Doctoral thesis

Doctoral theses at NTNU, 2021:311

Benjamin Cretois

Transforming the use of citizen science data for biodiversity conservation at different scales

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Social and Educational Sciences
Department of Geography

NTNU
Norwegian University of
Science and Technology

Benjamin Cretois

Transforming the use of citizen science data for biodiversity conservation at different scales

Thesis for the Degree of Philosophiae Doctor

Trondheim, October 2021

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Social and Educational Sciences
Department of Geography

ISBN 978-82-326-5998-2 (printed ver.)
ISBN 978-82-326-5528-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:311

Printed by NTNU Grafisk senter

Summary

Obtaining large amount of data on species' characteristics such as their distribution, abundance or movement patterns is not only important for scientists to better understand species' ecology but it is also fundamentally important for policy-makers and environmental managers because it provides a knowledge platform to ensure successful biodiversity conservation. Citizen science (i.e. the involvement of volunteers who collect and/or process data as part of a scientific inquiry) data has recently gained attention among researchers as it can help researchers tackle research questions that otherwise could not be addressed without the involvement of large numbers of professional data collectors. Nevertheless, because volunteer recorders are highly motivated to encounter interesting wildlife the spatial and temporal patterns of records are not random or systematic and hence very different from the kind of rigorous random sampling protocol that scientists are used to work with.

The aim of this thesis is to understand the citizen science observation generation process and then assess the performance of citizen science observations to infer ecological properties at different ecological scales. This thesis consists of five articles. The first study takes the example of hunters as a special case of citizen scientists and assesses their importance for biodiversity monitoring. In articles 2 and 3 I compile distribution maps of large mammals in Europe using citizen science data in tandem with other source of data to study species' ecology at a macro scales (distribution). Finally, articles 4 and 5 show that if potential biases are accounted for using appropriate statistical methods, citizen science observations can give a good approximation of species' ecology at micro scales (habitat patches).

From this thesis two main conclusions emerge. Firstly, to transform how citizen science is used it is critical to understand the data generation process underlying the creation of the geo-localised citizen science observations in order to fully grasp the extent of the potential biases in citizen science data. Then citizen science data can be used at multiple scales if biases are accounted for using proper methodology. It is important to realize that the methodologies used to account for biases in citizen science data have to be scaled to the research question and to the data available as results could lead to misleading conclusions about the species' ecology. Overall, the future of citizen science remains very bright and this thesis contributes to further understanding and making better use of it.

Acknowledgment

As Batman says in the infamous comic *The Dark Night Returns*, *"the world only makes sense if you force it to"*. This is also true for data: it makes sense only if we force it to and this is exactly what this thesis is about: making sense of noisy citizen science data. This work would have never been possible without the support and help of many persons.

I remember having sent about 200 CVs and motivation letters to randomly selected companies to find my masters' internship. Among the 3 answers I received only one was positive and literally changed the course of my career. Thus, I would like to first thank John Linnell (not the singer) to have offered me the opportunity to do an internship with him. He has been a great mentor and helped me to understand that science is a story to tell. Thanks John!

I would also like to thank Jan Ketil Rød and Erlend B. Nilsen who were great supervisors and who I cannot thank enough for letting me have almost complete freedom to complete my PhD and whose comments and suggestions helped me become a better scientist.

The ideas I had during the PhD matured thanks to the researchers, post-docs and students I met and whose conversation helped me clarify my thoughts. I do not have the place to thank all of them but I would like to mention a few, namely Wouter Koch, Caitlin Mandeville, Kwaku, Jorge Sicacha Parada, Jan Borgelt, Emily Simmonds and Mathew Grainger.

Naturally, this thesis would never have existed without the project "Transforming Citizen Science for Biodiversity" initiated by Bob O'Hara which has been funded by the Norwegian University of Science and Technology (NTNU).

I have been lucky to be part of the Department of Geography at NTNU and the Department of Terrestrial Biodiversity at the Norwegian Institute for Nature Research (NINA) which are both very dynamic environments with brilliant persons. In particular, the administration at the Department of Geography helped me tremendously with what I dislike most: administrative duties.

I would like to thank my family who gave me the opportunity to go to University and who always gave me the freedom to choose the path I wanted to follow, and my old friends Christopher and Gautier who heavily contributed to the person I am now.

The hours spent boxing, climbing and surfing were as important as the hours spent at the office as they contributed to releasing the pressure and I am grateful to the persons who have helped me improving my skills in all these activities.

Most importantly, I would like to thank my partner Ana Silva, one of the best and most dedicated scientist I know who has always supported me and also helped me be become a better person.

List of articles

This thesis contains the following articles:

1. **Cretois, B.**, Linnell, J. D., Grainger, M., Nilsen, E. B., and Rød, J. K. (2020). Hunters as citizen scientists: Contributions to biodiversity monitoring in Europe. *Global Ecology and Conservation, 23*, e01077. https://doi.org/10.1016/j.gecco.2020.e01077

2. Linnell, J. D. * ,**Cretois, B.** * , Nilsen, E. B., Rolandsen, C. M., Solberg, E. J., Veiberg, V., Kaczensky, P., van Moorter, B., Panzacchi, M., Rauset. G., and Kaltenborn, B. (2020). The challenges and opportunities of coexisting with wild ungulates in the human-dominated landscapes of Europe's Anthropocene. *Biological Conservation, 244*, 108500. https://doi.org/10.1016/j.biocon.2020.108500

* *co-first authors*

3. **Cretois, B.**, Linnell, J. D., van Moorter, B., Kaczensky, P., Nilsen, E. B., Parada, J. S., and Rød, J. K. (2020). Coexistence of large mammals and humans is possible in Europe's anthropogenic landscapes. Submitted to *Anthropocene*.

4. Sicacha-Parada, J., Steinsland, I., **Cretois, B.**, and Borgelt, J. (2020). Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. *Spatial Statistics*, 100446. https://doi.org/10.1016/j.spasta.2020.100446

5. **Cretois, B.**, van Moorter, B., Linnell, J. D., Simmonds, E. G., Rolandsen, C. M., Solberg, E. J., Strand. O., Gundersen, V., and Rød, J. K. Identifying and correcting spatial bias in citizen science data for wild ungulates in Norway. Submitted to *Methods in Ecology and Evolution*.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Background

1.1 Towards a different way of knowledge development

It has been argued that traditionally, the development of scientific knowledge has largely been an autonomous process in relation to other societal and institutional domains: scientists do their work in research centres and at universities, after which their findings are communicated to society via science popularisation or specialist journals (Regeer and Bunders 2009). This model is criticized as the "information deficit model" because of the gap between research results and stakeholder needs (Burgess et al. 1998) and assumes that the relationship between science and society is limited to the fact that scientific progress, even though theoretical, can engender societal and economic progress. How scientific knowledge is used in the societal domain is the responsibility of societal actors, not scientists, as they are only responsible for the production of objective knowledge. This paradigm primarily derived from the monodisciplinary scientific approach (Regeer and Bunders 2009). Though enjoying some intuitive appeal, this paradigm have repeatedly been shown, by experience and in research, to be only a construction (Owens 2000). Obvious examples of alternative models can be found in mathematics where the work of Leibniz and Pascal were important for theological purposes (i.e. to prove the existence of God), the natural sciences with the work of Leonardo da Vinci closely linked to improving arts and painting in particular, or in ecology where most discoveries were made by citizens during their leisure time and the idea of professional scientists was non-existent until recently (Miller-Rushing et al. 2012).

This traditional way of thinking about science nevertheless shifted towards more integrative ideas as society began to question the idea that scientific development automatically leads to prosperity, progress and therefore societal relevance (Regeer and Bunders 2009). It was thought that knowledge and technology development should be influenced via policy and financing structures, so that the output of science and technology better meets the needs of society. This has led stakeholders to fund science and technology related to environmental pollution, biotechnology or synthetic materials, to name a few (Regeer and Bunders 2009). Epistemologically, this way of thinking assume that the role of scientific knowledge development is instrumental in resolving societal problems and stimulating the economy. This has led to the development of multi- or trans- disciplinary research (Centre for Educational Research and Innovation and OECD 1972).

In today's paradigm the development of scientific knowledge is thought to be inseperable from the societal demand for solutions to real-world problems and societal needs are the basis for scientific development (Lemos et al. 2018). There has been a shift of focus from generating scientific knowledge (based on the assumption that this will lead to progress), to resolving unstructured problems through integrated societal and scientific (technological) innovations (Lemos et al. 2018). The switch in paradigm also changed the way knowledge is produced as it is now accepted that scientists and other societal actors can generate new knowledge via mutual interaction by clarifying the problems and generate possible solutions. With this paradigm, scientific development becomes more of a socio-technological innovation process. There is no question of 'knowledge transfer' or of applying (or making applicable) scientific knowledge in a specific problem context: the knowledge will be generated through a socio-technological innovation process (Regeer and Bunders 2009). This implies that the demarcations between the different actors are much less clear; actors merge and are closely interwoven in a network of knowledge. This paradigm has heavily contributed to the democratization of science and the inclusion of citizens, or non-experts, in the scientific process of knowledge generation. Their contribution is now obvious and has been proven helpful in numerous disciplines including physics, geography, and a wider breadth of topics. For instance, the project Galaxy Zoo invites the general public to visually inspect and classify the shape of galaxies via internet (Lintott et al. 2008). Citizens can also contribute to monitoring noise pollution using mobile phones (Maisonneuve et al. 2010), monitoring water quality (Kinchy 2017), or engage in different project in ecology and conservation, such as butterfly or bee monitoring (Silvertown 2009). Figure 1.1 summarizes the current paradigm regarding the role of different stakeholders in the process of knowledge generation.

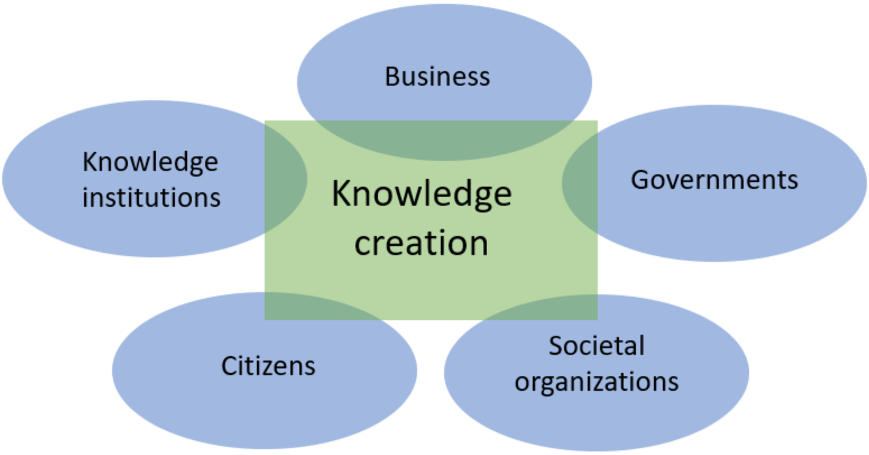


Figure 1.1: Current societal paradigm regarding the role of different stakeholders in the process of knowledge generation. Adapted from Lemos et al. (2018).

1.2 The opportunities of citizen science for biodiversity conservation

The increased use of natural resources due to the rapidly growing human population has caused a severe decline in global biodiversity (Díaz et al. 2019). Nevertheless, the international community has been pushing for the development of policy goals to achieve biodiversity conservation and human-wildlife coexistence such as the Aichi Biodiversity Targets set by the Convention on Biological Diversity and the United Nations Sustainable Development Goals (UN General Assembly 1992; 2015, Cretois et al. 2019). Fulfilling these goals requires tremendous amounts of data to get reliable information to assess the status and trends of global biodiversity. For instance, the EU's Habitat and Bird Directives require changes in species status to be reported every six years, and progress against the Convention of Biological Diversity targets is reported on a decadal basis. To help track global biodiversity change, the Group on Earth Observations Biodiversity Observation Network (GEO BON) proposed a candidate set of the key characteristics of biodiversity refereed to as Essential Biodiversity Variables or EBVs (Pereira et al. 2013). EBVs represent the minimum set of measurements needed to capture major dimensions of biodiversity change, and are now being used to set up frameworks for global monitoring of biodiversity (Kissling et al. 2015). Despite these efforts, there is still a lack of knowledge about species' ecology, and more data is needed to ensure successful conservation. The lack of data is partly due to the challenges in developing monitoring tools that are efficient and not too costly to researchers. Fortunately, citizen science has repeatedly been proven to be a cost-effective solution to gather massive amounts of information on biodiversity at different scales and resolutions (Thornhill et al. 2021).

Citizen science means different things to different people, causing confusion about its nature and utility. In fact, the wording differs among disciplines and topics, and citizen science is also refereed to as citizen sensing, spatial crowdsourcing or volunteered geographic information (Sheth 2009, Zhao and Han 2016, Goodchild and Li 2012). As noted in Goodchild (2007), spatial crowdsourcing, citizen sensing, volunteered geographical information and citizen science are all similar terms relating to citizens acting as sensors to collect spatial information that are made available for the common good. In this thesis I use a widespread definition of citizen science and define the term as *the practice of engaging the public in a scientific project that produces reliable data and information usable by scientists, decision-makers, or the public and that is open to the same system of peer review that applies to conventional science* (McKinley et al. 2017). Citizen science data has recently gained traction among researchers as it can be interpreted as a way of addressing research questions that otherwise could not be addressed without the involvement of large numbers of professional data collectors, at great cost (Silvertown 2009). Nevertheless, this is not a new phenomenon since prior to the professionalization of science in the late 19th century, nearly all scientific research was conducted by amateurs (i.e. by people who were not paid as scientists, Vetter 2011). These amateurs themselves recruited other non-experts passionate to contribute to natural history observations. For instance in the mid-18th century, a Norwegian bishop created a network of clergymen and asked them to contribute observations and collections of natural objects throughout Norway to aid his research. This resulted in the first Norwegian flora and close to thirty papers on natural history topics (Brenna 2011). Recruiting volunteers was a common way for early ecologists, such as John Ray and Carl Linnaeus, to collect specimens and observations from across the known world (Miller-Rushing et al. 2012).

The popularity citizen science is gaining has even resulted in the establishment of professional citizen science societies around the world. The largest are based in the United States (the Citizen Science Association, `citizenscience.org`), in Europe (the European Citizen Science Association, `ecsa.citizen-science.net`) and in Australia (Australian Citizen Science Association `citizenscience.org.au`). This popularity has also helped the creation in mid-2016 of a new peer-reviewed journal for citizen science researchers and practitioners: *Citizen Science: Theory and Practice* (`theoryandpractice.citizenscienceassociation.org`). The recent explosion in citizen science activity can be partly explained by the current widespread accessibility of technology (e.g. internet, portable devices such as mobile phones and tablets) which has contributed to the increase of citizen science records by substantially increasing project visibility, functionality, and accessibility (Bonney et al. 2014). People who are passionate about a subject can quickly locate a citizen science project they are interested in, follow its instructions, submit data directly to online databases, and join a community of like-minded people. Particularly, mobile applications have made it easy for citizen scientists to submit data about any species they see at any time and in any location (Bonney et al. 2014). Mobile applications have been shown to be very effective tools in engaging many citizens passionate about biodiversity conservation, and AI recognition algorithms have made citizen science accessible for any novice. For instance, the eBird application (A mobile application initiated by the Cornell Ornithology Lab) is used by more than 500,000 birders who collected a total of more than 737 million bird observations (eBird 2019). The success of citizen science is not limited to birders as iNaturalist, a more generalist citizen science platform, is used by more than 1,000,000 citizen scientists (Scott 2020). In fact, it has recently been estimated that as much as half of the species occurrence records stored in the Global Biodiversity Information Facility (`www.gbif.org`) have been collected by Citizen Scientists, and the proportion of records is still increasing every year (Walker 2019). Citizen scientists are also able to collect information on almost all EBVs proving that citizen scientists make substantial contributions to large-scale international biodiversity monitoring for a wide range of taxa (Chandler et al. 2017). Hunters in particular are able to collect data on biodiversity in all its key dimensions (article 1, Figure 1.2).

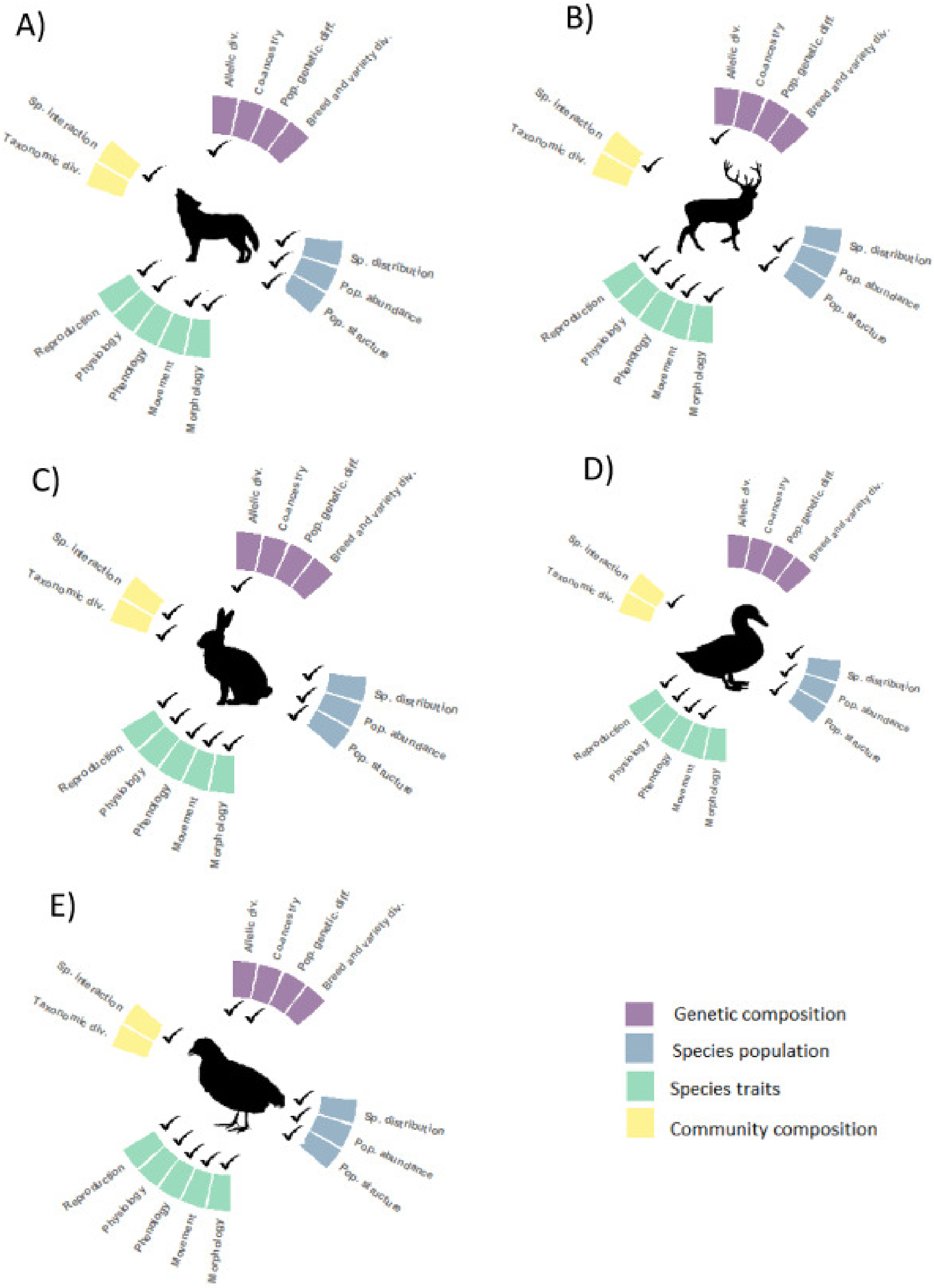


Figure 1.2: Diversity of Essential Biodiversity Variables (EBVs) covered by hunter-based monitoring. A) large carnivores, B) ungulates, C) small game, D) other game birds and E) waterfowl. A tick indicates if the Essential Biodiversity Variable is being monitored by hunters. Colors represent the Essential Biodiversity Class. Extracted from article 1.

Besides providing scientific data, citizen science is also increasingly seen as a way to engage the public in science, improve scientific literacy and interest in science, and inform participants about particular topics, such as butterfly ecology, conservation, or climate change (Lowman et al. 2009). Citizen science also has a strong impact on the perception citizens have of research by facilitating and increasing the engagement of public in conservation science and management (Ellwood et al. 2017). Citizen science also builds knowledge legitimacy as it leads to longer-lasting effects on stakeholder relationships, social learning, and implementation of environmental legislation which are necessary conditions for successful conservation strategy implementation (Armitage et al. 2015, Sullivan et al. 2017). Citizen science is a cooperative approach that has been shown to strongly contribute to improve co-production of knowledge and evidence as for instance in 2019 only, researchers published 58 peer reviewed publications using eBird data (eBird 2019).

1.3 The challenges of using citizen science data in biodiversity conservation

Despite the wealth of information emerging from citizen science projects, the practice is not universally accepted as a valid method for resolving scientific questions. Scientific papers presenting volunteer-collected data sometimes have trouble getting reviewed and are often placed in outreach sections of journals or education tracks of scientific meetings (Bonney et al. 2014). Statements about the value of citizen science data are often qualified with references to the biases and limitations opportunistic volunteer-collected data are associated with (Bird et al. 2014). Records are usually made only for species that were observed and do not contain any information about whether other species were **not** observed (henceforth the name "presence-only" data), limiting the inference researchers can extract from them (Tingley and Beissinger 2009). A main problem in citizen science data is that volunteer recorders are highly motivated to encounter interesting wildlife and visit pleasant areas. This means that the spatial and temporal patterns of records are not random and hence very different from the kind of rigorous random sampling protocol that professional ecologists are used to work with (Tulloch et al. 2013, article 4). This "recorder effort problem" (Hill 2012) result in four major biases in records data: (1) uneven sampling intensity over time, (2) uneven spatial coverage, (3) uneven sampling effort per visit, and (4) uneven detectability across time and space (Isaac et al. 2014).

Uneven sampling over time is the best-known form of bias. The number of records being generated has increased tremendously in recent years, and for many groups the growth is approximately exponential (Isaac and Pocock 2015). Citizen science observations are often aggregated into grid cells to give information about species' distribution and abundance (see maps used in article 2). However, as recording intensity increases, the number of grid cells that appear to be occupied is likely to increase, even for species with stable distributions (Telfer et al. 2002). If naively used to infer trends in species distribution, not accounting for this bias can be highly misleading. Uneven spatial coverage occurs because most recorders tend to submit records in areas that are easily accessible such as near where they live, near roads and even in places they enjoy to visit (Tiago et al. 2017, article 4, article 5). Patterns of variation in species richness may therefore be misleading because people choose to visit places which are especially diverse for their taxon of interest (Prendergast et al. 1993). Sampling effort per visit refers to the degree to which any one set of records is an accurate reflection of the organisms that were actually present (Isaac and Pocock 2015). Sampling effort can be decomposed into two components: the intensity of the search effort and the set of species that were surveyed and reported. The first of these reflects the standard concept of a species accumulation curve: the more time a person spends searching, the more species this person finds. The second component is best illustrated by the concept of the 'complete list': a complete list of birds means that the citizen scientist reports each bird species observed during a particular trip. This does not mean that all bird species are reported because the citizen scientist might not have seen all of them (detection is always less than perfect and some species are easier to spot than others, Kéry et al. 2009) but he / she reported all the species detected. Incomplete lists occur because many biological records are not visit lists (i.e. there is no 'survey'), but rather 'incidental records' (e.g. just the 'interesting' species, Isaac and Pocock 2015).

Common approaches used by researchers to handle biases coming from the "recorder effort problem" (Hill 2012) include for instance aggregating the data to certain spatio-temporal scale at which the biases in space and time might be argued to be averaged out. For instance, in a grid cell of 10km^2 small scale biases such as distance to roads might not matter, and it is possible to get a species' distribution map by counting the number of occupied grid cells across the whole country (Telfer et al. 2002, Thomas 2004). Another approach consists of selecting records which are thought to be free of biases (Roy et al. 2012). New approaches do not attempt to remove the biases but rather correct for them by explicitly modelling the data collection process (Altwegg and Nichols 2019, Yuan et al. 2017, Hochachka et al. 2012, Johnston et al. 2018). While the first two methods throw away a large amount of information they may be well suited for certain ecological questions such as studies at species' distribution at a continental scale which do not need accurate measurements (article 2, article 3). The latter approaches on the contrary attempt to retain a maximum of information to make use of citizen science observations at smaller scale (article 4, article 5)

While some citizen science projects such as eBird allow citizen scientists to submit "checklists" which can be used by researchers to control for some biases such as effort, most citizen projects do not collect such information and biases are more difficult to account for. In fact, complete lists are of great help for ecologists and statisticians. We would expect the identity of species on the list to be some function of their local population density and visual apparency (Royle and Nichols 2003). The length of the list is then an indication of the duration and intensity of the survey (Szabo et al. 2010). Nevertheless, citizen science projects usually have a very simple sampling design that violates the fundamental principles of sound experimental designs, we refer to these programs as being "unstructured". Most data collected by volunteers are unevenly distributed in space (i.e. often concentrated close to roads and urban centers), time (i.e. collected during daylight and during the weekends). Moreover, observers differ in their abilities to recognize species and the effort they spend to detect certain species (Isaac et al. 2014).

1.4 Aim of the thesis

This thesis assesses the potential of citizen science observations to infer ecological properties at different scales. The point of departure is the understanding of the data collection process and the diversity of data collected by citizen scientists for biodiversity conservation. I will emphasize the role of hunters, a specific group of citizen scientist who are one of the main contributor of data for wildlife monitoring and management which collect a broad range of data on diverse biodiversity characteristics or EBVs (article 1). Establishing knowledge on how citizen science data are created is fundamental to understanding the biases contained in the observations, and creating solutions to use citizen science data to its full potential. I will assess the extent to which citizen science can be used for deriving inference about species' ecology at different ecological scales. I will show that at a continental scale (i.e. at macro habitat level) citizen science data can provide a good description of species' distributions when pooled with distributional data from other sources (article 2) and that from this description it is possible to derive useful results for the science of human-wildlife interactions (article 2 and 3). I will also downscale the study area and show that using citizen science data at a fine scale (i.e. at a micro habitat level) possibly results in misleading conclusions if the spatial biases inherent to citizen science observations are not accounted for (article 4 and 5). Figure 1.3 shows how the articles in the thesis relate to each other.

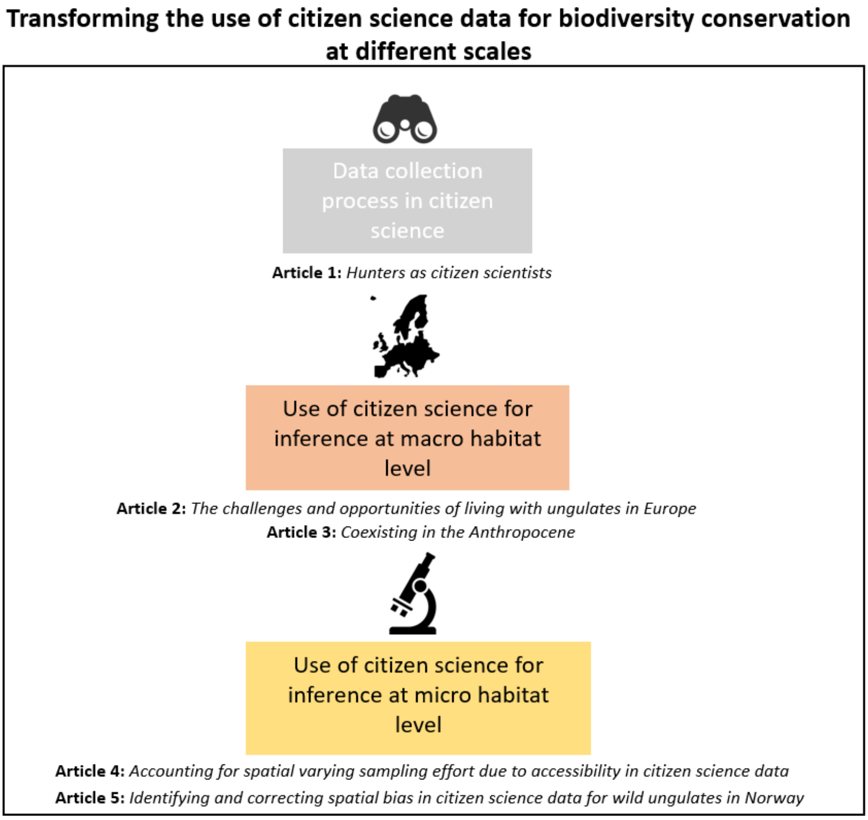


Figure 1.3: Components of the PhD project and how they relate to each other.

1.5 Structure of the thesis

The thesis is comprised of two parts and is organized in the following way. The first part includes a comprehensive introduction that elaborates on the theoretical and methodological aspects of the thesis. In addition, I discuss the key findings and overall conclusions of the thesis. In Chapter 2, I introduce the concepts of data generation processes and biases in citizen science observations. I also introduce to the concept of scale and how this relates to the biases found in citizen science observations. In Chapter 3, I outline my methodological choices, and describe the data material I analyzed. In the concluding chapter (Chapter 4), I draw a brief concluding summary of the thesis and summarize the key findings from the five research articles. The second part of the thesis comprises the five research articles.

Chapter 2

Conceptual issues of citizen science data and scale

2.1 Collection processes and citizen science data

To understand how biases in citizen science data might influence inference we first need to understand how citizen science observations are generated (i.e. the data collection process). Figure 2.1 is an attempt to represent this process.

First, species presence or absence is determined by the species niche, a fundamental concept in ecology defined as the subset of n-dimensional space of all possible environmental conditions in which a species can survive (Hutchinson 1991). Nevertheless, a species can be absent from a place where environmental conditions meet the species' requirements for diverse reasons including dispersal limitations, historical factors, or anthropogenic interactions such as overharvesting or poaching (Price and Kirkpatrick 2009, Heim et al. 2019). If the species is present it needs to be observed either by a person such as a citizen scientist as represented on the figure or by a device such as a camera trap, acoustic recorder, or a drone. Even though both the species and the citizen scientists are present, the latter has to see or hear the animal to observe it. Once the species has been observed, it needs to be correctly reported (i.e. uploaded on a citizen science platform), which is not always the case as the reporting process might also depends on multiple factors such as the ability to identify the species or the motivation of the citizen scientist to do so as seen in article 1 in the case of hunters. It should be noted that the observation can be reported prior to identification, for instance a citizen scientist posting a picture of a bird on a social media platform to get help identifying the species. Failure to either observe or report the observation results in a

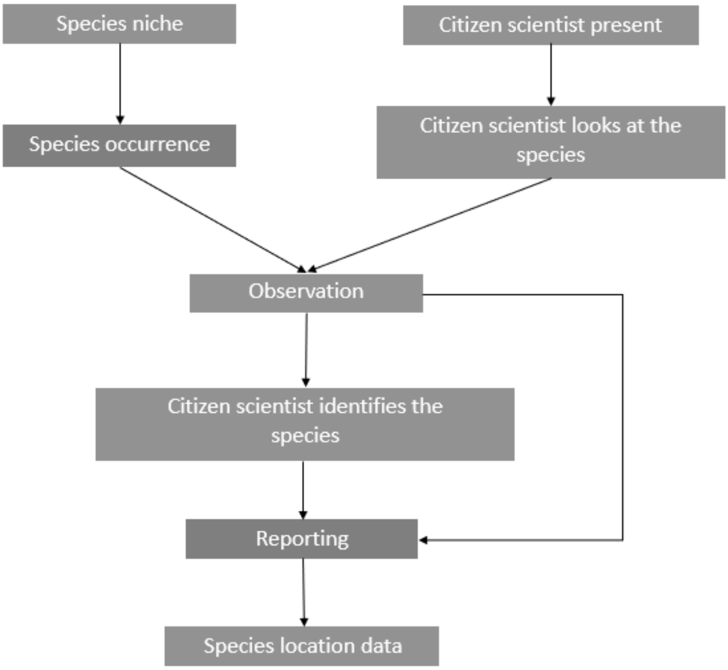


Figure 2.1: The process of generating a citizen science observation.

"false absence" which constitutes the most important source of uncertainty in species conservation (Guillera-Arroita et al. 2017). The process finally results in a geocoded data point (i.e. points with coordinates from a certain spatial reference such as latitude and longitude or UTM) that is often stored in a spreadsheet and that is used by researchers for diverse biodiversity related research. It is also possible that the data point in the database has been incorrectly identified, resulting in a false presence. In this thesis I do not account for this source of bias as the species studied in all articles are easily recognizable (i.e. large ungulates), greatly minimizing this source of bias. Moreover, accounting for this bias requires very different conceptual and statistical tools that are outside the scope of this thesis.

Below we go in detail through the steps leading to the creation of the citizen science observation.

2.1.1 Observation process

Generally in data collection, understanding the observation process is critical as it gives indications about the quality of the data and thus information on its limits for drawing inferences. In citizen science the observation process is driven by multiple factors which can be grouped into four broad categories:

Where

To be able to map or make inference about species' distribution, it is usual for researchers to attempt to cover a large portion of the landscape to observe whether a species is present in a certain area or not. However, citizen scientists tend to be influenced by factors facilitating accessibility through the landscape. This results in spatial biases. Citizen science observations occur more often in locations that are more easily accessible (i.e. not in the deep forest) and distance to roads and public transport networks are variables which have been shown to be influential as illustrated in Figure 2.2 (Tiago et al. 2017, article 4, article 5). Moreover, citizen science data are usually located near areas that are close to human infrastructure and distance to urban settlements and population density are all variables which are positively related to the number of citizen science observations in an area (Tiago et al. 2017, Mair and Ruete 2016, article 5). For example, Figure 2.3 shows the distribution of moose GPS telemetry locations (i.e. wild moose equipped with tracking devices by biologists), citizen science observations, and available locations within the animal home range. Figure 2.3 clearly demonstrates that citizen science observations are on average closer to urban settlements and on paths that are more heavily utilized than telemetry locations ($median_{distance\,to\,urban\,settlements} = 3672, 11541, 4020\,meters$ and $median_{log\,path\,use} = 2.89, 0, 0$ for citizen science observations, available locations and telemetry locations respectively).

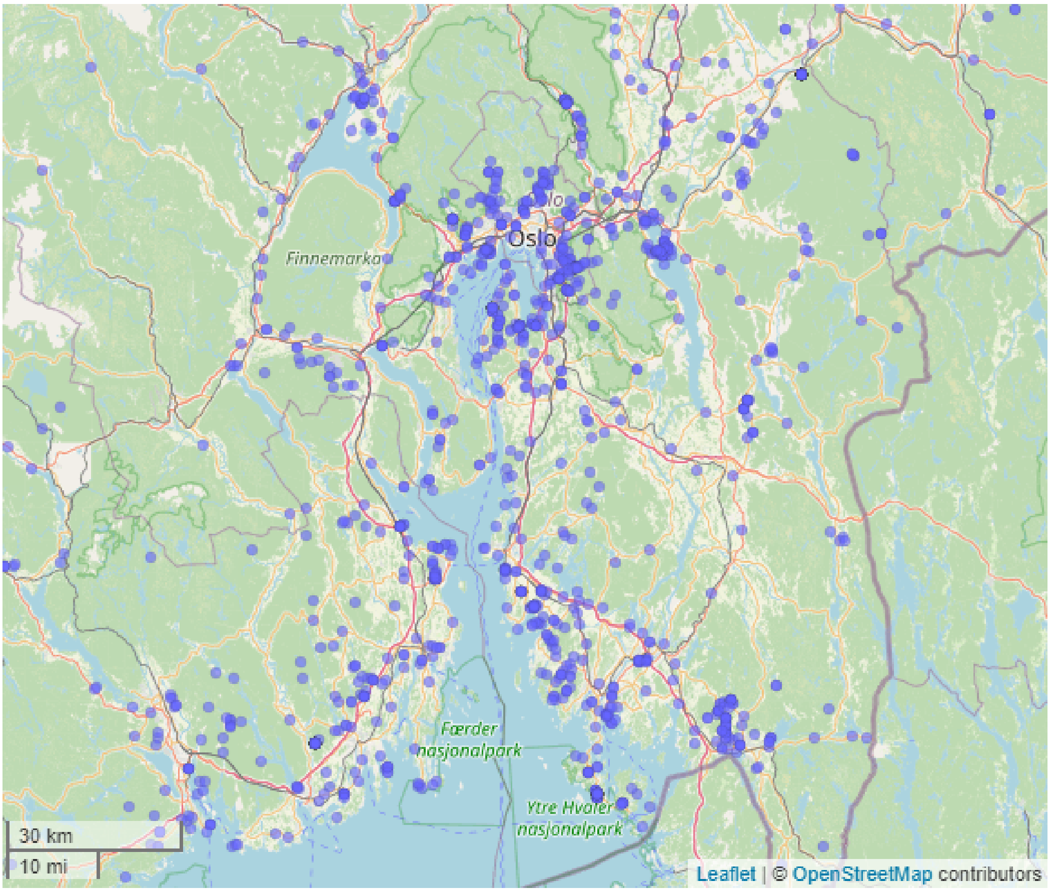


Figure 2.2: Distribution of roe deer citizen science observations in Oslo area. Purple dots represent roe deer sightings. The basemap is taken from OpenStreetMap.

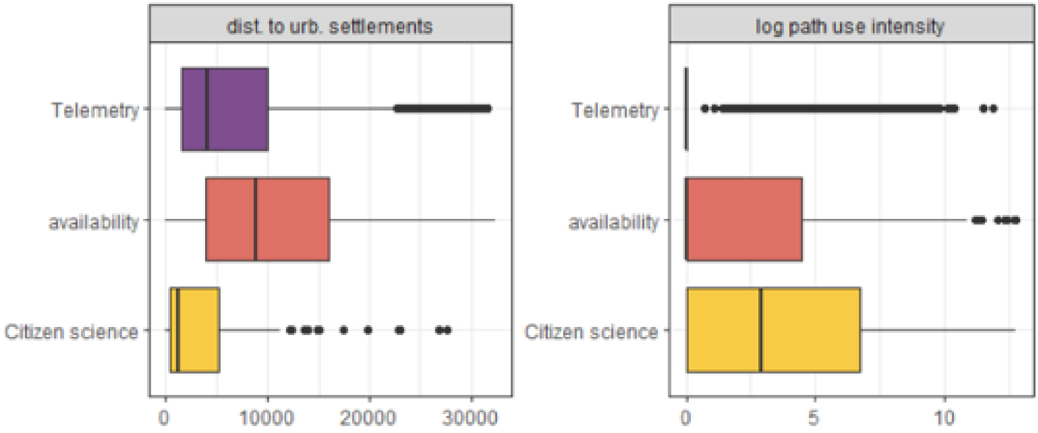


Figure 2.3: Boxplots of the distribution of telemetry locations, available locations and citizen science observations within the distance to urban settlements (in meters) and log path use intensity (calculated as the log number of persons having registered a STRAVA activity on each path) spectrum for moose. Figure made with the data used in article 5.

Even for citizen scientists who are less influenced by human infrastructures (e.g. hikers, skiers) for whom we could assume that observations are less spatially bias, studies observe that most locations occur close to footpaths, hiking trails and ski tracks, and are negatively influenced by landscape variables such as terrain ruggedness and elevation (Mair and Ruete 2016).

Finally, the perception of aesthetic values in a landscape might also drive citizen scientists to prefer certain places that are considered more "natural" or for which the citizen scientist feels more attached culturally (Bubalo et al. 2019, Setten 2004).

When

Spatial bias in citizen science dataset is well studied and well understood (Tiago et al. 2017, Mair and Ruete 2016, Geldmann et al. 2016). Nevertheless, it is important to be aware of the uneven distribution of observation in time that also exist in citizen science datasets.

Because most citizen scientists have more time for leisure activities during non-working day more observations are reported during the weekends than during weekdays. This can have dramatic influences for scientists studying birds' migratory pattern using citizen science data for instance (Courter et al. 2013).

The volume of citizen science data is also growing at a very fast pace every year and for instance, the amount of citizen science data coming into the GBIF database has increased exponentially for over a decade and nowadays citizen science observations make up around 50% of the biodiversity knowledge on the GBIF network (Figure 2.4). These between year biases lead to distorted views of biodiversity trends. For instance, a comparison of pre-1930 and post-1990 data for all species across the Indian Subcontinent suggests dramatic biodiversity loss from the central plains, for example, but not from the Himalayas. However, this may simply be a sampling artefact. Ornithologists no longer tend to visit low-biodiversity areas, making it difficult for researchers to infer these regions' biodiversity status (Boakes et al. 2010).

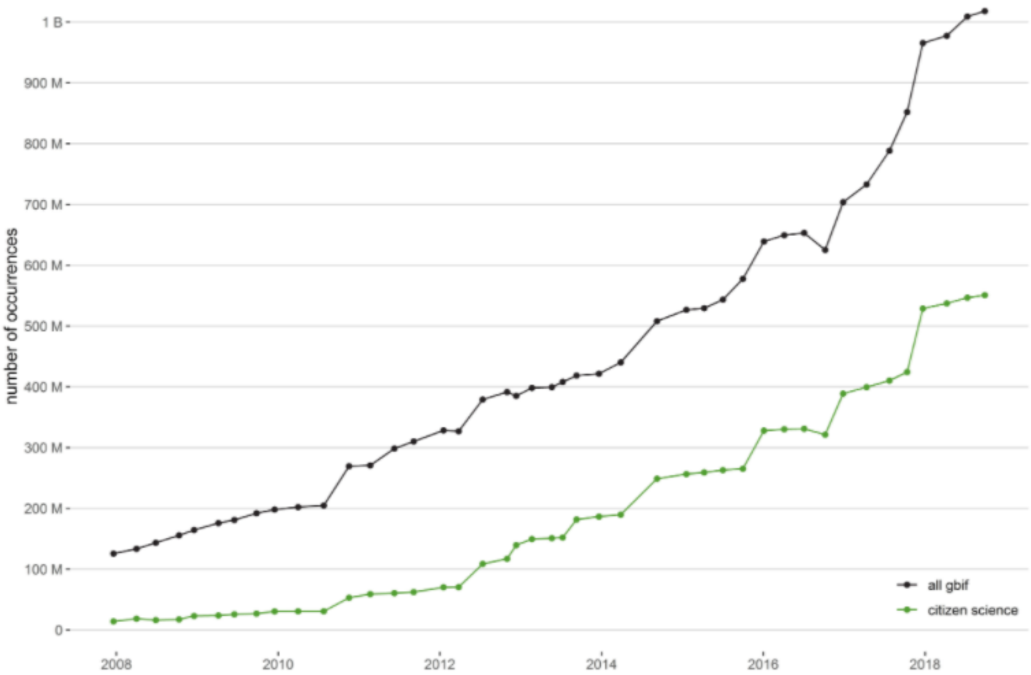


Figure 2.4: The growth of citizen science in GBIF database. Extracted from Walker (2019).

Temporality also affects species behavior and species have different daily and seasonal activity patterns. Some species such as bats are more active during the night, making them harder to spot for any citizen scientist going into the field during daylight (Razgour et al. 2016). On a yearly basis animals' activity patterns might also change and some species such as wild reindeer, will migrate or other, such as bears, hibernates during winter modifying the probability of their detection. Moreover, breeding period might be beneficial to the observation as animals are typically more active and emit loud sounds (Juillet et al. 2012).

Why

If scientists are to see citizen scientists as colleagues rather than ‘data-drones’ and if they want to get a grasp on how the data are generated, they must gain an understanding of volunteer motivations for participating. In fact, experience has shown that different segments of the population are more prone to participate than others and their motivation might impact the range of species that are observed, the effort put into searching for species, and hence the probability of detecting a species (Soleri et al. 2016, Dickinson et al. 2012).

It has been suggested that biodiversity recording can function as an excuse for spending quality time outdoors and “indulging in nature, or fleeing from family and social demands” (Lawrence and Turnhout 2010). Such citizen scientists might not search as hard as dedicated citizen scientists and might record more sporadically than other types of citizen scientists. This sort of motivation can influence the degree of expertise of a citizen scientist. A very motivated citizen scientist will learn faster than an unmotivated citizen scientist and might be able to identify and spot species more accurately, equating the quality of data collected by professionals (Kosmala et al. 2016).

In addition to learning about and connecting with nature and biodiversity in general, some authors have called attention to the importance of specific places in motivating nature-based citizen science. Researchers must pay attention to the importance of sense of place, i.e. the meaning of, and attachment to, specific sites, which is a possible motivation for, and result of, participation in citizen science. In fact, studies show that a stronger sense of attachment to place induced repeated sampling visits from the citizens, data of great value for ecologists and increased attention to species threatening the "naturalness" of the landscape such as invasive species (Haywood et al. 2016, Niemiec et al. 2017).

How

Finally, the way citizen science is done is very much linked to the motivation and can also induce biases in the detection probability of a species (Altwegg and Nichols 2019, van Strien et al. 2013). The length of time spent at a location has tremendous consequences on the probability of detection. This is especially the case for elusive and rare species which require rigorous searches in the area of interest to be detected (MacKenzie 2018). The method used to detect a species is also an important factor. While most citizen science observation are done by direct sighting, birders (i.e. citizen scientists who have a specific interest for birds) might also use sound as a detection method. Therefore, a citizen scientist with hearing deficiency might not be able to detect the target species (Wood et al. 2011).

2.1.2 Reporting process

On one hand reporting is facilitated by new technologies and tools such as the gamification of citizen science which has been shown to be beneficial to engage citizen scientists in reporting more species. Researchers have identified “rewards", such as online gaming badges and competitions as a promising method for motivating and retaining volunteers (Cooper et al. 2010). For other groups of citizen scientists (i.e. hunters) reporting is mandatory as public authorities have to control their activity and use their data as a primary source of wildlife monitoring (article 1). Most citizen scientists do citizen science as a leisure and gaming or being part of a citizen science "community" might induce a sense of camaraderie that make scientific exploration and discovery enjoyable (Newman et al. 2012).

Several authors have found that connecting with, and learning about, local nature and biodiversity is a key motivation for participation in nature volunteering and Bruyere and Rappe (2007) found that to ‘help the environment’ was by far the highest-scoring motivation in their factor analysis of environmental volunteer motivations. This can have an influence on the taxon recorded as citizens who are concerned about biodiversity loss could disproportionately record endangered or rare species even though they observe other taxa. This can result in a distorted picture of the species’ abundance and distribution in a certain area (Dickinson et al. 2012).

Nevertheless, multiple reasons can induce the non-reporting of citizen science observations. While contributing to science is an important motivation for volunteer biodiversity recorders, they also value the context of their sightings and need to trust that their data will be used respectfully and for the right purposes (i.e. for biodiversity conservation and not for monetary purposes, Lawrence and Turnhout 2010). In fact, records might be considered public property by many volunteers, but if these geocoded "nature experiences" are used for the reasons that do not match citizen scientists’ expectations they might ultimately decide to withhold them or to purposely induce biases (Ganzevoort et al. 2017). Trust between citizen scientists and researchers is therefore a prerequisite for rigorous reporting and it has been highlighted in several studies that researchers should keep the volunteers updated on the how and why of use of their data (e.g. Groom et al. 2017, Dickinson et al. 2012).

Other more practical reasons negatively influencing the report of observations exist. A fisherman or a hunter might be reluctant to share the location of areas in which the harvest has been successful and thus purposely fail to report or geographically mis-reference the data point. In fact, it has been shown that anglers are particularly concerned about applications revealing secret locations, advertising high catch rates, and being used by agencies to spy on anglers (Venturelli et al. 2017). Hunters, on the other hand might report a higher harvest rate or a higher number of observations to artificially boost population numbers, potentially increasing the quota for the following year (Popescu et al. 2016, article 1).

2.1.3 The diversity of data and spatial coverage of citizen science

Biases explained in the previous sections mainly impact work on spatial-temporal characteristics of the species including distribution, abundance or phenology. However, it is recognized that citizen science has the potential to obtain data on a very wide range of species' characteristics such as species' body mass or ecosystem properties such as productivity and habitat structure over a wide geographic range and can contribute to large-scale international biodiversity monitoring (Chandler et al. 2017). Nevertheless, even though this diversity of data is possible, citizen scientists are biased toward monitoring species distribution particularly for birds in North America and Europe and there is a critical lack of data for less charismatic species such as insects, with the exception of butterflies (Amano et al. 2016).

However, not all groups of citizen scientists are equal in their contribution to the diversity of species' characteristics and hunters' data have been used extensively as a reliable source of knowledge by wildlife researchers for multiple characteristics over very large areas (Europe, North America). This rational has made us inquire about the extent of the role of hunters in biodiversity conservation and in article 1 we summarized the contribution of hunters to biodiversity conservation in Europe through a systematic literature review, a method aiming to summarize the state of the art in a subject field (Rowley and Slack 2004). For the review we gathered published literature including papers and books in tandem with expert knowledge. To gather the literature I made use of different online search engines and NINA's library. Moreover, collecting expert's knowledge has been made easy through my co-authors vast networks of conservation professionals.

2.2 The notion of scale in ecology and its relation to citizen science bias

While I have described the biases found in citizen science which are primarily driven by factors influencing the observation and reporting process, to fully grasp the contribution of this thesis it is important to introduce the concept of *scale*. It is important to note that the term scale has many meanings, and it is strongly dependent on context and discipline. For instance, in human-geography, the term scale has been recognised as having three aspects, namely size of the study area (e.g. census tract, province, continent), level (e.g. local, regional, national), and relation (as a complex mix that includes space, place and environment, Dabiri and Blaschke 2019). On the other hand, ecologists and physical geographers defined spatial scale as having two main components: (1) *grain*, which is referring to the size of the individual units of observation (i.e., the smallest entities that can be distinguished) and (2) *extent* defined as the size of the study area or landscape (O'Neill et al. 1989). The extent is simply the spatial domain over which the system is studied and for which data are available. The choice of scale is influenced by both the research question and the availability of data (it is not always possible to have access to fine scale data). In this thesis I use the ecologists' definitions of scale.

Macrohabitat and microhabitat are two sub-components of the extent and are common terms in ecology that relate to the landscape level at which a study is being conducted. Generally, macrohabitat refers to landscape-scale features such as distribution areas or zones of specific vegetation associations (Block and Brennan 1993). On the other hand, microhabitat usually refers to a small, localized habitat patch within an ecosystem such as a species home range or at even finer scales like a decomposing log. Johnson (1980), a key reference paper in spatial ecology, recognized this hierarchical nature of habitat use where an animal will be located in a higher order habitat if it has first selected the lower order. For instance, for a wild reindeer to be able to forage and breed in the plateau of Dovrefjell (high order habitat), it first needs to thrive within Norway (lower order habitat). Johnson summarized four natural ordering habitat selection processes (Johnson 1980, Figure 2.5).

- **First-order selection**. This is essentially the selection of the physical or geographical range of a species.
- **Second-order selection**. This is the home range of an individual or social group within their geographical range.
- **Third-order selection**. This relates to how the habitat components or patches within the home range are used (i.e., areas used for foraging).
- **Fourth-order selection**. This order of habitat selection relates to how components of a habitat are used. If third-order selection determines a foraging site, the fourth-order would be the actual procurement of food items from those available at that site.

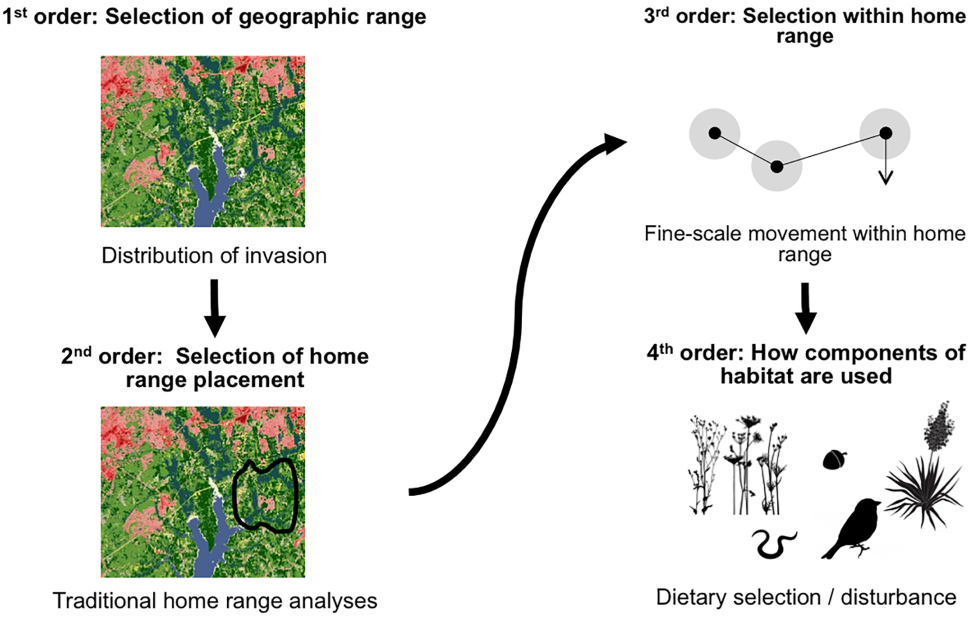


Figure 2.5: Diagram based on Johnson (1980), indicating four scales of habitat selection. The picture illustrating the first order can be represented as grid cells where the species has been considered "present". Within this "presence" grid cell the species will mostly stay within its home range as shown by the polygon in the second figure (2nd order). Within its home range the species will use only certain areas (3rd order). Finally, within the areas intra-home range, the species will select only certain components to fulfill its biological and physiological needs. Extracted from Froehly et al. (2020).

Based on this hierarchy, macrohabitat is the first-order of habitat selection and microhabitat is the combination of the second, third, and fourth levels (Krausman 1999). Understanding these levels can have profound influences on the management of a species.

2.2.1 The problem of scale: the Modifiable Areal Unit Problem

Spatial data are often encoded within a set of spatial units that exhaustively partition a region, where individual level data are aggregated, or continuous data are summarized. Census data and species abundance data are obvious examples. For instance, it is possible to survey the human population of a region to know if they have already been infected by COVID-19. While information is available at the individual level, it is more informative for health authorities to sum the information at a municipality level to evaluate infection rate. The same principle holds for species abundance: it is possible to extensively survey a region to locate all species and infer their presence. It is then possible to aggregate the presences of all species into N x N km squares and sum the number of species found in this particular square to obtain species' abundance. Partitioning schemes (i.e. the way space is divided) can vary by scale, where one partitioning scheme spatially nests within another. This is shown in Figure 2.6 a) where a single unit of the third figure can host four units of the first figure. Partitioning schemes can also vary by zoning, where two partitioning schemes have the same number of units but the unit shapes and boundaries differ. This is demonstrated in in Figure 2.6 b). While rectangles in the first figure have equal areas, they have different shapes and areas in the last figure. The Modifiable Areal Unit Problem (MAUP) refers to the fact that the nature of spatial partitioning (both scale and zoning) can affect the interpretation and results of visualization and statistical analysis (Openshaw 1981).

Even though different disciplines have developed their own notions of the "scale problem" (Marceau 1999) the Modifiable Areal Unit Problem, a concept developed initially by geographers, is certainly the most comprehensive treatment of the sensitivity of analytical results to the definition of data collection units found in the statistical and geographical literature. The MAUP arises from the fact that areal units are usually arbitrarily determined and "modifiable". The term "modifiable" refers to the choice of number of spatial units (i.e. the scale of the analysis) and their configuration (i.e. the shape of the partitioning) which is subjective as other choices could have been made (Haining 2003). Thus, the MAUP has two related but distinctive components relating to the geographic scale and the zoning effect (i.e. the placement of zonal boundaries Haining 2003).

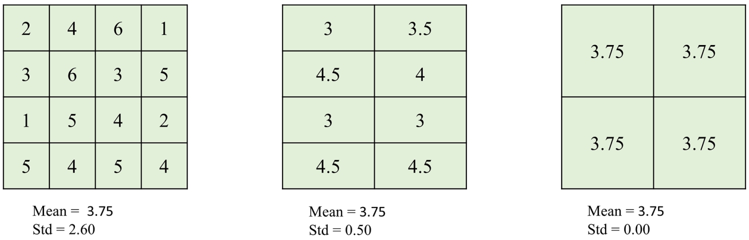
- **The scale effect** is attributed to variation in numerical results owing strictly to the number of areal units used in the analysis of a given area (Openshaw 1981).

- **The zonation effect** is attributed to changes in numerical results owing strictly to the manner in which a larger number of smaller areal units are grouped into a smaller number of larger areal units (Openshaw 1981).

The quantitative effects of the MAUP are illustrated in Figure 2.6 . On a) we observe that the choice of the scale influences the standard deviation of the area. At a fine scale (i.e. more spatial units) we observe differences in the numbers contained in each unit, resulting in a standard deviation of 2.60. Nevertheless, aggregating the smaller units within larger spatial units erodes differences between spatial units and at the largest scale numbers in each spatial areas are equals, resulting in a standard deviation of 0.

Zoning has an influence on both the standard deviation and the mean as illustrated in Figure 2.6 b). On the first figure the area is divided into units with equal areas and the mean and standard deviation are 3.75 and 0.93 respectively. When the same area is divided into units that are different in shape and size such as on the third figure, the mean and standard deviation change and in this particular example the mean decreases while the standard deviation increases when comparing to the first figure (mean = 3.17, Std = 2.11).

a) Effects of aggregation



b) Effects of zoning systems

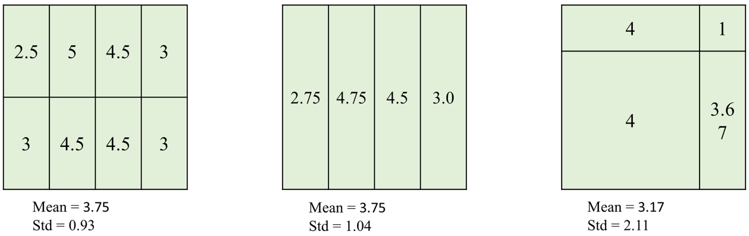


Figure 2.6: The modifiable areal unit problem and its two components: the scale effect and the zoning effect. Both the choice of scale and zoning can affect the mean and standard deviation of the study area, possibly influencing the way results are interpreted Figure from Da Re et al. (2020).

2.2.2 Using citizen science data at macro and micro habitat level

MAUP is closely related to the *ecological inference fallacy* concept which refers to the assumption that an individual has the same attributes as the aggregate group to which it belongs. Usually, species' distribution maps are made on a network composed of regularly distributed grid cell and zonation could be an issue as the grid can be directed in different directions (i.e. east-west or north-south). Nevertheless, in the case of articles 2 and 3 our grid cell resolution is small enough so that potential biases due to the zone effect is diluted and we are not concerned with this problem. Biodiversity knowledge is scale-dependent also in terms of grain size (i.e. the scale effect in MAUP), the extent of data coverage and also depends on biases contained in the data (Whittaker et al. 2005). At the largest possible grain size (the entire Earth), biases in citizen science occurrences totally disappear and we have perfect, but in most cases unhelpful, knowledge of the distribution of any species that has been described (e.g. an elephant has been seen in Africa therefore there are elephants on Earth). However, at smaller grain sizes, increasingly precise information about distribution is required to know about the true presence of the species over the landscape (Riddle et al. 2011, Pineda and Lobo 2012). At very small grain sizes, it becomes difficult to even confirm the presence/absence of a species, especially for highly mobile animals that range over wide areas and multiple habitat types (e.g. Boyce et al. 2002). For instance, one can imagine that if we study roe deer distribution in a grid cell of 100x100km^2 biases in citizen scientists observations of a roe deer (i.e. closer to roads than where the species is in reality) will be negligible whereas at smaller grain size biases might interfere with the information provided by the dataset (Kadmon et al. 2004).

The scale problem can also create a misleading view of species' diversity at smaller grain sizes. The sensitivity of diversity between species (i.e. beta diversity) to sampling grain reflects the effect of local environmental heterogeneity: As sampling grain increases, species communities appear more homogeneous (Mac Nally et al. 2004). This also applies for phenological acharacteristics of the species: at smaller spatial resolutions distributions can fluctuate to different degrees depending (for instance if the species is migratory, Thomson et al. 2007).

The concept of scale-dependent knowledge is illustrated in Figure 2.7 and 2.8. Compared to the accepted distribution of roe deer in Norway (article 2) we clearly see that at a coarse grain size (i.e. at county levels), the distribution of the species is very plausible although not detailed. At a county level, biases such as distance to trails or distance to roads do not influence this knowledge as there is road network in all counties in Norway. However, with finer grain sizes comes more uncertainty regarding roe deer distribution and map a) of Figure 2.7 only partially captures the distribution of roe deer in Norway. This is more obvious when map a) is compared to the accepted distribution map of roe deer presented in article 2.

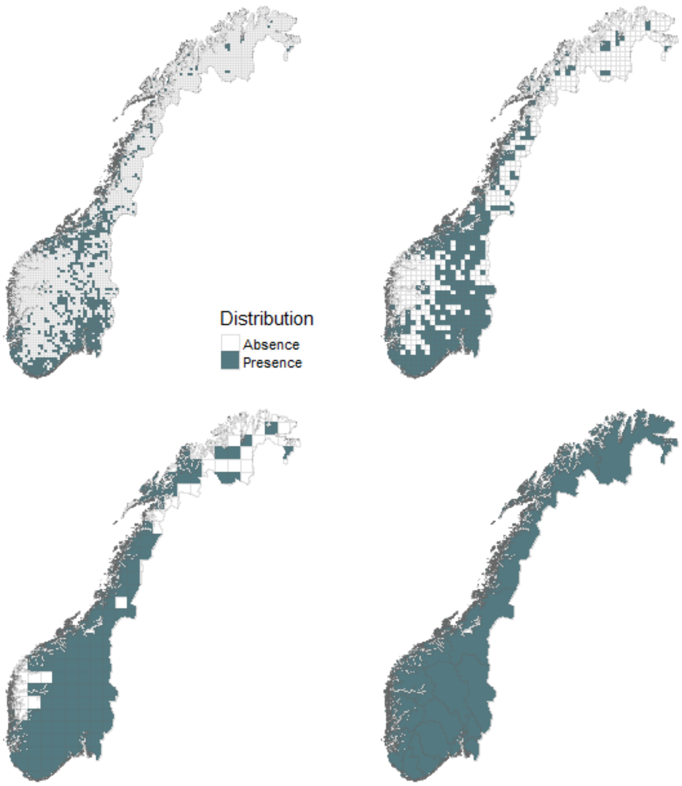


Figure 2.7: Roe deer distribution given by using citizen science observations from the Norwegian Species Observation Service at different grid size. a) uses a grid of 10x10km, b) 20x20km, c) 50x50km, and d) counties.

Due to their different generation properties, different dataset carry differents biases and the information brought by each dataset will differ as the grain size decreases. Biases gain more importance, and it becomes more obvious that each dataset carry different information with regard to the species' distribution. For instance, while using citizen science observations might indicate that roe deer sporadically appear near populated areas, it won't capture the fact that roe deer are also and primarily located in forested areas. On the contrary, while hunters' data might capture the affinity of roe deer for forested areas, it won't capture the sporadic presence of the species' near urban settlements. This fact implies that at fine-enough scales different dataset complement each other. Figure 2.8 shows this inter-relation between datasets and while ungulate-car collision datasets are by definition largely biased towards roads (map b), hunters' data are bound to hunting grounds (map c) which exclude urban areas, and parts of the species distribution where density is so low that they have not opened for hunter harvest. Following this rational we gathered a large range of dataset (see section 3.2) to create the species' distribution maps used in articles 2 and 3.

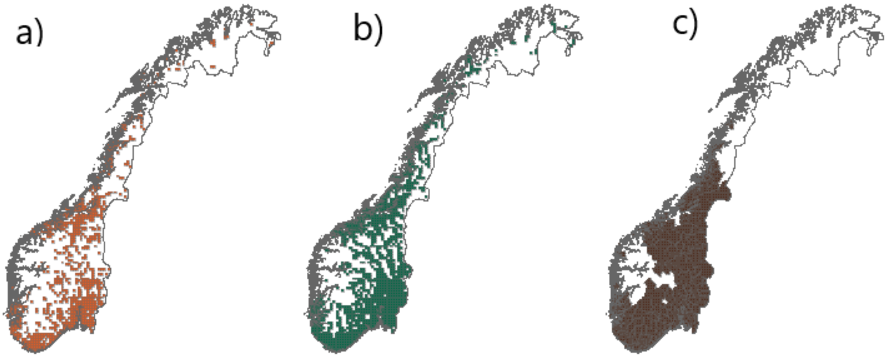


Figure 2.8: Roe deer distribution at a 10x10km resolution using a) citizen science observations, b) roe deer and car collision data, and c) hunting records.

Nevertheless, reducing the influence of biases in citizen science by combining this source of data with complementary datasets is only logical for questions about species at coarse scales (i.e. 1st order of selection, Figure 2.5, Johnson 1980). Studying biodiversity at finer scales such as their home range requires data that are more precise and important habitat-related information might not be captured at a 10x10km resolution. In fact, at fine grain, the number of thematic classes (e.g. agricultural area, urban area or forested area) can increase since small patches of rare land-use types will become distinguishable from dominant land-use types (Dunford and Freemark 2005). Using grain size that corresponds to known spatial characteristics of the species' habitat is likely to improve model accuracy (Gottschalk et al. 2011). Thus, in article 4 and 5 we used citizen science observations at the finest possible grain size namely the point location where an animal occurred, as this provides a fine enough resolution to study animals' habitat selection (Boyce et al. 2002) and we provide guidelines to use citizen science data at microscale.

Chapter 3

Statistical approach and considerations about the data material

In line with the quantitative tradition of the discipline of Geography and for the purpose if this thesis, I primarily used *quantitative methods* including Geographical Information Systems (GIS), descriptive statistics and spatial statistics using a Bayesian framework. Quantitative methods are powerful methods used to infer correlation or causation under certain properties (Pearl 2009) and allowed me to properly use citizen science observations at large scale and develop general methods for better use of citizen science data at a finer scale. Finally, using a combination of data visualization techniques such as maps and plots it was possible to display summary statistics and species' distribution maps in a clear and detailed way.

3.1 Statistical approach

3.1.1 Using the Bayesian framework

Even though it is not yet the dominant paradigm in statistics within Geography or Ecology, I chose to take a Bayesian instead of a frequentist stance. This choice has been made based on both ideological and pragmatic grounds.

In a nutshell, the major difference between the frequentist and Bayesian approach is how the unknown parameters are treated. Specifically, the frequentist's approach assumes that the observed data is from a specific likelihood model and the unknown parameters are “fixed and unknowable” (Carlin et al. 2009). In contrast, the Bayesian approach assumes that the unknown parameters follow prior distributions and uses these prior distributions to obtain the posterior distributions of unknown parameters (Eq 3.1).

$$P(\theta) = \frac{P(D \mid \theta) P(\theta)}{P(D)} \tag{3.1}$$

A usual way for frequentist to make inference is to do "statistical tests" of significance. Nevertheless, from the beginning of the PhD I began to see that the hypothesis testing framework wasn’t necessarily the best approach to science and that in research the typical hypothesis was an artificial construct (i.e. we accept that a subjective threshold $\alpha < 0.05$ indicates whether a result is worthy of attention or not). I came to see that hypothesis testing is a very binary way of thinking and while it can certainly have values in some circumstances (e.g. is the drug safe or not?), I found that estimation was even more important.

On a more pragmatic side Bayesian methods are more adapted to spatial models, ubiquitous in geography and ecology. Spatial modeling is essentially concerned with three issues: estimation and inference of parameter estimates (e.g. the environmental variables influencing the distribution of a given species), model specification and comparison (i.e. spatial models being hierarchical by nature), and prediction. It has been demonstrated that Bayesian spatial modeling can address these issues and has more attractive features in contrast to the conventional approach (i.e., frequentist) to spatial modeling (Banerjee et al. 2004). For example, the Bayesian spatial modeling approach offers a more solid foundation as the uncertainties and/or existing knowledge of unknown parameters are taken into account. In addition, the statistical inference of the posterior distribution of Bayesian spatial modeling is more intuitive and directly corresponds to the concept of probability.

A drawback in using Bayesian inference is that it is extremely intensive computationally to apply Bayes Theorem to complex models with large amount of data. It often takes days of computer time even on reasonably powerful personal computers to conduct an analysis. Fortunately, NTNU researchers Rue et al. (2009) have developed a novel computational method for Bayesian inference termed the *Integrated Nested Laplace Approximation (INLA)*. Due to the complexity of my models as well as the amount of distributional data I had to model I used INLA for inference in article 3, 4 and 5.

3.1.2 Simulation studies

To fully understand the process underlying the statistical methods developed in article 4 and 5, we first made use of simulations before using any empirical data. Simulation studies are computer experiments that involve creating data by pseudo-random sampling from known probability distributions. They are an invaluable tool for statistical research, particularly for the evaluation of new methods, the comparison of alternative methods, and are used to obtain empirical results about the performance of statistical methods in controlled scenarios (Morris et al. 2019). This is especially the case with methods aiming to improve the use of very noisy data such as citizen science data as it is often not possible to control for all variables influencing the data generation process as explained in section 2.1.

In articles 4 and 5 our simulation studies aim to show: (i) the implications of not accounting for variations on sampling effort when CS data is modeled, (ii) how accounting for this variation can contribute to improve the inference made about the point process underlying the spatial distribution of a species and (iii) see how mis-specification of the suggested method model can affect the quality of the inference.

3.2 Overview of the data material

3.2.1 Using unstructured citizen science

Citizen science projects differ in the way they are done. While most of them aim at collecting broad scale biodiversity data, their data falls along a continuum, from structured to unstructured, based on the objectives, survey design, flexibility, rigorousness, and detail collected about the observation process (Kelling et al. 2019). Projects with clear objectives, clearly planned data analysis, and rigorous protocols, for instance, are classified as structured projects. This is for instance the case of the Finnish Wildlife Triangle program in which participants are required to follow a certain route (i.e. a 4km sided triangle, article 1). In the middle of this continuum are projects which are flexible but give some indications about the effort of the absence of species. This is the case of iNaturalist and eBird which both provide a direct measure of effort and from which it is possible to obtain complete lists (i.e. presence *and* absence). On the last end of the continuum there are projects with open and flexible recruitment and with a general lack of protocols. They are classified as unstructured projects (Callaghan et al. 2019). This is the case of the Norwegian Species Observation Service (`www.artsobservasjoner.no`) which does not provide any measure of effort.

In articles 2 and 3 we used the Norwegian Species Observation Service data combined with other sources of data at broader scales to dilute the biases inherent to such unstructured citizen science dataset (as seen in section 2.2.2). However, in articles 4 and 5 we showed the potential of using only unstructured citizen science observations in an effort to legitimize this type of data, often neglected by researchers but also to make the methods we designed more accessible as there is usually more unstructured citizen science data available.

3.2.2 Choice of study organisms

The thesis exclusively takes the case of citizen science records of large mammal observation. This choice has been made on both pragmatic and emotional ground. First, my collaboration with researchers at the Norwegian Institute for Norwegian Research (NINA) has given me access to a wealth of information regarding the distribution of large mammals. This made comparisons between citizen science observations and professionally collected data such as GPS-telemetry possible (article 5). Moreover, due to the proximity between researchers at NINA and hunting authorities, collecting data on large mammal distributions was easier. Finally, being surrounded by mammal specialists helped me to make sense of the results I was obtaining.

Another important argument for the choice of large mammals was to increase the generalization of the work I was doing. Citizen scientists mainly record the location of bird species (Amano et al. 2016) and naturally, most of the methods for improving the use of citizen science has been developed with birds as case study (Follett and Strezov 2015). For instance, in the Norwegian Species Observation Service dataset used in articles 2, 3, 4 and 5, birds observations represent 82% of all observations (more than 20 millions records). Mammals being a charismatic taxon I was hoping that my work would grasp the attention of both specialists and generalists researchers working in biodiversity conservation thus increasing the visibility of the methods I was developing. The attention mammals provoke also made it easy to ground my work into well-known frameworks such as the wildlife management, human-wildlife conflict, and coexistence frameworks.

Finally, working with large mammals makes it easier to give a concrete aspect to my work as there are multiple management issues associated with these species. The presence of large mammals in our landscape provokes a wide variety of interactions that occur between humans and wild ungulates in the shared space of modern day European landscapes (article 2). Wild ungulates for instance serve as a prey for hunters, are in some cases a tourist attraction, and are considered ecosystem engineers. The presence of wild ungulates also has a human and economic cost as their presence can result in car accidents and damage to agriculture and forestry (article 2, Figure 3.1). Apart from the human-large mammal interactions, the presence of these species has also consequences on human-human relationships as there are many conflicts between stakeholders over the way that ungulates should be managed to reduce their impacts (Bredin et al. 2015). Both the human-ungulate and human-human interactions resulting from the presence of these species, as well as the current legal framework on encouraging human-wildlife coexistence, largely promote the collection of data on these species by hunters in particular (article 1, article 2). As a result, data on the distribution and habitat use of these species is of direct utility for wildlife managers in managing conflicts and promoting sustainable relationships, ensuring that my research will make a real-world contribution.

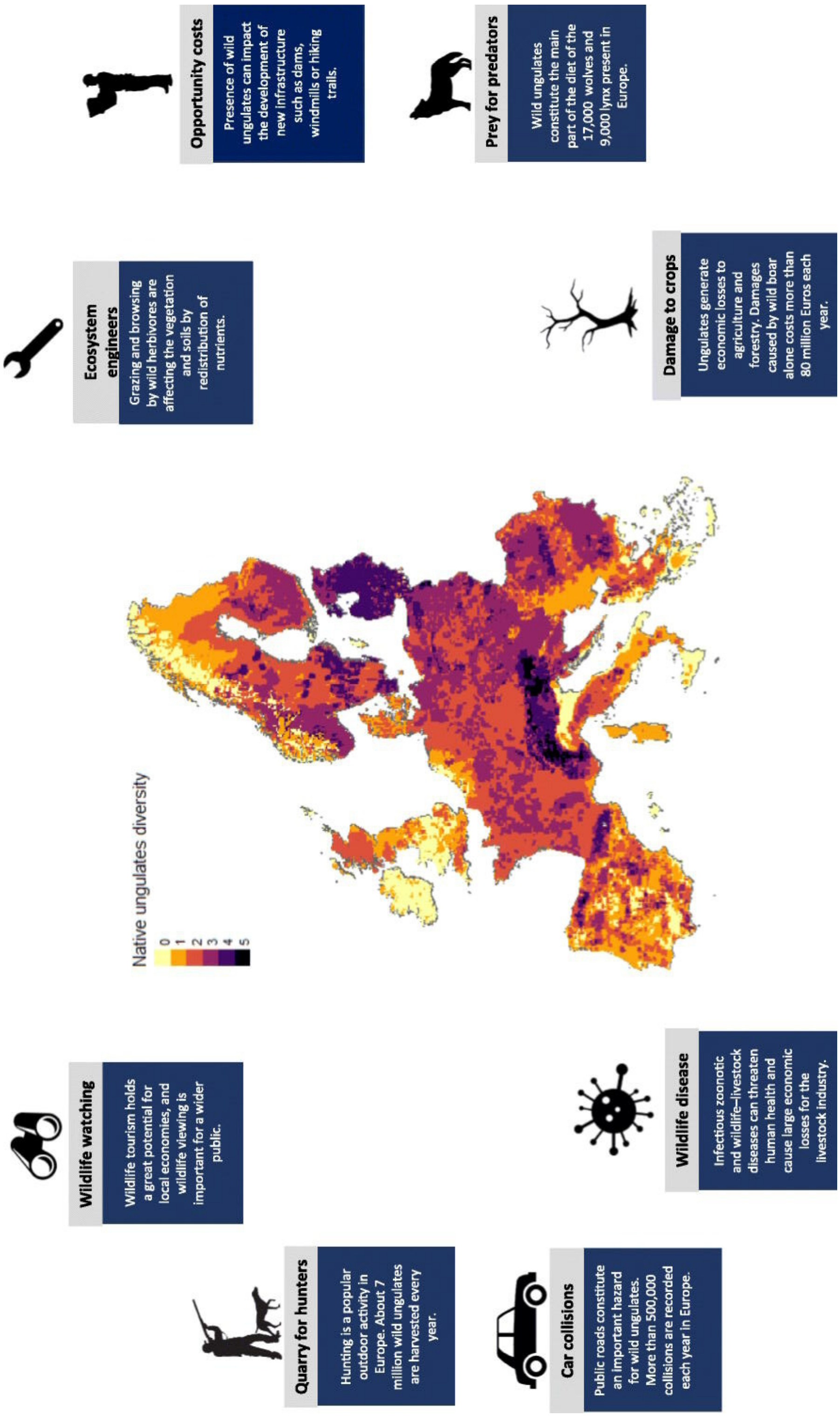


Figure 3.1: Diversity of interactions between human and ungulates in the European landscape. The figure is the graphical abstract of article 2.

3.2.3 Diversity of covariates used

Evaluating the use of citizen science data at different orders of selection implies using a large diversity of environmental covariates. The scale-dependence of biodiversity patterns has often been associated with different environmental factors playing a more significant role in shaping diversity at different scales (Belmaker and Jetz 2011). Factors like climate and energy are often recognized as main drivers at coarse continental and global scales, while other factors like disturbance, productivity, topography, land cover and habitat diversity seem to be more important at intermediate scales, and still other factors like biotic interactions and micro-environmental characteristics are more important at fine scales (Kallimanis et al. 2008, Kleijn et al. 2009).

In article 3 we used environmental and landscape related covariates (i.e. terrain ruggedness, summer and winter severity) that are thought to be influential drivers of species distribution at biogeographic scales (1st order of selection, Johnson 1980) while in articles 4 and 5 we used environmental covariates known to influence species habitat preference including forest coverage, slope and altitude (3rd order of selection, Johnson 1980). The same rational was made for human related variables and in articles 2 and 3 we used the Human Footprint Index (HWI) as a proxy for human influence over the landscape. The HFI is an aggregation of multiple variables related to human disturbance (e.g. the extent of built environment, cropland, pasture land, human population density, nighttime lights, railways, roads and navigable waterways, Venter et al. 2016). On the contrary, because it is important to understand the mechanistic influence of human variables on species habitat preference we used distance to roads and urban settlements in article 4 and 5. Finally in article 5, we availed of a novel dataset of fine-scaled recreationist movement (Strava) collected by citizens using digital applications on their smartphones and smartwatches.

3.2.4 Diversity of response variables used

Because the different articles had different aims I had to work with very different data material. While article 1 is a literature review on the contribution of hunters to biodiversity monitoring, maps used in articles 2 and 3 were made using a wide range of distribution data and article 5 used animal GPS-telemetry data. Nevertheless, citizen science data including hunters data and mammal records from the Norwegian Species Observation Service dataset were the primary source of information for all papers.

As mentioned in section 2.2.2 the importance of biases in citizen science data are reduced by both the grain size and the combination with other datasets containing other biases. For this reason, we collected a wide array of data sources that we thought complemented hunting and other citizen science data to create the maps used in articles 2 and 3. This included the three key volumes on ungulate management in Europe which contains ungulate distribution maps mainly drawn from hunting statistics data (Apollonio et al. 2010, Putman et al. 2011, Putman and Apollonio 2014), data from national mammal atlases, citizen science databases, vehicle collisions, scientific papers that provided the location of their study sites or which reviewed the status and distribution of various species and populations, and expert assessments.

One of the most common forms of evaluation implemented in citizen science projects is the comparison of data collected, or conservation projects completed, by volunteers to those done by professional researchers. In article 5 I used GPS-telemetry tracking data to evaluate the performance of citizen science.

3.3 Aiming at fully reproducible studies

In an era where it is becoming more and more important to do robust science I adhered and attempted to stick to the FAIR data principles throughout the PhD (Wilkinson et al. 2016). FAIR is an acronym which stands for Findability, Accessibility, Interoperability, and Reuse of digital assets and which refers to the guiding principles for scientific data management and stewardship.

- **F**indable means that enough metadata should be written so that the data can easily be found by both humans and computers.
- **A**ccessible implies that once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.
- **I**nteroperable. The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.
- **R**eusable. The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

The data were findable and accessible as I stored each used dataset on Open Science Framework for articles 1, 2, 3 and Zenodo for article 5. Moreover, I made sure of licensing the data used in article 2, 3 and 5 under Creative Common Attribution 4.0 International so that anyone can copy and redistribute the data as long as they give appropriate credit. Data used in article 5 were only partially shared (i.e. the shared dataset do not contain the raw data) as there were privacy concerns. We did not store data used in article 4 on an open science plateform as we made use of a GBIF dataset which is freely available and used the R package *rgbif*. Thus, access to the scripts should allow anyone to carry out a replicate of our analysis.

In tandem with data sharing, I made sure that the studies were entirely reproducible by writing a lengthy description of the data collection and processing in the articles' annexes and shared a well annotated version of the R scripts along with the data that has been used in the analysis. Naturally, adhering to the FAIR principles and the reproducible principles is a process and I believe it is possible to see improvements in the way data and code sharing has been done from article 1 to article 5. It is possible to find the link to the online repository containing data and scripts in Table 3.1.

article 1	`https://osf.io/gkazm/`
article 2	`https://osf.io/n5p2u/`
article 3	`https://osf.io/xv8nh/`
article 4	`https://github.com/jorgesicacha/VSE-paper`
article 5	`10.5281/zenodo.4590153`

Table 3.1: Links to the online repositories containing data and scripts used in each article.

In line with the FAIR principles and with an open source mindset I favored the use of free and open source software (i.e. software with source code that anyone can inspect, modify, and enhance) and primarily used R in combination with R Studio even though other softwares such as MatLab or Stata were available. This had multiple benefits including a large support community and software availability to all, ensuring reproducibility. GIS tasks were mainly performed in R but due to the inefficiency of the software to process large spatial data I made use of certain algorithms in ArcGIS Pro or if possible QGIS (Open source software).

Chapter 4

Summary of the articles

This thesis consists of five independent but inter-related articles. In article 1 I examine the tremendous importance of hunters for biodiversity monitoring due to the wide range of characteristics they report and their knowledge on the species they hunt. In articles 2 and 3 I use hunter data on species' distribution in combination with more generalist citizen science and other source of data to summarize knowledge about the distribution of large mammals in Europe, demonstrating that citizen science observations can be used at coarse scales (i.e. in the case of these two articles at continental scale) to provide robust results about species' conservation status, management and ecology. Finally, articles 4 and 5 show that if biases are accounted for by using appropriate statistical methods, citizen science observations can give a good approximation of species' ecology at finer scales.

4.1 Article 1: Hunters' contribution to biodiversity monitoring

Cretois, B., Linnell, J. D., Grainger, M., Nilsen, E. B., and Rød, J. K. (2020). Hunters as citizen scientists: Contributions to biodiversity monitoring in Europe. *Global Ecology and Conservation, 23*, e01077. `https://doi.org/10.1016/j.gecco.2020.e01077`

Monitoring biodiversity characteristics at large scales and with adequate resolution requires considerable effort and resources. Overall, there is clearly a huge scope for European hunters, a special and often overlooked group of citizen scientist, to contribute even more to biodiversity monitoring, especially because of their presence across the entire European landscape. In Cretois et al. (2020) we used the Essential Biodiversity Variables (EBVs) framework and we reviewed hunters' contributions to biodiversity monitoring. We reviewed the published and grey literature and contacted experts in wildlife management working with hunters. This gave us a comprehensive overview of hunters' contribution to biodiversity monitoring. We examined the methods used to collect data in hunter-based monitoring, the geographic and taxonomic extent of such contributions and the scientific output stemming from hunter-based monitoring data. We provide evidence that hunter-based monitoring is widely distributed across Europe and across taxa as we found records for 32 out of the 36 European countries included in our analysis involving hunters in the monitoring of at least one species group. Ungulates and small game species groups which have the widest hunter-based monitoring coverage. We found that it is possible to infer characteristics on Genetic composition, Species population, Species traits and Community composition with data that are being routinely collected by hunters in at least some countries. The main types of data provided are hunting bags data, biological samples including carcasses of shot animals and non-invasive sampling (e.g. scats) and observations for different counts and indices. Hunters collect data on biodiversity in its key dimensions. We conclude that collaborations between hunters and scientists are fruitful and should be considered a standard partnership for biodiversity conservation. Finally, we suggest that to overcome the challenges in the use of hunter data, more rigorous protocols for sampling data should be implemented and improvements made in data integration methods.

4.2 Article 2: The challenges and opportunities of living with wild ungulates

Linnell, J. D. * ,**Cretois, B.** * , Nilsen, E. B., Rolandsen, C. M., Solberg, E. J., Veiberg, V., Kaczensky, P., van Moorter, B., Panzacchi, M., Rauset. G., and Kaltenborn, B. (2020). The challenges and opportunities of coexisting with wild ungulates in the human-dominated landscapes of Europe's Anthropocene. *Biological Conservation, 244*, 108500. `https://doi.org/10.1016/j.biocon.2020.108500`

* *co-first authors*

The cumulative impact of human activities has driven many species into severe declines across the globe. However, the recent focus on conservation optimism has begun to highlight case studies that go against this trend. Reforestation, agricultural abandonment, reintroduction and legislative change have led to a situation where large mammals have recovered and are now widespread across the European continent. In Linnell et al. (2020) we summarize the knowledge about wild ungulate distribution in Europe and review the diversity of ways in which they interact with humans. Drawn from a wide range of sources, we compiled distribution maps of European wild ungulates at a resolution of 10x10km. These maps provide an updated version of the previous International Union for Conservation of Nature (IUCN) ones. Results show that 90% of Europe is home to at least 1 species of wild native ungulate, with roe deer and wild boar occupying 74% and 64% of Europe respectively. In contrast, wild native mountain ungulates only occupy 5% of Europe, and are often associated with protected areas. The wide distribution of most European ungulates combined with the extensive human activity within Europe result in a wide range of interactions between ungulates and humans. These interactions can be variously classified as services or disservices depending on the value orientation and economic position of the various stakeholders perceiving this relationship. Overall, Linnell et al. (2020) highlights the success of wildlife management policies in Europe and the potential for continental scale conservation of large mammals in human-dominated landscapes. However, maintaining the success of wild ungulate conservation requires actions from national and European institutions to improve coordinated management across jurisdictional borders and sectorial coordination for the whole landscape.

4.3 Article 3: Coexisting with large mammals in Europe

Cretois, B., Linnell, J. D., van Moorter, B., Kaczensky, P., Nilsen, E. B., Parada, J. S., and Rød, J. K. (2020). Coexistence of large mammals and humans is possible in Europe's anthropogenic landscapes. Submitted to *Anthropocene*.

A critical question in the conservation of both large carnivores and wild ungulates in the Anthropocene is to know the extent to which they can tolerate human disturbance. In Europe, large mammals have persisted for millennia, and recently expanded, alongside humans, but surprisingly little quantitative data is available about large scale effects of human disturbance on their distribution. In article 3, we quantify the relative importance of human land use and protected areas as opposed to biophysical constraints on large mammal distribution. We analyse recently compiled data on large mammals' distribution (data from article 2 for large ungulates and Chapron et al. 2014 for large carnivores) using Bayesian hierarchical models along with dominance analysis to quantify the relative effect of anthropogenic variables on species' distribution. We finally quantify the effect of anthropogenic variables on the size of the species' niche by simulating a scenario where we assumed no anthropogenic pressure on the landscape. We find that the broad scale distribution of most large mammals in Europe includes areas of high to very high human disturbance. Their distribution is primarily constrained by biophysical constraints rather than the human footprint or the presence of protected areas. Furthermore, our counterfactual scenario provides evidence that the human footprint and protected area coverage hardly influence the area of species' distribution. We suggest that coexistence between large mammals and humans is primarily limited by non-environmental aspects of human pressure such as the willingness of humans to share multi-use landscapes with wildlife rather than the ability of wildlife to tolerate humans. This finding offers grounds for optimism concerning wildlife conservation in the Anthropocene.

4.4 Article 4: Accounting for spatial varying sampling effort due to accessibility

Sicacha-Parada, J., Steinsland, I., **Cretois, B.**, and Borgelt, J. (2020). Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. *Spatial Statistics*, 100446. `https://doi.org/10.1016/j.spasta.2020.100446`

Unknown and varying sampling effort is a major issue when making inference based on citizen science data. In Sicacha-Parada et al. (2020) we propose a modeling approach for accounting for variation in sampling effort due to accessibility. The article first uses a simulation study to provide evidence that the suggested method is theoretically sound and then use a case study involving citizen science data of moose occurrence in the former Hedmark county, in southeastern Norway. In the case study, we aim to make proper inference about the importance of two geographical properties known to influence moose occurrence; terrain ruggedness index (i.e. rugged areas are difficultly accessible for moose) and solar radiation (i.e. a general proxy for snow cover duration and plant productivity). As expected, explanatory analysis shows that moose occurrences are over-represented close to roads. This led us to use distance to roads as a proxy for accessibility. We propose a model based on a Bayesian Log-Gaussian Cox Process specification for occurrence. The model accounts for accessibility through two functional forms, the half-normal function that assumes an exponential decay of the probability of accessing a location as the distance to the closest road increases and a semi-parametric approach that explains the decay of this probability as a function of a linear combination of I-spline basis functions. This approach can be seen as a thinning process where probability of thinning, i.e. not observing, increases with increasing distances. We conclude that the proposed model performs better than a model not accounting for accessibility and that there are misleading changes in the results if varying sampling effort due to accessibility is not accounted for.

4.5 Article 5: Identifying and correcting spatial bias in citizen science data

Cretois, B., van Moorter, B., Linnell, J. D., Simmonds, E. G., Rolandsen, C. M., Solberg, E. J., Strand. O., Gundersen, V., and Rød, J. K. Identifying and correcting spatial bias in citizen science data for wild ungulates in Norway. Submitted to Method in Ecology and Evolution

Many publications make use of opportunistic citizen science observation data, that is, observations of species presence collected without standardized field protocols and without explicit sampling designs. Most of these publications use citizen science to infer large-scale properties of species' distribution. However, other publications use citizen science data to study animal ecology at a habitat patch level without accounting for spatial biases in citizen science records or using methods that are difficult to generalize. In article 5, we use Resource Selection Functions to show that citizen science observations can be used to make inference about species' ecology at the microhabitat scale. Resource Selection Functions are a common way of characterizing species' habitat use and compares environmental covariates at locations visited by an animal to environmental covariates at a set of locations assumed available to the animal (availability locations). A usual way of conducting Resource Selection Function is to randomly sample availability locations across the landscape. However, rather than randomly sampling availability locations in the species home range, we suggest a method that selectively samples availability locations by accounting for biases in citizen science. We first inspect the biases in the citizen science dataset for three species of wild ungulate: wild reindeer, roe deer and moose in Norway. Then, we test the method on simulated data to evaluate whether the method is sound. After making sure the method works in ideal conditions we tested it on empirical dataset for the three wild ungulates (roe deer, moose and wild reindeer). We create an "observer model" estimating the extent of biases in citizen science observations and sampled availability locations with regard to this model. Then we compare how a model using citizen science observations with and without correction in availability locations perform as opposed to unbiased GPS-telemetry data to estimate species' Resource Selection Function. Finally, we draw and compare habitat suitability maps obtained using each of these models. We show that citizen science observations are more affected by human access and visibility variables than animal telemetry locations. This has consequences for drawing inferences about a species' ecology, as models using citizen science observations in tandem with random availability locations in habitat use studies can result in spurious inferences. However, sampling availability locations with regard to spatial biases in citizen science drastically improves the estimation of the species' Resource Selection Function and the species' habitat suitability maps. This analysis highlights the challenges and the opportunities of using citizen science observation in habitat-use scale studies. In particular we suggest the possibility of using an observer model developed from one species to correct for the observer biases of another species. While our method is not foolproof it is a first step towards unlocking the potential citizen science holds in finer scale studies.

Chapter 5

Concluding remarks and future research agenda

In the Oxford Language dictionary *to transform* means to "make a marked change in the form, nature or appearance of" something. In this thesis I have sought *to transform the way citizen science is used* or in other words to make a change in the nature of this use by better understanding and accounting for the biases in citizen science observations for better inference about species' ecology at different scales or order of selection (Johnson 1980) and therefore for better biodiversity conservation. Figure 5.1 summarizes how each article contributes to transforming the use of citizen science data.

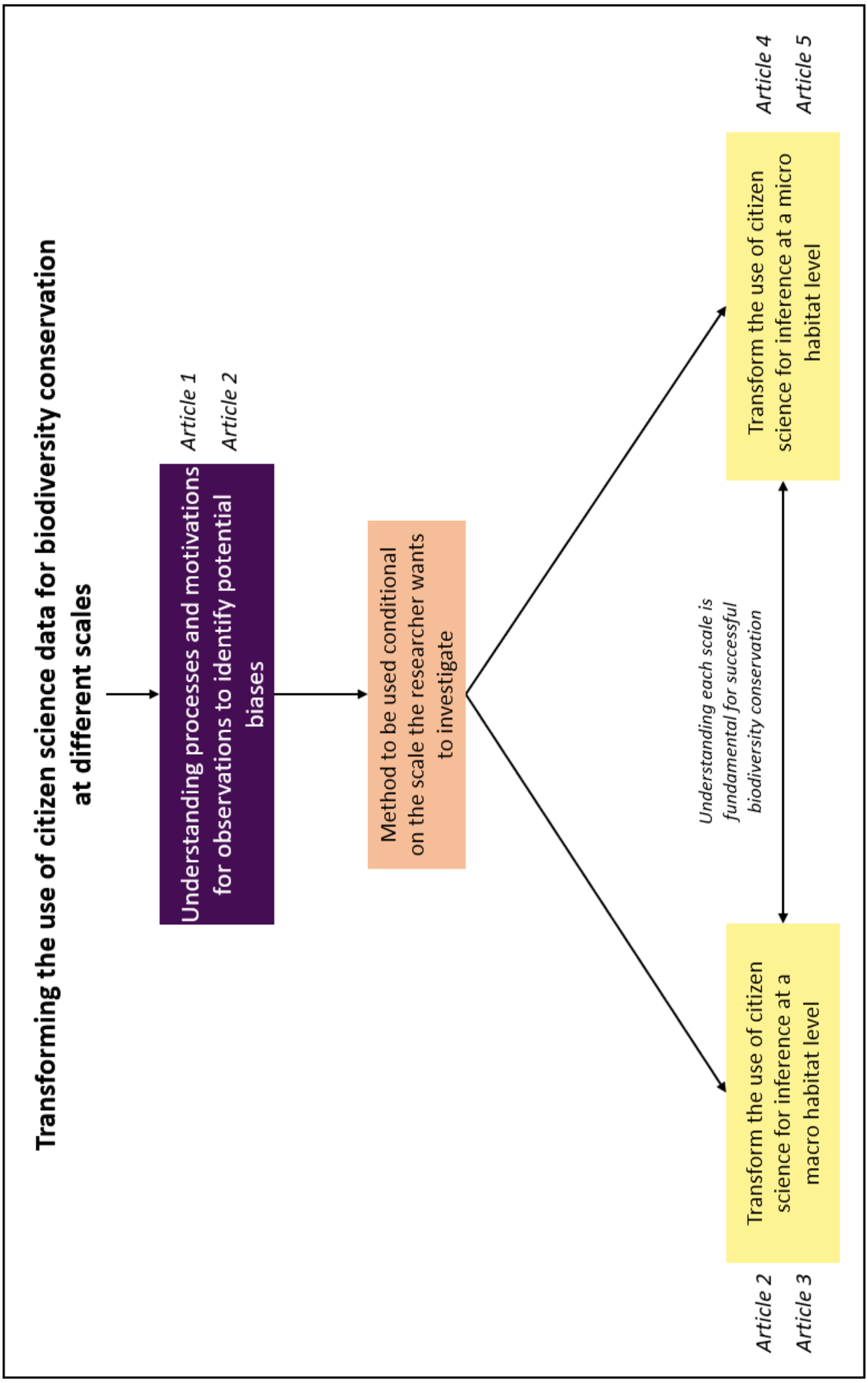


Figure 5.1: Contribution of each articles to transforming the use of citizen science data for biodiversity conservation.

To transform or change the use of citizen science data we first need to understand how the behavior of the citizen scientists generates and affects biases in citizen science data (Figure 5.1). In section 2.1 on the data collection process I went back to the root of the problem and have shown how the biases are generated and how this translates into biases in the geo-referenced data point itself. In article 1, I took this understanding further by better evaluating the contribution of hunters to biodiversity dataset and I have concluded that even though their contribution is of critical importance for wildlife researchers as they gather a wide range of Essential Biodiversity Variables for all game species as illustrated in Figure 1.2, the data they gather are not exempt from biases. However, these biases are different from other citizen scientists and this can induce several challenges regarding the use of hunter-based monitoring. As elaborated in article 1, institutionalizing the data can potentially lead to some extent of misreporting. Because each country has its own reporting system, misreporting can be facilitated or hindered depending on the system. While some systems are very systematic such as the Norwegian Moose Monitoring Program or the Finnish Wildlife Triangle, many other programs across Europe are much less structured (i.e. they do not follow a robust sampling scheme) and are based on hunter reports of total numbers of animals believed to occur on their hunting grounds, which are then aggregated at other administrative levels. Finally, obtaining biodiversity data from hunters from a certain area is conditional on whether the area of interest allows hunting. Hunters are thus a fundamental actor in successful human-wild ungulate coexistence as the data they collect is of direct interest for wildlife management structures (article 1). Moreover, harvesting wild ungulates also has a role in regulating human-ungulate interactions as decreasing ungulates population can reduce ungulate-vehicle collisions, reduce the damage to agricultural crops and control the spread of wildlife diseases (article 2). Nevertheless, the legitimacy of hunter-based monitoring is nowadays challenged as some members of the public view this practice as morally unacceptable (Gamborg and Jensen 2017).

Understanding the behaviors of both hunters and citizen scientists helped me realize the extent and the type of biases a researcher faces when using citizen science observations. While there are a broad array of biases as shown in section 2.1 and article 1, I chose to emphasize on the potential for transforming the use of citizen science data with regards to spatial biases in articles 2, 3, 4 and 5. In fact, spatial biases are more relevant in the case of species' spatial distribution, an Essential Biodiversity Variable often use in biodiversity conservation plans and fundamental to ensure successful human-wildlife coexistence (article 2). As mentioned in section 3.2.2 monitoring ungulate distribution has direct implications for spatial planning for wildlife management authorities. This is why hunters, the main provider of data on ungulates, are an important actor not only of wildlife monitoring but of operationalizing human-wildlife coexistence (article 1).

In section 2.2 of this thesis I suggest that spatial biases can be minimized by combining datasets which have different spatial biases such as hunting bag data (data representing huntable areas), citizen science observations (i.e. data close to areas of human activities), ungulate-vehicle collisions (i.e. located on the roads). Maps constructed in this way are accurate representations of species' distribution at a large enough grid size (10x10km in the case of articles 2 and 3) and can be of critical importance for biodiversity conservation planning and human wildlife coexistence as illustrated in articles 2 and 3. In fact, the results from articles 2 and 3 clearly demonstrate the ability of wild large mammals to occupy heavily human-dominated landscapes, which is an ecological prerequisite for coexistence. Nevertheless, these results also illustrate the complexity of the human-wildlife interactions that stem from co-occupation of shared space. Interactions between human and ungulates in the landscape may be positive and/or negative, or both, and depend on the economic interests and basic values of different stakeholders. The extent to which we can regard the co-occupation with wild ungulates as coexistence depends very much on how these interactions are managed. I highlight that because wild large mammals are very mobile and very tolerant of humans there is a need to coordinate management at continental, national and sub-national levels. This coordination should concern both the public and the private sector and concern a broad range of sectorial interests such as agriculture, transport, energy, veterinary and animal health sectors. In article 2 in particular I highlight that good maps are needed to ensure effective spatial planning of human activities and management of wild ungulate populations.

The maps of wild ungulate distribution on a continental scale are a first step in identifying and communicating the almost universal relevance of these issues across Europe. However, as expressed in article 3 and shown on Figure 5.1, successful human-wildlife coexistence requires much finer scale and locally adapted mapping of areas of distribution and movement and the method consisting of minimizing biases by pooling different dataset does not translate well at finer scale as only few dataset are available at such specific scale (e.g. telemetry data, professional surveys in certain areas). As stated in article 3, large scale studies (e.g. with a continental scope) can produce results that apparently contradict finer scale studies (e.g. with a sub-national scope) and failure to consider scale can lead to misinterpretation of results (Johnson 1980). For instance, while many fine scale studies find that the presence or habitat use of large mammals is mainly negatively affected by proximity to human infrastructure such as roads or cities (article 3) we found that large scale distribution of wild ungulates is primarily affected by biophysical factors such as terrain roughness, winter and summer severity and not human factors. Thus, in article 3 I suggest that conservation scientists should be careful about the scale used to answer their research questions and adapt their research methods to these questions (Figure 5.1).

As pooling datasets becomes less relevant to make proper inference at finer scale for biodiversity conservation, in articles 4 and 5 I contribute to the transformation of the use of citizen science by developing novel methodologies to make citizen science observations more relevant on their own (i.e. without relying on multiple source of data). In particular, in article 4 we show that it is possible to account for spatial biases by considering citizen science observations as a thinned spatial point pattern with an intensity λ degrading as it moves away from a road and in article 5 we show that specifying an observer model prior to fitting a habitat preference analysis (Resource Selection Function in the case of article 5) dramatically improves the inferences. Even though taking different statistical approaches (i.e. article 4 considering citizen science observations as a thinned point pattern and article 5 as presence / pseudo absence) results from both articles clearly demonstrate that *accessibility* is one of the main drivers of citizen science observation locations and needs to be accounted for in fine scale analysis. Nevertheless, both methodologies rely on assumptions which are not always met in other systems. In article 4 we assume that citizen science observations are only biased with regard to the distance to roads, however, as mentioned in previous sections other variables such as visibility and distance to urban settlements heavily affect the location of citizen science observations. We extended this idea in article 5 by specifying an observer model which is able to account for all known variables that are known to be drivers of citizen science observation locations. Nevertheless, the methodology presented in article 5 also relies on the assumption that citizen science observations are representative of the environmental space of the species (i.e. the n-dimensional space characterizing the biotic and abiotic conditions of a location). This may be the case for large mammals during certain period of the year as discussed in article 5 (i.e. when the species is not migrating) but more research is needed to determine if this is also the case for other taxon. The concept of mismatch between the citizen science and species environmental space is represented in Figure 5.2. In this conceptual figure, citizen scientists and the target species do not use the landscape in the same way and citizen science observations only partially capture the 2nd and 3rd orders of selection (Johnson 1980). In contrast, telemetry observations are in theory able to capture the 3rd order of selection. In this hypothetical example, citizen scientists would use steeper slope and trails that are heavily used compared to the species that prefer less steep landscapes that contains trails that are moderately used. Thus, the distribution of the citizen scientists and the species only partially overlap. This implies that citizen science only partially captures the species' behavior at different scales and therefore only partially the Resource

Selection Function of the species.

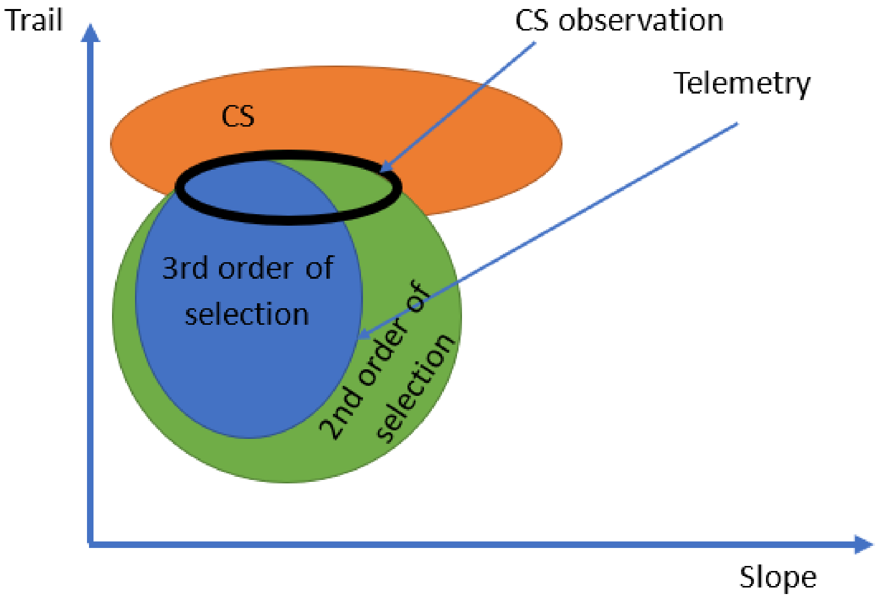


Figure 5.2: Conceptual figure representing the reasoning underlying the use of citizen science observations to infer species habitat preference along two potential environmental gradients. The thick line represents the area where citizen science observations correctly identify species ecological properties, including the RSF. Extracted from article 5.

This thesis demonstrates that citizen science data can be used at multiple scales if biases are accounted for using proper methodology. This thesis has aimed to *transform the use of citizen science for biodiversity conservation* by making researchers using such type of data aware of its drawbacks and introducing the reader to novel methodologies which account for such biases. A main conclusion of the thesis is that the methodologies used to account for biases in citizen science data have to be scaled to the research question and to the data available: it is impossible to infer species' fine scale movement within home ranges (3rd order of selection, sensu Johnson 1980) with presence / absence data on a grid of 20x20km. It is important to note that the work done during this thesis has been carried out on wild large mammals in Europe which include countries with a rich tradition of citizen science and where it is relatively easy to obtain high quality data on large mammal distribution allowing us to gain new insights into citizen science. The methodologies presented in this thesis might require changes in places where data are scarce or of poor quality and it would be important to pursue how we can complement the lack of information with modeling approaches.

In conclusion, the future of citizen science remains very promising and research to understand and make better use of it is of great interest for biodiversity researchers. As technology provides so much support to citizen science and as researchers continue to develop methods to make good use of their data, so too will the capacity of citizen science to contribute to science (Thornhill et al. 2021).

Bibliography

Altwegg, R. and Nichols, J. D. (2019). Occupancy models for citizen-science data. *Methods in Ecology and Evolution*, 10(1):8–21.

Amano, T., Lamming, J. D. L., and Sutherland, W. J. (2016). Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. *BioScience*, 66(5):393–400.

Apollonio, M., Andersen, R., and Putman, R., editors (2010). *European ungulates and their management in the 21st century*. Cambridge University Press, Cambridge, UK ; New York. OCLC: ocn401147378.

Armitage, D., de Loë, R. C., Morris, M., Edwards, T. W. D., Gerlak, A. K., Hall, R. I., Huitema, D., Ison, R., Livingstone, D., MacDonald, G., Mirumachi, N., Plummer, R., and Wolfe, B. B. (2015). Science–policy processes for transboundary water governance. *Ambio*, 44(5):353–366.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. Number 101 in Monographs on statistics and applied probability. Chapman & Hall/CRC, Boca Raton, Fla.

Belmaker, J. and Jetz, W. (2011). Cross-scale variation in species richness-environment associations: Richness-environment scaling. *Global Ecology and Biogeography*, 20(3):464–474.

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., and Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173:144–154.

Block, W. and Brennan, L. (1993). The habitat concept in ornithology. *Current Ornithology*, pages 35–91.

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., and Mace, G. M. (2010). Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biology*, 8(6):e1000385.

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., and Parrish, J. K. (2014). Next Steps for Citizen Science. *Science*, 343(6178):1436–1437.

Boyce, M. S., Vernier, P. R., Nielsen, S. E., and Schmiegelow, F. K. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157(2-3):281–300.

Bredin, Y. K., Lindhjem, H., van Dijk, J., and Linnell, J. D. (2015). Mapping value plurality towards ecosystem services in the case of Norwegian wildlife management: A Q analysis. *Ecological Economics*, 118:198–206.

Brenna, B. (2011). Clergymen Abiding in the Fields: The Making of the Naturalist Observer in Eighteenth-Century Norwegian Natural History. *Science in Context*, 24(2):143–166.

Bruyere, B. and Rappe, S. (2007). Identifying the motivations of environmental volunteers. *Journal of Environmental Planning and Management*, 50(4):503–516.

Bubalo, M., van Zanten, B. T., and Verburg, P. H. (2019). Crowdsourcing geo-information on landscape perceptions and preferences: A review. *Landscape and Urban Planning*, 184:101–111.

Burgess, J., Harrison, C. M., and Filius, P. (1998). Environmental Communication and the Cultural Politics of Environmental Citizenship. *Environment and Planning A: Economy and Space*, 30(8):1445–1460.

Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B., and Major, R. E. (2019). Improving big citizen science data: Moving beyond haphazard sampling. *PLOS Biology*, 17(6):e3000357.

Carlin, B. P., Louis, T. A., and Carlin, B. P. (2009). *Bayesian methods for data analysis*. Chapman & Hall/CRC texts in statistical science series. CRC Press, Boca Raton, 3rd ed edition. OCLC: ocn227205756.

Centre for Educational Research and Innovation and OECD, editors (1972). *Interdisciplinarity: problems of teaching and research in universities ; this report is* *based of a Seminar on Interdisciplinarity in Universities which was organised by CERI (Centre for Educational Research and Innovation) in coll. with the French Ministry of Education at the Univ. of Nice, Sept. 7.-12. 1970.* OECD, Paris. OCLC: 1207595.

Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., and Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213:280–294.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., and players, F. (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760.

Courter, J. R., Johnson, R. J., Stuyck, C. M., Lang, B. A., and Kaiser, E. W. (2013). Weekend bias in Citizen Science data reporting: implications for phenology studies. *International Journal of Biometeorology*, 57(5):715–720.

Cretois, B., Linnell, J. D., Grainger, M., Nilsen, E. B., and Rød, J. K. (2020). Hunters as citizen scientists: Contributions to biodiversity monitoring in Europe. *Global Ecology and Conservation*, 23:e01077.

Cretois, B., Linnell, J. D. C., Kaltenborn, B. P., and Trouwborst, A. (2019). What form of human-wildlife coexistence is mandated by legislation? A comparative analysis of international and national instruments. *Biodiversity and Conservation*, 28(7):1729–1741.

Da Re, D., Gilbert, M., Chaiban, C., Bourguignon, P., Thanapongtharm, W., Robinson, T. P., and Vanwambeke, S. O. (2020). Downscaling livestock census data using multivariate predictive models: Sensitivity to modifiable areal unit problem. *PLOS ONE*, 15(1):e0221070.

Dabiri, Z. and Blaschke, T. (2019). Scale matters: a survey of the concepts of scale used in spatial disciplines. *European Journal of Remote Sensing*, 52(1):419–434.

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R. R., Shin, Y.-J., Visseren-Hamakers, I., Willis, K. J., and Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, 366(6471):eaax3100.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., and Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, 10(6):291–297.

Dunford, W. and Freemark, K. (2005). Matrix Matters: Effects of Surrounding Land Uses on Forest Birds Near Ottawa, Canada. *Landscape Ecology*, 20(5):497–511.

eBird (2019). ebird 2019 year in review. *https://ebird.org/news/ebird-2019-year-in-review*.

Ellwood, E. R., Crimmins, T. M., and Miller-Rushing, A. J. (2017). Citizen science and conservation: Recommendations for a rapidly moving field. *Biological Conservation*, 208:1–4.

Follett, R. and Strezov, V. (2015). An Analysis of Citizen Science Based Research: Usage and Publication Patterns. *PLOS ONE*, 10(11):e0143687.

Froehly, J. L., Beane, N. R., Evans, D. E., Cagle, K. E., and Jachowski, D. S. (2020). Using multi-scale behavioral investigations to inform wild pig (Sus scrofa) population management. *PLOS ONE*, 15(2):e0228705.

Gamborg, C. and Jensen, F. S. (2017). Attitudes towards recreational hunting: A quantitative survey of the general public in Denmark. *Journal of Outdoor Recreation and Tourism*, 17:20–28.

Ganzevoort, W., van den Born, R. J. G., Halffman, W., and Turnhout, S. (2017). Sharing biodiversity data: citizen scientists' concerns and motivations. *Biodiversity and Conservation*, 26(12):2821–2837.

Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., and Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, 22(11):1139–1149.

Goodchild, M. (2007). Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International journal of spatial data infrastructures research*, 2(2):24–32.

Goodchild, M. F. and Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1:110–120.

Gottschalk, T. K., Aue, B., Hotes, S., and Ekschmitt, K. (2011). Influence of grain size on species–habitat models. *Ecological Modelling*, 222(18):3403–3412.

Groom, Q., Weatherdon, L., and Geijzendorffer, I. R. (2017). Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology*, 54(2):612–617.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Rooyen, A. R., Weeks, A. R., and Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, 8(9):1081–1091.

Haining, R. P. (2003). *Spatial data analysis theory and practice*. Cambridge University Press, Cambridge, UK; New York. OCLC: 1162465968.

Haywood, B. K., Parrish, J. K., and Dolliver, J. (2016). Place-based and data-rich citizen science as a precursor for conservation action: Citizen Science and Conservation Action. *Conservation Biology*, 30(3):476–486.

Heim, N., Fisher, J. T., Volpe, J., Clevenger, A. P., and Paczkowski, J. (2019). Carnivore community response to anthropogenic landscape change: species-specificity foils generalizations. *Landscape Ecology*, 34(11):2493–2507.

Hill, M. O. (2012). Local frequency as a key to interpreting species occurrence data when recording effort is not known: *A key to interpreting species occurrence data*. *Methods in Ecology and Evolution*, 3(1):195–205.

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W.-K., and Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution*, 27(2):130–137.

Hutchinson, G. E. (1991). Population studies: Animal ecology and demography. *Bulletin of Mathematical Biology*, 53(1-2):193–213.

Isaac, N. J. B. and Pocock, M. J. O. (2015). Bias and information in biological records: Bias and information in biological records. *Biological Journal of the Linnean Society*, 115(3):522–531.

Isaac, N. J. B., Strien, A. J., August, T. A., Zeeuw, M. P., and Roy, D. B. (2014). Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5(10):1052–1060.

Johnson, D. H. (1980). The Comparison of Usage and Availability Measurements for Evaluating Resource Preference. *Ecology*, 61(1):65–71.

Johnston, A., Fink, D., Hochachka, W. M., and Kelling, S. (2018). Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution*, 9(1):88–97.

Juillet, C., Choquet, R., Gauthier, G., Lefebvre, J., and Pradel, R. (2012). Carry-over effects of spring hunt and climate on recruitment to the natal colony in a migratory species. *Journal of Applied Ecology*, 49(6):1237–1246.

Kadmon, R., Farber, O., and Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, 14(2):401–413.

Kallimanis, A., Tsiafouli, M., Pantis, J., Mazaris, A., Matsinos, Y., and Sgardelis, S. (2008). Arable land and habitat diversity in Natura 2000 sites in Greece. *Journal of Biological Research*, 9:55–66.

Kelling, S., Johnston, A., Fink, D., Ruiz-Gutierrez, V., and Bonney, R. (2019). Using semistructured surveys to improve citizen science data for monitoring biodiversity. *BioScience*, 69:170–179.

Kinchy, A. (2017). Citizen Science and Democracy: Participatory Water Monitoring in the Marcellus Shale Fracking Boom. *Science as Culture*, 26(1):88–110.

Kissling, W. D., Hardisty, A., García, E. A., Santamaria, M., De Leo, F., Pesole, G., Freyhof, J., Manset, D., Wissel, S., Konijn, J., and Los, W. (2015). Towards global interoperability for supporting biodiversity research on essential biodiversity variables (EBVs). *Biodiversity*, 16(2-3):99–107.

Kleijn, D., Kohler, F., Báldi, A., Batáry, P., Concepción, E., Clough, Y., Díaz, M., Gabriel, D., Holzschuh, A., Knop, E., Kovács, A., Marshall, E., Tscharntke, T., and Verhulst, J. (2009). On the relationship between farmland biodiversity and land-use intensity in Europe. *Proceedings of the Royal Society B: Biological Sciences*, 276(1658):903–909.

Kosmala, M., Wiggins, A., Swanson, A., and Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10):551–560.

Krausman, P. (1999). Some basic principles of habitat use. *Grazing behavior of livestock and wildlife*, 70:85–90.

Kéry, M., Dorazio, R. M., Soldaat, L., Van Strien, A., Zuiderwijk, A., and Royle, J. A. (2009). Trend estimation in populations with imperfect detection: *Trend estimation with imperfect detection*. *Journal of Applied Ecology*, 46(6):1163–1172.

Lawrence, A. and Turnhout, E. (2010). Personal meaning in the public sphere: The standardisation and rationalisation of biodiversity data in the UK and the Netherlands. *Journal of Rural Studies*, 26(4):353–360.

Lemos, M. C., Arnott, J. C., Ardoin, N. M., Baja, K., Bednarek, A. T., Dewulf, A., Fieseler, C., Goodrich, K. A., Jagannathan, K., Klenk, N., Mach, K. J., Meadow, A. M., Meyer, R., Moss, R., Nichols, L., Sjostrom, K. D., Stults, M., Turnhout, E., Vaughan, C., Wong-Parodi, G., and Wyborn, C. (2018). To co-produce or not to co-produce. *Nature Sustainability*, 1(12):722–724.

Linnell, J. D., Cretois, B., Nilsen, E. B., Rolandsen, C. M., Solberg, E. J., Veiberg, V., Kaczensky, P., Van Moorter, B., Panzacchi, M., Rauset, G. R., and Kaltenborn, B. (2020). The challenges and opportunities of coexisting with wild ungulates in the human-dominated landscapes of Europe's Anthropocene. *Biological Conservation*, 244:108500.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. (2008). Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189.

Lowman, M., D'Avanzo, C., and Brewer, C. (2009). ECOLOGY: A National Ecological Network for Research and Education. *Science*, 323(5918):1172–1173.

Mac Nally, R., Fleishman, E., Bulluck, L., and Betrus, C. (2004). Comparative influence of spatial scale on beta diversity within regional assemblages of birds and butterflies. *Journal of biogeography*, 31(6):917–929.

MacKenzie, D. I. (2018). *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. OCLC: 1011094347.

Mair, L. and Ruete, A. (2016). Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. *PLOS ONE*, 11(1):e0147796.

Maisonneuve, N., Stevens, M., and Ochab, B. (2010). Participatory noise pollution monitoring using mobile phones. *Information Polity*, 15(1,2):51–71.

Marceau, D. (1999). The Scale Issue in the Social and Natural Sciences. *Canadian Journal of Remote Sensing*, 25(4):347–356.

McKinley, D., Miller-Rushing, A., Ballard, H., Bonney, R., Brown, H., Cook-Patton, S., and Soukup, M. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, 208:15–28.

Miller-Rushing, A., Primack, R., and Bonney, R. (2012). The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, 10(6):285–290.

Morris, T. P., White, I. R., and Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., and Crowston, K. (2012). The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6):298–304.

Niemiec, R. M., Ardoin, N. M., Wharton, C. B., and Brewer, F. K. (2017). Civic and natural place attachment as correlates of resident invasive species control behavior in Hawaii. *Biological Conservation*, 209:415–422.

O'Neill, R. V., Johnson, A. R., and King, A. W. (1989). A hierarchical framework for the analysis of scale. *Landscape Ecology*, 3(3-4):193–205.

Openshaw, S. (1981). The modifiable areal unit problem. *Quantitative geography: A British view*, pages 60–69.

Owens, S. (2000). 'Engaging the public': information and deliberation in environmental policy. *Environment and Planning*, 32(7):1141–1148.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Hoft, R., Hurtt, G., Jetz, W., Karp, D. S., McGeoch, M. A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J. P. W., Stuart, S. N., Turak, E., Walpole, M., and Wegmann, M. (2013). Essential Biodiversity Variables. *Science*, 339(6117):277–278.

Pineda, E. and Lobo, J. M. (2012). The performance of range maps and species distribution models representing the geographic variation of species richness at different resolutions: Performance of range maps and species distribution models. *Global Ecology and Biogeography*, 21(9):935–944.

Popescu, V. D., Artelle, K. A., Pop, M. I., Manolache, S., and Rozylowicz, L. (2016). Assessing biological realism of wildlife population estimates in data-poor systems. *Journal of Applied Ecology*, 53(4):1248–1259.

Prendergast, J. R., Wood, S. N., Lawton, J. H., and Eversham, B. C. (1993). Correcting for Variation in Recording Effort in Analyses of Diversity Hotspots. *Biodiversity Letters*, 1(2):39.

Price, T. D. and Kirkpatrick, M. (2009). Evolutionarily stable range limits set by interspecific competition. *Proceedings of the Royal Society B: Biological Sciences*, 276(1661):1429–1434.

Putman, R. and Apollonio, M., editors (2014). *Behaviour and management of European ungulates*. Whittles Publishing, Dunbeath, Scotland. OCLC: ocn883426047.

Putman, R., Apollonio, M., and Andersen, R., editors (2011). *Ungulate Management in Europe: Problems and Practices*. Cambridge University Press, Cambridge.

Razgour, O., Rebelo, H., Di Febbraro, M., and Russo, D. (2016). Painting maps with bats: species distribution modelling in bat research and conservation. *Hystrix*, 27(1):1–8.

Regeer, B. and Bunders, J. (2009). Knowledge co-creation: Interaction between science and society. A Transdisciplinary Approach to Complex Societal Issues. Technical report, Den Haag: Advisory Council for Research on Spatial Planning, Nature and the Environment/Consultative Committee of Sector Councils in the Netherlands [RMNO/COS].

Riddle, B., Laddle, R., Lourie, S., and Whittaker, R. (2011). Basic biogeography: estimating biodiversity and mapping nature. *Conservation biogeography*, pages 45–92.

Rowley, J. and Slack, F. (2004). Conducting a literature review. *Management Research News*, 27(6):31–39.

Roy, H. E., Adriaens, T., Isaac, N. J. B., Kenis, M., Onkelinx, T., Martin, G. S., Brown, P. M. J., Hautier, L., Poland, R., Roy, D. B., Comont, R., Eschen, R., Frost, R., Zindel, R., Van Vlaenderen, J., Nedvěd, O., Ravn, H. P., Grégoire, J.-C., de Biseau, J.-C., and Maes, D. (2012). Invasive alien predator causes rapid declines of native European ladybirds: Alien predator causes declines of native ladybirds. *Diversity and Distributions*, 18(7):717–725.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, 84(3):777–790.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Scott, L. (2020). We've reached 1,000,000 observers! *www.inaturalist.org*.

Setten, G. (2004). The habitus, the rule and the moral landscape. *cultural geographies*, 11(4):389–415.

Sheth, A. (2009). Citizen Sensing, Social Signals, and Enriching Human Experience. *IEEE Internet Computing*, 13(4):87–92.

Sicacha-Parada, J., Steinsland, I., Cretois, B., and Borgelt, J. (2020). Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. *Spatial Statistics*, page 100446.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9):467–471.

Soleri, D., Long, J., Ramirez-Andreotta, M., Eitemiller, R., and Pandya, R. (2016). Finding pathways to more equitable and meaningful public-scientist partnerships. *Citizen Science Theory and Practice*, 1(9):1.

Sullivan, B. L., Phillips, T., Dayer, A. A., Wood, C. L., Farnsworth, A., Iliff, M. J., Davies, I. J., Wiggins, A., Fink, D., Hochachka, W. M., Rodewald, A. D., Rosenberg, K. V., Bonney, R., and Kelling, S. (2017). Using open access observational data for conservation action: A case study for birds. *Biological Conservation*, 208:5–14.

Szabo, J. K., Vesk, P. A., Baxter, P. W. J., and Possingham, H. P. (2010). Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. *Ecological Applications*, 20(8):2157–2169.

Telfer, M. G., Preston, C., and Rothery, P. (2002). A general method for measuring relative change in range size from biological atlas data. *Biological Conservation*, 107(1):99–109.

Thomas, J. A. (2004). Comparative Losses of British Butterflies, Birds, and Plants and the Global Extinction Crisis. *Science*, 303(5665):1879–1881.

Thomson, J. R., Mac Nally, R., Fleishman, E., and Horrocks, G. (2007). Predicting Bird Species Distributions in Reconstructed Landscapes. *Conservation Biology*, 21(3):752–766.

Thornhill, I., Cornelissen, J. H. C., McPherson, J. M., MacBride-Stewart, S., Mohamad, Z., White, H. J., and Wiersma, Y. F. (2021). Towards ecological science for all by all. *Journal of Applied Ecology*, 58(2):206–213.

Tiago, P., Ceia-Hasse, A., Marques, T., Capinha, C., and Pereira, H. (2017). Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Scientific Reports*, 7(1):1–9.

Tingley, M. W. and Beissinger, S. R. (2009). Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends in Ecology & Evolution*, 24(11):625–633.

Tulloch, A. I. T., Mustin, K., Possingham, H. P., Szabo, J. K., and Wilson, K. A. (2013). To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale. *Diversity and Distributions*, 19(4):465–480.

UN General Assembly (1992). UN General Assembly. The Convention on Biological Diversity of 5 June 1992.

UN General Assembly (2015). UN General Assembly. Transforming Our World : the 2030 Agenda for Sustainable Development.

van Strien, A., van Swaay, C., and Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50(6):1450–1458.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., and Watson, J. E. M. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, 7(1):12558.

Venturelli, P. A., Hyder, K., and Skov, C. (2017). Angler apps as a source of recreational fisheries data: opportunities, challenges and proposed standards. *Fish and Fisheries*, 18(3):578–595.

Vetter, J. (2011). Introduction: Lay Participation in the History of Scientific Observation. *Science in Context*, 24(2):127–141.

Walker, J. (2019). Will Citizen Science take over? *https://data-blog.gbif.org/post/gbif-citizen-science-data/*.

Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., and Willis, K. J. (2005). Conservation Biogeography: assessment and prospect: Conservation Biogeography. *Diversity and Distributions*, 11(1):3–23.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018.

Wood, C., Sullivan, B., Iliff, M., Fink, D., and Kelling, S. (2011). eBird: Engaging Birders in Science and Conservation. *PLoS Biology*, 9(12):e1001220.

Yuan, Y., Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., Buckland, S. T., Rue, H., and Gerrodette, T. (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics*, 11(4):2270–2297.

Zhao, Y. and Han, Q. (2016). Spatial crowdsourcing: current state and future directions. *IEEE communications magazine*, 54(7):102–107.

Part II

Articles

Article 1

Global Ecology and Conservation 23 (2020) e01077

Contents lists available at ScienceDirect

Global Ecology and Conservation

journal homepage: http://www.elsevier.com/locate/gecco

Original Research Article

Hunters as citizen scientists: Contributions to biodiversity monitoring in Europe

Benjamin Cretois [a], *, John D.C. Linnell [b], Matthew Grainger [b], Erlend B. Nilsen [b], Jan Ketil Rød [a]

[a] *Department of Geography, Norwegian University of Science and Technology, 7491, Trondheim, Norway*
[b] *Norwegian Institute for Nature Research, PO Box 5685, Torgard, 7485, Trondheim, Norway*

ARTICLE INFO

Article history:
Received 23 March 2020
Received in revised form 23 April 2020
Accepted 23 April 2020

Keywords:
Citizen science
Hunter
Biodiversity monitoring
Europe
Essential biodiversity variables (EBV)
Biodiversity conservation
Mammal
Birds
Game species

ABSTRACT

Monitoring biodiversity characteristics at large scales and with adequate resolution requires considerable effort and resources. Overall, there is clearly a huge scope for European hunters, a special and often overlooked group of citizen scientist, to contribute even more to biodiversity monitoring, especially because of their presence across the entire European landscape.

Using the Essential Biodiversity Variables (EBVs) framework we reviewed the published and grey literature and contacted experts to provide a comprehensive overview of hunters' contributions to biodiversity monitoring. We examined the methods used to collect data in hunter-based monitoring, the geographic and taxonomic extent of such contributions and the scientific output stemming from hunter-based monitoring data.

Our study suggests that hunter-based monitoring is widely distributed across Europe and across taxa as 32 out of the 36 European countries included in our analysis involve hunters in the monitoring of at least one species group with ungulates and small game species groups which have the widest hunter-based monitoring coverage. We found that it is possible to infer characteristics on Genetic composition, Species population, Species traits and Community composition with data that are being routinely collected by hunters in at least some countries. The main types of data provided are hunting bags data, biological samples including carcasses of shot animals and non-invasive samplings and Observations for counts and indices.

Hunters collect data on biodiversity in its key dimensions. Collaborations between hunters and scientists are fruitful and should be considered a standard partnership for biodiversity conservation. To overcome the challenges in the use of hunters' data, more rigorous protocols for sampling data should be implemented and improvements made in data integration methods.

1. Introduction

Global biodiversity is undergoing severe declines (Díaz et al., 2019). This situation has led the international community to take action to alter this trend by setting policy frameworks and objectives. For example, the Aichi Biodiversity Targets set by

---

* Corresponding author.
E-mail address: benjamin.cretois@ntnu.no (B. Cretois).

https://doi.org/10.1016/j.gecco.2020.e01077

the Convention on Biological Diversity (UN General Assembly, 1992) and the United Nations Sustainable Development Goals (UN General Assembly, 2015) are globally accepted frameworks which set targets for progress toward a more sustainable world. Biodiversity monitoring is an essential component of measuring progress towards these goals. However, monitoring biodiversity at large scales and with adequate resolution requires considerable effort and resources and represents a logistical challenge for researchers. This is one driver behind the recent enthusiasm about involving of the public in the data collection process (Silvertown, 2009).

Citizen science, here defined as the involvement of citizens in scientific research and knowledge production, has repeatedly demonstrated its ability to gather massive amounts of data at a spatial scale unattainable by research teams and biodiversity management structures such as state authorities and other societal groups active in biodiversity monitoring (Silvertown, 2009). Even though some citizen science projects are able to provide data the quality of which equals expert-based data, most citizen science biodiversity programs focus solely on species abundance and distribution, limiting its use for assessing some international biodiversity targets (Kosmala et al., 2016; Chandler et al., 2017). Other concerns about citizen science include observational biases such as ‘false absences’ or misidentification or uneven spatial and temporal coverage. This raises concerns when making inference using this kind of data, despite the fact that advances in modelling based on such data are currently being made (Hochachka et al., 2012). The extent to which citizen science dataset are biased depends on both the sampling regime used in citizen science programs and the expertise of the recorder which can reduce some of the biases mentioned (e.g. bird watchers data are less clustered around urban areas; Geldmann et al., 2016; Isaac and Pocock, 2015).

Here we study the monitoring of biodiversity characteristics by volunteer citizen scientists taking the special, and often overlooked, case of European hunters. For the purposes of this review we focus on the monitoring activities that hunters engage in that are specific to their hunting activity. We do not include other citizen science activities that they might engage in outside of hunting. Hunters collect data during their activity both voluntarily through cooperation with veterinary or other research institutes, and as a parts of compulsory programs when countries’ hunting regulations mandate such reporting through hunting statistics or the collection of other data (see for instance http://artemis-face.eu/ for an overview of the European hunting bag regulations; Mörner et al., 2014). To a large extent hunter collected data is formally institutionalised into wildlife management structures that are intended to support sustainable harvest. Virtually the entire European landscape is utilised for some form of hunting, and most hunting systems are tied to some form of property rights that ensure a broad distribution of hunters across the whole landscape (Linnell et al., 2015). These factors combined make Europe’s estimated seven million hunters a potentially valuable resource for citizen science data collection (www.face.eu/).

In this study we use the Essential Biodiversity Variables (EBVs) framework to categorise the different types of data coming from hunter-based monitoring and hence assess their contribution to biodiversity monitoring. EBVs are a set of variables that aim to represent biodiversity across its key dimensions (space, time and biological organisation) and that can accurately document biodiversity change (Kissling et al., 2018). EBVs are being defined and refined by GEO BON, a global biodiversity network that contributes to effective management policies for biodiversity. They provide a first level of aggregation computed from the raw data and can be used to compute more complex biodiversity indicators that can be used to measure the achievement of policy goals (Pereira et al., 2013). GEO BON has divided the EBVs into 21 candidates grouped among 6 classes (i.e. genetic composition, species population, species traits, community composition, ecosystem functioning and ecosystem structure, www.geobon.org/ebvs).

We aim at providing a comprehensive overview of hunters’ contributions to biodiversity monitoring in Europe, and review the methods used to collect data in hunter-based monitoring. We also examine the geographic and taxonomic extent of such contributions and the scientific output stemming from hunter-based monitoring data.

2. Method

2.1. Systematic literature search

The first step of the review process was to define the scope of research that focuses on the research question (Booth et al., 2016). In the present study we aimed at identifying which Essential Biodiversity Variables are possible to derive using hunter-based monitoring. We initially developed a list of keywords listing the actor (i.e. hunter or hunting team), the full list of EBVs as defined by GEO BON, the taxonomic scope and the geographic scope. The refinement of this list was done in an iterative fashion, running the list of keywords through Scopus and Web of Science Core Collection and adding new keywords that emerged, and then re-running the search until we reached a plateau in the number of papers returned by the databases (for the full list of keywords see Document A1 in Appendix).

Before any screening the search string returned 1335 papers that we exported to create a dataset. The dataset was initially reduced to 962 papers after screening for duplicates. The search returned many papers that were outside the scope of our study and that concerned anthropological studies on hunter gatherers, studies on hunter-based monitoring outside Europe, or sociological studies on hunters such as hunters’ perception of management decisions or hunters’ willingness to contribute to species monitoring. We excluded these studies after screening for titles and abstracts and reduced our dataset to 493 papers. If doubts remained regarding the potential contribution of a paper to our study, we kept it in our dataset for final screening. We finally screened the paper’s full text and rejected studies in which hunters’ contribution was unclear such as if hunters were only mentioned in the acknowledgement of the papers or if it was unclear how hunter-based monitoring was used to compute the EBV. After this final step we retained a total of 277 papers (Figure A1 in Appendix). The screening process was facilitated using the R package 'revtools' (Westgate, 2019).

2.2. Non-systematic literature search

The non-systematic search of our study was divided into two parts; a targeted search on Web of Science Core Collection and Scopus and a search in the grey literature. Literature collection databases such as Scopus and Web of Science only screen through the title, the abstract and keywords. Because hunters' contribution to data collection is in some cases only mentioned in the method section of peer reviewed papers, we expected the systematic literature search to return incomplete information. To be able to include such sources of information, we used 'snow-ball sampling' (Goodman, 1961), whereby we sampled the references found in the systematic literature search searching for certain authors or countries that we suspected were commonly using hunters' as their main data providers. This added 89 additional unique papers to our dataset.

Secondly, we manually accessed a sample of the proceedings of the International Union of Game Biologists (IUGB). This sample was restricted to the volumes available from our own institutional archives and, did not constitute the whole collection. The included volumes spanned a time period ranging from 1957 to 2011. In total we analyzed 14 out of the 35 existing proceedings. Some of the articles were written in a language that no authors in this paper were able to read (i.e. Russian or German) and were directly excluded. We selected papers based on the inclusion and exclusion criteria previously defined and the search in the IUGB resulted in the addition of 92 papers to the dataset.

2.3. Expert knowledge

To further complete our search of hunters' contribution to wildlife monitoring, we requested information from our contact network of experts, wildlife managers and national hunting associations in different European countries. The list of informants was primarily based on the authors' professional networks, but in some cases we were redirected to more competent contacts they personally knew. We first asked for information related to any hunter-based monitoring programs or other monitoring schemes involving hunters in their respective country and if they could provide documentations (i.e. scientific papers or other documents such as official reports or links to websites), for which species and which method was used. A total of 28 contact persons returned 89 inputs including 23 published papers used to complete our dataset.

2.4. Classification of EBVs

Following the recommendations given by Stewart et al. (2005) and based on the pre-defined keywords, we compiled a database of inputs found in different sources i) a systematic search of the peer-reviewed literature, ii) a non-systematic gathering of literature, including 'snow-ball' search of the scientific and technical grey-literature (which included our own knowledge and libraries), and iii) a survey among our professional networks (i.e. experts in wildlife management or wildlife research, and among national hunting associations).

For each input provided we documented the country, the species, the EBVs computed from the data collected by hunters, the methods used by hunters to collect the data and the source of the case. If there was any doubt about the EBV computed we referred to the 'Measurement and scalability' section of each EBV candidate on the GEO-BON website (https://geobon.org/) to find the EBV that most closely proximated the data collected by hunters (e.g. if jaw bone length was measured, we deduced the EBV candidate 'morphology' was used, although with the implicit understanding that this was ultimately used as a metric to monitor demography/life history). To facilitate the analysis and interpretation we pooled species into five broad functional species groups: "ungulates", "large carnivores", "waterfowl", "other birds" and "small game" (including lagomorphs and medium sized carnivores, Table A1 in Appendix for the full description of each group). We aggregated the diverse data

Table 1
List of criteria to classify hunters' data collection methods.

Method	Criteria
Bag	If the EBV was calculated from data taken from official harvest numbers
Camera trap	If the EBV was calculated from data collected through camera traps operated by hunters
Carcass	If the EBV was calculated from the carcass of the shot animal including any invasive samples
Direct Observation	If the EBV was calculated from hunters' direct Observations of a species
Indirect Observation	If the EBV was calculated from hunters' indirect Observations of signs such as snow tracking, faeces sampling, observations of dens
Questionnaire	If the EBV was calculated from data coming from any sort of questionnaires, including written or digital questionnaires, distributed to hunters
Ringing	If the EBV was calculated as a result of the ringing of animals by hunters, this include birds rings or animal tags
Other	If the EBV was calculated from data collected through any sort hunters' cooperation such as direct interview with hunters, non-invasive sampling or bringing shot ringed or tagged animals

collection methods into eight categories based on the criteria shown in Table 1, namely “Bag”, “Camera trap”, “Carcasses”, “Direct Observations”, “Indirect observations”, “Questionnaire”, “Ringing” and “Others”.

3. Results

3.1. Geographic and taxonomic extent of hunter-based monitoring

Our study suggests that hunter-based monitoring is widely distributed across Europe and across taxa (Fig. 1). We found that 32 out of 36 countries involve hunters in the monitoring of at least one species groups. With respect to the species group present in those countries we found that 16 countries use hunter-based monitoring for all potential species groups (UK, Ireland, the Netherlands and Iceland do not host any populations of large carnivores). In four countries (Albania, Kosovo, Macedonia and Liechtenstein), we did not find any published evidence of an organised hunter involvement in the monitoring of game species.

Based on our review, ungulates and small game are the species groups which have the widest hunter-based monitoring coverage, as nearly all European countries (80% for ungulates and 86% for small game; Fig. 1) involve hunters in the monitoring of these species. Even though geographically widespread, we found that waterfowl and large carnivores are the groups which receive less attention from hunter-based monitoring coverage as 63% and 66% respectively of European countries use some sort of hunter-based monitoring for these groups (Table A2). We did not find any hunter-based monitoring scheme for large carnivores in countries such as France or Lithuania, even though they have population of large carnivores (Linnell and Cretois, 2018).

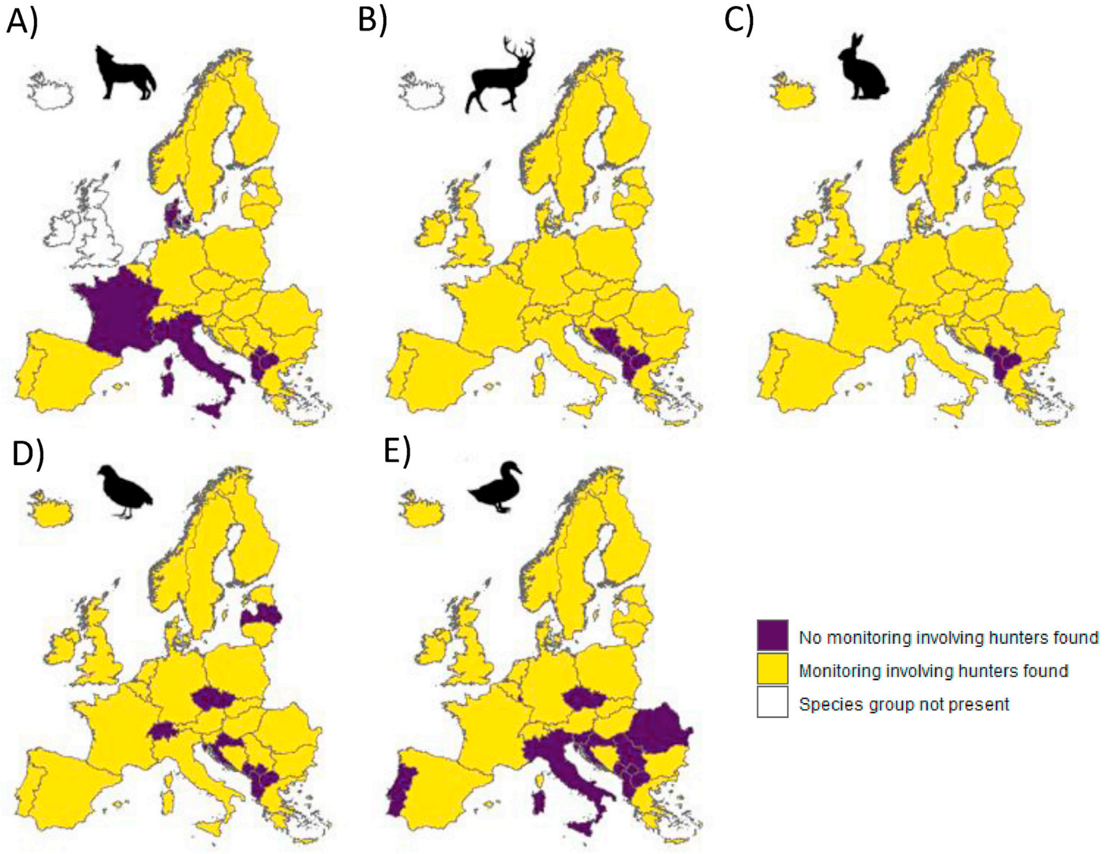


Fig. 1. Geographic extent of hunter-based monitoring per species group. A) Large carnivores, B) ungulates, C) small game, D) other game birds and E) waterfowl.

Table 2
Use of monitoring methods to infer different Essential Biodiversity Variables. (X) representing the finding of a source documenting the use of the method to infer the EBV, (–) if no sources have been found.

EBV class	EBV candidate	Bag	Cam	Carcass	D	I	Help	Questionnaire	Ringing	Other
Genetic composition	Allelic diversity	–	–	X	–	X	–	–	–	X
	Co-ancestry	–	–	X	–	–	–	–	–	–
Species population	Population abundance	X	X	X	X	X	–	X	–	X
	Species distribution	X	X	X	X	X	–	X	–	X
	Population structure	X	–	X	X	X	–	X	–	X
Species traits	Morphology	–	–	X	–	X	–	X	–	–
	Movement	–	–	X	–	–	X	–	X	X
	Phenology	X	–	X	X	X	–	–	–	–
	Physiology	–	–	X	–	X	–	X	–	X
	Reproduction	X	X	X	X	X	–	–	–	X
Community composition	Species interaction	X	–	X	X	X	–	–	–	X
	Taxonomic diversity	–	–	X	–	–	–	X	–	–

3.2. *Diversity of biodiversity characteristics recorded by hunter-based monitoring*

Overall, we found that a wide range of biodiversity characteristics are being derived from hunter-based monitoring programs (Fig. 2). In fact, our study suggests that researchers and wildlife managers infer characteristics on Genetic composition, Species population, Species traits and Community composition with data that are being routinely collected by hunters. We did not find any evidence of hunter-based monitoring schemes directly gathering information on Ecosystem function and Ecosystem structure.

Other game birds and small game were the taxonomic groups for which hunter-based monitoring was the most diverse, with 79% of the species centric EBV candidates being monitored in at least some countries (if we exclude the 7 EBV candidates of the EBV classes Ecosystem function and Ecosystem structure). Hunter based monitoring for other groups of species was less diverse, with 64% of EBVs being recorded for ungulates and 57% for large carnivores and waterfowl (Table A2).

For all species groups, all characteristics concerning their population (i.e. species distribution, population abundance and structure) were recorded except for ungulates, for which we did not find any inputs explicitly documenting the monitoring of their distribution. Nevertheless, it has been argued that species distribution can be inferred from species abundance (Kéry and Royle, 2015), hence it is an implicit by-product of other monitoring. Characteristics at the individual level were also very well monitored by hunters, and examples of monitoring of traits such as physiology, morphology and reproduction were found for all species groups. Even traits that are normally hard to obtain with traditional citizen science such as species phenology and movement were monitored by hunters for 4 out of the 5 species groups.

Our results also suggest that studies use hunters' data for monitoring of genetic composition through studies on allelic diversity and studies on co-ancestry for large carnivores, ungulates, small game and other game birds on samples collected by hunters. It should, however, be noted that we did not find evidence of hunter-based monitoring for the EBV candidates population genetic differentiation and breed and variety diversity for any taxon, although tissue samples from animals shot and non-invasive scat samples are often used by geneticists to study these topics.

3.3. *Methods used to obtain species characteristics*

Hunters contribute to the collection of data relevant for monitoring in many ways, which vary greatly with respect to data volume, coverage and quality (Table 2). The main types of data provided include;

Hunting bags: Information on the numbers of individual animals of different species that are killed by hunters is recorded in most countries, although the spatial resolution of the information varies. Under assumptions of more or less similar effort and similar quotas, between year variation in the numbers of animals shot is being used to infer broad scale spatio-temporal changes in abundance and thus species demographic attributes (e.g. Aebischer, 2019; Massei et al., 2015), especially if combined with secondary data sources (Moleón et al., 2008, 2013; van der Jeugd and Kwak, 2017). As well as being used to follow single population trends, the analysis of such data from multiple populations is used to map changes in distribution and elucidate the relative impacts of multiple drivers of population change (Hagen et al., 2014; Grøtan et al., 2005; Reimoser et al., 2014). We also found that hunting bag data are being used to infer species interaction characteristics through studying fluctuations in small game hunting bags (Smedshaug, 1999).

Biological samples including carcasses of shot animals and non-invasive samplings: Shot animals are used to yield a wide variety of information relevant for monitoring. For mammals, this demographic data is made even more valuable when animals can be aged from tooth sectioning. Data on age and sex can be used to infer population structure and survival rates via analyses like life-tables or population reconstruction (Nilsen et al., 2012; Solberg et al., 1999) and for spatial population structure (Swenson et al., 1998; Kojola and Laitala, 2000). Data on reproduction can be obtained from the analysis of reproductive organs. Body weights and measurement of jawbones or femurs are used to infer body size and condition. Bird wings are used to infer age and sex of animals killed (Pöysä and Väänänen, 2018). Tissue samples are also collected for disease

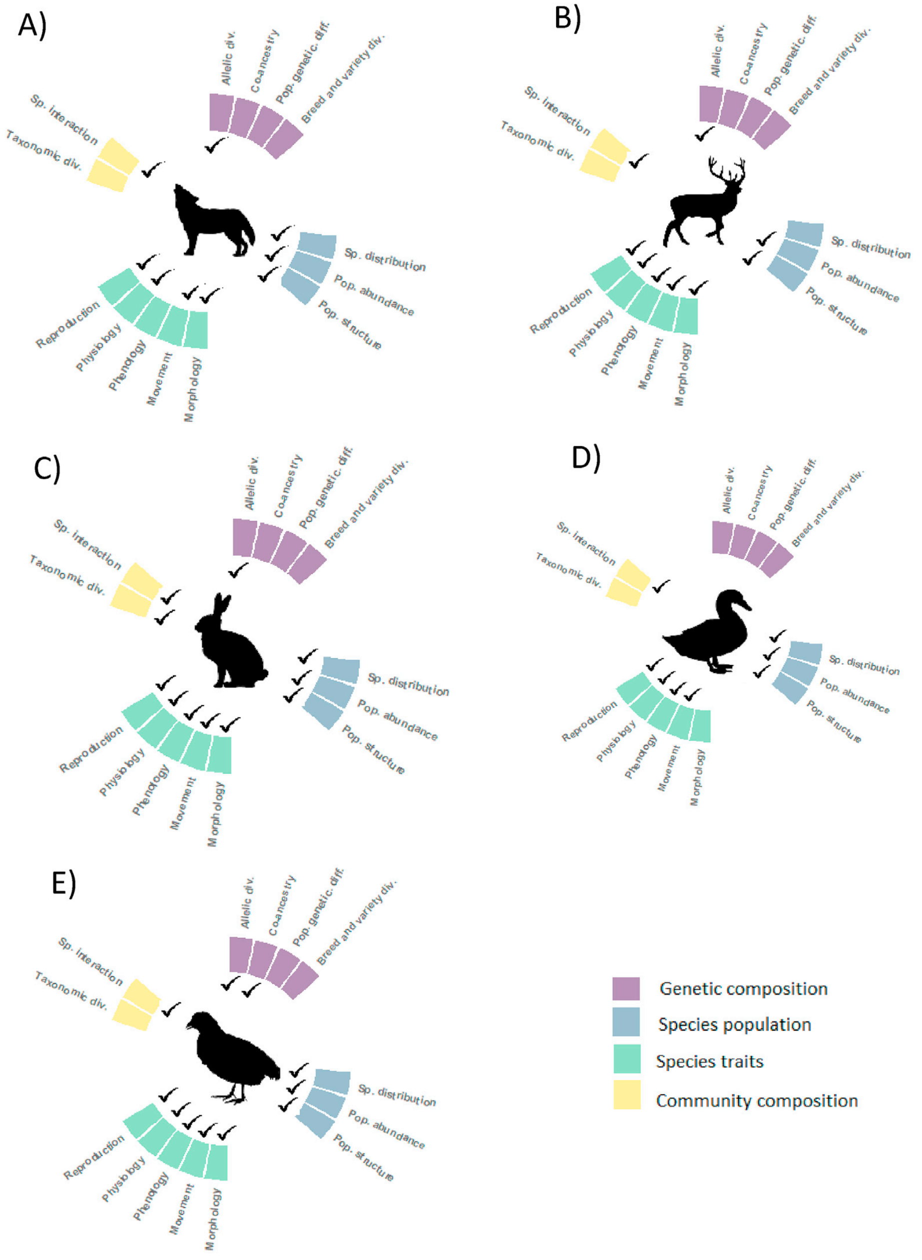


Fig. 2. Diversity of EBV monitored through hunter-based monitoring. A) Large carnivores, B) ungulates, C) small game, D) other game birds and E) waterfowl. A tick indicates if the Essential Biodiversity Variable has been found in our review. Colors represent the Essential Biodiversity Class. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

and parasite screening, ecotoxicology screening, or for genetical analysis (Garbarino et al., 2017; Jelenko and Pokorny, 2010; Tallmon et al., 2004). The rise in non-invasive DNA methods has opened a whole new avenue for collaboration as hunters can collect samples such as faeces for use in population census. For example, the collection of bear scats for DNA based census depends on hunters in Norway, Sweden, Slovenia and Croatia who annually collect thousands of samples (Kindberg et al., 2011; Skrbinšek et al., 2019).

Observations for counts and indices: Hunters also observe many animals while hunting or tending to their hunting areas. These observations, whether direct or indirect, are used to obtain very valuable data on abundance, distribution and structure if there is a robust design and analysis. For example, in Scandinavia the number of moose and bear observations per hunter per hunting day during the hunting season constitutes a robust index of relative abundance (Ericsson and Wallin, 1999; Kindberg et al., 2011; Solberg and Saether, 1999; Swenson et al., 1994). French roe deer hunters report numbers of roe deer seen along transects (Vincent et al., 1991). Hunters in much of Scandinavia and Finland also take part in structured distance sampling-based surveys of abundance of ptarmigan and forest-living grouse (Lindén, 1996). Bear hunters in Slovenia and Croatia record data on sex ratio and reproductive rates of bears observed on feeding stations (Reljic et al., 2018) as well as using simultaneous observations to produce relatively robust minimum counts of the size of the bear population (Bordjan et al., 2019). Hunters all across the Nordic and Baltic countries submit records of lynx and wolf tracks (and increasingly camera trap images) that are used to produce minimum counts of lynx and wolf populations (Linnell et al., 2007, 2010).

Other types of data: Hunters also collect other type of data used to infer a wide range of characteristics regarding their species of interest. Questionnaires and interviews are used for disease detections through documentation of what hunters observe on the hunting ground such as scabies infestation for red fox, hair loss in moose, or inferring species distribution or abundance based on their experience and past Observations (Gortázar et al., 1998; Llaneza and Núñez-Quirós, 2009; Madslien et al., 2011). Hunter cooperation with researchers and management authorities also includes their willingness to help ring birds or tag mammals, and return the carcass of shot ringed and tagged animals (Guzmán et al., 2017; Jensen, 1973).

3.4. Origin of the information

Overall, we found that more than 70% of the diversity in EBV monitoring was documented in both the systematic and unstructured literature search for large carnivores, other game birds, small game and ungulates (Fig. 3). The result is slightly more contrasted for waterfowl as this number goes down to about 60%. Inputs from the unstructured search generally brought as much or more information than the systematic search for large carnivores, small game, ungulates and waterfowl (respectively contributing to 20, 18, 10 and 25%) and didn't add any unique information for small game.

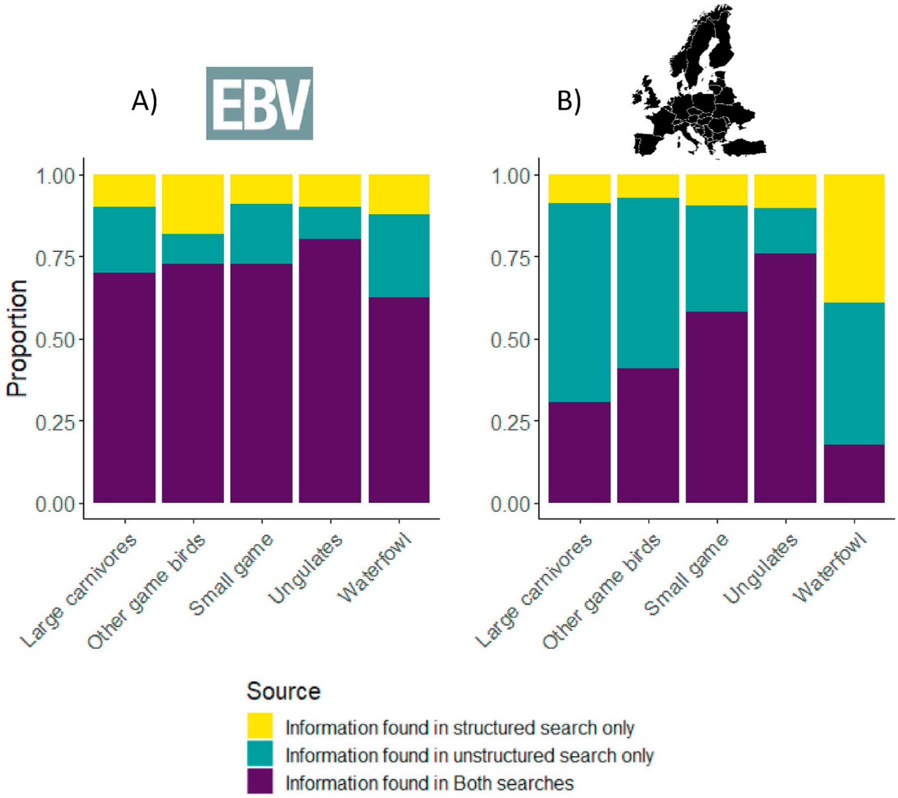


Fig. 3. Proportion of unique information yielded by the systematic search, unstructured search and both sources with regards to A) EBVs and B) Countries.

Regarding geographic extent we found that the unstructured search provided a large amount of unique information for other birds (40%), waterfowl (31%), small game (25%) and large carnivores, for which other sources contributed to 50%. The use of a structured search only yielded more information than the unstructured search for waterfowl and ungulates (respectively 47 and 21% of total information).

4. Discussion

Our study highlights the taxonomic and geographical potential hunter-based monitoring holds, with evidence of nearly all European countries (32 out of 36) using hunter-based monitoring for at least one species group. Moreover, in about half of Europe (including the Scandinavian and Baltic countries) there were examples of hunter-based monitoring to monitor of all species groups. We found that the overwhelming majority of hunter-based monitoring programs focus on multiple aspects of the EBVs grouped under species characteristics (i.e. species population, species traits, genetic composition and species community). However, we found that not all countries use hunter-based monitoring for all species groups and some gaps remain. More specifically, we did not find any evidence of hunter-based monitoring in many of the Balkan countries. Concerning species groups, waterfowl is less monitored by hunters than any other groups. This could possibly be due to the A lack of literature about the use of hunter-based monitoring due to a lack of acknowledgement of hunters' work from the scientific community or due to language barriers could be two other reasons for these gaps given that some of our data has been found only through experts' inputs. Up to 35% of the conservation literature is not written in English (Amano et al., 2016) and given the diversity of languages in Europe we did not expect to get the full picture of the contribution of hunters in biodiversity monitoring through the systematic search of the literature. There is almost certainly a major geographic and species-specific bias in the extent to which hunter derived data is analyzed and published by scientists in English. In fact, the emails sent to wildlife managers and hunting associations added a significant number of inputs in our dataset which were not documented in academic databases through the form of scientific papers or grey literature. This result highlights that we would have dramatically underestimated the extent to which hunters take part in monitoring across Europe without the input of experts. It is almost certain that there are more examples that our search was not able to uncover. The scale of our study (i.e. pooling species into groups and studying hunter based monitoring at country level) might also hide certain fine scale particularities. Even though there is monitoring of a certain EBV in ungulates in a given country, this does not mean that all harvested ungulates are monitored, nor that the whole country is included. As such, our review gives a broad overview of the potential and some examples, but it may not give a full picture of how widespread the use is.

4.1. Particularities of hunter-based monitoring

Our results have highlighted the particularity of hunters as citizen scientists due to their access to certain forms of data. Even though highly controversial in some countries in Europe (Fischer et al., 2013), we have shown that hunting delivers data that can be beneficial to researchers and management authorities through the submission of body parts (ovaries, jaw bones, femurs, wings, tissues) from the carcass of harvested animals from which certain species characteristics would have been otherwise unobtainable.

Hunting data can provide a unique time series in some countries and we have found studies using roe deer antlers over a 67-year period to study change in environmental pollution (Kierdorf and Kierdorf, 2000), or studies using 30 years of bag data (Jansson and Pehrson, 2007). Even though the hunting season is normally limited to only certain periods (i.e. hunting seasons, depending on the taxa and the countries' regulations), some programs make hunters monitor species characteristics throughout the entire year. For instance, the Finnish wildlife triangle is carried out once during summer and once during winter, allowing the creation of time-series dataset useful for ecological research.

Hunting grounds are also widely spread across almost the entire European continent, and even most protected areas are usually open to hunting (Linnell et al., 2015). Hunting is an opportunity for monitoring the status of species population in these areas as hunter-based monitoring has the potential to supplement traditional citizen science which is highly biased towards environments with easy access such as human infrastructures, or nature reserves (Tiago et al., 2017).

Hunters monitor characteristics primarily on species they hunt and hence species they can recognise easily (i.e. for most countries in Europe, getting a hunting licence requires passing examinations, including assessments of species knowledge). This feature is especially important when making inference from hunters' data as species misidentification is presumably less of an issue than in other citizen science data. Hunters' data are also characterised by their institutionalisation. Most European countries oblige hunters to report their harvest to estimate the relative population of game species and to set quotas for the following year. This system involves hunters directly in the management loop, motivating the data collection. Unfortunately, this degree of institutionalisation does not always extend to scientific analysis and publication, with much information remaining within management organisations where it is not readily available.

4.2. The challenges with hunter-based monitoring

There are however several challenges regarding the use of hunter-based monitoring.

Institutionalising the data can potentially lead to some extent of misreporting. Hunters might report a higher harvest rate or a higher number of Observations to artificially boost population numbers, increasing the quota for the following year

(Popescu et al., 2016) or to pretend that they are active on their concession and that they follow the wildlife management plan. This can lead to growing population of game species leading to an increasing human-wildlife interface and resulting in more damages to properties. The same issues could also lead to over-reporting to hide a real decline. Even when misreporting is not purposeful, the discrepancies between population indexes resulting from hunting game bags and other more systematic methods can result in mistrust in data provided by hunters from both institutions and other stakeholders, potentially increasing the negative perception of hunters.

Moreover, because each country has its own reporting system, misreporting can be facilitated depending on the system. For instance, some of the very structured monitoring programs are able to report data that provide precise indices of the variation of game populations over large areas (Ueno et al., 2014; for a selected sample of these programs see Box 1). However, many programs across Europe are much less structured (i.e. they do not follow a robust sampling scheme) and are based on hunter reports of total numbers of animals believed to occur on their hunting grounds, which are then aggregated at other administrative levels. These procedures have poorly described methodology, no robust measures to prevent multiple counts of the same individuals and are highly prone to misinterpretation or even potential abuse (Popescu et al., 2016). At best they may provide a rough relative index of temporal change in abundance (Bragina et al., 2018) and an indication of broad scale distribution. However, despite their somewhat ad hoc nature their utility in guiding sustainable hunting practices over the last 50 years in many areas must be acknowledged. There is potential to add value to these systems if the underlying concrete Observations can be separated from the interpretation, and if some transparent structure can be placed onto both the observation and interpretation processes (e.g. ENETWILD consortium et al., 2018).

Finally, hunting based monitoring programs potentially encounter similar biases as any other citizen science programs. For instance, hunter-killed birds rarely constitute a random subset of a population as juveniles or older birds are more likely to be shot, resulting in age biases in hunting bags (Madsen, 2010). Geographical biases also exist with hunting-based monitoring because of different regional management practices, making some hunting areas more popular than others which render abundance indices less reliable (Ranta et al., 2008). Nevertheless, obtaining accurate measures of game characteristics is possible using rigorous methods and standardised protocols such as the collection of ‘indicators of ecological change’ as done in France for ungulates (Morellet et al., 2007). Recent statistical developments regarding the integration of different data sources and types also allowed the researcher to combine hunters’ bag statistics with other sources to overcome some of the biases inherent to hunting bag data (Isaac et al., 2019; Rutten et al., 2019).

5. Conclusion

With limited resources and requests from governments to monitor diverse biodiversity characteristics at large scales, there is a growing need for scientists and wildlife management authorities to use cost effective methods to collect data. Our study

Box 1

Examples of highly structured hunter-based monitoring schemes

Norwegian moose monitoring program: Besides reporting the harvest, Norwegian moose hunters are asked to report all moose observed during the hunting season on a standardised form (Solberg and Saether, 1999). This system was started in a few municipalities in the late 1960s and extended to cover the entire country during the 1980s. On a daily basis, the leader of each moose hunting team records the number, sex (male, female, unknown) and age (calf, adult, unknown) of all moose observed by the team members but removes individuals that with certainty are observed by more than one member of the team. In addition, they record the number of members that are hunting each day of the hunting season. Data are later reported to the municipality wildlife board and the national deer register (www.hjorteviltregisteret.no/) and used to generate various indices of moose population density and structure for use by the wildlife management (Solberg and Saether, 1999). In addition, hunters in a selection of monitoring sites are required to submit jawbones, ovaries and body weight information of the animals harvested. These data go back to the 1980’s.

Danish bird wing survey: This survey consists of collecting wings from bird shots during the hunting season and is based on voluntary contributions from hunters across Denmark. Every year thousands of wings are forwarded and are used to infer survival, population abundance and structure of the Danish game birds. More information can be found on the website of Aarhus university (http://fauna.au.dk/en/hunting-and-game-management/wing-survey/)

Finnish wildlife triangles: The Finnish Wildlife Triangle scheme was developed by the Finnish Game and Fisheries Research Institute in cooperation with the Hunters’ central organization in 1988. It provides a wide range of information on species population distribution, abundance and structure for 30 wildlife species. This scheme is highly structured and consists of equilateral triangles with 4 km sides distributed across the whole Finnish landscape. These transects are travelled in winter where tracks in the snow are counted, and during the summer when species seen are counted. Annually the census is carried out for 800 to 1000 triangles and involve 7000 volunteers, mainly hunters (Pellikka et al., 2005).

shows that collaborations between hunters and scientists can contribute to biodiversity monitoring in nearly all its key dimensions, except for habitat indicators. Nevertheless, hunter based monitoring is not a panacea as geographical and taxonomical gaps exist in the information brought by hunter based monitoring in Europe, possibly due to the low acceptance for the use of hunter-based monitoring within some conservation circles because of the societal and ethical challenges hunting is now facing (Fischer et al., 2013).

Furthermore, apart from some very structured programs, the unsystematic nature of hunting-based monitoring poses challenges concerning the use of these data. Statistical developments in data integration as well as more rigorous protocols for data collection when using hunting-based monitoring is needed to unlock further the potential that hunters' data holds.

Authors' contribution

All authors participated in the conception and design of the study. BC collected and analyzed the data. BC and JDCL drafted the manuscript with the help of all authors. All authors gave final approval for publication.

Data availability statement

The R script and the full dataset used to carry out the analysis are both made available to ensure full reproducibility and can be found at DOI 10.17605/OSF.IO/GKAZM. The dataset lists all the papers from the structured and unstructured searches as well as the answers from the authors' contacts. The distribution of data from expert consultations of was done under full consent from the respondants.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

BC was funded by a PhD scholarship from the Norwegian University of Science and Technology. The involvement of JDCL, EBN and MG was funded by the Research Council of Norway (grant 251112). We thank David Scallan from FACE – The European Federation for Hunting and Conservation for facilitating access to Europe's national hunting associations. We also thank the wildlife managers and scientists who kindly answered our survey and gave us precious information and the reviewer who provided useful inputs that improved the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gecco.2020.e01077.

References

Aebischer, N.J., 2019. Fifty-year trends in UK hunting bags of birds and mammals, and calibrated estimation of national bag size, using GWCT's National Gamebag Census. Eur. J. Wildl. Res. 65 (4) https://doi.org/10.1007/s10344-019-1299-x.

Amano, T., González-Varo, J.P., Sutherland, W.J., 2016. Languages are still a major barrier to global science. PLoS Biol. 14 (12).

Booth, A., Sutton, A., Papaioannou, D., 2016. Systematic Approaches to a Successful Literature Review. Sage.

Bordjan, D., Javornik, J., Jerina, K., 2019. Cost Benefit Analysis of Different Monitoring Approaches and Guidelines for Optimized Monitoring of Brown Bears. Report prepared within the LIFE DINALP BEAR Project (LIFE13 NAT/SI/0005).

Bragina, E.V., Ives, A.R., Pidgeon, A.M., Balčiauskas, L., Csányi, S., Khoyetskyy, P., Radeloff, V.C., 2018. Wildlife population changes across Eastern Europe after the collapse of socialism. Front. Ecol. Environ. 16 (2), 77–81. https://doi.org/10.1002/fee.1770.

Chandler, M., See, L., Copas, K., Bonde, A.M.Z., López, B.C., Danielsen, F., Turak, E., 2017. Contribution of citizen science towards international biodiversity monitoring. Biol. Conserv. 213, 280–294. https://doi.org/10.1016/j.biocon.2016.09.004.

Díaz, S., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S., Chan, K., Garibaldi, L., Ichii, K., Liu, J., Subrmanian, S., Midgley, G., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razz, C., 2019. No TiSummary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Servicestle.

ENETWILD consortium, Oliver, Keuling, Sange, Marie, Acevedo, Pelayo, Podgorski, Tomasz, Smith, Graham, Scandura, Massimo, Marco Apollonio, E.F., Vicente, J., 2018. Guidance on estimation of wild boar population abundance and density: methods, challenges, possibilities. EFSA Support. Publ. 15 (7), 1449E.

Ericsson, G., Wallin, K., 1999. Hunter Observations as an index of moose Alces alces population parameters. Wildl. Biol. 5 (1), 177–185. https://doi.org/10.2981/wlb.1999.022.

Fischer, A., Kereži, V., Arroyo, B., Mateos-Delibes, M., Tadie, D., Lowassa, A., et al., 2013. (De) legitimising hunting—Discourses over the morality of hunting in Europe and eastern Africa. Land Use Pol. 32, 261–270.

Garbarino, C., Interisano, M., Chiatante, A., Marucci, G., Merli, E., Arrigoni, N., Pozio, E., 2017. Trichinella spiralis a new alien parasite in Italy and the increased risk of infection for domestic and wild swine. Vet. Parasitol. 246 (July), 1–4. https://doi.org/10.1016/j.vetpar.2017.08.021.

Geldmann, J., Heilmann-Clausen, J., Holm, T.E., Levinsky, I., Markussen, B., Olsen, K., Tøttrup, A.P., 2016. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. Divers. Distrib. 22 (11), 1139–1149. https://doi.org/10.1111/ddi.12477.

Goodman, L.A., 1961. Snowball sampling. Ann. Math. Stat. 148–170.

Gortázar, C., Villafuerte, R., Blanco, J.C., Fernández-De-Luco, D., 1998. Enzootic sarcoptic mange in red foxes in Spain. Eur. J. Wildl. Res. 44 (4), 251–256.

Grøtan, V., Sæther, B.E., Engen, S., Solberg, E.J., Linnell, J.D.C., Andersen, R., Lund, E., 2005. Climate causes large-scale spatial synchrony in population fluctuations of a temperate herbivore. Ecology 86 (6), 1472–1482. https://doi.org/10.1890/04-1502.

Guzmán, J.L., Caro, J., Arroyo, B., 2017. Factors influencing mobility and survival of Eurasian Woodcock wintering in Spain. Avian Conserv. Ecol. 12 (2) https://doi.org/10.5751/ace-01096-120221.

Hagen, R., Heurich, M., Kröschel, M., Herdtfelder, M., 2014. Synchrony in hunting bags: reaction on climatic and human induced changes? Sci. Total Environ. 468–469, 140–146. https://doi.org/10.1016/j.scitotenv.2013.08.022.

Hochachka, W.M., Fink, D., Hutchinson, R.A., Sheldon, D., Wong, W.K., Kelling, S., 2012. Data-intensive science applied to broad-scale citizen science. Trends Ecol. Evol. 27 (2), 130–137. https://doi.org/10.1016/j.tree.2011.11.006.

Isaac, N.J., Jarzyna, M.A., Keil, P., Dambly, L.I., Boersch-Supan, P.H., Browning, E., et al., 2019. Data integration for large-scale models of species distributions. Trends Ecol. Evol. 35 (1), 56–67.

Isaac, N.J.B., Pocock, M.J.O., 2015. Bias and information in biological records. Biol. J. Linn. Soc. 115 (3), 522–531. https://doi.org/10.1111/bij.12532.

Jansson, G., Pehrson, Å., 2007. The recent expansion of the brown hare (Lepus europaeus) in Sweden with possible implications to the mountain hare (L. timidus). Eur. J. Wildl. Res. 53 (2), 125–130. https://doi.org/10.1007/s10344-007-0086-2.

Jelenko, I., Pokorny, B., 2010. Historical biomonitoring of fluoride pollution by determining fluoride contents in roe deer (Capreolus capreolus L.) antlers and mandibles in the vicinity of the largest Slovene thermal power plant. Sci. Total Environ. 409 (2), 430–438. https://doi.org/10.1016/j.scitotenv.2010.10.012.

Jensen, B., 1973. Movements of the red fox (Vulpes vulpes L.). In: Denmark Investigated by Marking and Recovery. Vildtbiologisk Station.

Kéry, M., Royle, A.J., 2015. Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS: Volume 1: Prelude and Static Models. Academic Press.

Kierdorf, H., Kierdorf, U., 2000. Roe deer antlers as monitoring units for assessing temporal changes in environmental pollution by fluoride and lead in a German forest area over a 67-year period. Arch. Environ. Contam. Toxicol. 39 (1), 1–6. https://doi.org/10.1007/s002440010072.

Kindberg, J., Swenson, J.E., Ericsson, G., Bellemain, E., Miquel, C., Taberlet, P., 2011. Estimating population size and trends of the Swedish brown bear Ursus arctos population. Wildl. Biol. 17 (2), 114–123. https://doi.org/10.2981/10-100.

Kissling, W.D., Ahumada, J.A., Bowser, A., Fernandez, M., Fernández, N., García, E.A., Hardisty, A.R., 2018. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biol. Rev. 93 (1), 600–625. https://doi.org/10.1111/brv.12359.

Kojola, I., Laitala, H.M., 2000. Changes in the structure of an increasing brown bear population with distance from core areas: another example of pre-saturation female dispersal? Ann. Zool. Fenn. 37 (1), 59–64.

Kosmala, M., Wiggins, A., Swanson, A., Simmons, B., 2016. Assessing data quality in citizen science. Front. Ecol. Environ. 14 (10), 551–560. https://doi.org/10.1002/fee.1436.

Lindén, H., 1996. Wildlife triangle scheme in Finland: methods and aims for monitoring wildlife populations. Finn. Game Res. 49, 4–11.

Linnell, J.D.C., Cretois, B., 2018. Research for AGRI Committee-The Revival of Wolves and Other Large Predators and its Impact on Farmers and Their Livelihood in Rural Regions of Europe.

Linnell, J.D.C., Broseth, H., Odden, J., Nilsen, E.B., 2010. Sustainably harvesting a large carnivore? Development of eurasian lynx populations in Norway during 160 years of shifting policy. Environ. Manag. 45 (5), 1142–1154. https://doi.org/10.1007/s00267-010-9455-9.

Linnell, J.D.C., Kaczensky, P., Wotschikowsky, U., Lescureux, N., Boitani, L., 2015. Framing the relationship between people and nature in the context of European conservation. Conserv. Biol. 29 (4), 978–985. https://doi.org/10.1111/cobi.12534.

Linnell, J.D.C., Odden, J., Andersen, R., Brøseth, H., Andrén, H., Liberg, O., Okarma, H., 2007. Distance rules for minimum counts of Eurasian lynx Lynx lynx family groups under different ecological conditions. Wildl. Biol. 13 (4), 447–455. https://doi.org/10.2981/0909-6396(2007)13[447:drfmco]2.0.co;2.

Llaneza, L., Núñez-Quirós, P., 2009. Distribution of the iberian wolf (Canis lupus signatus) in Galicia (NW Spain): concordance between field sampling and questionnaires. Wildl. Biol. Pract. 5 (1), 23–32. https://doi.org/10.2461/wbp.2009.5.5.

Madsen, J., 2010. Age bias in the bag of pink-footed geese Anser brachyrhynchus: influence of flocking behaviour on vulnerability. Eur. J. Wildl. Res. 56 (4), 577–582.

Madslien, K., Ytrehus, B., Vikøren, T., Malmsten, J., Isaken, K., Hygen, H.O., Solberg, E.J., 2011. Hair-loss epizootic in moose (Alces alces) associated with massive deer ked (Lipoptena cervi) infestation. Journal of wildlife diseases 47 (4), 893–906.

Massei, G., Kindberg, J., Licoppe, A., Gačić, D., Šprem, N., Kamler, J., Náhlik, A., 2015. Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. Pest Manag. Sci. 71 (4), 492–500. https://doi.org/10.1002/ps.3965.

Moleón, M., Almaraz, P., Sánchez-Zapata, J.A., 2008. An emerging infectious disease triggering large-scale hyperpredation. PloS One 3 (6), 12–17. https://doi.org/10.1371/journal.pone.0002307.

Moleón, M., Almaraz, P., Sánchez-Zapata, J.A., 2013. Inferring ecological mechanisms from hunting bag data in wildlife management: a reply to Blanco-Aguiar et al. Eur. J. Wildl. Res. 59 (4), 599–608. https://doi.org/10.1007/s10344-013-0711-1, 2012.

Morellet, N., Gaillard, J.M., Hewison, A.M., Ballon, P., Boscardin, Y.V.E.S., Duncan, P., et al., 2007. Indicators of ecological change: new tools for managing populations of large herbivores. J. Appl. Ecol. 44 (3), 634–643.

Morner, T., Fischer, J., Bengis, R., 2014. The value of increasing the role of private individuals and organisations in One Health. Sci. Tech. Rev. 33 (3), 605–613.

Nilsen, E.B., Brøseth, H., Odden, J., Linnell, J.D.C., 2012. Quota hunting of Eurasian lynx in Norway: patterns of hunter selection, hunter efficiency and monitoring accuracy. Eur. J. Wildl. Res. 58 (1), 325–333. https://doi.org/10.1007/s10344-011-0585-z.

Pellikka, J., Rita, H., Lindén, H., 2005. Monitoring wildlife richness - Finnish applications based on wildlife triangle censuses. Ann. Zool. Fenn. 42 (2), 123–134.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., et al., 2013. Essential biodiversity variables. Science 339 (6117), 277–278.

Popescu, V.D., Artelle, K.A., Pop, M.I., Manolache, S., Rozylowicz, L., 2016. Assessing biological realism of wildlife population estimates in data-poor systems. J. Appl. Ecol. 53 (4), 1248–1259. https://doi.org/10.1111/1365-2664.12660.

Pöysä, H., Väänänen, V.M., 2018. Changes in the proportion of young birds in the hunting bag of Eurasian wigeon: long-term decline, but no association with climate. Eur. J. Wildl. Res. 64 (2) https://doi.org/10.1007/s10344-018-1179-9.

Ranta, E., Lindström, J., Lindén, H., Helle, P., 2008. How reliable are harvesting data for analyses of spatio-temporal population dynamics? Oikos 117 (10), 1461–1468.

Reimoser, S., Smidt, S., Reimoser, F., Wildauer, L., 2014. Changes of hunting bag and habitat in the southern Vienna-Woods since 1891. Allgemeine Forst-Und Jagdzeitung 185 (1/2), 16–27.

Reljic, S., Jerina, K., Nilsen, E.B., Huber, D., Kusak, J., Jonozovic, M., Linnell, J.D.C., 2018. Challenges for transboundary management of a European brown bear population. Glob. Ecol. Conserv. 16, e00488 https://doi.org/10.1016/j.gecco.2018.e00488.

Rutten, A., Casaer, J., Swinnen, K.R., Herremans, M., Leirs, H., 2019. Future distribution of wild boar in a highly anthropogenic landscape: models combining hunting bag and citizen science data. Ecol. Model. 411, 108804.

Silvertown, J., 2009. A new dawn for citizen science. Trends Ecol. Evol. 24, 467–471. https://doi.org/10.1016/j.chemosphere.2018.03.203.

Skrbinšek, T., Luštrik, R., Majić-Skrbinšek, A., Potočnik, H., Kljun, F., Jelenčič, M., Trontelj, P., 2019. From science to practice: genetic estimate of brown bear population size in Slovenia and how it influenced bear management. Eur. J. Wildl. Res. 65 (2) https://doi.org/10.1007/s10344-019-1265-7.

Smedshaug, C.A., Selås, V., Lund, S.E., Sonerud, G.A., 1999. The effect of a natural reduction of red fox Vulpes vulpes on small game hunting bags in Norway. Wildl. Biol. 5 (1), 157–166. https://doi.org/10.2981/wlb.1999.020.

Solberg, E.J., Saether, B.E., Strand, O., Loison, A., 1999. Dynamics of a harvested moose population in a variable environment. J. Anim. Ecol. 68 (1), 186–204. https://doi.org/10.1046/j.1365-2656.1999.00275.x.

Solberg, J., Saether, B., 1999. Hunter Observations of moose Alces alces as a management tool. Wildl. Biol. 5 (1), 107. https://doi.org/10.2981/wlb.1999.014.

Stewart, G.B., Coles, C.F., Pullin, A.S., 2005. Applying evidence-based practice in conservation management: lessons from the first systematic review and dissemination projects. Biol. Conserv. 126 (2), 270–278. https://doi.org/10.1016/j.biocon.2005.06.003.

Swenson, J.E., Sandegren, F., Bjärvall, A., Söderberg, A., Wabakken, P., Franzén, R., 1994. Size, trend, distribution and conservation of the brown bear Ursus arctos population in Sweden. Biol. Conserv. 70, 9–17. https://doi.org/10.1016/0006-3207(94)90293-3.

Swenson, J.E., Sandegren, F., Söderberg, A., 1998. Geographic expansion of an increasing brown bear population: evidence for presaturation dispersal. J. Anim. Ecol. 67 (5), 819–826. https://doi.org/10.1046/j.1365-2656.1998.00248.x.

Tallmon, D.A., Luikart, G., Waples, R.S., 2004. The alluring simplicity and complex reality of genetic rescue. Trends Ecol. Evol. 19 (9), 489–496. https://doi.org/10.1016/j.tree.2004.07.003.

Tiago, P., Ceia-Hasse, A., Marques, T.A., Capinha, C., Pereira, H.M., 2017. Spatial distribution of citizen science casuistic Observations for different taxonomic groups. Sci. Rep. 7 (1), 1–9. https://doi.org/10.1038/s41598-017-13130-8.

Ueno, M., Solberg, E.J., Iijima, H., Rolandsen, C.M., Gangsei, L.E., 2014. Performance of hunting statistics as spatiotemporal density indices of moose (Alces alces) in Norway. Ecosphere 5 (2). https://doi.org/10.1890/ES13-00083.1.

UN General Assembly, 1992. The Convention on Biological Diversity of 5 June 1992.

UN General Assembly, 2015. Transforming Our World : the 2030 Agenda for Sustainable Development, 21 October 2015, A/RES/70/1, available at: (Accessed 1 December 2019) https://www.refworld.org/docid/57b6e3e44.html.

van der Jeugd, H.P., Kwak, A., 2017. Management of a Dutch resident barnacle goose Branta leucopsis population: how can results from counts, ringing and hunting bag statistics be reconciled? Ambio 46 (s2), 251–261. https://doi.org/10.1007/s13280-017-0900-3.

Vincent, J.-P., Gaillard, J.-M., Bideau, E., 1991. Kilometric index as biological indicator for monitoring forest roe deer populations. Acta Theriol. 36, 315–328. https://doi.org/10.4098/at.arch.91-33.

Westgate, M.J., 2019. revtools: an R package to support article screening for evidence synthesis. Res. Synth. Methods 606–614. https://doi.org/10.1002/jrsm.1374.

Article 2

Biological Conservation 244 (2020) 108500

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Review

The challenges and opportunities of coexisting with wild ungulates in the human-dominated landscapes of Europe's Anthropocene

John D.C. Linnell[a,1], Benjamin Cretois[b,*,1], Erlend B. Nilsen[a], Christer M. Rolandsen[a], Erling J. Solberg[a], Vebjørn Veiberg[a], Petra Kaczensky[a], Bram Van Moorter[a], Manuela Panzacchi[a], Geir R. Rauset[a], Bjørn Kaltenborn[c]

[a] *Norwegian Institute for Nature Research, PO Box 5685 Torgard, NO-7485 Trondheim, Norway*
[b] *Department of Geography, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*
[c] *Norwegian Institute for Nature Research, Vormstuguvegen 40, NO-2624 Lillehammer, Norway*

ARTICLE INFO

Keywords:
Ungulate
Cervidae
Bovidae
wild boar
Caprinae
conservation

ABSTRACT

The cumulative impact of human activities has driven many species into severe declines across the globe. However, the recent focus on conservation optimism has begun to highlight case studies that go against this trend. Reforestation, agricultural abandonment, reintroduction and legislative change have led to a situation where large mammals have recovered and are now widespread across the European continent. This study summarizes the knowledge about wild ungulate distribution in Europe and review the diversity of ways in which they interact with humans. Drawn from a wide range of sources, we built distribution maps of European wild ungulates. Results show that 90% of Europe is home to at least 1 species of wild native ungulate, with roe deer and wild boar occupying 74% and 64% of Europe respectively. In contrast, wild native mountain ungulates only occupy 5% of Europe, and are often associated with protected areas. The wide distribution of most European ungulates combined with the extensive human activity within Europe result in a wide range of interactions between ungulates and humans. These interactions can be classified as services or disservices depending on the value orientation and economic position of the various stakeholders perceiving this relationship. Overall, our survey highlights the success of wildlife management policies in Europe and the potential for continental scale conservation of large mammals in human-dominated landscapes. However, maintaining the success of wild ungulate conservation requires actions from national and European institutions to improve coordinated management across jurisdictional borders and sectorial coordination for the whole landscape.

1. Introduction

There are currently many debates ongoing within conservation science concerning the best models for human – nature interactions. These include the debates about land sparing vs land sharing (Fischer et al., 2014), the role of protected areas vs multi-use landscapes (Sayer, 2009), and sustainable use vs protectionist ideals (Cretois et al., 2019). The science of spared landscapes (i.e. protected areas) is well developed, at least in part due to its robust conceptual foundations in island biogeography and the small population paradigm which were central to the early days of conservation biology (Caughley, 1994). However, critics of this approach note that the conceptual confinement of wildlife into “human-free” areas impedes our capacity to envision conservation strategies that do not include remoteness (López-Bao et al., 2017). In contrast, the science of coexistence, which promotes the presence of wildlife in multi-use landscapes, remains ad hoc and fragmented, and both the strategic utility and practicality of the whole approach are being contested. Despite recent attempts to conceptualise the approach (Carter and Linnell, 2016), there remains a dearth of good studies and analyses within the conservation science literature to illuminate the ongoing discussions.

The practice of wildlife conservation began many decades, even centuries, before the development of conservation science (Leopold, 1933). The results of these efforts can contribute valuable insights to inform ongoing debates. This is especially evident for species such as large mammals with which humans have a long and complex relationship, and which have been both directly exploited and directly managed. The relationship between Europeans and large ungulates goes

---

* Corresponding author.
E-mail addresses: john.linnell@nina.no (J.D.C. Linnell), benjamin.cretois@ntnu.no (B. Cretois).
[1] Joint first authors.

https://doi.org/10.1016/j.biocon.2020.108500
Received 8 November 2019; Received in revised form 3 February 2020; Accepted 4 March 2020

back 30,000–40,000 years with the species serving as important contributions to early diet (Morin, 2008) and to early cultures (as objects of art; Fosse and Philippe, 2005). The distributions of species and the structures of communities have fluctuated dramatically during the post-Pleistocene and Holocene, driven by climate, human exploitation and human land-use (Boivin et al., 2016). As for most large mammals, the cumulative impact of human activities had driven most species into severe declines and regional extinctions by the end of the Holocene (i.e. late 19th and early 20th centuries). A trend that has continued for many of the planet's wild ungulates (Ripple et al., 2015). However, somewhat paradoxically with respect to global trends in species endangerment, the status of most European wild ungulates has dramatically improved during the 20th century's transition to the Anthropocene (Linnell and Zachos, 2010). For example the Alpine ibex (*Capra ibex*) was reduced to a few individuals localised in one hunting preserve (now Gran Paradiso National Park) and is now widespread all over the Alps with a population numbering more than 34,000 individuals in the late 1980's (Stüwe and Nievergelt, 1991). A combination of reforestation, agricultural-abandonment, rural-urban migration, legislative change, the development of wildlife management institutions and active reintroduction (mainly driven by hunters) has led to a situation where wild ungulates are now widespread across the European continent, making this an ideal case study to explore the challenges and opportunities associated to human wildlife coexistence.

Our objectives here are to take the case of wild ungulates in Europe, to summarise knowledge on their distribution across the continent, and examine the complex ways with which they interact with humans through the lens of the emerging coexistence discourse which is ongoing within conservation biology. This will include reviewing both the tangible and intangible aspects of their interactions with humans, which will also be viewed through the lens of ecosystem services and disservices.

2. Data gathering

2.1. Distribution maps and analysis

To our best knowledge, the only detailed wild ungulate range maps available are the ones published by the Global Mammal Assessment Group in 2008 and hosted by the IUCN Red List of Threatened Species (https://www.iucnredlist.org/). Although these maps are now a milestone for studies on ungulate ecology and conservation, updating them is demanding. However, ignoring new information can be detrimental to species conservation (Hughes, 2017). Moreover, it has already been noted that the IUCN species range maps lack precision at local scales (Ficetola et al., 2014).

To visualise the extent to which humans and ungulates presently share space, we have produced up-to-date continent-wide distribution maps of all of the native species of wild ungulates (see Fig. A1 in the online Appendix for a workflow illustration). This includes the following species: Red deer (*Cervus elaphus*), roe deer (*Capreolus capreolus*), moose (*Alces alces*), wild reindeer (*Rangifer tarandus*), European bison (*Bison bonasus*), Alpine and Iberian ibex (*Capra ibex* and *C. pyrenaica*), Northern chamois (*Rupicapra rupicapra*), Pyrenean chamois (*R. pyrenaica*) and wild boar (*Sus scrofa*) (Fig. 1). We also include other species like fallow deer (*Dama dama*) and mouflon (*Ovis ammon*), whose native status is unclear and contested (Chapman and Chapman, 1980; Lever, 1985), and species that are without doubt introduced, like white-tailed deer (*Odocoileus virginianus*), sika deer (*Cervus nippon*), Chinese water deer (*Hydropotes inermis*), Reeves's muntjac (*Muntiacus reevesi*), muskox (*Ovibos moschatus*) and barbary sheep (*Ammotragus lervia*) (Fig. A2 in the online Appendix). We do not include feral or extensively grazed free-ranging livestock like horses and semi-domestic reindeer. Data on distribution were drawn from many sources. The three volumes on ungulate management in Europe (Apollonio et al., 2010; Putman et al., 2011a; Putman and Apollonio, 2014) were central starting points. In addition, we added data from national mammal atlases, hunting statistics, citizen science databases, vehicle collisions, scientific papers that provided the location of their study sites or which reviewed the status and distribution of various species and populations, and expert assessments. Full details are provided in the online Appendix. Data were digitalised and represented on a 10 × 10 km grid, although it must be borne in mind that the real resolution of some underlying data may have been coarser. For example, some data were only available as polygon data representing counties, hunting grounds or other administrative units. From the 10 × 10 km grid we derived a spatial dataset containing 48,499 observations that we used to carry out GIS analysis and visualize the results. All analyses have been conducted using R (R Core team, 2018), and maps created using the R package tmap (Tennekes, 2018).

2.2. Hunting bags, ungulate–car collisions, and agriculture and forestry damage data

Because no overall European hunting bag and ungulate-car collision data exist, we gathered data from websites and reports publishing national statistics. If no data were found, we contacted wildlife managers and other administrative staff who had access to these data. Tables 1 and 2 summarize the data collected from national sources (see Table A1 in the online Appendix for a complete overview). Data on wildlife damages to agriculture and forestry are highly fragmented and hard to access. This illustrates that such costs are challenging both to map and quantify. For the purpose of this study, we mainly rely on the compilation found in Apollonio et al. (2010) and selected scientific case studies to illustrate the general types and magnitudes of issues.

3. Current distribution of wild ungulates in Europe

From the pool of 18 species that we included, we recognised 10 species as being native and 8 as being introduced (Linnell and Kaltenborn, 2019). About 90% of Europe's land area hosts from one to five species of wild native ungulate (Fig. 2A and C), with mountainous areas being generally the most species diverse. It should be noted that even though northern Scandinavia appears ungulate free or has a low ungulate diversity, semi-domestic reindeer thrive in most of this range (see Pape and Löffler, 2012 and Fig. A3 in the online Appendix). The introduced species tended to have more limited distributions, although a few areas have up to 4 sympatric species (Fig. 2B and D). Species had highly variable distribution areas. Roe deer and wild boar occupying respectively 74% and 64% (more than 3 million km^2) of the continent and the mountain ungulates (i.e. Northern and Pyrenean chamois, and Alpine and Iberian Ibex) occupying 5% or less of the European land area (less than 250,000 km^2). These differences in distribution area are also reflected in differences in the extent to which distributions were linked to protected areas. With the exceptions of the European bison and the muskox which are most associated with protected areas, the mountain ungulates tended to have the largest proportions of their distribution ranges within protected areas. However, the Pyrenean chamois was the only species to have more than 50% of their range within protected areas. For the other species, on average more than 70% of their distribution was outside protected areas.

4. Interactions with humans and human interests

The fact that most of the distribution area of wild ungulates is outside protected areas, combined with the fact that there is extensive human activity, including hunting, within all European protected areas (Linnell et al., 2015; van Beeck Calkoen et al., 2020), results in a wide range of interactions between wild ungulates and humans. In the following sections we review these interactions and their consequences on humans and their shared ecosystems and discuss how to improve coexistence with wild ungulates.

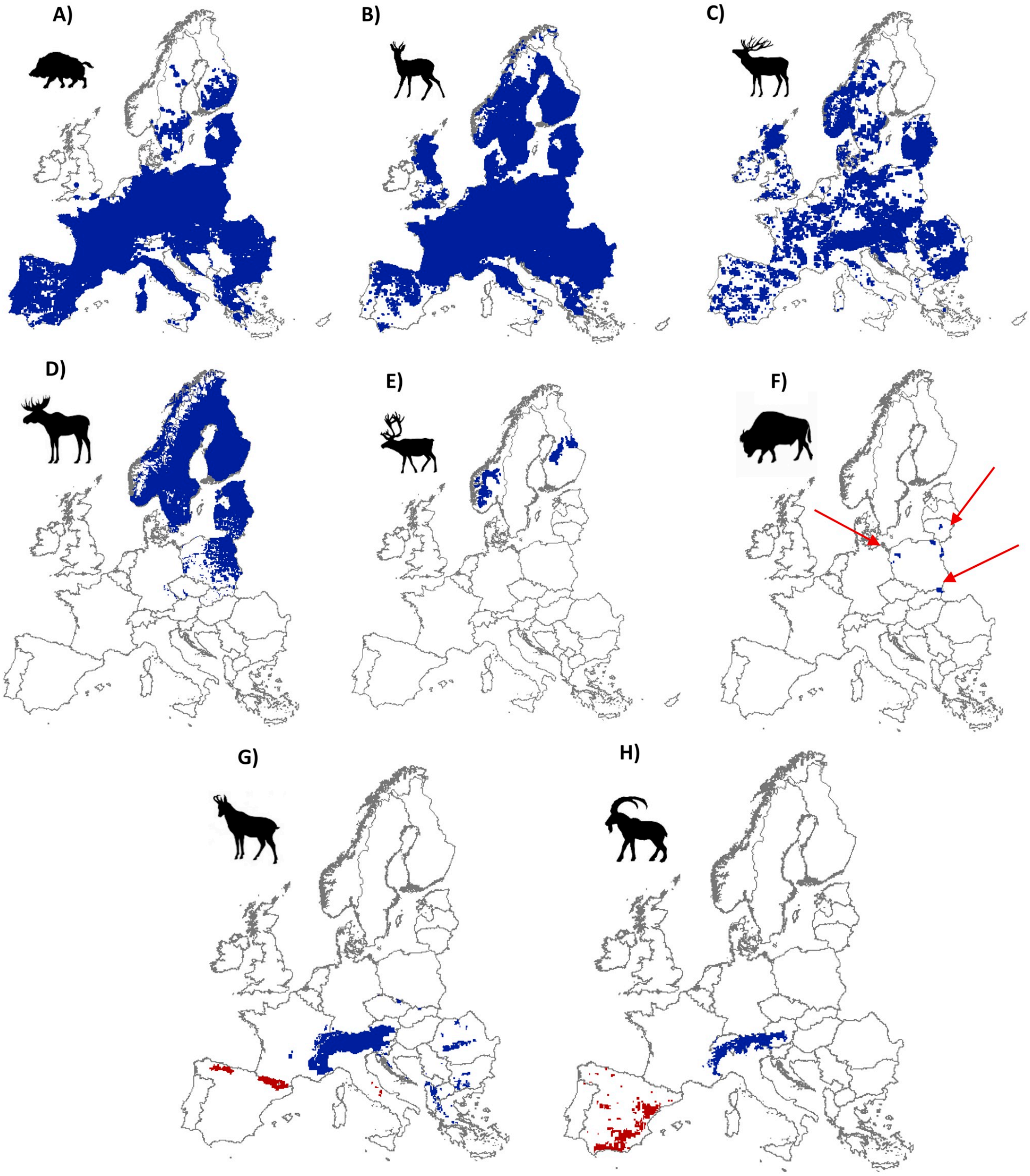


Fig. 1. Distribution maps for wild native ungulates in Europe. A) wild boar, B) roe deer, C) red deer, D) moose, E) wild reindeer, F) European bison, G) northern chamois (in blue) and Pyrenean chamois (in red), H) alpine ibex (in blue) and Iberian ibex (in red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1. Prey for predators and carrion for scavengers

The recovery of Europe's wild ungulates during the 20th century (Linnell and Zachos, 2010) preceded, and facilitated, the recovery of Europe's large carnivores in the latter half of the century (Chapron et al., 2014). There are currently an estimated 17,000 grey wolves (*Canis lupus*) and 9000 Eurasian lynx (*Lynx lynx*) in Europe (Linnell and Cretois, 2018). Presently, wild ungulates constitute the main part of the diet of both species across most of Europe, with the exception of some areas in the extreme north of Fennoscandia and the Mediterranean south where domestic animals constitute the main prey of lynx and wolves respectively (Breitenmoser et al., 2008; Zlatanova et al., 2014). Wolves prey extensively on moose, red deer, roe deer and wild boar, with lynx preying on roe deer, Northern chamois and red deer. Brown bears (*Ursus arctos*) in Scandinavia also prey on moose calves during spring (Rauset et al., 2012) and occasionally on adult moose. Wild

Table 1
Number of wild ungulates harvested in Europe for the most recent year available.

Countries	Years	Total shot per year	Wild boar	Roe deer	Red deer	Moose	Other deer[a]	Mountain ungulates
Germany	2017	2,111,695	836,865	1,190,724	76,794		2429	4883
France	2017	1,421,958	756,149	585,925	62,418		1516	15,950
Poland	2017	660,100	341,400	214,800	94,400		9500	
Austria	2017	413,355	40,297	285,718	61,545		1749	24,046
Czech Republic	2017	383,584	229,182	103,455	27,878		23,069	
Spain	2010	371,377	207,159	29,935	112,252		10,846	11,185
Sweden	2017	354,305	114,831	103,396	10,494	84,754	40,830	
Hungary	2017	249,608	102,600		101,464		34,725	10,819
Slovakia	2016	132,106	53,788	25,627			45,861	6830
Bulgaria	2013	123,654	49,513	64,889	6295		2205	752
Denmark	2017	116,086	190	96,193	9927		9703	73
Scotland	2017	115,037		42,543	61,640		10,854	
Norway	2017	111,816	226	33,280	42,541	31,613	4156	
Finland	2017	107,849	571	9392		56,581	41,285	20
Switzerland	2017	83,200	11,346	44,394	14,611			12,849
Italy (Lombardie)	2017	187,059	114,831	46,507	7978		4424	13,319
Latvia	2017	70,066	25,549	22,135	15,330	7052		
Lithuania	2016	63,792	32,624	23,828	5266	1796	278	
Slovenia	2017	56,150	12,238	34,156	7425			2331
Croatia	2017	49,645	30,000	15,400			4245	
Estonia	2015	46,969	32,580	6264	1252	6873		
Belgium (Wallonie)	2016	45,930	21,721	19,272	4635		139	163
Romania	2012	29,496	19,965	7457	1617		457	
Luxembourg	2017	14,052	6520	6868	468		144	52
Belgium (Flanders)	2017	7379	1550	5762	5		62	
Total		7,326,268	3,041,695	3,017,920	726,235	188,669	248,477	103,272

[a] Includes wild reindeer, sika deer, fallow deer and white-tailed deer. No available harvest data for Reeves' muntjac and Chinese water deer.

ungulates help reduce conflicts between predators and livestock producers as high livestock depredation rates are often linked to low wild prey densities (e.g. Gervasi et al., 2014; Odden et al., 2013).

Although Europe's large carnivores now prey on wild ungulates across much of the continent, the dynamics of these predator-prey relationships are highly modified by the direct and indirect impacts of humans on all trophic levels, i.e. harvest and modification of vegetation through forestry and agriculture, and hunting and traffic collisions that kill both wild ungulates and carnivores (Boitani and Linnell, 2015; Kuijper et al., 2016). This implies that Europe has established large predator – ungulate dynamics within a wide range of different novel ecosystem contexts rather than re-establishing natural dynamics.

4.2. *Ecosystem engineers*

In addition to providing a prey base for increasingly viable large carnivore populations in Europe, the carcasses of ungulates killed by carnivores, or dying from other causes, as well as the gut-piles left by hunters, provide a crucial source of carrion for a diverse guild of vertebrate and invertebrate scavengers (Melis et al., 2004, 2007; Wikenros et al., 2013). Many scavengers such as vultures are of conservation concern and often require the artificial provisioning of carrion, for example in so called “vulture restaurants” (Piper, 2005).

Wild ungulates are also directly associated with a wide range of ecological interactions with vegetation and soil that change the structure and distribution of plant species and soil nutrients. These processes include grazing and browsing, physical disturbance of the soil surface, seed dispersal, and redistribution of nutrients via urination and defecation. In turn, these ungulate-plant-soil interactions have cascading effects on a wide range of invertebrate and vertebrate species (e.g. Danell et al., 2006 and chapters therein).

Although the presence of wild ungulates is essential for restoring ecosystem processes, they have recovered to densities that probably surpass those from historical times in many parts of Europe. This is because of high access to anthropogenic food sources (from agriculture, forestry, and supplementary feeding) and from reductions in predator densities. These high-density ungulate populations have been widely shown to influence vegetation, including species composition, ground vegetation and tree regeneration (Bernes et al., 2018; Kuijper, 2011), as well as associated vertebrate and invertebrate communities (Foster et al., 2014). In some cases, these high grazing or browsing rates may have negative impacts on species of conservation concern and on ecosystem productivity (Ramirez et al., 2018; Velamazán et al., 2017). Grazing could also hinder ecological succession, hence preventing overgrowth in open ecosystems, which can be viewed as both a positive or negative effect depending on conservation and economic objectives.

4.3. *Quarry for hunters*

Most of the early restoration of wild ungulates was initiated and motivated by hunters with the intention of increasing hunting opportunities. Wild ungulates (with the exception of bison and muskox) are currently hunted in virtually all parts of their distributional range, including most protected areas (van Beeck Calkoen et al., 2020). Wild ungulate hunting in Europe is managed in different ways, with major differences existing between and within countries. Hunting management is usually delegated to local (sub-national) administrative levels where different socio-economic and political traditions and cultures lead to different approaches (Linnell and Kaltenborn, 2019). Even though some initiatives (e.g. project Artemis; artemis-face.eu) promote collaboration on hunting management between countries, there are no overall European hunting bag statistics. Nevertheless, Table 1 gives a fairly good estimate of the number of wild ungulates being harvested each year (see Table A2 in Appendix for detailed number of large ungulates harvested in Europe). The approximatively 7.3 million ungulates being harvested represent a considerable meat resource, as well as a source of income for landowners (that may offset some of the damages caused by ungulates) and rural communities, and many recreational opportunities for the estimated 7 million hunters in Europe (https://www.face.eu/).

4.4. *Wildlife viewing for rural residents and tourists*

Wildlife tourism, and in particular wildlife watching, is currently a

Table 2
Number of ungulate-vehicle collisions (!), number of ungulates killed in vehicle collision (†) or both (!†) each year in a selection of European countries.

Country	Years	Yearly number	Roe deer	Wild boar	Moose	Red deer	Fallow deer	Mountain ungulates	Source
Germany	2017–2018	233,070!	191,590!	34,550!		2920!	4010!		Deutscher Jagdverband (www.jagdverband.de/content/wilde-zeiten-auf-der-straße)
United Kingdom	2016	74,000!							The deer initiative (www.deeraware.com/uploads/publications/press-release-motorists-warne-28.pdf)
Sweden	2017	60,857†!	45,863†!	6082†!	5941†!	425†!	2546†!		Nationella Viltolycksrådet (www.viltolycka.se)
France	2009	53,418!	27,991!	20,665!		4762!			Christine St Andrieux, Personal communication
Austria	2017–2018	42,279†	40,897†	602†		663†	27†	85†	Project Roadkill (http://globalroadkill.net)
Norway	2017–2018	22,396†!	14,872†!		4226†!	3298†!			Norwegian deer registry (www.hjorteviltregisteret.no)
Estonia	2014–2018	18,401!	1300	2000	2000				Estonian Road Administration, pers. comm
Poland	2014–2018	18,000!							Jasinska K, pers. comm
Czech Republic	2017	12,043!	75%	15%					Bíl et al., 2017; Bil M, Personal communication
Spain	2006–2012	9042!	3441!	5005!		581!	15!		Sáenz-de-Santa-María and Tellería, 2015
Hungary	2011–2014	7650†	4500–5000†	2500–2000†		500–800†			András. W, Personal communication
Denmark	2017	6608!	85%			4%	7%		Miljø-og Fødevareministeriet (schweiss.dk/sites/default/files/trafikeftersoegninger_2017.pdf)
Latvia	2017	3190!							Lama A, Personal communication
Croatia	2016	2863 †	2457†	334†		72†			Bil. M, Personal communication
Finland	2017	1824!			1824!				Liikennevirasto (julkaisut.vayla.fi/pdf8/lti_2018-06_hirvionnettomuudet_2017_web.pdf)
Lithuania	Per year on average	1000–2000†							Balciauskas. L - Personal communication
Italy(South Tyrol)	2014	776†	686†			90†			Favilli et al., 2018
Belgium (Flanders)	2017	179†	127†	44†			8†		Waarnemingen.be (Citizen science project: waarnemingen.be/Observation.be)
Italy(Piedmont)	2011	148!							Tizzani P, Personal communication
Total		568,244†!	360,203†!	73,178†!	13,831†!	13,825†!	7068†!	85†	

growing sector in Europe with wild ungulates serving as a tourist attraction in many remote areas. For instance, the rising interest in 'moose safaris' in Norway and Sweden show the potential wild ungulates hold for local economies (Margaryan and Wall-Reinius, 2017; Thulin et al., 2015). Muskoxen in central Norway provide novel opportunities for wildlife watching and the basis for new nature tourism companies. This is both as direct targets for viewing, and more indirectly as part of the allure or branding of natural areas with multiple values (Schirpke et al., 2018). Furthermore, it is evident that many rural residents take pleasure in seeing wild ungulates close to their houses and during recreational activity in more remote areas. However, these intangible and non-consumptive interactions between humans and ungulates remain grossly understudied and unquantified.

4.5. *Vehicle collisions*

Europe is a highly and increasingly fragmented continent (Piorr et al., 2011), with an estimated 7.3 million km of public roads of various sizes criss-crossing the landscape (Central Intelligence Agency, 2018). This provides a consequent interface between traffic and ungulates (Bruinderink and Hazebroek, 1996; Putman et al., 2011b). Although up-to-date statistics are hard to find for the whole continent, Table 2 shows that for the selected countries or a subset of regions within some countries, more than half a million collisions with ungulates are recorded every year. These collisions result in damage to vehicles, human injuries and loss of human life, and represent a major source of ungulate mortality (Putman et al., 2011a). They also constitute an animal-welfare issue as an unknown proportion of ungulates are injured but not found after the accidents. In Norway, for instance, the number of moose-vehicle collisions is at least two times as high as the number of moose recorded killed in these accidents (http://www.hjorteviltregisteret.no). Many mitigation measures have been tested, but few have been consistently documented to effectively reduce collision frequency apart from fencing longer road sections in combination with wildlife crossing structures (Huijser et al., 2016; Rytwinski et al., 2016). With high, and in some countries and regions increasing, abundance of ungulates and increasing traffic there is a growing need to reduce the number of collisions (e.g. Rolandsen et al., 2011; Massei et al., 2015).

4.6. *Damage to crops and forests*

Although browsing and grazing are essential ecosystem processes provided by wild ungulates, they can be major sources of conflict with humans when they exploit species of importance to agriculture or forestry. In the absence of protection measures, browsing and grazing can have substantial economic consequences for some farmers. Regions of high moose, red deer and roe deer in regions of high densities are also frequently causing damage to tree species of interest to foresters, preventing the recruitment of some species and, through browsing or bark stripping, seriously reducing the growth rate of others. Estimating the costs of such damages is notoriously difficult, but in some countries, where damages are compensated, payments can reach 13 million euros (e.g. for Poland, see Table 3). Wild boar is one of the species that is most often causing damage to crops (Barrios-Garcia and Ballari, 2012) and is estimated to generate an economic loss of more than 30 million euros in the agricultural and forestry sector in Italy and France alone (Apollonio et al., 2010).

4.7. *Diseases*

There is a risk of increased pathogen transmission when people, livestock and wild ungulates share a landscape. An extensive investigation of emerging infectious disease found that 60% of disease events were caused by zoonoses. More than 70% of these originate in wildlife (Jones et al., 2008), and changing climate may increase the

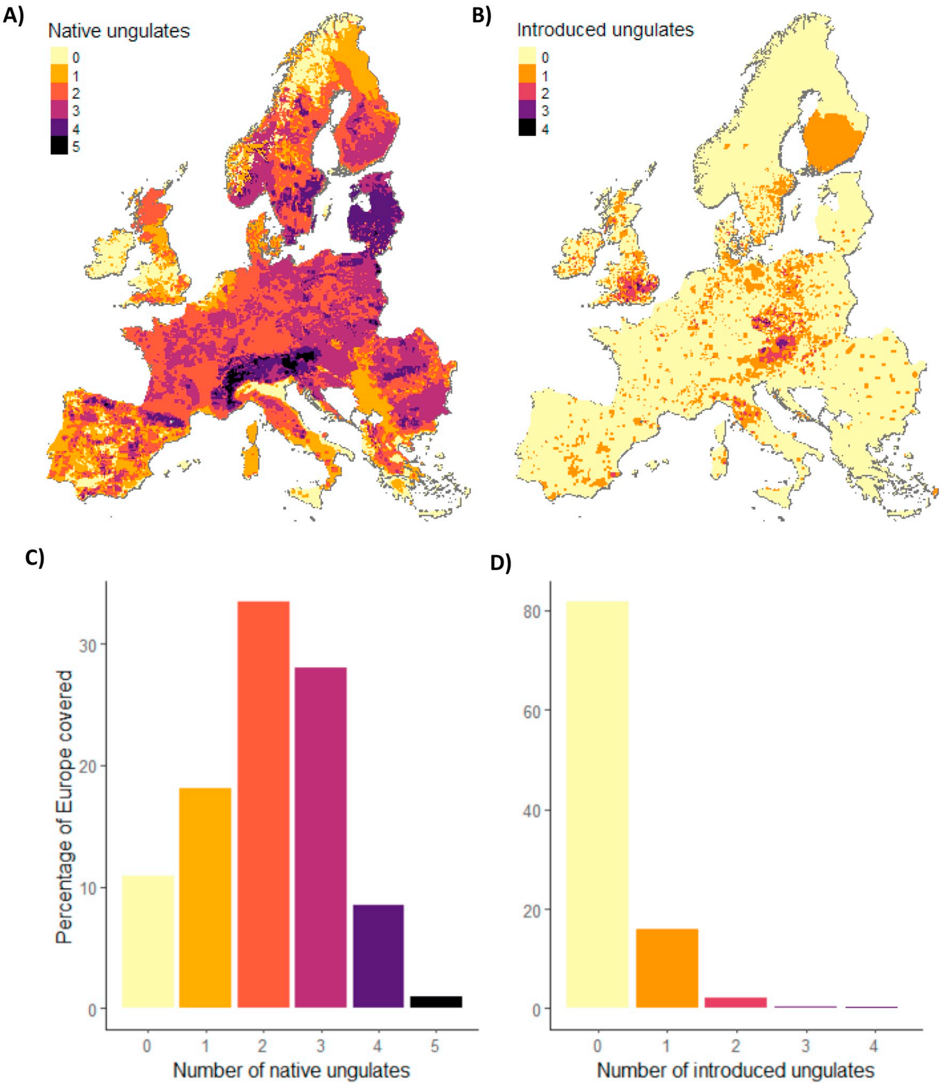


Fig. 2. Distribution maps and histograms of the number of native (A) (C) and introduced (B) (D) diversity in Europe.

occurrence of diseases (Altizer et al., 2011; Altizer et al., 2013). Lyme disease and tick-borne encephalitis, where deer are important tick reproduction hosts, are examples of common zoonoses with an increasing geographic distribution. Accordingly, ungulate density is one of several factors which influences their prevalence (Mysterud et al., 2016). Other zoonoses, involving ungulates as reservoirs, include tuberculosis, and brucellosis (Martin et al., 2011), while trichinellosis is an example of a disease that can be directly transmitted to humans by wild boar meat consumption (Rostami et al., 2018).

Livestock production may be affected by a large number of diseases including anaplasmosis, tuberculosis, paratuberculosis, African swine fever (ASF), classical swine fever, brucellosis, blue-tongue virus and Aujeszky's disease virus (Martin et al., 2011; Gortázar et al., 2007), all of which are on the World Organisation for Animal Health's (OIE) notifiable disease list (OIE 2019 - https://www.oie.int/animal-health-in-the-world/oie-listed-diseases-2019/). Diseases can also be transmitted from livestock to wild ungulate populations. For example, domestic sheep populations can act as a reservoir for *Mycoplasma conjunctiva*, an infectious eye disease that is normally not self-maintained in wild northern chamois (Belloy et al., 2003).

The recent geographic expansion of chronic wasting disease (CWD), into Norwegian wild reindeer (Benestad et al., 2016), moose (Pirisinu et al., 2018), and red deer (Vikøren et al., 2019), was listed among the top 15 most important global issues for biodiversity and conservation research (Sutherland et al., 2018). ASF and CWD are a cause of major concerns for wildlife and livestock health in Europe (Gavier-Widén et al., 2015; EFSA Panel on Biological Hazards, 2016), and may have effects that go far beyond the direct impact on wildlife and livestock as authorities often respond to outbreaks with powerful measures. This may include actions such as eradication of entire subpopulations (Mysterud and Rolandsen, 2018), building of fences to mitigate transmission (Mysterud and Rolandsen, 2019), and significant population

Table 3
Some examples of damage caused by wild ungulates to agricultural crops and forestry at various scales in Europe (most data from Apollonio et al., 2010).

Country	Damage to agriculture	Damage to forestry	Source
Europe	Estimated annual costs from wild boar of €80,000,000.		Apollonio et al. (2010)
Luxembourg	Compensation for wild boar damages was €900,000 in 2004.		Schley et al. (2008)
Slovenia	Compensation for wild boar damages was €575,000 in 2013.		Slovenia Forest Service (2014)
Italy	No consistent estimates available, but total annual costs are expected to exceed €10,000,000. In north eastern Sardinia, a total of €483,982 have been attributed for compensating wild boar damages to crops over a 7 years period.		Apollonio et al. (2010) Lombardini et al. (2016)
Lithuania	With more than 130 ha of crops damaged in 2017, the growth of the endangered European bison populations has led to increase costs in compensation.		Marozas et al. (2019)
Hungary	Compensation payment for crop damage reached €6,000,000 in 2008.	Compensation paid for forest damages in 2005 reached up €585,000.	Bleier et al. (2012) Csányi and Lehoczki (2010)
Poland	Compensation payment for crop damages was €13.7 million in 2010,in which €90,000 was attributed to bison damage.	Costs exceeded €15,000,000 in 2003.	Hofman-Kamińska and Kowalczyk (2012) Wawrzyniak et al. (2010)
France	Compensation for crop damages was more than €21,000,000 in 2005 (mainly wild boar).		Maillard et al. (2010)
Croatia	Compensation payment for damage to crops, forests and vehicles estimated to €685,000 yearly (mainly wild boar and red deer).		Kusak and Krešimir (2010)
Czech Republic		Annual costs estimated to reach up to €1,500,000 per year.	Bartoš et al. (2010)
Finland	Compensation payment in 2006 was €260,000 (mainly moose and white-tailed deer).	Amount of compensation payment in 2006 was €3,200,000 (mainly moose and red deer).	Ruusila and Kojola (2010)
England	Estimated annual costs at €6,560,000.		Putman (2010)
Slovakia	Compensation to crops was €320,000in 2005 (mainly wild boar).	Annual costs exceed €1,490,000 (mainly red deer, roe deer, mouflon and fallow deer).	Findo and Skuban (2007)
Sweden		Estimated annual loss €150,000.	Liberg et al. (2010)

reductions (EFSA, 2018). These actions can in extreme cases also influence entire countries' abilities to trade livestock and livestock products (EFSA, 2018; Schulz et al., 2019), and can be controversial both among the public, scientists and conservationists (Vicente et al., 2019; Mysterud and Rolandsen, 2018).

The interface between wild and domestic ungulates is influenced by livestock husbandry and wildlife management practices. For example, the widespread practice of supplementary feeding of wild ungulates (e.g. Mysterud et al., 2019; Putman and Staines, 2004) may create situations with high local densities and high interaction rates (Gortázar et al., 2006). Providing artificial waterpoints or saltlicks, may have similar effects. Another issue is the long-distance transfer of wild animals between hunting areas (Apollonio et al., 2017). The herding practice of livestock (i.e. enclosed, herded or free-ranging) may also affect the potential for disease transmission with wildlife and hence the level of biosecurity. Sheep, goats and cattle that are extensively grazing in alpine, heathland or forest pastures, have the greatest interface, while those kept on fields close to farms or indoors have the least (Gortázar et al., 2007).

4.8. Opportunity costs

Although many wild ungulates show an ability to tolerate anthropogenic land-use and habitat modifications, the level of tolerance needs to be considered in land-use planning of new developments. Including wild ungulates in the equation may incur opportunity costs for landowners and other property rights holders. One classic example concerns the vulnerability of wild mountain reindeer to virtually all infrastructures (e.g. roads, energy production plants, tourist resorts) and forms of human disturbances (e.g. hiking, skiing; Panzacchi et al., 2016; Beyer et al., 2016). Taking the needs of reindeer into account may require that such human activities are reduced and/or directed away from critical habitats (e.g. Gundersen et al., 2019). Mountain ungulates like ibex and chamois are also vulnerable to such disturbances (Pęksa and Ciach, 2015), and the potential mountains hold for the energy sector (wind and hydropower) have to account for the survival and wellbeing of these species. High ungulate densities may also impose opportunity costs on foresters and farmers as intense browsing pressure may exclude the possibility of planting certain tree species or growing certain crops.

5. A challenge or an opportunity? It's all in the eye of the beholder

The previous sections outlined the diversity of interactions that occur between humans and wild ungulates in shared landscapes. However, it is almost impossible to classify any as being unambiguously positive or negative (challenges vs. opportunities or services vs. disservices) because it depends on the basic value orientation and economic position of the various stakeholders. Conservationists may view predation on ungulates as a positive ecological process, whereas hunters may view this as competition for a valuable quarry (Bisi et al., 2010). Likewise, some conservationists may view herbivory as a positive ecological disturbance process, whereas foresters will view it as an economic loss. However, high herbivory pressure can also compromise conservation values and may as such be considered negative by some conservationists. While hunters obviously view the opportunity to hunt as a positive value, a certain proportion of the public will view it as morally unacceptable (Gamborg and Jensen, 2017). People may even view wildlife diseases in different ways depending on their basic view on whether these are important components of ecological processes and biodiversity (both of which include parasites and diseases as intrinsic components).

The result is that while there are many situations where ungulates cause costs to human economic assets and interests, there are also many conflicts between stakeholders over the way that ungulates should be managed (Bredin et al., 2015; Redpath et al., 2013). Interestingly, in the European context this is not only between conservationists and other stakeholders, but also within the conservation community, as there are many competing approaches to conservation in Europe (Linnell et al., 2015). Even individual stakeholders can have to face complex trade-offs in cases where they engage in the pluriactivity which characterises many European rural residents; for example, it is not uncommon for many farmers to also be hunters and/or foresters and/or engage in rural tourism. In such cases wild ungulates may have very different impacts on each of these different income streams. These complex trade-offs challenge all governance structures. Although economic valuation through ecosystem services is a popular approach to address such issues, it is likely to fail with regard to wild ungulate management. This is because so many of the costs and benefits are of an intangible nature, and not conducive to economic valuation. In addition, the distributions of costs and benefits fall on widely different spatial scales (Linnell, 2015) and within different value domains (Arias-Arévalo et al., 2018). As an example we point to the challenges of scale concerning the impact of ungulate-vehicle collisions. The benefits of high ungulate densities fall on local hunters and landowners, whereas the costs of accidents and mitigation measures are carried at wider societal scales. Finally, the density of ungulates varies across Europe and the impact that ungulates have is often density dependent implying that the relative strengths of different interactions will vary across the continent.

6. Looking into the future

The results of our mapping exercise illustrate that more than 70% of the European continent is now occupied by two or more species of native wild ungulates, with almost 1% hosting 5 species. This is a clear demonstration of the ability of these species to occupy heavily human-dominated landscapes, which is a prerequisite for coexistence. Although we have not formally examined the potential for further increases in distribution, it is highly likely that some species, especially wild boar, roe deer and red deer, have the potential to increase their distributions further in large parts of southern Europe, and in parts of Fennoscandia (Rosvold and Andersen, 2008; Vingada et al., 2010). However, the results of our examination of management issues illustrate the complexity of the human-wildlife and human-human interactions that stem from co-occupation of space. These interactions may be positive and/or negative, or both, and depend on the economic interests and basic values of different stakeholders. The extent to which we can regard the co-occupation with wild ungulates as coexistence depends very much on how these interactions are managed (Fig. 3).

Endangered species and those of clear conservation concern (like large carnivores) are subject to continental level management coordination due to their inclusion under the legal frameworks of the Habitats Directive and the Bern Convention (Linnell and Kaltenborn, 2019). In contrast, the wild ungulates that we consider in this article are normally managed at national or sub-national levels that do not always correspond to ecological scales (Linnell et al., 2015; Linnell and Kaltenborn, 2019). In some cases, this scale can be as low as the municipality or even the individual landowner. Furthermore, there is enormous variation in the extent to which management authorities provide research and administrative support to, and supervision of, these local decision-making structures. This decentralised and delegated approach may have functioned to acceptable levels in the past and has facilitated the recovery of ungulates whose results we are seeing today. However, there are many indications that changes in management structure are going to be needed to meet the challenges of the future.

The drivers of change are diverse, but include the following:

(1) Changing human pressure on the landscape through infrastructure development (transport, recreation, renewable energy production, Venter et al., 2016).
(2) Global change, including climate change, the re-emergence of diseases once thought to be under control, and the appearance of new diseases (Lindgren et al., 2012).
(3) Increased diversity of stakeholder perspectives with divergent, and often conflicting, perspectives on wild ungulate management. The increase in focus on new ideologies like animal rights and rewilding, for example, represent considerable challenges for conventional management structures that are centred around hunting as both an objective and a tool to reach other ecosystem goals (Kennedy and Koch, 2004).
(4) The return of large carnivores as predators on ungulates and competitors with hunters (Chapron et al., 2014).
(5) Constant changes in agricultural and forestry practices in response to shifting policy priorities (Persson et al., 2016).
(6) The controversial impacts of the increasing densities of ungulates, and their expansion into many areas, especially urbanised areas and those with intensive agricultural production create a number of challenges associated with the success of their conservation (Stillfried et al., 2017).
(7) New knowledge about movement patterns, demography, ecological interactions and disease processes that challenge existing management paradigms.
(8) The impact of shifting political directions that are currently dismantling, or restructuring, many of the wildlife management institutions that have developed during the 20th century.

6.1. The need for a bigger picture

Many of the drivers mentioned above affect human-wildlife interactions on large spatial scales, and their consequences can be seen in the need for increased coordination in several areas:

6.1.1. Coordinated management across jurisdictional borders

It is clearly challenging to achieve divergent management goals at spatial scales that are smaller than the biological scales at which the species in question operate (Linnell, 2015). A wide range of telemetry studies conducted during the last 40 years have demonstrated how mobile these species can be (Tucker et al., 2018). Therefore, there is a need to coordinate management actions across administrative and jurisdictional borders. This may include cooperation between multiple landowners, between state and private properties, between protected areas and neighbouring unprotected areas, or between different municipalities, counties, cantons, federal states or even countries.

It has been noted that the degree of transboundary management is currently unsatisfactory (Apollonio et al., 2010) so that there is considerable scope for improvement. Unfortunately, moving management up in scale removes it from the local acceptance that may be instrumental in bringing it to its present favourable situation (Linnell, 2015; Linnell and Kaltenborn, 2019).

Creating central databases of parameters like hunting bags, vehicle collisions and damage payments would be a crucial tool in this endeavour.

6.1.2. Sectorial coordination for whole landscape planning

Traditionally, ungulate management has been an issue mainly limited to the agricultural and forestry sectors. The new challenges require cooperation with a far broader set of sectorial interests. Their importance as prey for large carnivores and their interaction with wider biodiversity make wild ungulates a group of major concern for nature conservation authorities (Ripple et al., 2015). The transport sector represents another example because of the conflicts associated with vehicle collisions and the need to integrate crossing structures like green bridges or viaducts and safety structures like fences into road construction plans to constrain and facilitate ungulate movements (Gunson et al., 2011). The move towards renewable energy sources like solar, wind and hydropower leads to a rapid development of infrastructure, with associated human access, disturbance and barrier effects, in often previously undeveloped areas. These developments can have negative impacts on ungulates, requiring coordination with the energy sector (Northrup and Wittemyer, 2013). Likewise, the growth in recreation, and its associated infrastructure, can create disturbances in previously undisturbed areas, and requires that the wildlife management coordinate with the tourism and recreation sectors. The recent trend to build and refurbish border security fencing (Linnell et al., 2016) which can inhibit ungulate movement, also underlines the need to coordinate with the border security or defence sectors. A similar coordination is needed with the agricultural, veterinary and animal health authorities, given the potential role of wild ungulates in disease transmission.

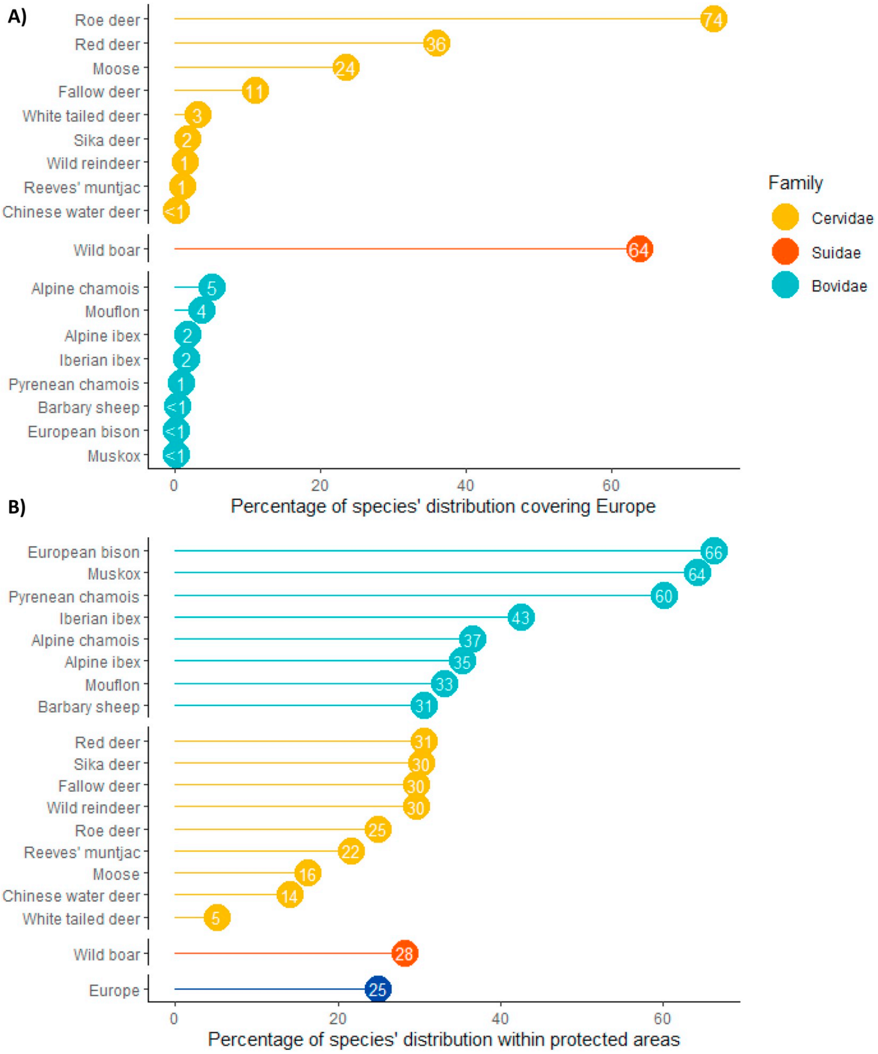


Fig. 3. A) The area of distribution of wild ungulates in Europe (in % of total area) and B) the percentage of this distribution that is within protected areas.

The need for multi-sectorial coordination is not only necessary for the sake of wild ungulate conservation, but also for the possibility for other sectors to succeed in their own priorities. For example, the transport sector can build good roads, but failing to consider wild ungulates can lead to high accident rates. Similarly, a failure by veterinary authorities to consider the possible role of wildlife in disease transmission will frustrate attempts to eradicate or control diseases in livestock.

6.1.3. *The need for planning human activity*

The result of the issues highlighted in the last two sections is the need for both effective spatial planning of human activities and management of wild ungulate populations (Sandström, 2012). Spatial planning is the only existing policy arena where the interests and concerns of multiple sectors can be easily integrated. If the different sectors are aware of the need to consider wild ungulate issues, then spatial planning documents can help them identify areas of overlap and conflict. However, this presupposes an awareness of the need to consider these issues, which is often lacking. Rectifying this lack will require a great deal of awareness raising, and the provisioning of policy guidelines that outline the relevance of wild ungulates for the different sectors. A related aspect of this process also concerns the explicit integration of wild ungulate issues into environmental impact assessment procedures. Our maps of wild ungulate distribution on a continental scale are a first step in identifying and communicating the almost universal relevance of these issues across Europe. However, operationalising these concerns in planning processes will require much finer scaled and locally adapted mapping of areas of distribution and movement.

The other clear finding lies in the need for management of wild ungulates and of regulation of the ways humans interact with them. Hunting is the only readily available tool for influencing wild ungulate numbers on large scales, but the impacts of hunting extend beyond the numbers killed and also embrace the practices that hunters use to influence density, demography and distribution of animals (Milner et al., 2007). The need for hunting regulations to prevent over-harvest is obvious from a conservation point of view. However, it is also important to consider the conflicts (with other sectors) that can occur if too few animals are harvested, or if measures like supplementary feeding lead to too high densities, or aggregations in undesirable areas.

7. Conclusions – insights into coexistence and living with success

There are three main conclusions from this synthesis. Firstly, the survey identifies the high potential for conserving diverse wild ungulate communities across the entire landscape of the European continent, irrespective of whether it is private or public land, protected or unprotected areas. In other words, it reveals that there is a huge scope for a land-sharing form of conservation. Secondly, the diversity of interactions that occur with the humans who share this space underlines the need for a range of active management policies to deal with conflicting objectives and trade-offs. This implies that it is practically meaningless to expect a hands-off (i.e. "let nature take its course") management strategy to work or to use any form of "naturalness" (i.e. an absence of human influence) as a benchmark for setting management objectives (outside of some exceptional areas) as many rewilding or animal rights advocates would desire (Manfredo et al., 2019). Finally, because of the increasing diversity of interests and increasing conflict over objectives, there is a need to build on existing management structures to ensure greater transparency, scientific robustness and social legitimacy. The integration of all these elements is needed to ensure that this very successful cohabitation can continue to be termed coexistence (Carter and Linnell, 2016).

Data availability statement

The R script and the full dataset use to carry out the analysis are both made available to ensure full reproducibility and can be found at https://doi.org/10.17605/OSF.IO/N5P2U.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was mainly funded by the Research Council of Norway (grant 251112). BC was funded by a PhD scholarship from the Norwegian University of Science and Technology. In addition, we thank the two reviewers who provided useful comments which helped to improve the quality of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2020.108500.

References

Altizer, S., Bartel, R., Han, B.A., 2011. Animal migration and infectious disease risk. science 331 (6015), 296–302.

Altizer, S., Ostfeld, R.S., Johnson, P.T., Kutz, S., Harvell, C.D., 2013. Climate change and infectious diseases: from evidence to a predictive framework. science 341 (6145), 514–519.

Apollonio, M., Andersen, R., Putman, R., 2010. European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, UK (604p).

Apollonio, M., Belkin, V.V., Borkowski, J., Borodin, O.I., Borowik, T., Cagnacci, F., ... Gaillard, J.M., 2017. Challenges and science-based implications for modern management and conservation of European ungulate populations. Mamm. Res. 62 (3), 209–217.

Arias-Arévalo, P., Gómez-Baggethun, E., Martín-López, B., Pérez-Rincón, M., 2018. Widening the evaluative space for ecosystem services: a taxonomy of plural values and valuation methods. Environ. Values 27 (1), 29–53.

Barrios-Garcia, M.N., Ballari, S.A., 2012. Impact of wild boar (Sus scrofa) in its introduced and native range: a review. Biol. Invasions 14 (11), 2283–2300.

Bartoš, L., Kotrba, R., Pintíř, J., 2010. Ungulates and their management in the Czech Republic. In: European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, UK, pp. 243–261.

Belloy, L., Janovsky, M., Vilei, E.M., Pilo, P., Giacometti, M., Frey, J., 2003. Molecular epidemiology of Mycoplasma conjunctivae in Caprinae: transmission across species in natural outbreaks. Appl. Environ. Microbiol. 69 (4), 1913–1919.

Benestad, S.L., Mitchell, G., Simmons, M., Ytrehus, B., Vikøren, T., 2016. First case of chronic wasting disease in Europe in a Norwegian free-ranging reindeer. Vet. Res. 47 (1), 88.

Bernes, C., Macura, B., Jonsson, B.G., Junninen, K., Müller, J., Sandström, J., ... Macdonald, E., 2018. Manipulating ungulate herbivory in temperate and boreal forests: effects on vegetation and invertebrates. A systematic review. Environ. Evid. 7 (1), 13.

Beyer, H.L., Gurarie, E., Börger, L., Panzacchi, M., Basille, M., Herfindal, I., ... Matthiopoulos, J., 2016. 'You shall not pass!': quantifying barrier permeability and proximity avoidance by animals. Journal of Animal Ecology 85 (1), 43–53.

Bíl, M., Kubeček, J., Sedoník, J., Andrášik, R., 2017. Srazenazver. cz: a system for evidence of animal-vehicle collisions along transportation networks. Biol. Conserv. 213, 167–174.

Bisi, J., Liukkonen, T., Mykrä, S., Pohja-Mykrä, M., Kurki, S., 2010. The good bad wolf—wolf evaluation reveals the roots of the Finnish wolf conflict. Eur. J. Wildl. Res. 56 (5), 771–779.

Bleier, N., Lehoczki, R., Újváry, D., Szemethy, L., Csányi, S., 2012. Relationships between wild ungulates density and crop damage in Hungary. Acta Theriol. 57 (4), 351–359.

Boitani, L., Linnell, J.D., 2015. Bringing large mammals back: large carnivores in Europe. In: Rewilding European Landscapes. Springer, Cham, pp. 67–84.

Boivin, N.L., Zeder, M.A., Fuller, D.Q., Crowther, A., Larson, G., Erlandson, J.M., ... Petraglia, M.D., 2016. Ecological consequences of human niche construction: examining long-term anthropogenic shaping of global species distributions. Proceedings of the National Academy of Sciences 113 (23), 6388–6396.

Bredin, Y.K., Lindhjem, H., van Dijk, J., Linnell, J.D., 2015. Mapping value plurality towards ecosystem services in the case of Norwegian wildlife management: AQ analysis. Ecol. Econ. 118, 198–206.

Breitenmoser, U., Breitenmoser-Würsten, C., Jobin, A., 2008. Der Luchs. Salm-Verlag.

Bruinderink, G.G., Hazebroek, E., 1996. Ungulate traffic collisions in Europe. Conserv. Biol. 10 (4), 1059–1067.

Carter, N.H., Linnell, J.D., 2016. Co-adaptation is key to coexisting with large carnivores. Trends Ecol. Evol. 31 (8), 575–578.

Caughley, G., 1994. Directions in conservation biology. J. Anim. Ecol. 215–244.

Central Intelligence Agency, 2018. The World Factbook: European Union. Extracted from. https://www.cia.gov/library/publications/the-world-factbook/geos/ee.html.

Chapman, N.G., Chapman, D.I., 1980. The distribution of fallow deer: a worldwide review. Mammal Rev. 10 (2–3), 61–138.

Chapron, G., Kaczensky, P., Linnell, J.D., von Arx, M., Huber, D., Andrén, H., ... Anders, O., 2014. Recovery of large carnivores in Europe's modern human-dominated landscapes. Science 346 (6216), 1517–1519.

Cretois, B., Linnell, J.D., Kaltenborn, B.P., Trouwborst, A., 2019. What form of human-wildlife coexistence is mandated by legislation? A comparative analysis of international and national instruments. Biodivers. Conserv. 28 (7), 1729–1741.

Csányi, S., Lehoczki, R., 2010. Ungulates and Their Management in Hungary. European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, pp. 291–318.

Danell, K., Bergström, R., Duncan, P., Pastor, J., 2006. Large Herbivore Ecology, Ecosystem Dynamics and Conservation. vol. 11 Cambridge University Press, Cambridge, UK.

EFSA Panel on Animal Health and Welfare (AHAW), More, S., Miranda, M.A., Bicout, D., Bøtner, A., Butterworth, A., ... Michel, V., 2018. African swine fever in wild boar. EFSA Journal 16 (7), e05344.

EFSA Panel on Biological Hazards, 2016. Chronic Wasting Disease (CWD) in cervids. EFSA J. 15, 4667.

Favilli, F., Bíl, M., Sedoník, J., Andrášik, R., Kasal, P., Agreiter, A., Streifeneder, T., 2018. Application of KDE+ software to identify collective risk hotspots of ungulate-vehicle collisions in South Tyrol, Northern Italy. Eur. J. Wildl. Res. 64 (5), 59.

Ficetola, G.F., Rondinini, C., Bonardi, A., Katariya, V., Padoa-Schioppa, E., Angulo, A., 2014. An evaluation of the robustness of global amphibian range maps. J. Biogeogr. 41 (2), 211–221.

Findo, S., Skuban, M., 2007. Ungulates and Their Management in Slovakia. European Ungulates and Their Management in the 21st CENTURY. Cambridge University Press, Cambridge, UK, pp. 262–290.

Fischer, J., Abson, D.J., Butsic, V., Chappell, M.J., Ekroos, J., Hanspach, J., ... von Wehrden, H., 2014. Land sparing versus land sharing: moving forward. Conservation Letters 7 (3), 149–157.

Fosse, P., Philippe, M., 2005. La faune de la grotte Chauvet: paléobiologie et anthropolzoologie. In: Bulletin de la société préhistorique française, pp. 89–102.

Foster, C.N., Barton, P.S., Lindenmayer, D.B., 2014. Effects of large native herbivores on other animals. J. Appl. Ecol. 51 (4), 929–938.

Gamborg, C., Jensen, F.S., 2017. Attitudes towards recreational hunting: a quantitative survey of the general public in Denmark. J. Outdoor Recreat. Tour. 17, 20–28.

Gavier-Widén, D., Ståhl, K., Neimanis, A.S., av Segerstad, C.H., Gortázar, C., Rossi, S., Kuiken, T., 2015. African swine fever in wild boar in Europe: a notable challenge.

Vet. Rec. 176 (8), 199–200.

Gervasi, V., Nilsen, E.B., Odden, J., Bouyer, Y., Linnell, J., 2014. The spatio-temporal distribution of wild and domestic ungulates modulates lynx kill rates in a multi-use landscape. J. Zool. 292 (3), 175–183.

Gortázar, C., Acevedo, P., Ruiz-Fons, F., Vicente, J., 2006. Disease risks and overabundance of game species. Eur. J. Wildl. Res. 52 (2), 81–87.

Gortázar, C., Ferroglio, E., Höfle, U., Frölich, K., Vicente, J., 2007. Diseases shared between wildlife and livestock: a European perspective. Eur. J. Wildl. Res. 53 (4), 241.

Gundersen, V., Vistad, O.I., Panzacchi, M., Strand, O., van Moorter, B., 2019. Large-scale segregation of tourists and wild reindeer in three Norwegian national parks: Management implications. Tour. Manage. 75, 22–33.

Gunson, K.E., Mountrakis, G., Quackenbush, L.J., 2011. Spatial wildlife-vehicle collision models: a review of current work and its application to transportation mitigation projects. J. Environ. Manag. 92 (4), 1074–1082.

Hofman-Kamińska, E., Kowalczyk, R., 2012. Farm crops depredation by European bison (Bison bonasus) in the vicinity of forest habitats in northeastern Poland. Environ. Manag. 50 (4), 530–541.

Hughes, A., 2017. Mapping priorities for conservation in Southeast Asia. Biol. Conserv. 209, 395–405.

Huijser, M.P., Fairbank, E.R., Camel-Means, W., Graham, J., Watson, V., Basting, P., Becker, D., 2016. Effectiveness of short sections of wildlife fencing and crossing structures along highways in reducing wildlife–vehicle collisions and providing safe crossing opportunities for large mammals. Biol. Conserv. 197, 61–68.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. Nature 451 (7181), 990–993.

Kennedy, J.J., Koch, N.E., 2004. Viewing and managing natural resources as human-ecosystem relationships. Forest Policy Econ. 6 (5), 497–504.

Kuijper, D., 2011. Lack of natural control mechanisms increases wildlife–forestry conflict in managed temperate European forest systems. Eur. J. For. Res. 130 (6), 895.

Kuijper, D., Sahlén, E., Elmhagen, B., Chamaillé-Jammes, S., Sand, H., Lone, K., Cromsigt, J., 2016. Paws without claws? Ecological effects of large carnivores in anthropogenic landscapes. Proc. R. Soc. B Biol. Sci. 283 (1841), 20161625.

Kusak, J., Krešimir, K., 2010. Ungulates and their management in Croatia. In: European Ungulates and Their Management in the 21st century. Cambridge University Press, Cambridge, UK, pp. 527–539.

Leopold, A., 1933. Game management. Charles Scribner's sons, New York, pp. 481.

Lever, C., 1985. Naturalized Mammals of the World. Longman.

Liberg, O., Bergström, R., Kindberg, J., Von Essen, H., 2010. Ungulates and their management in Sweden. In: European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, UK, pp. 37–70.

Lindgren, E., Andersson, Y., Suk, J.E., Sudre, B., Semenza, J.C., 2012. Monitoring EU emerging infectious disease risk due to climate change. science 336 (6080), 418–419.

Linnell, J.D., 2015. Defining scales for managing biodiversity and natural resources in the face of conflicts. In: Conflicts in Conservation Navigating Towards Solutions, pp. 212–225.

Linnell, J.D., Cretois, B., 2018. Research for AGRI Committee-The Revival of Wolves and Other Large Predators and its Impact on Farmers and Their Livelihood in Rural Regions of Europe. European Parliament.

Linnell, J.D.C., Kaltenborn, B.P., 2019. Institutions for Achieving Human Wildlife Coexistence: The Case of Large Herbivores and Large Carnivores in Europe. Human Wildlife Interactions: Turning Conflict Into Coexistence. Cambridge University Press, Cambridge, UK.

Linnell, J., Zachos, F., 2010. Status and distribution patterns of European ungulates: genetics, population history and conservation. In: Ungulate Management in Europe: Problems and Practices, pp. 12–53.

Linnell, J.D., Kaczensky, P., Wotschikowsky, U., Lescureux, N., Boitani, L., 2015. Framing the relationship between people and nature in the context of European conservation. Conserv. Biol. 29 (4), 978–985.

Linnell, J.D., Trouwborst, A., Boitani, L., Kaczensky, P., Huber, D., Reljic, S., ... Potocnik, H., 2016. Border security fencing and wildlife: the end of the transboundary paradigm in Eurasia? PLoS biology 14 (6), e1002483.

Lombardini, M., Meriggi, A., Fozzi, A., 2016. Factors influencing wild boar damage to agricultural crops in Sardinia (Italy). Curr. Zool. 63 (5), 507–514.

López-Bao, J.V., Bruskotter, J., Chapron, G., 2017. Finding space for large carnivores. Nat. Ecol. Evol. 1 (5), 140.

Maillard, D., Gaillard, J.-M., Hewison, M., Ballon, P., Duncan, P., Loison, A., ... Garel, M., 2010. Ungulates and their management in France. In: European Ungulates and Their Management in the 21st Century, pp. 441–474.

Manfredo, M.J., Urquiza-Haas, E.G., Carlos, A.W.D., Bruskotter, J.T., Dietsch, A.M., 2019. How anthropomorphism is changing the social context of modern wildlife conservation. Biol. Conserv. 108297.

Margaryan, L., Wall-Reinius, S., 2017. Commercializing the unpredictable: perspectives from wildlife watching tourism entrepreneurs in Sweden. Hum. Dimens. Wildl. 22 (5), 406–421.

Marozas, V., Kibiša, A., Brazaitis, G., Jõgiste, K., Šimkevičius, K., Bartkevičius, E., 2019. Distribution and habitat selection of free-ranging European Bison (Bison bonasus L.) in a Mosaic Landscape—a lithuanian case. Forests 10 (4), 345.

Martin, C., Pastoret, P.P., Brochier, B., Humblet, M.F., Saegerman, C., 2011. A survey of the transmission of infectious diseases/infections between wild and domestic ungulates in Europe. Vet. Res. 42 (1), 70.

Massei, G., Kindberg, J., Licoppe, A., Gačić, D., Šprem, N., Kamler, J., ... Cellina, S., 2015. Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. Pest management science 71 (4), 492–500.

Melis, C., Teurlings, I., Linnell, J.D., Andersen, R., Bordoni, A., 2004. Influence of a deer carcass on Coleopteran diversity in a Scandinavian boreal forest: a preliminary study. Eur. J. Wildl. Res. 50 (3), 146–149.

Melis, C., Selva, N., Teurlings, I., Skarpe, C., Linnell, J.D., Andersen, R., 2007. Soil and vegetation nutrient response to bison carcasses in Białowieża Primeval Forest, Poland. Ecol. Res. 22 (5), 807–813.

Milner, J.M., Nilsen, E.B., Andreassen, H.P., 2007. Demographic side effects of selective hunting in ungulates and carnivores. Conserv. Biol. 21 (1), 36–47.

Morin, E., 2008. Evidence for declines in human population densities during the early Upper Paleolithic in western Europe. Proc. Natl. Acad. Sci. 105 (1), 48–53.

Mysterud, A., Rolandsen, C.M., 2018. A reindeer cull to prevent chronic wasting disease in Europe. Nat. Ecol. Evol. 2 (9), 1343.

Mysterud, A., Rolandsen, C.M., 2019. Fencing for wildlife disease control. J. Appl. Ecol. 56 (3), 519–525.

Mysterud, A., Easterday, W.R., Stigum, V.M., Aas, A.B., Meisingset, E.L., Viljugrein, H., 2016. Contrasting emergence of Lyme disease across ecosystems. Nat. Commun. 7, 11882.

Mysterud, A., Viljugrein, H., Solberg, E.J., Rolandsen, C.M., 2019. Legal regulation of supplementary cervid feeding facing chronic wasting disease. J. Wildl. Manag. 83 (8), 1667–1675.

Nationella Viltolycksrådet, 2017. Extracted from. https://www.viltolycka.se/.

Northrup, J.M., Wittemyer, G., 2013. Characterising the impacts of emerging energy development on wildlife, with an eye towards mitigation. Ecol. Lett. 16 (1), 112–125.

Odden, J., Nilsen, E.B., Linnell, J.D., 2013. Density of wild prey modulates lynx kill rates on free-ranging domestic sheep. PLoS One 8 (11), e79261.

Panzacchi, M., Van Moorter, B., Strand, O., Saerens, M., Kivimäki, I., St. Clair, C.C., ... Boitani, L., 2016. Predicting the continuum between corridors and barriers to animal movements using Step Selection Functions and Randomized Shortest Paths. Journal of Animal Ecology 85 (1), 32–42.

Pape, R., Löffler, J., 2012. Climate change, land use conflicts, predation and ecological degradation as challenges for reindeer husbandry in Northern Europe: what do we really know after half a century of research? Ambio 41 (5), 421–434.

Pęksa, Ł., Ciach, M., 2015. Negative effects of mass tourism on high mountain fauna: the case of the Tatra chamois Rupicapra rupicapra tatrica. Oryx 49 (3), 500–505.

Persson, Å., Eckerberg, K., Nilsson, M., 2016. Institutionalization or wither away? Twenty-five years of environmental policy integration under shifting governance models in Sweden. Eviron. Plann. C. Gov. Policy 34 (3), 478–495.

Piorr, A., Ravetz, J., Tosics, I., 2011. Peri-urbanisation in Europe. In: Towards European Policies to Sustain Urban-Rural Futures. Syntesis Report. PLUREL Consortium, Berlin.

Piper, S.E., 2005. Supplementary feeding programmes: how necessary are they for the maintenance of numerous and healthy vulture populations. In: Conservation and management of vulture populations, pp. 41–50 Thessaloniki, Greece.

Pirisinu, L., Tran, L., Chiappini, B., Vanni, I., Di Bari, M.A., Vaccari, G., ... Mitchell, G., 2018. Novel type of chronic wasting disease detected in moose (Alces alces), Norway. Emerging infectious diseases 24 (12), 2210.

Putman, R., 2010. Ungulates and Their Management in Great Britain and Ireland. European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, UK, pp. 129–164.

Putman, R., Apollonio, M., 2014. Behaviour and management of European ungulates. Whittles Publishing Dunbeath, UK.

Putman, R.J., Staines, B.W., 2004. Supplementary winter feeding of wild red deer Cervus elaphus in Europe and North America: justifications, feeding practice and effectiveness. Mammal Rev. 34 (4), 285–306.

Putman, R., Apollonio, M., Andersen, R., 2011a. Ungulate Management in Europe: Problems and Practices. Cambridge University Press.

Putman, R., Langbein, J., Green, P., Watson, P., 2011b. Identifying threshold densities for wild deer in the UK above which negative impacts may occur. Mammal Rev. 41 (3), 175–196.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ramirez, J.I., Jansen, P.A., Poorter, L., 2018. Effects of wild ungulates on the regeneration, structure and functioning of temperate forests: a semi-quantitative review. For. Ecol. Manag. 424, 406–419.

Rauset, G.R., Kindberg, J., Swenson, J.E., 2012. Modeling female brown bear kill rates on moose calves using global positioning satellite data. J. Wildl. Manag. 76 (8), 1597–1606.

Redpath, S.M., Young, J., Evely, A., Adams, W.M., Sutherland, W.J., Whitehouse, A., ... Watt, A., 2013. Understanding and managing conservation conflicts. Trends in Ecology & Evolution 28 (2), 100–109.

Ripple, W.J., Newsome, T.M., Wolf, C., Dirzo, R., Everatt, K.T., Galetti, M., ... Lindsey, P.A., 2015. Collapse of the world's largest herbivores. Science advances 1 (4), e1400103.

Rolandsen, C.M., Solberg, E.J., Herfindal, I., Van Moorter, B., Sæther, B.E., 2011. Large-scale spatiotemporal variation in road mortality of moose: Is it all about population density? Ecosphere 2 (10), 1–12.

Rostami, A., Riahi, S.M., Ghadimi, R., Hanifehpour, H., Hamidi, F., Khazan, H., Gamble, H.R., 2018. A systematic review and meta-analysis on the global seroprevalence of Trichinella infection among wild boars. Food Control 91, 404–411.

Rosvold, J., Andersen, R., 2008. Wild boar in Norway–is climate a limiting factor. NTNU Vitesnskapsmuseet Rapp. Zool 1, 1–23.

Ruusila, V., Kojola, I., 2010. Ungulates and their management in Finland. In: European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, UK, pp. 86–102.

Rytwinski, T., Soanes, K., Jaeger, J.A., Fahrig, L., Findlay, C.S., Houlahan, J., ... van der Grift, E.A., 2016. How effective is road mitigation at reducing road-kill? A meta-analysis. PLoS one 11 (11), e0166941.

Sáenz-de-Santa-María, A., Tellería, J.L., 2015. Wildlife-vehicle collisions in Spain. Eur. J. Wildl. Res. 61 (3), 399–406.

Sandström, C., 2012. Managing large ungulates in Europe: the need to address institutional challenges of wildlife management. Hum. Dimens. Wildl. 17 (5), 320–332.

Sayer, J., 2009. Reconciling conservation and development: are landscapes the answer? Biotropica 41 (6), 649–652.

Schirpke, U., Scolozzi, R., Da Re, R., Masiero, M., Pellegrino, D., Marino, D., 2018. Recreational ecosystem services in protected areas: A survey of visitors to Natura 2000 sites in Italy. J. Outdoor Recreat. Tour. 21, 39–50.

Schley, L., Dufrêne, M., Krier, A., Frantz, A.C., 2008. Patterns of crop damage by wild boar (Sus scrofa) in Luxembourg over a 10-year period. Eur. J. Wildl. Res. 54 (4), 589.

Schulz, K., Oļševskis, E., Staubach, C., Lamberga, K., Seržants, M., Cvetkova, S., ... Sauter-Louis, C., 2019. Epidemiological evaluation of Latvian control measures for African swine fever in wild boar on the basis of surveillance data. Scientific reports 9 (1), 4189.

Slovenia Forest Service, 2014. Yearly hunting management plans for all Slovene hunting management districts for 2014. (in Slovene).

Stillfried, M., Gras, P., Börner, K., Göritz, F., Painer, J., Röllig, K., ... Kramer-Schadt, S., 2017. Secrets of success in a landscape of fear: urban wild boar adjust risk perception and tolerate disturbance. Frontiers in Ecology and Evolution 5, 157.

Stüwe, M., Nievergelt, B., 1991. Recovery of alpine ibex from near extinction: the result of effective protection, captive breeding, and reintroductions. Appl. Anim. Behav. Sci. 29 (1–4), 379–387.

Sutherland, W.J., Butchart, S.H., Connor, B., Culshaw, C., Dicks, L.V., Dinsdale, J., ... Jiang, Z., 2018. A 2018 horizon scan of emerging issues for global conservation and biological diversity. Trends in ecology & evolution 33 (1), 47–58.

Tennekes, M., 2018. tmap: thematic maps in R. J. Stat. Softw. 84 (6), 1–39.

Thulin, C.-G., Malmsten, J., Ericsson, G., 2015. Opportunities and challenges with growing wildlife populations and zoonotic diseases in Sweden. Eur. J. Wildl. Res. 61 (5), 649–656.

Tucker, M.A., Böhning-Gaese, K., Fagan, W.F., Fryxell, J.M., Van Moorter, B., Alberts, S.C., ... Bartlam-Brooks, H., 2018. Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. Science 359 (6374), 466–469.

van Beeck Calkoen, S.T.S., Mühlbauer, L., Andren, H., Apollonio, M., Balciauskas, L., Belotti, E., ... Heurich, M., 2020. Ungulate management in European national parks: why a more integrated European policy is needed. Journal of Environmental Management 260, 110068.

Velamazán, M., San Miguel, A., Escribano, R., Perea, R., 2017. Threatened woody flora as an ecological indicator of large herbivore introductions. Biodivers. Conserv. 26 (4), 917–930.

Venter, O., Sanderson, E.W., Magrach, A., Allan, J.R., Beher, J., Jones, K.R., ... Fekete, B.M., 2016. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. Nature communications 7, 12558.

Vicente, J., Apollonio, M., Blanco-Aguiar, J.A., Borowik, T., Brivio, F., Casaer, J., ... Gortázar, C., 2019. Science-based wildlife disease response. Science (New York, NY) 364 (6444), 943.

Vikøren, T., Våge, J., Madslien, K.I., Røed, K.H., Rolandsen, C.M., Tran, L., ... Neves, C.G.D., 2019. First detection of chronic wasting disease in a wild red deer (Cervus elaphus) in Europe. J. Wildl. Dis. 55 (4), 970–972.

Vingada, J., Fonseca, C., Cancela, J., Ferreira, J., Eira, C., 2010. Ungulates and their management in Portugal. In: European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, pp. 392–418.

Wawrzyniak, P., Jędrzejewski, W., Jędrzejewska, B., Borowik, T., 2010. Ungulates and their management in Poland. European Ungulates and Their Management in the 21st Century. Cambridge University Press, Cambridge, UK, pp. 223–242.

Wikenros, C., Sand, H., Ahlqvist, P., Liberg, O., 2013. Biomass flow and scavengers use of carcasses after re-colonization of an apex predator. PLoS One 8 (10), e77373.

Zlatanova, D., Ahmed, A., Valasseva, A., Genov, P., 2014. Adaptive diet strategy of the wolf (Canis lupus L.) in Europe: a review. Acta Zool. Bulg. 66 (4), 439–452.

Article 3

Coexistence of large mammals and humans is possible in Europe's anthropogenic landscapes

Authors: Benjamin Cretois[1]; John D. C. Linnell[2]; Bram Van Moorter[2]; Petra Kaczensky[2]; Erlend B. Nilsen[2]; Jorge Parada[3]; Jan Ketil Rød[1]

Affiliations:

[1] Department of Geography, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[2] Norwegian Institute for Nature Research, PO Box 5685, Torgard, 7485 Trondheim, Norway

[3] Department of Mathematical sciences, Norwegian University of Science and Technology,
7491 Trondheim, Norway

Corresponding author:

Benjamin Cretois

e-mail: benjamin.cretois@ntnu.no

keywords:

sustainability, Europe, ungulates, carnivores, human disturbance, human-wildlife coexistence

Abstract

A critical question in the conservation of both large carnivores and wild ungulates in the Anthropocene is to know the extent to which they can tolerate human disturbance. In Europe, large mammals have persisted for millennia, and recently expanded, alongside humans, but surprisingly little quantitative data is available about large scale effects of human disturbance on their broad scale distribution. In this study, we quantify the relative importance of human land use and protected areas as opposed to biophysical constraints on large mammal distribution. We analyse recently compiled data on large mammals' distribution (large ungulates and large carnivores) using Bayesian hierarchical models along with dominance analysis to quantify the relative effect of anthropogenic variables on species' distribution. We finally quantify the effect of anthropogenic variables on the size of the species' niche by simulating a scenario where we assumed no anthropogenic pressure on the landscape. We find that the broad scale distribution of most large mammals in Europe includes areas of high to very high human disturbance. Their distribution is primarily constrained by biophysical constraints rather than the human footprint or the presence of protected areas. Furthermore, our counterfactual scenario provides evidence that the human footprint and protected area coverage hardly influence the area of species' distribution. We suggest that coexistence between large mammals and humans is primarily determined by non-environmental aspects of human pressure such as the willingness of humans to share multi-use landscapes with wildlife rather than the ability of wildlife to tolerate humans. This finding offers grounds for optimism concerning wildlife conservation in the Anthropocene.

keywords:

sustainability, Europe, ungulates, carnivores, human disturbance, human-wildlife coexistence

1. Introduction

One of the hallmarks of the Anthropocene is the loss of biodiversity in most parts of the globe as a consequence of a growing human population and increasing per capita consumption with all its associated direct and indirect impacts on wildlife and habitats (Pimm et al. 2014). This trend is accelerating, and the most recent IPBES report estimates that 25 per cent of assessed animal and plant species are threatened by local or global extinction (Díaz et al. 2020). Species diversity supports a wide range of ecosystem services (e.g pollination, food, aesthetic appreciation) including large mammals which have a range of consumptive and non-consumptive values for people (Linnell et al. 2020, Ripple et al. 2014) and its loss would have dramatic impacts on our societies.

Even though most conservation actions have the primary objective of safeguarding the long-term persistence of wildlife, there is substantial disagreement about the most effective strategies to achieve these goals (e.g. land sharing vs land sparing, Phalan et al. 2011). Some conservationists advocate for implementing a spatial dichotomy, where “wild areas” would be subject to minimal human intervention (land sparing) acting as refugia for wildlife against human disturbance. Another paradigm consists of a diversity of coexistence strategies (land sharing), which envisions the possibility of shared landscapes where human and wildlife interactions are allowed, managed and sustained by effective institutions (Carter & Linnell 2016, Linnell & Kaltenborn 2019). Current legislation in most countries mandates wildlife conservation throughout the entire multi-use landscape with the possibility for sustainable use of wildlife, while it additionally sets aside a minimum proportion of the land as protected areas (Cretois et al. 2019).

Adopting a land sharing strategy requires a mutual adaptation in behavior from both humans and wildlife (Carter & Linnell 2016). This may seem especially challenging for large animals as they are more likely to be negatively impacted directly (e.g. through persecution and exploitation) and indirectly (loss and fragmentation of habitats) by human activities due to their larger spatial and resource requirements and the potential for human-wildlife conflicts (Redpath et al. 2013). Because of their size, large animals with wide-ranging behavior and slow reproductive rates are frequently viewed as being at a disproportionally high risk of extinction (Ripple et al. 2014, Ripple et al. 2015).

Coexistence with large mammals has been a historical challenge in Europe. Large carnivores were extensively persecuted in retaliation for killing livestock while large ungulates were overexploited for sport and meat hunting and to minimize damage to crops and forests (Ripple et al. 2014, Ripple et al. 2015). This resulted in populations of both taxa being driven to the edge of a near continent-wide extinction in the 19th and early 20th centuries (Chapron et al. 2014, Apollonio et al. 2010). Even though European landscapes are among the most affected by humans (Venter et al. 2016), strict regulations, reintroduction programs, effective wildlife management institutions, reforestation and agricultural abandonment have allowed most large mammal species to recover. Nowadays, these species are again found across very large areas of the European landscape (Chapron et al. 2014, Linnell & Zachos 2010, Linnell et al. 2020).

Another factor which could potentially have contributed to the re-establishment of these species and their widespread distribution is the widespread protected area network created throughout Europe. However, due to the diverse legislative framework and multiple goals (i.e., encouraging tourism and allowing hunting and different forms of traditional land use) and their small sizes, the conservation effectiveness of protected areas in Europe has been widely disputed for highly mobile, large mammals (Linnell et al. 2015, Gaston et al. 2008).

Although there is an increasing body of literature addressing the influence humans have on large mammals (Tucker et al. 2018, Carter et al. 2012, Alexander et al. 2016), we are not aware of any attempts to quantify the extent to which the contemporary recovering distributions of large predators and their prey in Europe are constrained by the presence of humans' modification in their habitat as opposed to underlying biophysical constraints. The issue is important in order to understand the factors limiting the potential for large scale land-sharing in a crowded and human-modified continent.

In this study we evaluate the relative effects of both the human footprint, a proxy for human disturbance levels widely used in large scale ecological studies (Belote et al. 2020, Tucker et al. 2020, Tucker et al. 2018) and protected areas. We then compare the effect of these two human variables with the effects of biophysical environmental variables such as climate and terrain on large mammal distribution at a continental scale. We use Bayesian hierarchical models to estimate the importance of these variables on species' distributions and compare the environmental niche of these species with and without accounting for human variables by simulating a scenario where the European landscape is free of human influence.

2. Material and Methods

2.1. Distribution data

In this paper we focus on wild large mammals which are native to Europe and whose distribution is not intensively managed (i.e. doesn't depends on heavy human management such as the European bison *Bison bonasus,* Linnell et al. 2020). This includes ten large ungulates: roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*), moose (*Alces alces*), wild reindeer (*Rangifer tarandus*), Alpine chamois (*Rupicapra rupicapra*), Pyrenean chamois (*Rupicapra pyrenaica*), Alpine ibex (*Capra ibex*), Iberian ibex (*Capra pyrenaica*) and wild boar (*Sus scrofa*). We extracted the distribution data provided in Linnell et al. (2020) for all these species. Because the distribution of the mountain ungulates is restricted and because several species belong to the same genus and have similar ecological requirements, we merged the distribution of the Iberian and Alpine ibex, and the distribution of the Alpine and Pyrenean chamois creating *Capra* spp. and *Rupicapra* spp. distributions, respectively. Distribution data for the four species of large carnivore present in Europe; wolves (*Canis lupus*), Eurasian lynx (*Lynx lynx*), brown bears (*Ursus arctos*) and wolverines (*Gulo gulo*) were derived from published data (Chapron et al. 2014). Distribution data for all species had a spatial resolution of 10km x 10km and take the value 0 if the species is absent and 1 if the species is present. As the underlying distribution data is of widely varying quality and resolution, 10km x 10 km is the finest resolution we would advocate for large scaled studies as it erases uncertainty related to the location of a species observation and is computationally manageable. In addition, the 10km x 10km resolution allows the results of our analysis to be comparable to other large scaled studies such as Tucker et al. (2018) or Chapron et al. (2014). We included data on both herbivore and carnivore distribution from 31 countries, consisting of all EU countries (excluding Cyprus and Malta), plus Norway, Switzerland, Serbia, Albania, Northern Macedonia and Kosovo.

2.2. Explanatory variables

We collected three abiotic covariates, two related to climate and one to terrain relief that are thought to be influential biophysical drivers of species distribution (Araújo & Peterson 2012). In addition, we included the two anthropogenic covariates human footprint (HF) and protected area. The biophysical drivers represent potential large-scale and long-term constraints on species' potential distributions (i.e. bioclimatic envelopes) operating through physiological tolerance, rather than fine-scaled and temporally variable environmental factors that typically represent vegetation or habitat patch quality.

Terrain Ruggedness Index and the Potential Evapotranspiration for the Warmest Quarter (PETWQ) were acquired from the ENVIREM dataset (Title & Bemmel 2018) at a spatial resolution of 2.5 arc minutes (i.e. about 3km x 3km at 50°N). The mean snow cover duration (SCD) was derived from the Global SnowPack, a 14-year average available at a 0.25km x 0.25km resolution (from 2000 to 2014, Dietz et al. 2015). We used PETWQ and SCD as proxies for summer and winter severity respectively. Snow cover is widely viewed as being a major limiting factor for species latitudinal and altitude distributions as it correlates with cold winter temperatures, and the physically inhibition of animal movement and access to forage (Leblond et al. 2010). Evapotranspiration serves as a proxy for hot, dry, unproductive summer conditions that also limit species through thermal stress, and poor forage conditions (Tattersall et al. 2012). Terrain ruggedness is widely viewed as being an important escape terrain for species (especially ibex and chamois, and potentially wild reindeer) to avoid disturbance and predation (Nellemann et al. 2007). These three biophysical variables were all obtained as raster data.

As a measure of human disturbance, we chose the Human Footprint Index (HFI, Venter et al. 2016). Ranging from 0 to 50, the HFI is a composite raster built from multiple variables related to human disturbance (e.g. the extent of built environment, cropland, pasture lands, human population density, nighttime lights, railways, roads and navigable waterways; Venter et al. 2016). The HFI has been recently used in multiple continent-wide comparisons of mammal movement rates (e.g. Tucker et al. 2018, Tucker et al. 2020).

Finally, we obtained the protected area coverage from the World Database on Protected Areas: https://protectedplanet.net/). We selected all terrestrial designated protected areas in Europe (i.e. we excluded protected areas whose category was either “Not Reported”, “Not Applicable” or not “Assigned”). Data was available as vector data and was rasterized at a resolution of 1km^2 using ArcGIS Pro for ease of computation. We finally used aggregation to sum the total number of 1km x 1km pixels of protected area within each 10km x 10km grid cell (i.e. the grid cell value for protected area varying from 0 for a grid cell containing no protected area to 100 for a grid cell entirely covered by a protected area) Although European protected areas are almost never wilderness areas (Calkoen et al. 2020, Linnell et al. 2015) they are expected to be associated with greater restrictions on human activities that could potentially better limit human impacts on wildlife, and less intensive forms of land use. However, we did not separate the different IUCN categories as previous studies show that there is little difference in human footprint between categories (Leroux et al. 2010) and there is a high degree of variation between European countries in how they manage protected areas of different IUCN categories (Gaston et al. 2008).

We assessed the extent of collinearity between the covariates. Winter and summer severity were negatively related (r = -0.71), as both display strong coastal-inland and north-south gradients. However, we opted to include both as they reflect different mechanisms for species' ecology. Following Dormann et al. 2013 we made sure to carefully interpret the results of these two variables by interpreting the combined effects of all environmental variables (More detailed explanations in Annexes). Other covariates were not significatively correlated with each other (r < 0.70; Table A1 in Annexes). We aggregated all the explanatory variables to the same 10km x 10km grid cell resolution.

2.3. Statistical analysis

2.3.1. Model specification

Because the residuals of the non-spatial models were strongly spatially correlated, we fitted an intrinsic conditional autoregression (iCAR) model using hierarchical Bayesian models for each of the 13 species. The probability of presence (π) of a given species in a given grid cell was calculated using a Bernoulli distribution and the following model:

$$y_i \sim Bernouilli(\pi_i)$$

$$logit(\pi_i) = \alpha_i + x_i\beta + u_i$$

Where x_i is the vector of covariates for cell *i*, β the vector of parameters to be estimated and u_i the spatially correlated random effect whose prior is defined as:

$$u_i|u_k \sim normal(\frac{\sum_{i\neq k} w_{i,k}\, u_k}{n_i}, \frac{\sigma_u^2}{n_i})$$

Where $w_{i,k} = 1$ if grid cells *i* and *k* are neighbors and 0 otherwise. n_i is the total number of neighbors of grid cell *i*. We define two cells as being neighbors if they directly share a single boundary point. All models assume a vague prior for the regression parameters $\beta \sim normal(Mean = 0,\ SD = 1000)$ and we used a penalized complexity prior on the spatial effect to avoid risks of overfitting (See Annexes).

As we expect species to have an optimal niche for environmental variables, we included linear and quadratic terms for winter and summer severity and ruggedness (Svenning et al. 2011). We also included linear and quadratic terms for human footprint as we suspected certain species to have an optimal niche in the moderate human disturbance level. We only included a linear effect for protected area coverage as we only expected a linear response.

To fit the spatial models, we used the Integrated Nested Laplace Approximation (INLA) approach with the package R-INLA (Lindgren & Rue 2015). INLA is a faster alternative to Markov Chain Monte Carlo approaches and yields similar, if not identical, results (Beguin et al. 2012). We standardized the covariates to enable direct comparison between the regression coefficients. All analyses were conducted in R 3.6.1.

We validated the models by plotting residual values against covariates for each model. We also plotted the leave-one out cross validation scores (conditional predictive ordinate CPO in our case) to estimate model fit.

2.3.2. Evaluation of variables’ importance for species’ distribution

We estimated the relative importance of both environmental and anthropogenic variables using dominance analysis (Azen & Budescu 2003), which is a procedure to quantify the importance of a random variable through examination of the R^2 values (or similar metrics) for all possible subset models of a predefined full model. In a dominance analysis, the higher the dominance score the more useful is the random variable in predicting the response variable. Because the number of models required to estimate the importance of a single random variable grows exponentially with the total number of random variables, we did not quantify the importance of each single variable, but rather the importance of the combined effect of summer and winter severity and ruggedness ("environmental variables") and human footprint and protected area coverage ("anthropogenic variables"). Thus, we fitted 3 models for each of the 11 species; a full model containing all variables, a model containing only the environmental variables and a model containing only the anthropogenic variables. For all models we computed the R^2_{glmm}, a modified version of the classic R^2 which is suitable for mixed models (Nakagawa & Schielzeth. 2013). We sampled 1,000 values from the posterior distribution of the model parameters and bootstrapped the R^2_{glmm} 1,000 times. We finally rescaled the dominance score for it to range from 0 to 100%.

2.3.3. Quantifying the effect of anthropogenic variables on the size of the species' suitable habitat

To further assess the results of the dominance analysis we assessed the relative extent to which anthropogenic variables influence the realized distribution of the studied large mammals we quantified the geographic representation of the suitable habitat for each species (i.e. the potential suitable area available due to environmental predictors only, Guisan & Thuiller 2005). We predicted the probability of a species' occurrence within a grid cell both when anthropogenic variables were set at their minimum value (i.e. we simulated a landscape free of all human influence: no human footprint and no protected areas) and when anthropogenic variables are set to their observed values. We summed these predicted occurrences across Europe to estimate the expected number of occupied cells (i.e. the size of a species' suitable area in Europe). A sum of predictions in a human-free landscape higher than a sum of prediction for the full model implies that the species increase its range in absence of human influence in the landscape. We sampled 1,000 values from the posterior distribution of the model parameters and bootstrapped the niche area 10,000 times.

3. Results

For ease of interpretation we consider five disturbance levels (Venter et al. 2016). A ‘no human disturbance’ area has a human footprint of 0; a ‘low disturbance’ area a human footprint of 1-2; a ‘moderate disturbance’ area a human footprint of 3-5; a ‘high disturbance’ areas a human footprint of 6-11; and ‘very high disturbance’ area a human footprint of 12-50, following the definition by Venter et al (2016).

With a median human footprint of 12.2, summary statistics show that more than 50% of Europe’s area is in an area of very high human disturbance, while less than 8% of Europe have no, to low, human footprint (Figure 1). Protected areas are spread throughout Europe with the median area of protected areas per 100 km^2 (i.e. per 10km x10 km grid cell) being 9 km^2 (Q1 = 0 km^2, Q3 = 41 km^2). Grid cells containing at least 50 km^2 of protected areas tended to have on average a slightly lower human footprint than grid cells containing less than 50km^2 of protected areas (median = 10.04 and 12.98 respectively).

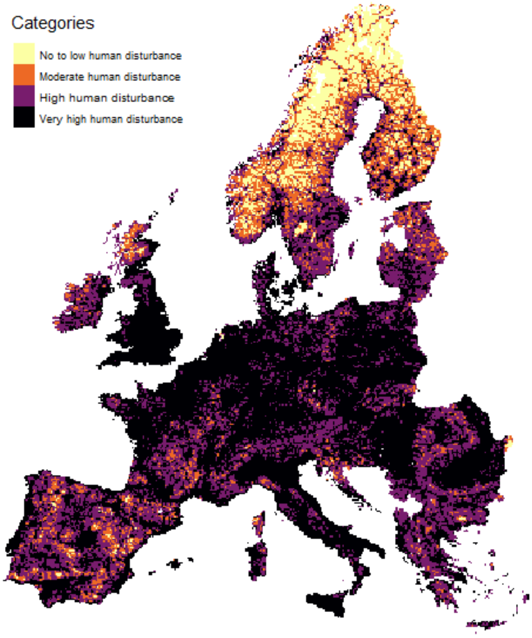


***Figure 1**: Map of the distribution of human disturbance levels in Europe*

The seven large ungulates and four large carnivores demonstrate great variability in their presence across the human footprint gradient (Figure 2). Roe deer (median of 12.8, Q1 = 8.2, Q3 = 18.2) and wild boar (median = 13.5, Q1 = 9.2, Q3 = 18.7) are the species present at the highest human footprints. These statistics show that more than 50% of the roe deer and wild boar distribution occurs in areas of very high human footprint. Wild reindeer (median = 3.9, Q1 = 2.1, Q3 = 4.8) and wolverines (median = 2.7, Q1 = 1.1, Q3 = 4.4) are at the other end of the spectrum with distributions in places that are least impacted by human disturbance. Our data also shows that wolves are not restricted to "wild" remote places but can live in areas where human disturbance is high (median = 9.6, Q1 = 6.8, Q3 = 13). More than 25% of their distribution is in areas where human disturbance is very high.

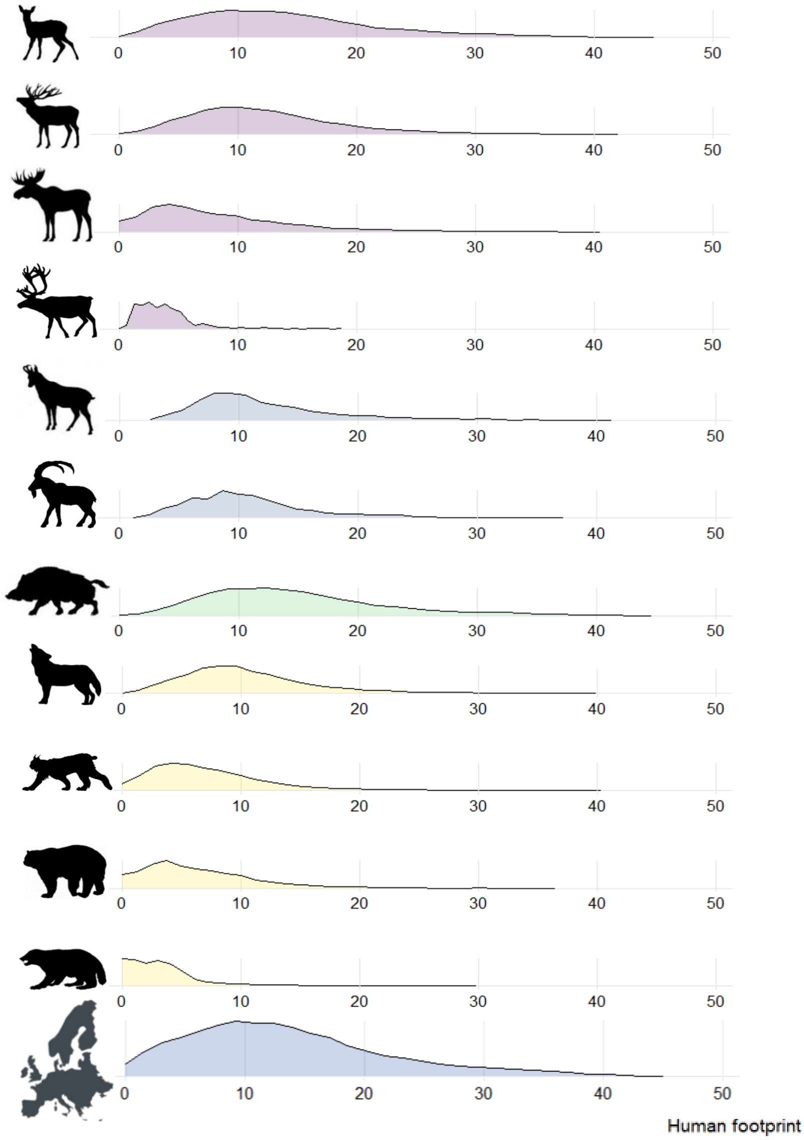


Figure 2: *Species' distributions across the human footprint gradient. From top to bottom: roe deer, red deer, moose, wild reindeer, chamois, ibex, wolf, lynx, bear, wolverine, and the European human footprint distribution.*

Results from the dominance analysis show that the distributions of all 11 species are largely explained by the biophysical variables (Figure 3). In fact, biophysical variables consistently dominate the models (with a relative importance close to 100%) and the influence of anthropogenic variables in our models is shown to be close to 0% or even negative (i.e. the R^2 of the model gets worse as we include these variables). Only for red deer and wolf do anthropogenic variables increase the models' R^2 values (median = 3.3% and median = 12% respectively), although their effects were still considerably lower than those of the biophysical variables.

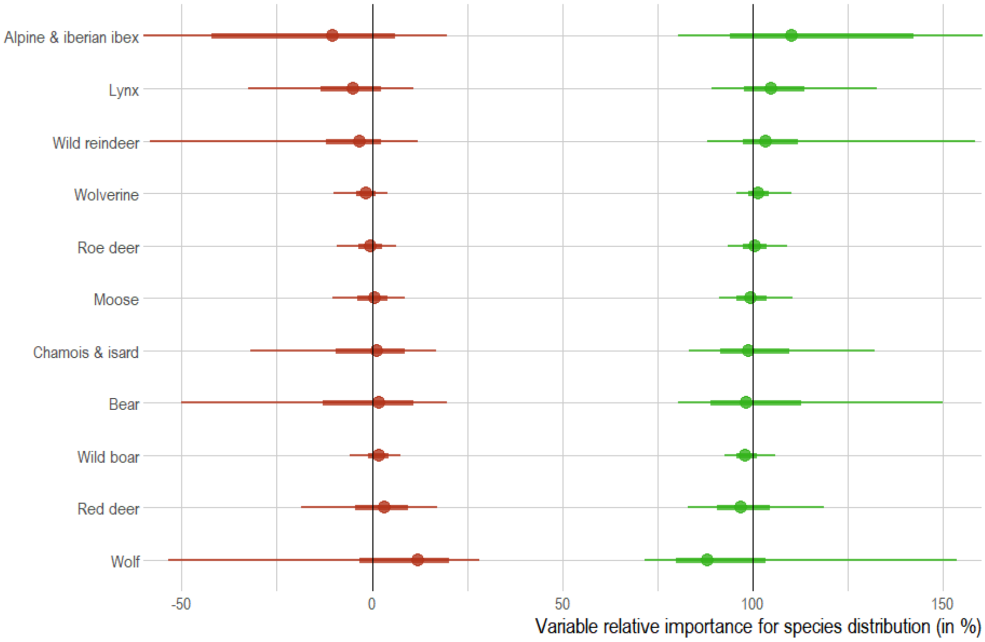


***Figure 3**: Relative importance for model fit (in percentage) of anthropogenic variables (human footprint and protected area coverage; in red) and* biophysical *variables (winter and summer severity and terrain ruggedness; in green) to species distribution. Negative importance indicates a drop in the R^2 when the variable is included in the model. Points represent the median value, thick lines the 50% credible interval and thin lines the 95% credible interval.*

Finally, in Figure 4 we show that human modifications on the landscape hardly influence the area of species' potential distribution. The suitable area for most studied mammals (i.e. ibex, wild reindeer, bears, wolverines, red deer and moose) is weakly influenced by setting both human footprint and protected areas to zero. Only in the case of chamois and roe deer did we observe a strong decrease of predicted suitable area when setting the anthropogenic variables to zero (median = -13,900 and -284,400 km^2 respectively). We also observed a decrease of the predicted suitable area for wolverine, wild reindeer and ibex when removing anthropogenic effects (median = -12,900 and -6,200 km^2 respectively), due to the removal of protected areas (see Figure S1 and S2 in the Annexes). In contrast, the total predicted suitable area available for wolf, lynx and wild boar increases when anthropogenic effects are set to zero (median = 50,700, 133,400 and 131,200 km^2 respectively). These predicted gains represent 17%, 6% and 4% of the actual lynx, wolf and wild boar distributions.

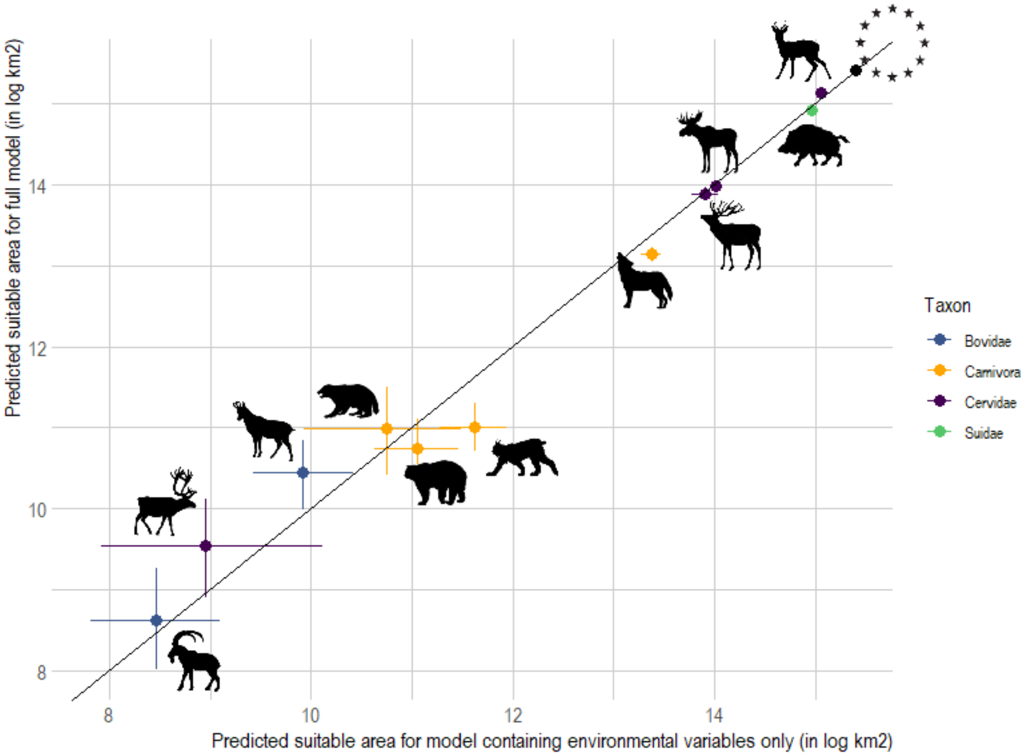


Figure 4: *Predicted environmental niche (in log in km^2) of European large mammals in the presence (y axis) and absence (x axis) of anthropogenic variables. A value below the iso-line indicates an increase in potentially suitable area when removing anthropogenic variables. Thin lines represent the 95% credible intervals. The stars representing the area of our study area, Europe. Species symbols, from bottom left, are ibex, wild reindeer, chamois, wolverine, brown bear, Eurasian lynx, wolf, red deer, moose, wild boar, roe deer.*

4. Discussion

In this study we have demonstrated that the large-scale distributions of Europe's main large mammalian species include large areas of high to very high human disturbance. Even though there is a wide distribution of high human disturbance combined with a rarity of wild places in the European landscape (Venter et al. 2016) these results show that large mammals can maintain a presence in these heavily modified multi-use landscapes. We have further shown that human disturbance and protected area coverage are only minor drivers of large mammal distributions at the continental scale. Overall, for all large mammals, our results show that the anthropogenic variables are poor predictors of species distribution compared to the other biophysical environmental variables.

Large scale studies (e.g. with a continental scope) and finer scale studies (e.g. with a sub-national scope) do not answer the same questions, and their results can apparently be in contradiction. Failure to consider scale can lead to misinterpretation of results (Johnson 1980) and conservation scientists should be careful about the scale used to answer their research questions. While our models of first order habitat selection (distribution range) suggest that anthropogenic factors such as protected area coverage and human disturbance are minor drivers of large ungulate and large carnivore distribution in Europe, results should not be generalized to higher order habitat selection at finer scales (sensu Johnson 1980). Indeed, many fine scale studies find that the presence or habitat use of large mammals is mainly negatively affected by their proximity to human infrastructure such as trails, roads or cities (for red and roe deer see D'Amico et al. 2016, Polfus et al. 2011 for moose, Lesmerises et al. 2013 for wolves, Gundersen et al. 2019 for wild reindeer, May et al. 2006 for wolverines and Pęksa & Ciach 2018 for chamois). Furthermore, studies demonstrate that species are often forced to adapt to the proximity of humans through temporal segregation (e.g. animals become primarily night active, Gaynor et al. 2018). As different ecological processes drive distributions at different scales, it is therefore not surprising that results will vary across studies at different scales. For instance, while mountain ungulates forage on steep slopes, human settlements are usually located in the valley bottoms, allowing a vertical coexistence in close proximity. Thus, topographic complexity can provide refuge areas that facilitates human wildlife proximity (Richard & Côté 2016). The Human Footprint Index is an aggregated metric of human pressure appropriate for analysis of coarse scale data like ours. Finer scale analyses of other data sets would benefit from breaking it down its component layers to explore mechanistic relationships between the different aspects of human activity and landuse.

Similar to other large-scale studies such as Belote et al. (2020) or Pacifici et al. (2020), our analysis is also limited to distributional data whose quality is highly variable and coarse, and we do not analyze effects on density, behavior or demography. Therefore, while our results document the ability of populations of ungulates and carnivores to persist and use areas in the general proximity to areas of high human footprint, this does not mean these species are not influenced by humans in other ways and at finer spatio-temporal scales. Another challenge is the lack of historical distribution data which makes inferences about causal relationships between human activities and land uses with changes in distributions and population of ungulates populations. While some attempts to reconstruct large mammals' historical distribution are made, they generally rely on current distribution (Belotte et al. 2020)

The low effect of anthropogenic variables in our models also implies a weak effect of protected areas on large mammal distributions in Europe. (for ungulates see Linnell et al. 2020, for carnivores Chapron et al. 2014). A main reason is the small size of most European protected areas relative to the spatial requirements of large mammals (for ungulates see Linnell et al. 2020, for carnivores Chapron et al. 2014). Moreover. although European protected areas have on average a lower human footprint, they are not free of human disturbance. In fact, most European protected areas permit harvesting or culling of large herbivores as well as livestock grazing, extensive agriculture and forestry (Calkoen et al. 2020, Linnell et al. 2015), and they encourage tourism. It should be noted that these disturbances are not captured by the Human Footprint Index which focuses on infrastructure, implying that the actual disturbance level of protected areas might be higher than the ones used in this analysis. Only in the case of the wolverine and the wild reindeer does protected area coverage increase the suitable area available because their actual distribution is largely located within protected areas. The mechanistic relationship between the presence of these species and protected area management is however unclear, although for both species human activity and infrastructure has been shown to have negative effects (Nellemann et al. 2000, May et al. 2012).

This demonstration of the weak effect of human footprint on species distribution compared to the effect of biophysical covariates indicates that most of the large mammals included in our study are flexible enough to adapt to the dramatic anthropogenic impacts which have occurred within their bioclimatic envelope in the European landscape during recent centuries. This is reflected by the overall generalist behavior of these species. For instance, moose seem to adapt to road presence and associated forage in their proximity (Eldegard et al. 2012), while agricultural landscapes help roe deer to supplement their diet (Abbas et al. 2011).

5. Conclusion

Our results contribute to advancing the science of human-wildlife coexistence in the heavily modified landscapes that are typical of the Anthropocene. Although several papers rightly point out that large mammals are threatened by human impacts in many parts of the world (Ripple et al. 2014, Ripple et al. 2015) we argue that the European experience demonstrates that coexistence between humans and wild large mammals at broad scales, and continental scale recovery, are both possible. We suggest that it is impossible for nature conservation authorities to rely on a land-sparing policy for large mammals because protected areas large enough to support viable populations of these space demanding species don’t exist. Ultimately, the challenge of coexistence may not be about whether species are able to cope with human modification to the landscape but whether humans are willing to share their landscape and host wildlife in their backyards (Title & Bemmel 2018). Europe has multiple layers of formal and informal institutions at continental, national and local scales that effectively manage wildlife and human-wildlife interactions and which appear to have an instrumental role in facilitating this coexistence (Linnell & Kaltenborn 2019). Overall, the results permit cautious optimism concerning the possibility for wildlife conservation in the Anthropocene.

6. Data availability statement

The dataset and scripts used to conduct all analyses presented in this manuscript are available at DOI 10.17605/OSF.IO/XV8NH. For access to the raw data concerning large ungulates distribution please refer to the DOI link: DOI 10.17605/OSF.IO/N5P2U.

7. References

Abbas, F., Morellet, N., Hewison, A. M., Merlet, J., Cargnelutti, B., Lourtet, B., ... & Verheyden, H. (2011). Landscape fragmentation generates spatial variation of diet composition and quality in a generalist herbivore. *Oecologia*, *167*(2), 401-411.

Alexander, J. S., Gopalaswamy, A. M., Shi, K., Hughes, J., & Riordan, P. (2016). Patterns of snow leopard site use in an increasingly human-dominated landscape. *PLoS One*, *11*(5).

Apollonio, M., Andersen, R., & Putman, R. (Eds.). (2010). *European ungulates and their management in the 21st century*. Cambridge University Press.

Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*, *93*(7), 1527-1539.

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological methods*, *8*(2), 129.

Beguin, J., Martino, S., Rue, H., & Cumming, S. G. (2012). Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution*, *3*(5), 921-929.

Belote, R. T., Faurby, S., Brennan, A., Carter, N. H., Dietz, M. S., Hahn, B., ... & Gage, J. Mammal species composition reveals new insights into Earth's remaining wilderness. *Frontiers in Ecology and the Environment*.Carter, N. H., & Linnell, J. D. (2016). Co-adaptation is key to coexisting with large carnivores. *Trends in Ecology & Evolution*, *31*(8), 575-578.

Carter, N. H., Shrestha, B. K., Karki, J. B., Pradhan, N. M. B., & Liu, J. (2012). Coexistence between wildlife and humans at fine spatial scales. *Proceedings of the National Academy of Sciences*, *109*(38), 15360-15365.

Chapron, G., Kaczensky, P., Linnell, J. D., von Arx, M., Huber, D., Andrén, H., ... & Balčiauskas, L. (2014). Recovery of large carnivores in Europe’s modern human-dominated landscapes. *science*, *346*(6216), 1517-1519.

Cretois, B., Linnell, J. D., Kaltenborn, B. P., & Trouwborst, A. (2019). What form of human-wildlife coexistence is mandated by legislation? A comparative analysis of international and national instruments. *Biodiversity and Conservation*, *28*(7), 1729-1741.

D'Amico, M., Périquet, S., Román, J., & Revilla, E. (2016). Road avoidance responses determine the impact of heterogeneous road networks at a regional scale. *Journal of Applied Ecology*, *53*(1), 181-190.

Díaz, S., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., ... & Chan, K. (2020). Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

Dietz, A. J., Kuenzer, C., & Dech, S. (2015). Global SnowPack: a new set of snow cover parameters for studying status and dynamics of the planetary snow cover extent. *Remote Sensing Letters*, *6*(11), 844-853.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46.

Eldegard, K., Lyngved, J. T., & Hjeljord, O. (2012). Coping in a human-dominated landscape: trade-off between foraging and keeping away from roads by moose (Alces alces). *European journal of wildlife research*, *58*(6), 969-979.

Gaston, K. J., Jackson, S. F., Nagy, A., Cantú-Salazar, L., & Johnson, M. (2008). Protected areas in Europe: principle and practice. *Annals of the New York Academy of Sciences*, *1134*(1), 97-119.

Gaynor, K. M., Hojnowski, C. E., Carter, N. H., & Brashares, J. S. (2018). The influence of human disturbance on wildlife nocturnality. *Science*, *360*(6394), 1232-1235.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters*, *8*(9), 993-1009.

Gundersen, V., Vistad, O. I., Panzacchi, M., Strand, O., & van Moorter, B. (2019). Large-scale segregation of tourists and wild reindeer in three Norwegian national parks: Management implications. *Tourism Management*, *75*, 22-33.

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

Leblond, M., Dussault, C., & Ouellet, J. P. (2010). What drives fine-scale movements of large herbivores? A case study using moose. *Ecography*, *33*(6), 1102-1112.

Leroux, S. J., Krawchuk, M. A., Schmiegelow, F., Cumming, S. G., Lisgo, K., Anderson, L. G., & Petkova, M. (2010). Global protected areas and IUCN designations: Do the categories match the conditions?. *Biological conservation*, *143*(3), 609-616.

Lesmerises, F., Dussault, C., & St-Laurent, M. H. (2013). Major roadwork impacts the space use behaviour of gray wolf. *Landscape and urban planning*, *112*, 18-25.

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1-25.

Linnell, J. D., Cretois, B., Nilsen, E. B., Rolandsen, C. M., Solberg, E. J., Veiberg, V., ... & Kaltenborn, B. (2020). The challenges and opportunities of coexisting with wild ungulates in the human-dominated landscapes of Europe's Anthropocene. *Biological Conservation*, *244*, 108500.

Linnell, J.D.C. & Kaltenborn, B. (2019) Institutions for achieving human-wildlife coexistence. The case of large herbivores and large carnivores in Europe. *Human wildlife interactions: turning conflict into coexistence* (eds B. FRank, J.A. Glikman & S. Marchini), pp. 288-310. Cambridge University Press, Cambridge, United Kingdom.

Linnell, J. D., Kaczensky, P., Wotschikowsky, U., Lescureux, N., & Boitani, L. (2015). Framing the relationship between people and nature in the context of European conservation. *Conservation Biology*, *29*(4), 978-985.

Linnell, J. D. C., & Zachos, F. E. (2010). Status and distribution patterns of European ungulates: genetics, population history and conservation. *Ungulate management in Europe: problems and practices*, 12-53.

May, R., Gorini, L., Van Dijk, J., Brøseth, H., Linnell, J. D. C., & Landa, A. (2012). Habitat characteristics associated with wolverine den sites in N orwegian multiple-use landscapes. *Journal of Zoology*, *287*(3), 195-204.

May, R., Landa, A., van Dijk, J., Linnell, J. D., & Andersen, R. (2006). Impact of infrastructure on habitat selection of wolverines Gulo gulo. *Wildlife Biology*, *12*(3), 285-295.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, *4*(2), 133-142.

Nellemann, C., Støen, O. G., Kindberg, J., Swenson, J. E., Vistnes, I., Ericsson, G., ... & Ordiz, A. (2007). Terrain use by an expanding brown bear population in relation to age, recreational resorts and human settlements. *Biological Conservation*, *138*(1-2), 157-165.

Nellemann, C., Jordhøy, P., Støen, O. G., & Strand, O. (2000). Cumulative impacts of tourist resorts on wild reindeer (Rangifer tarandus tarandus) during winter. *Arctic*, 9-17.

Pacifici, M., Rondinini, C., Rhodes, J. R., Burbidge, A. A., Cristiano, A., Watson, J. E., ... & Di Marco, M. (2020). Global correlates of range contractions and expansions in terrestrial mammals. *Nature Communications*, *11*(1), 1-9.

Pęksa, Ł., & Ciach, M. (2018). Daytime activity budget of an alpine ungulate (Tatra chamois Rupicapra rupicapra tatrica): influence of herd size, sex, weather and human disturbance. *Mammal Research*, *63*(4), 443-453.

Phalan, B., Onial, M., Balmford, A. & Green, R.E. (2011) Reconciling food production and biodiversity conservation: land sharing and land sparing compared. *Science,* **333,** 1289-1291.

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187), 1246752.

Polfus, J. L., Hebblewhite, M., & Heinemeyer, K. (2011). Identifying indirect habitat loss and avoidance of human infrastructure by northern mountain woodland caribou. *Biological Conservation*, *144*(11), 2637-2646.

Redpath, S. M., Young, J., Evely, A., Adams, W. M., Sutherland, W. J., Whitehouse, A., ... & Gutierrez, R. J. (2013). Understanding and managing conservation conflicts. *Trends in ecology & evolution*, *28*(2), 100-109.

Richard, J. H., & Côté, S. D. (2016). Space use analyses suggest avoidance of a ski area by mountain goats. *The Journal of Wildlife Management*, *80*(3), 387-395.

Ripple, W. J., Newsome, T. M., Wolf, C., Dirzo, R., Everatt, K. T., Galetti, M., ... & Macdonald, D. W. (2015). Collapse of the world's largest herbivores. *Science advances*, *1*(4), e1400103.

Ripple, W. J., Estes, J. A., Beschta, R. L., Wilmers, C. C., Ritchie, E. G., Hebblewhite, M., ... & Schmitz, O. J. (2014). Status and ecological effects of the world's largest carnivores. *Science*, *343*(6167), 1241484.

Svenning, J. C., Fløjgaard, C., & Baselga, A. (2011). Climate, history and neutrality as drivers of mammal beta diversity in Europe: insights from multiscale deconstruction. *Journal of Animal ecology*, *80*(2), 393-402.

Tattersall, G. J., Sinclair, B. J., Withers, P. C., Fields, P. A., Seebacher, F., Cooper, C. E., & Maloney, S. K. (2012). Coping with thermal challenges: physiological adaptations to environmental temperatures. *Comprehensive Physiology*, *2*(3), 2151-2202.

Title, P. O., & Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, *41*(2), 291-307.

Tucker, M. A., Santini, L., Carbone, C., & Mueller, T. (2020). Mammal population densities at a global scale are higher in human-modified areas. *Ecography*.

Tucker, M. A., Böhning-Gaese, K., Fagan, W. F., Fryxell, J. M., Van Moorter, B., Alberts, S. C., ... & Bartlam-Brooks, H. (2018). Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. *Science*, *359*(6374), 466-469.

van Beeck Calkoen, S. T., Mühlbauer, L., Andrén, H., Apollonio, M., Balčiauskas, L., Belotti, E., ... & Hetherington, D. (2020). Ungulate management in European national parks: why a more integrated European policy is needed. *Journal of environmental management*, *260*, 110068.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., ... & Levy, M. A. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature communications*, *7*(1), 1-11.

Supplementary material for:

Coexistence of dense human population and large mammals is possible in Europe’s anthropogenic landscape

This document includes:

Supplementary material and method

References

Table S1: Pearson correlation coefficients between covariates

Table S2: Models’ coefficients

Table S3: Table of marginal R2

Table S4: Table of conditional R2

Figure S1: Environmental niche of European large mammals in presence (y axis) and absence (x axis) of human footprint variables in log km^2. Protected area coverage at normal values.

Figure S2: Environmental niche of European large mammals in presence (y axis) and absence (x axis) of human footprint variables in log km^2. Protected area coverage at maximal value across the entire landscape.

Supplementary material and method

Penalized Complexity prior

When we first made the models we found σ_{CAR} to be very large, letting the random effect u_i the freedom the be very large as well, possibly resulting in overfitting[1]. To reduce the risk of overfit we used a penalized complexity prior on σ_{CAR} reducing its probability to be very large. We specified the PC prior as follow:

$$Prob(\sigma_{CAR} > 0.1) = 0.001$$

Model checking

Logistic regression requires 1) the dependent variable to be binary and ordinal logistic regression requires the dependent variable to be ordinal, 2) the observations to be independent of each other, 3) little or no multicollinearity among the independent variables and 4) linearity of independent variables and log odds.

The first assumption is fulfilled as presence / absence data are binary. We account for the second assumption adding the spatial term in the models, removing the pseudo-replication due to the spatial autocorrelation between observations. Multicollinearity comes from the mathematical problem formulated below:

$$Y = \beta 1 + \beta 2 * A + \beta 3 * B \quad eq.1$$

Where Y is a response variable, $\beta 1$ the intercept and $\beta 2$ and $\beta 3$ the coefficients for the covariates A and B respectively. If A and B are highly correlated, we can approximate eq. 1 by eq. 2:

$$Y = \beta 1 + \beta (A + B) \quad eq.2$$

This result in highly uncertain $\beta 2$ and $\beta 3$ coefficients[2].

In our study we account for multicollinearity by interpreting the combine effect of environmental variables instead of the individual coefficients, resolving the problem formulated. Finally, we checked the fourth assumption by plotting the residuals against each variables to evaluate the trend in the residuals (Fig S1 – S11). A non-linear trend would indicate that the model do not respect the assumption.

We also evaluated the predictive by computing the Conditional Predictive Ordinal (CPO) which is a leave-one-out cross validation score for each observation (Fig S12 – S22). A CPO_i close to one implies a good fit for observation *i*.

References

1. Zuur, A. F., Ieno E. N. (2018). Beginner's Guide to Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA, Volume II. Highland Statistics Ltd
2. McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

Table S1: Pearson correlation coefficients between covariates

	Winter severity	Summer severity	Terrain ruggedness	Human footprint
Summer severity	-0.71			
Terrain ruggedness	0.14	-0.16		
Human footprint	-0.55	0.3	-0.2	
Protected area coverage	-0.01	-0.02	0.21	-0.15

Table S2: Models coefficients

Variable	mean	0.025% credible interval	97.5% credible interval
Bear			
Intercept	-5.393	-5.901	-4.886
Winter severity	2.344	1.845	2.843
Winter severity^2	-0.529	-0.731	-0.328
Terrain ruggedness	1.096	0.769	1.422
Terrain ruggedness^2	-0.215	-0.309	-0.121
Summer severity	-0.143	-0.662	0.376
Summer severity^2	-0.549	-0.807	-0.29
Human Footprint	-0.393	-0.583	-0.203
Human Footprint^2	-0.031	-0.175	0.112
Protected area	0.03	-0.076	0.137
Alpine and Pyrenean chamois			
Intercept	-6.605	-7.109	-6.102
Winter severity	1.728	1.216	2.239
Winter severity^2	-0.765	-1.032	-0.499
Terrain ruggedness	2.03	1.712	2.349
Terrain ruggedness^2	-0.191	-0.275	-0.107
Summer severity	0.319	-0.19	0.828
Summer severity^2	-0.074	-0.329	0.18
Human Footprint	0.32	0.082	0.559
Human Footprint^2	-0.06	-0.183	0.063
Protected area	0.154	0.031	0.278
Ibex Tot			
Intercept	-8.7	-9.432	-7.967
Winter severity	0.538	-0.026	1.102
Winter severity^2	-0.047	-0.297	0.203
Terrain ruggedness	1.027	0.726	1.328
Terrain ruggedness^2	-0.033	-0.09	0.025
Summer severity	-0.477	-0.935	-0.021
Summer severity^2	0.045	-0.127	0.217
Human Footprint	0.023	-0.2	0.246
Human Footprint^2	-0.078	-0.215	0.059
Protected area	0.153	0.025	0.282
Lynx			
Intercept	-3.307	-3.588	-3.026
Winter severity	3.123	2.806	3.44

Winter severity^2	-0.818	-0.966	-0.67
Terrain ruggedness	0.625	0.367	0.882
Terrain ruggedness^2	-0.219	-0.293	-0.145
Summer severity	-1.028	-1.425	-0.631
Summer severity^2	-0.92	-1.178	-0.661
Human Footprint	-0.505	-0.645	-0.365
Human Footprint^2	0.086	-0.006	0.179
Protected area	-0.124	-0.216	-0.032
Moose			
Intercept	-2.219	-2.528	-1.909
Winter severity	6.127	5.807	6.446
Winter severity^2	-2.029	-2.176	-1.882
Terrain ruggedness	-1.264	-1.525	-1.003
Terrain ruggedness^2	0.155	0.077	0.233
Summer severity	-1.493	-1.991	-0.996
Summer severity^2	-0.943	-1.241	-0.645
Human Footprint	-0.009	-0.158	0.139
Human Footprint^2	0.078	-0.005	0.161
Protected area	-0.216	-0.296	-0.135
Red deer			
Intercept	-0.305	-0.515	-0.095
Winter severity	1.403	1.194	1.611
Winter severity^2	-1.466	-1.604	-1.328
Terrain ruggedness	0.521	0.343	0.7
Terrain ruggedness^2	-0.088	-0.146	-0.03
Summer severity	0.782	0.628	0.935
Summer severity^2	0.146	0.019	0.273
Human Footprint	-0.161	-0.281	-0.041
Human Footprint^2	-0.008	-0.069	0.052
Protected area	0.271	0.202	0.341
Roe deer			
Intercept	5.389	5.106	5.672
Winter severity	6.039	5.753	6.326
Winter severity^2	-2.387	-2.529	-2.244
Terrain ruggedness	-0.291	-0.481	-0.102
Terrain ruggedness^2	0.196	0.123	0.269
Summer severity	3.396	3.241	3.551
Summer severity^2	-0.452	-0.579	-0.325
Human Footprint	0.32	0.18	0.46
Human Footprint^2	-0.033	-0.095	0.03
Protected area	0.223	0.143	0.304

Wild boar			
Intercept	2.429	2.229	2.628
Winter severity	1.671	1.483	1.859
Winter severity^2	-1.289	-1.43	-1.148
Terrain ruggedness	0.593	0.428	0.758
Terrain ruggedness^2	0.08	0.033	0.128
Summer severity	4.276	4.128	4.424
Summer severity^2	-0.299	-0.422	-0.176
Human Footprint	-0.315	-0.427	-0.202
Human Footprint^2	-0.012	-0.067	0.043
Protected area	0.316	0.243	0.39
Wild reindeer			
Intercept	-15.861	-19.23	-12.495
Winter severity	3.897	2.156	5.638
Winter severity^2	-0.632	-1.065	-0.2
Terrain ruggedness	-0.772	-1.419	-0.127
Terrain ruggedness^2	-0.004	-0.182	0.174
Summer severity	-6.104	-10.994	-1.217
Summer severity^2	-1.555	-3.252	0.14
Human Footprint	-0.747	-1.532	0.036
Human Footprint^2	0.104	-0.402	0.609
Protected area	0.382	0.173	0.591
Wolverine			
Intercept	-14.683	-17.772	-11.596
Winter severity	2.481	1.525	3.437
Winter severity^2	-0.378	-0.656	-0.1
Terrain ruggedness	0.669	0.212	1.125
Terrain ruggedness^2	-0.191	-0.322	-0.059
Summer severity	-12.353	-17.121	-7.589
Summer severity^2	-4.414	-6.136	-2.693
Human Footprint	-0.416	-0.993	0.161
Human Footprint^2	0.136	-0.214	0.485
Protected area	0.18	0.04	0.321
Wolf			
Intercept	-4.422	-4.882	-3.962
Winter severity	2.072	1.583	2.56
Winter severity^2	-1.04	-1.341	-0.739
Terrain ruggedness	1.328	1.041	1.616
Terrain ruggedness^2	-0.21	-0.287	-0.132
Summer severity	1.048	0.606	1.489
Summer severity^2	-0.459	-0.683	-0.234
Human Footprint	-0.949	-1.124	-0.774

Human Footprint2	0.029	-0.088	0.146
Protected area	0.318	0.217	0.419

Table S3: Table of marginal R^2 for the full models, models containing environmental variables only and models containing anthropogenic variables only.

	Full model	Environmental variables only	Anthropogenic variables only
Wolverine	0.925	0.937	0.069
Wild reindeer	0.887	0.911	0.119
Moose	0.660	0.660	0.190
Wild boar	0.401	0.397	0.165
Lynx	0.326	0.337	0.038
Roe deer	0.263	0.264	0.001
Bear	0.167	0.166	0.034
Alpine & Pyrenean chamois	0.146	0.146	0.030
Alpine & Iberian ibex	0.113	0.123	0.025
Wolf	0.097	0.090	0.046
Red deer	0.048	0.047	0.004

Table S4: Table of conditional R^2 for the full models, models containing environmental variables only and models containing anthropogenic variables only.

	Full model	Environmental variables only	Anthropogenic variables only
Wolverine	0.988	0.990	0.881
Wild reindeer	0.967	0.974	0.784
Moose	0.936	0.937	0.789
Red deer	0.933	0.934	0.936
Lynx	0.921	0.923	0.926
Wild boar	0.918	0.918	0.552
Wolf	0.918	0.922	0.922
Roe deer	0.916	0.914	0.797
Bear	0.905	0.906	0.907
Alpine & Pyrenean chamois	0.858	0.860	0.893
Alpine & Iberian ibex	0.822	0.824	0.846

Figure S1: Environmental niche of European large mammals in presence (y axis) and absence (x axis) of human footprint variables in log km^2. Protected area coverage at current value.

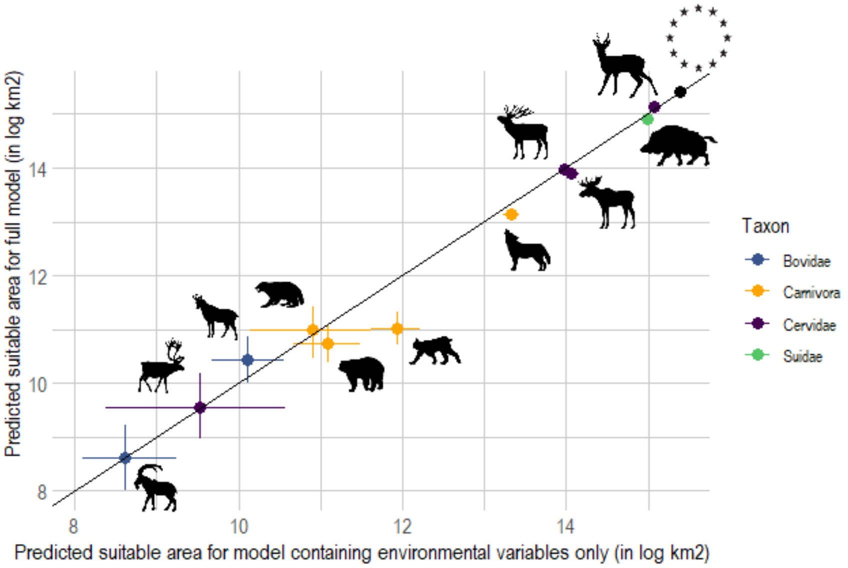
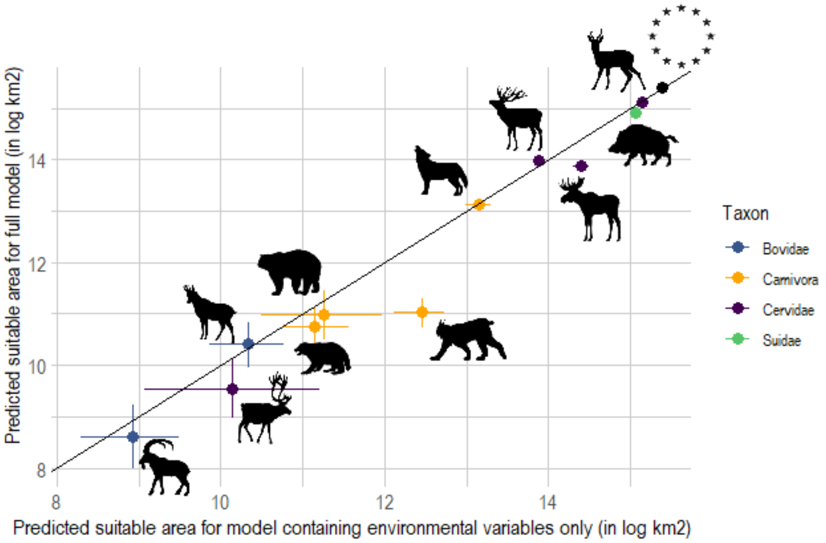


Figure S2: Environmental niche of European large mammals in presence (y axis) and absence (x axis) of human footprint variables in log km^2. Protected area coverage at maximal value across the entire landscape.



Article 4

ARTICLE IN PRESS

Spatial Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway

Jorge Sicacha-Parada [a,*], Ingelin Steinsland [a], Benjamin Cretois [b], Jan Borgelt [c]

[a] Department of Mathematical Sciences, NTNU (Norwegian University of Science and Technology), Norway
[b] Department of Geography, NTNU, Norway
[c] Department of Energy and Process Engineering, NTNU, Norway

ARTICLE INFO

Article history:
Received 1 November 2019
Received in revised form 5 March 2020
Accepted 8 April 2020
Available online xxxx

Keywords:
Bayesian modeling
Citizen Science
Variation in sampling effort
Thinned point process models
Integrated Nested Laplace Approximation (INLA)
Log-Gaussian Cox process (LGCP)

ABSTRACT

Citizen Scientists together with an increasing access to technology provide large datasets that can be used to study e.g. ecology and biodiversity. Unknown and varying sampling effort is a major issue when making inference based on citizen science data. In this paper we propose a modeling approach for accounting for variation in sampling effort due to accessibility. The paper is based on an illustrative case study using citizen science data of moose occurrence in Hedmark, Norway. The aim is to make inference about the importance of two geographical properties known to influence moose occurrence; terrain ruggedness index and solar radiation. Explanatory analysis shows that moose occurrences are overrepresented close to roads, and we use distance to roads as a proxy for accessibility. We propose a model based on a Bayesian Log-Gaussian Cox Process specification for occurrence. The model accounts for accessibility through two functional forms. This approach can be seen as a thinning process where probability of thinning, i.e. not observing, increases with increasing distances. For the moose case study distance to roads are used. Computationally efficient full Bayesian inference is performed using the Integrated Nested Laplace Approximation and the Stochastic Partial Differential Equation approach for spatial modeling. The proposed model as well as the consequences of

---

* Corresponding author.
E-mail addresses: jorge.sicacha@ntnu.no (J. Sicacha-Parada), ingelin.steinsland@ntnu.no (I. Steinsland), bernjamin.cretois@ntnu.no (B. Cretois), jan.borgelt@ntnu.no (J. Borgelt).

https://doi.org/10.1016/j.spasta.2020.100446

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

not accounting for varying sampling effort due to accessibility are studied through a simulation study based on the case study. Considerable biases are found in estimates for the effect of radiation on moose occurrence when accessibility is not considered in the model.

1. Introduction

With the expansion of technology, information and data have become readily available not only for the scientific community, but also for society in general. Citizen Science (CS), i.e. the engagement of the public in activities formerly exclusive of trained people in scientific projects, has emerged as a consequence, (Newman et al., 2012). The convenience offered by technology has encouraged people to contribute to different fields of scientific research ranging from social sciences (www.ancientlives.org, www.oldweather.org) or astronomy (www.galaxyzoo.org) to biodiversity (e.g. www.artsobservasjoner.no, www.eBird.org and www.iNaturalist.org).

According to the typology of Citizen Science introduced in Strasser et al. (2019), CS projects in biodiversity are regarded as “sensing” projects. It means that the role of volunteers is to collect information and submit it to a large database. These projects take advantage of the participants local knowledge on their environment and reach high spatial coverage. The impact of these projects can be measured in the amount of observations that are stored in their databases. For example, by September 2019, about 1.3 billion of occurrences had been reported in the global biodiversity information facility (GBIF). The Norwegian biodiversity information centre (Artsdatabanken) has about 21 million of occurrences reported. Despite being cost-efficient, easy to retrieve and its massive amount, CS data have some drawbacks. Given their “open” nature, there is no systematic sampling design to collect data, meaning citizens record observations at convenient sampling locations and times. Additionally, no scientific background is required to be part of a CS project, which implies that some species may get misidentified, (Kelling et al., 2015).

The differences in knowledge and expertise of participants in CS projects is only one of the potential sources of bias. As described in Isaac et al. (2014), the biases in the sampling processes can be classified in four groups: temporal bias, understood as varying activity of observation and reporting across time; geographical bias, meaning more reports in more convenient locations, (Mair and Ruete, 2016); uneven sampling effort per visit and differences in detectability. Preference for reporting a specific type of species constitutes another typical bias in CS sampling designs. All these biases yield in uneven sampling effort across space and time. Moreover the sampling process is not always independent of the variable intended to be measured or observed, known as preferential sampling, (Diggle et al., 2010). An issue that is not exclusive to CS records and that needs to be considered when uncertain about the independence between observation and sampling design.

Furthermore, ideally citizens record both locations where species have been observed and locations where species have been absent. This type of data is known as presence–absence data. In this case the locations are fixed and presence or absence of a species is recorded. However, CS databases in biodiversity contain mostly presence-only data. Hence, the only information given is the presence of a species in random locations whereas the rest of the landscape remains unknown. They can be actual absences or locations that have not been sampled yet. Then, there is an evident necessity of modeling CS data in a way that acknowledges the randomness of the number and the location of the observations and that accounts for different biases in the underlying sampling process.

The focus of this paper is on presence-only data and geographical bias due to accessibility. A common approach to model this data is turning some of the unobserved locations into pseudo-absences, then the available observations could be modeled as presence–absence data, (Ferrier et al.,

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

2002) and (Barbet-Massin et al., 2012) However, it does not account for the spatial autocorrelation for presences and absences across space, (Gelfand and Shirota, 2019). Arguably the most common approach for modeling presence-only data is Maxent, Phillips et al. (2009, 2006). This is an algorithmic strategy that aims to find an optimal species density subject to some constraint. Given its nature, Maxent does not account for the uncertainty of the predictions. Furthermore, it provides the relative chance of finding a species in comparison to other locations rather than a probability of presence or absence at each location. In Chakraborty et al. (2011) presence-only data is regarded as a realization of a spatial point process which, for the particular case of CS data, is subject to degradation. This approach was proven to perform better than Maxent in terms of goodness-of-fit statistics in a scenario with biased sampling.

The source of variation in sampling effort targeted in this paper is spatial bias due to differences in accessibility. It has been discussed in Gelfand and Shirota (2019) and addressed in Monsarrat et al. (2019) that studies historical large mammal records in South Africa where accessibility depends on proximity to freshwater and European settlements. There, an accessibility index is computed as the average of two functions defined as the half-normal function, characteristic of distance sampling. This functional form is also mentioned in Yuan et al. (2017) as an approach to model the probability of detection as a function of the perpendicular distance to a transect line segment.

In this paper we aim to emphasize the importance of accounting for differences in accessibility when CS data is modeled. We do it by making use of the Bayesian spatial approach proposed in Chakraborty et al. (2011) and Gelfand and Shirota (2019) to model the intensity of the point process associated to the distribution of a species. It means the observed point process is understood as the resulting process after the potential point process has been degraded by the probability of having access to each location. Our working hypothesis is that the distance to the road system is a good indicator of accessibility. Thus, we account for accessibility by making use of two functional forms introduced in Yuan et al. (2017): (a) the half-normal function that assumes an exponential decay of the probability of accessing a location as the distance to the closest road increases and (b) a semi-parametric approach that explains the decay of this probability as a function of a linear combination of I-spline basis functions, (Ramsay, 1988). These functional forms are then included as part of the models that explain the observed intensity. We refer to these models as the Varying Sampling Effort (VSE) model and the Extended Varying Sampling Effort (EVSE) model. A common goal of ecological studies is to explore the importance of geographical, climatic or biological quantities that drive the distribution of a species. Hence, we also aim to see how accounting for accessibility impacts the parameters estimates in a Bayesian spatial model, changing then the way the dynamics of a species is understood. Gelfand and Shirota (2019) uses a Markov chain Monte Carlo (MCMC) sampling for inference, which is computationally expensive. The Integrated Nested Laplace Approximation (INLA), (Rue et al., 2009) is a non-sampling approach to full Bayesian inference. INLA can also be used for spatial models based on Gaussian Matern Processes using the stochastic partial differential equation (SPDE) approach, (Lindgren et al., 2011) , also in point process modeling, (Simpson et al., 2016). We use INLA for inference, and its computational efficiency enable us to do a simulation study.

We consider an illustrative case study of CS presence data of moose (*Alces alces*) in the county of Hedmark, Norway. Moose is a large ungulate distributed across most of the Norwegian landscape. It utilizes a wide variety of environments, including forests, wetlands and farmland, (Hundertmark, 2016). The species contributes to ecosystem health parameters by providing key ecological processes such as browsing on both broad-leaved and needle-leaved trees as well as shrubs (for a review see Shipley (2010)). Moose survival and fitness are highly determined by competition for food, e.g. Messier (1991). Hence, moose tend to avoid areas dominated by steep slopes, deep and enduring snow cover as well as poor food availability. In order to proxy this knowledge , we use two explanatory variables: solar radiation (RAD) and terrain ruggedness index (TRI). Solar radiation has been shown to influence fine scale movement of moose due to its effects on air temperature, snow cover and plant phenology, (Pomeroy et al., 1998). Moose are more likely to select areas receiving higher levels of solar energy as snow cover is shallow and plant productivity higher. Ruggedness, or terrain heterogeneity also has a major role in moose distribution as a high ruggedness increase their energy expenditure, (Leblond et al., 2010).

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

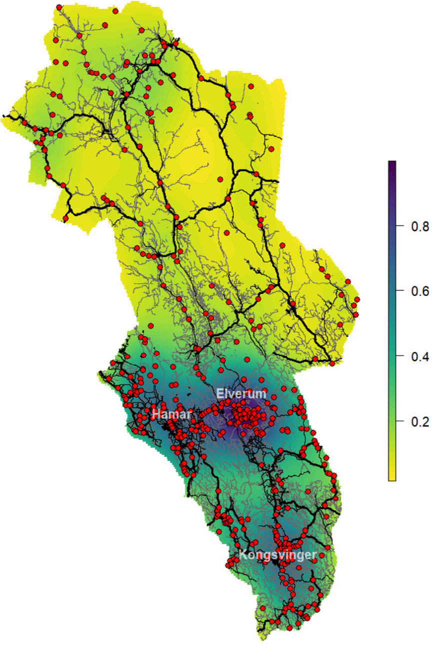


Fig. 1. Moose observations (red points) and road system (lines) in the county of Hedmark, Norway. Bold lines indicate main roads.

This paper is organized as follows: In Section 2, the dataset of the case study is introduced and explored. In Section 3, models are presented, as well as the inference method and measures for evaluating and comparing them. In Section 4, we perform a simulation study comparing the models that account for variation in sampling effort and a model not accounting for it. In Section 5 results of both the simulation study and the moose case study are shown. The paper finishes in Section 6 with the discussion of the results and concluding remarks.

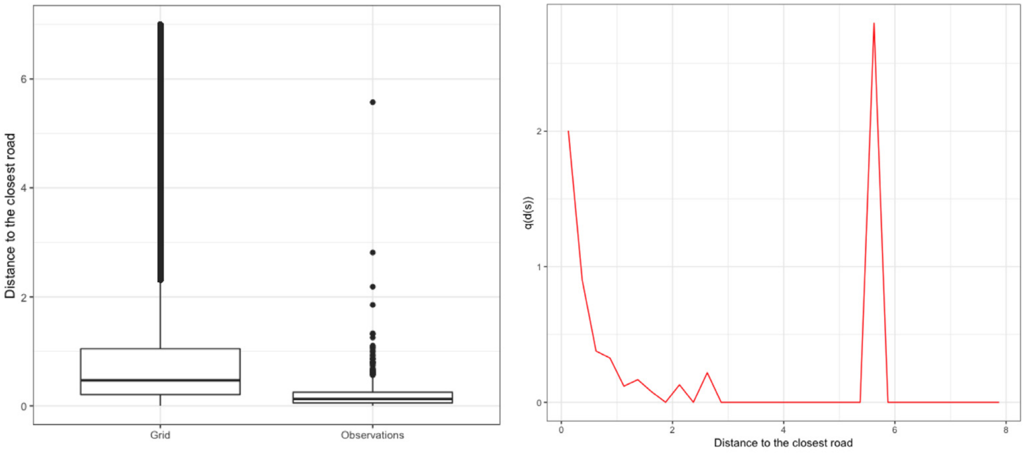
2. Case study: Moose in Hedmark and exploratory analysis

In this paper we study moose distribution using locations recorded by citizen scientists and retrieved from GBIF (https://gbif.org). It corresponds to 472 observations product of human observation from 2000 to 2019, NBIC (2019b,a), Blindheim (2019) and iNaturalist.org (2019). These observations correspond to locations of moose in the county of Hedmark, Norway, see Fig. 1. Further, we have two explanatory variables available: RAD and TRI. RAD is computed as the yearly average of the monthly solar radiation retrieved from WorldClim (http://worldclim.org/version2), Fick and Hijmans (2017). TRI was obtained from the ENVIREM dataset (https://envirem.github.io). Both variables are available at approximately 1 km × 1 km resolution, Title and Bemmels (2018).

Our working hypothesis is that spatial variation in sampling effort can be partly explained by accessibility due to distance to roads. In order to determine whether or not it happens, we used the road system of Hedmark retrieved from the spatial crowd-sourcing project OpenStreetMap (https://www.openstreetmap.org). This dataset includes a detailed network of roads that ranges from highways to footways. Fig. 1 shows the roads as well as reported moose presences in Hedmark. Most of the observations are made in southern Hedmark and near populated zones of the region, such as Hamar, Elverum and Kongsvinger, or in zones with many roads.

To explore if the observed locations are more accessible than the mass of locations in the region, we compare the citizen science dataset that contains the 472 observed points with a grid of about 400 thousand evenly distributed points. We computed the closest distance to the road network for

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.



(a) (b)

Fig. 2. (a) Boxplots of distance to the road system. Left: Dense grid of about 400 thousand points. Right: 472 reports of moose in Hedmark (b) Relationship between the observed ratio $\hat{q}(\mathbf{s}_d)$ and the distances to closest road, $\mathbf{s}_d$.

both datasets. The boxplots of these distances for each set of points are displayed in Fig. 2a. 91% of the observations reported are located less than 500 m away from a road. On the other hand, the grid has points that are more distant from the road system. The boxplots show that locations further away than 1 km are not represented in the observed point pattern. A Kolmogorov–Smirnov test was performed on the two sets of distances in order determine if these two sets of distances follow the same distribution or not. The result ($p - value < 2.2e - 16$) let us conclude that, as suspected, the sets of distances do not follow the same distribution. This is an indication of a non-random sampling process. Following our working hypothesis we explore the relationship between the distance to the closest road, $d(\mathbf{s})$ and $q(\mathbf{s})$, the probability of retaining a point located at distance $d(\mathbf{s})$ (i.e. not thinning) in the observed pattern. To proxy $q(\mathbf{s})$, we grouped both sets of distances into bins, $\mathbf{s}_d$, of width 0.25 and for each of them we computed:

$$\hat{q}(\mathbf{s}_d) = \frac{\hat{p}_{obs}(\mathbf{s}_d)}{\hat{p}_{grid}(\mathbf{s}_d)}$$

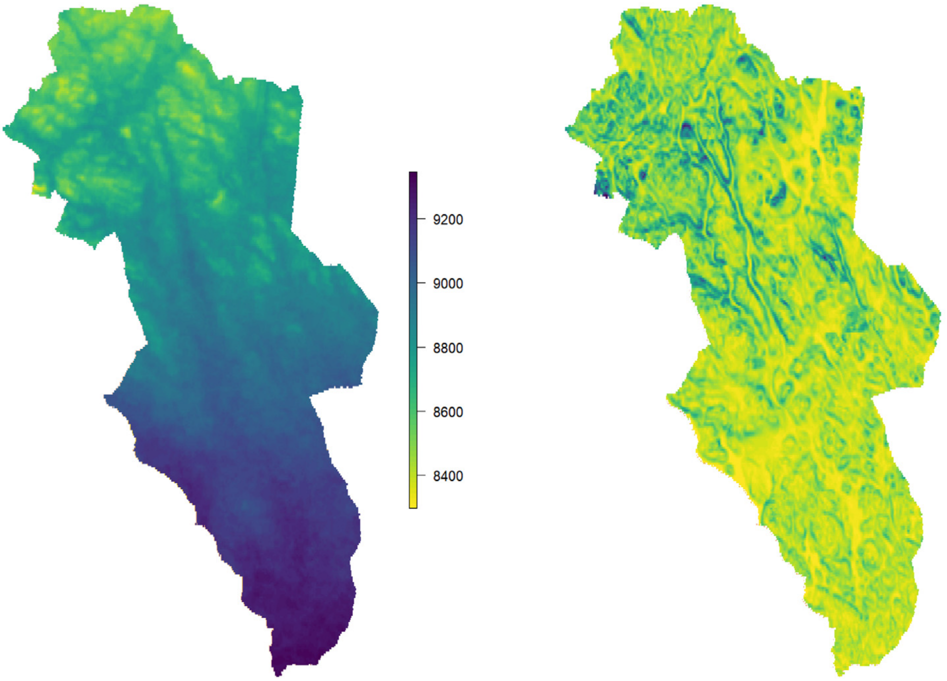
with $\hat{p}_{obs}(\mathbf{s}_d)$ and $\hat{p}_{grid}(\mathbf{s}_d)$, the proportion of points that are part of the bin $\mathbf{s}_d$ in the observed pattern and the dense grid, respectively. In Fig. 2b we observe a considerable decrease of $\hat{q}(\mathbf{s}_d)$ from $\mathbf{s}_d = [0, 0.25]$ to $\mathbf{s}_d = (1.5, 1.75]$. After this distance, $\hat{q}(\mathbf{s}_d)$ becomes 0, except for $\mathbf{s}_d = \{(2, 2.25]; (2.5, 2.75]; (5.5, 5.75]\}$ where few observations were reported.

According to the shape of $\hat{q}(\mathbf{s}_d)$ obtained from our sample, an exponential decay function as the one introduced in Yuan et al. (2017) arguably describes well the relationship between $d(\mathbf{s})$ and $q(\mathbf{s}_d)$. In addition to it, a semi-parametric approach also presented in Yuan et al. (2017) could be used. Both approaches are explained in more detail in Sections 3.1.2 and 3.1.3.

3. Modeling and inference approach

In this section we introduce three models that will be fitted and compared. They are based on the specification of a Log-Gaussian Cox Process. The first of them, the naive model, does not account for any difference in accessibility, while the second and third model account for accessibility as a potential source of variation in sampling effort. Then, we briefly describe the inference methods we will use. Finally, we introduce the criteria to assess and compare these models.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.



(a) (b)

Fig. 3. (a) Solar Radiation (RAD) and (b) Terrain Ruggedness Index (TRI) in the county of Hedmark, Norway.

3.1. Models

3.1.1. Naive model

The observed data are regarded as a realization of a point process. It means both the number of points and their locations are random. The intensity measure, understood as the mean number of points per area unit, is the variable we are interested in modeling. In what follows, we will assume the observed point pattern is a realization of an inhomogeneous Poisson Process (NHPP), Illian et al. (2008), over the region $D \subset \mathbb{R}^2$. Thus, the number of points in D is assumed to be random and to have a Poisson distribution with mean $\int_D \lambda(x)dx$. We assume the point process is a Log-Gaussian Cox Process (LGCP). Hence, $\lambda(\mathbf{s})$, $\mathbf{s} \in D$ can be expressed as:

$$\log(\lambda(\mathbf{s})) = \mathbf{x}^T(\mathbf{s})\beta + \omega(\mathbf{s}) \tag{1}$$

with $\mathbf{x}(\mathbf{s})$ a set of spatially-referenced covariates and $\omega(\mathbf{s})$ a zero-mean Gaussian process that accounts for residual spatial autocorrelation between locations in D. For our case study the set of spatial covariates $\mathbf{x}(\mathbf{s})$ are: TRI and RAD, displayed in Fig. 3. A flexible family of covariance functions is the Matérn class:

$$\frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa\|s_i - s_j\|)^{\nu}K_{\nu}(\kappa\|s_i - s_j\|) \tag{2}$$

with $\|s_i - s_j\|$ the Euclidean distance between two locations $s_i, s_j \in D$. σ^2 stands for the marginal variance, and K_ν represents the modified Bessel function of the second kind and order $\nu > 0$. ν is the parameter that determines the degree of smoothness of the process, while $\kappa > 0$ is a scaling parameter.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

3.1.2. Variation in sampling effort (VSE) model

Degeneration of the point process has to be considered in the model. We associate it to a thinned intensity. That is, we now assume that the intensity of the observed point process is $\lambda(\mathbf{s})q(\mathbf{s})$ with $\lambda(\mathbf{s})$ the intensity modeled in the naive model, named in Chakraborty et al. (2011) as the potential intensity and $q(\mathbf{s})$ the thinning factor which ranges between 0 and 1, with 0 representing total degradation and 1 no degradation. In our application, the degradation is associated to accessibility based on distances to a road network. Thus, as $d(\mathbf{s})$ approaches 0, $q(d(\mathbf{s}))$ approaches 1.

The way $q(\mathbf{s})$ can be specified is still an open question, and several alternatives are available, depending on the sources of variation in sampling effort that are considered in the model. For example, in the case of moose distribution in Hedmark, $q(\mathbf{s})$ could be associated to accessibility to the road system, (Gelfand and Shirota, 2019), to populated areas and freshwater, (Monsarrat et al., 2019), or land transformation, (Chakraborty et al., 2011). As pointed out in Yuan et al. (2017), in case $q(\mathbf{s})$ is not log-linear, the estimation of the parameters is not part of the latent Gaussian model framework of INLA. Thus, following the half normal detection function in distance sampling, (Yuan et al., 2017), we aim to account for differences in accessibility by making use of the functional form:

$$q(\mathbf{s}) = \exp(-\zeta \cdot d(\mathbf{s})^2/2); \quad \zeta > 0 \tag{3}$$

where ζ is a scale parameter and $d(\mathbf{s})$ is the closest distance from location $\mathbf{s}$ to the road system. Thus, the model we propose, which accounts for differences in accessibility is:

$$\log(\lambda(\mathbf{s})q(\mathbf{s})) = \mathbf{x}^T(\mathbf{s})\beta + \omega(\mathbf{s}) + \log(q(\mathbf{s})) \tag{4}$$

This model requires that the variables that are used to explain $q(\mathbf{s})$, in our application distance to the road system, are available at every $\mathbf{s} \in D$.

3.1.3. Extended variation in sampling effort model (EVSE)

Even if the VSE model accounts for variation in sampling effort, the functional form of $q(\mathbf{s})$ does not offer enough flexibility in situations with thinning processes that do not follow an exponential functional form. A natural, convenient way of overcoming this issue and still keeping a log-linear relationship between $d(\mathbf{s})$ and $q(\mathbf{s})$, is by means of a non-parametric approach. We can specify $-\log(q(\mathbf{s}))$ as a linear combination of basis functions as proposed in Yuan et al. (2017). In order to guarantee the monotonicity of $-\log(q(\mathbf{s}))$, we should use a basis of monotone functions, $B_k(\mathbf{s}), k = 1, \ldots, p$ in the linear combination:

$$-\log(q(\mathbf{s})) = \sum_{k=1}^{p} \zeta_k B_k(\mathbf{s}) \tag{5}$$

with ζ_k a set of parameters constrained to be positive, (Yuan et al., 2017) and (Ramsay, 1988). Since this specification of $q(\mathbf{s})$ is only implemented in INLA for independent ζ_k, p should not be more than 2 or 3. Otherwise the resulting $q(\mathbf{s})$ would not be smooth, (Yuan et al., 2017). A graphical overview of the relationship between the basis function $B_k(\mathbf{s})$ and $q(\mathbf{s})$ is available in Appendix A.

3.1.4. Prior specification

The parameter ν in the Matérn covariance function (2) is fixed to be 1. On the other hand, the interest is put on the spatial range ρ and on σ, with ρ related to κ in (2) through $\rho = \sqrt{8}/\kappa$. These two parameters are specified by making use of PC priors, (Fuglstad et al., 2019). In this case we set $P(\rho < 15) = 0.05$ and $P(\sigma > 1) = 0.05$. It means that under this prior specification a standard deviation greater than 1 is regarded as large, while a spatial range less than 15 is considered unlikely. The parameters in $\boldsymbol{\beta}$ have Normal prior with mean 0 and precision 0.01. Finally, let $\zeta = \exp(\theta)$. For the hyperparameter θ a Normal prior distribution with mean 1 and precision 0.05 is specified. In (5), let $\zeta_k = \exp(\theta_k)$, $k = 1, \ldots, p$. Each θ_k has a normal prior with mean 1 and precision 0.05.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

3.2. Inference and computational approach

The models introduced in Section 3.1 will be fitted making use of the Integrated Nested Laplace Approximation (INLA), (Rue et al., 2009), the SPDE approach, (Lindgren et al., 2011), and the approach introduced in Simpson et al. (2016) for fitting spatial point processes.

3.2.1. The Integrated Nested Laplace Approximation (INLA)

The traditional approach for performing Bayesian inference for latent Gaussian models is Monte Carlo Markov Chains (MCMC). However, the Integrated Nested Laplace Approximation (INLA), (Rue et al., 2009), has emerged as a reliable alternative, (Illian et al., 2013; Humphreys et al., 2017) and (Sadykova et al., 2017). While MCMC requires considerable time to perform Bayesian inference for complex structures such as those inherent to spatial models, INLA requires less time to do the same task since, unlike MCMC which is simulation based, INLA is a deterministic algorithm, (Blangiardo and Cameletti, 2015). The aim of INLA is to produce a numerical approximation of the marginal posterior distribution of the parameters and hyperparameters of the model. In addition to its computational benefits, implementing INLA is simple by making use of the R-INLA library.

3.2.2. The SPDE approach

A useful and efficient way to represent a continuous spatial process based on a discretely indexed spatial random process is the Stochastic Partial Differential Equation (SPDE) approach, (Lindgren et al., 2011). This is based on the solution to the SPDE:

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}}(\tau\xi(\mathbf{s})) = \mathbf{W}(\mathbf{s}) \tag{6}$$

where $\mathbf{s}$ is a vector of locations in $\mathbb{R}^2$, Δ is the Laplacian. ν, $\kappa > 0$ and $\tau > 0$ are parameters that represent a control for the smoothness, scale and variance, respectively. $\mathbf{W}(\mathbf{s})$ is a Gaussian spatial white noise process. The solution for this equation, $\xi(\mathbf{s})$, is a stationary Gaussian Field with Matérn covariance function (2). This solution can be approximated through a basis function representation defined on a triangulation of the spatial domain D:

$$\xi(\mathbf{s}) = \sum_{g=1}^{G} \phi_g(\mathbf{s})\tilde{\xi}_g \tag{7}$$

where G is the total number of vertices of the triangulation, $\{\phi_g\}$ is the set of basis functions, and $\{\tilde{\xi}_g\}$ are zero-mean Gaussian distributed weights. This way of representing the Gaussian Random Field has been proven to make more efficient the fitting process. Fig. 4a displays the triangulation for the moose distribution example.

3.2.3. Approach for modeling LGCPs

The traditional way of fitting point process models is by gridding the space and then modeling the intensity on a discrete number of cells. However, this approach becomes unfeasible and computationally expensive as the number of grids increases. Given that gridding the space also implies approximating the location of the observations, it also represents a waste in information in contexts such as Citizen Science where the locations of the observations are collected with considerable precision. Since a better approximation of the continuous random field is achieved by making the size of the cells as small as possible, lattice-based methods become unfeasible as stressed in Simpson et al. (2016). The approach there introduced is especially useful in situations with uneven sampling effort since the resolution of the approximation can be locally adapted in those regions with low sampling. Some additional details of this approach are now presented.

Let $\boldsymbol{\omega}(\mathbf{s})$ be a finite-dimensional continuously specified random field defined as:

$$\boldsymbol{\omega}(\mathbf{s}) = \sum_{i=1}^{n} \omega_i\phi_i(\mathbf{s}) \tag{8}$$

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

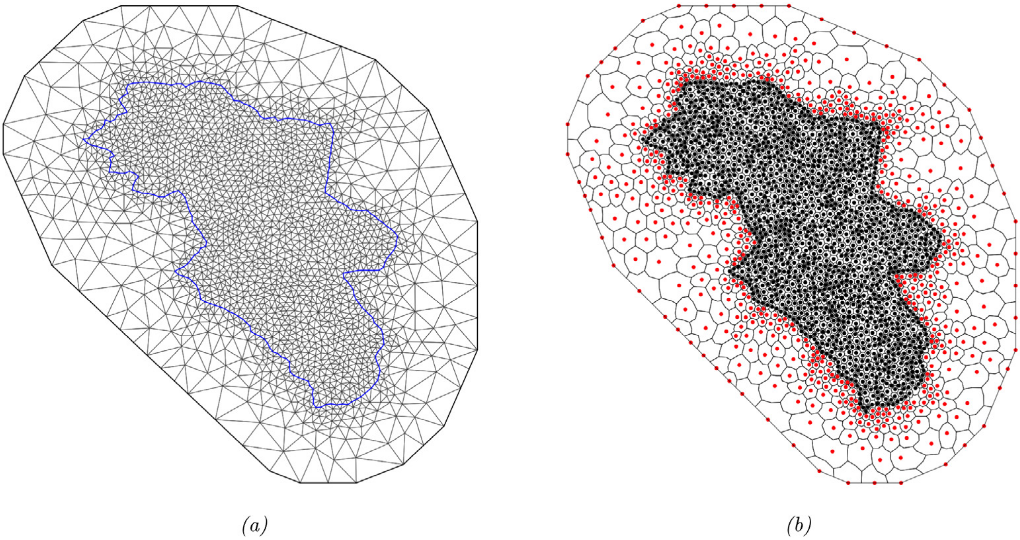


Fig. 4. (a) Triangulation of Hedmark according to the SPDE approach (b) Dual mesh for approximating the likelihood of the LGCP associated to moose distribution in Hedmark. The points are the locations $\tilde{s}_i$ in Eq. (10) and the areas of the polygons are the weights $\tilde{\alpha}_i$ in Eq. (10).

Based on this specification, the likelihood of a LGCP conditional on a realization of $\boldsymbol{\omega}$:

$$\log(\pi(\lambda(\cdot)|\boldsymbol{\omega})) = |\boldsymbol{\omega}| - \int_{\boldsymbol{\omega}} \exp(\boldsymbol{\omega}(\mathbf{s}))d\mathbf{s} + \sum_{i=1}^{N} \boldsymbol{\omega}(s_i) \tag{9}$$

can be approximated by :

$$\log(\pi(\lambda(\cdot)|\boldsymbol{\omega})) \approx C - \sum_{i=1}^{p} \tilde{\alpha}_i \exp\left\{\sum_{j=1}^{n} \omega_j \phi_j(\tilde{s}_i)\right\} + \sum_{i=1}^{N}\sum_{j=1}^{n} \omega_j \phi_j(s_i) \tag{10}$$

with $\tilde{a}_i$ and $\tilde{s}_i$ a set of deterministic weights and locations that can be obtained from a dual mesh with polygons centered at each node of the mesh. Then, $\tilde{\mathbf{s}} = \{\tilde{s}_1, \ldots, \tilde{s}_n\}$ are the nodes of the mesh and $\tilde{\mathbf{a}} = \{\tilde{a}_1, \ldots, \tilde{a}_n\}$ the areas of the polygons linked to each centroid. These polygons are constructed by making use of the midpoint rule, (Simpson et al., 2016). The dual mesh for our application is shown in Fig. 4b.

3.3. Model assessment

In order to assess and compare competing models such as the ones we are fitting in upcoming sections, we employ the Deviance Information Criterion (DIC), (Spiegelhalter et al., 2002), the Watanabe–Akaike Information Criterion (WAIC), Watanabe (2010), and the logarithm of the pseudo marginal likelihood (LPML). DIC makes use of the deviance of the model

$$D(\theta) = -2\log(p(\mathbf{y}|\boldsymbol{\theta}))$$

to compute the posterior mean deviance $\bar{D} = E_{\boldsymbol{\theta}|\mathbf{y}}(D(\boldsymbol{\theta}))$. In order to penalize the complexity of the model, the effective number of parameters,

$$p_D = E_{\boldsymbol{\theta}|\mathbf{y}}(D(\boldsymbol{\theta})) - D(E_{\boldsymbol{\theta}|\mathbf{y}}(\boldsymbol{\theta})) = \bar{D} - D(\bar{\boldsymbol{\theta}})$$

is added to $\bar{D}$. Thus,

$$DIC = \bar{D} + p_D.$$

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

The Watanabe–Akaike Information Criterion is based on the posterior predictive density, which makes it preferable to the Akaike and the deviance information criteria, since according to Gelman et al. (2014) it averages over the posterior distribution rather than conditioning on a point estimate. It is empirically computed as

$$-2\left[\sum_{i=1}^{n}\log\left(\frac{1}{S}\sum_{s=1}^{S}p(y_i|\theta^s)\right)+\sum_{i=1}^{n}V_{s=1}^{S}(\log p(y_i|\theta^s))\right]$$

with θ^s a sample of the posterior distribution and $V_{s=1}^{S}$ the sample variance

Another criterion to compare the models is LMPL,defined as:

$$LPML=\sum_{i=1}^{n}\log(CPO_i)$$

It depends on CPO_i, the Conditional Predictive Ordinate at location i, (Pettit, 1990), a measure that assesses the model performance by means of leave-one-out cross validation. It is defined as:

$$CPO_i=p(y_i^*|y_f)$$

with y_i^* the prediction of y at location i and $y_f=y_{-i}$.

4. Simulation studies

Our simulation studies aim to show: (i) the implications of not accounting for variations on sampling effort when CS data is modeled, (ii) how accounting for at least one source of variation in sampling effort can contribute to improve the inference made about the point process underlying the spatial distribution of a species and (iii) see how misspecification of $q(\mathbf{s})$ in the VSE model can affect the quality of the inference. In order to do it, we make use of the same region map, the road system in the application, the covariate Solar Radiation (RAD), given its association with the sampling process (82% of the reports are made in locations whose solar radiation is above the median solar radiation of the entire region) and its negative correlation, (−0.43), with the distance to the road system. Then a zero-mean Gaussian random field with Matérn covariance function is simulated.

A point pattern whose intensity depends on RAD is simulated. This is specified as a Log-Gaussian Cox Process, $Y(\mathbf{s})$, with log-intensity given by:

$$\log(\lambda(\mathbf{s}))=\beta_0+\beta_1 RAD(\mathbf{s})+\omega(\mathbf{s}) \tag{11}$$

It is simulated with $\beta_0=-4.25$ and $\beta_1=0.82$. The parameters of the Matérn covariance associated to the zero-mean Gaussian field, $\omega(\mathbf{s})$, are assumed to be $\nu=1$, $\kappa\approx\sqrt{8}/\rho=\sqrt{8}/34$, (Lindgren et al., 2011), with ρ the practical range, and $\sigma^2=0.7$.

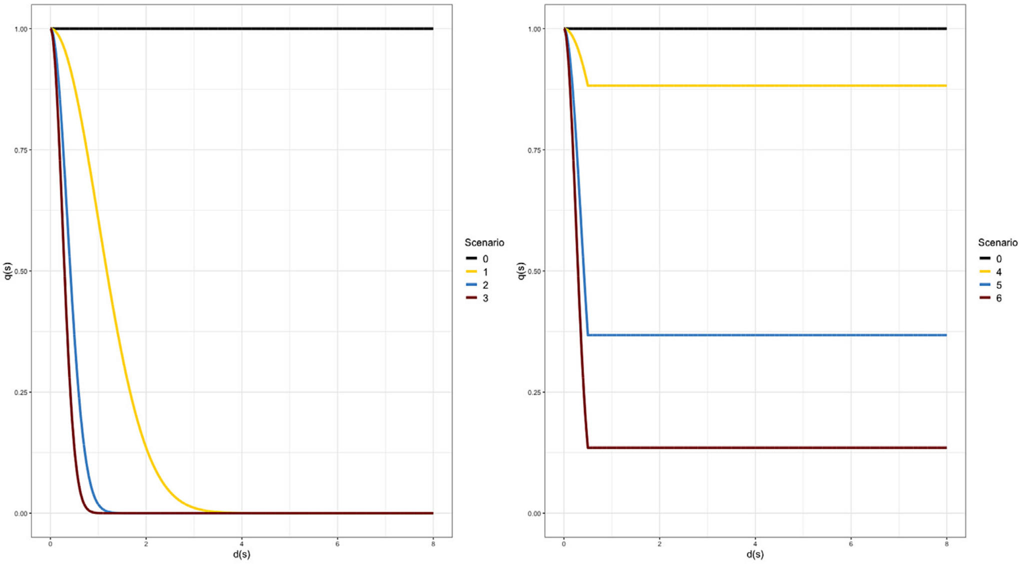
After simulating the LGCP, we thin the point pattern using two functional forms. For the first of them a point located at a distance $d(\mathbf{s})$ from the nearest road is retained with probability given by the half-normal function in (3). We create 4 scenarios based on the value of ζ: scenario 0, when $\zeta=0$; scenario 1, when $\zeta=1$; scenario 2, when $\zeta=8$ and scenario 3, when $\zeta=16$. $\zeta=0$ corresponds to the case with no thinning. The other three values of ζ represent increasing levels of thinning that result in about 13%, 39% and 50% of observations removed, respectively.

The second functional form is a mix between the half-normal function and a constant probability of retention. In this case the probability of retaining a point follows the same functional form as in (3) until a distance d_1. After this, the probability becomes constant. That is,

$$q(\mathbf{s},d_1)=\exp\left(-\frac{\zeta}{2}d^2(\mathbf{s})\right)\mathbb{1}_{[0,d_1)}(d(\mathbf{s}))+\exp\left(-\frac{\zeta}{2}d_1^2\right)\mathbb{1}_{[d_1,\infty)}(d(\mathbf{s})) \tag{12}$$

With $d_1=0.5$, three simulation scenarios were created: scenario 4, when $\zeta=1$; scenario 5, when $\zeta=8$ and scenario 6, when $\zeta=16$.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.



(a) *(b)*

Fig. 5. (a) Relationship between $d(\mathbf{s})$ and $q(\mathbf{s})$ for the simulation scenarios 0,1,2 and 3 (b) Relationship between $d(\mathbf{s})$ and $q(\mathbf{s})$ for the simulation scenarios 0,4,5 and 6.

Table 1
Simulation scenarios.

Scenario	Thinning	ζ	d_1
0	No thinning	0	–
1	Half-normal	1	–
2	Half-normal	8	–
3	Half-normal	16	–
4	Mixed	1	0,5
5	Mixed	8	0,5
6	Mixed	16	0,5

Fig. 5a displays how the functional form of $q(\mathbf{s})$ in Eq. (3) varies as ζ increases, while Fig. 5b shows $q(\mathbf{s})$ as a function of $d(\mathbf{s})$ in each of the proposed scenarios when the functional form associated to the thinning is (12) . The process of simulating a LGCP and thinning it according to ζ and d_1 was made for 100 different simulated point patterns. All the simulation scenarios are summarized in Table 1.

To assess the performance of each model for each scenario, we simulate 10 000 realizations $\{\theta^p_{jkl}\}, j\ =, 1 \dots, 10\,000$, from the posterior distribution of each parameter θ for point pattern $k = 1, \dots, 100$ in scenario $l = 0, 1, 2, 3, 4, 5, 6$. Then, the bias and the Root Mean Square Error (RMSE) for point pattern k in scenario l are computed as:

$$bias_{kl} = \frac{1}{10000}\sum_{j=1}^{10000}(\theta^p_{jkl} - \tilde{\theta})$$

$$RMSE_{kl} = \sqrt{\frac{1}{10000}\sum_{j=1}^{10000}(\theta^p_{jkl} - \tilde{\theta})^2}$$

with $\tilde{\theta}$ the actual value of parameter θ.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

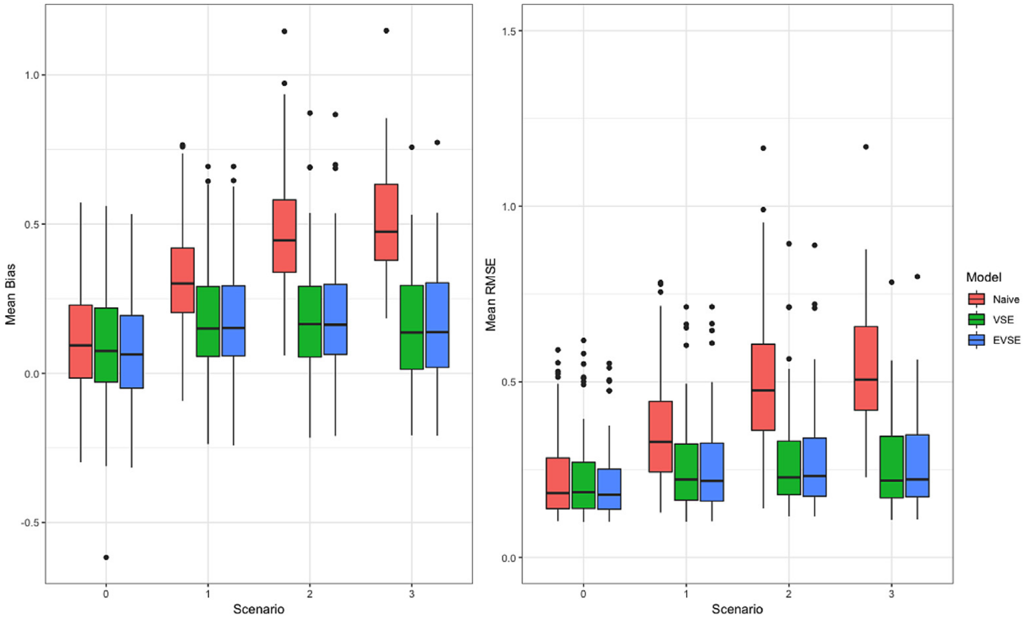


Fig. 6. Boxplots of mean bias (left) and mean RMSE (right) of β_1 for all datasets in each scenario (scenarios 0,1,2,3) and for each model.

5. Results

5.1. Simulation study

5.1.1. Results for half-normal form of $q(\mathbf{s})$

The point patterns obtained for each of the 100 simulations in each scenario described in Section 4 were fitted using the naive, the VSE and the EVSE model with $p = 3$ as suggested in Section 3. The chosen basis functions are plotted in Appendix B. The results are summarized by using measures of performance such as bias, RMSE, already introduced in Section 4, and frequentist coverage.

The parameter β_1 is the parameter of our interest. Fig. 6 presents both the mean bias and the mean RMSE at all simulated datasets for each scenario and model for this parameter. We first notice that when there is no thinning (scenario 0) the models perform similarly according to their mean bias and RMSE. However, as the original process becomes thinned (scenarios 1,2 and 3), the naive model shows poorer results than the models that account for variation in sampling effort. In scenario 3, for example, for 50% of the simulated datasets the mean RMSE for the naive model exceeds 0.5, while for less than 10% of the simulated datasets the mean RMSE is greater than 0.5 for the VSE and the EVSE models.

Table 2 introduces the mean bias and RMSE of parameters β_0, β_1, ρ and σ for the three models. The only parameters for which the bias and RMSE are not considerably different between the naive and the other two models are ρ and σ. However, ρ is clearly overestimated by all the models. The spatial variance and the range are the most difficult parameters to estimate and prior distributions that provide more information about these parameters may be useful to improve the accuracy of their estimates, (Cameletti et al., 2019) and (Bakar et al., 2015).

As an additional comparison measure we used the frequentist coverage of the equal-tailed $100(1-\alpha)$% Bayesian credible intervals for each parameter. Table 3 presents the frequentist coverage of the parameter β_1 for the three models, the results for the other parameters are available in Appendix B. The coverage of the spatial parameters does not differ between models and scenarios.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

Table 2
Mean bias and RMSE for the parameters of the naive, VSE and EVSE models in the 4 scenarios simulated. In parenthesis the standard deviation of each measure.

Scenario	Approach	β_0		β_1		ρ		σ	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0	Naive	0,132 (0,150)	0,265 (0,083)	0,109 (0,186)	0,223 (0,112)	13,698 (7,994)	17,437 (7,698)	−0,055 (0,113)	0,160 (0,054)
	VSE	0,187 (0,357)	0,309 (0,322)	0,089 (0,199)	0,223 (0,114)	13,507 (8,272)	17,410 (7,672)	0,349 (4,016)	0,559 (3,994)
	EVSE	0,192 (0,158)	0,300 (0,099)	0,082 (0,184)	0,213 (0,103)	13,849 (8,019)	17,590 (7,705)	−0,051 (0,112)	0,159 (0,053)
1	Naive	−0,157 (0,154)	0,285 (0,096)	0,310 (0,179)	0,352 (0,154)	14,480 (8,754)	18,594 (8,475)	−0,033 (0,127)	0,170 (0,058)
	VSE	0,125 (0,165)	0,277 (0,084)	0,168 (0,188)	0,258 (0,128)	14,420 (8,190)	18,641 (7,701)	−0,049 (0,121)	0,169 (0,054)
	EVSE	0,121 (0,168)	0,277 (0,081)	0,169 (0,187)	0,258 (0,128)	14,302 (8,394)	18,401 (7,914)	−0,047 (0,121)	0,167 (0,056)
2	Naive	−0,648 (0,166)	0,685 (0,162)	0,463 (0,187)	0,494 (0,177)	15,187 (9,211)	20,263 (9,121)	−0,022 (0,129)	0,179 (0,057)
	VSE	0,025 (0,163)	0,253 (0,075)	0,179 (0,196)	0,276 (0,137)	15,593 (10,625)	21,463 (12,145)	−0,075 (0,131)	0,190 (0,058)
	EVSE	−0,007 (0,166)	0,254 (0,076)	0,182 (0,196)	0,278 (0,138)	14,803 (12,403)	20,063 (13,296)	−0,074 (0,149)	0,193 (0,069)
3	Naive	−0,890 (0,168)	0,918 (0,167)	0,503 (0,181)	0,534 (0,174)	14,856 (10,104)	20,600 (10,055)	−0,016 (0,129)	0,183 (0,055)
	VSE	−0,025 (0,158)	0,252 (0,067)	0,161 (0,193)	0,271 (0,130)	15,371 (10,847)	22,397 (11,051)	−0,094 (0,135)	0,203 (0,064)
	EVSE	−0,068 (0,160)	0,259 (0,079)	0,164 (0,194)	0,272 (0,131)	15,174 (14,158)	20,900 (13,846)	−0,087 (0,151)	0,195 (0,077)

Table 3
Frequentist coverage of the equal-tailed 95% Bayesian credible interval for β_1. In parenthesis, mean length of the intervals.

Scenario	Model		
	Naive	VSE	EVSE
0	0,76 (0,49)	0,76 (0,48)	0,79 (0,49)
1	0,43 (0,55)	0,73 (0,54)	0,72 (0,54)
2	0,19 (0,63)	0,81 (0,61)	0,79 (0,61)
3	0,16 (0,67)	0,81 (0,64)	0,81 (0,64)

It is worth noting that smaller coverages are obtained for β_0 for the naive model in comparison to the other two models as the parameter ζ increases.

The model comparison methods based on the deviance and on the predictive distribution as the ones introduced in Section 3 are used to compare the results of the three models. In the scenario with $\zeta = 0$ the naive model is the true model and, as expected, it performed better than the other two models in about 40% of the simulated point patterns. This situation changes as the thinning parameter increases, the models that account for variation in sampling effort perform better than the naive one for all the simulated datasets.

5.1.2. *Results for mixed functional form of $q(\mathbf{s})$*

As explained in Section 4, we now thin differently the simulated point processes. The function $q(\mathbf{s})$ is now half-normal up to a distance d_1, where it becomes constant. We fit the resulting observations using the same three models. Fig. 7 displays the mean bias and RMSE for the three models in each scenario.

In scenarios with low values of the thinning parameter ($\zeta = 0, 1$), there are not large differences in terms of bias and RMSE for the posterior median of β_1 for the three approaches. On the other

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

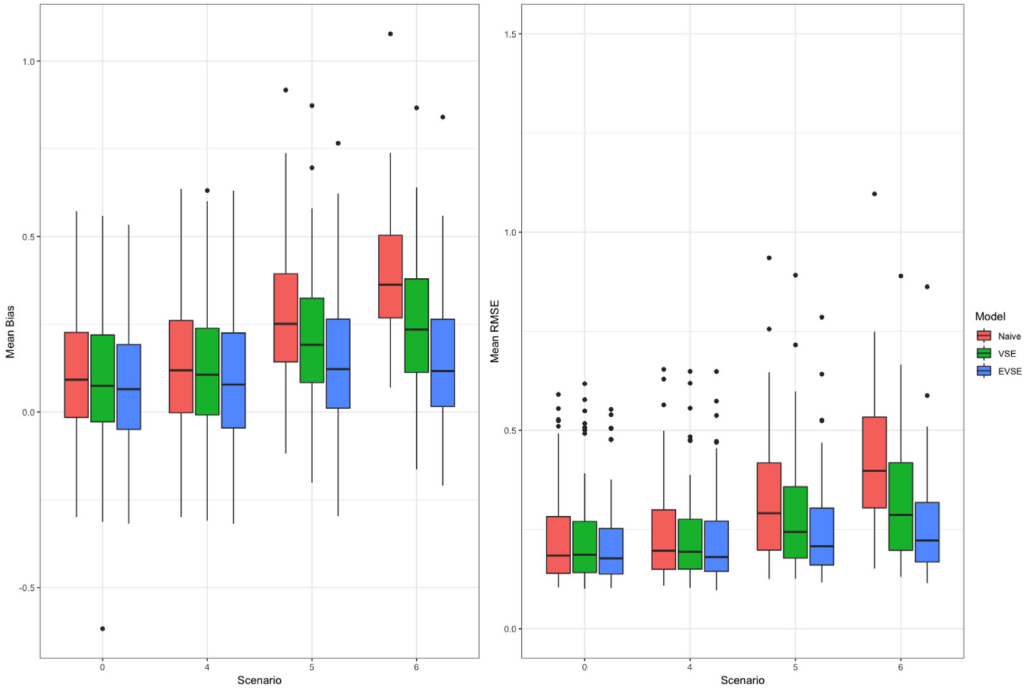


Fig. 7. Boxplots of mean bias (left) and mean RMSE (right) of β_1 for all datasets in each scenario (scenarios 0,4,5,6) and for each model.

hand, as ζ increases the differences between the three models become evident. While the EVSE model produces mean bias and RMSE consistent with scenarios of low thinning, the mean bias and RMSE of the VSE model increase, but not as much as for the naive model. Table 4 has the mean bias and RMSE of the parameters β_0, β_1, ρ and σ.

The same pattern described for the bias and RMSE of the parameter β_1 occurs for the intercept β_0. In contrast, for the spatial hyperparameters, ρ and σ, there are not considerable differences in mean bias or RMSE between the three models. As made for scenarios 0,1,2 and 3, the frequentist coverage of each parameter in each scenario was computed. In Table 5, the frequentist coverage of β_1 is reported. The frequentist coverage for the other parameters is available in Appendix B.

The frequentist coverage of β_1 is very similar between the three models when the thinning is moderate, i.e. scenarios 0 and 1. However, as more observations are removed from the original point pattern, the differences between the models become larger, with the EVSE model having about 80% of coverage, while the VSE model has less than 70% and the naive model less than 60%. Finally, in terms of DIC, WAIC and CPO, the EVSE model outperforms the other two models when the thinning of the model is high.

5.2. Results for moose distribution in Hedmark application

The models introduced in Section 3 are fitted for the dataset introduced in Section 2 . Table 6 reports the posterior mean and standard deviation of the parameters for each of these models. Terrain Ruggedness Index (TRI) is negatively related to the intensity, while Solar Radiation (RAD) has positive association with it for all the models. This suggests, as expected, that moose occurrences are more likely found in locations with higher solar radiation and where the terrain is less rough . The variability and range of the Gaussian field have right skewed posterior distributions based on their posterior medians and means. There is a difference in the posterior mean of RAD coefficient

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

Table 4
Mean bias and RMSE for the parameters of the naive and the VSE model under the 3 scenarios simulated with mixed thinning. In parenthesis the standard deviation of each measure.

Scenario	Approach	β_0		β_1		ρ		σ	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
4	Naive	0,059	0,244	0,132	0,235	13,749	17,640	−0,056	0,163
		(0,151)	(0,068)	(0,188)	(0,117)	(8,344)	(8,035)	(0,115)	(0,054)
	VSE	0,088	0,255	0,117	0,229	13,793	17,696	−0,057	0,163
		(0,157)	(0,073)	(0,188)	(0,112)	(8,311)	(7,952)	(0,114)	(0,054)
	EVSE	0,142	0,277	0,097	0,221	14,045	17,923	−0,051	0,161
		(0,162)	(0,086)	(0,187)	(0,107)	(8,377)	(8,050)	(0,115)	(0,054)
5	Naive	−0,369	0,427	0,265	0,321	14,110	18,823	−0,062	0,172
		(0,158)	(0,137)	(0,191)	(0,155)	(8,378)	(8,040)	(0,118)	(0,055)
	VSE	−0,264	0,348	0,207	0,284	14,344	19,049	−0,067	0,172
		(0,167)	(0,128)	(0,190)	(0,139)	(8,121)	(7,772)	(0,116)	(0,053)
	EVSE	−0,047	0,248	0,132	0,245	14,403	19,027	−0,068	0,172
		(0,162)	(0,077)	(0,191)	(0,119)	(8,134)	(7,758)	(0,117)	(0,055)
6	Naive	−0,733	0,763	0,391	0,428	13,587	19,145	−0,044	0,177
		(0,169)	(0,167)	(0,184)	(0,171)	(9,549)	(9,338)	(0,124)	(0,055)
	VSE	−0,465	0,514	0,247	0,315	14,155	19,905	−0,071	0,181
		(0,184)	(0,167)	(0,184)	(0,145)	(9,629)	(9,432)	(0,120)	(0,056)
	EVSE	−0,147	0,281	0,145	0,258	14,607	20,545	−0,085	0,189
		(0,164)	(0,103)	(0,191)	(0,126)	(9,920)	(9,546)	(0,126)	(0,063)

Table 5
Frequentist coverage of the equal-tailed 95% Bayesian credible interval for β_1. In parenthesis, mean length of the intervals.

Scenario	Model		
	Naive	VSE	EVSE
0	0,76 (0,49)	0,76 (0,48)	0,79 (0,49)
4	0,72 (0,50)	0,76 (0,50)	0,77 (0,50)
5	0,53 (0,56)	0,66 (0,56)	0,81 (0,56)
6	0,36 (0,62)	0,63 (0,62)	0,82 (0,61)

between the models. It is larger when differences in accessibility are not considered in the model. In addition to it, both parameters associated to the Matérn Gaussian field have lower posterior medians for the models that account for variation in sampling effort.

RAD is the most influential parameter for the three models. We see from Fig. 8 and Table 6 that the posteriors of this parameter shift considerably between the models. While the naive model has the largest posterior mean for RAD, the EVSE model has the smallest posterior mean.

The parameter ζ in the VSE model with posterior median 0.87 indicates that the observed point pattern is a thinned version of the real one, while the posterior medians of ζ_1, ζ_2 and ζ_3 seem to give more weight to the first basis function. The basis functions used for modeling $q(\mathbf{s})$ are presented in Appendix C. Fig. 9 shows the estimated relationship between distance (in kilometers) to the road system and $q(\mathbf{s})$ for the VSE and the EVSE models. According to the results of the VSE model a point located more than 3 km away from the road system can be regarded as inaccessible for citizen scientists. On the other hand, the EVSE model does not consider any location as inaccessible for citizen scientists. Instead, it assigns constant $q(s) \approx 0.05$ for locations more than 1.5 km away from the nearest road.

Fig. 10 displays the map of differences in posterior median and standard error of the logarithm of the intensity between the EVSE and the naive model. The maps with the differences in posterior median and standard error between all the models are available in Appendix C. The largest differences occur in zones that are distant to the nearest road and that have no occurrences of moose recorded. These places have lower solar radiation than the rest of the region and have considerable elevation in some locations. For the zones that are more observed, accounting for differences in

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

Table 6
Posterior summaries of the parameters of the naive and the VSE model for the moose presence data in Hedmark, Norway.

Parameter	Model														
	Naive					VSE					EVSE				
	Mean	Sd	0.025q	0.50q	0.975q	Mean	Sd	0.025q	0.50q	0.975q	Mean	Sd	0.025q	0.50q	0.975q
Intercept	−4,87	0,23	−5,32	−4,87	−4,41	−4,56	0,21	−4,97	−4,56	−4,14	−4,17	0,20	−4,57	−4,17	−3,77
TRI	−0,20	0,08	−0,35	−0,20	−0,04	−0,20	0,08	−0,35	−0,20	−0,04	−0,16	0,08	−0,32	−0,16	−0,01
RAD	1,01	0,19	0,64	1,01	1,38	0,73	0,18	0,37	0,73	1,10	0,57	0,17	0,23	0,57	0,91
ζ	–	–	–	–	–	0,88	0,21	0,52	0,87	1,33	–	–	–	–	–
ρ	39,78	9,06	26,14	38,32	61,37	37,73	7,68	25,88	36,59	55,74	36,45	7,04	25,26	35,52	52,78
σ	1,12	0,16	0,85	1,10	1,48	1,04	0,13	0,81	1,03	1,34	0,99	0,12	0,78	0,98	1,26
ζ_1	–	–	–	–	–	–	–	–	–	–	2,79	0,35	2,20	2,75	3,57
ζ_2	–	–	–	–	–	–	–	–	–	–	0,09	0,14	0,00	0,04	0,45
ζ_3	–	–	–	–	–	–	–	–	–	–	0,09	4,32	0,00	0,03	0,51

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

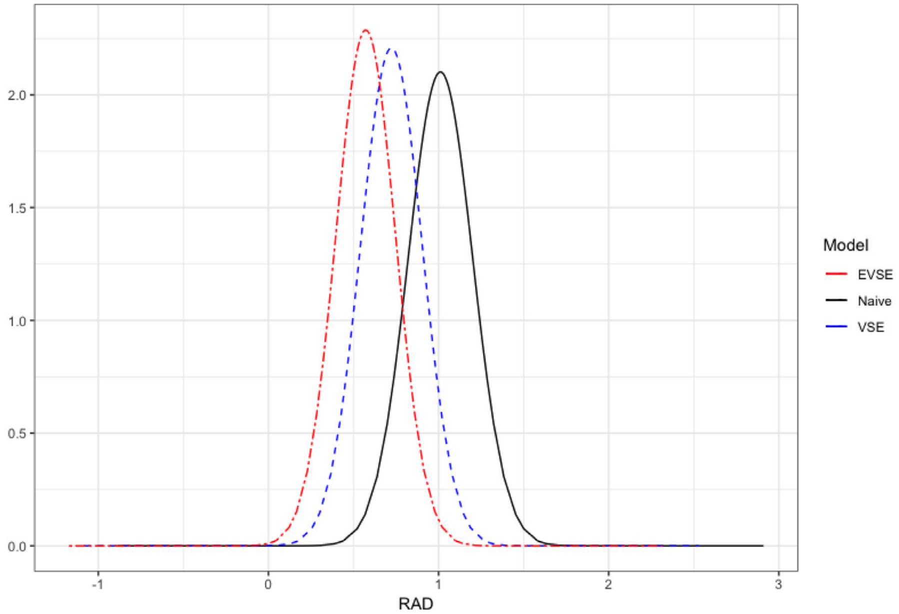


Fig. 8. Posterior density of RAD for the three models.

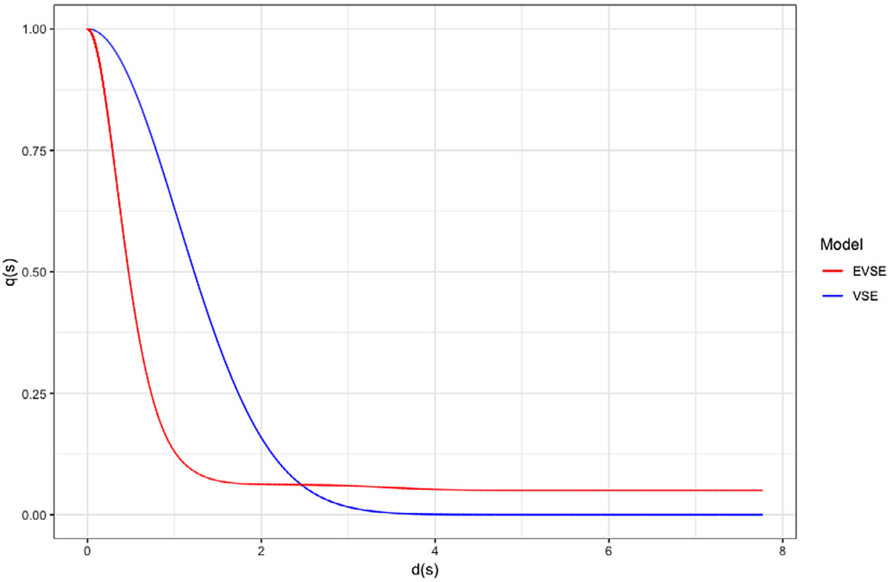
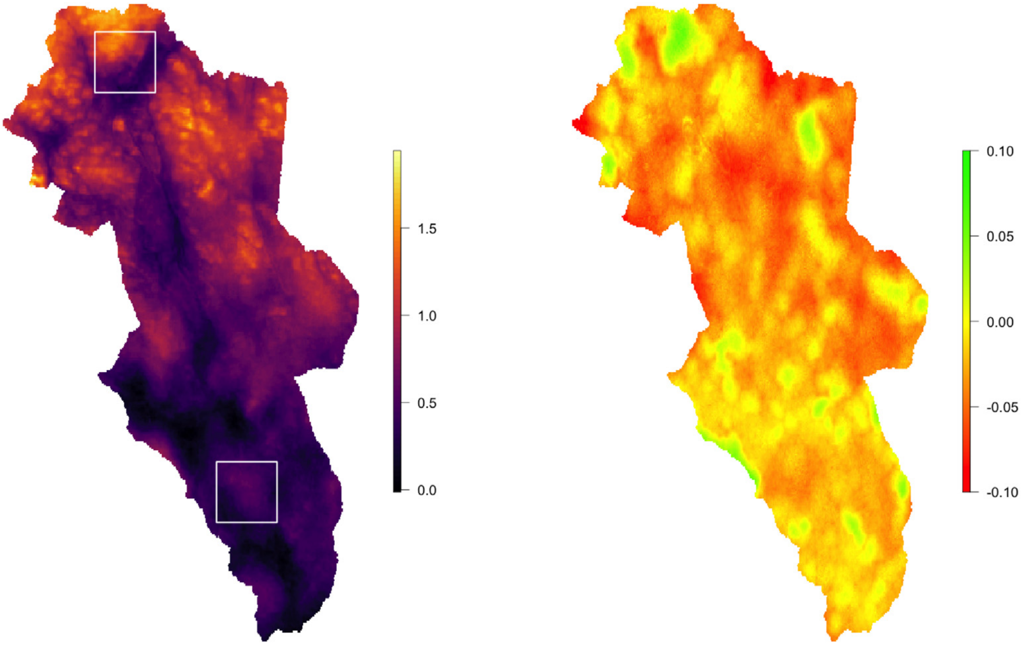


Fig. 9. Estimated relation between distance to the road system, in kilometers, and the probability of having access to location **s**.

accessibility does not affect the posterior median intensity and the uncertainty. The uncertainty is smaller for the EVSE model in most of the locations, except for some that include bodies of water such as lakes Mjøsa and Femun and national parks like Forollhogna national park.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.



(a) (b)

Fig. 10. (a) Differences in posterior median intensity and (b) differences in standard error of the posterior median intensity obtained through the VSE model and the naive model. In (a) the two squares represent the zones that are focused in Fig. 11 (south of Hedmark) and 12 (north of Hedmark).

Table 7
Comparison criteria for the naive and VSE model fitted to moose location reports.

	Model		
	Naive	VSE	EVSE
DIC	4377,51	4344,90	4265,77
WAIC	4505,36	4471,39	4400,91
LPML	−2467,61	−2446,98	−2428,182

The magnitude of the differences in the posterior median intensity between the VSE and the naive model is lower than between the EVSE and the naive model. The places with the highest differences in intensity and uncertainty are the same as between the EVSE and the naive model. The differences between the VSE and the EVSE model are considerably small. The three models are compared by making use of the DIC, the WAIC and the LPML. Table 7 introduces the value of each criterion for each model.

For the case of moose in Hedmark the results in Table 7 indicate that accounting for variation in sampling effort represents an improvement in terms of goodness of fit since both the DIC and WAIC are smaller, and the LPML is larger for the VSE and the EVSE model, with the latter showing better results in this sense than the former model.

Now we will focus on two specific zones of Hedmark to see with more detail how the posterior median and its associated uncertainty vary between the models. The two zones are bounded by a 30 km × 30 km square and are highlighted in Fig. 10. The first zone is located on the southern half of Hedmark between Kongsvinger and Hamar. It is accessible only through service roads, which are

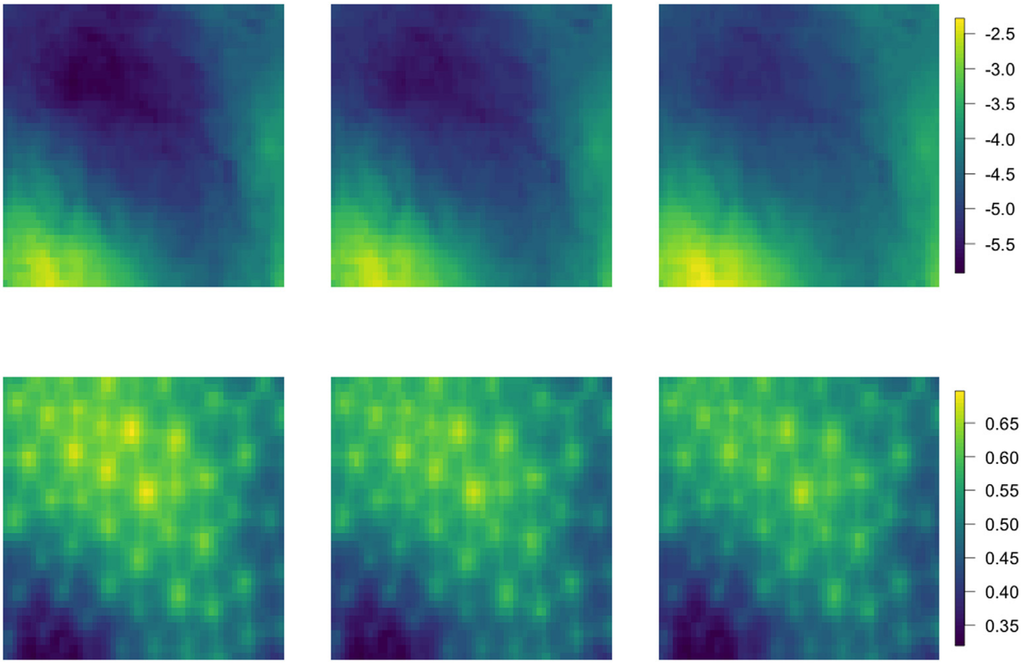


Fig. 11. Posterior median intensity (top) and associated standard error (bottom) for the naive model (left), the VSE model (middle) and the EVSE model (right) in zone 1.

not as visited as the main roads of the region, while the second square corresponds to one of the most distant zones of the region, which is located on the northern border of Hedmark. For zone 1 the posterior median intensity and its associated standard error for all the models are displayed in Fig. 11. The posterior median intensity is similar for the three models as well as the associated uncertainties. Given that the zone is regarded as highly accessible, considerable differences are not expected. In contrast, for zone 2 the EVSE model increases the intensity in most locations compared to the other two models. In terms of uncertainty the three models produce similar results. However, it becomes larger in some few zones under the VSE model, see Fig. 12.

6. Discussion and conclusions

The main goal of this paper was to highlight the importance of accounting for sources of variation in sampling effort for CS data. Bayesian spatial models that account for variation in sampling effort by including proxies for external processes that degrade the intensity of the point process have been introduced.

This paper focused on differences in accessibility across space. In the simulation studies performed in Section 4, we created scenarios where the only source of degradation for the actual point pattern was the distance to the nearest road. Two of the functional forms presented in Yuan et al. (2017) were used to link it to the intensity of the point pattern. The first of them is the half-normal function, characteristic of distance sampling. The second one is a function of a linear combination of a set of monotone functions with strictly positive coefficients. The aim of ecological studies is often to learn about the effect of covariates. The results of both the simulation study and the real data application suggest that in situations with some evidence of uneven sampling effort accounting for differences in accessibility improves performance indices, such as bias and RMSE, and model selection indices, such as DIC, WAIC and LMPL. In the scenario with no thinning on the point pattern

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

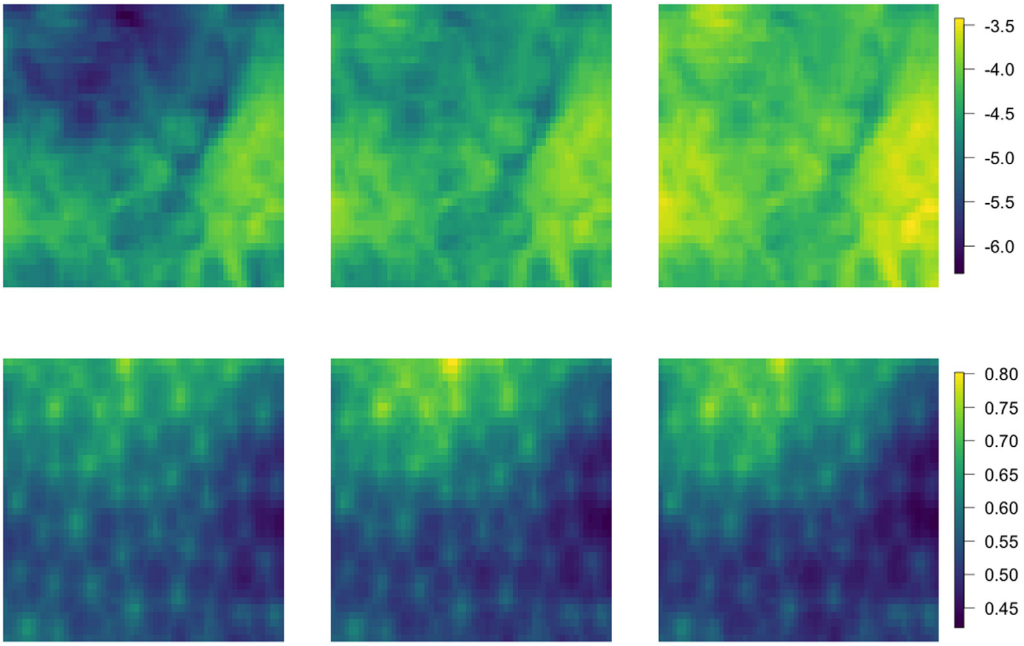


Fig. 12. Posterior median intensity (top) and associated standard error (bottom) for the naive model (left), the VSE model (middle) and the EVSE model (right) in zone 2.

due to variation in sampling effort, we found that including a term that accounts for it does not affect the quality of the inference. Furthermore, differences in the covariates posterior summaries in the simulation study showed that in cases with sampling biases the effect of an explanatory variable may be incorrectly estimated if they are not considered in the model. It is also important to note that the VSE model was proved not robust to misspecification of the relationship between $d(\mathbf{s})$ and $q(\mathbf{s})$ in scenarios with considerable thinning.

In our case study we focused on two zones of Hedmark. The large difference in intensity between the naive and the other two models in Zone 2 shows how the models that account for variation in sampling effort regard some locations on the west of this zone as possibly thinned given that they are located above 2 km away from a road and their geographical characteristics make them suitable for moose presences. The differences and the uncertainty on the north side indicate a need for increased sampling effort in this region, marking the area around Forollhogna national park. This area is one of the few mountainous areas in Norway with relatively gentle slopes and is therefore called the "friendly mountains". Moose occasionally passes through this area, however, only few CS observations have been made so far which might partly be due to a low accessibility and therefore low CS activity. In contrast, the road network in zone 1 is rather dense. Therefore, the values of $q(\mathbf{s})$ are estimated to be relatively high and the model assumes high CS activity in this area. However, the road network here is mainly composed of service roads and small tracks. Therefore, no CS observations of moose in this area might be a result of a low visiting rate of people rather than moose being absent. However, we only accounted for differences in accessibility of sampling locations in space, therefore, the habitat is predicted to be not suitable, which seems to be wrong from an ecological perspective. Accounting for differences in visits of sampling locations in time, for instance by using spatially refined information on type of road or population data could further increase modeling performance. The results highlight, that not only accessibility (e.g. roads) are important features for quantifying preferential sampling in CS data, but also how frequent sampling sites are being visited. Small service roads and hiking tracks are likely to have a lower turnover of

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

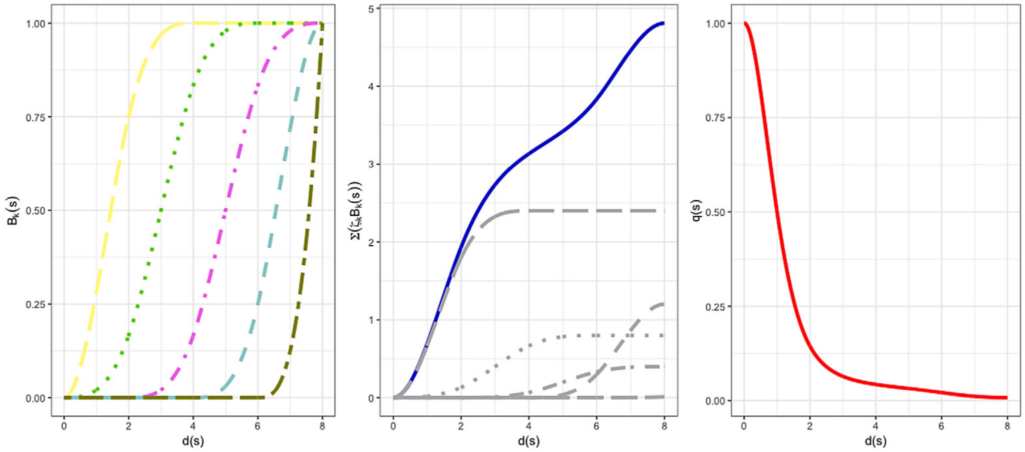


Fig. A.13. Illustration of the relationship between the basis functions and $q(\mathbf{s})$ in the EVSE model. Left, basis functions $B_k(\mathbf{s}), k = 1, \ldots, 5$. Middle, weighted basis functions by the coefficients $\zeta_k, k = 1, \ldots, 5$ (gray); linear combination of the weighted basis functions (solid, blue). Right, estimated $q(\mathbf{s})$ computed as Eq. (A.1).

visiting people than larger roads, and hence, CS more frequently register observations close to larger roads than close to small and remote roads.

An important part of the VSE and the EVSE models are the parameters ζ and $\zeta_k, k = 1, \ldots, 3$, which are necessary to determine to what extent the differences in accessibility affect the observed process. Interpreting and including them in the model is more difficult for the EVSE model given that the basis functions need to be chosen. The prior specification of the parameters that are part of the spatial Gaussian field $\omega(\mathbf{s})$ is a complex task in spatial statistics. In this paper PC priors were used as a way to incorporate prior knowledge about these parameters in a straightforward way. Alternative prior specifications using PC priors are introduced in Sørbye et al. (2019).

The VSE and EVSE models are a first step for modeling CS data in a way that accounts for its inherent sources of bias. More effort is required for e.g. extending the sampling effort model to more quantities (e.g. cell phone coverage or geographical parameters). Extending the VSE and the EVSE to more spices would be an interesting approach for learning more about citizen science sampling effort in general.

Acknowledgments

This work is part of the Transforming Citizen Science for Biodiversity project, funded by the NTNU digital transformation initiative .

Appendix A. Illustration of the EVSE model

In the EVSE model we assume

$$q(\mathbf{s}) = \exp\left(-\sum_{k=1}^{p} \zeta_k B_k(\mathbf{s})\right) \tag{A.1}$$

That is, $q(\mathbf{s})$ is assumed as a function of a linear combination of p basis functions $B_k(\mathbf{s}), k = 1, \ldots, p$. As mentioned in Section 3, $B_k(\mathbf{s}), k = 1, \ldots, p$ are a set of monotone nondecreasing functions . In addition to it, the coefficients $\zeta_k, k = 1, \ldots, p$ are constrained to be positive in order to guarantee monotonicity, (Yuan et al., 2017) and (Ramsay, 1988). Fig. A.13 illustrates, similarly as made in Yuan et al. (2017), how the relationship between these basis functions and $q(\mathbf{s})$ works .

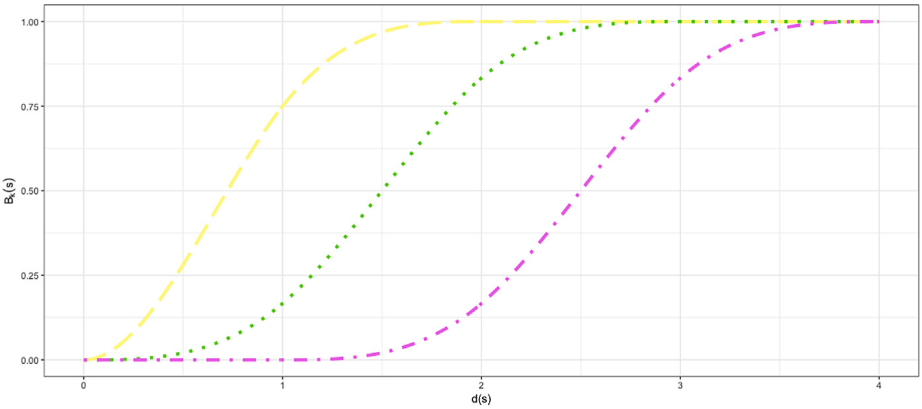


Fig. B.14. Basis functions used to fit the EVSE model in the simulation study.

Table B.8
Frequentist coverage of the equal-tailed 95% Bayesian credible interval for all the parameters in the simulations. In parenthesis, mean length of the intervals.

Parameter	Scenario	Model		
		Naive	VSE	EVSE
β_0	0	0,93 (0,74)	0,91 (0,74)	0,85 (0,75)
	1	0,92 (0,79)	0,92 (0,78)	0,94 (0,77)
	2	0,09 (0,83)	0,99 (0,8)	0,99 (0,8)
	3	0 (0,85)	0,99 (0,8)	0,99 (0,8)
	4	0,97 (0,75)	0,95 (0,75)	0,92 (0,75)
	5	0,53 (0,77)	0,73 (0,77)	0,98 (0,76)
	6	0,01 (0,8)	0,35 (0,79)	0,94 (0,78)
β_1	0	0,76 (0,49)	0,76 (0,49)	0,79 (0,49)
	1	0,43 (0,55)	0,73 (0,54)	0,72 (0,54)
	2	0,19 (0,63)	0,79 (0,61)	0,79 (0,61)
	3	0,16 (0,67)	0,81 (0,64)	0,81 (0,64)
	4	0,72 (0,5)	0,76 (0,5)	0,77 (0,5)
	5	0,53 (0,56)	0,66 (0,57)	0,81 (0,56)
	6	0,36 (0,62)	0,63 (0,62)	0,82 (0,61)
ρ	0	0,75 (39,88)	0,73 (39,56)	0,7 (39,96)
	1	0,72 (42,94)	0,75 (43,23)	0,72 (42,21)
	2	0,79 (49,35)	0,74 (47,08)	0,74 (47,08)
	3	0,87 (52,19)	0,73 (47,15)	0,73 (47,15)
	4	0,75 (40,67)	0,75 (40,75)	0,72 (40,88)
	5	0,8 (45,67)	0,79 (46,15)	0,7 (45,57)
	6	0,88 (49,07)	0,87 (50,96)	0,68 (52,26)
σ	0	0,87 (0,43)	0,86 (0,43)	0,88 (0,43)
	1	0,92 (0,47)	0,9 (0,46)	0,89 (0,45)
	2	0,92 (0,51)	0,73 (0,44)	0,73 (0,44)
	3	0,94 (0,54)	0,74 (0,43)	0,74 (0,43)
	4	0,87 (0,44)	0,87 (0,44)	0,87 (0,44)
	5	0,88 (0,46)	0,88 (0,46)	0,82 (0,46)
	6	0,91 (0,5)	0,88 (0,49)	0,76 (0,5)

Appendix B. Simulation study: Extra tables and figures

See Fig. B.14 and Table B.8.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

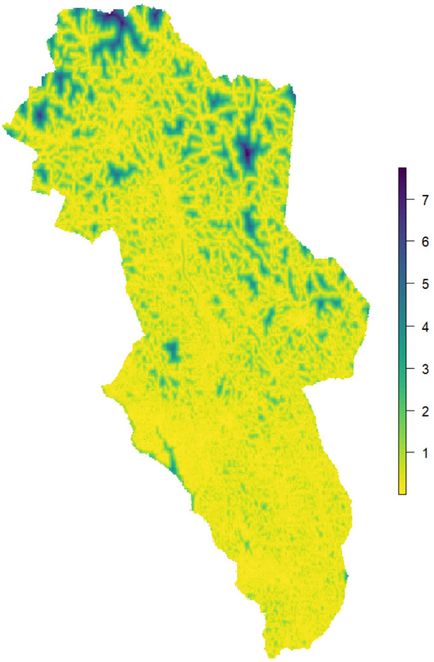


Fig. C.15. Distance to the nearest road for all locations in Hedmark.

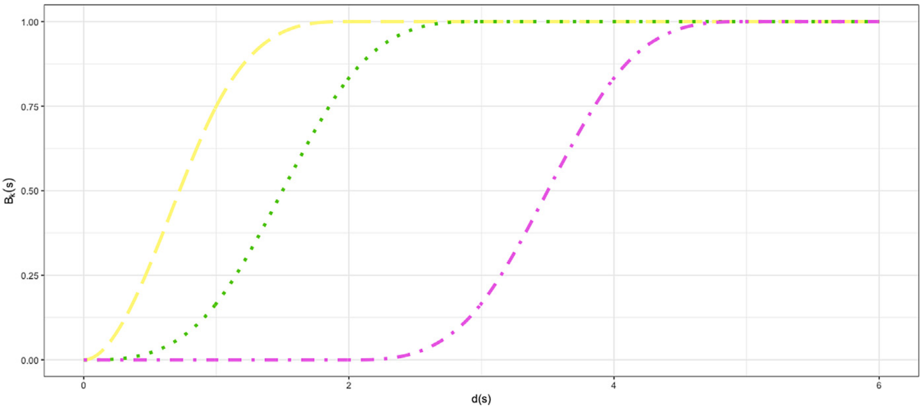


Fig. C.16. Basis functions used to fit the EVSE model for the real dataset application.

Appendix C. Moose in Hedmark application: Extra figures

See Figs. C.15–C.17.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

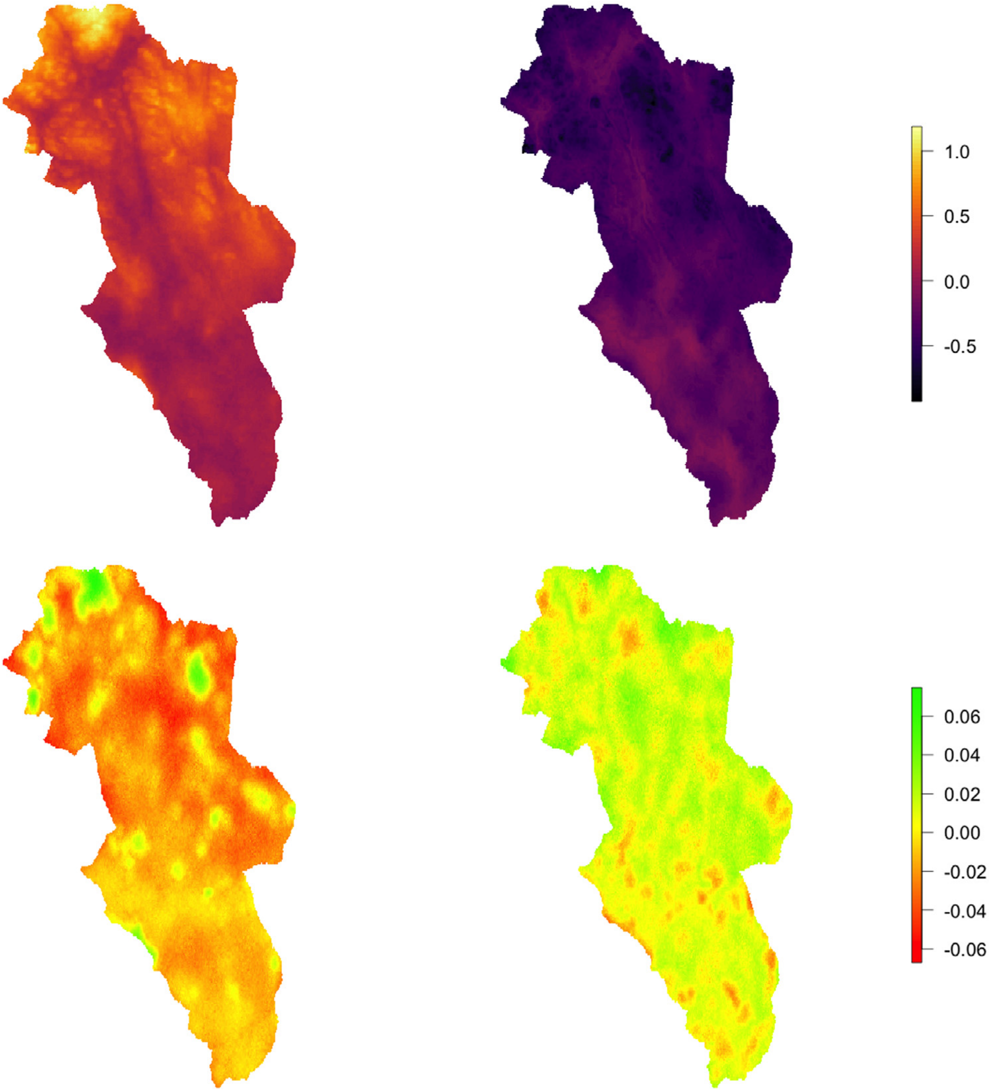


Fig. C.17. Differences in posterior median (top) and standard deviation (bottom), in log-scale, between the naive and the VSE model (left) and between the VSE and the EVSE model (right).

References

Bakar, K.S., Sahu, S.K., et al., 2015. Sptimer: Spatio-temporal bayesian modelling using R. J. Stat. Softw. 63 (15), 1–32.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3 (2), 327–338, URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-210X.2011.00172.x.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models With R-INLA. John Wiley & Sons.

Blindheim, T., 2019. Biofokus. http://dx.doi.org/10.15468/jxbhqx, Accessed via GBIF.org on 2019-12-02.

Cameletti, M., Gómez-Rubio, V., Blangiardo, M., 2019. Bayesian modelling for spatially misaligned health and air pollution data through the INLA-spde approach. Spat. Statist. 31, 100353, URL http://www.sciencedirect.com/science/article/pii/S2211675318301799.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M., Silander, J.A., 2011. Point pattern modelling for degraded presence-only data over large regions. J. R. Stat. Soc. Ser. C. Appl. Stat. 60 (5), 757–776, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2011.00769.x.

Diggle, P.J., Menezes, R., Su, T.-l., 2010. Geostatistical inference under preferential sampling. J. R. Stat. Soc. Ser. C. Appl. Stat. 59 (2), 191–232, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2009.00701.x.

Ferrier, S., Drielsma, M., Manion, G., Watson, G., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast new south Wales. II. Community-level modelling. Biodivers. Conserv. 11 (12), 2309–2338. http://dx.doi.org/10.1023/A:1021374009951.

Fick, S.E., Hijmans, R.J., 2017. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086.

Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. J. Amer. Statist. Assoc. 114 (525), 445–452. http://dx.doi.org/10.1080/01621459.2017.1415907.

Gelfand, A.E., Shirota, S., 2019. Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. Ecol. Monograph 89 (3), e01372, URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecm.1372.

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Stat. Comput. 24 (6), 997–1016. http://dx.doi.org/10.1007/s11222-013-9416-2.

Humphreys, J.M., Elsner, J.B., Jagger, T.H., Pau, S., 2017. A Bayesian geostatistical approach to modeling global distributions of lygodium microphyllum under projected climate warming. Ecol. Model. 363, 192–206, URL http://www.sciencedirect.com/science/article/pii/S0304380017304064.

Hundertmark, K., 2016. Alces alces. the iucn red list of threatened species 2016: e.T56003281a22157381.. Downloaded on 29 October 2019. URL http://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T56003281A22157381.en.

Illian, J.B., Martino, S., Sørbye, S.H., Gallego-Fernández, J.B., Zunzunegui, M., Esquivias, M.P., Travis, J.M.J., 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. Methods Ecol. Evol. 4 (4), 305–315, URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210x.12017.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns, Vol. 70. John Wiley & Sons.

iNaturalist.org, 2019. INaturalist research-grade observations. http://dx.doi.org/10.15468/ab3s5x, Accessed via GBIF.org on 2019-12-02.

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., Roy, D.B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods Ecol. Evol. 5 (10), 1052–1060, URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12254.

Kelling, S., Johnston, A., Hochachka, W.M., Iliff, M., Fink, D., Gerbracht, J., Lagoze, C., La Sorte, F.A., Moore, T., Wiggins, A., et al., 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? PLoS One 10 (10), e0139600.

Leblond, M., Dussault, C., Ouellet, J.-P., 2010. What drives fine-scale movements of large herbivores? A case study using moose. Ecography 33 (6), 1102–1112. http://dx.doi.org/10.1111/j.1600-0587.2009.06104.x.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x.

Mair, L., Ruete, A., 2016. Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PLOS ONE 11 (1), 1–13. http://dx.doi.org/10.1371/journal.pone.0147796.

Messier, F., 1991. The significance of limiting and regulating factors on the demography of moose and white-tailed deer. J. Anim. Ecol. 377–393.

Monsarrat, S., Boshoff, A.F., Kerley, G.I., 2019. Accessibility maps as a tool to predict sampling bias in historical biodiversity occurrence records. Ecography 42 (1), 125–136. http://dx.doi.org/10.1111/ecog.03944, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.03944.

NBIC, 2019a. Norwegian biodiversity information centre - other datasets. http://dx.doi.org/10.15468/tm56sc, Accessed via GBIF.org on 2019-12-02.

NBIC, 2019b. Norwegian species observation service. http://dx.doi.org/10.15468/zjbzel, Accessed via GBIF.org on 2019-12-02.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., Crowston, K., 2012. The future of citizen science: emerging technologies and shifting paradigms. Front. Ecol. Environ. 10 (6), 298–304, URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/110294.

Pettit, L., 1990. The conditional predictive ordinate for the normal distribution. J. R. Stat. Soc. Ser. B Stat. Methodol. 52 (1), 175–184, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1990.tb01780.x.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3), 231–259, URL http://www.sciencedirect.com/science/article/pii/S030438000500267X.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19 (1), 181–197, URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/07-2153.1.

Pomeroy, J.W., Gray, D.M., Shook, K.R., Toth, B., Essery, R.L.H., Pietroniro, A., Hedstrom, N., 1998. An evaluation of snow accumulation and ablation processes for land surface modelling. Hydrol. Process. 12 (15), 2339–2367. http://dx.doi.org/10.1002/(SICI)1099-1085(199812)12:15<2339::AID-HYP800>3.0.CO;2-L.

Ramsay, J.O., 1988. Monotone regression splines in action. Statist. Sci. 3 (4), 425–441. http://dx.doi.org/10.1214/ss/1177012761.

Please cite this article as: J. Sicacha-Parada, I. Steinsland, B. Cretois et al., Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics (2020) 100446, https://doi.org/10.1016/j.spasta.2020.100446.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x.

Sadykova, D., Scott, B.E., De Dominicis, M., Wakelin, S.L., Sadykov, A., Wolf, J., 2017. Bayesian joint models with INLA exploring marine mobile predator–prey and competitor species habitat overlap. Ecol. Evol. 7 (14), 5212–5226, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.3081.

Shipley, L., 2010. Fifty years of food and foraging in moose: lessons in ecology from a model herbivore. Alces:J. Devot. Biol. Manage. Moose 46, 1–13.

Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H., 2016. Going off grid: computationally efficient inference for log-Gaussian cox processes. Biometrika 103 (1), 49–70. http://dx.doi.org/10.1093/biomet/asv064.

Sørbye, S.H., Illian, J.B., Simpson, D.P., Burslem, D., Rue, H., 2019. Careful prior specification avoids incautious inference for log-Gaussian cox point processes. J. R. Stat. Soc. Ser. C. Appl. Stat. 68 (3), 543–564, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12321.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (4), 583–639, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00353.

Strasser, B., Baudry, J., Mahr, D., Sanchez, G., Tancoigne, E., 2019. “Citizen science”? Rethinking science and public participation. Sci. Technol. Stud. 32 (ARTICLE), 52–76.

Title, P.O., Bemmels, J.B., 2018. Envirem: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography 41 (2), 291–307, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.02880.

Watanabe, S., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11, 3571–3594, URL http://dl.acm.org/citation.cfm?id=1756006.1953045.

Yuan, Y., Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T., Rue, H., Gerrodette, T., et al., 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. Ann. Appl. Stat. 11 (4), 2270–2297.

Article 5

Identifying and correcting spatial bias in citizen science data for wild ungulates in Norway

Authors:

Benjamin Cretois[1]; Bram van Moorter[2]; John D.C. Linnell[2,3]; Emily G. Simmonds[4]; Christer M. Rolandsen[2]; Erling J. Solberg[2]; Olav Strand[2]; Vegard Gundersen[2]; Ole Roer[5]; Jan Ketil Rød[1]

Affiliations:

[1] Department of Geography, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[2] Norwegian Institute for Nature Research, PO Box 5685, Torgarden, 7485 Trondheim, Norway

[3] Department of Forestry and Wildlife Management, Inland Norway University of Applied Sciences, 2480 Koppand, Norway

[4] Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[5] Faun Naturforvaltning AS, Klokkarhamaren 6, 3870 Fyresdal

Corresponding author:

Benjamin Cretois

e-mail: benjamin.cretois@ntnu.no; bencretois@gmail.com

Keywords:

citizen science; resource selection function; habitat selection; ungulates; preferential sampling; spatial bias

Abstract

1. Many publications make use of opportunistic citizen science observation data, observations of species presence collected without standardized field protocols and without explicit sampling designs. Most of these publications use citizen science data to infer large-scale properties of species' distributions. However, the few publications that use citizen science data to study animal ecology at a habitat level do so without accounting for spatial biases in citizen science records or using methods that are difficult to generalize. In this study, we explore the biases that may exist in citizen science observations and suggest an approach to correct for them.
2. We first examined the extent of the biases in citizen science observations of large ungulates by comparing them to data from GPS-telemetry. We then quantified the extent of the biases by specifying a model of the biases, known as an observer model. From the observer model we sampled availability locations within the species' home range. In tandem with citizen science observations, we used the corrected availability locations to estimate a Resource Selection Function (RSF). We tested this method with simulations and empirical datasets for three wild ungulates (roe deer, moose and wild reindeer) in Norway. We compared the results of our correction method to RSFs derived using citizen science observations without correction of availability locations and to RSFs using the GPS-telemetry data. Finally, we compared habitat suitability maps obtained using each of these models.
3. As expected, citizen science observations are more affected by human access and visibility variables than locations derived from GPS-telemetry. This has consequences for drawing inferences about species' ecology. Models using citizen science observations naïvely in habitat use studies, with simple random availability locations, can result in spurious inferences. However,

sampling availability locations based on the spatial biases in citizen science drastically improves the estimation of the species’ RSFs and the species’ habitat suitability maps.

4. This study highlights the challenges and opportunities of using citizen science observations in habitat-use scale studies. While our method is not foolproof it is a first step towards unlocking the potential citizen science holds for finer scale studies of habitat selection.

1. Introduction

In species' habitat studies, ecologists compare habitat availability with habitat use (Krausman 1999). While studies at large scales (i.e. characterizing the macrohabitat) are useful for global assessment, more detailed studies (i.e. characterizing the microhabitat) are necessary to understand how animals interact with their environment, select their foraging areas, and the resulting fitness consequences of different space-use strategies (i.e. the 3rd and 4th order of selection, Johnson, 1980). Modern biotelemetry devices such as Very High Frequency (VHF) and Global Positioning System (GPS) transmitters have made it possible to study the habitat use of multiple animals at fine spatial and temporal scales, providing unique opportunities to study how species use their environment without observer bias (Frair et al. 2010). Nevertheless, such devices are expensive, often logistically difficult to deploy, and require specialist training in addition to the welfare considerations associated with animal capture. The result is that these approaches are often only used in study sites of limited size or with limited number of study animals, which may lead to poor population-level inferences (Hebblewhite & Haydon 2010) and cannot be applied to all species for which such information is desirable. Ideally it should be possible to use biotelemetry data where they exist to understand, and correct for, biases associated with the use of more extensive data types, such as those associated with citizen science data sources.

Citizen science (i.e. the engagement of volunteers in data collection) has the potential to provide tremendous amounts of data over large temporal and spatial scales that can potentially transform the study of ecology (Tewksbury et al. 2014, Bela et al. 2016). It has recently been estimated that as much as 50% of the species occurrence records stored in the Global Biodiversity Information Facility (GBIF) have been collected by Citizen Scientists (Walker 2019). Many citizen science projects have a long history (e.g. hunters recording harvest numbers, Cretois et al. 2020; records of the timing of cherry blossom in Japan, Aono & Kazui 2008; the UK Butterfly Monitoring Scheme, Brereton et al. 2014; the Christmas Bird Count, Kobori et al. 2016), and the development of web-based recording with user-friendly interfaces and associated databases is leading to an increase in the number of initiatives and an increasing uptake by the scientific community (Dickinson et al. 2012). Where these datasets have a sufficient spatial and temporal resolution, they represent a cost-effective tool for monitoring biodiversity at large scales (i.e. at the 1st order of habitat selection; Johnson 1980). However, the use of citizen science data does not come without costs as the sampling design usually violates the fundamental principles of data sampling. For instance, most data collected by volunteers are unevenly distributed in both space (i.e. off and on trails, close to roads and human settlements, Westekemper et al. 2018) and time (i.e. collected during daylight and during the weekends). Moreover, observers differ in their abilities to recognize species and the effort they spend to detect certain species (Isaac et al. 2014). There is also a question whether the observed individuals are representative of the main wildlife population, or if they have deviant behavior, e.g. because they are sick or is more than usual habituated to human activities (Reimers et al. 2010). Even though at fine scales these biases can lead to misleading conclusions if not accounted for (Sicacha-Parada et al. 2020), some studies chose to simply ignore them (Weisshaupt & Rodríguez-Pérez 2017), uncritically combine citizen science records with other source of data (Mononen et al. 2018), or use methods difficult to replicate in other systems (Todd et al. 2016)

The assumption underlying the use of citizen science observations in fine scale habitat studies is that the hypervolume occupied by citizen science data in environmental space (i.e. n-dimensional space characterizing the biotic and abiotic conditions of a location) is the intersection of the characteristics determining occupancy by observers and those by the species of interest (see a two-dimensional example in Figure 1). The environmental conditions determining occupancy by a species result from a hierarchical selection process (Johnson 1980), while the fact that citizen science data are conditional upon the presence of an observer and their ability to see and identify the animal is a source of bias. In contrast, the hypervolume in environmental space occupied by telemetry data results only from the space use of marked individuals from the species of interest. For instance, in Figure 1, citizen scientists and the target species do not use the landscape in the same way and citizen science observations only partially capture the 2nd and 3rd orders of selection (Johnson 1980). In contrast, telemetry observations

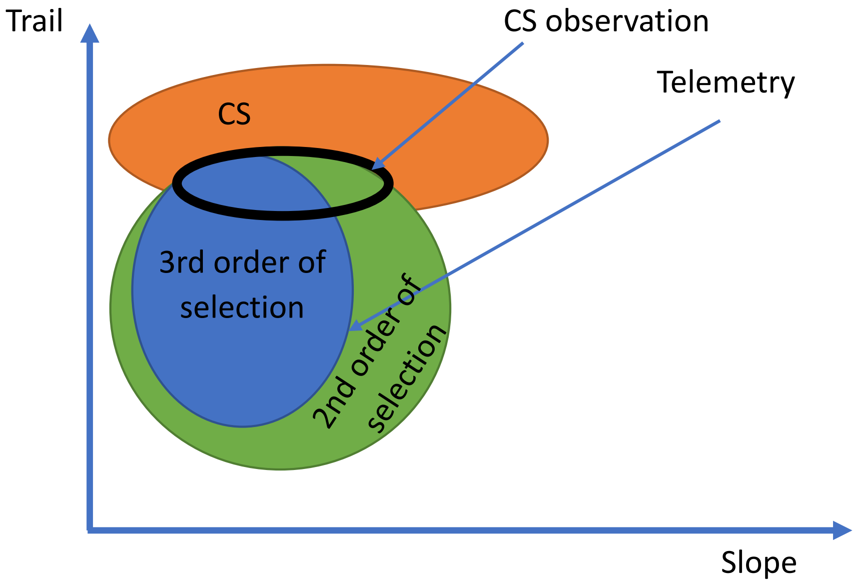


Figure 1. Conceptual figure representing the reasoning underlying the use of citizen science observations to infer species' habitat preference along two potential environmental gradients. The thick line represents the area where citizen science observations correctly identify species' ecological properties, including the RSF.

are in theory able to capture the 3rd order of selection. In this example, citizen scientists use steeper slope and heavily used trails compared to the target species that prefers less steep landscapes that contains trails that are moderately used. Thus, the distribution of the citizen scientists and the species only partially overlap. However, uneven sampling of research individuals and study site with respect to the overall population distribution may be a different source of bias in this case. Under the assumption of a representative sample of individuals in the telemetry data within a given site, it is possible to combine citizen science and telemetry data to estimate the hypervolume occupied by the observers, which could be used to correct observer bias in the citizen science data.

In this study, we present a novel method which aims to account for spatial biases in citizen science observations to get a more accurate characterization of species' habitat selection. We first explore the potential biases in citizen science observations for three widespread and easily recognizable wild ungulate species (i.e. to limit the extent of the misidentification bias), roe deer (*Capreolus capreolus*), moose (*Alces alces*) and wild mountain reindeer (*Rangifer tarandus*) in southern Norway. We then quantify the biases by specifying a model of the biases (i.e. an observer model) that estimates how accessible areas within the species' home ranges are to a citizen scientist. From the observer model we sample availability locations with regard to accessibility that are used in tandem with citizen science observations to estimate a Resource Selection Function (RSF). We tested this method with simulations and empirical datasets for the three wild ungulates. We compared the results of our novel method to a RSF using citizen science observations without correction in availability locations and to RSFs derived from the unbiased telemetry data.

2. Material and Methods

2.1. Empirical data

2.1.1. Telemetry data

We used GPS-telemetry data collected between 2008 and 2017 from a total of 501 individuals (n_{moose} = 116, $n_{roe\ deer}$ = 49, $n_{wild\ reindeer}$ = 294) located in southern Norway (more details on data collection and study site locations in Roer et al. 2018 for moose; Peters et al. 2017 for roe deer, and Panzacchi et al. 2015 for wild reindeer). Because of the geographical particularities of the area in which the roe deer GPS collar data were located, we complemented the GPS dataset with VHF-telemetry data from Viken county in order to represent a wider diversity of landscape types. The VHF data were collected using either ground based triangulation or aerial locations. Even though VHF data are not as accurate as GPS telemetry, we chose to include them due to their wider coverage. The VHF data was obtained from 41 individuals and were collected between 1995 and 2004.

For all species, the GPS data sampling interval ranged between 1 and 12 relocations per day. However, because of the large number of data points, which caused computational inefficiency, and to avoid risks of temporal and spatial autocorrelation, we resampled the telemetry dataset using the R package amt (Signer et al. 2019). We selected 1 GPS location every 5 hours for both moose and roe deer and 1 GPS location every 10 hours for wild reindeer as more observations were available. Then, we selected location data that were recorded during summer (i.e. 21st of June to 22nd of September) and during hours of normal human activities (i.e. between 8 and 22) and daylight for a fair comparison with citizen science observations, which are more numerous in the summer months (in our dataset) and during daylight. Focusing on summer only also removed complications arising from variable migration behavior and possible confounding effects of proximity to winter feeding stations that are often used by moose and roe deer, as well as issues related to the increased grouping behavior of moose and roe deer in winter and the reduced human access to habitats caused by snow (Fryxell et al. 1988).

2.1.2. Citizen Science data

We extracted moose, roe deer and wild reindeer citizen science records from the mammal records of the Norwegian Species Observation Service (https://www.artsobservasjoner.no/) dataset that we downloaded from GBIF (The Norwegian Biodiversity Information Centre 2020). “Artsobservasjoner” is the most popular citizen science platform in Norway and is maintained by the Norwegian Biodiversity Information Centre (https://www.artsdatabanken.no/). We retained all observations that were recorded during the summer season. The time of observation was not recorded in the database.

Because of the low number of citizen science records in the telemetry study sites of both roe deer and moose, we built a 10 km buffer around the distribution of telemetry observations and extracted the citizen science records located within this buffer. For wild reindeer, enough citizen science records were available within the areas from which telemetry locations were available (i.e. no buffer was needed). We sorted all citizen science observations by the name of the observer and deleted observations made by “super-observers”, persons contributing to more than 50% of the dataset. These super-observers were often employed by wildlife management institutions, and therefore were not considered to be representative of typical citizen scientists. The citizen science dataset resulted in 160 records for moose, 316 for roe deer and 183 for wild reindeer.

2.1.3. Explanatory variables

2.1.3.1. Explanatory variables for the observer model

To quantify the biases in the citizen science records, we used variables related to human infrastructure (that might influence access to wildlife habitat), human activities (that might influence the potential number of observers), and visibility (that might influence the detectability of a species to an observer), which are factors that are presumed to be the main drivers of biases in citizen science records (Geldmann et al. 2016, Tiago et al. 2017).

We extracted map layers on roads and human settlement from Open Street Map (https://www.openstreetmap.org; OSM) and Statistics Norway (https://kart.ssb.no/), respectively. Human settlements are defined as a cluster of buildings inhabited by at least 200 persons and the distance between buildings is less than 50m (https://www.ssb.no/en/klass/klassifikasjoner/110). In the OSM dataset we selected the main segments of the road network: motorway, trunk, primary, secondary, tertiary, unclassified and residential. Datasets were then used to compute 10m resolution rasters of distance to roads and distance to human settlements. Both rasters were created in ArcGIS Pro.

To represent human activity intensity, we used path use intensity and population number. For both moose and roe deer we used Strava Metro data (https://metro.strava.com/) for southern Norway to compute a path use intensity raster. The Strava Metro product is a shapefile composed of OSM trails and roads. In its attribute table, each segment contained the number of users who recorded an activity, and calibration of the STRAVA activity counts using fixed-point counter station estimates revealed a strong overall correlation (Venter et al. 2020). Data was available between 2017 and 2020. We summed the number of users who recorded an activity within 1km x 1km grid cells and rasterized the results. Because of a lack of Strava users in high mountain habitats (due to poor telephone network coverage and battery constraints on mobile devices) we used a trail use index derived from count data and surveys for wild reindeer (Gundersen et al. 2019). Human population density (residential) at a resolution of 100m x 100m was extracted from the national database.

Finally, the Corine Land Cover dataset (available at a 100m x 100m resolution raster) was used to calculate the amount of forested area within a 1km x 1km grid cell as a proxy for visibility. We assumed that the more forested area in a grid cell, the harder it would be for an observer to spot an animal.

2.1.3.2. Explanatory variables for the Resource Selection Function

Explanatory data used to estimate habitat selection for moose, roe deer and wild reindeer were classified into two categories, related to habitat (i.e. environmental data) and to human activity. These data were chosen based on previous fine-scale studies of habitat selection of these species (for roe deer see Bouyer et al. 2015, for moose see Bjørneraas et al. 2011, 2012, and for wild reindeer see Panzacchi et al. 2015; Table 1). Slope and altitude were computed from a 20m resolution Digital Elevation Model extracted from Geonorge (https://www.geonorge.no/). Path use intensity, distance to roads and urban settlements and forest coverage were the same variables we used in the observer model. We also included agricultural area coverage, which was computed by filtering the pixels labeled “agricultural areas” in the Corine Land Cover dataset. We then summed the proportion of agricultural area in each 1km x 1km grid cell.

Variables	Explanation	Resolution	Observer model	RSF
Environment				
alt	mean altitude in each pixel	$50m^2$	no	yes
slope	mean slope in each pixel	$50m^2$	no	yes
n_forest	proportion of forested areas within each pixel	$1km^2$	yes	yes
Human activity				
d_roads	distance to roads	$10m^2$	yes	yes
d_urb	distance to human settlements	$10m^2$	yes	yes
path_use	number of users who recorded an activity in each pixel	$1km^2$	yes	yes
n_agr	proportion of agricultural fields within each pixel	$1km^2$	no	yes
pop	number of inhabitants in each pixel	$100m^2$	yes (no reindeer)	no

Table 1. Description of the covariates used in the observer model and in the estimation of the Resource Selection Function.

2.2. Simulation study

2.2.1. Simulating environmental variables

We created a simulated landscape by generating different environmental variables with the nlmr package in R (Sciaini et al. 2018) on a grid composed of 200 x 200 regularly spaced cells. *Distance to human settlements* was generated by calculating the distance from each grid cell to a location placed on the upper left corner of the study area. *Forest* and *nice viewpoints* were simulated as clusters across the simulated landscape and each grid cell was coded as 0 for “absence of forest” or “absence of nice viewpoints” and 1 as “presence of forest” or “presence of nice viewpoints”. For *Other gradient* and *distance to roads,* we generated a segment across the landscape and the distance to that segment was computed for each grid cell. It should be noted that the variable *Other gradient* was generated to be highly correlated with *distance to roads* (Pearson’s r = -0.80). This allowed us to estimate the effect of the method on a variable not used in the observer model but closely correlated to a variable used in the observer model. Finally, we simulated *elevation* across the landscape using a Gaussian random field (mean = 1 and sd = 1).

2.2.2. Simulating animal locations and citizen science observations

We simulated animal locations which were linearly dependent on the value of the simulated environmental conditions. More specifically, we simulated a species which was more likely to be situated in forests ($\beta_{\text{forests}} = 2.5$), in lower altitude ($\beta_{\text{altitude}} = -2$), away from roads ($\beta_{\text{d\_roads}} = 4.5$) and attracted by an unknown gradient ($\beta_{\text{other\_gradient}} = 4.5$). The probability of presence of the species in a specific grid cell is given by Eq. 1.

$$\pi = logit^{-1}(-7 + \beta_{forest} * \text{forest} + \beta_{altitude} * \text{altitude} + \beta_{d_{roads}} * \text{d}_{\text{roads}} + \beta_{other_{gradient}} * \text{other}_{\text{gradient}}) \qquad \text{Eq. 1}$$

Accessibility is known as one of the main sources of variation in citizen science observations across the landscape and depends on multiple variables (Sicacha-Parada et al. 2020). We gave an accessibility score to each grid cell, the higher the score the more accessible the grid cell was. A high score (i.e. high accessibility) was given to grid cells located close to roads ($\beta_{d\_roads} = -6$), close to densely populated areas ($\beta_{d\_urb} = -3$) and if there were nice viewpoints ($\beta_{nice\_viewpoints} = 1$). The calculation of the accessibility score (α) is given by Eq. 2.

$$\alpha = logit^{-1}(\beta_{d_{roads}} * \text{d\_roads} + \beta_{d_{urb}} * \text{d\_urb} + \beta_{nice_{viewpoints}} * \text{nice\_viewpoints}) \qquad \text{Eq. 2}$$

Finally, for each grid cell we calculated the probability ψ of having a citizen science observation. We defined ψ as the product of $inv.logit(Accessibility)$ and $logit(\pi)$, the probability of presence of the species in a specific grid cell. Thus, the probability of having a citizen science observation in a specific grid cell was higher if the grid cell was easily accessible and if the probability of presence of the species was high (eq 3)

$$\psi = \alpha\, \pi \qquad \text{Eq. 3}$$

2.3. Analysis

2.3.1. Sampling random availability locations

RSFs are commonly used to characterize species' habitat use (Boyce & McDonald 1999). RSFs compare environmental covariates at locations visited by an animal to environmental covariates at a set of locations assumed to be available to the animal (Manly et al. 2007). Concretely, RSFs are evaluated by fitting a logistic regression to observed and available locations with available locations consisting of points sampled randomly or systematically from within an animal's estimated home range (Manly et al. 2007).

Sampling available locations is a crucial step in habitat selection studies, and different choices of available locations may influence the quantification of selection (Beyer et al. 2010). Usually “available” areas are defined as “available” if they are found within a Minimum Convex Polygon (i.e. MCP, the smallest possible convex polygon around data points) drawn around the area from which “use” locations are derived. We consequently randomly sampled availability points for citizen science and telemetry observations from within the MCP built around the telemetry observations.

2.3.2. Sampling availability locations with regards to biases

Although areas are theoretically available, they are not all equally accessible to the citizen scientist within any given part of the species’ range. Drawing availability locations at random implies the assumption of homogeneous accessibility through the species’ home range. This assumption is however not realistic as citizen scientists’ movements are influenced by a variety of factors, and not accounting for accessibility could lead to biases in the parameters estimated by any analysis such as an RSF (Sicacha-Parada et al. 2020).

We represented the spatial biases contained in citizen science observations due to the observer behavior using a model of the biases known as the *observer model* (Table 2). The observer model estimates the differences in locations between telemetry data and citizen science observations, in other words the model estimated the biases in the observation process. For both telemetry and citizen science data we extracted the value of the covariate that has been demonstrated to influence the observation process (Table 1). We then estimated the parameters of the observer model" using a logistic regression where citizen science records were coded as 1 and telemetry observations coded as 0. Thus, estimated parameter values are different from 0 if there is mismatch in the environmental space for the variables we are testing between citizen science observation locations and telemetry locations. For instance, if citizen science observations were on average located closer to the roads than the telemetry observations the observer model would return a negative parameter value.

We then randomly sampled 100,000 locations within the polygon surrounding the telemetry points (plus buffer) and predicted the probability that a point would be "used" by the citizen scientist based on the parameter values estimated by the "observer model". Finally, we sampled the corrected available locations (n = 3 x CS observations, Table 2) from the CS "use distribution" to estimate the RSF.

2.3.3. Model fitting

For each species we estimated an RSF with (1) telemetry data and randomly sampled availability points within the MCP drawn around the telemetry locations, (2) citizen science data and availability points randomly sampled within the MCP surrounding all observed locations, and (3) citizen science data with availability points sampled with regard to the "observer model" (see section 2.3.2). In the following (1) is referred as the *telemetry model*, (2) as the *naïve CS model*, and (3) as the *corrected CS model* (Table 2). To account for individual-specific variation in the telemetry dataset we used a random slope for all coefficients (Muff et al. 2019).

Following results and recommendations from previous studies (Muff et al. 2019), we used an Infinitively Weighted Logistic Regression (using weights = 1000) instead of a standard logistic regression to estimate the RSF. Theoretically, results from an Infinitively Weighted Logistic Regression are similar to those of a logistic regression with a large number of availability points (Muff et al. 2019). Using an Infinitively Weighted Logistic Regression thus reduces computational burden and is less subject to under-sampling the right amount of availability points (Filthian & Hastie 2013, Muff et al. 2019).

We used the Integrated Nested Laplace Approximation (INLA) approach with the package R-INLA for all models (Lindgren & Rue 2015). INLA is a faster alternative to Markov-Chain Monte Carlo approaches and yields similar, if not identical, results (Beguin et al. 2012). We log-transformed all variables except for slope because of their skewed distributions. All statistical analysis was conducted in R version 4.0.2 (R Core Team).

Term	Definition
Observer model	Model quantifying accessibility within species home range for a citizen scientist by evaluating differences in locations between citizens science and telemetry data
Corrected availability	Available locations sampled and that are used in the corrected CS model in tandem with citizen science observations
Corrected CS model	Resource selection function estimated by an infinitively weighted logistic regression using both corrected availability locations and citizen science observations
Naïve CS model	Resource selection function estimated by an infinitively weighted logistic regression using both availability locations randomly sampled across the species' home range and citizen science observations

Table 2: Terms specific to the method presented in this study and their associated definitions.

2.3.4. Habitat suitability maps

In order to visualize the consequences of different approaches we predicted habitat suitability maps on a grid placed over the MCP drawn around the telemetry locations of each species, extracting covariate values at each grid cell and multiplied by the mean parameter estimates of the fitted RSF. This procedure was done for the telemetry model, the naïve CS model, and the corrected CS model. Each grid cell has an area of 1km x 1km for roe deer and moose, and 2km x 2km for wild reindeer as their range was larger. Finally, we calculated the Pearson correlation coefficients between each pair of habitat suitability maps.

3. Results

3.1. Difference in the locations of citizen science observations and animal telemetry

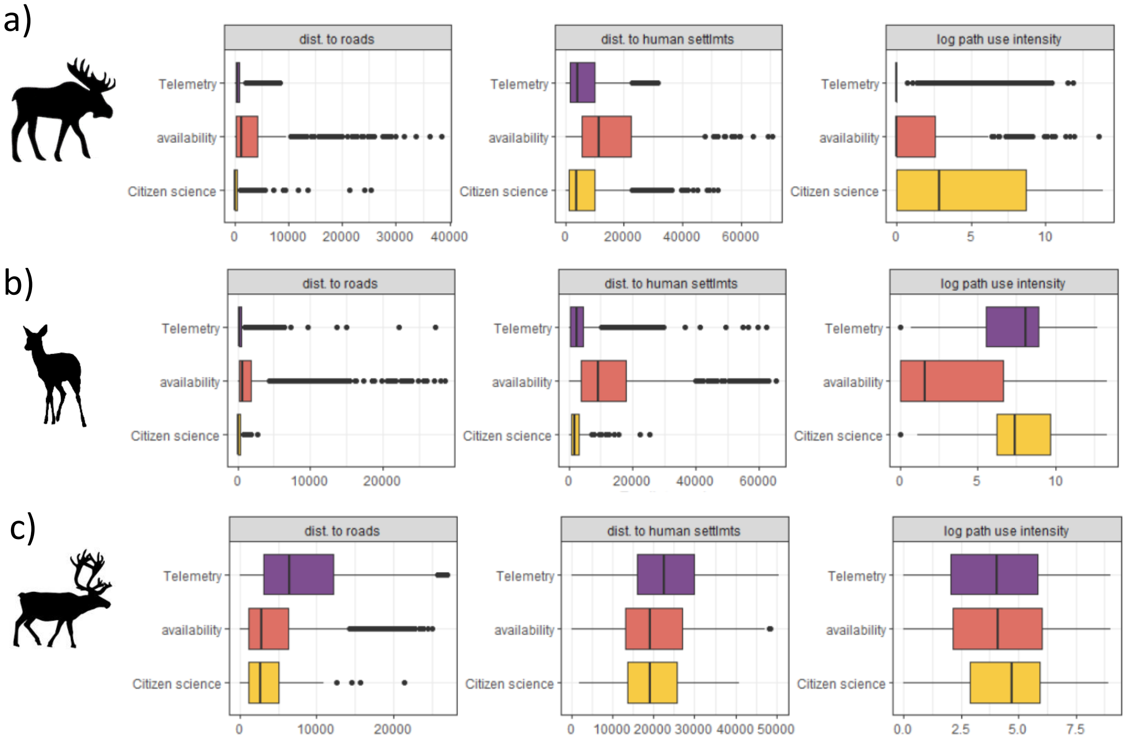


Figure 2. Boxplots of the distribution of telemetry locations, availability locations and citizen science observations within the distance to roads, distance to human settlements and path use intensity spectrum for a) moose, b) roe deer and c) wild reindeer.

We first estimated the potential biases in citizen science observations by comparing them with telemetry data and available locations within the species’ home range. Figure 2 shows that, human activity variables (i.e. distance to roads, distance to human settlements and path use intensity) influence the location of citizen science observations and telemetry observations differently, and that the contrast between telemetry and citizen science is species' specific. While citizen science and telemetry locations are similarly distributed with regard to distance to roads for moose ($\text{mean}_{\text{dist to roads moose}}$ = 741 and 757 meters for citizen science and telemetry locations, respectively), summary statistics indicate that citizen science observations are on average closer to the roads for both roe deer and wild reindeer ($\text{mean}_{\text{dist to road roe deer}}$ = 262 and 422 meters and $\text{mean}_{\text{dist to roads wild reindeer}}$ = 3574 and 8283 meters for citizen science and telemetry locations, respectively).

In contrast, citizen science observations are on average closer to human settlements than telemetry observations for all species, although the discrepancy is particularly strong for moose ($\text{mean}_{\text{dist to human settlements moose}}$ = 3693, 6842; $\text{mean}_{\text{dist to human settlements roe deer}}$ = 2222, 4022; $\text{mean}_{\text{dist to human settlements wild reindeer}}$ = 19417, 23536 for citizen science and telemetry locations respectively). Finally, the descriptive statistics shows that citizen science observations are located on average closer to more utilized paths than telemetry locations for all three species ($\text{mean}_{\text{log path use intensity moose}}$ = 3.54, 1.06; $\text{mean}_{\text{log path use intensity roe deer}}$ = 7.43, 6.77; $\text{mean}_{\text{log path use intensity wild reindeer}}$ = 4.41, 3.91 for citizen science and telemetry locations respectively).

3.2. Results of the simulation study

The simulation study was used to test our method under ideal conditions and confirm the intuition that accounting for biases improves inference.

The results of the simulation (displayed on Figure 3) show that the model accounting for observer bias by sampling corrected available locations (i.e. RSF CS corrected) returns coefficients that are more consistent with the parameter values used to simulate the species presence (i.e. Simulated parameter value in Figure 3) than the model sampling random availability locations (i.e. RSF CS naïve). This is particularly visible for the simulated distance to roads where the RSF CS naïve model returns a parameter value of opposite direction ($\beta_{\text{Simul. dist. to roads RSF CS naive}}$ = -2.098). In contrast, the $\beta_{\text{Simul. dist. to roads RSF CS corrected}}$ = 3.699, which is consistent with and close to the simulated parameter value of 4.5.

Also, the parameter estimate for other gradients seems to benefit from the correction in availability as the mean estimated parameter value is close to the true parameter value of 3 used to simulate species presence ($\beta_{\text{Simul other gradient RSF CS naïve}}$ = 1.438, $\beta_{\text{Simul other gradient RSF CS corrected}}$ = 2.535).

Finally, our simulation suggests that both the corrected RSF and naïve RSF return estimated parameter values that are very similar to the parameter value used to simulate species presence for variables influencing species occurrence only (i.e. β_{forest} = 2.5, 2.471, 2.328 and β_{altitude} = -2, -1.617, -1.496 for the parameter value used to simulate species' presence, RSF CS naïve and RSF CS corrected respectively).

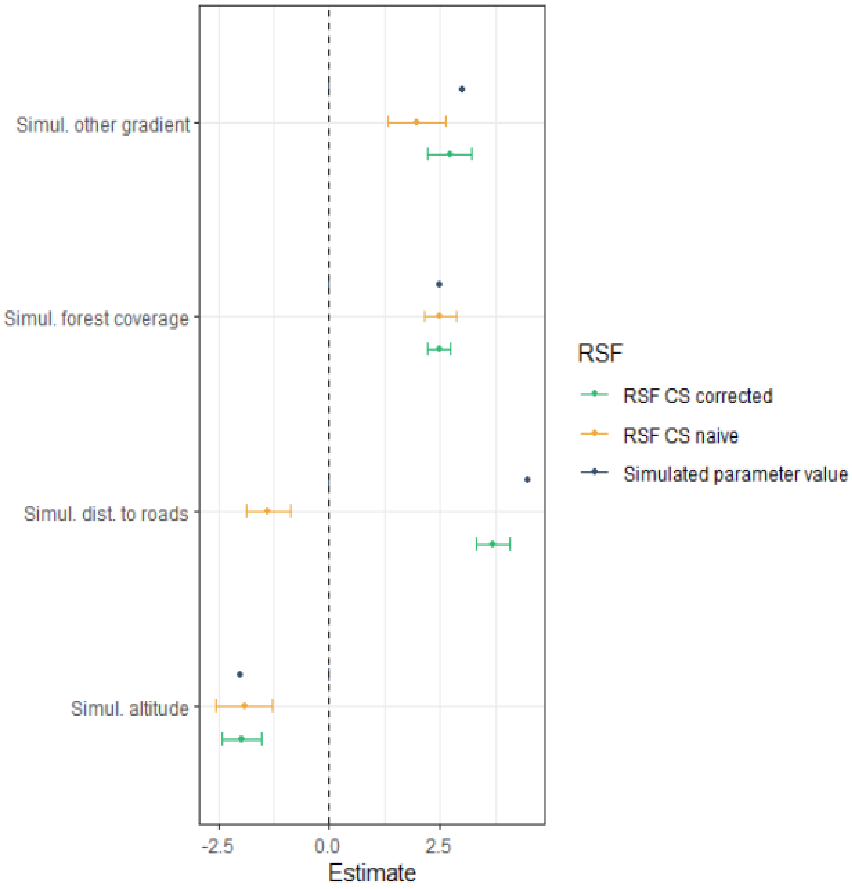


Figure 3. RSF parameter estimates obtained in the simulation for the naïve CS model (in yellow) and the corrected CS model (in green). Blue dots represent the parameter value used to simulate species presence across the simulated landscape. Dots represent the mean parameter estimates and bars the 95% credible intervals.

3.3. Results from modelling the biases in citizen science observations: the observer models

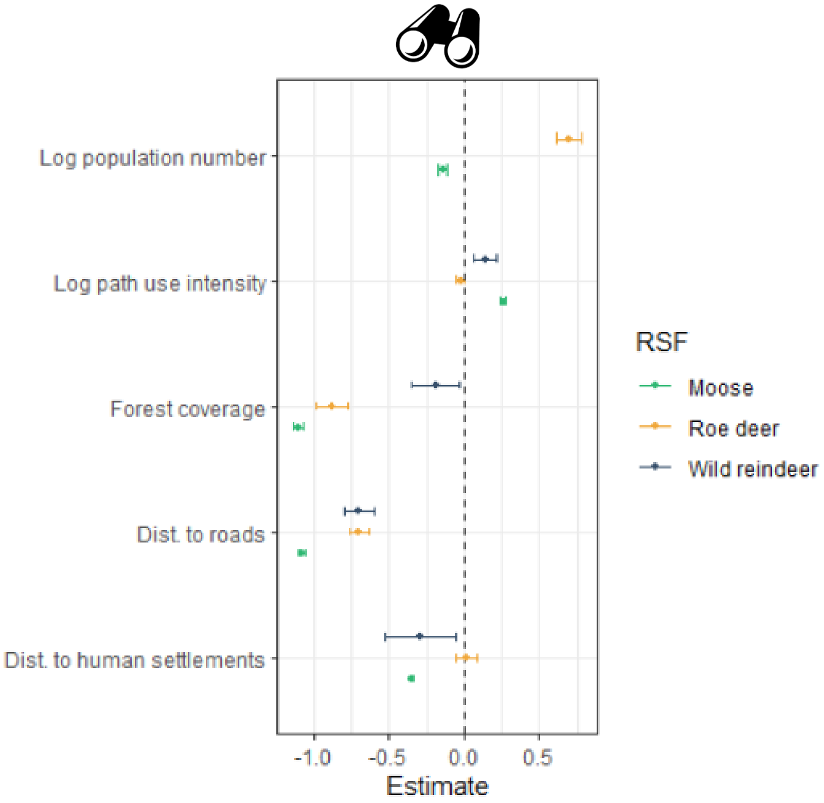


Figure 4. Parameter estimates from the moose, roe deer and wild reindeer observer models. The further away from 0 is the estimate, the more CS observations are affected by the variable compared to telemetry observations. Dots represent the mean parameter estimate and bars the 95% Credible intervals.

Even though the extent of the biases were qualitatively examined in section 3.1, the method we suggest in this paper relies on the quantification of these biases. We quantified the biases by specifying an observer model and quantifying the strength of the variables influencing the accessibility within the species' home range for a citizen scientist.

The observer model (Figure 4) confirms that there are spatial discrepancies between citizen science and telemetry locations. Results indicate that citizen science observations are on average located within a different ecological space. Nevertheless, while there are similarities between species, the extent to which the different variables influence citizen science observation compared to telemetry observations differ. Citizen science observations are on average closer to the roads than telemetry locations for both roe deer and wild reindeer ($\text{mean}_{\text{dist. to roads =}}$ -0.700, -0.729 for roe deer and wild reindeer respectively), but we only found a very weak effect for moose ($\text{mean}_{\text{dist to roads}}$ = -0.027). Nevertheless, citizen science observations are closer to human settlements than telemetry locations for moose ($\text{mean}_{\text{dist to human settlements =}}$ -0.405).

The observer models show that citizen science observations are more likely to be located near a heavily utilized path than telemetry locations for moose and wild reindeer $\text{mean}_{\text{log path use intensity}}$ = 0.247, 0.147 for moose and wild reindeer respectively). It does not make a difference for roe deer, where citizen science observations were located with a similar frequency as telemetry locations regarding path use intensity ($\text{mean}_{\text{log path use intensity}}$ = -0.027). Citizen science observations of moose and roe deer are also likely to be made in areas with higher human density than telemetry locations ($\text{mean}_{\text{log population number}}$ = 0.204, 0.700).

Finally, visibility (i.e. approximated by forest coverage within a grid cell) also had a major role as citizen science observations for all three species were more likely to be found in grid cells containing less forests ($\text{mean}_{\text{Forest coverage}}$ = -0.771, -0.876, -0.149 for moose, roe deer and wild reindeer respectively).

3.4. RSF obtained with a model using telemetry, citizen science with random availability and citizen science using a corrected availability

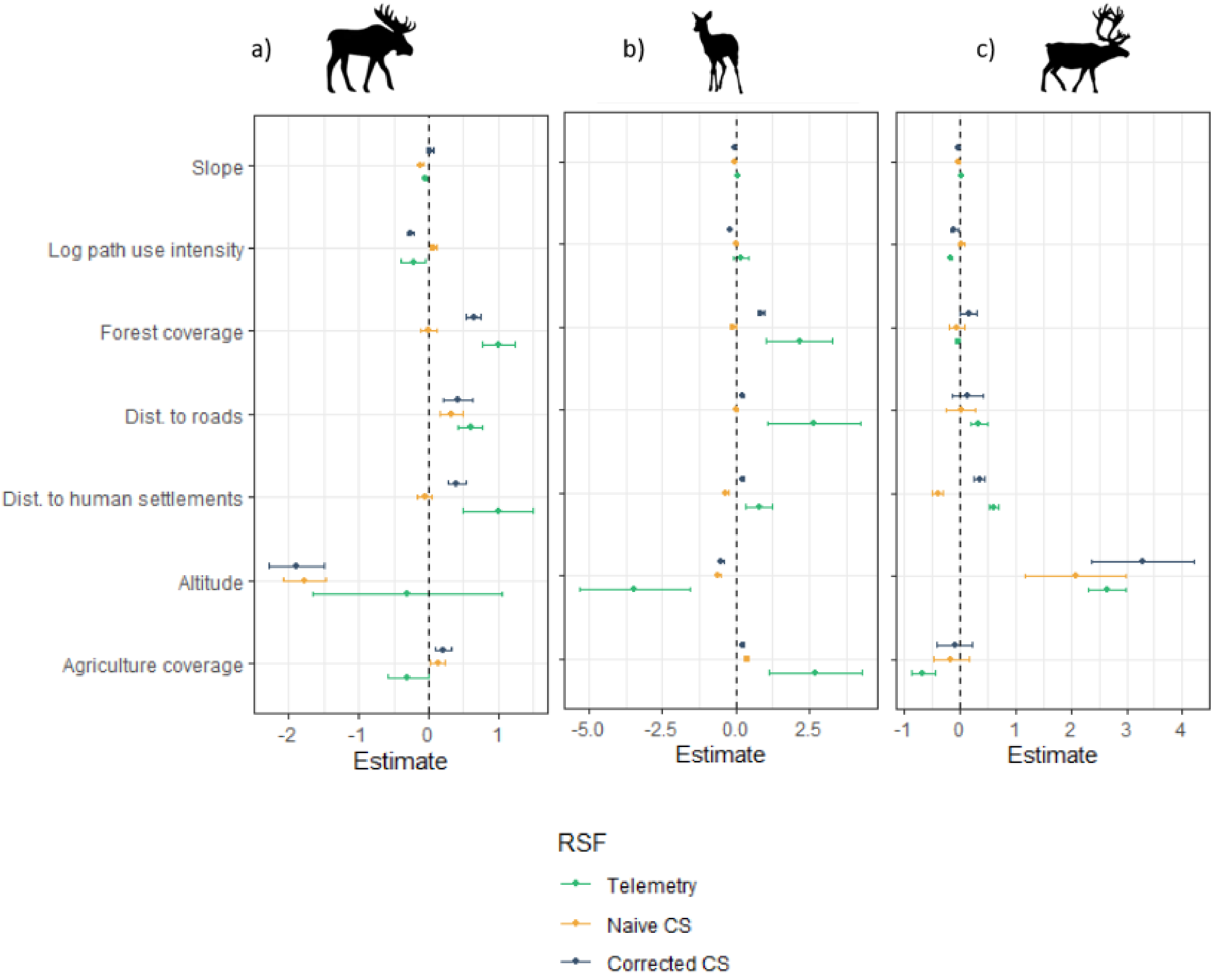


Figure 5. RSF parameter estimates for a) moose, b) roe deer and c) wild reindeer. In green are the parameter estimates obtained with the telemetry model, in yellow with the naïve CS model and in purple with the corrected CS model. Dots represent the mean parameter estimates and bars the 95% credible intervals.

Finally, after quantifying the extent of the biases in 3.3 and sample availability locations regarding the observer model, we were able to estimate and compare the RSFs for the different models.

Discrepancies between the citizen science observations and the telemetry locations had substantial effects on the naive CS model (Figure 5). Especially, in the model using telemetry data the coefficient was positive for distance to roads and distance to human settlements for all three species ($\text{mean}_{\text{dist to roads telemetry}}$ = 0.604, 0.779, 0.613, $\text{mean}_{\text{dist. human settlements telemetry}}$ = 1.005, 2.653, 0.341 for moose, roe deer and wild reindeer respectively). In contrast, the coefficients for distance to roads were negative for roe deer and wild reindeer ($\text{mean}_{\text{dist. to roads naïve CS}}$ = -0.036, -0.406 for roe deer and wild reindeer respectively) in the naïve CS model, and negative for distance to human settlements for moose ($\text{mean}_{\text{dist. to human settlements naïve CS}}$ = -0.055). In the telemetry model we can also see that the coefficient for path use intensity is negative for both moose and wild reindeer ($\text{mean}_{\text{log path use intensity telemetry}}$ = -0.219; -0.166 for moose and wild reindeer respectively), but in the naïve CS model this value is positive for moose and very close to 0 for wild reindeer ($\text{mean}_{\text{log path use intensity telemetry}}$ = 0.066, 0.029 for moose and wild reindeer respectively). Finally, the coefficient for the proportion of forested area per grid cell is positive in the telemetry model for moose and roe deer ($\text{mean}_{\text{forest coverage telemetry}}$ = 1.001; 2.155 for moose and roe deer respectively) while it is close to 0 in the naïve CS model for both species ($\text{mean}_{\text{forest coverage naïve CS}}$ = 0.004; -0.099 for moose and roe deer respectively).

Interestingly, correcting for the availability points used in the RSF brings the coefficients closer to the telemetry coefficients in most cases. Particularly, in the corrected CS model the sign of the coefficients for distance to roads and distance to urban settlements is consistently in the direction of the telemetry ones for all three species ($\text{mean}_{\text{dist. to roads corrected CS}}$ = 0.419, 0.240, 0.359, $\text{mean}_{\text{dist to human settlements corrected CS}}$ = 0.409, 0.216, 0.137 for moose, roe deer and wild reindeer respectively). The coefficient for path use intensity is also of the same sign of the telemetry model for moose and wild reindeer ($\text{mean}_{\text{log path use}}$

$_{\text{intensity corrected CS}}$ = -0.248, -0.103 for moose and wild reindeer respectively) while the coefficient for forest cover has the same sign as the telemetry value for moose and roe deer.

While the corrected model does better than the naïve model for most of the coefficients, surprisingly we see in Figure 5 that the corrected CS model coefficients does not get closer to the telemetry model for either proportion of agricultural coverage or altitude and even seems to perform worse than the naïve CS model for these variables (mean$_{\text{agricultural coverage corrected CS}}$ = 0.204, 0.217, -0.099, mean$_{\text{altitude corrected CS}}$ = -1.896, -0.493, 3.297 for moose, roe deer and wild reindeer respectively).

3.5. Suitability maps obtained with a model using telemetry, citizen science with random availability and citizen science using a corrected availability

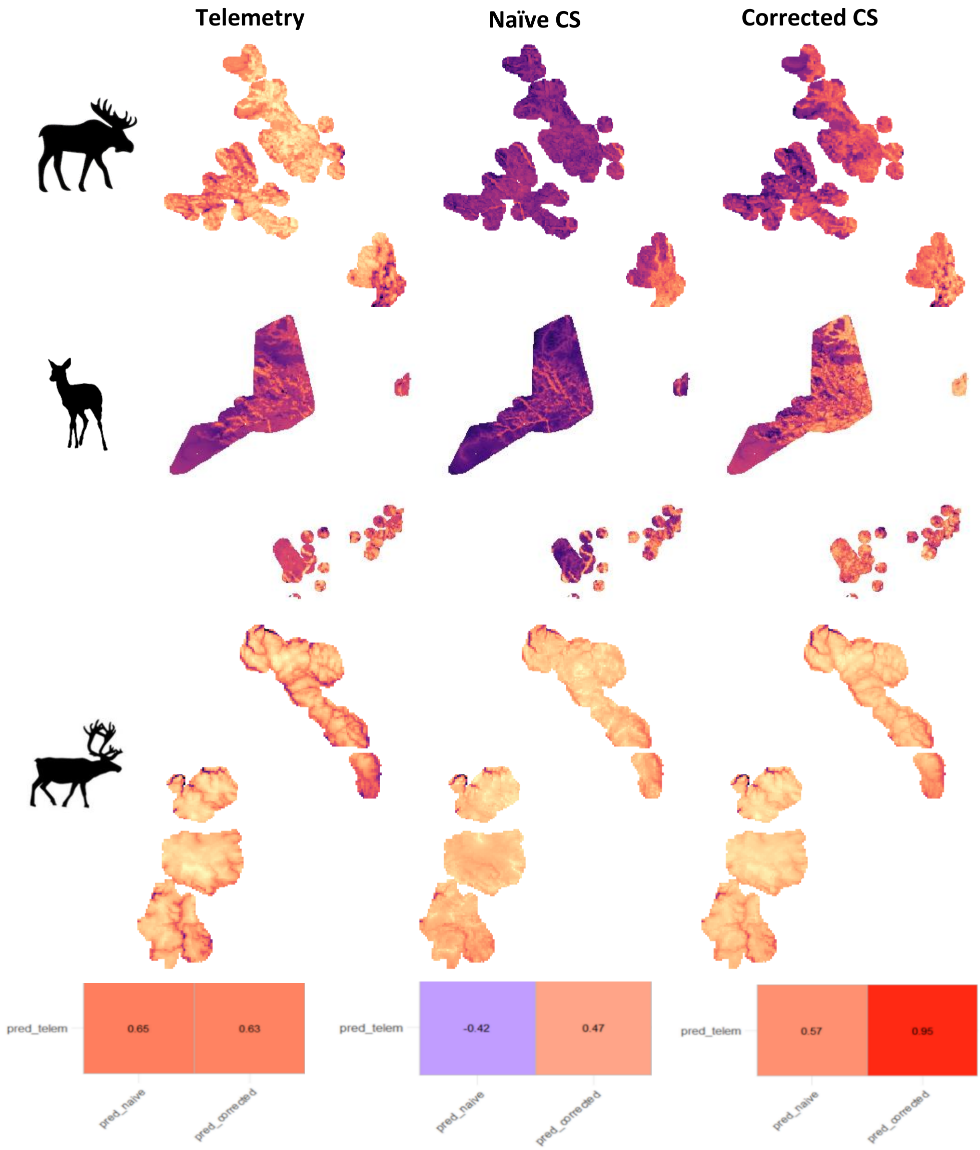


Figure 6. Suitability maps obtained using the mean coefficients of the telemetry (column 1), naïve CS (column 2) and corrected CS models (column 3) for roe deer (row 1), moose (row 2) and wild reindeer (row 3). Map values are the log-odds probabilities. On row 4 are the Pearson correlation coefficients between the habitat suitability maps obtained with the telemetry model and the ones obtained with the naïve CS model and the corrected CS model.

The habitat suitability maps obtained with the naïve CS model are similar to the habitat suitability maps obtained with telemetry location for both roe deer and wild reindeer (Figure 6). In fact, the Pearson correlation coefficients between the habitat suitability maps are 0.67 and 0.57, respectively, which can be considered as a moderately strong positive correlation. However, for moose the naïve CS model does not represent well the habitat suitability map obtained with telemetry locations as the correlation coefficient is negative (r = -0.42).

The corrected CS models perform better than the naïve CS model for both moose and wild reindeer and as good as the naïve CS model for roe deer. The improvement is particularly obvious for the moose, for which the corrected CS model return a moderately strong positive correlation with the habitat suitability map obtained with telemetry locations (r = 0.47). Furthermore, for wild reindeer the corrected CS model returns a habitat suitability map strongly correlated to a suitability map obtained with telemetry locations (r = 0.95).

Discussion

In this study we show that citizen science observations are skewed towards areas that are more easily accessible and used by humans (i.e. citizen scientists). This result is consistent with many studies pointing out the biases in opportunistic citizen science data (Sicacha-Parada et al. 2020, Tiago et al. 2017, Geldmann et al. 2016), and indicate that a naive use of citizen science observations in fine-scale habitat selection studies can lead to misleading results. We show how the RSF obtained with a naïve CS model can return parameter values inconsistent with the species habitat preferences as estimated from telemetry data (i.e. opposite habitat selection to that obtained in RSFs using telemetry data). However, by implementing an observer model that accounts for citizen scientist accessibility within the species' home range, we were able to account for some of the biases in citizen science observations.

Under ideal conditions (i.e. the simulation) our method returns parameter estimates consistent with the values used to simulate species' presence data. Empirically, an RSF estimated with our correction method returns parameter values consistent with an RSF based on telemetry. While the correction seems less helpful for roe deer, it gave substantial improvement for species that are more sensitive to human activities such as moose and wild reindeer. In fact, except for a few parameters, the estimates had consistently the same direction and were within the 95% Credible Interval of the telemetry models' parameter estimates. Drawing habitat suitability maps from the corrected CS model also drastically improved the accuracy, with the exception for roe deer, which was already quite good. In contrast, the naïve CS model returned habitat maps that, in the worst cases, were totally misleading (e.g. moose in our study).

Even though the method clearly shows that there is potential for using corrected citizen science data for fine scale habitat selection studies we can see some concerns notably with the estimated parameter values for both altitude and agricultural field coverage which do not get closer to the coefficients of the telemetry model. In the simulation, the corrected CS model should always correct coefficients related to any bias variables and the discrepancies might only be due to the stochasticity inherent to any statistical model and bootstrapping could be used to stabilize the estimates (See Figures A1 and A2 in the Annexes). Alternatively, these results may be due to our failure to correctly understand how these environmental layers influence citizen scientists' movements and observation pattern.

Finally, the suggested method relies on a good estimation of observer bias. This requires reliable information about the species' ecology which can be obtained with GPS telemetry data. Nevertheless, GPS telemetry studies are costly and thus cannot be conducted everywhere on all species. Nevertheless, our results suggest that using telemetry observation from an auxiliary species with similar habitat preference to a target species could be used to estimate observer bias, correct for availability locations used in the logistic model, and thus partially correct parameter estimates (See Figure A6 – A11 in the Annexes).

There are multiple reasons for why citizen science observations do not represent accurately Resource Selection Functions and more generally species' ecology including spatial and temporal biases (Isaac et al. 2014). While the method suggested in this paper account for these biases to improve ecological inference, it is also necessary to improve citizen science data collection. Citizen scientists usually report observations from quite human dominated areas for instance when surprised to find a species in a place where they are not used to be seen, capturing only certain species' individual behavior that are not representative of the species usual range. Encouraging other types of citizen scientists, such as hunters or other outdoor enthusiast, could improve the coverage of the dataset, and improve inference (Cretois et al. 2019).

Despite its limitations, our method is a first step towards using citizen science data in habitat selection studies. In fact, we do not present our method to correct for availability as an infallible technique but rather as a way to initiate conversations and research among ecologists to account for spatial biases in citizen science data for more accurate inference at fine scales. Methods to account for variation in the observation process in citizen science observations are developing and improving, notably with the potential of occupancy models (Strien et al. 2013, Altwegg & Nichols 2018). Nevertheless, these developments account for biases in citizen science at the distribution level (4th order of selection; Johnson 1980) and to our knowledge our study is the first attempting to find a general solution for using citizen science data at finer scale.

Conclusion

In this paper we explore the challenges and the opportunities of using citizen science data in habitat preference studies. We show that citizen science used in a naïve way can be misleading and result in spurious ecological inference. Accounting for the observation process reduces this risk. Our is a first step towards using citizen science data for finer scale analyses. We also show that it is possible to use an auxiliary species observation model to correct the habitat preference model of a target species.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

BC was funded by a PhD scholarship from the Norwegian University of Science and Technology. JDCL and BVM were funded by the Research Council of Norway (grant 251112).

Authors contribution

JDCL, BC and BVM conceived the study idea, BC and BVM designed the methodology and compiled the telemetry data that were provided by JDCL, CMR, EJS and OS. VG provided the path layers for wild reindeer. BC, BVM and EGS worked on the simulations. BC analyzed the data and interpreted the results. All authors participated in the writing of the manuscript and gave final approval for submission.

Data availability

Code and data to run the analysis and the simulation study can be found at https://zenodo.org/record/4590153#.YFHDv69KguU.

References

Altwegg, R., & Nichols, J. D. (2019). Occupancy models for citizen-science data. *Methods in Ecology and Evolution*, *10*(1), 8-21.

Beguin, J., Martino, S., Rue, H., & Cumming, S. G. (2012). Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution*, *3*(5), 921-929.

Bela, G., Peltola, T., Young, J. C., Balázs, B., Arpin, I., Pataki, G., ... & Keune, H. (2016). Learning and the transformative potential of citizen science. *Conservation Biology*, *30*(5), 990-999.

Beyer, H. L., Haydon, D. T., Morales, J. M., Frair, J. L., Hebblewhite, M., Mitchell, M., & Matthiopoulos, J. (2010). The interpretation of habitat preference metrics under use–availability designs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2245-2254.

Bjørneraas, K., Herfindal, I., Solberg, E. J., Sæther, B. E., van Moorter, B., & Rolandsen, C. M. (2012). Habitat quality influences population distribution, individual space use and functional responses in habitat selection by a large herbivore. *Oecologia*, *168*(1), 231-243.

Bjørneraas, K., Solberg, E. J., Herfindal, I., Van Moorter, B., Rolandsen, C. M., Tremblay, J. P., ... & Astrup, R. (2011). Moose Alces alces habitat use at multiple temporal scales in a human-altered landscape. *Wildlife Biology*, *17*(1), 44-54.

Bouyer, Y., Rigot, T., Panzacchi, M., Van Moorter, B., Poncin, P., Beudels-Jamar, R., ... & Linnell, J. D. (2015, April). Using zero-inflated models to predict the relative distribution and abundance of roe deer over very large spatial scales. In *Annales Zoologici Fennici* (Vol. 52, No. 1–2, pp. 66-76). Finnish Zoological and Botanical Publishing Board.

Boyce, M. S., & McDonald, L. L. (1999). Relating populations to habitats using resource selection functions. *Trends in Ecology & Evolution*, *14*(7), 268-272.

Cretois, B., Linnell, J. D., Grainger, M., Nilsen, E. B., & Rød, J. K. (2020). Hunters as citizen scientists: Contributions to biodiversity monitoring in Europe. *Global Ecology and Conservation*, e01077.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., ... & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291-297.

Gundersen, Vegard, et al. "Large-scale segregation of tourists and wild reindeer in three Norwegian national parks: Management implications." *Tourism Management* 75 (2019): 22-33.

Fithian, W., & Hastie, T. (2013). Finite-sample equivalence in statistical models for presence-only data. *The annals of applied statistics*, *7*(4), 1917.

Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J., & Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2187-2200.

Fryxell, J. M., Greever, J., & Sinclair, A. R. E. (1988). Why are migratory ungulates so abundant?. *The American Naturalist*, *131*(6), 781-798.

Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., ... & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, *22*(11), 1139-1149.

Gundersen, V., Vistad, O. I., Panzacchi, M., Strand, O., & van Moorter, B. (2019). Large-scale segregation of tourists and wild reindeer in three Norwegian national parks: Management implications. *Tourism Management*, *75*, 22-33.

Hebblewhite, M., & Haydon, D. T. (2010). Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2303-2312.

Isaac, N. J., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, *5*(10), 1052-1060.

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

Krausman, P. R. (1999). Some basic principles of habitat use. *Grazing behavior of livestock and wildlife*, *70*, 85-90.

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1-25.

Manly, B. F. L., McDonald, L., Thomas, D. L., McDonald, T. L., & Erickson, W. P. (2007). *Resource selection by animals: statistical design and analysis for field studies*. Springer Science & Business Media.

Muff, S., Signer, J., & Fieberg, J. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology*, *89*(1), 80-92.

Panzacchi, M., Van Moorter, B., Strand, O., Loe, L. E., & Reimers, E. (2015). Searching for the fundamental niche using individual-based habitat selection modelling across populations. *Ecography*, *38*(7), 659-669.

Peters, W., Hebblewhite, M., Mysterud, A., Spitz, D., Focardi, S., Urbano, F., ... & Cagnacci, F. (2017). Migration in geographic and ecological space by a large herbivore. *Ecological Monographs*, *87*(2), 297-320.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reimers, E., Røed, K. H., Flaget, Ø., & Lurås, E. (2010). Habituation responses in wild reindeer exposed to recreational activities. *Rangifer,* 30, 45-59. doi.org/10.7557/2.30.1.781

Roer, O., Rolandsen, C. M., Meland, M., Gangsei, L.E., Panzacchi, M., Van Moorter, B., … & Solberg, E. J. 2018. Moose project Akershus – Part 1. Cameramonitoring of wildlife crossing structures and area use of moose in Øvre Romerike. Statens vegvesen. 65 p. + appendix

Sciaini, M., Fritsch, M., Scherer, C., & Simpkins, C. E. (2018). NLMR and landscapetools: An integrated environment for simulating and modifying neutral landscape models in R. *Methods in Ecology and Evolution*, *9*(11), 2240-2248.

Sicacha-Parada, J., Steinsland, I., Cretois, B., & Borgelt, J. (2020). Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. *Spatial Statistics*, 100446.

Signer, J., Fieberg, J., & Avgar, T. (2019). Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. *Ecology and evolution*, *9*(2), 880-890.

Tewksbury, J. J., Anderson, J. G., Bakker, J. D., Billo, T. J., Dunwiddie, P. W., Groom, M. J., ... & Del Rio, C. M. (2014). Natural history's place in science and society. *BioScience*, *64*(4), 300-310.

The Norwegian Biodiversity Information Centre (NBIC). Occurrence dataset https://doi.org/10.15468/zjbzel accessed via GBIF.org on 2020-09-01.

Tiago, P., Ceia-Hasse, A., Marques, T. A., Capinha, C., & Pereira, H. M. (2017). Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Scientific Reports*, *7*(1), 1-9.

van Strien, A. J., van Swaay, C. A., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, *50*(6), 1450-1458.

Venter, Z. S., Barton, D., Gundersen, V., Figari, H. & M. Nowell. 2020. Urban nature in a time of crisis: recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. Environmental Research Letters. https://doi.org/10.1088/1748-9326/abb396

Walker, J. (2019). Will Citizen Science take over? https://data-blog.gbif.org/post/gbif-citizen-science-data/

Westekemper, K., Reinecke, H., Signer, J., Meißner, M., Herzog, S., & Balkenhol, N. (2018). Stay on trails–effects of human recreation on the spatiotemporal behavior of red deer Cervus elaphus in a German national park. *Wildlife Biology*, *2018*(1).

Annexes for:

Identifying and correcting spatial bias in citizen science data for wild ungulates in Norway

This document includes:

Supplementary material and method

Supplementary results and discussion

The data generation process of the simulation is stochastic. Therefore, we ran a bootstrap of the data generation process to create 50 different realisations of the simulated environmental and occurrence data. We fit all of the models (telemetry, naïve and corrected RSF) to each simulated dataset and recorded if the estimates of the coefficients in the final models were improved by the correction. Improvement was indicated as true or false, where true arises when the coefficient estimate from the corrected model was closer to the telemetry model coefficient estimate than the naïve model estimate. No measure of strength of improvement was included.

In addition to stochasticity in the data generation process, there is also stochasticity in the selection of background points for the citizen science models (both naïve and corrected). Therefore, we also ran a bootstrap of this process, repeating the model fitting 50 times for a single simulated dataset. Improvement in coefficient estimates was recorded in the same way as for the bootstrap of the data generation.

The results of the bootstrapping of the data generation process indicate that there is some stochasticity in how well the correction improves the estimates of the effect of certain covariates, when the selection of background points is only run once. The results show that while the estimates for the effect of forest and other gradient were consistent improvements regardless of the exact data generated, the results for altitude and distance to roads were less consistent. These latter two effects were heavily influenced by the exact data set that was generated in the simulation. However, Figure SX2 demonstrates that bootstrapping of the selection of background points for the corrected model could improve estimates of these coefficients. This is particularly true for the coefficient estimates for variables that are related to bias in the citizen science data, such as the distance to roads, bootstrapping of the background point selection led to 100% correct estimation in this particular realisation (note that this is for just one simulated dataset). However, for other variables, such as altitude, no improvement was achieved from 50 bootstraps of the selection of background points. It should also be noted that a lack of improvement in the coefficient estimate from the naïve to the corrected RSF does not necessarily indicate any substantial change to the estimate of the effect. It only indicates that the estimate from the corrected RSF is not closer to the telemetry model estimate, all could be very close. It is not surprising that those variables that do not influence bias are least impacted by the correction.

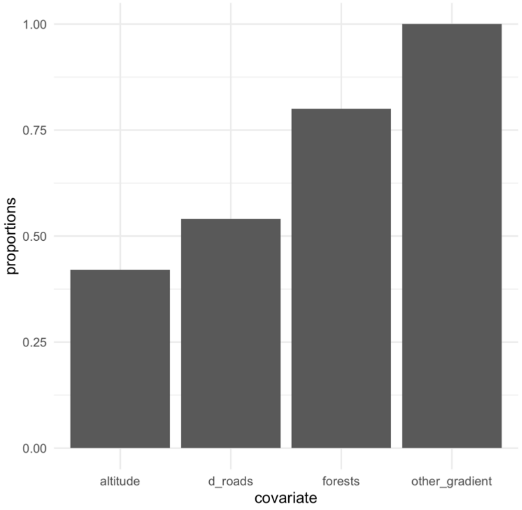


Figure A1: Proportion of times that the covariate estimate was improved by the corrected model. Results from 50 bootstraps of the data generation process.

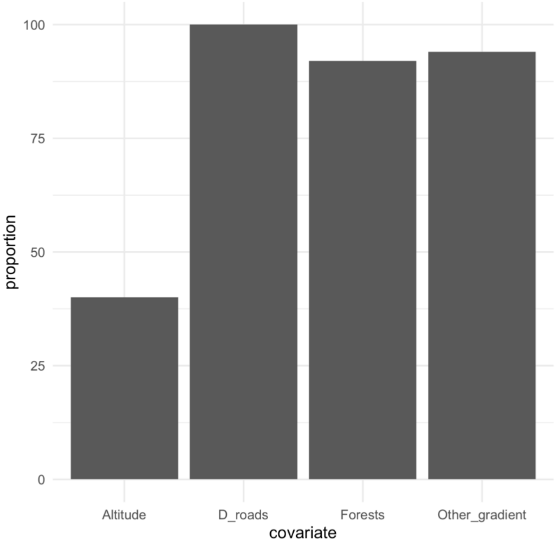


Figure A2. Proportion of times that the covariate estimate was improved by the corrected model. Results from 50 bootstraps of the background point selection process.

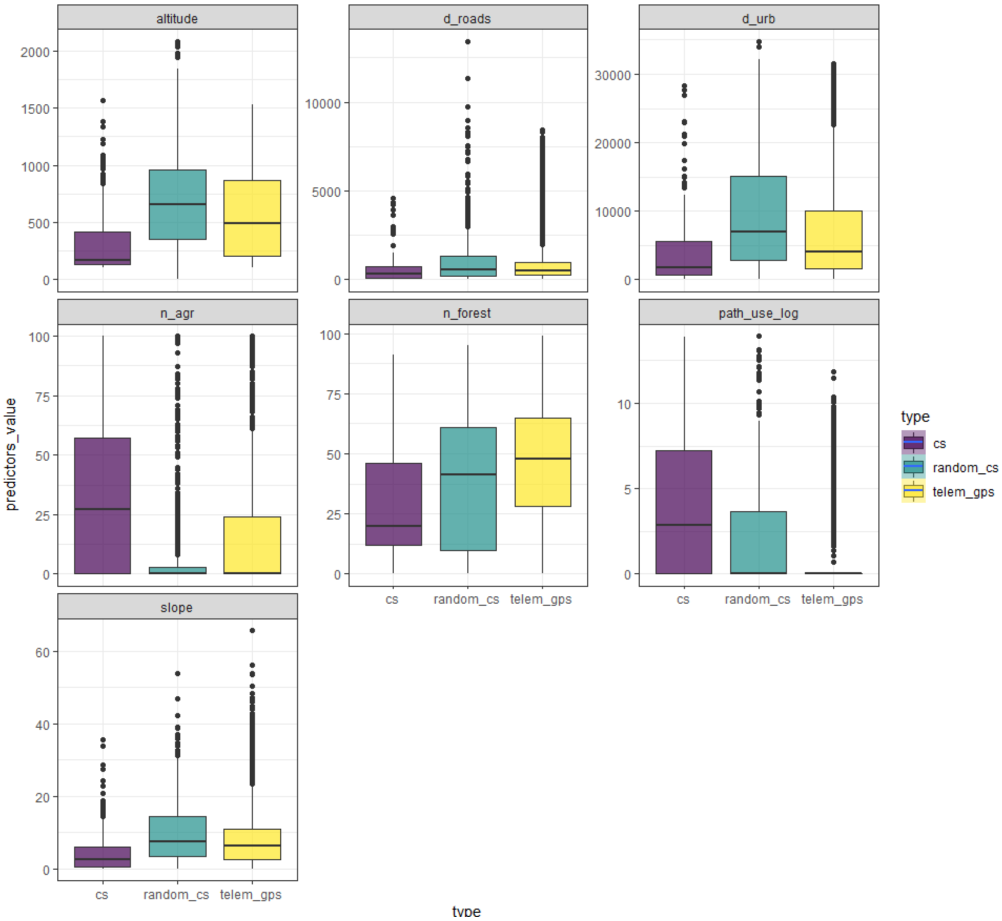


Figure A3. Histograms describing the distribution of the covariates used in the moose observer model and the moose RSF.

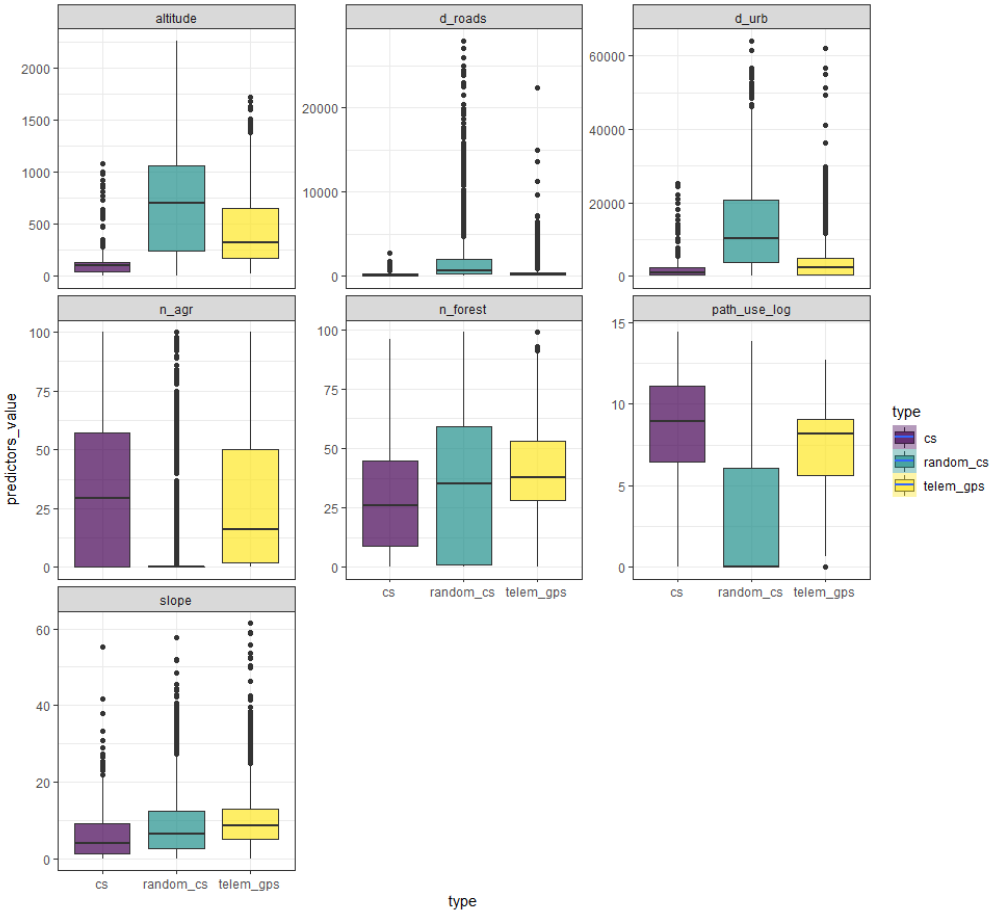


Figure A4. Histograms describing the distribution of the covariates used in the roe deer observer model and the roe deer RSF.

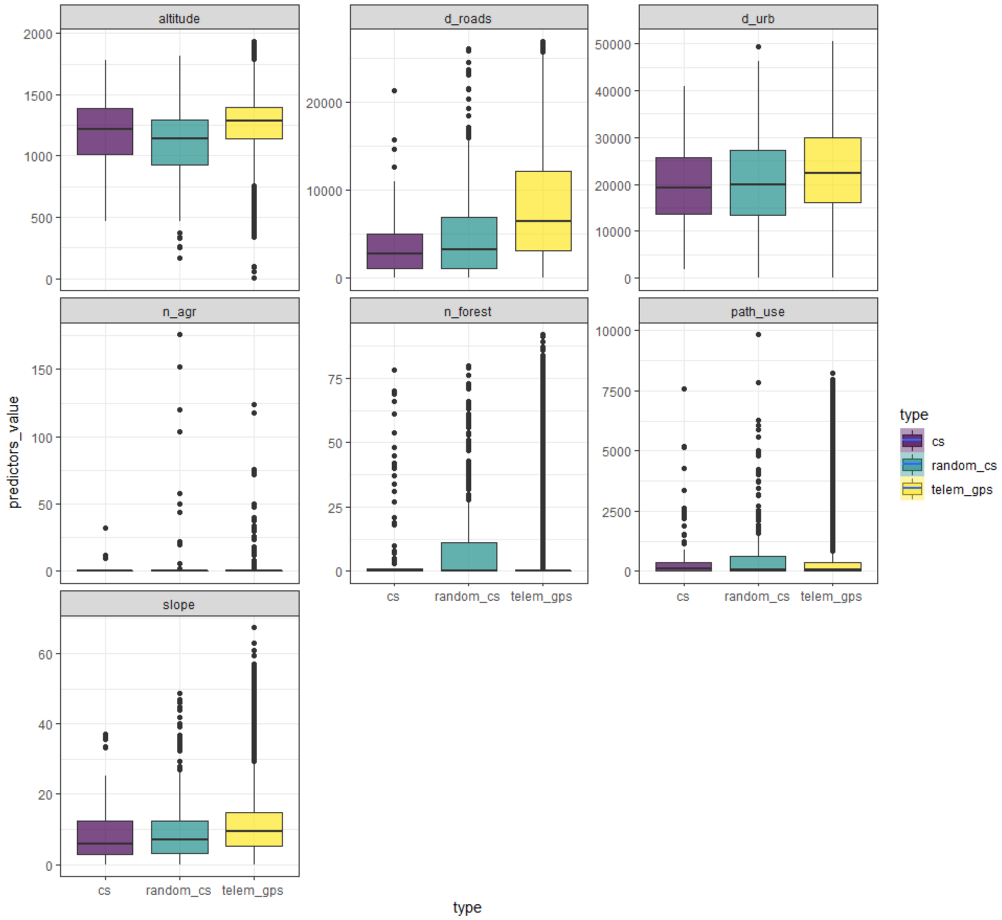


Figure A5. Histograms describing the distribution of the covariates used in the roe deer observer model and the roe deer RSF.

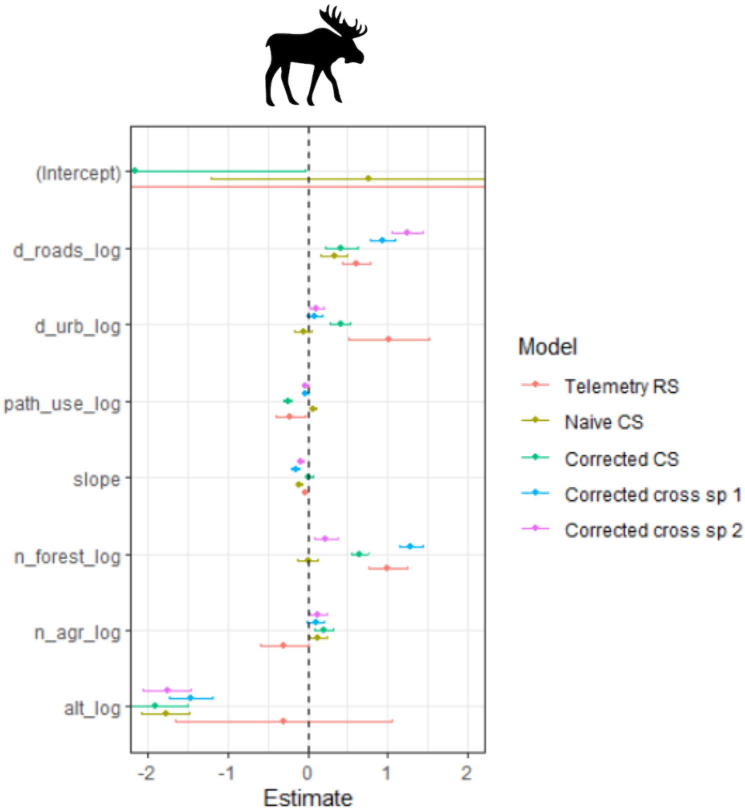
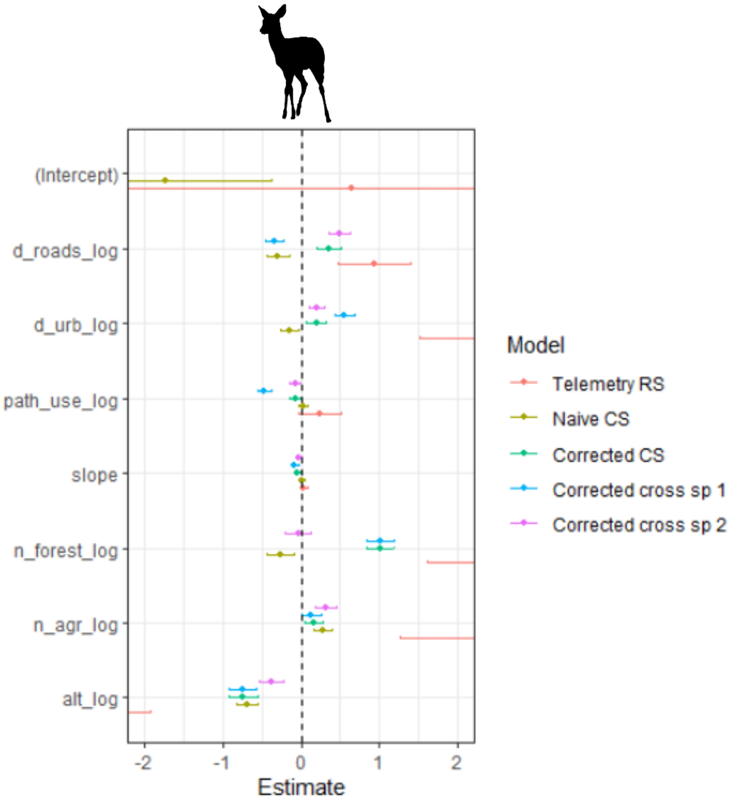


Figure A6. RSF parameter estimates for moose. In red are the parameter estimates obtained with the telemetry model, in yellow with the naïve CS model, in green with the corrected CS model, in blue with the corrected CS model using the roe deer observer model and in purple with corrected CS model using the wild reindeer observer model. Dots represent the mean parameter estimate and bars the 95 Credible intervals.



,

Figure A7. RSF parameter estimates for roe deer. In red are the parameter estimates obtained with the telemetry model, in yellow with the naïve CS model, in green with the corrected CS model, in blue with the corrected CS model using the moose observer model and in purple with corrected CS model using the wild reindeer observer model. Dots represent the mean parameter estimate and bars the 95 Credible intervals.

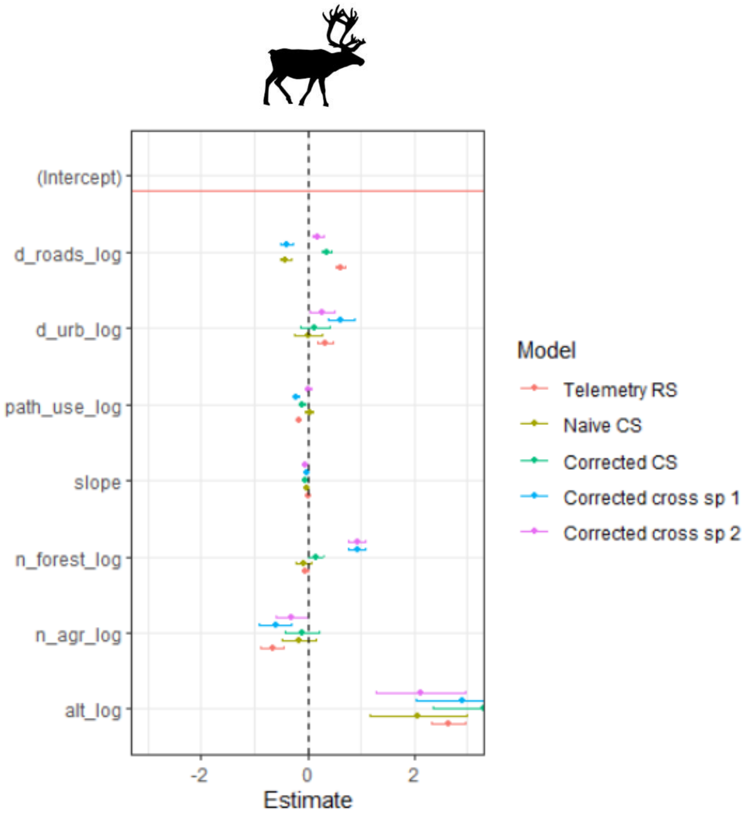


Figure A8. RSF parameter estimates for roe deer. In red are the parameter estimates obtained with the telemetry model, in yellow with the naïve CS model, in green with the corrected CS model, in blue with the corrected CS model using the moose observer model and in purple with corrected CS model using the roe deer observer model. Dots represent the mean parameter estimate and bars the 95 Credible intervals.

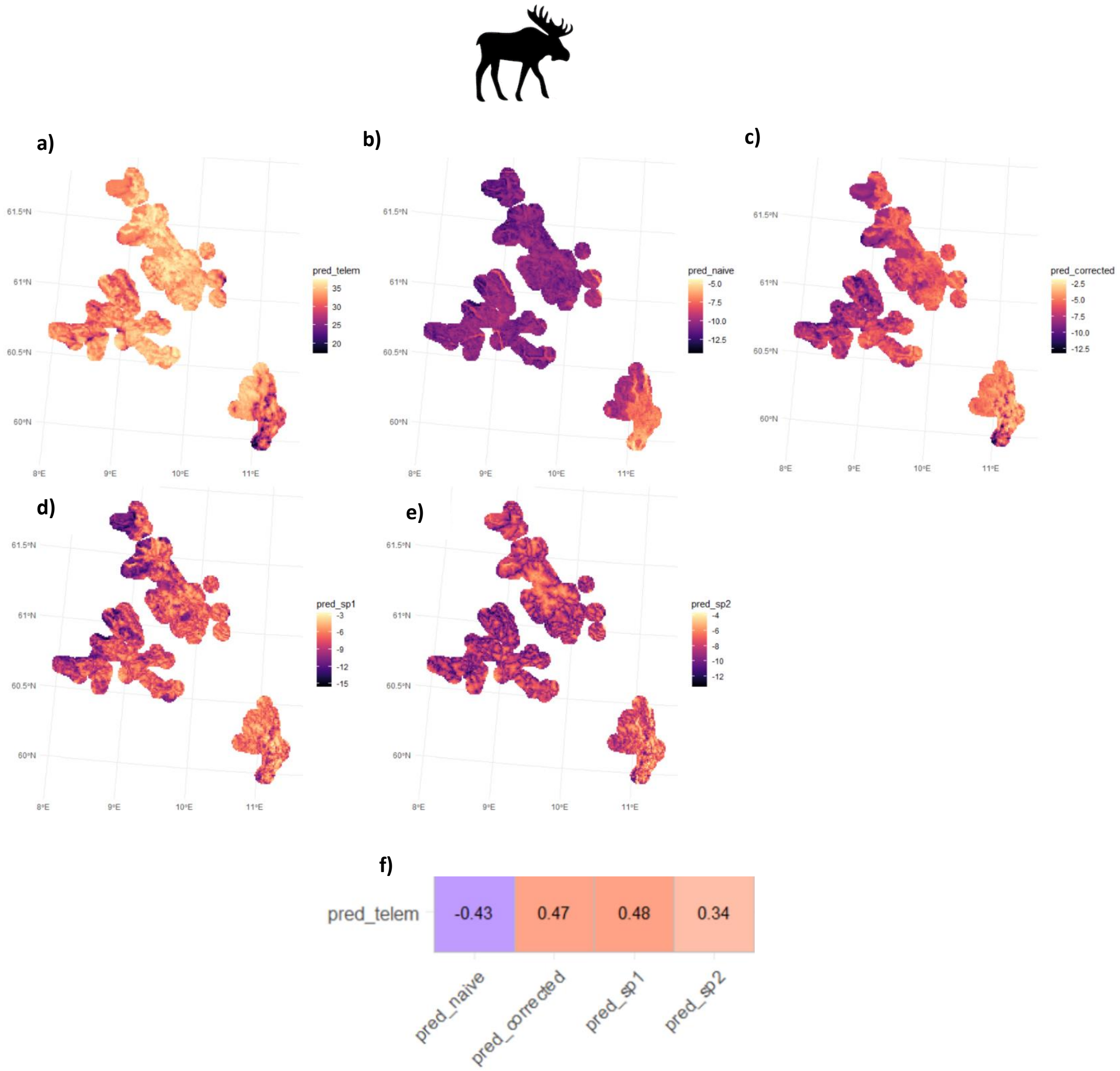


Figure A9. Moose suitability maps obtained using the mean coefficients of the a) telemetry model, b) naïve CS model, c) corrected CS model, d) corrected CS model using roe deer observer model and e) corrected CS model using wild reindeer observer model. On f) are the Pearson correlation coefficients between the habitat suitability maps obtained with the telemetry model and the ones obtained with the naïve CS model, the corrected CS model, the corrected CS model using roe deer observer model and the corrected CS model using wild reindeer observer model.

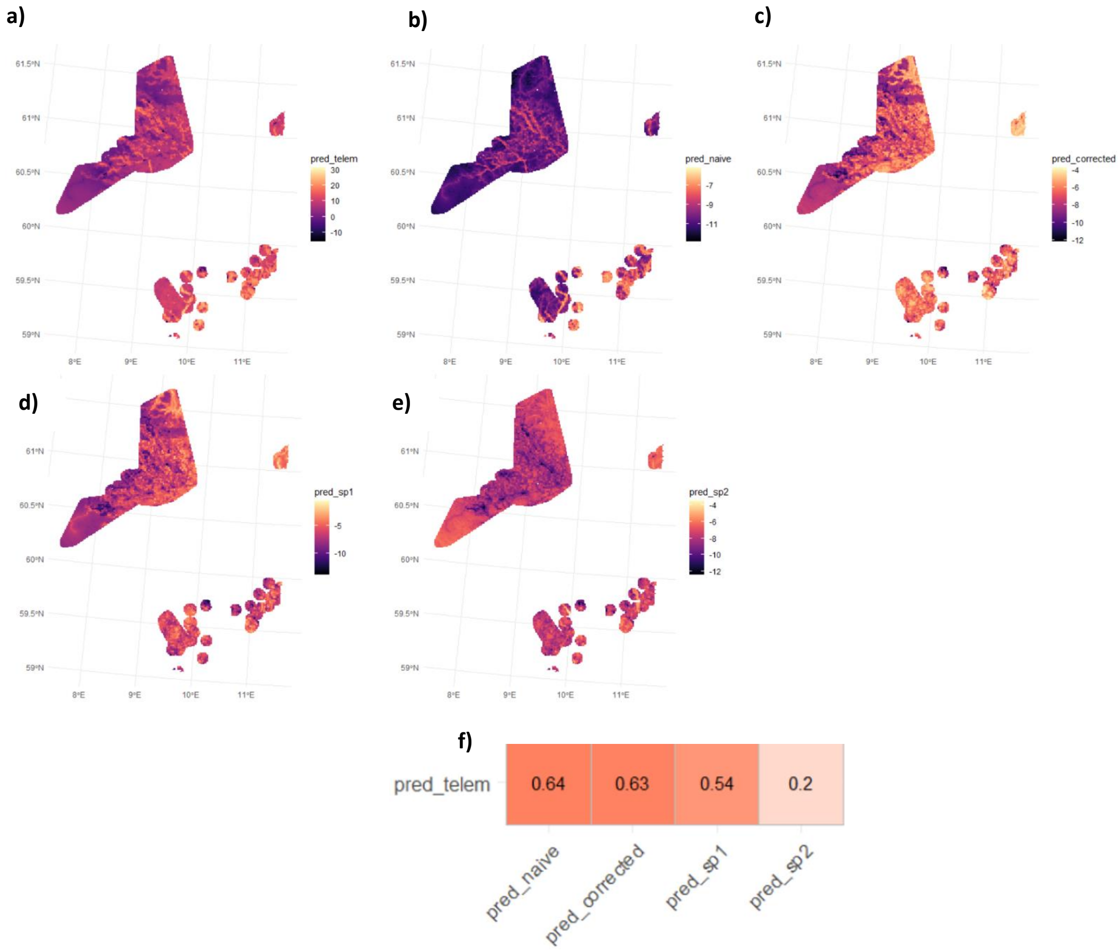


Figure A10. Roe deer suitability maps obtained using the mean coefficients of the a) telemetry model, b) naïve CS model, c) corrected CS model, d) corrected CS model using moose observer model and e) corrected CS model using wild reindeer observer model. On f) are the Pearson correlation coefficients between the habitat suitability maps obtained with the telemetry model and the ones obtained with the naïve CS model, the corrected CS model, the corrected CS model using moose observer model and the corrected CS model using wild reindeer observer model.

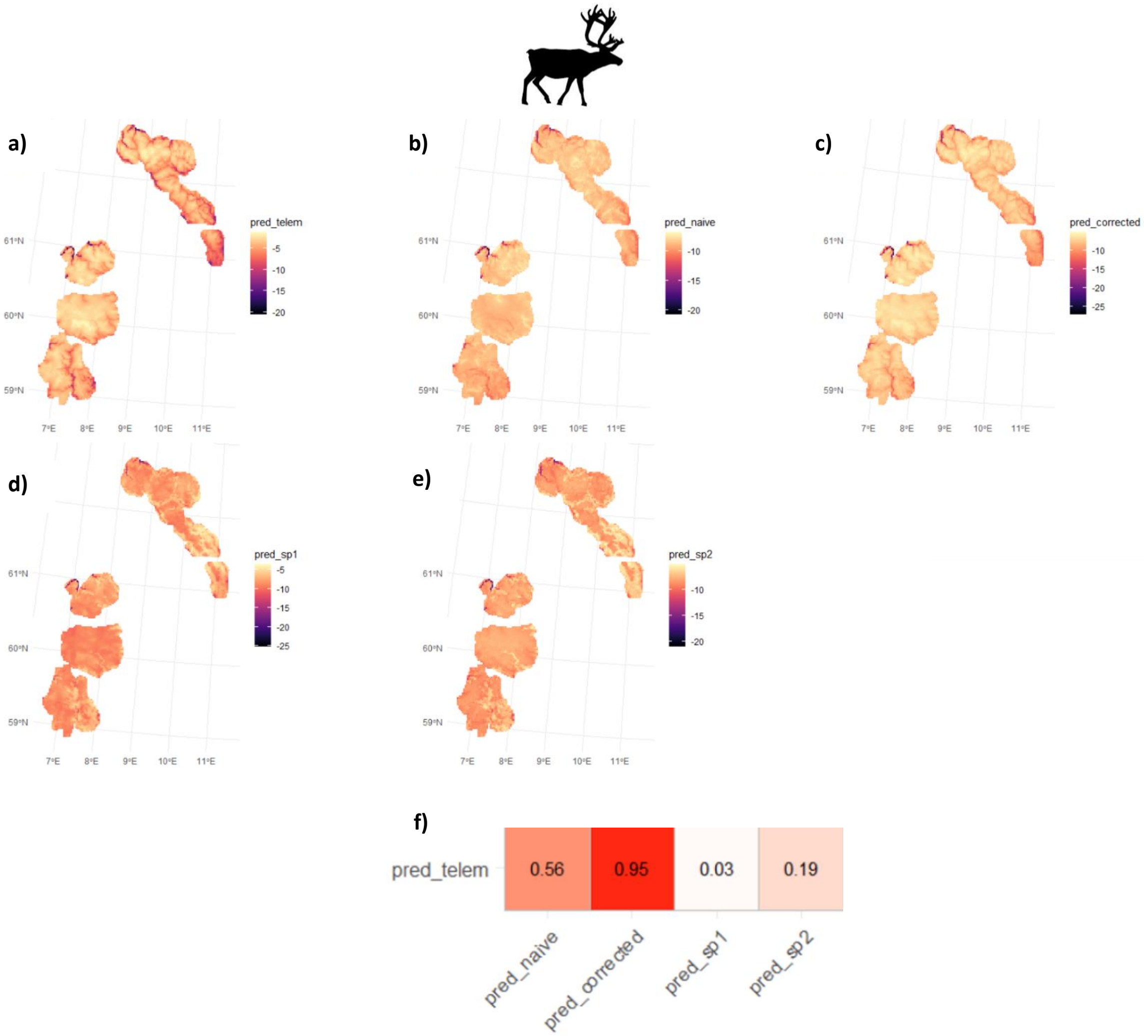


Figure A11. Wild reindeer suitability maps obtained using the mean coefficients of the a) telemetry model, b) naïve CS model, c) corrected CS model, d) corrected CS model using moose observer model and e) corrected CS model using wild reindeer observer model. On f) are the Pearson correlation coefficients between the habitat suitability maps obtained with the telemetry model and the ones obtained with the naïve CS model, the corrected CS model, the corrected CS model using moose observer model and the corrected CS model using roe deer observer model.

ISBN 978-82-326-5998-2 (printed ver.)
ISBN 978-82-326-5528-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology