nature portfolio

Peer Review File

Mean daily temperatures predict the thermal limits of malaria transmission better than hourly rate summation

Corresponding Author: Dr Courtney Murdock

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Shocket et al. present compelling evidence that daily temperature fluctuations have a significant impact on mosquito life-history traits, but that nonlinear averaging -to quantitatively model the effect of temperature fluctuations on traits- often fail to adequately predict performance under fluctuating conditions. They performed elegant experiments, considered several models, and created thermal suitability models to prove their point. It is a very well-written manuscript, and I could not find a flaw in their study design, models or conclusions.

One thing I would like to authors to consider however, is a discussion about future avenues where rate summation may work. First, one could work with a lethal temperature in addition to CTmax, which will (at least) ensure the rate summation predictions at higher DTRs will stop at temperatures below Topt. Second, the authors mention others have worked with TCPs that are not truncated below zero (Eppley curve), which could be an option worth exploring in temperature-mosquito interactions. Third, the authors mention that thermal stress and the accumulation and subsequent repair of cell damage occur near CTmin and CTmax. Could you build in a time lag in future models (i.e. development does not immediate continue when temperatures are suitable again, but starts e.g. 30min later). These were just some things I was wondering when I read your manuscript.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Review of “Mean daily temperatures can predict the thermal limits of malaria
transmission better than rate summation” by Shocket et al.

Understanding how thermal performance curves (TPCs) for pathogen and vector traits are shaped by realistic temperature variations is increasingly important for assessing the effects of environmental temperature on vector-pathogen interactions and disease transmission. Since properly generating experimental data for various fluctuating temperature regimes is impractical, evaluating the accuracy of theoretical predictions using the 'rate summation' method is crucial. In this study, the authors present evidence suggesting that rate summation may not work for certain mosquito traits. While this is a timely contribution, the following issues should be addressed before publication:

1. Limited temperature range of experimental data – The authors compare TPC parameters such as Tmin, Topt, Tmax, and Tbreadth (Figures 2, 3, and Table 1) to evaluate how well rate summation predicts the experimental data and TPCs under fluctuating temperature regimes. However, this method can be problematic when the experimental temperature range is not broad enough, as it reduces the accuracy of Tmin and Tmax extrapolations (and consequently Tbreadth) and the overall TPC estimates. Currently, the data range is limited to 16°C–32°C, while predicted Tmin and Tmax are often far beyond this range for the three mosquito traits (Figure 2). This is particularly important given the increased non-linearity in thermal performance at extreme temperatures, which is commonly observed in insects and other systems. Without data points at temperature extremes, along with properly spaced intervals, the predicted TPC parameters may be unreliable, complicating

the interpretation of the results. Preliminary experiments to identify temperatures where performance values reach zero could better inform the experimental design to ensure that the full thermal breadth is adequately covered. Alternatively, the authors could focus on the existing data range to discuss the accuracy of rate summation.

2. Unmatched data points – A related concern is the inclusion of data points at 36°C for constant temperature regimes, while the highest temperature used in the fluctuating regimes was only 32°C. The use of non-matching data points is not justified in the manuscript, and their inclusion is concerning as it may affect the overall TPC fitting and particularly Tmax, which appears higher in the constant temperature regime especially for bite rate and lifespan, possibly due to the inclusion of the 36°C data points (Figure 2A and 2B). I suggest removing these data points from all analyses, ensuring a proper assessment of the accuracy of rate summation for each trait along with predicted thermal suitability for malaria transmission.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
This study assesses whether rate summation can be used to predict the effects of diurnal temperature variation on the biting rate, longevity and fecundity of the malaria vector Anopheles stephensi in laboratory experiments. The authors present novel empirical evidence for the effects of diurnal temperature variation on these traits from laboratory studies and argue that rate summation does not accurately predict the effects of diurnally fluctuating temperatures, which are both noteworthy results. The findings are interesting and relevant to understanding the role of temperature on transmission.

The discussion around the mechanisms driving the effects of diurnal temperature variation on trait performance in numerous species was very detailed and interesting.

Comments

• Thermal performance curve fits:
1. Firstly, the fitted Brière curve biting rate Tmax seems to decline because of a lack of data and your prior choice. I am, therefore, not convinced the uncertainty is fully captured. I understand this is because when held at the constant temperatures mosquitoes are dying at these temperatures, but this does not necessarily mean the biting rate would be zero at these temperatures if the mosquito was alive. This could be important because the constant temperature TPCs are used in the rate summation where the temperatures do temporarily go above these temperatures, but the mosquito does not necessarily die – for example at the mean of 32 with DTR the mosquitoes will experience those high temperatures but some survive (Figure 2A). I don't think this limitation is acknowledged in the paper.
2. Secondly, for the biting rate TPC a truncated normal likelihood was used in the fitting. Why were the numbers of bites (or gonotrophic cycles) in a mosquitoes life not modelled with a Poisson distribution offset by the time, or geometric distribution for the gonotrophic cycle lengths of individual mosquitoes and then the biting rate estimated from this?
3. In the supplementary (trait TPC model specifications section) "i" corresponds to each individual level observation, I'm guessing this refers to the temperature because of ui, so am I correct in thinking the TPC fits do not account for uncertainty in the numbers of mosquito sampled at each temperature in the TPC fits?
4. For the priors a uniform distribution was used, meaning outside the minimum and upper bounds the posterior is zero – the values seem reasonable (except maybe Tmax), but I think it should be made clear that other parameter values are not possible.
5. Similarly, there are no data for lifespan when it has declined at the lower limits, (Figure 2B) meaning the uncertainty in Tmin for the lifespan might not be fully captured.

• Figure 3: how was goodness of fit between constant and rate summation approaches to the empirical fit compared? Some of the differences do not seem very clear, could they just be due to sampling or noise in the system?

• Currently, the rate summation value is predicted using the mean temperatures over each hour (lines 557 – 569). I'm guessing the temperatures in the incubators varied between the two temperatures between the hours? In which case, how sensitive is the estimated value to this choice of difference in time when calculating the rate summation? Rather than the results being due to the lagged effects of temperature or other unknown variables, are they just an artefact of this (potentially) arbitrary choice to use an hour?

• Five versions of the S(T) model are calculated, including (4) rate summation applied to all traits and (5) rate summation applied to the constant temperature S(T). How the level of rate summation impacts the S(T) estimates (model 4 vs model 5) is tested. I don't really understand what value this comparison adds, could it not be demonstrated mathematically that the rate summation of the S(T) parameters is not the equivalent to the rate summation of S(T)?

• Figure 6: the temperature-dependent R0 is scaled, so am I correct in thinking that S(T) > 0.001 or 0.5 is just the S(T) value relative to the minimum value? If so, I don't understand how this translates to months when "transmission is possible". Theoretically, whether transmission is possible depends on the minimum R0 value used in the scaling. I think this figure could therefore be easily misinterpreted.

• Line 319: should this be "thermal suitability" instead of environmental suitability?

• Lines 31 – 33 and 326 – 328. There is no assessment of the rate summation approach against transmission intensity or known transmission seasonality in the field, so I don't understand how the constant temperatures are more accurate for predicting and mapping the thermal limits of transmission? Indeed, laboratory strains of mosquito are used, but there is evidence of local adaptation of mosquitoes.

• With regards to the established literature, previous work on the effects of temperature on malaria vector and parasite traits in the laboratory (mostly published by the authors) is discussed in detail. The significance of these studies and this research to understand transmission in the field is noted (lines 68 -75, lines 402 - 404) and the results from this paper mapped (Figure 6). There has, however, also been lots of debate into the effects of temperature on malaria transmission using empirical malaria prevalence or incidence data. I think it is, therefore, necessary to either include this research to give a broader overview of the scientific consensus or change the narrative accordingly?

(Remarks on code availability)
Code and data are available on GitHub. The code provides a README file and details of package versions used. An R project was used making it easier to run the code.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
All my concerns have been addressed. Great job.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Thank you for responding to my points - I believe these have all been addressed.

(Remarks on code availability)

Dear Nature Communications Editorial Team,

We thank all three reviewers for their close reading and thoughtful comments, which have improved the manuscript. Please find a detailed description of our changes and our responses to each point that they raised below.

New text and substantially revised text in the revised manuscript is highlighted in yellow. We also lightly edited the abstract to improve clarity.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Shocket et al. present compelling evidence that daily temperature fluctuations have a significant impact on mosquito life-history traits, but that nonlinear averaging - to quantitatively model the effect of temperature fluctuations on traits - often fails to adequately predict performance under fluctuating conditions. They performed elegant experiments, considered several models, and created thermal suitability models to prove their point. It is a very well-written manuscript, and I could not find a flaw in their study design, models or conclusions.

We are glad the reviewer felt positively about our study design and manuscript.

One thing I would like the authors to consider however, is a discussion about future avenues where rate summation may work. First, one could work with a lethal temperature in addition to CTmax, which will (at least) ensure the rate summation predictions at higher DTRs will stop at temperatures below Topt. Second, the authors mention others have worked with TCPs that are not truncated below zero (Eppley curve), which could be an option worth exploring in temperature-mosquito interactions. Third, the authors mention that thermal stress and the accumulation and subsequent repair of cell damage occur near CTmin and CTmax. Could you build in a time lag in future models (i.e. development does not immediately continue when temperatures are suitable again, but starts e.g. 30 min later). These were just some things I was wondering when I read your manuscript.

We agree wholeheartedly with the reviewer on each of these points, and in fact pursued some related analyses during an earlier stage of the project. However, we ultimately decided not to include them in this paper because we felt this was beyond the scope of an already complex and long analysis.

We appreciate the reviewer's suggestion that making these points more explicitly in the manuscript would be helpful for readers.To do so concisely, we have added the following concluding sentence to the paragraph where we discuss the various factors that could affect the accuracy of rate summation (lines 473-5):

> "Accordingly, although rate summation was not accurate for mosquitoes here – using the most basic and common approach – it is possible that more complex approaches (e.g., incorporating threshold penalties, lags, or other effects) could potentially improve its performance."

Reviewer #2 (Remarks to the Author):

Review of "Mean daily temperatures can predict the thermal limits of malaria transmission better than rate summation" by Shocket et al.

Understanding how thermal performance curves (TPCs) for pathogen and vector traits are shaped by realistic temperature variations is increasingly important for assessing the effects of environmental temperature on vector-pathogen interactions and disease transmission. Since properly generating experimental data for various fluctuating temperature regimes is impractical, evaluating the accuracy of theoretical predictions using the 'rate summation' method is crucial. In this study, the authors present evidence suggesting that rate summation may not work for certain mosquito traits. While this is a timely contribution, the following issues should be addressed before publication:

1. Limited temperature range of experimental data – The authors compare TPC parameters such as Tmin, Topt, Tmax, and Tbreadth (Figures 2, 3, and Table 1) to evaluate how well rate summation predicts the experimental data and TPCs under fluctuating temperature regimes. However, this method can be problematic when the experimental temperature range is not broad enough, as it reduces the accuracy of Tmin and Tmax extrapolations (and consequently Tbreadth) and the overall TPC estimates. Currently, the data range is limited to 16°C–32°C, while predicted Tmin and Tmax are often far beyond this range for the three mosquito traits (Figure 2). This is particularly important given the increased non-linearity in thermal performance at extreme temperatures, which is commonly observed in insects and other systems. Without data points at temperature extremes, along with properly spaced intervals, the predicted TPC parameters may be unreliable, complicating the interpretation of the results. Preliminary experiments to identify temperatures where performance values reach zero could better inform the experimental design to ensure that the full thermal breadth is adequately covered. Alternatively, the authors could focus on the existing data range to discuss the accuracy of rate summation.

We recognise that our experiments did not empirically determine the critical Tmin and Tmax (which are typically measured with a different type of experiment, e.g. knock-down assays) and instead inferred them from somewhat limited data. However, we believe that our data are suitable for asking our specific questions regarding the accuracy of rate summation for the following reasons:

First, this experiment was substantially larger than most experiments measuring temperature-dependent traits in mosquitoes. Thus, the thermal performance curves presented here are some of the best parameterized to date. Rate summation is routinely applied to less well informed thermal performance curves, and we think this is a good opportunity to discuss the potential limitations of rate summation predictions of performance near the estimated thermal limits. Additionally, we experimentally determined that the incubators simulating fluctuating conditions are not reliable at temperatures beyond those used in our hottest DTR 12 treatment.

Second, while it's true that our specific estimates of Tmin and Tmax are not directly measured or validated, Tmax is informed by real, observed decreases in performance at

warmer temperatures (see comment below re: bite rate). While the experimental data are less clear at the cooler end of the temperature gradient, our suitability and mapping results related to Tmin do not address differences between our empirical models (we did not find any), but rather the effects of rate summation itself on these types of TPCs more generically. Accordingly, these results do not depend on our particular empirical data set used to fit the TPCs.

Nonetheless, based on the comments from reviewers 2 and 3, we appreciate that a more explicit discussion of the limitations of our data and their impact on our conclusions is warranted.

We have added the following paragraph to the end of the Discussion to address comments about data limitations from both Reviewer #2 and Reviewer #3 (lines 476-97):

> "The trait data used to parameterise our TPCs were limited in several important ways. Although our experiments were larger than most comparable studies, we did not directly measure Tmin and Tmax for each trait so these parameters should be considered estimates only. This could have several implications. First, we are lacking data at cooler temperatures to inform Tmin, particularly for lifespan, and it is possible there are significant differences between the empirical treatments that we were unable to detect. Second, Tmax for daily biting rate is largely inferred because we did not observe decreases at high temperatures. However, because the Tmin of lifespan was lower than for most other traits, and the Tmax for the daily bite rate exceeded the Tmax for lifespan (which is better informed by our data), these less reliable parameters do not affect the predicted Tmin and Tmax for overall suitability. Due to incubator constraints we included a 36ºC treatment for the constant temperature treatments but not the fluctuating treatments (see Materials and Methods). Additionally, our data were collected using a lab-adapted strain of mosquitoes whose thermal responses may differ from natural populations of An. stephenesi. More work is needed to definitively demonstrate how thermal performance in the lab is different than the field, because the evidence for mosquito systems is mixed (Lyons et al. 2012; Dennington et al. 2024). Finally, we did not attempt to validate the predictions from our various models with field data of human cases or entomological inoculation rates (EIR). Instead, we judged the "accuracy" of the rate summation models (models 3, 4, and 5) and the constant temperature model (model 1) by comparing them both to the empirically-fit fluctuating temperature model (model 2), which we assume most accurately reflects the impacts of natural temperature variation. Future work could use disease incidence or EIR data to better evaluate the accuracy of these models and to determine if incorporating fluctuating temperatures improves their performance."

2. Unmatched data points – A related concern is the inclusion of data points at 36°C for constant temperature regimes, while the highest temperature used in the fluctuating regimes was only 32°C. The use of non-matching data points is not justified in the manuscript, and their inclusion is concerning as it may affect the overall TPC fitting and particularly Tmax, which appears higher in the constant temperature regime especially for bite rate and lifespan, possibly due to the inclusion of the 36°C data points (Figure 2A and 2B). I suggest removing these data points from all analyses, ensuring a proper assessment of the accuracy

of rate summation for each trait along with predicted thermal suitability for malaria transmission.

We agree that the unmatched 36ºC treatment is another potential limitation of our data.

We included a 36ºC treatment for the constant temperature treatments but not the fluctuating treatments because 1) fluctuations around 36ºC go beyond the range in which our incubators can reliably regulate their temperature, and 2) we felt it was more important to get the best and most accurate possible fit for the constant temperature TPCs (which are used for the rate summation calculations) than to have a completely matched design.

We have added the following sentence to the Methods section (lines 549-51):

> We were unable to include fluctuating treatments around 36ºC because the incubators did not reliably regulate their temperature above 40°C.

This point is also mentioned in our new Discussion paragraph on data limitations (lines 485-7).

Reviewer #3 (Remarks to the Author):

This study assesses whether rate summation can be used to predict the effects of diurnal temperature variation on the biting rate, longevity and fecundity of the malaria vector Anopheles stephensi in laboratory experiments. The authors present novel empirical evidence for the effects of diurnal temperature variation on these traits from laboratory studies and argue that rate summation does not accurately predict the effects of diurnally fluctuating temperatures, which are both noteworthy results. The findings are interesting and relevant to understanding the role of temperature on transmission.

The discussion around the mechanisms driving the effects of diurnal temperature variation on trait performance in numerous species was very detailed and interesting.

Comments

• *Thermal performance curve fits:*

1. Firstly, the fitted Brière curve biting rate Tmax seems to decline because of a lack of data and your prior choice. I am, therefore, not convinced the uncertainty is fully captured. I understand this is because when held at the constant temperatures mosquitoes are dying at these temperatures, but this does not necessarily mean the biting rate would be zero at these temperatures if the mosquito was alive. This could be important because the constant temperature TPCs are used in the rate summation where the temperatures do temporarily go above these temperatures, but the mosquito does not necessarily die – for example at the mean of 32 with DTR the mosquitoes will experience those high temperatures but some survive (Figure 2A). I don't think this limitation is acknowledged in the paper.

We agree with this point, and have included it in our new paragraph discussing data limitations (see response to the first comment from Reviewer #2 above). Additionally, we

want to point out that the predictions for the Tmax of bite rate exceeds the Tmax for lifespan, which is based on more informative data (and obviously dead mosquitoes are unable to bite).

2. Secondly, for the biting rate TPC a truncated normal likelihood was used in the fitting. Why were the numbers of bites (or gonotrophic cycles) in a mosquitoes life not modelled with a Poisson distribution offset by the time, or geometric distribution for the gonotrophic cycle lengths of individual mosquitoes and then the biting rate estimated from this?

For our study we measured the actual biting rate: blood meals were offered daily, any bites were noted, and thus our data are not confounded with gonotrophic cycle duration. Indeed, the mosquitoes did take blood meals more frequently than dictated by their gonotrophic cycle.

We have added text to the methods section as excerpted below to clarify (lines 561-4):

> Each day until found dead, individuals were provided with a whole human blood meal for 15 minutes and inspected visually for imbibed blood to calculate the bite rate. This directly-measured bite rate was to determine if mosquitoes were biting more frequently than what would be inferred by the assumption of one bite per gonotrophic cycle.

3. In the supplementary (trait TPC model specifications section) "i" corresponds to each individual level observation, I'm guessing this refers to the temperature because of ui, so am I correct in thinking the TPC fits do not account for uncertainty in the numbers of mosquito sampled at each temperature in the TPC fits?

It is correct that in the model files, "i" represents the temperature and estimated parameter values for each individual-level observation. It is also correct that the model does not include any uncertainty in the number of mosquitoes sampled at each temperature, as that is fixed by the data, and has a very small (< 2%) percent error across temperature treatments (59-61 individuals, except for 28C constant, which had one extra block and 90 total individuals).

4. For the priors a uniform distribution was used, meaning outside the minimum and upper bounds the posterior is zero – the values seem reasonable (except maybe Tmax), but I think it should be made clear that other parameter values are not possible.

We have revised the Methods so the text now reads (lines 596-9):

> We assumed low-information uniform priors (Tmin ~ uniform (0, 20), Tmax ~ uniform (28, 45), c ~ uniform (0, 10), variance ~ uniform (0,1000), rate ~ uniform (1,100), r ~ uniform (1,100)) that restricted the range of parameters to biologically or statistically meaningful values (i.e., values outside of those ranges are not possible).

5. Similarly, there are no data for lifespan when it has declined at the lower limits, (Figure 2B) meaning the uncertainty in Tmin for the lifespan might not be fully captured.

We agree with this point and have included it in our new paragraph discussing data limitations (see response to the first comment from Reviewer #2 above). Additionally, we want to note that Tmin for this trait is very low. Thus this estimate is conservative and not does not play a role in constraining suitability, which is instead limited by other traits.

• *Figure 3: how was goodness of fit between constant and rate summation approaches to the empirical fit compared? Some of the differences do not seem very clear, could they just be due to sampling or noise in the system?*

Based on the 95% CIs (traits: Table S3; suitability Table S4), we have very strong statistical support that the major differences in critical TPC parameters (Tmax, Topt) are not just due to sampling or noise.

Additionally, we did not include goodness-of-fit measures comparing the constant and rate summation TPCs to model 2, because we felt like it would confuse readers by undermining our overarching point: there is no single "best model", but rather the best model depends on the specific research question and goal, and more specifically, what part of the temperature gradient is most relevant:

- If you care more about accurately estimating thermal limits, the constant temperature model is better because rate summation systematically distorts them.
- If you care more about the predicted absolute magnitude of transmission in the middle of the thermal gradient, then the rate summation model is better because it partially accounts for reduced performance in those temperatures.

We particularly appreciate this comment, because it made us re-evaluate our messaging throughout the manuscript, and have revised the text in several places to emphasize our research goals and this context-dependence re: when rate-summation performs better or worse than constant temperature TPCs:

- Introduction, lines 133-135: "this discrepancy particularly impacts the predicted thermal limits for transmission, which are often used for mapping applications (Ryan et al. 2015; Ryan et al. 2017; Ryan et al. 2021)."
- Results, lines 205-8: "As expected, rate summation always predicted that fluctuations would reduce performance in the middle of the temperature gradient, which was observed for two of the three traits, bite rate (a) and lifetime egg production (B) (Figure 3). However, for bite rate (a) rate summation only captured a small proportion of the observed decrease (Figure 3)."
- Results, lines 245-7: "Overall, compared to the original TPCs fit to data from constant temperatures, the TPCs predicted by rate summation were generally more accurate in the middle of the temperature gradient but less accurate near the upper thermal limits (Figure 3)."
- Discussion, lines 340-2: "Although RS did correctly predict decreased performance over much of the temperature gradient for some traits, it underestimated the magnitude of this effect and also incorrectly predicted increases in upper thermal limits (Tmax)."
- Discussion, lines 346-52: "This result stems from a general property of performing rate summation: thermal limits are distorted if TPCs cut-off at the x-axis, as is often the case for biological traits that cannot take negative values. Thus, we conclude that

while daily-scale temperature fluctuations have important impacts on organismal performance, for some specific applications (i.e. those focusing on thermal limits) it may be better to use thermal responses that are fit under constant temperature environments than to try to incorporate the impact of fluctuating temperatures using non-linear averaging." (new text underlined)

- Discussion, lines 514-20: "Our thermal suitability model based on data from constant temperatures was more accurate for mapping the thermal limits for malaria transmission than the model parameterized via rate summation, although the rate summation model did partially capture reduced transmission in the middle of the temperature gradient. Thus, for applications focusing on thermal limits it may be better to simply use thermal responses fit under constant temperature environments, while for applications focusing on transmission intensity in endemic areas it may be better to incorporate the impact of fluctuating temperatures using non-linear averaging."

• *Currently, the rate summation value is predicted using the mean temperatures over each hour (lines 557 – 569). I'm guessing the temperatures in the incubators varied between the two temperatures between the hours? In which case, how sensitive is the estimated value to this choice of difference in time when calculating the rate summation? Rather than the results being due to the lagged effects of temperature or other unknown variables, are they just an artefact of this (potentially) arbitrary choice to use an hour?*

The incubators gradually ramped up or down their temperatures between hourly set points. We know from preliminary work on another project (using a different data set) that performing rate summation at hourly or minute intervals does matter. Based on this comment, we have added the following text to the Discussion paragraph on factors affecting the accuracy in rate summation (lines 464-6):

> Third, we only performed rate summation calculations at one time scale (hourly). Averaging over different time scales may impact the results and could produce interactive effects with the time scale of thermal fluctuations.

• *Five versions of the S(T) model are calculated, including (4) rate summation applied to all traits and (5) rate summation applied to the constant temperature S(T). How the level of rate summation impacts the S(T) estimates (model 4 vs model 5) is tested. I don't really understand what value this comparison adds, could it not be demonstrated mathematically that the rate summation of the S(T) parameters is not the equivalent to the rate summation of S(T)?*

We chose this approach because even if one shows that they are not strictly mathematically equivalent, they still might be numerically equivalent (i.e., indistinguishable up to rounding errors) based on the amount of data, precision of measurement, range of temperatures, etc. Therefore, it's not just a question of strict mathematical equivalence, but whether the actual magnitudes are distinguishable. For that question, the easiest thing to do is compare them numerically. One could try to place bounds on measurement precision and how much data (i.e., statistical power) are needed in principle to be able to distinguish among models; however, showing it for a particular case still requires directly inputting empirical values, using an approach much like the one we used here.

• *Figure 6: the temperature-dependent R0 is scaled, so am I correct in thinking that S(T) > 0.001 or 0.5 is just the S(T) value relative to the minimum value? If so, I don't understand how this translates to months when "transmission is possible". Theoretically, whether transmission is possible depends on the minimum R0 value used in the scaling. I think this figure could therefore be easily misinterpreted.*

We apologise for a (now corrected) typo in the mapping methods that may have made our approach more confusing: 97.5% should have been 2.5%. We have also revised the text to hopefully make it clearer.

To address the specific questions here, for mapping we took the 2.5% lower CI contour for each model and scaled it by the maximum value of this contour so that it varied between 0 and 1. This means that S(T) is still relative to a maximum level of thermal suitability (= 1), but it has relatively narrow values for Tmin and Tmax compared to if we had used the median instead of the 2.5% CI contour. We chose this approach to match methods from previous papers and to be more conservative in our predictions and to minimize type 1 error (inclusion of unsuitable areas as suitable).

In this paper and previous papers, the phrase "transmission is possible" means the successful production of any new infections (i.e. R0 > 0), not the stable establishment of the parasite in a naive population (i.e R0 > 1). Thus, this is a very generous categorization of transmission, and for our scaled version of R0, that is S(T), the threshold for any possible transmission remains the same as for the unscaled R0 such that R0 > 0, where 0 is defined absolutely and needs no relative scale.

In this paper and previous papers, however, we use 0.001 as a threshold instead of 0 in order to exclude very small values that are not meaningfully distinguishable from 0. It is true that this threshold (unlike actual 0) will differ a small amount based on the scaling used, but the differences are small relative to the overall precision and uncertainty of the models.

Additionally, we added the following text to the discussion to emphasize that our mapping results are largely provided to illustrate how the impact of RS on thermal limits can affect mapping applications (lines 419-421):

> These mapping results are largely illustrative to demonstrate how rate summation can affect applications that are based on the thermal limits.

• *Line 319: should this be "thermal suitability" instead of environmental suitability?*

Yes, this has now been revised to say "thermal suitability."

• *Lines 31 – 33 and 326 – 328. There is no assessment of the rate summation approach against transmission intensity or known transmission seasonality in the field, so I don't understand how the constant temperatures are more accurate for predicting and mapping the thermal limits of transmission? Indeed, laboratory strains of mosquito are used, but there is evidence of local adaptation of mosquitoes.*

Our assessment of “accuracy” is based on comparing the rate summation models (models 3, 4, and 5) and constant temperature model (model 1) to the model fit to empirical data from fluctuating thermal environments (model 2), which is used as a benchmark of accuracy. We added text emphasizing this to the Model Overview section (lines 170-171):

> In general, we consider the empirical fluctuating model (model 2) to be most reflective of real systems and thus use it as a benchmark for the other models.

This study focuses on evaluating current, commonly-used methods for modeling and answering the questions:

1. Does rate summation actually improve model performance compared to constant temperatures?
2. What is the best way to model transmission using data collected from constant temperatures?

The former has not been tested previously, and the latter is an important question given the abundance of constant temperature trait data and lack of fluctuating temperature trait data. We believe our study directly and effectively addresses these specific questions in its current form, and that attempting to validate the models with field data is beyond the scope of this study and instead warrants a full follow-up study.

We agree that using a lab-adapted mosquito strain is potentially an important limitation, and we now include it in the paragraph discussing data limitations (see response to the first comment from Reviewer #2 above). However, in the literature there is limited and mixed evidence for this in mosquito systems. Dennington et al. 2024 (Global Change Ecology) and Cooper et al. 2024 (Proc Roy Soc) argue that they see evidence for local adaptation; however, no rigorous population genetics were included to rule out the effects of genetic drift, and in the former study the differences in thermal performance between populations were not related to the environment of origin (and thus could be random population differences or founder effects, not beneficial thermal adaptation). Additionally, when the direct comparison of thermal performance was made between field and long-standing lab colonies (see Lyons et al. 2012 Malaria Journal), thermal performance was conserved. Finally, previous temperature-dependent suitability models that were validated with entomological inoculation rates (Mordecai et al. 2017 PLoS NTDs) and spatial variation (Tesla et al. 2018 Proc Roy Soc) found that these temperature performance models are quite accurate at predicting overall suitability.

• *With regards to the established literature, previous work on the effects of temperature on malaria vector and parasite traits in the laboratory (mostly published by the authors) is discussed in detail. The significance of these studies and this research to understand transmission in the field is noted (lines 68 -75, lines 402 - 404) and the results from this paper mapped (Figure 6). There has, however, also been lots of debate into the effects of temperature on malaria transmission using empirical malaria prevalence or incidence data. I think it is, therefore, necessary to either include this research to give a broader overview of the scientific consensus or change the narrative accordingly?*

We agree that both of these are important points and added another paragraph to the Discussion to address them (lines 490-7). We were unsure what specific literature the

reviewer is referring to re: the debate over evidence for temperature impacts in empirical data, but if they provide citations we would be happy to include them in a future revision.

> We did not attempt to validate the predictions from our various models with field data of human cases or entomological inoculation rates (EIR). Instead, we judged the "accuracy" of the rate summation models (models 3, 4, and 5) and the constant temperature model (model 1) by comparing them both to the empirically-fit fluctuating temperature model (model 2), which we assume most accurately reflects the impacts of natural temperature variation. Future work could use disease incidence or EIR data to better evaluate the accuracy of these models and to determine if incorporating fluctuating temperatures improves their performance.

Reviewer #3 (Remarks on code availability):

Code and data are available on GitHub. The code provides a README file and details of package versions used. An R project was used making it easier to run the code.