General comments from Nicolas BECH

This preprint merits a revision

Dear authors of the article «Estimating abundance of a recovering transboundary brown bear population with capture recapture models »

This paper proposes a monitoring of the critically endangered Pyrenean brown bear population using the Pollock's closed robust design (PCRD), a capture recapture method. This study represents the first estimate of the abundance of the Pyrenean brown bear population and its trends since its reinforcement in 1996.

Based on the comments from 3 referees and from my own lecture, I believe that this paper fits well within the topic of PCI Ecology and represents a major contribution to global population studies and in particular to brown bear ones. Although this article may be acceptable for publication, it requires some clarifications, precisions and a proofreading by a native English speaker before publication. We suggest that authors consider the following suggestion as well as minor comments bellow.

Our answer:

Dear Nicolas BECH,
Thank you very much for having considered our manuscript entitled « Estimating abundance of a recovering transboundary brown bear population with capture recapture models », by Vanpé and colleagues. We are very happy to have received a positive evaluation, and we would like to express our appreciation to you and the three Reviewers for the helpful comments and suggestions.
We appreciated the three reviewers' valuable comments, which helped us to revise and considerably improve the manuscript. We fundamentally agree with all the comments made by the three Reviewers, so we paid heed to almost all of the reviewers' advice and suggestions and made every effort to address them.
We detailed below point-by-point responses to the three reviewers' comments. The corresponding revisions are highlighted into the revised manuscript by using the track changes mode of Word.
We believe that the content and the clarity of our manuscript have been considerably improved as a result of these revisions, and we hope that our revised manuscript is now suitable for being recommended for PCIEcology.
We would like to thank you once again for your consideration of our work and inviting us to submit the revised manuscript. We look forward to hearing from you.
Kinds regards,
For the authors,
Cécile Vanpé

Review by Susannah Woodruff

This paper reports on a new method for estimating abundance of the Pyrenean brown bear population. Given the increase in abundance and distribution in the population, new methods are warranted.

I have some concerns about some of the conclusions and the lack of transparency in the methods and results making it difficult to be confident in the conclusions that are reported. There are a lot of details provided for some of the methods but then wholly lacking for others.

Additionally, I think there are lot of more recent citations that should be included. Many of the citations are dated (very few are current except a couple with some of the same co-authors) and the citations are sparse throughout. The authors need to spend some time becoming familiar with the current literature and incorporating it into this manuscript.

There are many conclusions that are drawn that are inaccurate or could be compared to other, current literature but that does not occur. There is also a lot of important information that is not reported (see specifics and suggestions throughout my comments). It feels more like a report than a manuscript and I would not be comfortable citing this manuscript given the inaccuracies and lack of ability to draw conclusions given the incomplete reporting.

The primary author has a good handle on the English language but the manuscript would benefit immensely from proofreading and contribution by a native English speaker. Most journals prefer active voice when writing so instead of "sites were visited to collect samples", "we visited sites to collect samples". I recommend changing this throughout the manuscript for easier reading.

I also would recommend getting a paper in as "finished" as form as possible prior to submitting for peer-review, otherwise it is an immense amount of work for a reviewer. This took me more than 6 hours to review. It seemed like this was a pretty rough draft of a manuscript that is supposed to be ready for publication. PCI Ecology is specifically "not designed to be a free peer reviewing service for authors aiming to improve their articles before submission to a journal" but in this submission, that is what it seemed like. In the results and discussion, I spent less time making suggestions about how to reword sentences given the lengthy nature of the review already.

Because each page is numbered beginning with 1, it was a bit difficult to make comments since I had to check the page number before listing the line number. Continuous page numbering would be much easier to work with and is standard when submitting a manuscript.

Our answer: *We have answered point-by-point to the reviewer' comments, adding more information and details when lacking especially in the methods (see below). We asked a native English – speaker to check and correct the English language and grammar. We have now used active voice throughout the manuscript and a continuous page numbering, and incorporated more recent citations and compared our results to other current literature. We believe that the content and the clarity of our manuscript have been considerably improved as a result of these revisions and that our manuscript is now in as "finished" as form as possible.*

I have specific comments in the following pages.

Abstract Lines 18–20: This is not novel (i.e., that PCRD can provide reliable estimates) but it is written that this is something new that you found in this study. I would rephrase to say something to make it clear that this method worked for your species/study. Something like we used PRCD to reliably estimate abundance of Pyrenean brown bears, etc.

Our answer: *We have now rephrased some sentences of the Abstract section to made it clear that it is not novel that PCRD can provide reliable estimates but rather that our study provides evidence that this method works well and provides reliable estimates of size and trend for our brown bear population (see p.4, l.2-6).*

Main document:

Page 3

Line 7: I would change "almost impossible" to difficult.

Our answer: *We have now changed "almost impossible" to "difficult to implement" (see p.5, l.7-9).*

Line 7–8: Often relies on? How about something more like camera trapping or noninvasive/molecular techniques have been increasing or are commonly used methods now. And then you can cite some more recent studies that have used these techniques as examples.

Our answer: *We have now changed the sentence to "(…) so population monitoring consequently often needs to rely on non-invasive sampling methods (Long et al. 2008; Thompson 2013). Among them, molecular tools and camera trapping are commonly used methods now" and cited some recent reviews that provide evidence that these techniques are commonly used for large mammal monitoring"(e.g., Forsyth et al. 2022; Piel et al. 2022; Proctor et al. 2022)" (see p.5, l.8-11).*

Line 13: "so-called minimum detected size". Is there a citation for this? I don't know this as a "popular" monitoring method at least not by this name so I would include a citation here.

Our answer: *We have now made it clear that the common name that we can found in the literature is "minimum population size" as found for instance in Solberg et al. (2006) or Miotto et al. (2007) or Morin et al. (2022) but we chose to abbreviate it here as MDS for minimum detected (population) size (see p.5, l.15-19).*

Line 15–16: I would argue that this is changing. Costs are coming down for genetic analysis, analytical (statistical) techniques are improving so that fewer samples can be collected thus

reducing cost, time, and logistics. The cited paper is 9 years old at this point and a lot has changed since then.

Our answer: *We did not talk about non-invasive genetic sampling per se but about "MDS" or in other words "obtaining a MDS through exhaustive counts" that can be performed either using non-invasive genetic sampling but also using other methods (camera-trapping, visual observations…). As mentioned in Blanc et al. (2013), "exhaustive counts are (…) often expensive, time consuming and sometimes impractical". This is still true even though the costs of genetic analyses have somehow decreased since then. Whatever money available for population monitoring, exhaustive counts are difficult to obtain, even for small populations when the population is growing and spatially expanding. We have now made it clear (see p.5, l.21-23).*

Line 19: I would include more citations here aside from just Solberg et al. There are many relevant citations.

Our answer: *We have now included more citations aside from just Solberg et al., selecting two citations from 2022 and one citation from 2015 (see p.5, l.26).*

Line 22: remove the word exhaustively. It is unnecessary.

Our answer: *We have now removed the word "exhaustively" (see p.6, l.4).*

Lines 22–25 and onto page 4: Noninvasive DNA sampling does not "imply" these things you mention. You could, for example, know exactly the date a scat or hair sample was deposited (i.e., you saw the animal defecate or you had cleared the area of scat the day before and there was new scat when you resampled).

Our answer: *We have now changed this sentence to "non-invasive genetic CR models were specifically designed to account for issues such as…" to make it clear that not all non-invasive DNA-based sampling studies imply those issues (see p.5, l.6-7).*

Also you talk about CR surveys and then change to CR models. Maybe stick to models (i.e. change in line 21)?

Our answer: *We have now changed "CR surveys" to "CR models", as suggested by the reviewer (see p.6, l.2).*

I also suggest splitting up these lines and adding some citations. Could change to something like: Whereas CR models were originally limited to live-trapping studies, they have been adapted for use with non-invasive DNA sampling (insert citation here). Then go on to mention these issues and cite them: individual identification errors due to genotyping errors, uncertainty in the date of individual detection, and possibility of collecting multiple samples of the same individual across space within a single sampling occasion (Lukacs 2005; Lukacs & Burnham

2005). Although now as I get down to page 4, it seems that these lines and lines 8–10 on pg 4 should be combined. These 2 paragraphs should be restructured because they say much of the same thing.

Our answer: *We have now split up these lines, added some citations and changed those sentences to "Whereas CR models were originally limited to live-trapping studies, they have been adapted for use with non-invasive DNA-based sampling (Lukacs 2005; Lukacs & Burnham 2005). In particular, non-invasive genetic CMR models were specifically designed to to account for issues such as individual identification errors due to genotyping errors, uncertainty in the date of individual detection, and the possibility of collecting multiple samples from the same individual across space within a single sampling occasion (Lukacs & Burnham 2005; Petit & Valière 2006; Lampa et al. 2013)." (see p.6, l.4-11). However, we have not combined the two paragraphs since they talked about two completely different things: the first one talks about CR models adapted to non-invasive genetic sampling, whereas the second one deals with limitations of closed-population CR models (whether or not they have been adapted to non-invasive genetic sampling). We have now made this clearer (see p.6, l.12-13).*

Page 4:

Line 5: change "supposed" to assumed. You could also shorten by removing everything after detection probabilities.

Our answer: *We have now changed "supposed" to "assumed", but did not remove everything after "detection probabilities" since we think it is worth mentioning examples of factors ("individual attributes (e.g., age, body mass, social status) and habitat features (home-range location and composition)") that can lead to heterogeneities in detection probabilities among individuals (see p.6, l.15-17).*

Lines 3–8: I would split these sentences differently as the first one is long and then the sentence starting with "But…" seems out of place: In standard closed-population CR models, the population is assumed to be closed to changes in abundance both geographically (no immigration nor emigration) and demographically (no births nor deaths). Additionally, all individuals are assumed to have identical detection probabilities whatever their individual attributes (e.g., age, body mass, social status) and habitat features (home-range location and composition) (Otis et al. 1978), although these conditions are rarely fulfilled met in real populations of wild mammals (insert appropriate citation here).

Our answer: *We have now split the first sentence as suggested by the reviewer. However, we kept the sentence starting with "But..." (now starting with "However,") as before, since it refers to both the closure assumption and the identical detection probability assumption, and not only the second one as suggested by the new formulation of reviewer (see p.6, l.12-17).*

Line 9 and 13: remove probability and probabilities—i.e., change to detection heterogeneity or heterogeneity in detection

Our answer: *We have now removed "probability" and "probabilities" (see p.6, l.20 and 24).*

Line 11: what about citing Pollock here?

Our answer: *We have now cited Pollock (1982) (see p.6, l.22).*

Line 12: replace "study" with "estimate" since I think that is what you really mean here.

Our answer: *We have now replaced "study" with "estimate" (see p.6, l.23).*

Line 16–18: Simplify: In recent years, the implementation of Bayesian PCRD models has been made simpler by the development of user-specified models (insert citations here) (not sure this is the correct wording but something like this).

Our answer: *We have now changed this sentence to "However, it is only in the few last years that a Bayesian implementation of PCRD models has been made possible without ecologists having to code their own complex sampling algorithms (Rankin et al. 2016; Riecke et al. 2018)" (see p.7, l.10-13).*

Lines 19–21: Again, I would simplify this because it is very wordy: In the mid-1990s after decades of persecution, the brown bear (Ursus arctos) population in the Pyrenees Mountains at the border of France, Spain and Andorra (a study area figure would be good to reference here) had only five individuals remaining (Taberlet et al. 1997).

Our answer: *We have now rephrased the sentence and added the reference to the Figure 1 with the map of the study area, as suggested by the reviewer (see p.7, l.14-17).*

Line 24: Provide a reference where the Cantabrian population is (as in a country) because not everyone is familiar with this.

Our answer: *We have now indicated that the Cantabrian brown bear population is situated in north-western Spain (see p.7, l.20-21).*

Line 26: if you are going to use the word "high" here, I would state what the rate is. High is a subjective word so perhaps provide the rate along with each citation.

Our answer: *We have now stated what the consanguinity coefficient F is (0.132 in average 2020) (see p.7, l.22-23).*

Lines 26– 3 (on the next page): Simplify……. Thus implementing reliable methods to accurately estimate population abundance trend over time is crucial to monitor the conservation status of this population threatened with extinction and implement successful management plans. I also recommend changing one of the "implement" to develop or something else for smoother reading.

Our answer: *We have now simplified the sentence as suggested by the reviewer and changed the first "implement" to "develop" (see p.7, l.24-26).*

Page 5:

Line 4–5: Make it clear that this is what currently happens. So, currently, monitoring of the Pyrenean….

Our answer: *We have now added "currently" before "monitoring" to make it clear that this is what currently happens (see p.8, l.1).*

And again, I would simplify: Currently, monitoring of the Pyrenean brown bear population relies on either opportunistic collection of bear data or samples (e.g., scat or hair) by the public with no specific sampling design or a systematic sampling approach (Sentilles et al. 2021a; Sentilles, Vanpé & Quenette 2021).

Our answer: *We have now simplified the sentence as suggested by the reviewer (see p.8, l.1-8).*

Line 8: Change to: Similar to many large carnivore populations…..

Our answer: *We have now changed to "Similar to many large carnivore populations…" (see p.8, l.8).*

Lines 8–12: I think you can remove some words here such as "highly", "divided in", "specific" and it will still say the same thing, just simpler.

Our answer: *We have now removed "highly", but kept "divided across" "specific" as these words were important to understand the division of French regions in counties and the specific autonomous status of Val d'Aran inside the autonomous region of Catalonia in Spain (see p.8, l.8-13).*

Also, you don't specifically mention ANdorrra here but I assume you are referencing it. I would include Andorra here since you mention it in the next sentence but it sounds like somewhere different in that sentence.

Our answer: *We have now mentioned Andorra in the sentence (see p.8, l.10-11).*

Lines 16–19: Simplify: The aim of this study was to use cross-border non-invasive sampling data and collected from 2008 to 2020 in France, Spain and Andorra and PCRD and for which individual identification was possible through genetic analyses or visual evidence to provide the first published estimates of annual abundance of the Pyrenean brown bear population. You mentioned what the data is above so I don't think you need to specify it again here.

Our answer: *We have now simplified the sentence as follow: "The aim of this study was therefore to use cross-border non-invasive sampling data collected from 2008 to 2020 in France, Spain and Andorra, for which individual identification was possible through genetic analyses or visual evidence combined with PCRD modelling to provide the first published estimates of annual abundance of the Pyrenean brown bear population, while minimizing bias due to heterogeneity in detection probabilities among individuals." (see p.8, l.13-18).*

End of page 5 and onto page 6: I would remove the entire section on brown bear biology. I am not sure why this is included but it is not relevant to the current manuscript.

Our answer: *We have now removed the entire section on brown bear biology (see p.9, l.23-25 and p.10, l.1-12).*

Page 6:

Line 17: are you saying the area ranges over 10,000 sq km in 2020? Or the population? This isn't clear. Also, active voice would be preferred, i.e., We carried out our study….or our study area was….

Our answer: *We meant the population, hopefully this is not confusing. We have now used the active voice instead of the passive one (see p.10, l.15-18).*

Line 18: Same here—active voice and simplify: We used four different non-invasive methods……..

Our answer: *We have now used the active voice instead of the passive one (see p.10, l.19-20).*

Line 20: Simplify.. Systematic by trails (ST) corresponded to walking 8 to 10 km transects (from long),

Our answer: *We have now changed to "Systematic trail walking (ST), equivalent to transect surveys (from 8 to 10 km long)" (see p.10, l.21-22).*

Lines 23–25: I am not sure exactly what this means because of the French-English translation so this needs to be clarified. I can't figure out what this is describing: Trails were set in function of available bear habitats and passage areas detected using VHF and GPS collars or bear presence signs. I think you are talking about how the transects (i.e., trails?) were delineated and decided upon but it needs clarification.

Our answer: *To clarify, we have now rephrased this sentence as follow: "To optimize bear detection, we set transects in the most favourable bear areas in terms of habitat quality and in bear passage areas detected using VHF and GPS collars or presence signs" (see p.11, l.4-6).*

Line 25: accompanied occasionally by a scat detection dog? Like in certain years or months or locations? This needs more details here even if it is in the cited work so the reader can know the basics of the methods.

Our answer: *We have now removed the mention that the staff was "accompanied occasionally by a scat detection dog" since it concerns, as mentioned in Sentilles et al. (2021b),"planned operations (...) initiated during the year, (...) in order to search for bear scats around dens, diurnal resting places, in areas where females with cubs have been recently observed, or potential feeding areas during mast years" as well as opportunistic monitoring (as mentioned later in the opportunist section : "Since 2014, verification of testimonies and damage reports have been occasionally carried out with the help of a scat-detection dog trained to search for brown bear scats (Sentilles et al. 2021b)") but not "Systematic by trails (ST)" (see p.11, l.6-7).*

Page 7:

This entire section describing the 4 methods needs appropriate citations—a lot of people have used these methods and I am guessing you based your sampling on some of these studies.

Our answer: *We did cite several references in relation to the 4 methods (De Barba et al. 2010; Woods et al. 1999; Castro Arrellano et al. 2008; Gervasi et al. 2010; Mowat & Strobeck 2000; Boulanger et al. 2002; Sentilles et al. 2021b; Vanpé et al. 2021) (p.10, l.21-25; p.11, l.1-26; p.12, l. 1-25; p.13, l.1-25). We have now added a few more when appropriate: Parres et al. 2020 (see p.11, l.15), Burton et al. 2015 (p.12, l.16), Berezowska-Cnota et al. 2017 (see p.11, l.12).*

Line 1: What do you mean “immediate surroundings”? Was there some kind of delineation of how far off the trail you could go to look? Or how far off the trail you could see and collect sign or was this random?

Our answer: *“Immediate surroundings” mean that we search for bear indices by walking on the trail and checking trees and ground by eye at a distance of few meters from the trail. We have not added a definition of immediate surroundings so as not to make the text more cumbersome, but we have now added the reference of De Barba et al. (2010), who used a similar method and the same expression “Immediate surroundings” in their paper (see p.11, l.6-8).*

Line 2–3: “scattered along each itinerary”. Do you mean they were along the trail on each transect?

Our answer: *We mean that between five and seven hair traps were set up on trees found along the trail. We have now replaced “itinerary” by “trail” to make it clearer (see p.11, l.9).*

Line 4: what is smola?

Our answer: *We have now explained what “smola” is (beechwood tar) and added a reference (Berezowska-Cnota et al. 2017) of a study using similar approach and smola (see p.11, l.9-12).*

Line 6: when you say “similar to the camera method”, I would note that this is described below since you haven’t described it yet.

Our answer: *We have now mentioned that the systematic by camera traps method is described below (see p.11, l.13-14).*

Line 10: Why this height?

Our answer: *This is the optimal height for brown bear (see the reference Woods et al. 1999 just mentioned in next sentence). We have now added the Woods’ reference just after the mention of the cm height as well as 2 other references (Kendall & McKelvey 2008; Quinn et al. 2022) (see p.11, l.16-21).*

Lines 13–15: Where do these grid cell sizes come from? And what are the known female range areas and where do they come from? More examples of where citations are needed.

Oh, okay, now I got to the next lines and see where it came from. So I would put Lines 15–18 before lines 13-15. But you still need a citation of where the Pyrenees bears home range size comes from.

Our answer: *We have now put lines 15-18 before lines 13-15 and added a reference for the average adult female home range size in the Pyrenees (see p.11, l.21-26 and p.12, l.1-5).*

Lines 18–20: how did you predict the best bear habitat? "bear expert opinion" included what? Was this systematically determined in some kind of expert elicitation? Is it cited somewhere?

Our answer: *We have now mentioned that bear expert opinion included knowledge about tree types or tree species, with characteristics that make them more conspicuous for rubbing, adding a reference (González-Bernardo et al. 2021; Proctor et al. 2022) (see p.12, l.5-9).*

Line 22: remove the words "automatic-triggered" since this is clarified with "movement detection" on lines 254 and 25 and "essentially", as it is unnecessary

Our answer: *We have now removed the words "automatic-triggered" and "essentially" (see p.12, l.11-12).*

Line 25: how were these areas with frequent animal passages determined? And I assume by animal you mean bear? So areas of high use by bears? But clarify how you determined these areas.

Our answer: *These areas with frequent animal passages were determined by looking for animals' trails from all large mammals, which are visible in the vegetation and on the ground and that are often used by bears, as well as bear passage areas detected using VHF and GPS collars or bear presence signs. We have now clarified this (see p.12, l.16-19).*

Page 8

Lines 1–2: as you did in the previous section, state that each station was visited to collect samples and maintain cameras, or whatever it was you did there.

Our answer: *We have now clarified that visits were planned to collect samples and maintain cameras (see p.12, l.19-21).*

Lines 2-3: You could simplify this by saying you followed the same layout as above instead of repeating the 4 x4 km, etc. again.

Our answer: *We have now simplified by saying that we followed the same layout as above for SBHT (see p.12, l.21-22).*

Lines 4-7: I don't understand this. You need to describe this more. I understand if there was a radio collar or ear tag or a really distinguishing physical feature you didn't do genetic analysis so you just counted this as a "capture"? This would definitely affect the problem of heterogeneity since these individuals could potentially be more easily identified if, for example, you had a confirmed "capture" by the camera but if you had used the DNA and it failed, you would not have a detection. So any individuals that were not visually identifiable and then if the DNA didn't amplify they would be detected less often and with lower probability of detection. If you do not address this later on, it is something that needs to be addressed. How many "captures" or instances did this include?

Our answer: *Natural and/or artificial marks (colouration, scars, GPS collars, or VHF ear tag transmitters) may have helped temporally or permanently identifying some of the individuals of the population on photos or videos, causing potentially a bias in detection probabilities among individuals each month. Indeed, several individuals can use consecutively the same hair trap station. In that case, because of hair mixture issue, genetic analyses were not carried out. If after analysis of photographs or videos associated with the hair trap station, bear could be identified based on natural or artificial marks, this was considered as a capture. However, this issue concerned only a few individuals each year (for natural marks: between 0 and 3 individuals according to years; for artificial marks: 2 individuals in 2008-2009, 0 in 2010-2015, 1 in 2016-2018, 4 in 2019 and 1 in 2020) and a few indices per individual (since natural marks are cryptic and not always visible on photos and videos). And in the vast majority of cases, these individuals have also been detected independently each month through genetics on scats and hair. So we are confident this should not have affected significantly individual capture histories and monthly detections. We have clarified this in the Discussion section (see p.24, l.12-20).*

Lines 8–9: What is the bear potential range? How was this determined?

Our answer: *The bear potential range is defined as the area surrounding bear presence, allowing random locations (for bear absences) to fall where bears could have visited (15 km from the edge of presence) (see Martin et al. 2012). As explained in Martin et al. (2012), the study area was divided into 5x5 km cells, which approximately corresponds to the size of a bear's seasonal home range (Naves et al. 2003). Bear presence was classified as 1 in a cell where at least one observation occurred. We have now clarified this (see p.13, l.8-10).*

Line 10: What are "eating clues"? And scratches? How were these determined to be from brown bears?

Our answer: *Eating clues are carcasses of wild or domestic preys, overturning of a large stone or anthill and bee or wasp swarms burst open. To confirm that these were from brown bears, we look for evidence of bear clues (e.g., footprints, claw marks, hair, scats) next to or on the*

carcass, anthill and bee or wasp swarm. Bear signs of predation (punctures, bruises) and consumption on the carcass can also provide clues. The overturning of a large stone can be an indication of a bear looking for insect larvae under the stones. Scratches are claw marks on trees. Bear claw marks are parallel, regularly spaced 2 to 4 cm apart and comma-shaped. We have now clarified this (see p.13, l.11-22).

Lines 10–11: "gathered by mountain users". How were these reported and to whom? And what kind of data was collected in these reports? Were people trained in collection of data and samples. Did you provide sample collection materials to them? And how were these screened for verification of the data? And then how was this data used in the analysis?

Our answer: *We have now clarified that these mountain users are any mountain users (e.g., hikers, foresters, hunters, skiers, fishermen, shepherds) and so they are evidently not trained in collection of data and samples and we did not provide to them any sample collection materials. Mountain users report their observations to the bear team of the French Biodiversity Agency (OFB), who is in charge by the French Minister of Ecology of brown bear monitoring in the French Pyrenees. Testimonies are examined and approved by an expert agent from OFB. A conclusion as to its validity as bear evidence, "confirmed," "probable," "doubtful," or "false," is given to each putative bear presence sign that could be verified, on the same day or a few days after its transmission, according to the elements necessary for their verification. Bear observations are validated only if an indirect bear clue (scats, hair, footprints) is found at the sighting site or if a photo or video is provided by the observer. To confirm that eating clues are from brown bears, we specifically look for evidence of associated bear clues close by (e.g., footprints, claw marks, hair, scats). If the elements are not sufficient to make a decision or if the observer could not be found for the statement of his/her testimony, the evidence is classified as "impossible expertise". Only confirmed bear clues are included in our analyses. We have now clarified this (see p.13, l.3-22).*

Lines 16–18: I am assuming these are the areas you are referring to above when you mention autonomous regions and you should specify that above. Also, not all readers know where Catalonia and Aragon and Navarra are so specify this.

Our answer: *We have now clarified that Catalonia, Aragon and Navarra are autonomous regions situated in Spain (see p.8, l.12 in the Introduction section and p.14, l.1-3) as we can see also in Fig. 1.*

Lines 18–20: you say you focused on noninvasive data but what about the data from individuals that you identified on camera?

Our answer: *As mentioned (see p.14, l.17-22), "We used all validated non-invasive brown bear signs collected in the Pyrenees from 2008 to 2020 (Table S2) and for which individual identification was possible. Individual identification of bears was primarily based on genetic analyses of hair (stored dry in envelopes) and scats (stored in microtubes filled with 96% ethanol) non-invasively collected in the field, as well as visual evidence (colouration, scars,*

GPS collars, or VHF ear tag transmitters) obtained by remote cameras when available (Sentilles et al. 2021b). So indeed, data from individuals that we identified on camera were included in non-invasive data (see also our answer above for comment related to p. 18, Lines 4-7).

Line 21: You paid particular attention to the date for the opportunistic monitoring? Or for all monitoring?

Our answer: *We have now specified that we paid particular attention to the date for all the four monitoring methods (see p.14, l.10-13).*

Page 9:

Line 1: What do you mean "validated"? How did you validate it—and what was required for validation? I also thik you should include more details about your collection methods here. I am guessing you followed a standard protocol? And had citizens (mountain users) collecting samples in a systematic way too? Provide citations for these methods too.

Our answer: *We have now provided information and references on validation of putative bear signs and on citizens collecting bear signs above when describing the four monitoring methods (see p.13, l.13-22).*

Lines 6–7: This seems out of place here. Maybe up on page 6 would be a better location? Either starting on line 17 or 19 would be my suggestion.

Our answer: *We have now moved this to the previous section (see p.14, l.13-14).*

Line 8: This acronym (LECA-CNRS) needs defining.

Our answer: *We have now defined the acronym (see p.15, l.7).*

Line 11: what/where is "our laboratory"? Is that different from LECA-CNRS? And I would suggest that if you are going to specify how they were analyzed from 2008–2012 (i.e., multi-tube PCR) vs 2013–2016 (i.e., Illumina), you also say how they were analyzed in "our laboratory". Otherwise, you could simplify and say you analyzed the samples with multiple methods and provide the details for all methods in the supplementary materials.

Our answer: *We have now specified that our laboratory is at the ANTAGENE Company and that from 2017 to 2020, samples were analyzed using a new multiple-tubes Polymerase Chain Reaction (PCR) approach (see p.15, l.11-13). In Supplementary materials, we only provide details for methods used at the ANTAGENE Company since methods used by LECA-CNRS have already been published.*

Lines 12–13: What do you mean 4 repeats? You need to describe these methods. And what kind of genotyping errors are you referring to? How did you calculate genotyping error? You mention further information on genotyping error rate can be found in these references but are these the methods you followed?

Our answer: *We used "a multiple-tubes Polymerase Chain Reaction (PCR) approach (Taberlet et al. 1997)" and "a minimum of four repeats for each sample was carried out to avoid genotyping errors associated with low quantities of DNA (Miquel et al. 2006)". We provided relevant references to avoid describing in details those very classical methods used for rare DNA genotyping and genotyping error rate estimation. Taberlet et al. (1996) was cited for instance 1618 times and Taberlet et al. (1997) 655 times. Further information on genotyping error rate can be found in these references since indeed they are the methods we followed and samples from these papers are from our study population. We have thus now just briefly clarified that it consists in repeating each DNA amplification independently for each locus several times (see p.15, l.7-11).*

The entire section Population abundance estimation using capture-recapture models needs citations and better description. For example there are many more studies using PCRD than Kendall et al. And how specifically did you account for imperfect detection and temporary emigration?

Page 10: Lines 16-18: Another example of where citations should be used (along with, again, all throughout this section).

Our answer: *We have now extensively rewritten this section following the referee's suggestions. We have added references of the PCRD approach to bear populations, and a few other methodological references. We also clearly define what we mean by temporary emigration (see p.15, l.22-24, p.16, l.1-26 and p.17, l.1-26 and p.18, l.1-14).*

What do you mean you were exploring effects on survival etc? Effects of what? And you need more description of what is meant by "detection structures".

Okay I see now what you mean you were exploring. I would combine/restructure the first two sentences here to state that you built the 24 models to explore the effects on survival, etc.

A table would be really useful here to explain the models and the different "structures". It is difficult to follow in the text.

And now I found the table—you should mention Table 1 here and again in the results.

Our answer: *We have now restructured the first two sentences and added a reference to Table 1 here as recommended by the reviewer (see p.17, l.18-26). Table 1 was already mentioned in the Results Section.*

Line 24–25: You can delete that allows calling….and just state that you used RMark in R and cite both.

Our answer: *We would prefer not deleting "that allows calling the Mark program" to avoid confusion between the standalone Windows application Mark and the R package RMark; we do think it is important to specify that the RMark package just calls the Mark program, and to credit the work of all colleagues involved in the development of these pieces of software (see p.18, l.1-3)*

Same lines, I don't know what this means: "Because we run into boundary estimates issues…." like upper and lower bounds of abundance? Or physical boundaries across countries? I am guessing you mean with the estimates but why did you use both RMark and Bayesian? Why did you not just use Bayesian if you had these problems? You need to explain why you used both—or at least why you decided to report on both.

Our answer: *We agree that our wording was unclear. By boundary estimates issues, we meant that some probability estimates hit the boundary of the domain of define, and were estimated at 0 or 1 without any possibilities to quantify uncertainty with standard errors or confidence intervals. We used the frequentist approach to carry out model selection, which is much more stabilised in the frequentist framework and avoid prohibitive computation times, then the Bayesian approach for parameter estimation to avoid the boundary estimate issues. We have now clarified this (see p.18, l.3-10).*

Page 11

Results: I would suggest following the same flow as you did in the methods. So maybe state how many samples were collected and then point to Table S1. If there were only 2,524 genetic samples and 10,019 validated samples (still don't know what that means), what was the breakdown of the other samples? This should be in a table somewhere too—a complete breakdown of the samples that were collected by year and type. You also could include by year the number of individuals identified each year. And the range of times individuals were detected—maybe the median and min/max.

Our answer: *We have now added a new Table S2 (see p.52) providing information about the total number of validated non-invasive brown bear signs (e.g., scats, hair, tracks, visual observations, damages, photos / videos) collected in the Pyrenees, the total number of validated brown bear samples (i.e. scats and hair) collected in the Pyrenees, the number of samples (among collected sampled) genetically analysed by the French molecular laboratory LECA or ANTAGENE, the number of brown bear samples (among analysed samples) successfully genotyped and the number of different brown bear genotypes identified (among successfully genotyped samples) per year between 2008 and 2020. Individuals have been detected at up to 61 different secondary occasions (median = 4, min = 0) over the study period from 2008 to 2020 (which include 65 occasions of capture in total). We have now added this information (see p.18, l.18-25).*

Line 16: you should report the estimates ± SE or some confidence interval, not just say "the estimates were around. The estimates are nearly identical and if you report 2 decimal places, they are identical. I don't think the difference in SE in survival of subadults between 0.028 and 0.029 is important! It would make it simpler and correctly reported.

Our answer: *We have now reported estimates ± SE as recommended by the reviewer, although this repeats information from Table 2, and we have now rephrased the sentence to clarify that the estimates are nearly identical (see p.19, l.6-9).*

I didn't realize in Table 2 (until I looked at it several times) that 1 column is for model 1 and 1 for model 2. You should label the columns so that is clear.

Our answer: *We have now labelled the columns from Table 2 so that it is clearer that column 1 is for model with random temporary immigration and column 2 is for model with Markovian temporary immigration (see p.44).*

In your methods, you also don't mention that you are estimating anything except population abundance, but you also were estimating survival. Clarify this in the methods.

Our answer: *The main objective of our PCRD modelling is to estimate annual population abundance. However, as mentioned in the Materials and Methods section, we fitted 24 different models in total, with "four detection structures (constant, time-dependent considering variation between and within primary occasions and heterogeneous using finite mixtures), two survival structures (constant and age-dependent using three age classes: i.e. cubs < 2 year old, subadults = 2-3 years old and adults > 3 years old) and three emigration structures (constant, random and Markovian)" (see p.17, l.18-26). The parameter estimates for the best-supported model therefore necessarily provides survival estimates, as well as estimates for the detection probability, the probability of leaving the study area and the probability of remaining in the study area and we reported these estimates in the Results section (see p19, l.6-12).*

Table 2 refers to "class 1 of mixture" but this is never described anywhere else. I assume this is referring to the heterogeneity but the wording needs to be clear as to what you are estimating here.

Our answer: *The referee is completely right. We have added a description of what we mean by finite mixtures in the method section to make explicit the quantities we estimate (see p.17, l.20-26).*

Page 12:

The entire discussion lacks relevant citations. Typically in the discussion you would discuss your results and compare them to other studies. This is rarely done. And when it is, it is compared to very old studies—there are so many current published studies using this same analysis that would be good for comparison.

Our answer: *We have now included more citations and discussed more our results in comparison with those from other studies (see Discussion section from p.20, l.13 to p.27, l.20 and especially p.26, l.4-26 and p.27, l.1-6).*

Line 2: You reference fig 2 and include MRS in Fig 2 but it has not yet been mentioned. If you are going to include this in your results, it needs to be discussed in the methods. Given you also say you are monitoring the trend, you should mention it in the results. Something like it was a generally increasing trend through time.

Our answer: *We have now introduced MDS and MRS in the Introduction section (see p.8, l.17-26 and p.9, l.1-12) and moved the results concerning the comparison between PCRD, MDS and MRS in the Results section (see p.19, l.24-25, p.20, l.1-9). We have also mentioned the trend in the Results section (see p.19, l.21-23).*

Lines 5-10: you can substantially shorten this since you have already said all of this earlier in the manuscript. You could even remove it and jus start with "To assess the effectiveness of the translocation and……"

Our answer: *We have now substantially shorten the sentence (see p.20, l.14-15).*

Pages 12-13: Minimum retained size. This is the first mention of this and seemingly comes out of the blue. You need to include it earlier.

Our answer: *We have now included a definition of Minimum Retained Size (MRS) earlier in the Introduction section (see p.3, l.12-14).*

Page 13:

Line 4: this isn't clear—"individuals still alive" are more detectable than what? Dead individuals—of course they are so I am not understanding what you mean here.

Our answer: *We have now removed this unclear example (see p.21, l.25-26, p.22, l.1).*

Lines 11–12: I don't see how Table S1 clarifies this. The number of samples is not largely different from other years. This provides support perhaps for why you should include the

number of samples collected in each year and the number analyzed. And then if there is a big difference (as you say there is in 2017and 2018), you could then mention why that is the case.

Our answer: *To clarify, we have now provided information about the percentage of collected bear samples that have been genetically analysed for the years 2017 and 2018 compared to the average percentage over the study period (see p.23, l.6-22). We have also provided in a Table S2 full details on the total number of validated non-invasive brown bear signs (e.g., scats, hair, tracks, visual observations, damages, photos / videos) collected in the Pyrenees, the total number of validated brown bear samples (i.e. scats and hair) collected in the Pyrenees, the number of samples (among collected sampled) genetically analysed by the French molecular laboratory LECA or ANTAGENE, the number of brown bear samples (among analysed samples) successfully genotyped and the number of different brown bear genotypes identified (among successfully genotyped samples) per year between 2008 and 2020 (see p.52).*

Lines 21–23: This is not an accurate conclusion. Random temporary emigration does not impact abundance or survival estimates, only detection probability (see Schwarz and Stobo 1997. Estimating Temporary Migration Using the Robust Design). What about the detection heterogeneity in contributing to this bias? It makes sense if you have, for example, trap happy individuals coming to your baited sites (you never report any of this kind of information which would be useful), or you are detecting your tagged individuals more frequently (which would be unsurprising since you are more likely to have genotyping failure than not detect a radio collar), then your estimates will be negatively biased.

You do address this a bit in the next paragraph but not in the sense of the abundance estimate.

Our answer: *We agree with the reviewer and have now deleted the sentence: "However, note that MRS can be subject to sampling bias if some specific individual types (e.g., more detectable individuals or individuals still alive) are more prone to be detected a posteriori)" (see p.21, l.25-26, p.22, l.1).*

Page 14:

Line 9–10: You talk about 4 individuals with long detection histories, how about reporting in the results the range of length of time of monitoring.

Our answer: *We have now specified in the Results section that "individuals have been detected from 1 to 61 different capture occasions (median = 5.5, mean ± SD = 10.25 ± 12.23) over the study period from 2008 to 2020 (which include 65 occasions of capture in total)" (see p.18, l.23-25).*

Line 11: what do you mean "big males"? As in adult males? Big is subjective word and should not be used. I also suggest removing reference to specific "named" individuals, i.e., Pyros, Balou, etc. They are not referenced anywhere else and there is no context for "who" these bears

are. I also think it is not appropriate for professional scientific publication to include named animals.

Our answer: *We have now clarified that big males are large-sized adult males and we have also removed reference to specific named individuals (see p.24, l.3-9).*

Lines 18–20. I agree. SCR would be a much better idea in this analysis—particularly as the population size and distribution grow.

Our answer: *We have actually submitted a new manuscript focused on SCR modelling, check out https://www.biorxiv.org/content/10.1101/2022.05.13.491807v1. However, we consider only three years of data for which we could easily dig out the spatial locations. To assess trends in abundance and estimate demographic parameters like survival, only a non-spatial traditional approach was possible, which is totally fine.*

Lines 21–4 on next page: This could be simplified. For cubs, you don't need to list all of these reasons again (you already did that earlier) but can say that cubs have lower survival. You also should clarify that this is cub survival at older than a few months. You mention these mortality risks aren't limited to the early months, but you aren't measuring those first months either since you don't know actual cub survival (i.e., some die before you even detect them and you don't know about them).

Our answer: *In the previous version of the manuscript, we clearly indicated that the first months of life of the cubs were "excluded from our analyses as we considered months from May to September as secondary occasions". But to make this even clearer, we have now emphasized this limitation, stating that "Our estimate of cub survival is likely to be overestimated since our analyses do not take into account cub mortality at a very early age (< 4 months old) as we considered months from May to September as secondary occasions and births generally occur in the dens in January-February (Spady et al. 2007). As a result, some cubs may have died before we could even detect them for the first time." (see p.25, l.25-26 and p.26, l.1-2). As we did not listed before reasons of mortality in cubs (infanticides, predations, mother death or abandonments), we maintained our list of factors of cub mortality (see p.25, l.19-21).*

All of this discussion to detection heterogeneity is 100% relevant to the abundance estimate, so I suggest including it there. What about ease of access for heterogeneity? Particularly for the public collecting data. Locations that are more easily accessible will likely have more people out looking for sign. Some kind of accounting for effort would be reasonable to include in the models.

Our answer: *We did not move the discussion about detection heterogeneity, as suggested by the referee, but have now moved it to a separate paragraph and specified that it might affect both detection heterogeneity and abundance estimate (see p.24, l.21-26 and p.25, l.1-2). Regarding heterogeneity in the data collected by the public, we agree that it could result in some heterogeneity in the detection process. In our experience, using finite mixtures can*

efficiently account for most of this heterogeneity (https://oliviergimenez.github.io/pubs/Louvrieretal2018EcolModel.pdf), almost as efficiently as devising some proxies for sampling effort that would come with their own uncertainty anyway.

Page 15:

Lines 10–18: All of this has been said previously (this is the 4th mention of the fivefold increase and the second time in the discussion) so I suggest some rewording. If these are your major conclusions, perhaps they should not also be in the introduction.

Our answer: *We have now removed the redundant information (p.26, l.7-20) and transferred non redundant one in the Introduction section (see p.9, l.5-12) or at the beginning of the Discussion section (see p.21, l.1-15).*

Line 17: genetic aleas? Typo? Not sure what this word means.

Our answer: *We have now changed "aleas" to "stochasticity" (see p.21, l.14).*

Line 21: I would suggest sticking to either consanguinity or inbreeding. I assume you are using them interchangeably here but this is the first mention of inbreeding.

Our answer: *We have now changed "inbreeding" to "consanguinity" (see p.21, l.13).*

Page 16:

Line 6–7: I would suggest that you do not account enough for detection heterogeneity in this study—and you even say that several times. So I am surprised this is your final paragraph and conclusion. I am also surprised that you recommend using the PCRD method (as opposed to SCR) as one of your main conclusions—particularly since you say SCR would be a good method to use in an above paragraph.

Our answer: *PCRD and SCR should not be opposed actually. We do plan to implement a SCR modelling approach (see https://www.biorxiv.org/content/10.1101/2022.05.13.491807v1 for a first step), however, to quantify trends in abundance and estimate demographic parameters, we will need to go for open populations. Corresponding open SCR models actually implement a robust-design survey design. Our message was more for managers to abandon observed counts and use a capture-recapture approach. Now regarding heterogeneity, we agree that much remains to do to fully understand it, and SCR models will help by accommodating the spatial component. For the time being, we deleted the reference to detection heterogeneity in this sentence (see p.27, l.9-10).*

Supplementary materials:

Line 65: Did you have a freshness scale? And how did you use the freshness information—like did you subsample by freshness? Or not use samples determined to be not fresh?

Our answer: *We estimated the time interval when scats were dropped by evaluating the freshness of the scat (≤2 weeks) when collected in the field, using expert judgement. This classical evaluation (e.g., Sergiel et al. 2020 for a similar approach) is based on the color and appearance of the scat, recent weather conditions (rain, sunshine, snow, temperature, etc.) and type of habitat (directly exposed to sun, under vegetation cover, etc.). We used this information to assign the month (capture occasion) when bear defecated. We have now clarified how we evaluated freshness and added a reference (see p.48, l.74-78). As explained in Sentilles, Vanpé & Quenette (2021), we selected preferentially fresher scats (with less DNA degradation) to send to the molecular laboratory, allowing a better genotyping success and identifying more individuals genetically. We have now added this information (see p.48, l.81-85).*

Line 76–77: Another introduction of heterogeneity. These cubs in captivity are 100% detectable—unlike other bears. Why include these and not the relocation data from radiocollared bears? It is the same thing so justifying its use for cubs but not for radiocollared individuals does not make sense.

Our answer: *As explained, "Orphan cubs that were captured in the field and kept in captivity for a while for care before being released in the wild were considered as still present and detected in the population during the months of captivity (N = 1)." It concerns only 1 orphan cub during two months of captivity, whereas bears equipped with GPS or VHF collars concern respectively 5 bears and 1 bear for a period ranging from several months to a few years. We have now clarified this discrepancy (see p.49, l.95-96).*

Table S1: Was the number of samples analyzed different than the number collected?

Our answer: *We have now added full information in a new Table S2 (see p.52).*

Table S2: This table isn't really necessary. It doesn't add anything and isn't necessary information for this manuscript and you could simply report the PIDsibs in the text.

Our answer: *This table (now called S3, see p.53) is necessary since the molecular lab methods used by ANTAGENE to genotype bear samples have never been published before.*

Table S3 is never referenced in the text so I am not sure why it is included.

Our answer: *We have now referenced this Table (now called S4) in the text (see p.15, l.18 and l.20).*

Review by Romain Pigeault

Dear Recommender,

The manuscript entitled "**Estimating abundance of a recovering transboundary brown bear population with capture-recapture models**" submitted by Cécile Vanpé et al to PCI Ecology aimed to applied Pollock's robust design capture recapture models to estimate the abundance of the Pyrenean brown bear population and its trends over time. The subject of this article falls perfectly within the scope of PCI Ecology and is timely and very interesting. I really enjoyed reading this manuscript which is well written as well as very well structured. I do, however, have some comments and suggestions that I hope will help the authors bring some clarity to the manuscript. Indeed, the dataset as well as the analyses run by the authors are quite complex and some clarification seems to me necessary.

Our answer: *We acknowledge the reviewer for his positive evaluation of our manuscript.*

Main comments:

A) Abundance estimates in this study are in the majority of cases lower than "naïve" (MDS) and corrected counts (MRS, 10/13), which is not what we expect since the Bayesian Pollock's robust design capture-recapture approach should correct for detection imperfections and individual heterogeneity. Authors report that the differences observed between PCRD and MDS/MRS are explained by the fact that the PCRD framework includes temporary emigration, that samples that are difficult to date have been eliminated and that mortality is not managed in the same way in the different analyses, but how to explain the accentuation of the differences at the end of the study period (2017/2018 and 2019)? To what extent does the lack of funding to conduct genetic analysis in 2017 and 2018 impact the results?

Our answer: *From 2017 to 2019, EMR estimates of annual population abundance were much higher than both EMD and PCRD estimates that were quite similar, whereas for the years before 2017, PCRD estimates commonly differed from both EMD and EMR estimates, while EMD and EMR estimates were more similar. This suggests that the accentuation of the differences MRS and PCRD estimates at the end of the study period is likely linked to the increasing number of undetected individuals over a year. For 2017 and 2018, this may be partly explained by the fact that the number of genotyped samples decreased between 2016 and 2017 / 2018, while the population size increased. In 2019 however, the number of genotyped samples was much higher than in 2016, 2017 or 2018. Still with increasing Pyrenean brown bear population size, it becomes more and more difficult to detect all individuals over a year even with a good sampling and genotyping effort. We have now added this information in the Discussion section (see p.22, l.14-26 and p.23, l.1-22).*

B) In view of the significant differences observed between PCRD and MDS/MRS, it would be appropriate to mention in the conclusion that at present MSR remains the most accurate method to estimate bear abundance. But given that the bear population in the Pyrenees is constantly increasing, the development of new monitoring methods is timely because it gives the

possibility to compare the results obtained with the modeling approach with the more robust results obtained with the "naive" counts.

Our answer: *We agree that the MDS/MRS was probably close to the true abundance when the population was small because with our sampling effort, detection was almost perfect. However, as the referee writes, the population is constantly growing, and as was suggested by the 3rd referee Tim Coulson, it is impossible to decide now which method is the most accurate because detection is imperfect and true abundance is unknown. We have therefore tone down the last paragraph of the discussion. Because the MDS/MRS approach comes with at least one year's delay, and is logistically and financially costly, we conclude by recommending the capture-recapture approach which provides real-time abundance (and importantly survival) estimates, and comes with a quantification of uncertainty due to sampling (see p. 27, l.6-20).*

C) Some information concerning the non-invasive methods used to monitor the brown bear population is missing:

- Would it be possible to add on Figure 1 the location of the camera traps and the baited hair traps?

Our answer: *We have now added a figure in the supplementary materials (Fig. S1, see p.59) showing locations of the camera traps and transects used in 2020. Hair traps were set on transects.*

- Is the data collection effort uniformly distributed throughout the bear area range (e.g., walking transects, camera traps)? It seems at first sight that the data collection effort is more important in France, could this lead to an underestimation of the size of the bear population?

Our answer: *The data collection effort discrepancies among countries and administrative units should not affect bear detection and population abundance estimation, since the choice of the monitoring methods was not dictated by the country or administrative unit but rather by the regularity of bear presence in the area (ST was implemented only in areas of known, regular bear presence in France, Spain and Andorra, while OM was implemented everywhere within the potential brown bear presence area). We have now made this clear in the Materials and Methods section (see p.14, l.1-8).*

- The data collection effort has fluctuated a lot over the years. The number of camera traps has increased almost tenfold, and the SBHT has only been conducted over four years. In addition, there were years when not all samples could be analyzed due to lack of funding. Should such large variations in data collection effort not be accounted for in the analyses? For example, it is possible to compartmentalize the analysis into several sub-blocks and then test whether a model containing this compartmentalization obtains a better AICc.

Our answer: *We agree that heterogeneous sampling effort should be accounted for. However, we fear that by compartmentalizing the analyses, we would have too few data to be able to*

estimate abundance reliably. The solution we adopted was to use finite mixtures in the detection probability, which can efficiently account for most of this heterogeneity (https://oliviergimenez.github.io/pubs/Louvrieretal2018EcolModel.pdf), almost as efficiently as devising some proxies for sampling effort that would come with their own uncertainty anyway.

- Why the division of the study area, the installation of camera traps and baited hair traps (grid cell size, SBHT, SCT) are based only on the female range area and not on the male range area?

Our answer: *The different protocols (grid cell size, ST, SCT) are implemented both on female and male areas, but in areas where only males are present, we used a grid cell size of 8x8 km because of large home range. The SBHT protocol was tested for 4 years only in area were reproductive adult females were present (central Pyrenees). It was then abandoned because hair sample collections were insufficient.*

D) Analyses:

The analyses proposed in this study seems to me relevant and the github associated with this article is very useful to understand what was done. However, I have some questions/suggestions:

- It would have been interesting to add an additive time effect to study the yearly variation in survival rate.

Our answer: *If by additive, the referee means an additive effect of age and time, we did explore this avenue in preliminary analyses, and this effect received little if no support by the data.*

- Wouldn't it be relevant to study the effect of the sex of the bears on all the parameters tested (survival rate, emigration, detection)? Moreover, there is no information in the manuscript about the sex ratio in the bear population. Is it stable over time? Is the important increase in the number of individuals from 2018 not explained by a sex-ratio biased towards adult females?

Our answer: *It would be relevant to study the effect of sex on demographic parameters like survival or emigration, and to assess male/female differences in life-history trajectories. However, this is not the objective of our paper, which is mainly focused on the estimation of abundance. Now heterogeneity in detection due to differences between males and females could be an issue when estimating abundance. However, due to several individuals with unidentified sex, it was not possible to explicitly incorporate a male/female group covariate in the analyses. Despite this constraint, we were able to account for potential bias due to detection heterogeneity by considering finite mixtures à la Pledger in the detection probability.*

The sex ratio is not stable over time from 2008 to 2020. It varies from 0.57 in 2011 to 1.33 in 2008 (mean ± SD = 0.81 ± 0.21) among the whole population or 0.42 in 2018 to 1.25 in 2008 (mean ± SD = 0.75 ± 0.27) among adults. The important increase in the number of individuals

from 2018 cannot be explained by a sex ratio biased towards adult females since the sex ratio among adults has been systematically biased towards females since 2012. We have now clarified this in the Discussion section (see p.21, l.8-11) and added a new Table in Supplementary Materials (Table S7, see p.57) providing the evolution of the sex ratio over time.

- I am not familiar with PCRD models, but I'm wondering if there are goodness-of-fit tests for PCRD?

Our answer: *That's a good question. To our knowledge, there is unfortunately no goodness-of-fit tests for PCRD models. The way we proceeded was to include all effects in the analyses that were ecologically relevant like age, time, availability (temporary emigration) and heterogeneity effects.*

- Regarding the estimation of annual population abundance, authors used a Bayesian approach but there is no information in the main text. There is no information on the type of prior used, nor on how authors diagnose the fit of the bayesian model. It would be relevant to describe the analysis a little bit more (e.g., n.iter, n.burnin, thinned).

Our answer: *This information is provided in the appendix https://oliviergimenez.github.io/pyrenean-brown-bear-abundance/ (see p.18, l.10-11). Unless the referee/editor specifically asks for this information to be moved in the main text, which we would be happy to comply with, we would prefer keeping it in the appendix. Briefly speaking, we used uniform distribution between 0 and 1 for all probabilities, and a Dirichlet prior for the proportion of individuals in each class. We ran 3 chains with 20,000 iterations each, including 5,000 iterations for burn-in. In total posterior inference was done on 45,000 iterations.*

Minor comments

- Page 4, line 13: Add some additional information about the PCRD models.

Our answer: To address a similar comment by another referee, we added more information about PCRD models. Here we added the following text: "PCRD CR models rely on several so-called primary sampling occasions, each being composed of secondary occasions. The time interval between secondary sampling occasions must be short enough to meet the population closure assumption, while consecutive primary occasions should be sufficiently separated in time to allow the population to change." *(see p.6, l.24-26 and p.7, l.1-2)*

- Suggestion: I am wondering if the paragraph on page 12 (lines 18-25) and page 13 (lines 1-17) would not be more appropriate in the introduction. This paragraph is very important and sets the framework for the study. It alone justifies the importance of developing a new estimation method.

Our answer: *We have now moved most of the information found in this paragraph in the Introduction section (see p.8, l.17-26 and p.9, l.1-12).*

- Page 6, line 8: What is the average home range of males and females?

Our answer: *When we excluded recently translocated individuals, the average home range size (Kernel 85%) of brown bears in the Pyrenees was 84 km² in adult females (N = 6) and 1,551 km² in adult males (N = 6) (Halotel et al. unpubl. data). We have now added this information in the Materials and Methods section (see p.11, l.26 and p.12, l.1-3).*

- Page 8, line 15: What does "during the same period" mean? 2008 to 2020 or from May to November?

Our answer: *We have now clarified that "during the same period" means from May to November (see p.13, l.24-25).*

- Page 9, line 14: Why several sex markers were used?

Our answer: *Because the genetic sex marker described in the scientific publication De Barba et al. (2017) proved to be not very reproducible, the ANTAGENE laboratory uses a system of three pairs of primers allowing the amplification by PCR of two specific regions of the Y chromosome and one specific region of the X chromosome, according to a method developed and validated in all bear species (Bidon 2013). This system provides an internal positive control for all individuals, with the amplification of a region of the X chromosome present in males (XY) and in females (XX) and to amplify in duplicate a specific region of the Y chromosome present only in males (XY). This triple amplification guarantees an excellent recognition of the Y chromosome and therefore of males, and increases the reliability of characterization of the genetic sex, especially on DNA from degraded samples (hair, scats, etc.). We have now clarified this in the appendix section (see p.46, l.24-35).*

- Page 10, line 19: Could you please define “finite mixtures”

Our answer: *Sorry for that. We now define finite mixtures (see p.17, l.22-24).*

- Page 11, line 17: Do you have any information on individuals with a low detection probability? Is there an effect of sex? of age?

Our answer: *Among the 10 individuals with long detection history (N > 20 occasions) that had the lowest detection probability (< 30% of occasions), we had both males and females and we*

did not observe any age effect. We have now clarified this in the Discussion section (see p.24, l.9-20).

- Page 13, line 2-4: I don't understand this argument. Why is the increased likelihood of detection of specific types of individuals a bias in the estimation of the MRS?

Our answer: *We have now removed this argument (see p.21, l.25-26 and p.22, l.1).*

- Page 13, lines 10-11: Don't the funding restrictions also directly impact the PCRD models?

Our answer: *Due to financial constraints, only a subset of all collected hair and scat samples were genetically analysed to identify individuals each year (mean ± SD = 35.16 ± 12.29, min = 17.5 in 2015 and max = 59.5 in 2008; see Table S2). However, samples that were sent to the lab each year were carefully selected so that we optimised the detection of individuals (e.g., we favoured samples from cubs of the year or subadults, as well samples that were collected in colonisation front) and the genotyping success (e.g., freshest scats, avoidance of hair coming from different individuals). In addition, the difference between MRS and PCRD estimates was not positively correlated to the proportion of collected samples that were genetically analysed (F1,11 = 0.436, P = 0.52). We have now clarified this in the Materials and Methods (see p.15, l.1-6) and Discussion (see p.23, l.6-22) sections.*

- Page 15, lines 14-18: Is it possible to used PCRD models to make some predictions about abundance evolution in the coming years?

Our answer: *We don't feel comfortable in going that direction for two main reasons. First, PCRD models do not implement a process model for temporal trends in abundance, which makes it impossible to provide model-based forecasts with appropriate quantification of uncertainty. This step would require developing a framework for forecasting, possibly with state-space models, that deserves a full treatment in another paper. Second, we feel like producing these predictions is beyond the scope of our paper and its ecological objectives. In this paper, we argue for implementing capture-recapture abundance estimation to strengthen brown bear monitoring and aim at reporting past and current abundance estimates for informing species conservation.*

- Page 29, is it possible to do paternity analysis with your genetic data?

Our answer: *We know indeed the pedigree of our population since 1995 thanks to genotyping and parentage analyses.*

- Page 30, what is the proportion of samples not used because of a dating problem?

Our answer: *In France, we collected in total 4,022 hair or scat samples from 2008 to 2020, among which about 5.5% were excluded from our analyses due to inaccurate dating. We have now added this information in the Supplementary Materials section in "Dating of bear signs" (p.48, l.84-85).*

Review by Tim Coulson

The preprint "Estimating abundance of a recovering transboundary brown bear population with capture recapture models" by Vanpé et al. describes a statistical modelling exercise of European brown bear sightings, scats, and hair samples from the Pyrenees. More specifically a mix of hair samples and scats used for genetic identification, and camera trap data of recognised individuals, from a mixture of structured and opportunistic encounters with signs of bear activity, are used to construct mark-recapture histories from 2008 to 2020. These are then analysed in R Mark to provide detection and survival estimates. The final model is then used to provide Bayesian estimates of population estimates. The authors conclude that mark-recapture provide useful estimates of a population of a large and elusive carnivore.

The statistical modelling is appropriately conducted, the results make good biological sense, and I only have a few significant suggestions, plus several language and grammar edits included on the marked-up pdf.

Our answer: *We acknowledge the reviewer for his positive evaluation of our manuscript and his useful comments and meticulous reading of our manuscript for helping us with language and grammar edits. We have now answerer point by points to significant suggestions and used the marked-up pdf to correct our manuscript for language and grammar edits.*

Significant suggestions

1. In the abstract, and towards the end of the discussion, the authors conclude that the Pollock method they used provides accurate estimates of bear abundance. I suspect that this is true, but because the truth is not known, it is not possible to state this so strongly. Please tone down these statements to say that even in cases where sampling effort is large compared to population size, mark-recapture methods can provide estimates of survival and population size, having corrected for imperfect detection, that diverge from the minimum number known to be alive.

Our answer: *We agree with the reviewer and have now toned down this statement adding the sentence suggested by the reviewer in the Abstract (see p.4, l.2-6) and end of Discussion section (see p.27, l.6-10).*

2. I would like a bit more information on the analysis of individual photos to identify bears. Was this done visually by bear experts, or was pattern recognition software used? This section is a bit light on detail.

Our answer: *Individual identification of bears was mainly based on genetic analyses of hair and scats non-invasively collected in the field, as well as visual evidence (colouration, scars, GPS collars, or VHF ear tag transmitters) obtained by remote cameras (Sentilles et al. 2021b). This visual identification was performed by bear experts from OFB and was validated only if a consensus was released among all those experts without any doubt. We have now clarified this (see p.14, l.23-25).*

3. A frequentist framework is used for the mark-recapture analysis to estimate survival rates, and a Bayesian approach for population size estimates due to boundary issues encountered when estimating population size in a frequentist manner. I was left wondering: why not conduct all the analysis in a Bayesian framework? I am not requiring this to be done but would like some justification added as to the choice of an initial frequentist approach.

Our answer: *This is a question also asked by another referee. We used the frequentist approach to carry out model selection, which is much more stabilised in the frequentist framework and avoid prohibitive computation times, then the Bayesian approach for parameter estimation to avoid the boundary estimate issues. We have now clarified this (see p.18, l.7-10).*

4. At the beginning of the discussion there is some text I have flagged that should be moved into the methods and results. It is about calculation of the MRS estimates. It is not discussion but provides methodological approaches and new results. It could be removed completely given the following point.

Our answer: *We agree with the reviewer and have now decided to move most of this information in the Introduction (see p.8, l.17-26 and p.9, l.1-12) and Results (see p.19, l.24-25 and p.20, l.1-9) sections.*

5. I found the MRS estimates to be rather unnecessary. They depart from the MDS and PCRD estimates, suggesting the 'correction' used to calculate them from the MDS estimates is adding bias rather than insight. The authors set up a straw man by calculating a new index that differs from the MDS and PCRD estimates that in fact align quite well. Why invent a new index, the MRS index that is only introduced in the discussion as an apparent after thought, and then compare it to established method, discover it is wanting, but then it use to criticise the MRS and the MDS methods. It is all rather unnecessary. The MDS estimates are actually pretty good. Perhaps this lessens the conclusions of the paper with regards to the PCRD and MDS approach, but I would argue you should use the best statistical tools available - i.e. PCRD -- unless they significantly complicate analyses. I would recommend removing the MRS estimates, or if they are to be kept, much more strongly justifying their calculation and inclusion.

Our answer: *We apologize for the confusion about MRS and we realised that we need to be clearer about it as early as the Introduction section. The annual size of the Pyrenean brown bear population has been estimated since 1995 using the Minimum Detected Size (MDS) index, defined as the minimum number of different individuals detected (either by genetics or unique individual marks) inside the study area over the year, as well as the Minimum Retained Size (MRS) index, corresponding to a reassessment of the MDS in the light of the information newly collected in the following years (Sentilles, Vanpé & Quenette 2021; Sentilles et al. 2022). MRS was considered as a precise and accurate estimate of the true annual brown bear population size in the Pyrenees. However, the MRS estimate for year n is only available in years n+1 and even needs sometimes a reassessment on year n+2 or n+3. And as the population is growing in size and expanding spatially, the probability of not detecting some individuals each year using MDS and even MRS is growing. In addition, the MRS and MDS estimates are logistically and*

financially demanding, since they require intensive monitoring effort. We therefore decided to develop a new metric based on capture-recapture models to estimate the size if the Pyrenean brown bear population, while minimizing bias due to inter-individual heterogeneity in detection probabilities and allowing the quantification of sampling uncertainty surrounding these estimates. We have now made this clearer in the Introduction section (see p.8, l.17-26 and p.9, l.1-12).

I hope these comments are useful.

Our answer: *We thank a lot the reviewer for his useful comments.*