RESPONSE TO REVIEWERS

Evolutionary determinants of reproductive seasonality: a theoretical approach

---

Dear editor and reviewers,

We would like to thank the three reviewers and the editor for their useful comments and suggestions. Our point-by-point answers are presented below, in red and bold characters.

Three reviewers have now provided thoughtful and detailed comments on the preprint. All are broadly supportive of the simulations, and the way they are written up and reported. I consequently expect that I will ultimately recommend the preprint as it makes a valuable contribution to our understanding of reproductive decision-making in variable and seasonal environments. Despite these positives, there is work to do. All three reviewers make some major comments along with more minor suggestions. All the comments can be dealt with, some via rewriting or referring to additional literature. Other comments will require revisiting the simulations or the way results are reported.

In particular, all three reviewers note that using p-values to assess statistical significance from simulations is problematic, simply because you can increase the number of simulation runs until statistical significance of the output of interest is achieved. Something that explains a very tiny amount of variation then has a statistically significant p-value but is biologically unimportant. You really need to describe the difference in means and variance without recourse to p-values.

The use of p-values to identify the optimal strategies was indeed ill-advised as it is sensitive to the number of simulations. We instead followed the advice from reviewer 2 (Francois-Xavier Dechaume-Moncharmont) and used a 5% decrease on effect size criterion to select the optimal strategies. We also removed p-values from our result section and provide mean and standard deviation description instead. This new approach did not change our results.

You also need to better justify some assumptions, refer to a wider literature, and add clarity to some of the methods.

We followed the reviewers' suggestions regarding these changes (answers and modifications detailed below).

Reviewer 3 also makes a good point -- you need to justify why you are averaging over stochastic simulations. Doesn't this simply reveal the deterministic skeleton of your model?

It is true that stochasticity has a high computing cost and that we end up averaging over simulations. In order to remove stochasticity from our model, we would need to include the "average effect" of stochastic conception and mortality on reproductive seasonality. Yet, depending on the phenology strategy that we are testing, we believe that this average effect can be very different (see example in response to reviewer 3). Because such effects can be difficult to predict, we decided to make the minimum number of assumptions in order to reduce the risk of introducing a bias. In summary, we tried to have a model as realistic as possible and stochasticity in mortality and reproduction is part of the reality experienced by yellow baboons.

All these alterations are achievable, and i very much look forward to seeing a revised version in due course.

by Tim Coulson, 01 Oct 2022 07:38
Manuscript: *https://doi.org/10.1101/2022.08.22.504761*

Review by Nigel Yoccoz, 18 Sep 2022 15:43

I concur with the authors that our understanding of reproductive seasonality is relatively fragmentary, and that theoretical approaches are needed to complement empirical studies. The authors propose here an individual-based model derived from field studies on the yellow baboon, with energetics and resource availability being the main drivers of change in the model. The model framework is a first step that can lead to further developments.

My comments address mostly how the paper relates to the current literature and some of the underlying model assumptions. I have also to admit a "small mammal" bias.

L. 69: "and unpredictability (year-to-year variation in food availability) " : this is not the same thing - you may have large year-to-year variation which is at least in part predictable. For example, with el Nino-la Nina, or cyclic populations (small rodents, forest pest species). It is also different in terms of temporal scale: in this paper, the measure of variability combines both short-term (between weeks, as you use NDVI values that are given with a 2 weeks interval) and longer-term (between year) variability. Environmental change may affect both components in different ways. I understand that one cannot do everything in one paper, but it helps to make the assumptions clear – you partly discuss this l. 445-446 by referring to papers by Colwell and English, but without giving any clear idea of what it can mean in practice.

We clarified the terminology used by replacing "year-to-year variation" with "non-seasonal variation" which reflects better our decomposition of NDVI time series. We also insisted on the fact that non-seasonal variation of food availability is used as an *approximation* of environmental unpredictability and is later modelled as an added noise. In the context of our study on reproductive seasonality, this measure was chosen to capture environmental variation that would not be seasonal, which is acknowledged in the methods, when we first present this measure (l.337). As you rightfully point out, year-to-year variation / non-seasonal variation in NDVI could indeed be in part predictable and we report this limit in the discussion l.523. We also clarified in the methods (l.337) that our measure of variability also encompasses short-term measurement errors and not only between year variability.

L. 79: "However, only a few studies have investigated the effect of environmental unpredictability on reproductive seasonality". Not sure what you mean here – there is for example a large literature on trophic mismatch and environmental variability, as well as the importance of timing for population dynamics (eg Li et al. 2021).

We clarified the novelty of our work l. 81 by insisting on the difference between studying the effect of environmental unpredictability on the *timing* of birth seasonality (i. e. studying *when* the birth peak would occur) and on the *intensity* of reproductive seasonality (i.e. studying *why* there would be a birth peak and *how* narrow it would be).

l. 103: note that small multivoltine mammals are also "far from an integer number" (ie much less than 1), but that will have very different consequences as they have multiple generations per year, and different generations may have different decisions to make (when to start reproduction vs when to stop).

Thank you for this comment. We rephrased the sentence (now l. 95) to clarify our hypothesis, focusing on reproductive cycles above one year, as we are not able to investigate short and multiple cycles in this study.

l. 127: not sure this is correct even if I do not have any number to back it up. There are for example many studies on reproductive synchrony in large herbivores, and the costs and benefits of being born too early or too late. But the bias towards temperate species is clear, but not specific to reproductive seasonality.

We removed the reference to short-lived species to take better account of the literature on large herbivores and focused on the fact that baboons have multiple-years reproductive cycles (l.125). Even if it is true that the bias towards temperate species is not specific to reproductive seasonality, we think that this point is particularly relevant here as climate plays a critical role in shaping reproductive seasonality.

l. 142: you use NDVI as a proxy for resource availability but that is a rather strong assumption and you do not really provide empirical support by correlating NDVI to actual resources used by baboons (it seems to have some predictive ability but it is not clear exactly how). In different papers, you write that it is related to primary productivity (this paper l. 143), greenness and plant biomass (l. 188-189 this paper), vegetation cover (Dezeure et al. 2021 Breeding seasonality generates reproductive trade-offs, l 282). These different aspects of vegetation are of course somewhat correlated, but it would help developing better proxies of resources and be more explicit about what is relevant.

We homogenised our vocabulary and now use "vegetation productivity", following Pettorelli et al. (2005). We added precisions in the methods (l.194) about baboons' diet consisting mainly in plants and about how the model takes into account the use of alternative sources of food (fallback food). We otherwise agree that it would be good to develop better proxies of feeding resources used by baboons but this is particularly challenging in this omnivorous and highly generalist species, which is widely distributed and where populations likely show regional variation in diets. Since most baboon populations inhabit dry savannahs, NDVI variation is routinely used as a proxy for food availability in studies of baboon ecologies.

You would also need to check the values you have since in the papers you refer to (Dezeure et al. 2021 PRSB; eg Figure 1), you give values of NDVI around 0.10-0.15, whereas NDVI values so low would rather fit the environments I am more familiar with (ie high Arctic). Values of NDVI are ususall standardized and one would expect values around 0.5-0.8 in the study area.

The NDVI values around 0.10-0.15 given in Dezeure et al. (2021) correspond to a different location (Tasobis, Namibia) that is very arid indeed. The values used in this paper (Amboseli, Kenya) are slightly higher (mean value of 0.23, see supplementary l.155 and Fig S4) but still represent a semi-arid area with relatively low values of NDVI. Both Tsaobis and Amboseli NDVI ranges that we report are consistent with other studies (for instance: table 2 in Johnson et al (2015) - The ecological determinants of baboon troop movements at local and continental scales. Movement ecology). We clarified in the text l.129 that baboons' environment was semi arid. This is also why NDVI variation is a relatively good measure of food availability, which is extremely rainfall-dependent in such arid environments (see our response above).

L. 234: how do you decide on what is small vs large differences? You appear later on to use statistical tests but this is well known to be a very crude way to assess such differences. By increasing the number of simulations (eg 2 millions rather than 2000) I guess all differences would be statistically significant. Of course the P-values and the associated t-statistic you use is a function of the mean simulation difference and the simulation variance, so the question is what it brings in addition to just investigate the mean differences and possibly the variances if they differ a lot (which could be of interest)?

Thank you for this very important comment. We were not able to test extremely high numbers of simulation due to computational limits, and never reached a point where all differences were statistically significant, even if, as you (and the other reviewers) rightfully point out, such a significance would anyway have been meaningless. This nevertheless illustrates that, in this model, a unique optimal strategy never clearly emerged and explains why we needed to find a method to select a pool of optimal strategies. The use of p-values to do that was ill-advised as it is sensitive to the number of simulations. We instead followed the advice from reviewer 2 (Francois-Xavier Dechaume-Moncharmont) and used a 5% decrease on effect size criterion to select the optimal strategies (as described in the "Emergence, Adaptation" section l. 225). This new approach did not change our results.

L. 256: So in your model the component of adult survival (here approximated by lifespan) which is independent of energetics/resources is independent of environmental variation? But in the data, lifespan is partly determined by environmental variation, right? Just to be sure I understand the assumption here. One should not need to dig in the supplementary material to be sure.

We clarified this in the methods section of the main text (l.266): "Extrinsic mortality (external causes of death) is accounted for by assigning a maximal lifespan to each individual, also

randomly picked of empirical data (McLean et al., 2019) and independent from environmental variations. Individuals die if they reach their maximal lifespan, independently of their current energy resources. Yet, individuals can also die before their assigned lifespan is reached, if the environmental conditions do not provide them with enough energy to survive ("intrinsic" mortality). Overall, individuals in the model can either die from a stochastic external mortality or from an environmentally-dependent intrinsic mortality (see submodels "Death" and "Cycling" in supplementary for details)."

l. 324: as if the residual variation would represent "only" the biological variation... there is a large measurement error, 2 weeks variation is different from year to year variation... Just make this explicit.

The residual variation (NDVI_NS) indeed encompasses also a noise coming from measurement errors and that we cannot disentangle from the biological variation. We clarified that in the text l.340.

l. 331: this would clearly be "wrong" in terms of possible values for NDVI (i.e. NDVI=Normalized DVI is normalized between 0 and 1) - why not just remove completely NDVI and just use a theoretical measure of resources, as in Sun et al. 2020. I think the use of NDVI is more a distraction rather than anything really useful here.

We omitted to mention that we truncated the resulting time series between 0 and 1, thank you for catching this, it is now clearly said l.352. It is true that we could have used a theoretical measure of resources, instead of NDVI. Yet this measure would not have reflected the real variation of food abundance experienced by baboons. We have been very careful to use as many realistic values as possible to parametrize our model and shifting to an arbitrary time series for food availability does not seem consistent with this modeling approach. We believe that having a realistic model, despite the costs associated, is important to capture small shifts that could not be visible at a more global level, or with a different species, as well as to assess whether the theoretical variation of the parameters tested is realistic and could directly match existing environments or life histories.

Detailed comments:

L. 57: This seems a bit trivial to me - hard to imagine caribous breeding in the middle of the winter in Alaska... It would perhaps be more relevant to point out that some Arctic species may breed more or less year round (eg lemming under the snow) but others (caribou) are highly seasonal, and refer to tropical species that either breed non-seasonally or seasonally (for example the Serengeti species you mention).

We removed the trivial examples from the text and focused on the geographical "anomalies" and the sympatric differences in the following paragraph (l.61).

L. 63: even Arctic species can breed year-round, eg lemmings.

Thank you for this additional example that we added in the text l.63.

L. 89: We do not refer to delayed implantation in Langvatn et al. 2004 but to delayed ovulation at high densities.

Thank you for spotting this mistake. As suggested by reviewer 2 (Francois-Xavier Dechaume-Moncharmont), we decided to remove the paragraph about capital Vs income breeders that is not essential to understand our work and that would have needed further developments.

l. 90: Hellgren 1998 is not in reference list - is this Hellgren, E. C. 1998. Physiology of Hibernation in Bears. Ursus 10:467-477? This paper is more about physiology than ecology, perhaps use a more ecological/demographic paper. There is an abundant literature on bears of course but also marmots, etc.

We were indeed citing this paper, but the paragraph has now been removed from the introduction.

l. 359, figure 2: you cannot have negative values for r, perhaps find better regression lines, or cut them at 0. What is the grey colour on the figure?

We corrected the regression lines in figure 2. The grey cells represent non-viable environments and are characterised by a fitness of zero (regardless of the phenology strategy followed). Thank you for spotting that the information was missing, we added it in the mains text (l.356) and in the figure legends.

Figure 3: Please remove NS and stars from the figure, provide quantitative differences with some measure of uncertainties.

We removed NS and stars and replaced them with mean and standard deviation for each distribution.

l. 500-1: "intensity of reproductive seasonality was not associated with gestation length but rather with the size and number of litters per year (Heldstab, 2021)". For rodents it was the number of litters per year and it is a bit tautological – if one has many litters per year, it is difficult to image being highly seasonal?

Thank you for spotting the error regarding the results on rodents. We agree that an effect of the number of litters per year alone is not relevant and therefore we focused our argumentation on the effect of litter size l.572.

Nigel G. Yoccoz

References

Li, G., X. Wan, B. Yin, W. Wei, X. Hou, X. Zhang, E. Batsuren, J. Zhao, S. Huang, X. Xu, J. Liu, Y. Song, A. Ozgul, C. R. Dickman, G. Wang, C. J. Krebs, and Z. Zhang. 2021. Timing outweighs magnitude of rainfall in shaping population dynamics of a small mammal species in steppe grassland. Proceedings of the National Academy of Sciences 118:e2023691118.

Review by [Francois-Xavier Dechaume-Moncharmont](#), 27 Sep 2022 19:36

This article is a simulation-based investigation of the drivers of seasonal reproduction in the yellow Baboon. It aims at disentangling the complex effects of several ecological constraints and life-history traits. Most notably, it does a nice job in separating environmental productivity and seasonality which are inextricably correlated in studies investigating the effect of latitude: higher latitudes are both marked by high seasonality and low productivity, and this correlation is confusing when one want to understand the drivers of phenology. As a non-native English-speaking writer, I won't venture to criticize English or the syntax, which seem to be fine as far as I can tell. The MS is clearly written, with great deal of details, even if some important and non-trivial parts are to be found as supplementary materials. I also appreciated the fact that the source codes are developed in open-source languages and available on public depository. It certainly helps the reader to understand the model. The model is built on the Baboon biology, which is both a strength (the parameters are calibrated on field data) and a weakness (extrapolation of the results in non-ape species may be not straightforward). My overall impression about this MS is largely positive. Yet, several points need clarification or additional explanations.

Major comments

The introduction section should be partly rewritten. It should be more explicit about the differences between this study and previous literature. For instance, lines 79-81, one can read allusive statements: "However, only a few studies have investigated the effect of environmental unpredictability on reproductive seasonality (Dezeure et al., in press; English et al., 2012), with mixed results so far." First, the reference in press is submitted by co-authors of the present study, but the reader cannot appreciate the content of the cited.

The reference previously in press is now available online (https://doi.org/10.1086/722082). We modified the reference list accordingly.

Second, what do the authors mean by these mixed results? I recommend the authors to be more explicit about the novelty of their paper compared with existing literature.

We clarified the novelty of our work l.81 by insisting on the difference between studying the effect of environmental unpredictability on the *timing* of birth seasonality (i. e. studying *when* the birth peak would occur) and on the *intensity* of reproductive seasonality (i.e. studying *why* there would be a birth peak and *how* narrow it would be). We also clarified the mixed results found in previous studies on the subject.

The introduction is maybe too long at some points. For instance, the authors may reconsider the comments about capital VS. income breeders (lines 82 and following paragraph) as it is weakly relevant in the understanding of the analyses conducted in the model.

As suggested, we decided to remove the paragraph about capital Vs income breeders that is not essential to understand our work and that would have required further developments

The way of finding optimal strategy among simulated phenologies for a given set of simulation parameters should be detailed and illustrated at some point of the manuscript. I recommend to plot (possibly in supplementary material) the reproductive outcomes as function the phenologies for several sets of simulation parameters. Such plots allow for understanding the strength of selection pressure around the optimal phenology value. Even if a given value corresponds to the optimal phenology, the selection pressure towards this value could be very weak if the curve is rather flat around the optimal value. Mathematically the selection pressure corresponds to second-order derivative around the optimal value. Ideally, I would be very interested in finding quantitative information about theses selection pressure as a function of the environmental conditions. I expect that some set of condition leading to intermediate reproductive phenology are in fact under very weak selection pressure: following suboptimal phenology are poorly counter selected. This information is very important in the context of uncertain environments. It makes no sense to evolve highly efficient phenology leading to peaky optimal response curves because any uncertainty of stochasticity in the actual set of environmental condition would result in very suboptimal fitness gain. Finally, an additional interest of such plots is to assess the possible existence of local maxima leading to sympatric polymorphisms, or difficulty evolve from one phenology to another in absence of high-range mutation jumps.

Thank you for this very relevant recommendation. We now provide graphs of fitness as a function of phenology strategies in figure 3. We unfortunately cannot provide such graphs for each set of parameters studied (i.e. for each tile of each heatmap presented in the results), but we investigated several of them without observing any polymorphisms situations. We agree that it would be very interesting and informative to have quantitative measures of the strength of selection pressure for each of those cases. The difficulty lies in the topology of the space of possible phenology strategies. As you suggested, in order to quantify the strength of selection pressure, we would need to compute the second-order derivative of fitness around the optimal value, with respect to the strategies. Yet, the phenology strategies studied are characterized by two parameters (the start and length of the reproductive window), and constitute therefore a two-dimensional space. Furthermore, one of those parameters – the start date of the reproductive window – is a circular variable, which means that this two-dimensional space is not even a plan but a cylinder. In this context, computing the second order derivative in a cylinder-shaped space appears to be far from our area of expertise. It would also reinforce the complexity of this paper, in a context where, as you rightfully point out in one of your later comments, theoretical models already have little appeal from most ecologists.

Still on this question of finding the optimal strategy, I disagree of the method based on significant difference in simulation outcome. For instance, even the smallest differences become highly significant when the sample increase. It is only a question patience and brute force. The t-test were based on 2000 simulations. Fair enough, you just need to increase the number of simulations to finally obtain significant p-value (it is one major objection against the use of p-value calculation in

science). Please consider metrics insensitive to sample size, for instance 5% decrease of effect size index or similar arbitrary criterion. The use of **statistically significant** difference is misleading and questionable in simulation context.

Thank you for this very important comment. We were not able to test extremely high numbers of simulation due to computational limits, and never reached a point where all differences were statistically significant, even if, as you (and the other reviewers) rightfully point out, such a significance would anyway have been meaningless. This nevertheless illustrate that, in this model, a unique optimal strategy never clearly emerged and explains why we needed to find a method to select a pool of optimal strategies. The use of p-values to do that was ill-advised as it is sensitive to the number of simulations. We instead followed your advice and used a 5% decrease on effect size criterion to select the optimal strategies (as described in the "Emergence, Adaptation" section l. 225). This new approach did not change our results.

The way of modelling the reproductive cycle by artificially increase the number of days per year (lines 337 and followings) is rather hard to understand, and the results are difficult to interpret. I did my best to understand this modelling choice, but I am still puzzled by this analysis. I am not saying that the method is incorrect, I am just saying that, if conserved in the latter version of the MS, this non-trivial modelling choice deserves much clearer presentation.

With the added sentences and example l.376 and followings, we hope this procedure is clearer for the reader now:

"In other words, we "stretched" the NDVI time series so its seasonal variations would be perfectly synchronised with the reproductive cycle (year length = interbirth interval) or desynchronised (year length = 1.5 interbirth interval). For example, in the synchronised configuration, with a year length of 637 days, the annual good season would always fall during the same phase of the reproductive cycle. In this particular case, each month lasts 53 days and the raw NDVI values, originally spaced of 16 days, are spaced of 28 days."

We agree that this approach is not trivial, but it appears to us as the most biologically neutral solution, which is of key importance, as we explained more precisely in the main text l.368 and followings:

"Reducing or extending the reproductive cycle length (for example by modifying gestation length, growth rate, weaning mass…) would indeed be an alternative approach to synchronise the reproductive schedule with the annual cycle. Yet, by also modifying the energetical aspects of the reproductive cycle, this approach could have an additional effect on reproductive seasonality (e.g. the effect of daily reproductive energy expenditure (H4)). On the contrary, modifying the year length alone allows to isolate the effect of synchronisation (i.e. removal of the gaps between reproductive events), without altering the energetical transfers between mother and offspring.

Caveats. I recommend that the authors acknowledge and discuss several caveats about the present model. In its present form, it is a statistic optimisation model. The strategy of the female is fixed throughout her lifetime. Her investment does no vary dynamically as a function of her own reserve or age. For instance, the model does not allow for terminal investment or increase in reproduction investment toward the end of the life. Such analysis would require dynamic state modelling (dynamic programming).

We insisted on this point in our discussion l.613: "Additionally, and on top of sociality, dynamic individual strategies are not considered in this study: a female's investment does not vary dynamically as a function of her age and the model does not allow for terminal investment or increase in reproductive investment toward the end of life. Such variation could be best modelled with dynamic state modelling (dynamic programming)."

Similarly, the model optimizes the strategy of an isolated female in absence of any competitor. Yet, it is possible that stringent environmental conditions lead to polymorphism of strategies, some females choosing to delay or bring forward their reproduction to avoid competition for limited resources. Here, game theoretic models would be relevant to cope for frequency dependent effect.

I fully understand that such developments have been neglected in the present study for sake of simplicity. Yet, these caveats should be pointed out in the discussion section.

Following your comment, we insisted further on this limitation in our discussion l.607, where we discuss the effects of sociality in a context of reproductive competition:

"Such a polymorphism of strategies, where some females may choose to delay their reproduction or move it forward to minimize competition over limited resources, could not emerge from our model that only optimises the strategy of an isolated female in the absence of competitors. In order to integrate such frequency-dependent effects, other models, such as game theoretic models, should be developed."

Minor comments

Line 74: awkward wording choice "(…) disentangling environmental productivity, seasonality and their interaction". At least the analysis could disentangle **the effect of** environmental constraints, but the environmental parameters themselves.

Thank you for this correction that we included in the text l.75.

Line 83. Merge the two citations

Thank you for spotting this typo that is no longer present as this initial paragraph about income vs capital breeders has now been removed from the introduction.

Lines 87-93. This sentence is way to long and confuse. Even after several attempts, I still do not fully understand its point.

Following your suggestion above, we removed this paragraph that is not essential and would indeed have required further developments. The general idea was that energy storage in itself cannot be a reliable predictor of seasonal reproduction as it is a mechanism shared by both seasonal and non-seasonal breeders.

Line 101. While I understand what the authors mean by "close to an integer number of years", it is unnecessary obscure way of presenting the asynchrony of reproduction schedule across year. I recommend that the authors take greater care in presenting their analysis is a clearer way. Theoretical models are sufficiently unappealing for most ecologists (Fawcett TW, Higginson AD. 20212. Heavy use of equations impedes communication among biologists. Proceedings of the National Academy of Sciences 109:11735–11739 http://dx.doi.org/10.1073/pnas.1205259109) for the authors to indulge in needlessly obfuscating formulation.

We rephrased the sentence (now l.95) to clarify our hypothesis and simplify its presentation. Following comments from reviewer 1 (Nigel Yoccoz), we exemplified the hypothesis by focusing on reproductive cycles above one year, as we are not able to investigate short and multiple cycles in this study.

Line 102. "far from" is too vague a term. Be more specific.

We removed this term from the new sentence (see previous answer).

Line 124-125. "This species was chosen here because it belongs to a genus characterised by phenological flexibility with one in six Papio species breeding seasonally". This a rather weak justification for model species choice.

We further developed the explanation of our model species choice in the text (l.119), clarifying how it constitutes a new approach in this field:

"This choice of a long-lived tropical and non-seasonally breeding (Campos et al., 2017) model species contrasts with most studies on reproductive seasonality. First, focusing on a non-seasonal breeder and on how reproductive seasonality could emerge from this point is a new approach and the Papio genus is well suited for it because it is characterised by phenological flexibility, with one in six Papio species breeding seasonally (Petersdorf et al.,

2019). Second, studying a tropical species like baboons is of major interest because most studies of reproductive seasonality focus on organisms from temperate regions (Bronson, 1985; Vatka et al., 2014). Lastly, long-lived species such as baboons, where reproductive cycles spread over multiple years, tend to be under-represented in studies of breeding seasonality, but could bring important insights to understand our own reproductive phenology, namely why humans reproduce year-round."

p.329. I was strongly surprised by this modelling choice for the very reasons finally exposed toward the end of the MS (discussion section line 459 and followings). I recommend that the authors are more explicit of this equation and its limitation in this section.

We described the limitations of this decomposition in this section l. 342 : "In addition, unpredictability is a multidimensional entity that cannot be restricted to the NDVI_NS component alone: for example, one additional dimension is the unpredictability in the timing of the food peak. In other words, this decomposition describes the changes in NDVI magnitude for the same calendar date, but does not directly capture the changes in timing of the NDVI time series."

Similarly, I have the greatest difficulty to appreciate the relevance of the arbitrary range of variation for p, s and u parameters listed lines 330-332. Please, elaborate on this point.

We explained our choice of range of variation for the three parameters l.354: "The ranges of variation for the three parameters p, s and u were chosen in order to explore substantial variation in environmental productivity, seasonality and unpredictability while still simulating viable environments for baboons (i.e., with a sufficiently high productivity).

Line 334. Be more explicit: "daily **reproductive** energy expenditure".

Thank you, we clarified the expression for each occurrence in the manuscript

Line 341. What is MODIS data? Explain and give appropriate references.

The MODIS data are the raw NDVI values that we downloaded for the Amboseli location. They are cited earlier in the methods (l.203: "MOD13Q1 product") and this does not need to be repeated here. We replaced it with "raw data".

Line 353. Fig. 2c instead of Fig. 2b.

We removed Fig 2b, so it is now correct.

Fig 2b is not called in the main text. In addition, I do understand its relevance here.

We removed Figure 2b.

Fig. 2, 3 and 4. I do not understand the reason and meaning of the grey cells. Please explain them in the legend and the main text. In addition, the default parameter values for the simulation should be stated in the legend.

The grey cells represent non-viable environments and are characterised by a fitness of zero (regardless of the phenology strategy followed). Thank you for spotting that the information was missing, we added it in the mains text (l.356) and in the figure legends. We also added the default parameter values in each legend.

Line 425. New results, and particularly highly important ones, should not appear for the first time in the discussion section. This paragraph should be moved in the result section.

We moved the paragraph in the results section l. 387.

What is $\tho$ in supplementary p.28?

We added the definition of tho in the legend, in accordance with the definition given in appendix B (l.124)

Review by anonymous reviewer, 30 Sep 2022 13:42

The authors describe a model that explores the environmental conditions under which reproductive seasonality is favoured. I think it's an interesting piece of work and the results seem plausible. I do, however, have some queries about the methods:

Major comments:

I'm not sure I understand the necessity of statistical tests for comparing model outputs. Isn't it possible to just run more simulations and decrease uncertainty in the mean of the output to any desired degree of accuracy? Relatedly, have you tested the stability of your mean/variability estimates given different numbers of simulation runs (i.e. why 2000 runs)? There's a bit of literature on these issues see appendix B of (1) - it might be preferable to have an effect size rather than a probability based measure of significance.

Thank you for this very important comment. We were not able to test extremely high numbers of simulation due to computational limits, and never reached a point where all differences were statistically significant, even if, as you (and the other reviewers) rightfully point out, such a significance would anyway be meaningless. This nevertheless illustrates that, in this model, a unique optimal strategy never clearly emerged and explains why we needed to find a method to select a pool of optimal strategies. The use of p-values to do that was ill-advised as it is sensitive to the number of simulations. We instead followed the advice from reviewer 2 (Francois-Xavier Dechaume-Moncharmont) and used a 5% decrease on effect size criterion to select the optimal strategies (as described in the "Emergence, Adaptation" section l. 225). This new approach did not change our results. We increased the number of simulation runs until reaching stability in the results (it is now specified in the manuscript l. 298).

I might be missing a step here, but why include stochasticity in mortality/reproduction at all? You ultimately seem to use multiple simulations to converge to the non-stochastic answer anyway (although not completely e.g. Fig 2a). Is it to see if stochasticity would overcome small differences derived from optimal breeding time?

It is true that stochasticity has a high computing cost and that we end up averaging over simulations. In order to remove stochasticity from our model, we would need to include the "average effect" of stochastic conception and mortality on reproductive seasonality. Yet, depending on the phenology strategy that we are testing, we believe that this average effect can be very different. For example, without stochasticity, we would always have a fixed time period between the beginning of a reproductive cycle leading to the birth of an infant and its possible extrinsic death. If this period lasted 1 year and n months we would certainly slightly bias our results towards phenology strategies where reproduction is possible during a window lasting more than n months. Indeed, any female with a strategy where reproduction is restricted to a period shorter than n months would always need to wait for her next reproductive window after her infants' death, whereas females with longer windows could attempt to reproduce immediately after. Because such effects are difficult to predict, we decided to make the minimum number of possible assumptions in order to reduce the risk of introducing such a bias. In summary, we tried to have a model as realistic as possible and stochasticity in mortality and reproduction is part of the reality experienced by yellow baboons.

L. 334 To test an increase in daily expenditure you increase the growth rate which subsequently impacts the gestation period. Why not just increase expenditure through the constant of energy expenditure (equation 6 appendix)?

An increase in the constant of energy expenditure used to calculate the energy needed for maintenance (equation 6 in supplementary) would indeed increase the energy needed for reproduction through the maintenance of offspring. Yet, this would have an even bigger effect on the energy requirements of adult females that are not reproducing (because of their much higher body mass). Increasing the daily reproductive energy expenditure

through growth rate makes this increase more specific to reproduction phases. Additionally, increasing growth rate seems more realistic as important variations in growth rate are common between species with different life history paces while differences in energy for maintenance have been, to our knowledge, less documented. We added this information in the text l. 360

I also find the changes to the length of the year slightly odd. I think it's probably OK, but seems like a strange approach – isn't it possible to impose a non-breeding period that shifts the cycle through the year - or perhaps this would cause other issues?

Following comments from reviewer 2 (Francois-Xavier Dechaume-Moncharmont), we described more precisely the changes regarding the length of the year and gave an example l.376, We agree that this approach is not trivial, but it appears to us as the most biologically neutral solution, which is of key importance, as we explained more precisely in the main text l.368. Imposing a non-breeding period within the reproductive cycle would indeed be an alternative solution to artificially extend the cycle length. The main issue we see with this solution is that it only allows to lengthen the cycle (that is already longer than a year) and would not simulate the effect of a perfect synchronization where the reproductive cycle lasts exactly one year.

Minor comments:

L.47 'allows synchronizing the energetic costs...'

Thank you, we corrected the sentence.

L.306 Slightly puzzled by this paragraph, is the ABM 3d or is it 3 dimensions of parameter space? You mention '3d rep is difficult to apprehend and increasing computing time' does this mean not all levels of unpredictability were simulated? I might consider rewriting parts of this paragraph for clarity around what is in the model and what is being presented as a result (e.g. heatmaps)

We clarified the text in the methods section l.316. The 3 dimensions mentioned represent the 3 dimensions of the environmental parameters space: seasonality, productivity and unpredictability. We simulated only four levels of unpredictability.

L.318 the equation as written and described implies that the NDVI would be timeseries mean + daily mean which would give too much NDVI, whereas I think you calculate a daily offset from the mean (i.e. positive or negative)

Thank you, there was indeed an error in the description of NDVI_S which is the mean daily *offset*. It is now corrected l. 334.

L.351 I'm not sure if this sentence is correct, you see high birth seasonality with greater than the observed productivity (p=1.1 to 1.3)

We mean that we always see high birth seasonality for the lowest values of viable productivity (i.e. that keep the environment viable for baboon: non-grey cells). We clarified it in the text l.396.

L.411 and other figures: you don't typically need to say you 'plotted' you can just say 'Panel A shows ... '

We simplified the wording for all the figures, thank you.

(1)Lee, J.S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, J.G., Sun, Z. and Parker, D.C., 2015. The complexities of agent-based modeling output analysis. Journal of Artificial Societies and Social Simulation, 18(4).