated binomial model (Milleret et al., 2018), we assumed that detections of an individual at a given detector (Y_{ij}) followed a Binomial distribution such that:

$$Y_{ij} \sim \text{Binomial}(p_{ij}, K_j)$$

where K represents the number of subdetector cells (maximum 25) within a 5×5 km main detector grid cell. With this formulation, the number of detections at a given main detector is equal to the number of subdetector cells with at least one detection (Milleret et al., 2018).

NGS scenarios

For each of the 25 bear population realizations (2.1), we simulated four NGS datasets, each corresponding to a given NGS intensity. The first intensity (100%) represented 'business as usual'. In this scenario, we considered baseline detection probabilities ($pZero_c$) for the 6 counties (Dalarna, Gävleborg, Jämtland, Norrbotten, Västerbotten, Västernorrland) that led to an average number of detections per detected individual comparable to what was obtained in each county during the most recent NGS sampling (Table 1). We chose the average number of detections per detected individual as the main metric to calibrate our simulations against the bear detection data available in Rovbase because it is not influenced by the number of individuals present in the population, contrary to the total number of detections.

We then divided the chosen $pZero_c$ values in each county by 2, 4, and 8 to simulate an overall reduction in the intensity of the NGS. Taking 'business as usual' as a reference (100%), this corresponded to a 68%, 52% and 43% NGS intensity in terms of the number of detections per detected individuals, respectively (Table 1). In all scenarios, we simulated a positive effect of the *Effort* covariate using $\beta_{effort} = 1$. Finally, as for simulating bear distribution (2.1), we modelled σ and the negative effect of distance to roads β_{road} on $p0$, based on 25 randomly

selected posterior values obtained from the OPSCR model (Dupont et al., 2024).

For each simulated dataset of the four NGS intensity scenarios, we fitted two SCR models. In the first SCR model ("structured" sampling), we assumed that we had access to the true measure of spatial variation in search effort. This would be the case during structured sampling, where searchers record a GPS track during searches. In the second SCR model ("opportunistic" sampling), we only used a proxy of the spatial variation in search effort. This reflects the lack of a direct measure of effort during opportunistic sampling (Figure 2). Thus, the only differences between those two SCR models lied in the spatial covariate used for search effort (see Figure 2).

Table 1: Simulated sampling intensities expressed as the average number of brown bear NGS samples collected per individual detected. Four different sampling intensities were simulated (100%, 68%, 52%, 43%). For reference, the actual sampling intensity during the most recent monitoring bout in each county is shown in the column "True". Sampling intensities are shown by county and for the entire Swedish bear range. Simulated sampling intensities are accompanied by their respective 95% quantiles within brackets.

	True	100%	68%	52%	43%
Dalarna	2.9	2.9 (2.5-3.5)	2.1 (1.8-2.5)	1.6 (1.4-1.9)	1.3 (1.1-1.5)
Gävleborg	3.5	3.4 (2.7-4.4)	2.5 (2.0-2.9)	1.9 (1.5-2.4)	1.5 (1.2-1.9)
Jämtland	2.8	2.7 (2.5-2.9)	1.8 (1.6-2.0)	1.4 (1.3-1.5)	1.2 (1.1-1.3)
Norrbottn	3.0	3.0 (2.7-3.4)	1.9 (1.8-2.2)	1.5 (1.3-1.7)	1.2 (1.1-1.3)
Västerbotten	2.2	2.2 (1.9-2.4)	1.5 (1.4-1.7)	1.3 (1.1-1.4)	1.2 (1.1-1.3)
Västernorrland	3.9	3.9 (3.4-4.4)	2.3 (2.0-2.7)	1.7 (1.5-1.9)	1.3 (1.0-1.7)
Sweden	3.0	2.9 (2.7-3.0)	1.9 (1.8-2.1)	1.5 (1.4-1.6)	1.3 (1.2-1.3)

2.3 Model fitting

We used a single-season (i.e. demographically closed) SCR model for data analysis that contained the same features as the model used in the simulation.

Observation submodel

The observation submodel used for the *structured* sampling scenario assumed that:

$$\text{logit}(p0_j) = pZero_c + \beta_{road} * Road_j + \beta_{effort} * Effort_j$$

where *Effort* is the true covariate used for simulation.

For the *opportunistic* sampling scenario, we assumed that *p0* was a function of:

$$\text{logit}(p0_j) = pZero_c + \beta_{road} * Road_j + \beta_{effort} * EffortProxy_j$$

Where *EffortProxy* is a proxy of true search effort with three categories (low, medium, and high). This categorical proxy was generated using the following process. For each simulated dataset, we introduced random errors in the *Effort* covariate used to model spatial variation in detectability. The error at each detector *j* was randomly drawn from a normal distribution centered on 0 and with a standard deviation of 0.15. The resulting search effort proxy was then split into three categories, representing areas with low, medium, and high detection probability (Figure 2). Note that the amount of error and the categorisation were chosen arbitrarily.

Demographic submodel

We used a data-augmentation approach (Royle and Young, 2008) to estimate population size *N*. This means that the detection of an individual is conditional on the individual's state *z_i* (*z_i* = 1 when individual *i* is a member of the population), which is governed by the inclusion probability $z_i \sim \text{Bernoulli}(\psi)$. Population size can be then derived by: $N = \sum_{i=1}^M z_i$ where *M* is the number of individuals considered in the augmented pool ($M \gg N$).

Density submodel

We used a Bernoulli point process to model the distribution of individual activity centers (ACs, Zhang et al. 2022). The intensity of the point process was a function of both the locations of all bears recovered dead (*dead*) throughout the 2012-2021 period, and the presence/absence of all bear observations registered in SkandObs during the 2012-2021 period (*Obs*; as used in the simulation; see Dupont et al. 2024 for more details). The intensity of the Bernoulli point process was modelled as :

$$\lambda(s) = e^{\beta_{dead} * dead + \beta_{Obs} * Obs}$$

Implementation

The different scenarios led to 100 different simulated data sets (25 simulated populations with 4 sampling scenarios each). Two different models were fitted to each simulated data set: a model assuming opportunistic (using *EffortProxy*) and structured (using *Effort*) sampling. All models were fitted with NIMBLE (de Valpine et al., 2022; de Valpine et al., 2017) and nimbleSCR (Bischof et al., 2021) in R (R Core Team, 2021). We ran three MCMC chains, each with 7500 iterations, discarded the initial 2000 samples as burn-in. We assessed mixing of chains by inspecting traceplots, and we considered models as converged when the Rhat was ≤ 1.1 for all parameters (Gelman and Rubin, 1992).

2.4 Evaluation of model performance

We assessed the performance of SCR models in the different NGS scenarios by calculating the accuracy and precision of the Bayesian estimators of total (N) and region-specific population size. To obtain region-specific population size, we summed the number of predicted AC locations (individuals detected during sampling or predicted to be alive by the model) that fell within that region for each iteration of the MCMC chains. This produced a posterior distribution of abundance for that region. From such posteriors, abundance estimates and the associated uncertainty can be extracted for any spatial unit, including countries, counties or management regions. As a measure of accuracy, we used relative error $RE = \frac{(\hat{\theta} - \theta)}{\theta}$, where $\hat{\theta}$ is the posterior mean and θ is the true value of the parameter. As a measure of precision, we used the coefficient of variation $CV = \frac{SD(\hat{\theta})}{\hat{\theta}}$. We also calculated the coverage of the 95% credible interval (i.e., the probability that it contains the true value of the parameter) as the percentage of simulation replicates where the credible interval contained the true value.

2.5 Deriving the total number of DNA samples required

The simulation framework in this study does not produce a direct estimate of the total number of DNA samples that needs to be collected in the field for a given scenario. Instead, we expressed sampling intensity as the number of detections per detected individual. However, we can derive a rough estimate of the number of samples needed to be collected based on available information about samples collected and the number of detections actually used in the analysis.

We estimated the total number of DNA samples collected (S) associated with a given scenario using the following equation:

$$S = \frac{N * PD * DI}{GS * AL}$$

Where N is the total population size, PD the proportion of the population detected, DI the number of detections per detected individual, GS is the genotyping success rate, and AL is the proportion of samples retained after spatial aggregation. The numerator thus gives the

total number of detections provided to the SCR model in a given scenario, and the denominator accounts for the proportion of samples lost due to genotyping failures and to spatial aggregation. We can parameterize the equation above based on available information. The genotyping success (GS) of bear samples in Sweden averaged 68% in recent years (Åsbrink et al., 2020, 2021, 2022). Spatial aggregation of detections to a 1×1 km grid cell in the SCR model used for simulation resulted in retaining an average of 93% (AL) of the detections (see also Dupont et al. (2024)). Although the proportion of detected individuals (PD) depends on several factors, such as the spatial variation in effort, we here assume that range-wide sampling detects approximately 50-70% of the total population N . We assume that there are currently about $N = 3000$ bears in Sweden (Sköld and Åsbrink, 2023). We also assume that the same conversion between SCR detections and number of samples applied to both males and females. We can thus calculate S for the different levels of DI used in the simulations (1.3, 1.5, 1.9, and 2.9, Table 1) and PD (0.5, 0.6, and 0.7):

$$S = \frac{3000 * PD * DI}{0.68 * 0.93}$$

3 Results

3.1 Simulated detection data

The approach used here allowed us to simulate an average number of detections per individual detected comparable to what was obtained during the most recent NGS bout in each county (Table 1). We note that the values reported in the tables (Table 1, Table 2, Table 3) correspond to the number of successfully genotyped detections, after retaining only one detection per individual and per 1×1 km detector grid cell (Milleret et al., 2018). In all simulated datasets, the total number of detections and detected individuals were lower than the ones obtained during the last NGS sampling (Table 2, Table 3).

Table 2: Number of female bears detected by county and overall in Sweden. Four different sampling intensities were simulated (100%, 68%, 52%, 43%). For reference, the actual sampling intensity during the most recent monitoring bout in each county is shown in the column "True". Sampling intensities are shown by county and for the entire Swedish bear range. Simulated sampling intensities are accompanied by their respective 95% quantiles within brackets.

	True	100%	68%	52%	43%
Dalarna	164	156 (131-185)	111 (94-136)	73 (56-91)	44 (33-61)
Gävleborg	219	113 (96-126)	81 (68-98)	55 (42-67)	34 (25-48)
Jämtland	530	313 (265-355)	239 (204-277)	153 (129-180)	92 (72-116)
Norrbottn	242	190 (162-218)	145 (121-165)	99 (79-115)	58 (41-72)
Västerbotten	186	151 (123-176)	106 (83-126)	65 (51-83)	36 (27-49)
Västernorrland	227	79 (63-100)	67 (52-85)	48 (38-63)	30 (20-42)
Sweden	1568	1002 (908-1102)	750 (680-826)	492 (426-548)	295 (256-349)

Table 3: Number of detections of female bears by county and overall in Sweden. Four different sampling intensities were simulated (100%, 68%, 52%, 43%). For reference, the actual sampling intensity during the most recent monitoring bout in each county is shown in the column "True". Sampling intensities are shown by county and for the entire Swedish bear range. Simulated sampling intensities are accompanied by their respective 95% quantiles within brackets.

	True	100%	68%	52%	43%
Dalarna	473	458 (352-580)	232 (174-281)	118 (89-163)	58 (41-87)
Gävleborg	769	386 (295-519)	199 (156-258)	102 (71-136)	51 (37-74)
Jämtland	1476	846 (705-987)	430 (360-515)	212 (181-241)	109 (87-138)
Norrbottn	727	570 (476-674)	283 (230-331)	144 (115-165)	71 (51-87)
Västerbotten	410	326 (264-387)	165 (122-208)	81 (63-104)	42 (30-56)
Västernorrland	890	310 (250-394)	155 (119-202)	79 (59-105)	38 (27-51)
Sweden	4745	2896 (2550-3272)	1463 (1319-1639)	736 (655-819)	369 (312-442)

3.2 Abundance estimates

Range-wide estimates Simulations showed that opportunistic search effort led to a high risk of underestimating abundance and low coverage of the estimates (<20%, Figure 4). This pattern occurred at all NGS intensities, but the bias increased with decreasing NGS intensity (Figure 4). Structured NGS, in the other hand, led to robust inferences for most NGS intensities. As expected, the coefficient of variation increased with decreasing NGS intensity (Figure 4).

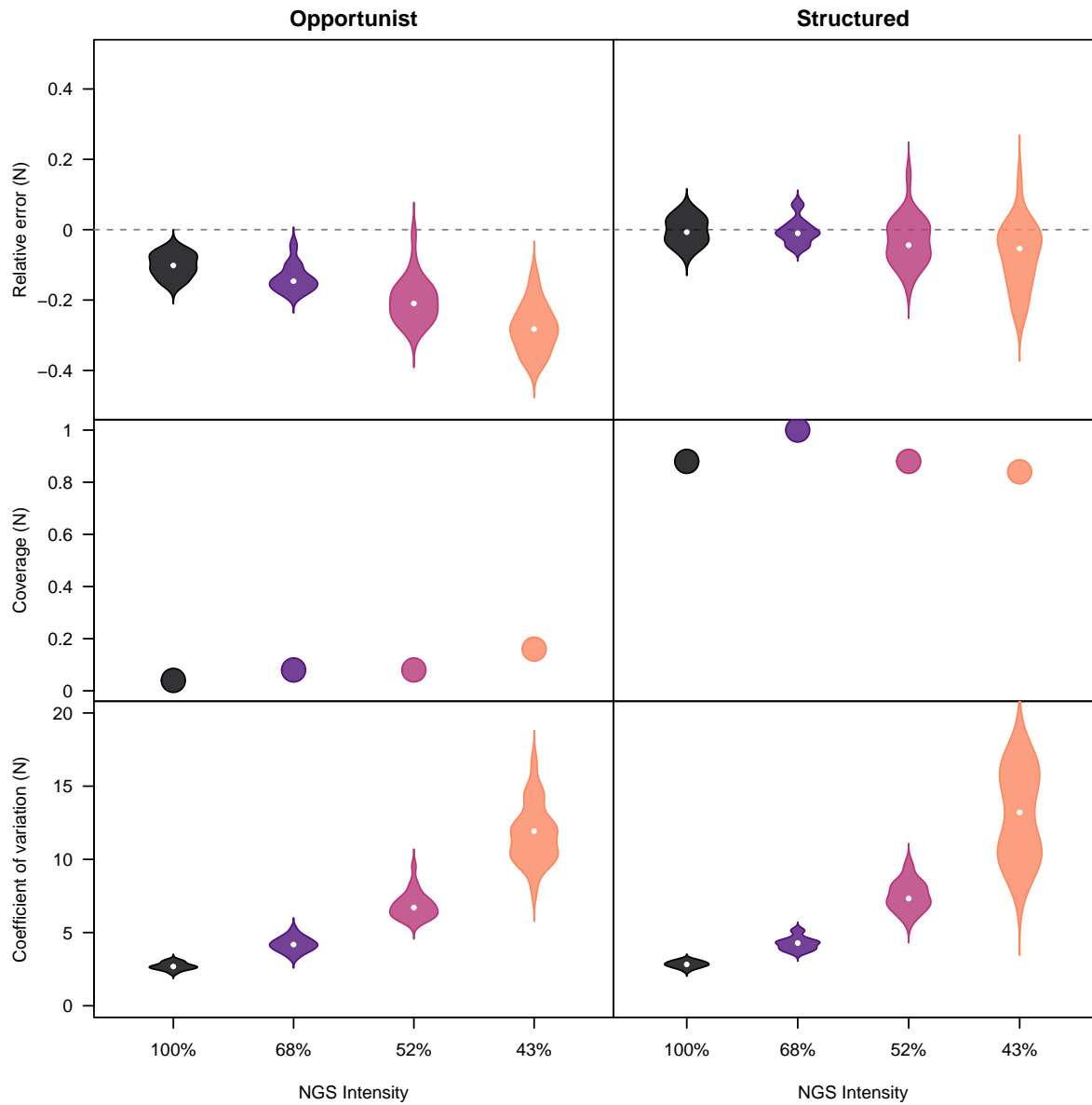


Figure 4: Relative error (top panel), coverage (middle panel), and coefficient of variation (bottom panel) of abundance estimates (N) for the 25 replicated data sets of each scenario. Left and right panels show results for the opportunistic scenario (i.e., proxy in search effort) and structured effort (i.e., known effort), respectively. Results are presented for the four non-invasive genetic sampling (NGS) intensity scenarios.

Regional estimates A comparable pattern was observed for regional abundance estimates. Abundance estimates were negatively biased in the opportunistic NGS scenario, and this bias tended to increase with decreasing sampling intensity (Figure 5). Structured NGS led to robust inferences at the regional level at most NGS intensities (Figure 5). In both the opportunistic and structured scenario, coefficients of variation of regional abundance estimates increased with decreasing NGS intensity (Figure 6).

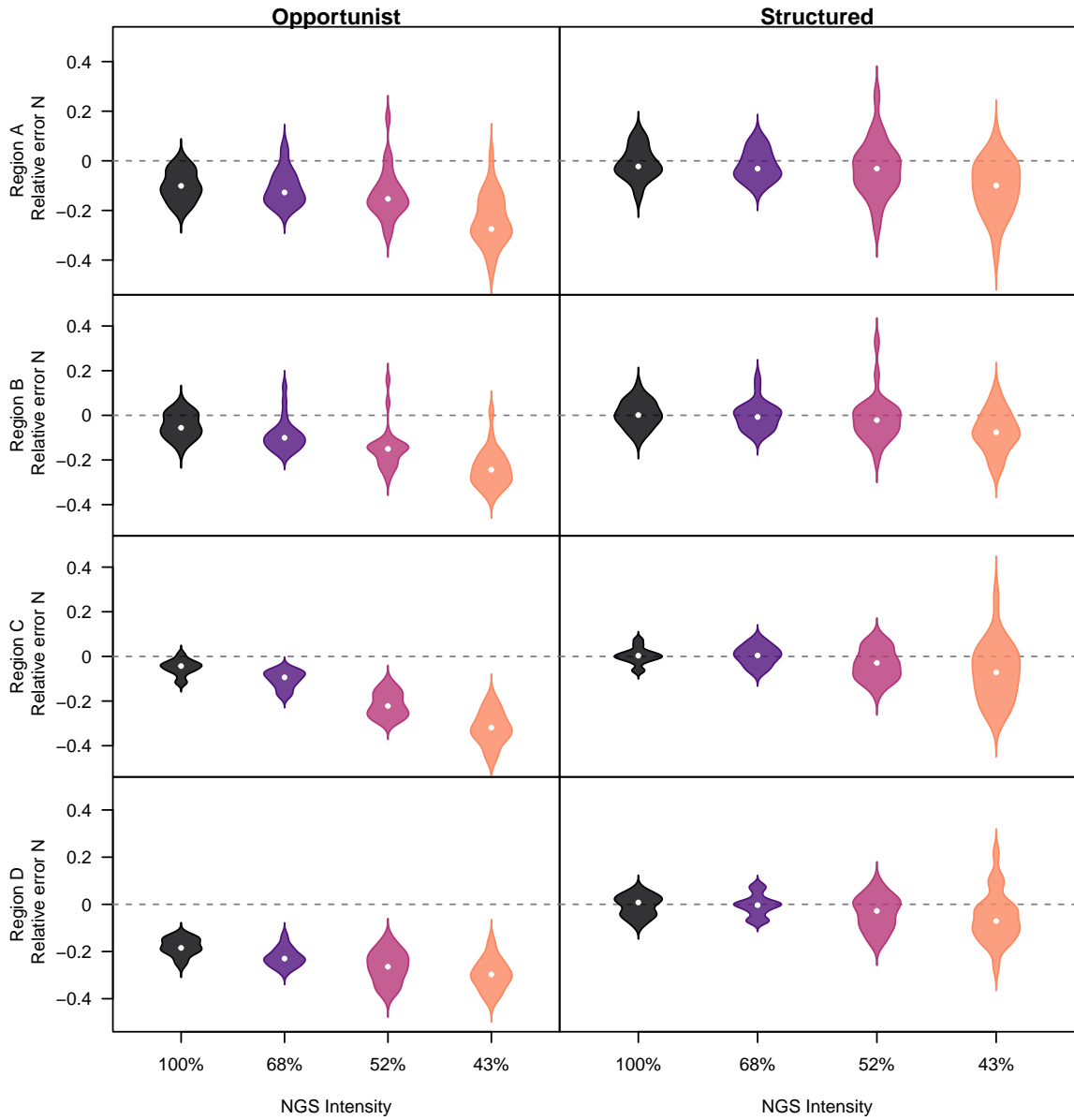


Figure 5: Violin plots of the relative error of the region-specific abundance estimates (N) for the 25 replicated data sets of each scenario. Left and right panels show results for the opportunistic scenario (i.e., proxy of search effort) and structured effort (i.e., known effort), respectively. Results are presented for the four non-invasive genetic sampling (NGS) intensity scenarios.

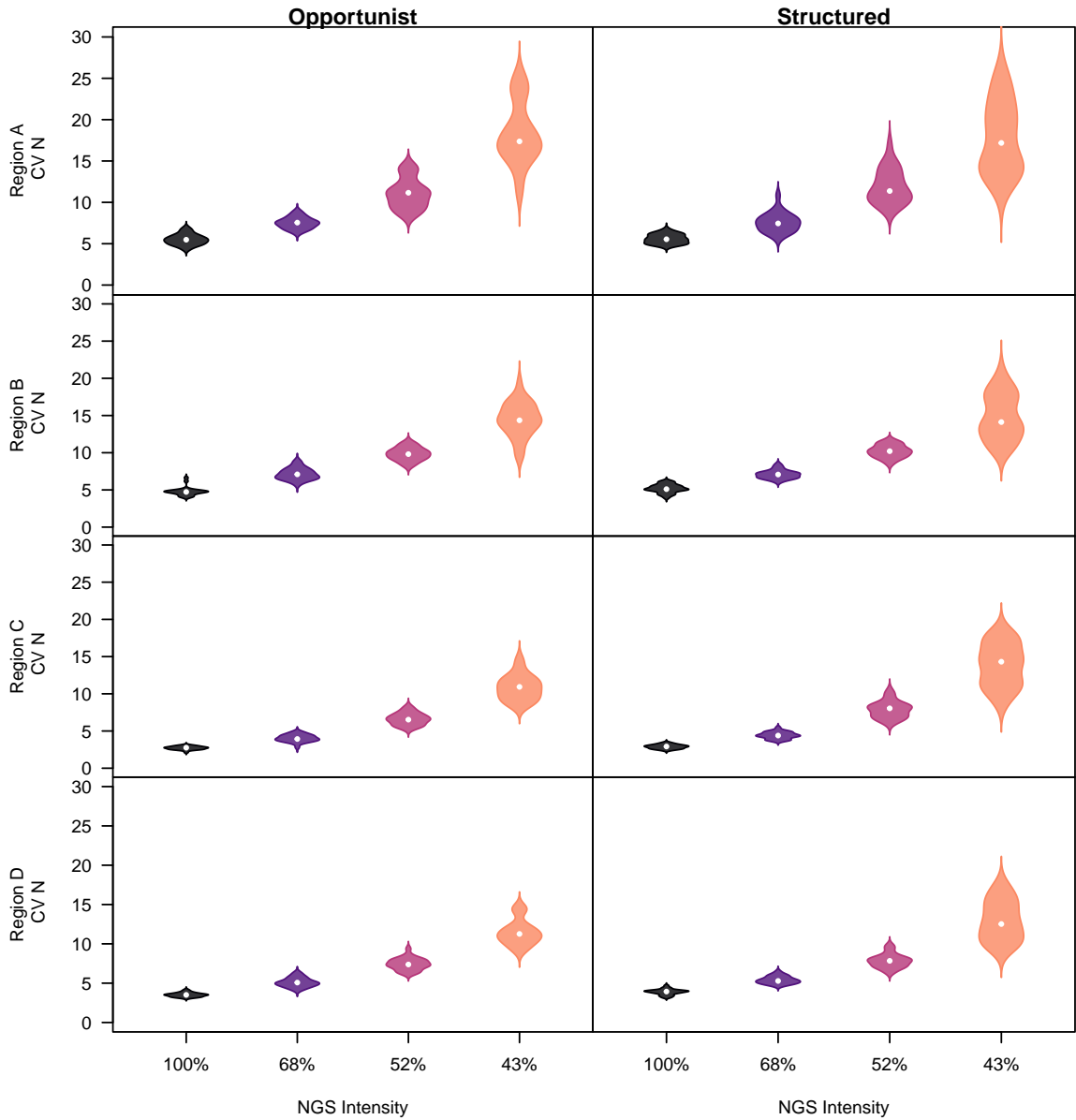


Figure 6: Violin plots of the coefficients of variation of the region-specific abundance estimates (N) for the 25 replicated data sets of each scenario. Left and right panels show results for the opportunistic scenario (i.e., proxy in search effort) and structured effort (i.e., known effort), respectively. Results are presented for the four non-invasive genetic sampling (NGS) intensity scenarios.

3.3 Total number of DNA samples required

The required number of genetic samples to be collected across the Swedish bear range varied between approximately 3000 and 10000, depending on the simulation scenario (Table 4). The required number of samples was substantially influenced by the expected proportion of individuals detected and the expected number of detections per detected individual. The latter is a key parameter determining the precision of the abundance estimates.

Table 4: Estimation of the total number of samples that should be collected in the field given that not all samples can be successfully genotyped (68% genotyping success), that approximately 93% of the detections are retained after aggregation in SCR models, and assuming that there are 3000 bears in Sweden at the time of sampling. Number of samples are estimated for all simulation scenarios tested in the report (100%, 68%, 52%, 43%), and for three levels of detectability (proportion of individuals detected). The calculation is described in Section 2.5.

Average detections per individual detected	Proportion of individuals detected		
	50%	60%	70%
1.3	3083	3700	4317
1.5	3558	4269	4981
1.9	4507	5408	6309
2.9	6879	8254	9630

4 Discussion

Our analysis revealed that the current NGS intensity, if applied to the entire Swedish brown bear range simultaneously, is unnecessarily high for obtaining robust population size estimates. On average, 1.5-2 spatial detections per individual detected should be sufficient to obtain robust range-wide and regional population size estimates. This would require 5000-6000 samples to be collected in the field throughout the entire Swedish bear population range. For reference, this sample size spread over the entire bear Swedish bear range (472000 km^2) is comparable to the number of samples collected within region C (75000 km^2) in 2020 (>5500 , Åsbrink et al. 2020, 2021, 2022). In addition, our study revealed that the lack of an accurate measure of search effort can lead to an underestimation of population size at the regional and national level. This finding confirms our previous work (Moqanaki et al., 2021). Structured sampling, with accurate documentation of search effort would allow more efficient and reliable monitoring of the bear population in Sweden.

The current NGS intensity results in an impressive average number of spatial detections per individual detected in each of the counties (between 2.2 and 3.9, Table 1). SCR models do not require such a high number of samples to obtain reliable population size estimates. As demonstrated here, reliable population size estimates can be obtained with 1.5 to 2.0 spatial detections per individual detected on average. These levels are in accordance with the general recommendations for SCR models (Royle et al., 2014).

Simulating detection data directly within the SCR model is a simplification of reality and did not allow us to capture all processes involved in data collection. To obtain an estimation of the total number of samples needed to be collected in the field, we had to rely on several assumptions. For example, we assumed that 68% of the collected samples can be successfully genotyped (Åsbrink et al., 2020, 2021, 2022, 2023) and that there was no temporal, spatial, and individual variation in genotyping success. We also assumed that the aggregation of detections to detectors caused an additional 7% loss in the number of samples. However, this rate was based on empirical data and may vary across sex, space, and depending on the sampling strategy. If all our assumptions hold, a total of 5000-6000 samples collected through the entire bear range in Sweden should allow reliable abundance estimation.

Our report confirms the importance of recording and accounting for spatial variation in effort to obtain unbiased estimates of population size, as previously shown in another simulation study (Moqanaki et al., 2021). The lack of search effort records generally associated with opportunistic data collection is therefore a strong drawback of such a sampling strategy. Our results showed that the use of a proxy to quantify spatial variation in effort can also lead to bias. Although

methods exist that can account for unmeasured variability in detection probability (including imperfectly known search effort, Dey et al. 2022), these approaches are currently computationally prohibitive at the scale of the Swedish bear population. The scope of our results rely on two main assumptions regarding spatial variation in effort: 1) The amount of error and loss of details used to create the proxy in search effort were chosen arbitrarily (Figure 2), and may not represent what could be obtained from an opportunistic data collection scheme; 2) We also do not know how well the observation records available in Skandobs and Rovbase reflect the spatial variation in effort. The lower number of detected individuals and detections simulated compared to what was collected suggest that the detection processes occurring during the opportunistic sampling were not entirely captured.

The distribution of search effort across Sweden is key from the point of view of sampling efficiency (Sun et al., 2014; Dupont et al., 2021a). However, opportunistic sampling provides no, or only limited, control over the spatial distribution of effort, a further hurdle to design and implement cost efficient sampling strategies. In this report, we did not aim at optimizing the spatial distribution of effort, but a stratified NGS sampling would be a good candidate to maximize the efficiency of the sampling and improve the performance of subsequent estimation. Stratification would allow the prioritization of certain areas, for example with known high bear density, while expending lower sampling intensities (and thus resources) in areas with lower bear density (Sun et al., 2014; Dupont et al., 2021a). In this regard, the combination of both opportunistic and structured sampling could be an interesting trade-off. It would combine the benefit of opportunistic sampling, with the possibility to direct a certain amount of systematic search effort to designated areas. Note that opportunistic and structured sampling are already integrated in the wolverine and wolf abundance estimations (Milleret et al., 2023b,a).

In this study, we focused on a single-season SCR model because it is known to be a robust method for obtaining abundance estimates (Dupont et al., 2019; Bischof et al., 2020; Theng et al., 2022). Open population capture-recapture (OPSCR) models could make use of more of the information contained in individual detections over several years (Milleret et al., 2020). These models also allow integrating other types of data such as dead recoveries (Dupont et al., 2021b), and may give access to a broader range of sampling strategies such as small spatio-temporal gaps in sampling (Milleret et al., 2020).

Sampling strategies tested in this report assumed that the logistical and financial challenges of conducting synchronous NGS in all Swedish regions can be overcome. For example, if NGS is conducted by hunters simultaneously in all regions, logistic and financial constraints may prevent genotyping all collected samples. Under this scenario, subsampling of collected samples will therefore be necessary, which would rely on hunters not losing motivation for sample collection if not all samples they collected are genetically analyzed. Although range-wide sampling of bears in Sweden presents its own set of logistic challenges, the present analysis provides an initial indication about its theoretical feasibility from a population estimation perspective. Without some form of range-wide sampling associated with a record of spatial effort in sampling, reliable population size estimates of the Swedish bear population will likely remain elusive.

5 Acknowledgements

This work was made possible by the large carnivore monitoring programs and the extensive monitoring data collected by Swedish and Norwegian wildlife management authorities, as well as the public. We also thank Swedish and Norwegian wildlife managers for feedback provided during project RovQuant and the Research Council of Sweden for partial funding (NFR 286886; project WildMap). The computations/simulations were performed on resources provided by

NMBU's computing cluster "Orion", administered by the Centre for Integrative Genetics and by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Sweden. We are grateful to the NIMBLE team (P. de Valpine and D. Turek) for help with the formulation of the OPSCR model. J. Vermaat provided helpful comments on drafts of this report.

6 Data availability

Data and R code to reproduce the analysis are available on GitHub <https://github.com/richbi/RovQuantPublic>.

References

- Åsbrink, J., Martin, S., Thomas, K., and Niclas, G. (2023). Resultat från inventering av brunbjörn i dalarnas, gävleborgs och värmlands län 2022.
- Åsbrink, J., Sköld, M., Källman, T., and Gyllenstrand, N. (2020). Resultat från inventeringen av brunbjörn i västerbottens län 2019.
- Åsbrink, J., Sköld, M., Källman, T., and Gyllenstrand, N. (2022). Resultat från inventeringen av brunbjörn i norrbottens län 2021.
- Åsbrink, J., Sköld, M., Thomas, K., and Niclas, G. (2021). Resultat från inventering av brunbjörn i jämtlands och västernorrlands län 2020.
- Bischof, R., Dupont, P., Milleret, C., Chipperfield, J., and Royle, J. A. (2020). Consequences of ignoring group association in spatial capture–recapture analysis. *Wildlife Biology*, 2020(1):1–10.
- Bischof, R., Milleret, C., Dupont, P., Chipperfield, J., Brøseth, H., and Kindberg, J. (2019). Estimating density, abundance and population dynamics of bears, wolverines, and wolves in Scandinavia. *MINA fagrapport 63*, 79pp.
- Bischof, R., Turek, D., Milleret, C., Ergon, T., Dupont, P., Dey, S., and de Valpine, P. (2021). *nimbleSCR: Spatial Capture-Recapture (SCR) Methods Using 'nimble'*. R package version 0.1.2.
- de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodriguez, A., Temple Lang, D., and Paganin, S. (2022). *NIMBLE User Manual*. R package manual version 0.12.2.
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2):403–413.
- Dey, S., Bischof, R., Dupont, P. P. A., and Milleret, C. (2022). Does the punishment fit the crime? consequences and diagnosis of misspecified detection functions in Bayesian spatial capture–recapture modeling. *Ecology and Evolution*, 12(2):e8600.
- Dupont, G., Royle, J. A., Nawaz, M. A., and Sutherland, C. (2021a). Optimal sampling design for spatial capture–recapture. *Ecology*, 102(3):e03262.
- Dupont, P., Milleret, C., Brøseth, H., Kindberg, J., and Bischof, R. (2024). Challenges in estimating range-wide brown bear density and abundance in sweden. *MINA fagrapport 95*, 47pp.
- Dupont, P., Milleret, C., Gimenez, O., and Bischof, R. (2019). Population closure and the bias-precision trade-off in spatial capture–recapture. *Methods in Ecology and Evolution*, 10(5):661–672.
- Dupont, P., Milleret, C., Tourani, M., Brøseth, H., and Bischof, R. (2021b). Integrating dead recoveries in open-population spatial capture–recapture models. *Ecosphere*, 12(7):e03571.
- Förordning (2009). Förordning (2009:1263) om förvaltning av björn, varg, jårv, lo och kungsörn, landsbygds- och infrastrukturdepartementet (rsl), svensk författningssamling (sfs).
- Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511. <http://www.stat.columbia.edu/~gelman/research/published/itsim.pdf>.
- Kindberg, J., Ericsson, G., and Swenson, J. E. (2009). Monitoring rare or elusive large mammals using effort-corrected voluntary observers. *Biological Conservation*, 142(1):159–165.
- Kindberg, J. and Swenson, J. E. (2017). Björnstammens storlek i jämtlans och västernorrlands län 2015. *Scandinavian Brown Bear Research Project, report*, 2.
- Milleret, C., Dupont, P., Brøseth, H., Flagstad, O., Kindberg, J., Svensson, L., and Bischof, R. (2023a). Estimates of wolf density, abundance, and population dynamics in scandinavia, 2013–2023. *MINA fagrapport 85*, 34pp.
- Milleret, C., Dupont, P., Brøseth, H., Flagstad, O., Kleven, O., Kindberg, J., and Bischof, R. (2023b). Estimates of wolverine density, abundance, and population dynamics in scandinavia, 2014–2023. *MINA fagrapport 89*, 33pp.
- Milleret, C., Dupont, P., Brøseth, H., Kindberg, J., Royle, J. A., and Bischof, R. (2018). Using partial aggregation in spatial capture recapture. *Methods in Ecology and Evolution*, 9(8):1896–1907.
- Milleret, C., Dupont, P., Chipperfield, J., Turek, D., Brøseth, H., Gimenez, O., de Valpine, P., and Bischof, R. (2020). Estimating abundance with interruptions in data collection using open population spatial capture–recapture models. *Ecosphere*, 11(7):e03172.
- Milleret, C., Dupont, P., Moqanaki, E., Brøseth, H., Flagstad, O., Kleven, O., Kindberg, J., and Bischof, R. (2022). Estimates of wolverine density, abundance, and population dynamics in Scandinavia, 2014–2022. *MINA fagrapport 79*, 35pp.
- Moqanaki, E. M., Milleret, C., Tourani, M., Dupont, P., and Bischof, R. (2021). Consequences of ignoring variable and spatially autocorrelated detection probability in spatial capture–recapture. *Landscape Ecology*, 36(10):2879–2895.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Royle, J. A., Chandler, R. B., Sollmann, R., and Gardner, B. (2014). *Spatial Capture-Recapture*. Academic Press.
- Royle, J. A. and Young, K. V. (2008). A hierarchical model for spatial capture–recapture data. *Ecology*, 89(8):2281–2289.
- Sköld, M. and Åsbrink, J. (2023). Björnpopulationens storlek och utbredning i sverige 2022. *Rapport från Naturhistoriska riksmuseet, 2023:3, Naturhistoriska riksmuseets småskriftserie*, 35pp.
- Sun, C. C., Fuller, A. K., and Royle, J. A. (2014). Trap configuration and spacing influences parameter estimates in spatial capture–recapture models. *PloS one*, 9(2):e88025.
- Theng, M., Milleret, C., Bracis, C., Cassey, P., and Delean, S. (2022). Confronting spatial capture–recapture models with realistic animal movement simulations. *Ecology*, n/a(n/a):e3676. e3676 ECY21-0499.R1.
- Zhang, W., Chipperfield, J. D., Illian, J. B., Dupont, P., Milleret, C., de Valpine, P., and Bischof, R. (2022). A flexible and efficient bayesian implementation of point process models for spatial capture–recapture data. *Ecology*, page e3887. e3887.