# Mark loss can strongly bias demographic rates in multi-state models: a case study with simulated and empirical datasets

F. Touzalin, Eric J. Petit, Emmanuelle Cam, Claire Stagier, Emma C. Teeling, Sébastien J. Puechmaille

July 2023

Dear Dr. Sylvain Billiard, Camille Jolivel and anonymous reviewers, thank you very much for your review and your comments on our manuscript, please find below the answers of these comments.

## General evaluation and major comments

1. Better define and give a precise formulation of your model, more in line with how such models are presented in the literature. I strongly suggest to follow reviewer's 1 advice to dedicate a whole section on the model, its formulation, hypothesis, link with previous literature, etc.

   *In order to better define the modeling process, all the distributions and formulations used to both simulating the data and analyzing them were inserted in the main text. Only priors used were let in the supp. information. However, we kept separated the description of the data generation and the description of models for a better understanding.*

2. All reviewers, myself and Camille Jolivel felt that a better use of supp. information should be made. Actually, they are too extensive, some are certainly worth to include in the main text, a suggestion by reviewer 3 to provide a tentative of synthesis of some supp. mat. by focusing on a specific case and show estimates in real parameter space rather than metrics summarizing the biases (for a better understanding of the effect and direction of the biases due to mark loss). Below, I also suggest another option to include some of your results in the main text.

   *We included now new figures in the main text on key aspects of the results, in particular the figure describing survival range across vertebrates, comparison of medians between the models accounting and not accounting for mark loss (on probability scale) for simulated data, and another for empirical data.*

3. I felt a bit frustrated that a very nice work of data compilation has been done, but which is maybe underutilised. It might be worth describing and discussing a bit some general patterns observed in Fig. S1 supp. information 1.

   *This has been done now.*

4. As several reviewers pointed out, I found interesting the conclusion and proposition that double mark should be used. I wonder if it could be possible to have a quantitative

assessment of that point with the methodological framework developed by the authors: compare biases with a single vs. two marks and evaluate the quantitative gain of estimation.

*A new estimate of the loss of tags was made using only partial data from the discovery of tags that had fallen into maternity roots, without accounting for a permanent mark (i.e. genotype). These tags found on the ground actually correspond to 61.5% of all the tags lost by the individuals studied, which can be deduce from the genotype. The results showed a lower estimate of tag loss (see Fig. 1) than the estimate from the model that did not take tag loss into account at all. Median tag lost rate for juveniles was estimated 0.23 (90%hdi [0.18,0.29]) if surgical glue was not used and 0.14 (90%hdi [0.11,0.18]) if the glue was used, versus 0.28 (90%hdi [0.23,0.33]) to 0.19 (90%hdi [0.16,0.22]), respectively. For the adults, the first year after tagging, tag shedding was estimated 0.05 (90%hdi [0.02,0.08]) if surgical glue was not used and 0.03 (90%hdi [0.01,0.07]) if the glue was used, versus 0.11 (90%hdi [0.08,0.15]) to 0.10 (90%hdi [0.06,0.14]), respectively. Then, the following year, after the first year of marking, tag shedding is 0.01 (90%hdi [0.00,0.01]) if the glue was not used and 0.02 (90%hdi [0.01,0.03]) if the glue was used, versus 0.02 (90%hdi [0.01,0.02]) to 0.03 (90%hdi [0.01,0.04]), respectively. Although we acknowledged that this quantitative estimation of the gain obtained via the double marking is interesting for our study system, the results does not add significantly to our general understanding of the effects of mark loss. The less information we have about tag loss, the closer the result will be to that of the model that did not take it into account, and on the contrary, the closer the number of tags found is to the truth, the closer the tag loss rate will be to 0. We are not convinced that this quantitative estimation per se is of interest to a wide audience. This is why we have decided not to include this new estimate in the article, which is already long, and because we feel that it does not provide any more information on the subject.*

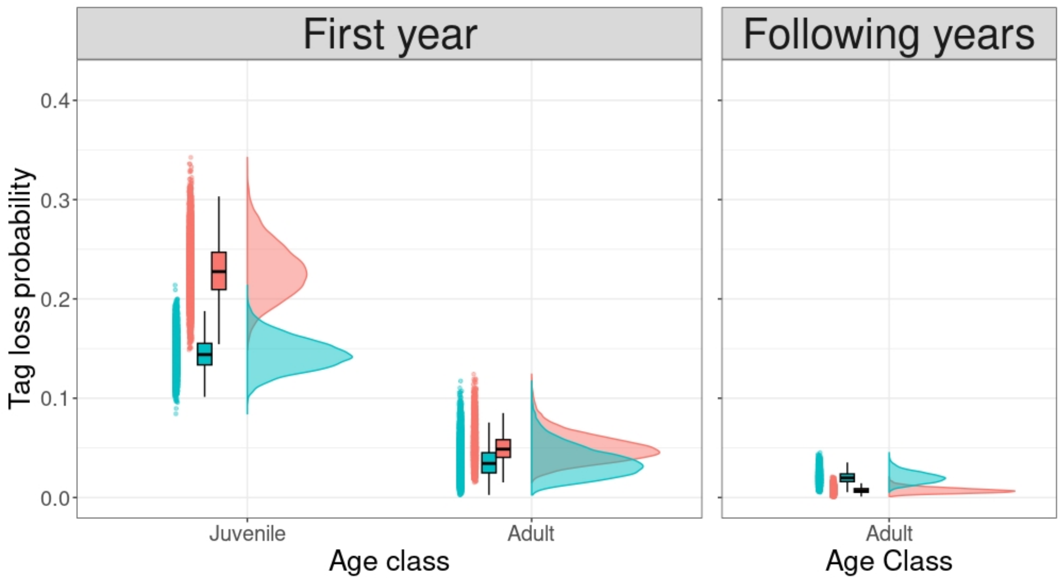


Figure 1: Posterior distribution of the tag loss probabilities according to age classes and time after marking in the Myotis myotis dataset. Left panel correspond to tag shedding rate during the year following the tag injection and the right panel for the following years (constant in time). In blue, distribution if surgical adhesive was used after tag injection and in red, without surgical adhesive.

## Additional comments

- Abstract: Most of the information given are about the simulation studies, especially sections 2., 3. and 4. The data analysis on bats is drown out in the abstract. We suggest to compile all information on bats in a 5th section.

  *A 5th section has been added accordingly.*

- Keywords: We suggest to remove from the keywords list the words already in the title of the paper.

  *The keywords list has been changed.*

- Introduction: Italicise all parameters and variables, do not italicise numbers in quotes; Homogenise the font for quotes.

  *The format has been homogenized.*

- Introduction: Suggestion of reorganisation of the first paragraph : Capture-mark-recapture (CMR) methods have [...] is the loss of marks (see Supported Information 1, Table S2). Mark loss has two consequences [...] which is called "recycling". Local survival is the product of true survival etc.

  *This suggestion has been accounted for.*

- L78: Explain the two parts of the equation

  *The equation has been explained in more details (L97-105).*

- L100-101: Homogenise the use of brackets for citations "(Nichols & Kendall, 1995 ; see also Cam 2009...)"

  *The format has been homogenised.*

- L91 vs. L.105: Incoherence between "has not been explored" vs. "is scarce"

  *We agree about this incoherence and we have reformulated this sentence and justified it (L125-126).*

- Material and method: You might improve the accuracy and the information content of the title of the sections. For instance: change "real data analysis" with "Application on a bats dataset"; "Computation details" is not sufficiently different to "data simulation". A title containing what is within this section could be "Estimation procedures and assessments".

  *We have modified the tittles of these sections.*

- L130: Consider explaining more precisely what you call "realistic". You might rather mean "representative"?

  *We agree that "representative" is a more appropriate word than "realistic" in this case and we modified it accordingly (L146).*

- L133: It is a bit odd to cite Fig. 1.b before Fig. 1.a

  *The figure 1 and the order of the figures cited in the main text have been modified according to different comments of the reviewers.*

- L140: remove comma after "Reptilia"

  *This has been modified (L155).*

- L145: It is not quite clear how you translated your data survey in these four scenario. Consider explaining more accurately. In addition, it might be worth making explicit the features you chose to neglect. Also discuss whether or not, by splitting the range of observation into four scenarios, your analysis is still adequate for any kind of species considered.

  *On the basis of our literature review, we have justified in greater detail our decision to study only 4 scenarios that are close to the extremes of the range of possible values* *for the survival and detection parameters. The full spectrum being virtually infinite (L153-160).*

- L169: add notation for the "mark loss probabilities" noted "ptl" in Fig 1.a

  *This notation has been added (Fig. 2 and L198).*

- L171: Not sure why 50 datasets here, and only ten in FigS1 supp Inf 2?

  *50 data sets were simulated for parameter estimation as a compromise between computation time and minimum number of replicates to obtain a distribution of estimated parameters. Fig. 1 in Supp. Inf. 2 describe all the parameters used for simulation and how a total of 1200 simulations were performed for achieving the simulation section of this paper.*

- L177: It would be more appropriate to call Ri,t a probability rather than a rate because rate refers to a speed, you could consider calling Ri,t the retention event probability.

  *For the sake of simplification, the retention probability is now designed as "pr" (L254-255, eq. 7).*

- L177: Not clear whether the "mark loss probabilities" = 1 - "retention probability" ?

  *The mark loss probability (ptl) is the complement of the retention probability (pr), see Fig. 2 and L198.*

- L177: Not easy to understand that j in... means over the different situation. The model should be better formalised.

  *The retention probability (pr) is now better defined (see Fig. 2 and L198) and subdivided in 3 categories depending of age and time since marking (detailed L254-255 and in equation 7).*

- L186: "recycled individuals" is defined here, whereas the term appears several times before. Consider defining explicitly what "recycled" means as soon as necessary.

  *This definition appears now at the beginning of the introduction (L74), corresponding to the first mention of the term.*

- L200: remove comma after "both"

  *Done (L272).*

- L219: delete extra dot

  *Done (L292).*

- L224: delete extra dot

  *This sentence as been modified (L295-297).*

- L234: Add missing dot

  *Done (L317).*

- L237: Consider adding some general and basic literature about JAGS, Gibbs, etc. for people that would not be familiar with these methods. e.g. Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

  *The literature corresponding to all mentioned software has been added (L117-119).*

- L238: Give the meaning of the acronym MCMC

  *This has been added (L327).*

- L242: "Chain convergence was assessed with the Gelman-Rubin statistic denoted R-hat"

  *This has been modified (L330).*

- L243: "converge" → "convergence"
  *This has been modified (L332).*
- L247: "expected", rather say "assumed"?
  *The term "expected" has been replaced by "considered" (L344).*
- L257: Only figure S43 is mentioned here, a series of figure is expected I guess.
  *Figure S43 became now figure S76 and it explain in details how the ROPE was calculated (L349-362).*
- Results: Homogenise notations of states. L277 "D" vs. L282 A and C.
  *These notations have been homogenised throughout the text.*
- L296: I feel a little frustrated that results in fig. S4-S11 are given in supp. mat. only. I think it would be worth moving these figures in the main text by providing a synthetic figure as a matrix as Fig 2 and 3.
  *A figure 7 has been added in the main text that syntheses fig. S4-S11.*

## Reviewer's 1 report.

### Main comments

*We thank reviewer 1 for his comments and provided answers below.*

- You have done an impressive work in gathering survival estimates and the magnitude of tag loss for hundreds of species. Congratulations on that. Tables S1 and S2 are awesome!
  *Many thanks!*
- I feel like lines 71-83 are too technical for an introduction. I would move this paragraph to Material and methods, and explain the problem with words here.
  *We have left the mathematical formulations, which are quite common for people working with CMR data, but we have described them in more detail in the text (L87-105).*
- Line 265: I find the use of ROPE disturbing because from the beginning of the paper we expect an assessment of bias. Why not simply consider posterior means/medians and compute bias in the frequentist way? I'm not asking to drop the ROPE metric, but to add a more traditional measure of bias (hopefully the simulations were saved and it will not take long).
  *We agreed that the use of ROPE is not so common and more specific to Bayesian approaches. As suggested we added 2 more common metrics used in such case: bias as difference between medians of the 2 different models (we included fig.4 in the main text with bias shown on the parameter scale) and precision as $bias^2 + variance$.*
- You're missing several important references. First, multistate models should not be called multistate CJS models. R. Cormack, G. Jolly and G. Seber developed models with a single alive state. Multistate models were developed by Neil Arnason and Carl Schwarz in a series of papers, and multistate models are sometimes referred to as the Arnason-Schwarz model. As of terminology, I guess multistate models or Arnason-Schwarz models do the job. The references that need to be cited are:

- Arnason, A. N. (1972) Parameter estimates from mark-recapture experiments on two populations
- subject to migration and death, Researches on Population Ecology, 13, pp. 97-113.
- Arnason, A. N. (1973) The estimation of population size, migration rates and survival in a stratified population, Researches on Population Ecology, 15, pp. 1-8.
- Schwarz, C. J., Schweigert, J. F. & Arnason, A. N. (1993) Estimating migration rates using tag-recovery data, Biometrics, 49, pp. 177-193.

*These references have been added.*

Second, about tag-loss, there are two papers that need to be included in your paper:

- Arnason, A. N., and K. H. Mills. 1987. Detection of handling mortality and its effects on Jolly–Seber estimates for mark-recapture experiments. Can. J. Fish. Aquat. Sci. 44: 64-73.
- Juillet, C., Choquet, R., Gauthier, G. et al. A Capture–Recapture Model with Double-Marking, Live and Dead Encounters, and Heterogeneity of Reporting Due to Auxiliary Mark Loss. JABES 16, 88–104 (2011).

*These references have been added.*

- Line 235: I've had a hard time identifying the model with tag-loss you're using. If I understand correctly, it is specified with Eqn 1. This equation should come with the relevant literature, unless you've developed this model yourself. I recommend having a specific section describing both models with and without tag loss, which are common to both the simulations and case study, so that the reader can easily goes back to it. In this new section, you should make it clear how the model with tag-loss you're using differs from existing models with double-marking. This new section could also include the literature review you do earlier in the paper, but if not, you should refer to it.

  *Now we have included all formulations for both data simulation and modeling in the main text, section 2.1 and 2.2 respectively. In particular, we described in the statistical model section (2.2) how mark retention was estimated and mentioned the difference between the 2 models used. References to previous methods described in the literature have been included to clarify the specificity of our model.*

## Minor comments

- Lines 31-33: I find the statement "little attention has been paid to the effect of common violations of the CMR model assumptions" a bit unfair. I understand that the abstract has to be punchy and attractive, but there are a lot of papers out there on the issues of heterogeneity, incorrect state assignment to individuals, and more broadly goodness-of-fit testing for multistate models. I would tone down this statement a little, and simply write that the tag-loss issue is tackled.

  *We fully agree that efforts have already been made to explore violations of the model assumption in CMR models and that many developments exist today to take them into account, but little has been done in the particular context of multi-state models. We reformulated this sentence (L31-33).*

- I noticed that you use both multi-state and multistate, I would homogeneise throughout the paper.

  *This has been homogenised.*

- Line 99: Bugs is not a software program, you mean WinBUGS, JAGS, OpenBUGS, STAN, NIMBLE, etc. I would also keep MARK and E-SURGE for user-friendly software here. Please, add the relevant references too, folks have put a lot of efforts in developping these pieces of software, and their dedication has served the community well, the least we can do is to acknowledge their work.

  *This has been modified accordingly (L117-119).*

- Sometimes you have initials for first names in the cited references. I guess we do not need them in the text, only in the list of references.

  *This has been homogenised.*

- Lines 104-108: You do a great job at reviewing the literature on tag loss in CJS models, and the applications of multistate models, but here we're missing what you expect to be new and/or different from the existing studies in terms of bias in demographic parameters. Actually, you have these predictions diluted in the paragraph lines 116-125, which is more about a specific case study. The general predictions should come with the simulation study.

  *We have now moved the general predictions after the description of the simulation study and the model assumptions of the AS model (L133-138).*

- Why going for a Bayesian approach? You don't have random effects in the simulations, nor do you have prior information to incorporate. Also, bias is kind of a frequentist notion, it would have been more natural to go for maximum likelihood estimation. Plus, this would have allowed for more than 50 simulated datasets (for each scenario or combination of parameters), which I guess you had to keep under a limit due to the computational burden of MCMC methods. Estimation of bias requires more simulations in general. I'm not asking to go for a frequentist approach, just to provide the reader with some justifications.

  *We decided to maintain a consistent approach for both the simulation part and the part using empirical data. As the analyses of the empirical data included random effects, we chose a Bayesian approach. Even if a frequentist approach would have made it possible to carry out more simulations than the computation time of the Bayesian approach would allow, the latter remains reasonable from our point of view. In addition, this enabled us to validate a parameterisation within the Bayesian framework of a model capable of estimating these parameters without bias.*

- Line 132: What about the dead state? It is absorbing too ;-)

  *We added the mention of the dead state which is of course an absorbing state as well as state 'D' in or scenario (L173).*

- Lines 125-126: Not sure there is a formal definition of bias in posterior distribution, isn't it?

  *Right, we reformulated our predictions (L134-138).*

- Line 193: What does 'To illustrate the simulations' mean? Some readers would be perfectly happy with simulations only. The case study brings less generality, because you don't know the truth (at least only 66% of it). I like the case study, it just needs to be better introduced: What does it add to the simulations?

  *This has been reformulated (L261-265). The empirical study provides a concrete illustration of the real estimate of the rate of mark loss, compared with the observed* *rate, when using a second permanent mark, which enabled us to assess the relevance of maintaining genotyping in this study. It turned out that a permanent mark such as the genotype is essential in this study to avoid estimation bias.*

- For the bat case study, you need to perform goodness-of-fit tests and cite the relevant literature. Bias in parameter estimates could be caused by transience or trap-dependence issues.

  *We added goodness-of-fit tests but despite them being useful, they are not sufficient to understand the structure of the data as they test restrictive hypotheses. In addition, it is often the case that some goodness-of-fit tests cannot be calculated due to the structure of the data (lack of data in contingency tables), and this is the case with our empirical data (L295-300).*

- Line 224: Bias on which parameter(s)?

  *This part was modified (L307-304).*

- Lines 243-248: It is difficult to follow you blindly here without knowing for which parameters convergence was not reached. Actually, with only 50 simulations, we might expect some 'bias' in the results due only to lack of convergence, not tag loss. I encourage you to be more specific.

  *This aspect has been detailed in the main text (L331-344) and R-hat values can be found in the supp. inf. 2, section 3.2.*

- Line 237: That's an old version of R ;-) I recommend updating JAGS, JagsUI and R.

  *We agree but this was the R version available on the cluster we had access to.*

- Line 274: Do we care about detection being biased?

  *In this paper we mainly focused on demographic parameters. We agree that detection is not really under concern here, we modified this sentences (L373).*

- Lines 298-299: I am not sure I follow, it's one thing to quantify bias on survival, it is another to demonstrate bias on the relationship between a covariate and survival. You might have a bias in survival, but if this bias is the same for all individuals or time intervals, then I guess there is no problem to assess the effect of individual or temporal covariate. Am I missing something?

  *True, this sentence has been removed.*

- Lines 311-328: The results are discussed with no reference whatsoever to the existing literature. What was found in other papers? Mobilizing the existing literature and comparing your results to previous findings will help you to emphasize what's new and original in your work.

  *This gap has now been filled (L423-436).*

- Lines 331-332: I've probably missed it, how can you be sure that you quantify bias in the case study with only two third of the individuals for which you know they've lost their tag?

  *We know that we only found around 60% of the tag lost on the ground of the maternity roosts because we use a permanent mark: the genotype of the individuals. As all tagged individuals are also genotyped, when we catch an unmarked individual we know if he has been marked before or if it is a new recruited individual. Our uncertainty about the status of tags concerns only individuals neither resighted again nor recaptured.*

- Line 345: Dispersal, not dispersion (I think).

  *This has been modified (L470).*

- Lines 402-404: I applaud the authors for their efforts of making data and code available. I would deposit the code and data on GitHub/GitLab for versioning and also to make it easy to re-use your code (copying and pasting from a PDF to R can produce funky behaviors sometimes).

  *This will be done after acceptation of the manuscript.*

- Legend of Table 1: You write 'U=Univariate distribution)', you mean 'Uniform distribution' I guess. I also find it unclear the meaning of the minus something you have in the short-live species column. A Uniform(a,b) - 0.5 is Uniform(a-0.5,b-0.5), and Normal(mu,sigma)-0.3 is Normal(mu-0.3,sigma). Did you simulate directly from these 'translated' distributions, or did you simulate from Uniform(a,b) and Normal(mu,sigma) then substract something?

  *This has been modified and clarified (see Table 2 and Supp. Inf. 2 part 1.4). In order to simulate We simulated from the distributions from which we subtracted a fix number as specified, e.g. as Uniform(a,b) - 0.5 but not as Uniform(a-0.5,b-0.5)). This corresponded to a translation of 0.5 unites on the parameter scale.*

- Figure 1: Nice work!

  *Thanks! Following suggestions from other reviewers, this figure is now divided into Fig.2 and Fig.3.*

- Figures 2 and 3: These figures cannot be read independently of the main text. I would remind the reader in the legend what the scenarios are, and what the numbers and letters on the axes refer to. Also I'd define ROPE.

  *This has been added in the figure description (Fig. 5 & 6).*

- Line 266: I find 'for most parameters in most situations' to be a vague statement. I encourage you to provide some quantitative information so that the reader can make a judgment by herself/himself.

  *This as been describe in more details in the main text (Line 332-340) and values were displayed in Supp. Inf. 2, part 3.2.*

- Figure S2: I am not sure the relationship will remain significant once you account for uncertainty in both adult and juvenile survival probabilities. You're doing statistics on statistics here.

  *We agree, this figure has been removed.*

## Reviewer's 2 report.

The manuscript presents a simulation and analysis of real data, assessing the bias due to tag loss in parameter estimates from multi-state CMR models. This is a timely contribution, as the effects of tag loss in CMR models have received some attention in the literature, but not in the case of multi-state models. I do not have the technical expertise to evaluate the correctness of the code presented (I trust other reviewers will do that), but I found the text clear and generally easy to understand what was done. The introduction is possibly the part of the manuscript that could do with a bit more work. Some sections, such as between lines 62-67 could use more references. The results and figures in the manuscript are clear, although I found that the supporting information is quite extensive and difficult to follow. I agree with the interpretation of the results, that tag loss is relevant and should be taken into account in CMR studies. To me it was also surprising that it biased the transition probabilities more strongly than survival estimates. The need for double marking is also an important message that the authors properly emphasize.

*We thank reviewer 2 for his comments and, as suggested, we have tried to improve the introduction, and we have added a 5th section following other comments. We have also added more references in the text in general and in particular in the section mentioned (L82-83).*

Reviewer's 3 report.

The authors created an extensive simulation experiment to study the effect of mark-loss on bias of parameter estimates in multistate capture-recapture models. Overall the author did a good job illustrating the fact that mark-loss can induce substantial parameter bias in multistate models. Moreover, this bias can present itself in an unintuitive fashion due to the complexity of the model and the interactions of the parameters in the likelihood. Although it is strictly a personal preference, the article might appeal to the ecological community that uses these models if the authors illustrated some of these bias effects on the real scale of the parameters in the main portion of the paper (not in an appendix). I.e., for perhaps with just the bat analysis, the authors could create a figure with, say survival or transition probabilities under each model. That way users can see the effects in real terms, rather than EMD or ROPE metrics that don't have a meaningful interpretation in real parameter space.
In addition to my overall comments, I have attached an annotated pdf with more specific comments and questions.

*We thank reviewer 3 for his comments and annotations. We agreed that assessment of bias on parameter scale is more appealing than ROPE. In particular, we have calculated bias in the median of the posterior distribution of the parameters estimated from simulated data and included a new figure in the main text illustrating them in a specific scenario (Fig. 4). We also added another new figure in the main text illustrating the difference between the medians of the 2 different models used for empirical data to compare the effect of accounting or not for tag loss on the parameter scale (Fig. 7).*

## Annotations in the manuscript

- Introduction: Should these marks be independent?
  *we mentioned the fact that double marks should be independent or specific formulations should be added in the model to account for non-independence in mark loss with the corresponding bibliography (L48-50 and L482-486)*
- L114: You can get around this if you have the appropriate covariates, e.g., sex or age, etc. Or, as in your bat analysis you used a random effect to account for heterogeneity. So, this is overstated here. You just need to be able to account for differences with covariates or random effects.
  *We fully agree that this was overstated and we removed this sentence. The goal was to remember the reader the general assumptions of the AS model.*
- L122: In your small example (line 81), you showed that transitions would be underestimated. Why do you expect overestimation here?
  *we rewrite this sentence and indeed indicated that we also expect, as illustrated, transitions probability to be underestimated. However, we could expect that specific*

*transitions such as emigration, where such data is available, will be overestimated, as the individuals who have lost their tag may be considered either to have died or emigrated. For the sake of clarity, we have not developed these considerations in the introduction.*

- L186-187: Did you do this for all individuals that lost a mark? It Seems that in a real analysis the size of the population would dictate how likely this is to happen. E.g., in a large population if an individual looses a mark, it may be very unlikely that that individual is captured again. Therefore recycling would be rare. In your model structure, you account for recapture with a Jolly-Seber like capture parameter, but it's not obvious here that is how the data were simulated.
  *we have now clarified the way data with recycling individuals were produced with an example (L221-227). Indeed, data sets with recycling individuals contain more individuals than the real population had, which could affect further recycling in relationship with recapture rate. We agreed that the proportion of recycling is dependent of population size and expect high recycling when population is small, tag loss high and recapture high. In case of large population, dilution effect will reduce mathematically recycling as probability to recapture an individual that lost its mark is low. We didn't explicitly play with population size but figure S3 in Supporting Information 2 illustrates the diversity of range of total recycled individual we explored. It also highlights the diversity of situations that can be encountered in the field, but even with a large cohort, we argue that this does not preclude simulating data to test the potential effect of mark loss on parameter estimates in any study where mark loss has occurred.*

- L294: Should this be removed, i.e., there was substantial bias? Based on the next sentence, it seems as though there was substantial bias.
  *This has been removed (L406).*

- Figure 2 & 3 : A short description of the scenarios should be added here so the reader can easily digest the figure. Same for Fig 3.
  *This has been described in more details (now Fig.5 & 6).*