Effects of climate warming on the pine processionary moth at the southern edge of its range: a retrospective analysis on egg survival in Tunisia

Revision 1

We are very grateful for the positive and constructive comments we received from all three reviewers as well as the recommender on this manuscript. Multiple points were raised throughout the manuscript that needed being clarified or improved. We have addressed issues to the best of our ability when the four reviews were sometimes hard to reconcile, giving priority to the concerns that several reviewers shared. We believe this round of reviews has significantly improved and strengthened the manuscript. We hope that this revised version will adequately answer your remarks and that you will consider this work suitable for publication in PCI Ecology. Our detailed reply to the comments and the detail of revisions are reported in the following pages.

> **Élodie Vercken (recommender)**
>
> Dear Authors,
>
> Thank you very much for submitting your work for recommendation in PCI Ecology. Your manuscript has been evaluated by 3 reviewers and myself, and we all concur that the topic of the study is most timely, and the use of collection data is an excellent way to address issues around long-term climate change and its effect on living organisms.
>
> The manuscript was described as well-written, scientifically sound, and likely to be of interest to a wide community of ecologists. I would be most willing to recommend this work for PCI Ecology, provided you can address the concerns raised by the different evaluators.

Thank you very much for the positive review of this work, as well as the highly relevant comments and suggestions you have made for the analysis and figures.

> Please pay particular attention to the following points that appear most critical for a better :
> - Use of NASA satellite data at lower scale to measure temperature variations
> - Cluster definition and inclusion/exclusion of some of the data (e.g. data from Thélepte in later years)
> - Analysis of time series for temperature
> - Use of GLM(M)s instead of non-parametric tests
> - Justification of the relevance of "length of egg masses" as a variable

Answers to first four points are detailed in other answers below. Regarding the last point, the length of PPM egg masses is interesting because the ratio between length and number of eggs in a clutch can be used to describe varying levels of egg density in an egg mass, which may relate to microclimatic conditions under the protective scales. Egg laying and egg density may also be affected by the host tree. In Tunisia, drought has reduced the length of pine needles, for this reason the length of egg masses here are quite lower than that in the northern parts of PPM distribution (Europe). However, we found here that clutch length and fecundity were highly correlated and redundant in our samples, therefore we present the data but focus analyses on fecundity. A sentence was added in the Results L311-315:

"The length of egg masses was measured in case the distance between eggs of a clutch would differ among areas

or periods, but this variable was found to be highly positively correlated to fecundity (Spearman correlation test, rs = 0.72, p < 0.001), thereby leaving little room for variations in the fecundity/length ratio. Therefore, we focused further analyses on fecundity only...”

Don’t hesitate to ask if any of the reviewers’ comments are unclear to you, or if I can help in any way. I sincerely hope our comments are helpful to you and contribute to improving the manuscript further.

Élodie Vercken’s review:

1. I liked very much the analysis of the climate clusters, I reckon it is a very good way to address the issue of unbalanced sampling design. Only concern here is that I don’t understand what are the data points on Fig 2. There are 30 points for each meteorological station, so it seems there is one point for each year, but what is it? TN, TX, their average? Please clarify in the text and/or figure legend.

Each point indeed corresponds to the PCA score for one sample year projected in the multivariate space based on its values in the 24 variables described in L199-204: the variables are monthly averages of TN and TX (one TN and one TX variables from January to December, resulting in a set of 24 variables). Since the PCA is only used as a statistical prerequisite for the PAM clustering procedure, not to describe climatic features (which are better described per cluster in Figures 3 and 4, and per climate station in charts in the Supplementary Material), the 24 variables are not reported on Figure 2. L258-259 in the legend of Figure 2 state that points are PCA scores after multivariate projection according to the monthly TN and TX variables, and that the PAM clustering results are overlaid onto the PCA scores (L260-261).

2. Once the clusters were identified, I don’t understand why clusters 2 and 4 are kept in the analysis although they don’t include any sampling point from the “early” period. There seems to be a problem in testing a cluster x period interaction with 2 clusters containing data from a single period only. Alternatively, if you are interested mostly in inter-cluster differences, then you should not analyze longitudinal data from the 2 periods. Choose your question here, and select the data that help you answering it. I suspect there is enough data within cluster 1 and 3 to address the issue of temporal variation in reproductive success across clusters. At present, you are trying to analyze temporal and cluster effects independently, but you end up describing partial interactions which makes your argumentation very case-specific (e.g. L. 295-303, L. 310-320 you mention a significant effect of period, but you describe only effects within clusters and never the shared trend across clusters).

We acknowledge this statistical issue and the text did not explain why all clusters were kept. While the main focus of this work was to investigate trends over time, we also wanted to take the opportunity to compare average egg phenotype among climate clusters, as it is equally crucial to indirectly decipher the effects of climate on PPM eggs. We were therefore also interested in the cluster effect ignoring the period effect. In order to achieve these joint goals while fixing the issue of unbalanced design you raised, we have split the analysis in two for each goal:

- In Analysis (1), we considered only clusters where both past and present samples are available (clusters 1 and 3), and the cluster:period interaction can be investigated.
- In Analysis (2), we considered all clusters and just the period they have in common, i.e., between clusters 1, 3, and 4 in the 1990s, and between clusters 1, 2 and 3 in the 2010s.

The paragraph of “*Egg phenotype comparisons*” in the third subsection of Materials and Methods and the Results have been modified accordingly (see lines 228- 247, 316-330 and 352-369).

3. I do have a remaining concern about including single samples from a given year within each

Cluster (i.e. 1992, 2017 and 2019 in Cluster 1 ; 1993 in Cluster 3). By including these points, you make the assumption that annual variation within each period is negligible, but then you do mention extreme events in 2017 in the discussion to help explaining your results. I would like to see how the inclusion or exclusion of these points affect the results of the study. On the whole, there is quite a lot of stress on these particular events in the manuscript (e.g in the abstract and the discussion), but they might not be reflecting any wider trend, and they are not directly supported by the climatic data you provide.

We understand that years represented by a single site can raise concerns. We used a statistical approach to group sites by climatic similarity and ensure egg masses originating a climatically consistent areas could be considered replicates for the questions addressed in this study. Discarding samples a posteriori is hard to reconcile with this approach, but we agree that the text should emphasize more clearly that caution must be taken to interpret results pertaining to the heatwave of 2017. We have greatly modified the results and discussion sections to:

(1) Better dissociate results specific to cluster 1 in 2017 and other overall trends from past to recent years.

(2) Add a complementary analysis on a subset of the data without the site sampled in cluster 1 in 2017 (see lines 370-379, see also SM8). We chose to keep data from 2019 because hatching rate at Thélepte in 2019 was similar to that of other sites within the same cluster in the recent period, further supporting that this population is not abnormal compared to other populations of the cluster. Additionally, it should be noted that the sample size from Thélepte in 2017 is one of the largest in the whole data series (43 clutches sampled that year), which is now explicitly stated in the manuscript.

(3) Elaborate on the characteristics of the 2017 heatwave, which also replies to two other reviewers. Philippe Louâpre was interested in these data but concerned about not seeing the heat days in Kasserine supplementary material, which was due to cold days (blue) being displayed on top of heat days (red) and concealing the latter, given their relatively low proportion in Kasserine. José Hódar also raised that samples from Thélepte 2017 could be outliers, a topic that is now better covered in the manuscript by showing that 2017 was indeed historically extreme but congruent with the general trend in Kasserine, and particularly relevant to locate the tipping point of egg survival in the PPM given that the stress endured in 2017 corresponds to thresholds observed in previous experimental studies. We therefore also rephrased some sentences to support the relevance of not neglecting these data to understand PPM dynamics under challenging climates, while better discussing that 2017 was a climatic anomaly.

We are confident that these changes better depict the importance of the data in 2017 as well as the limits to generalizing them to the whole recent period or cluster, and we hope that you and the other reviewers will agree. The changes are significant and scattered in several places of the MS, therefore they are not detailed here, but can be tracked in the main body of the MS with visible version history.

4. I am far from convinced that the non-parametric approach chosen here is the most appropriate for your data. I did not understand if you analyzed each egg mass as a replicate point, or if data were aggregated by sampling site. There seems to be an issue with pseudo-replication (uneven number of egg masses and/or year replicates for each sampling site), so a mixed model including sampling site as a random effect would be more appropriate here. Cluster and period should be included as fixed effects. The fact that the variables are not Gaussian should not be a problem, as Generalized Linear Mixed Models can be used (e.g. Poisson or Negative Binomial distribution for the number of eggs ; Gamma distribution for egg

length; Binomial distribution for hatching or parasitism rate).

We agree that GLMMs would be a robust approach to handle the non-normality of some variables as well as nesting data by using sampling site as a random effect. However, this was the approach we initially tried in our analyses before the first submission and couldn't improve the quality of residuals despite selecting appropriate distribution families in the models, which is why the non-parametric method appeared as more reliable for this data set. This approach didn't allow aggregating egg masses by sampling site as with a random effect in a mixed-model approach, but it is important to consider that sampling sites within a cluster are usually separated by only a few tens of kilometres (and that sampling within each site is also scattered geographically), which is a a regular sampling radius for studies of population dynamics in the PPM considering field variations in population density and adult dispersal power. The difference, however, is our sampling design is more patchy with multiple geographical centroids corresponding to different sites, which the current approach does not account for, but the low intra-cluster variance (small error bars) we observed for every egg variable when multiple sites are grouped together supports that clustered sites do indeed follow similar patterns.

5. I have strong doubts about the methods for the analysis of change in temperatures : (i) first, the 30 years of monthly averaged TX between 1990 and 2019 are analyzed with linear regression, instead of specific methods that exist for the analyses of time series (e.g., spectral analyses, auto-regressive models, etc) that are more powerful to detect underlying trends ; (ii) for the analysis of TX averaged over summer, the time period is split into two groups (one for 1990-2004, and one for 2005-2019 if I understood correctly) and period is modelled as a factor in an ANOVA. I really don't see the logic in this analysis, the summer-averaged TX could be analysis as a continuous time series, just like the monthly-averaged TX. It might make sense to split time into two discrete groups if it were to analyze the differences between the two periods of sampling (1992-1995 vs 2010-2014) that are indeed separated by a few years. The analysis chosen here has very low power and I am not surprised it fails to detect any effect of time.
In addition, the abstract mentions an increase in climatic variability, but I did not find this analysis in the text (I might have missed it)? Time series analyses can also be used to reveal temporal changes in the amount of stochastic noise in the data, which could support your argument here.

These suggestions are relevant and we have spent a large part of the revision time to try implementing them. We absolutely agree that the approach comparing average summer TX between two periods was weak, let alone for periods that were broader than those in the egg data series. Constraining the climatic periods in this comparison to the climatic periods when eggs have been sampled would have been impossible, since the sampling effort within clusters was not constant over time and some years are missing or sampled on a varying number of sites. After reviewing the whole manuscript in light of the changes it incurred to reply other queries from the reviewers, we realized that this analysis was little relevant and did not add any extra information compared to the distribution of daily TX within each month (now Fig. 3) or linear trends in each summer month (now Fig. 4), and consequently we decided to remove this analysis. We believe this removal did not impede the interpretation of climate trends over time within each cluster.

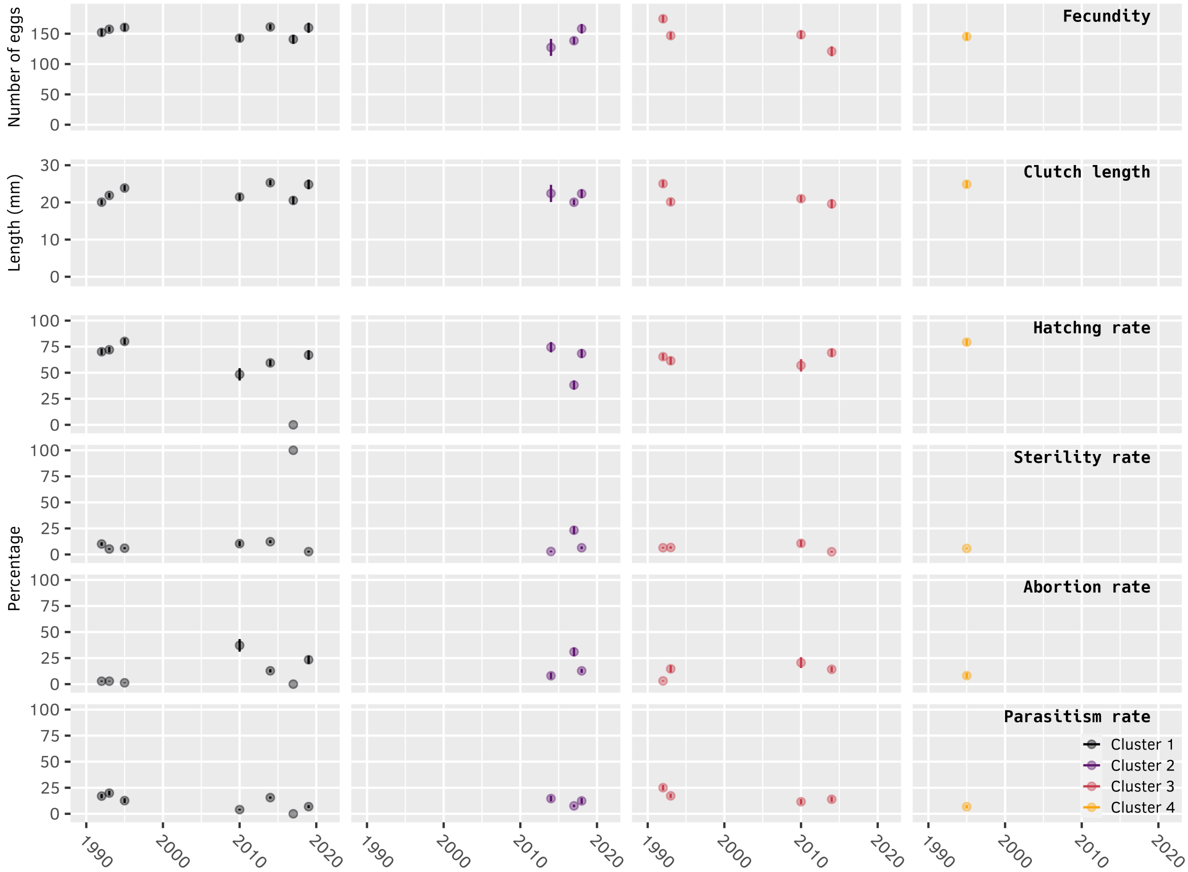
We kept the analyses performed on the 1990-2019 continuous data series (Fig. 4). Seeking linear trends may neglect some of the complexity of climate change over time, but our aim was to test trends and their magnitude at broad temporal and spatial scales, to keep the main focus of the manuscript on the season of interest for PPM eggs in Tunisia. Since egg data were not available for every year, more in-depth analyses of climate time-series may not have helped understanding egg results. While trying to implement the suggested statistical toolkit into this case study, we found that the new analyses would shift the focus of the interpretation towards other parameters of climate variability, such as seasonality or short series of autocorrelated days. Our simpler analysis here is performed on monthly means of TX across years and is therefore not subjected to the autocorrelation expected between consecutive individual daily points. The approach succeeded in discriminating contrasted trends among clusters, namely showing the comparatively fast warming observed July and August in some clusters, but not September, which is of critical importance in PPM eggs because their occurrence in each clusters varies along these months.

In addition to cited references suggesting the future increase in the frequency of stochastic events, we have added two figures to illustrate the frequency and duration of August heatwaves in Kasserine (Fig. 6, Fig. SM 9), the station closest to the population were eggs seem to have been impacted during the 2017 heatwave.

> 6. Some of the figures are poorly informative. The axis on Fig 4 is redundant with the legend box that already assigns colours to clusters, and it is misleading as it does not reflect the continuous nature of the temporal data. Fig 6 should be redone: bar plots are not helpful here to visualize small quantitative differences and the temporal dimension is not apparent on the graphs. I would rather suggest scatter plots of average values for each sampling point (with confidence intervals or quantiles) plotted against year on the x-axis, and if possible the different clusters plotted on the same graph if it not too crowded.

The x-axis labels and legend box in Fig. 4 were made redundant to keep figure colors consistent across the manuscript while making it legible in black and white, but this figure has been removed during this revision.

We totally agree with the suggested change in the axes of Fig. 6 (now Fig. 5), this was actually a discussion topic when preparing the manuscript for the first submission. We have completely rearranged this figure to now show years as a continuous x-axis, which clearly shows periods where no data is available, as well as a better illustration of the global timeframe of the data. Plotting clusters in the same graph panel turned out to make it hard to read. The version we have included in the manuscript is based on bars because we believe it is more legible than a scatter plot (small error bars, particularly, are harder to distinguish from average points in the scatter plot), but we have included a scatter plot variant below so that you can decide which one you believe will be most relevant for the final manuscript:



7. Tables are difficult to read and provide information that is not entirely necessary to understand the results. I would suggest to modify Figure 1 to include information about sampling periods (with a color code for instance) and to remove Table 1 to the Supplementary Information. Tables 2 and 4 are quite redundant with the text, the values of adjusted mean are difficult to interpret, I would also suggest to remove both Tables to the Supplementary Information and to focus on improving Fig 4 and 6 to make any trends more visible.

Our attempts to add this information to Fig. 1 with an extra colour code to each point made it too cluttered and really difficult to read, while not fully removing the need for table 1. This table is still necessary to show sample sizes for instance. We thus chose to keep Fig. 1 and Table 1. However, following your remarks, we moved (and modified) Table 2 to the Supplementary Information (see SM8), and we deleted Table 3 from the text because the results have been modified based on the changes we made in the analyses (see point 2 above).

8. I had the feeling that the discussion was a bit lengthy and it was sometimes hard to keep up with the main message. I reckon it could benefit from a bit of shortening, with a tighter focus on the main questions asked in the study (e.g. inter-cluster comparisons might not provide major insight here, the part on local adaptation seems a bit far from your results, detailed description of isolated data points make it difficult to see the higher trend).

We did our best to keep the discussion as concise as possible and shortened several sections, but at the same time had to elaborate others to address the different queries raised in all four reviews. We hope, however, than the discussion structure is now more robust and easier to follow.

Matt Hill

Overall, this is a well written and interesting study, I enjoyed reading it. The authors effectively use data from both historic (1990s) and current (2010s) time periods and present a novel study of range retraction due to lowered fecundity (and different aspects of this), likely associated with increases in temperature.

Different aspects of temperature in terms of variability and seasonality are well explored to help support the conclusions in this study, and alternate hypotheses around parasitism rates are well covered.

Thank you very much for the positive review on this work and for the helpful remarks to improve the manuscript. Below are detailed answers to your comments.

The one main issue I have is probably **only a point for discussion**. The climate clustering of the different egg masses relies on data that is collected quite a number of kilometres away from the sample sites. While this probably cannot be helped, it is worth mentioning that there is possibility of some margin of error around how the clusters are defined. On lines 149-160 the authors mention the use of gridded data to patch in missing weather information. It appears like this was just done at the eight weather stations? Wouldn't it also be worth looking at these gridded NASA data at each of the sampling localities? If these data are interpolated to 8km grid cells, then it may account for some of the distances between weather stations and sites, and perhaps give differences in clustering rather than relying on the values at 8 weather stations to define the clusters. While any interpolation would lose accuracy with distance away from stations, the smoothing of the temperature data in a gridded form may be more insightful than the raw station readings?

We found a confusion in the text when addressing your comment: the spatial resolution of the NASA climatic data is not 8 km x 8 km but 0.5 degree × 0.625 degree (about 50 km × 60 km). The mistake arose from a mix-up with the spatial resolution used for monitoring the PPM range in other studies that were conducted at the time the manuscript was written. Given this relatively coarse spatial resolution for NASA data, we feel that restricting their use to fill missing data only and keeping data from meteorological stations when available will ensure the most accurate depiction of the local conditions in the sites we considered. Considering grid NASA data only would however be an interesting approach for future studies at a large scale in Tunisia because it would allow finding robust climatic regions while explicitly taking into account the spatial dimension in a continuous fashion instead of isolated measurement points, but this would be at the cost of lower accuracy and higher risk of failure to detect deviations to actual local conditions experienced by eggs.

The Discussion is quite lengthy, mostly it is good however there are a couple of instances where results are presented before discussion of them. Some of these could be left in the results section, namely lines 338-342, 343-346, 363-365.

We removed lines 416-420 and 421-424 from the discussion but did not move them to the results to avoid redundancy there. We kept lines 459-460 to clarify that despite the higher probability to exceed 32 °C and 40 °C in cluster 2, we found that hatching rate was higher in cluster 2 than cluster 1, likely due to the influence of the low value recorded in 2017 in the latter, as well as phenological differences between the two clusters that may lower the summer heat risk exposure of eggs from cluster 2. The discussion has been shortened in multiple sections, but at the same time had to be completed to follow suggestions from other reviewers.

Some smaller points:

Line 57: Consider changing to “This average climate warming has already impacted phenology and the distributions of many…:

As we broadened the beginning of the introduction as suggested by Phillipe Louâpre, we have made some changes and the new sentence now reads as follows: “Climate warming may induce heritable as well as plastic changes…” (see lines 57-59)

Line 68: Consider changing “their combination” to “the combination of these”

Done.

Line 70: Please add the name of the describer, and the year.

Done.

Line 75: this instance of “Lepidopteran” is not needed

We deleted the word.

Line 82: Consider changing to: “Contrary to the beneficial”

Done.

Line 95: “Authors found” is not clear if referring to the previous sentence or the one references at the end of this sentence.

These are actually the references cited at the end of the sentence, we clarified this in the text (see line 110).

Line 140, Table 1 caption: add “climate” before clusters, it wasn’t clear at this stage of the MS what a cluster was.

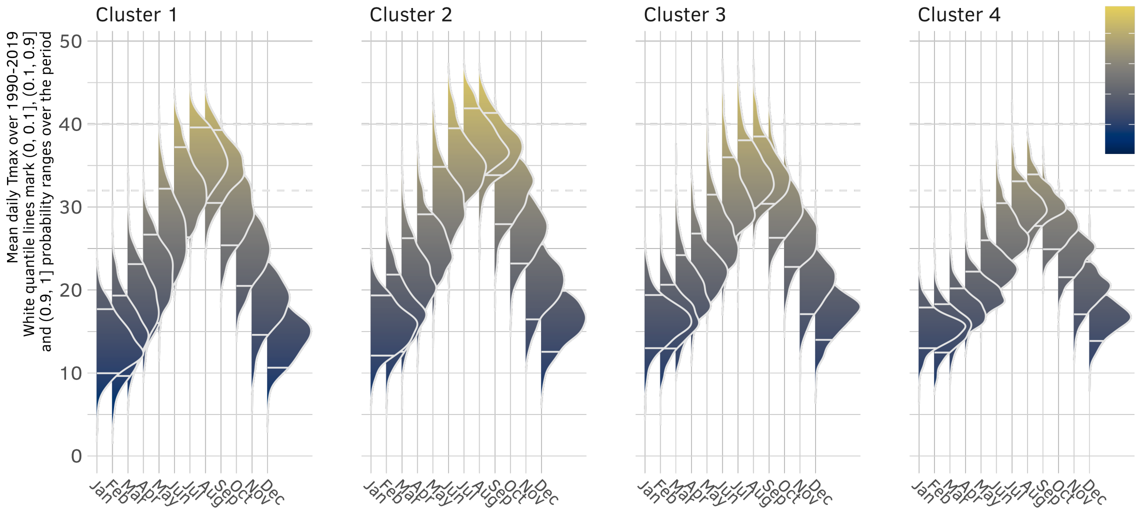
Done.

Line 333: comma after “in Tunisia”

Done.

Figure 3 – is it possible to reverse the axes and then have a single column of the 4 clusters? This would potentially make it easier to compare between clusters. If it makes it too cluttered, then this way is ok, just a little harder to compare across columns.

When reversing the x and y axes, the variance in Temperature is shown along the y-axis, which makes it difficult to compare the four clusters if they are arranged in a single column because they can no longer be projected along the same y-axis. One solution would be to arrange them in four distinct columns and the same y-axis, but we found that vertical density lines are not straightforward to read:



An alternative representation is to keep Temperature along the x-axis and density lines in the usual orientation, but arranging the four clusters in a single column (see proposed variant in the revised MS), as they now all share the same Temperature axis which facilitates comparison. The downside is this makes the figure footprint larger. In case you disapprove this change, we can revert back to the original figure, or to the rotated density lines shown above. Please note that the figure now represents daily TX instead of monthly TX (i.e., ~30 times more data included in the distribution curves) because this better depicts the actual variation within each month over the last 30 years, whereas the previous version plotted on montly TX failed to catch extreme days that would not reflect on monthly average despite their potential lethality for the PPM. This is the reason why tails in the distribution curves are now longer.

Figure 4: Is it possible to make the x-axis titles clearer, perhaps group under 1 – 4 and then have subtitles for time periods.

Done.

Figure 5: Please make the axis text larger

Done.

Philippe Louâpre

Among the various effects of climate change on life on earth, the modification of species ranges is probably one of the most integrative consequence one can observe at a large spatial and temporal scale. This is clearly a highly topical issue of major importance for scientific community with societal concerns. The other side of the coin is that revealing these complex and global effects is challenging as it requires handling often incomplete and fragmented data from long term studies. One alternative is to infer it from local effects of climatic conditions on key life history traits linked with population dynamic and distribution range.

I am especially sensitive to studies such as the one proposed here by Asma Bourougaaoui and colleagues. The authors combined two sets of data, from past and recent collections of the Pine Processionary Moth, Thaumetopoea pityocampa, in Tunisia, to explore potential relationship between thermal trends since 1990s and some key life history traits measured on clumped eggs, at the southern edge of the species range. They analyzed 30-years thermal data series across Tunisia and highlighted different climatic clusters within latitudinal and altitudinal discriminant situations. Overall, they suggest that thermal environments in Tunisia evolved through a shift toward warmer temperatures, especially in summer months, and sometimes overreached critical temperature, which negatively affects the population dynamic of T. pityocampa. This effect appears as small and statistically supported only for July, while confusing trends appears for June and August depending on the climate clusters considered.

I have no major concerns about the observations made and the statistics performed that would have mitigate the quality of the study. Rather, I found the study ambitious and well performed, with data rigorously analyzed. Overall, the manuscript is well written. The points I raised bellow are remarks and suggestions that would clarify some confusions and interrogations I had after reading the manuscript.

We are very grateful for the encouraging and relevant comments you made on the manuscript, and we hope our replies below will adequately answer your interrogations.

1- Thaumetopoea pityocampa provides a very interesting biological model to study the consequences of global warming on the expansion ranges, fully justifying its used in the present study. However, I would have appreciated the introduction to be set in a broader context giving the reader more conceptual elements about thermal biology and temperature-dependent mechanisms, thus linking individual responses to temperature to population dynamics and distribution range. I found the introduction too much focused on the biological system.

We broadened the beginning of the introduction by mentioning possible changes in traits due to climate warming, and we added examples of other temperature-induced distribution shifts, see lines 57-70.

2- Local thermal habitat is undoubtedly one of the key determinants of the spatial distribution of ectotherms. However, the manuscript gives the impression that biological trends highlighted here (hatching rate, sterility rate…) are caused by thermal events occurring in some specific areas, which may illustrate some of the well-known causal relationships between temperature and survival, growth rate, mortality… While these relationships are obviously true, it would have been more rigorous to look forward supplementary alternatives, including different environmental parameters (other than Temp) that may also explain the biological responses the authors highlighted here (depletion of food patches, exposure to natural enemies…). In what extend alternative hypothesis based on environmental cues other than temperature may explain the biological response the authors related in their study?

While not independent from temperature, we clarified in the text that in southern parts of the distribution range, intense solar radiation may also induce high embryonic mortality because females tend to lay their eggs on needles exposed to the sun (Démolin, 1969) (lines 644-647). Food availability rarely is a limiting factor in this defoliating species because it feeds on evergreen trees that are well distributed in the environment, from natural or semi-natural stands to urban areas where they often occur in relatively high numbers as ornamental trees. Natural enemies such as pathogens or predators (mostly insect parasitoids) at early larval stages have been suspected to snowball into increasing mortality during larval development because of the impact on the colony size and silk weaving effort to build a tent, however (1) there is no evidence that the enemy pressure would differ among areas investigated here and (2) temperature, particularly summer heat waves or early autumnal cold snaps, are often put forward as a major cause of early mortality (lines 647-656). This has been added to the manuscript as per suggested.

> 3- A significant part of the arguments provided in the manuscript is based on two temperatures (32 and 40°C) which have been suggested in 1969 by Huchon and Démolin as critical for T. pityocampa development and survival. Do you have updated information about the thermal curve and thermal range – at least for what it is known for this species or for populations in Tunisia? How do you evaluate the pertinency of using such pivotal thermal thresholds in your study, especially in relation with what it is discussed in the manuscript (thermal adaptation and thermal plasticity of individuals from populations very distant from each other)?

The two temperatures 32 and 40°C have frequently used in the literature as pivotal for range limits of the PPM. Those thresholds have not been experimentally determined but deduced from observations of consistent occurrence and/or density crashes in nature following meteorological events and/or among different bioclimatic areas. However, the upper thermal threshold is known to be higher, although it always depends on the duration of exposure, and to our knowledge no studies have investigated the duration component so far.

Recent studies showed that eggs from a French population were able to withstand a short transient exposure to a daily maximal temperature of 40°C during several consecutive days with no mortality impact (Robinet et al., 2013), while eggs from a nearby population could survive a single 6-hour-long exposure to up to 44°C (Poitou 2021). Meanwhile, mortality appeared on egg masses from a Tunisian population at 42°C after only three days of daily 4-hour exposure (Rocha et al., 2017), and steep survival drops were observed by Asma Bourougaaoui in Tunisian eggs exposed for 6 hours to 44°C (manuscript in prep.). While slightly different methods have been used and make it difficult to directly compare populations, those results show congruent upper limits for eggs among populations situated along a latitudinal gradient.

In larvae, the survivorship of L1 and L2 from Portuguese populations started to drop after 4-hour exposures to 36°C and 40°C, respectively (Santos et al., 2011), showing higher susceptibility in early larvae compared to eggs. In the French population, Poitou (2021) found *CTmax* values of 51-52°C in L1-L4 instars, and as high as 55°C in field-sampled L3. Although the *CTmax* is a proxy of thermal tolerance that hardly translates into survival to such temperatures in nature, those values are high compared to most temperate insects, and very similar to those found among Tunisian populations by Asma Bourougaaoui (51-56°C, manuscript in prep.). Finally, in Poitou et al. (2022), we determined thermal performance curves in development rate in the first four larval instars in the same French population, and showed that performance linearly increases up to a condition of 24°C on average, which in that fluctuating thermal regimes experiment corresponded to 6-hour daily exposures to a Tmax phase of 29°C, and started decreasing between that condition and the closest warmer condition of 30°C on average (6-hour daily exposures to Tmax = 35°C). Démolin's indicative 32°C estimate is comprised within this decreasing performance phase determined experimentally.

The compilation of those various results shows the rather complex thermal response of the PPM to temperature depending on instar and on whether stresses are acute or chronic, but also suggests that those stress parameters prevail, with population differences only being a secondary source of variance. The 32°C and 40°C thresholds proposed by Huchon and Démolin (1970) may appear as conservative considering the high tolerance found to short acute stresses, but those coarse estimates are only meant to be used as indicators of challenging conditions instead of exact survival thresholds. Being deduced from natural observations, they have the benefit of being good integrators of whether a population faces stressful conditions regardless of the exact duration of exposure and are not as sensitive as thresholds determined experimentally in very specific artificial conditions.

The above details are a synthesis of both published and unpublished data, therefore a shorter variation of this text focused on published data was added to the manuscript to better explain the use of those thresholds (lines 492-511).

4- The absence of hatching in samples collected in 2017 in cluster 1 is discussed in the light of a particularly hot summer that may have effected embryo development. I agree with the other plausible explanations, especially with a detrimental effect on reproductive physiology. I would suggest broadening the discussion as excessive high temperature may also influence negatively reproductive behaviors, independently of gamete production, maturity, and viability, which is well known in Lepidopteran species. My concern here is more about the correlation made – which sounds as obvious – between extreme temperature occurring during summer 2017 and high sterility rate. I apology if I missed something but Figure 5 did not reveal any specific climatic events or trends in 2017 that may support this hypothesis for the summer months, especially in cluster #1. Temperature reached higher values in the past, but data reported non-zero values for hatching rate.

Hot temperatures can indeed induce behavioural changes in Lepidopteran species, including alterations in important fitness-related behaviours, namely mating. We broadened the discussion to mention how thermal stress interferes with chemical communication between genders and may alter the quality of gametes (see lines 439-441).

The extreme days of August 2017 in cluster 1 did not show clearly in the figures we initially provided because figures like Fig. 5 (now Fig. 4, but Fig. 3 as well) illustrated a broader time scale (i.e., monthly means over years). However, the lethality of extreme heats on PPM eggs is more readily related to absolute daily values rather than trends consistent enough to shift monthly averages between years. Fig. 3 now shows daily TX instead of monthly TX, and better documents extreme days (distribution tails) over the last 30 years. Additionally, in Fig. SM4 and the right panel corresponding to Kasserine (the station closest to the population where hatching rate dropped to 0 in 2017), the number of days warmer than 40 °C was hard to see because the value is relatively low in this location compared to others (Gafsa for instance), and was hidden below the blue layer depicting the number of cold days. The number of cold days is now shown underneath the number of days reaching 40 °C, thereby making it easier to see the most relevant variable for this study. It is important to note that the number of days above this threshold in this figure is considered per year and not per month. This is critical because further investigation shows that, in 2017, most of these days occurred in August (see Fig. 6 and Fig. SM8), when eggs are already laid, while there has been other years in 1990-2019 with a similar yearly total number of days above 40 °C in Kasserine, but mostly occurring in July when eggs are not laid yet in that area. See also the detailed answer to Élodie Vercken above.

5- Parasitism rate was estimated on both the number of emerged parasitoids and dead ones inside the eggs. In insects, especially in lepidopterans, individuals, even at egg stage, can defense toward parasitoid eggs through various immune defense mechanisms. There is a growing amount of evidence that such immune response to parasitoid attack is temperature dependent. This remark does not attenuate the quality of the measure and the interpretation made, but I suggest the results on parasitism rate to be discussed in the light of the thermal-modulated response of immunity in insects

We fully agree with the reviewer and mentioned some of the immune defence mechanisms in the discussion to cover the possible effects of temperature changes on immune responses to parasitoids (lines 604-620).

6- The climate data appears to be rigorously handled and the strong correlation between data series from the Tunisian Institut National de Météorologie and the NASA PWER is reassuring. I wonder whether it would not have been a better way to use exclusively satellite measurements of daily temp in order to reduce undesired source of variance, and to have a better proxy of temp occurring at the collecting sites. By clustering daily temp in sites sometimes more than 80 kilometers away may have led to inaccurate association between local climates and populations (e.g. El Ayoun and Jebel Sidi Aich are associated to the same climatic data recorded at Kasserine meteorological station while these two sites are presumably exposed to different climatic conditions based on a north-south gradient).

We mistakenly mentioned a wrong scale for NASA data in the text, see above answer to Matt Hill. For this reason, using exclusively satellite measurements of a lower accuracy may increase the risk of failure to detect deviations to actual conditions eggs are exposed to.

José Hodar

This article analyses a process of great interest, which is the retraction of the equatorial border of the distribution of a circum-Mediterranean pest species, the pine processionary moth, through its decrease in fertility as a consequence of the increase in temperature. To do this, a long (but heterogeneous) series of PPM egg batches sampled over the last thirty years is used. The work is well done, the data well analysed, and the manuscript well written, but I think there are some issues that the authors should try to correct or acknowledge before attempting to publish it.

Thank you for the positive and relevant remarks you made on our work, as well as for your appreciation of the importance of long data series, despite the difficulties to associated to analysing such heterogeneous data. We tried below to address some of the issues you pointed and to better acknowledge the limitations we couldn't fully solve.

The first is the labelling of the clusters, the groups made with the locations and years sampled. The sample is quite heterogeneous, and to give greater coherence to the analysis the authors use a PCA procedure to group samples into four large clusters, labelled 1-4. This ranking is very unintuitive, and hampers the correct interpretation of the parameters studied. I think the 1-2-3-4 clusters should be labelled as 3-4-2-1, respectively. Ranked in this way, the samples sparse approximately from NE to SW, marking a coast-inland gradient. Thus, the samples in the new cluster 1, coastal, would have less extreme and less variable temperatures, and the heat and variability would increase towards the new 4. This is a very easy change to make without a substantial change to the manuscript, but for the reader unfamiliar with the area it can be an important help.

Reordering clusters would be easy to change in the manuscript and figures, but the current numbering has a statistical meaning: it sorts clusters in their order of appearance in the PAM clustering method, meaning that cluster 1 is the first group of individual points to emerge and 4 is the last to emerge when seeking 4 contrasting groups. Renaming clusters may thus prove confusing from a mathematical perspective and induce misinterpretations in the most important contrasts among clusters.

The wide heterogeneity of the set of samples gives the advantage of covering a very wide area for a long time, but it also has the disadvantage of having the samples very unbalanced between locations and years, which can potentially skew the results. I assume that the samples are what there are and that they cannot be ignored. But I think that, by handling the analysis, some of the problems of this heterogeneity can be solved. My suggestion would be merging the coastal clusters (3 and 4 according to the authors' nomenclature). I suggest this change because these clusters have only 3 and 2 locations respectively and, in particular, samples of cluster 4 came from 1995 only. Certainly climatic conditions at the stations of Carthage (cluster 3) and Kelibia (cluster 4) are quite contrasted (in fact, I wonder whether Carthage data will not be affected by the "heat island" phenomenon due to the city of Tunis), but it would be interesting to do the test. Also, data from Thélepte 2017 should be removed from the joint analysis, although keeping data in tables and graphs. The information of this point is very interesting, since it shows the effect of an extreme event, which is a heat wave, on PPM survival. But from 2014 onwards, cluster 1 only has data from Thélepte 2017 and 2019, so I think that the effect of mortality suffered in 2017 alters the analysis. The authors state that "hatching rate tended to decrease between the 1990s and the 2010s in cluster 1" (L343-344). Looking at the Fig. 6 this could be true, but the tendency is surely less apparent if this outlier is removed.

This is a valid concern but a posteriori regrouping clusters goes against our choice of an objectified approach. The clustering analysis could be forcibly restricted to a maximum of K=3 clusters only, which would indeed group Kélibia and Carthage together, but the climate contrast that appeared between those two relatively nearby stations (Fig. SM4), particularly in summer, suggested that forcing three clusters would introduce an important bias in the comparisons to be done. Hypothetically, those two stations are considered close to each other in a K=3 analysis because of their relatively low seasonality between summer and other seasons compared to other areas where the contrast is of higher magnitude (see Fig. SM6), but the absolute differences in summer averages and number of days above stress thresholds are marked (see namely Fig. SM4). Neglecting these average heat differences would allow for a more balanced data set in further statistical analyses, but would neglect one of the factors that we assumed as among the most prominent for egg survival: absolute temperatures. Our decision was therefore to keep the four groups, whether they are coastal or not, assuming that geographic features with a sensible effect would reflect on the climate data. Else, the merging of locations is supported by climate consistency in an objective and reproducible decision criterion, which was our target. Likewise, in clusters represented by only three or two sites, sampling size amounts for tens of egg masses from multiple trees scattered within each site. Therefore if multiple sites belong to a common cluster, then egg masses can reasonably be considered replicates of a common climate group (although we agree that testing the site effect as a random effect would have been best, but unfortunately residuals distribution after GLMM modelling were not satisfactory to ensure correct interpretations, and using Quade's non-parametric analysis proved to be more reliable).

That being said, we agree that other methodological choices than this one could be done and equally defended. There are downsides in each of them and this is the reason why we would rather stick to the original intent, given that altering the clusters would cascade into altering every results and figures that have in general been positively evaluated during this round of reviews, and the other limitations for this different choice would also have to be discussed. Finally, we now describe much more accurately in the results and discussion the trends that are carried by single sampling sites like Thélepte in 2017, and we completed the analysis by another one without these data to infer on the weight of these samples on the general conclusions (see above answer to Élodie Vercken). This should now present results, associated conclusions and whether they can be generalized in a more objective way.

> Given that the main interest of the manuscript is to verify how the phenotypic estimators of the clutches varied over time, I wonder if the database would support a GLMM-type analysis, using year as a predictor variable. I think that trends would be better reflected, as well as possible differences in trend between clusters, especially if clusters are reduced from 4 to 3. This would also allow eliminating figures such as 6, and developing graphs in which the year was the variable x and each of the phenotypic variables was on the y-axis, showing the trend of the model over time.

Unfortunately GLMMs did not yield satisfactory prediction of the data, which showed on uneven residuals distributions and ultimately led us to using a non-parametric approach. See the end of the first paragraph in the above answer, as well as the reply to Élodie Vercken about this (point 4). The Fig. 6 (now Fig. 5) has also been restructured. The unbalanced nature and the insufficient accuracy of the model predictions did not allow plotting predicted curves over time, but the figure is now arranged in a way that makes it easier to see the overall trend in clusters sampled during both periods, or among clusters within individual periods.

> I'm surprised about the lack of information regarding the identity of the parasitoids. A recent analysis on a similar gradient (altitude instead of continentality), Hódar et al. 2021 Ecosphere 10.1002/ecs2.3476, shows how the behaviour of the two main parasitoid species is quite different along the gradient. Although the focus of the article is PPM, the identity of the

> responsible for a significant loss in the initial number of larvae in egg batches could help to understand the patterns. Hódar et al. also show that the rate of parasitism decreases as temperature decreases with altitude, reaching the highest rates in coastal localities, thus a pattern quite different from that found by the authors in Tunisia. Furthermore, when comparing the parasitism ratios in Tunisia with European references, I think it would be better referring to figures from SE Spain, much more similar in climate and landscape to Tunisia than France or Bulgaria.

Absolutely, identifying parasitoids would have been interesting and could have allowed extra analyses on the causes of mortality. Egg masses were sampled too late in recent samples to observe all parasitoid emergences, and although a posteriori identification can be done to some extent using exit holes and meconia, the data was not necessarily available for past samples. That is why we only showed here the overall contribution of parasitism to natural mortality of egg PPM. We completed the manuscript by adding references to the example from SE Spain in the comparison (see line 632-633).

> Overall, I believe the dataset is very interesting, but I think it can be presented in a better way with some changes. I hope my comments contribute to improve the manuscript.

We are very grateful for the constructive comments you made about this study. We understand that the primary change you suggested in this review was not implemented in the revised manuscript, but we hope you will agree with us that consequences of the change would imply other limitations to the approach which could be equally criticized, due in part to the nature of historic data series (and inherent missing data) that are often hard to adequately balance for factors to be tested.