Imperial College London

CMEE Mini Project, February 2016

How well does a mechanistic model based upon biochemical principles fit a dataset of thermal responses of individual fitness in plants?

Calum Pennington

Word count: 3351

Introduction

As temperature affects all biological processes, the response of individuals, populations and communities to climate change is likely driven by the temperature sensitivity of traits. [Dillon et al., 2010, **?**]

Phemenological models simply describe the relationship between data; parameters do not have a mechanistic basis. In contrast, mechanistic models give insight into processes behind phenomena. They are more useful because they represent a quantitative hypothesis and parameters are estimates of actual system properties. However, choosing the wrong mechanistic model may give misguided conclusions about the mechanisms of interest. Probing the mechanisms of trait thermal sensitivity is important, to accurately model effects of climate change. But, it is to understand the effect of variation in temperature dependence of individuals, populations, and communities, where mechanistic models are most key [Pawar et al., 2016].

At the level of organisms, fitness and biological rates have a typical, unimodal, temperature dependence. Performance increases gradually with temperature to an optimum, then declines [Gillooly et al., 2001, Knies and Kingsolver, 2010]. This has been attributed to Arrhenius kinetics: rates depend on rate-limiting reactions, and reaction rates increase exponentially with temperature [**?**]. The decrease in fitness occurs as high temperatures disrupt the active site of and denature enzymes. Mechanistic models underpinned by physiological principles are proposed to improve predictions for temperature dependence and the biological effects of climate change [Dell et al., 2011, **?**].

The Boltzmann-Arrhenius model of biochemical kinetics is a mechanistic model for predicting trait thermal response [Knies and Kingsolver, 2010]:

– B - trait at a given temperature, T (K)

– B_0 is temperature-independent and accounts for body size. It controls the curve's vertical offset (due to variation among species) and is the trait value at a reference temperature.

– E - activation energy (in eV) of rate-limiting reactions; controls the curve's rise up to the peak (trait's thermal sensitivity).

– T_{pk} - temperature at which the trait peaks

– k - Boltzmann constant ($8.617 \times 10^{-5} \cdot K^{-1}$)

Some models simply use the Arrhenius equation, however, the Schoolfield allows for thermal denaturation of rate-limiting enzymes [Knies and Kingsolver, 2010, Schoolfield et al., 1981]. An extra parameter, E_d (deactivation energy) controls the curve's fall.

$$B = \frac{B_0 e^{\frac{-E}{k}(\frac{1}{T} - \frac{1}{283.15})}}{1 + \frac{E}{E_d - E} e^{\frac{E_d}{k}(\frac{1}{T_p k} - \frac{1}{T})}} \quad (1)$$

The proposition that thermal sensitivity follows Arrhenius kinetics - that it corresponds to the activation energy of underlying rate-limiting enzymes - has been criticised; these models are contentious. A particular issue is whether E reflects the activation energy of one or more enzymes, and studies exhibit E values that vary from the reported interspecific average, 0.6 eV [Pawar et al., 2016, Dell et al., 2011]. The development of mechanistic models of trait temperature dependence rely on resolving this.

To this end, I aim to evaluate the fit of the Schoolfield and Boltzmann-Arrhenius models to a dataset of thermal performance curves. I contrast a phemenological alternative, a cubic polynomial model, which has the same number of parameters as the Schoolfield.

Questions

1. Do the models fit the data well?

2. Overall, which model best fits the data?

3. Does the best model vary across traits and habitats?

4. Are the data consistent with the mechanism of Arrhenius kinetics? Does the Schoolfield model generate accurate, realistic best-fit parameter values?

Method

Dataset

The dataset contains (2314) thermal responses of growth, photosynthesis, and respiration rates in 629 species, across 120 orders. (To record a species' thermal response, you measure a certain trait across temperatures within a given range). Nearly all species are algae or plants (Cyanophyta, Chromista, Plantae and Viridiplantae) (aquatic and terrestrial), but there are data on six Euglenophytes, twenty-two fungi, and six Metazoans. These were generated by field and lab experiments across the world, and compiled by various people into the Global Biotraits Database [i]. Each row of the dataset contains a trait and temperature value - an x and y point of a thermal response. Points of the same response share a unique ID. Temperature is the organism's body temperature, not ambient (which was not recorded for nearly all responses).

[i] published online at biotraits.ucla.edu

Data Wrangling

I removed (63) rows with missing trait values (none were missing temperature values). The dataset's authors standardised values of the same trait from different sources (converted them to the same SI unit). Except for growth rate, this was either not completed or done wrongly. Often, an original value is missing a standardised one. For many IDs, standardised values are equal when original ones differ. Maybe the first value was standardised and mistakenly copied to subsequent rows. So, for growth rate, I used standardised values. Otherwise, I used original ones. This is preferable to discarding data - if more time was available, I would standardise these values.

I log-transformed trait values, as I used logarithmic versions of the Boltzmann and Schoolfield models. Working in linear space is preferable to exponential, as:

- constant rates of change are more easily interpreted than exponential ones
- it is simpler to interpret model parameters, e.g. in linear space, E is the gradient of the rising function
- residuals are more easily calculated in linear space.

Accordingly, I had to deal with negative and zero trait values (log of a negative is not a number; of 0 is infinite). Instead of discarding these, I transformed all trait values again. To deal with negatives, I subtracted the minimum of all traits. To deal with zeros, I added one (substract the minimum first, as this makes more zeros). Importantly, I transformed original and standardised values separately. The minimum original trait was `~-285`; the standardised, `~-0.003`. As the maximum standardised value was 6.7, I felt substracting -285 from these was drastic, and maybe invalid. After data wrangling, there were 2308 of the 2314 original curves (21 915 of the original 21 978 data points).

Analysis

I performed all analysis in R, version 3.3.2, using the 'minpack.lm' package for nonlinear regression. To fit the cubic polynomial model, I used linear regression [ii]. To fit the Schoolfield and Boltzmann, I used nonlinear (least-squares) regression.

Nonlinear Regression (NLLS)

Regression finds values of the parameters that are most likely to be correct - those that minimize the sum of the squared vertical distance between data points and the curve (Motulusky & Christopoulos 2004). It

[ii] while the graph of y vs x is curved, it is a linear equation. This is because a graph of any parameter vs y would be linear (Motulusky & Christopoulos 2004).

is the sum of squares that is useful for comparing models.

Nonlinear regression is an iterative procedure, so you need to define initial values for each parameter. I chose 10°C as the reference temperature (T_{ref}), which is within the range organisms commonly operate [Schoolfield et al., 1981]. The start value of B_0 was the trait value at T_{ref}. If there was no recording at T_{ref}, I used the temperature closest to it. I chose the trait value at the minimum temperature, if this temperature was bigger than T_{ref}. Otherwise, I chose the trait at the maximum temperature, for temperatures below T_{ref}. I set the start value of T_{peak} to the temperature at which the trait value peaked. The maximum trait may occur more than once and at multiple temperatures. If so, I set the start value to the maximum temperature at which the trait peaked.

To set the start value of E (E_{st}), I approximated the curve rise's gradient, taking the rise as data up to and including T_{peak}. In log space, I calculated linear regression of an Arrhenius plot (log trait vs the reciprocal of temperature, $\frac{-1}{kT}$) [Pawar et al., 2016, Schoolfield et al., 1981]. As E models the curve's rise, it must be positive, so I used the gradient's absolute value (its magnitude without regard to its sign). If the curve rise had one data point or was absent, regression was not possible, so I set E_{st} to 0.6, a universal average [Dell et al., 2011]. As thermal performance curves have exponential rises, linear regression in log space is an estimate. To account for error in the initial estimate of E_{st}, I generated a hundred random deviants, using E_{st} as the mean of a normal distribution. For nonlinear regressions that failed with the original starting estimate, I tried again, up to a hundred times, with a deviant.

I excluded (361) curves without at least two unique temperature and five trait values. This was to ensure the data captured at least part of a thermal response, and also to avoid good fits due to overfitting (the Schoolfield and cubic have four parameters). Similarly, to fit the Boltzmann, I did not run NLLS for (117) curves whose rise did not have at least three unique temperatures. After filtering, 1947 of the original 2314 curves were analysed.

It is important not just to know the best-fit value of each parameter, but how certain that value is. How well did the Schoolfield and Boltzmann determine the best-fit values? Accordingly, I looked at the standard error of the best-fit values. Nonlinear regression finds parameters that make a model fit the data as closely as possible, but does not ask if another model might work better. Moreover, the model with the smallest sum-of-squares is not necessarily the best. A more complicated model (more parameters) has more flexibility to fit the data. For model comparison, I used Akaike's Information Criteria, which balances goodness-of-fit with the number of parameters.

Assumptions of Linear and Nonlinear Regression

Results

NLLS

In most (1288) cases, NLLS to fit the Schoolfield converged in fewer than 100 iterations. For 646 curves, NLLS succeeded but did not converge. For 13, NLLS gave an error. For 1691 curves, NLLS succeeded with the start value of E originally calculated; 239 needed a different value. NLLS to fit the Boltzmann succeeded for all but one curve.

R^2: do the best-fit curves come close to the data?

The cubic polynomial had an R^2>0.6 for 95.3% of curves (n=1947); Schoolfield, 93.5% (n=1927[iii]); and Boltzmann, 93.7% (n=1819[iv]) (Figure 1).

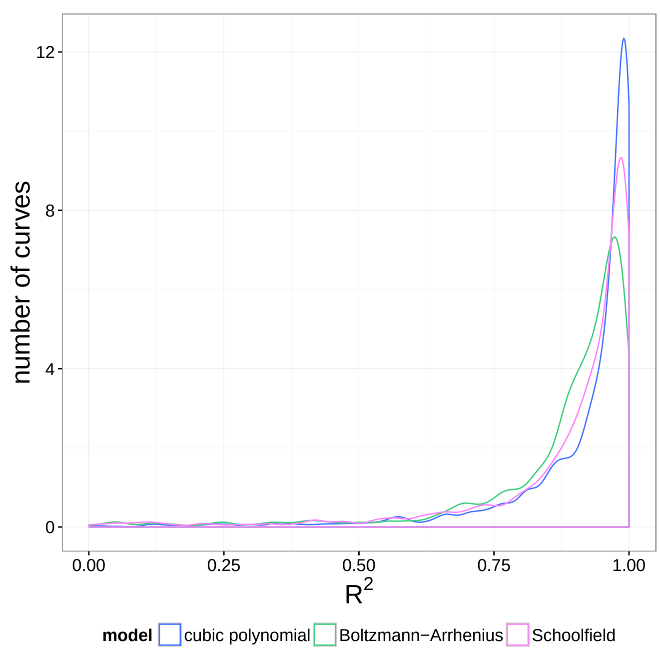


Figure 1: *Density of R^2 goodness-of-fit values*

I fit models to the thermal responses of growth, photosynthesis, and respiration rates in algae and plants. Using linear regression, I fit the cubic polynomial to 1947 curves. Using nonlinear regression, I fit the Schoolfield to 1938 curves, and Boltzmann to 390.

[iii]NLLS succeeded for 1938 curves, 11 of which I excluded for having an anomalous R^2 (see *Anomalies*).

[iv]I did not run NLLS for 117 curves, and excluded 10 more for having an R^2<0 (see *Anomalies*).

Comparing Models using Akaike's Information Criteria (AIC)

Overall, for most (63.2% of) curves, the cubic polynomial model had the lowest AIC. The Schoolfield had the lowest for a notable proportion (31.0%). The Boltzmann had the lowest in only 5.8% of cases. This pattern is consistent across data subsets, but the cubic was particularly dominant for respiration rates (81.5%) and terrestrial habitats (73.6%) (Figure 2). Growth rate was another notable deviation: for a higher proportion (39.1%), the Schoolfield had the lowest AIC.

If the Boltzmann AIC was lowest (110 curves), I compared the cubic and Schoolfield. The cubic AIC was lower for nearly all (95.5%) of these. Curves where the Schoolfield AIC was lower, were mainly data on photosynthesis rate or freshwater species (lower Schoolfield for 14.8 and 13.8% of these subsets respectively). For respiration rates and terrestrial habitats, the cubic had the lowest AIC in all cases.

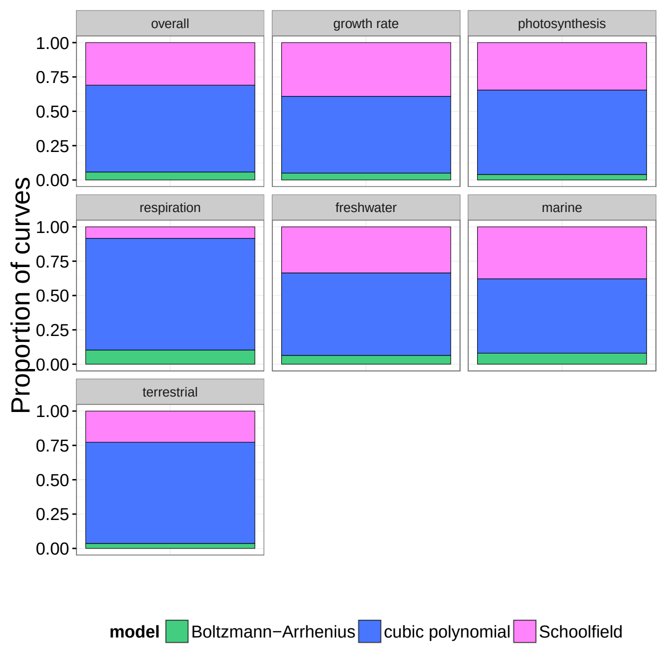


Figure 2: *Per model, the proportion of curves for which it had the the lowest AIC. Differences across habitat and trait categories are shown.*

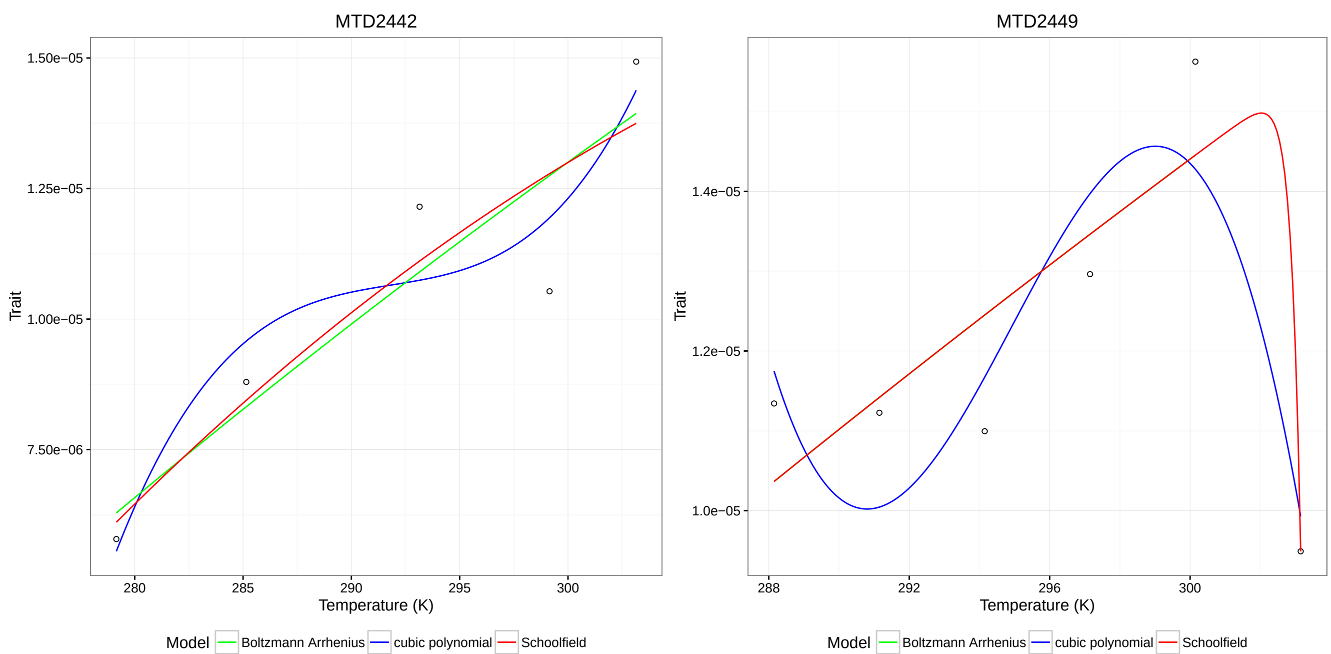
I fit models to the thermal responses of growth, photosynthesis, and respiration rates in algae and plants. Using linear regression, I fit the cubic polynomial to 1947 curves. Using nonlinear regression, I fit the Schoolfield to 1938 curves, and Boltzmann to 390. I compared the models using Akaike's Information Criteria.

What was the cubic's goodness-of-fit, when the Schoolfield had a lower AIC, and vice versa?

Where the Schoolfield's AIC was lower (n = 607), the cubic had an R^2>0.9 for 69.2% of curves. For 3.1% (19) of curves, it had an R^2<0.6. The Schoolfield also had an R^2<0.6 for 2.6%. Where the cubic's AIC was lower (n = 1329), the Schoolfield had an R^2>0.9 for 61.7% of curves. For 8.9% of curves, it had an R^2<0.6.

Are curves that only represent the rise of a thermal response best fit by the Boltzmann?

390 curves only represent the rise of a thermal response (e.g. Figure 3a). For most of these, the cubic had the lowest AIC (74.6%), similarly to the overall results. However, in contrast, a much higher proportion were best fit by the Boltzmann (24.1%). What was the cubic's goodness-of-fit, where the Boltzmann's AIC was lower? 86.2% of curves had an R_2 >0.6; 61.7% had an R_2 >0.9. Where the Boltzmann had the overall lowest AIC, 80.3% of curves only represent the rise.



(a) Data only representing the rise of a thermal response

(b) Data with an initial decrease in the trait value

Figure 3: Models: cubic polynomial (blue), Boltzmann-Arrhenius (green), Schoolfield (red)

Are curves with an initial decrease in the trait value best fit by the cubic polynomial?

180 curves had an initial decrease (e.g. Figure 3b). The cubic had the lowest AIC for 82.8% of these: a higher proportion than that for all curves. For 25.4%, the Schoolfield had an R^2 <0.6.

Do the best-fit parameter values of the Schoolfield make sense?

For 93.0% of curves, the best-fit value of E was <0.1 eV . The minimum value was 2.3 x 10^{-13} eV; the maximum, 2.7 eV. E was >0.6 for 1.4% of curves. For 99.6% of curves, the best-fit E_d value was <1 eV. The minimum was 0.002 eV; the maximum, 144.9 eV. For 63.3% of curves, the best-fit T_{pk} value was between 20 and 60°C. For 17.1%, it was >100°C; for 4.2%, <10°C.

Are plausible best-fit T_{pk} values generated for curves that only represent the rise of a thermal response?

For IDs with a best-fit T_{pk} >400 K, 81.7% only represented the rise of a thermal response (n = 317).

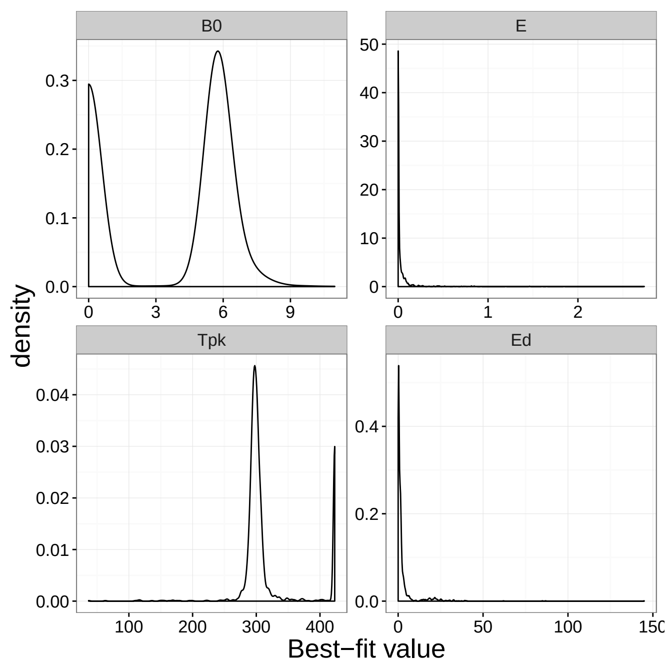


Figure 4: *Density of the Schoolfield model's best-fit parameter values*

Using nonlinear regression, I fit the model to 1938 thermal-response curves of growth, photosynthesis, and respiration rates in algae and plants.

How precise are the Schoolfield's best-fit parameter values?

The highest standard error (SE) of a best-fit value for B_0 was 11.0 (Figure 5). For all 1938 fits, the SE of T_{peak} was between 0 and 1. For E, 92.3% of fits gave a SE <2. However, twelve fits gave a SE >100. Seven of these curves lacked a well-defined rise, having two or fewer points before T_{peak} (e.g. Supplementary Figure). Three lacked any form (e.g. Supplementary Figure), and two seemed to show much variation among replicates (e.g. Supplementary Figure) (judged by manually inspecting plotted data [v]; I cannot give quantitative measures).

For E_d, 97.0% of fits had a SE <10. However, fifty-nine fits had a SE >10. Thirty-six of these curves lacked a well-defined fall (e.g. Supplementary Figure), having two or fewer points after T_{peak}. The other twenty-three, though, seemed to be well-defined (judged by manually inspecting plotted data).

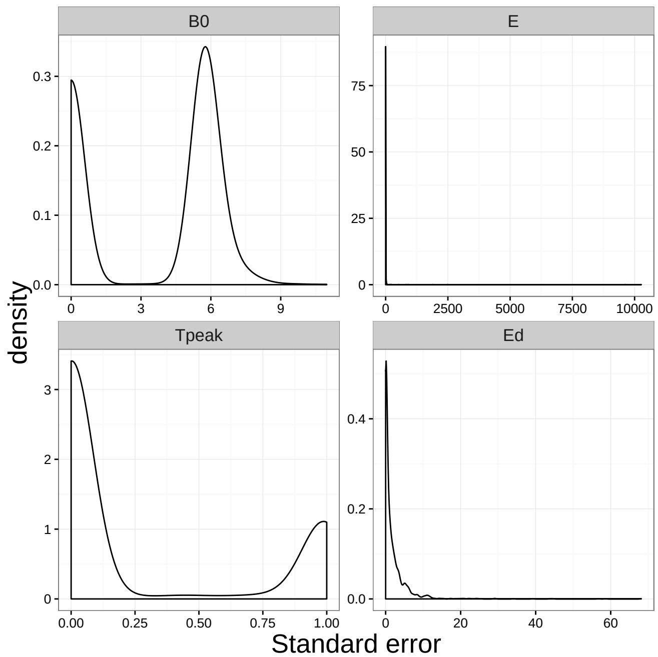


Figure 5: *Density of the standard error of the Schoolfield model's best-fit parameter values*

Using nonlinear regression, I fit the model to 1938 thermal-response curves of growth, photosynthesis, and respiration rates in algae and plants.

Discussion

Does the cubic polynomial describe the general shape of the dataset?

The cubic polynomial model best fit most curves. This was most striking for respiration rates. However, a notable proportion were not best fit by the cubic. So, a key question is, where did these exceptions occur? While many occurred when the data only represent the rise of a thermal response, in most of these cases, the cubic was still the best fit. Moreover, where it was not, it had sufficient goodness-of-fit. The split between curves best fit by the cubic, and those by the Schoolfield, was generally consistent across studied habitats and traits. For growth rate, there was an increased proportion of Schoolfield best fits, although most were still best fit by the cubic. So, the cubic describes the general shape of the dataset's

[v] Curve IDs: MTD3954, MTD4304, MTD4312, MTD4342, MTD4375

TPCs and could be used to interpolate unknown values. Yet, crucial exceptions do exist: attention must be paid to the trait and habitat of interest. Exactly if, how and why the cubic is an unreliable fit of certain data, must be further probed, especially for a wider range of traits/habitats. Conclusions using the cubic should be drawn cautiously and, where precision is key, it is worth considering other models.

Does the Schoolfield fit the dataset, and are its best-fit parameter values accurate and realistic?

While the Schoolfield model was not the best fit of most curves, it was for a notable fraction. Furthermore, similarly to the cubic, where it was not the best, it had sufficient goodness-of-fit. (As previously discussed, this was especially stark for respiration, but less so for growth rate). A marked departure were curves that had an initial decrease in trait value. Not only were more best fit by the cubic, but the Schoolfield was a bad fit to many. Yet, this does not invalidate the Schoolfield for these data. The decrease is likely due to error amongst replicate measurements, or a lack of replicates. It is presumably not a true decrease, but low-temperature deactivation. At very low temperature, enzymes work slowly (substrate molecules have less energy and move into the active site more slowly) - the gradient of the curve's rise is less steep. An extended Schoolfield models this with an extra parameter, E_l [Schoolfield et al., 1981]. With another parameter, E_h, it can model high-temperature deactivation, a similar effect. This study demonstrates that in most cases, a simplified Schoolfield is appropriate - measurements were not made at sufficiently low temperatures to detect deactivation.

Most best-fit E values were much lower than 0.65 eV, the reported average [Dell et al., 2011]. Given there are so many curves, this is notable. It may be due to differences in the studied datasets [Pawar et al., 2016], or the units presented. Here, growth rate (43% of responses) was standardised to 'per second' units, which is small, lowering the trait values and thus slope. On the other hand, E is the slope of the response's rise (when plotted on ln scale), so should be independent of the rate's units. Further clarification is needed. Nevertheless, upon viewing curves plotted with best-fit parameters, all go through the raw points (see *Supplementary Information*). The best-fit values of T_{pk} were biologically realistic (most were around 27°C). While a big proportion were implausibly high, most of these occurred when only the response's rise was represented. T_{pk} occurs at the inflection between the response's rise and fall; its estimation requires both aspects. Moreover, these data may not capture the response's true peak. Thus, in these cases, it is unreasonable to estimate T_{pk} reliably.

A severe limitation of the analysis is that I did not calculate confidence intervals of the best-fit parameter values. While confidence intervals are based on SE, you must consider the value's magnitude to interpret its SE (e.g. a SE of 1 for a value of 10 is permissible, but for a value of 0.5, it is a lot of uncertainty). That is hard in a big dataset. So, I cannot conclude this key question, but only allude to general trends. Given the small best-fit values of E, it is especially hard to evaluate SE, but the uncertainty is not excessive. Exceptions are explained by a lack of a curve rise. Most best-fit T_{pk} values were between 290 and 330 K; all had a SE 0-1 K. T_{pk} values were not only plausible, but had high certainty.

Generally, the Schoolfield is a reliable model. However, the study emphasises the contention of whether thermal sensitivity can be expressed as one activation energy. Variation among traits and habitats is critical for modelling climate change effects, and a challenge to the Schoolfield's efficacy. However, as thermal responses seem to have a marked basis in Arrhenius kinetics, the Schoolfield is a good platform upon which to explore more complicated models.

To improve the analysis, I should have used the corrected AIC (AICc), as each thermal response had a small value of N. For small sample sizes, the calculated AIC is too small and AICc, more accurate. Moreover, it would be worthwhile considering the magnitude of difference among AICs, for a response. Also, plotting the residuals of fits would have enabled me to assess if curves deviated systematically from data. In line with nonlinear regression's assumptions, if a model is appropriate, data are randomly scattered around the best-fit curve. Clusters of points would indicate data are not randomly distributed, and the fit is inappropriate (Motulsky & Christopoulos 2004).

References

[Dell et al., 2011] Dell, A. I., Pawar, S., and Savage, V. M. (2011). Systematic variation in the temperature dependence of physiological and ecological traits.

[Dillon et al., 2010] Dillon, M. E., Wang, G., and Huey, R. B. (2010). Global metabolic impacts of recent climate warming. *Nature*, 467(7316):704–706.

[Gillooly et al., 2001] Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., and Charnov, E. L. (2001). Effects of size and temperature on metabolic rate. *Science (New York, N.Y.)*, 293(5538):2248–2251.

[Knies and Kingsolver, 2010] Knies, J. L. and Kingsolver, J. G. (2010). Erroneous Arrhenius: modified Arrhenius model best explains temperature dependence of ectotherm fittness. *The American Naturalist*, 176(2):227–233.

[Pawar et al., 2016] Pawar, S., Dell, A. I., Savage, V. M., and Knies, J. L. (2016). Real versus Artificial Variation in the Thermal Sensitivity of Biological Traits. *The American Naturalist*, 187(2):E41–E52.

[Schoolfield et al., 1981] Schoolfield, R. M., Sharpe, P. J. H., and Magnuson, C. E. (1981). Non-linear Regression of Biological Temperature-dependent Rate Models Based on Absolute Reacdon-rate Theory. *Journal of theoretical biology*, 88:719–731.