Modelling Thermal Responses of Metabolic Traits

Jacob Griffiths

Imperial College London

jacob.griffiths18@imperial.ac.uk

March 3, 2019

Abstract

Hello this is the Abstract

I. Introduction

Temperature is a fundamental parameter in almost all processes and its importance in metabolic biology is well documented (Montoya et al. 2012, Dell et al. 2011, DeLong et al. 2017). Metabolic processes are catalysed by enzymes which depend on kinetic, and ultimately heat, energy to function. As temperature decreases, atoms and molecules move progressively slower and thus metabolic rates decrease accordingly whereas when temperature increases, metabolic rates increase rapidly until the thermal optimum is reached (DeLong et al. 2017, Dell et al. 2011). That is, the temperature at which optimal metabolic rate occurs (T_{pk} or T_{opt}). Beyond this, increasing temperature will start to hinder the metabolic rate as the proteins that enzymes are composed of will start to denature until the process stops entirely. These metabolic responses to temperature exhibit a remarkably similar pattern, often referred to as Thermal Performance Curves (TPCs), across an array

of metabolic processes and taxa. This makes the study of TPCs a useful tool of comparison for all life on Earth as all species depend on metabolism for their energy. Furthermore, an increased understanding of how species respond to temperature is imperitive in a rapidly warming world. If we can find some plasticity in a species' temperature tolerance then perhaps it will have a better chance of avoiding the mass extinction that is sweeping our planet, although some recent findings suggest the scope for adaptation may be limited (Tüzün & Stoks 2018). Of course, there are other ways a species may adapt through latitudinal range shifting or evolution, but the latter seems unlikely in the time-frame available and the former is only possible for mobile species with suitable habitats to move to.

i. Models

Three models were used in this study to compare their ability to fit to each dataset within BioTraits. Firstly, the cubic polynomial was used as a phenomenological model with the following form:

$$B = B_0 + B_1 T + B_2 T^2 + B_3 T^3 \qquad (1)$$

Where B is the responding trait value and T is the temperature. Secondly, the Briere model (Briere et al. 1999) was used as an alternative phenomenological model:

$$B = B_0 T (T - T_0) \sqrt{T_m - T} \qquad (2)$$

Where T_0 and T_m are the minimum and maximum tolerances for the trait, B, and B_0 is a normalisation constant. Whilst this model is phenomenological, it can still provide useful biological information when fit as it provides an estimate of the minimum and maximum thermal tolerances of a particular trait for a particular organism. However, it falls short of the definition of a mechanistic model as the model provides no insight into the underlying biological mechanisms at work. Therefore, the third model used in this study was a simplified version of the Sharpe-Schoolfield (Schoolfield et al. 1981) model to provide a mechanistic comparison as it was formulated from thermodynamic and enzyme kinetic theory. The full

model is given by:

$$B = \frac{B_0 e^{\frac{-E}{k}(\frac{1}{T} - \frac{1}{283.15})}}{1 + e^{\frac{E_l}{k}(\frac{1}{T_l} - \frac{1}{T})} + e^{\frac{E_h}{k}(\frac{1}{T_h} - \frac{1}{T})}} \quad (3)$$

And the simplified:

$$B = \frac{B_0 e^{\frac{-E}{k}(\frac{1}{T} - \frac{1}{283.15})}}{1 + e^{\frac{E_h}{k}(\frac{1}{T_h} - \frac{1}{T})}} \quad (4)$$

Where k is the Boltzmann constant (8.617×10^{-5} eV $\dot{K}^{-1}$), B_0 is the trait value at a reference temperature (283.15 K in this study), E_l is the low-temperature deactivation energy (eV) of the enzyme and controls the behaviour of the curve at very low temperatures and T_l is the temperature at which 50% of the enzyme is low-temperature deactivated. E_h is the high-temperature deactivation energy of the enzyme and controls the behaviour of the curve at high temperatures and T_h is the temperature at which 50% of the enzyme is high-temperature deactivated. E is the activation energy which controls the behaviour of the curve in the enzyme's 'normal operating range', that is before $T_p k$ but not at low temperatures. The simplified version was chosen for this study as low-temperature deactivation is weak and a lot of datasets within BioTraits lacked sufficient data at low temperatures. It also allows for more datasets to be used as the minimum number of datapoints required for the six-parameter full model would be larger than for the four-parameter simplified version.

II. Methods

i. Data

The database used in this study, BioTraits, was provided by my supervisor, Dr. Samraat Pawar and is an extension of the database used by Dell et al. (2011). It consists of 2165 unique thermal responses of metabolic processes from 1010 publications. Predominantly, respiration, growth and photosynthetic rate are the metabolic process being measured against temperature. BioTraits includes species from many Phyla and with diverse life histories but a majority of representitives are terrestrial species, often Arthropods.

As this dataset contains 155 columns and 25826 rows, it was first refined to a handful of relevant columns for this study to improve computational speed, namely the trait value and temperature. Rows with missing values

for these columns were removed and any sub-dataset with less than five datapoints was removed as this is the minimum required to estimate four parameter models like the cubic polynomial and simplified Schoolfield.

ii. Parameter estimation

Using R, starting parameters for every unique sub-dataset in BioTraits were calcualted. For the cubic polynomial model, starting parameters of 1 were used for all four parameters. For Briere, estimates for T_0 and T_m were made using the minimum and maximum recorded temperatures respectively. For Schoolfield, A reference temperature of 10 degrees Celsius (283.15 K) was used for as this has been used effectively in previous publications (Dell et al. 2011) and B_0 was estimated as the recorded trait value nearest to this temperature, by definition. The peak metabolic rate B_{max} was then calculated (T_{pk} being the corresponding temperature at this trait value) and the dataset was split around this value. If B_{max} occurred at the highest recorded temperature (i.e. the rate had not started descending yet) the dataset was not split and the following regression was carried out on the whole dataset. The trait values of each side were logged and the reciprocals of the temperature values were multiplied by the boltzmann constant ($8.617 \times 10^{-5} eV \cdot K^{-1}$). Linear regression was carried out on the left-hand (below T_{pk}) data and the estimate for E was taken to be the gradient of this line, with the Eh estimate being twice this value. If regression failed, default estimates of 0.65 for E and 1.3 for Eh were used as recommended defaults from the literature and E was given bounds of 0 to 3 while Eh was bounded between 0 and 6 (Montoya et al. 2012, Dell et al. 2011). Th was estimated by calculating the nearest recorded temperature to $B_{max}/2$ as Th is the temperature at which half the enzyme units have been made inactive so this provides a good estimate, with a lower bound of T_{pk} and an upper bound of 400 Kelvin applied (Sal et al. 2018). For datasets with no datapoints after B_{max}, Th was given a starting value equal to T_{max}.

iii. Model comparison

The Akaike Information Criterion (AIC) (A 1974) was used to compare model fits within each dataset and is given by the formula assuming the model is univariate, is linear in its parameters and has normally-distributed residuals:

$$AIC = 2k - 2\ln(\hat{L}) \tag{5}$$

Where $\hat{L}$ is the maximum likelihood estimation and k is the number of parameters for the model. This method was chosen as it rewards the relative goodness of fit between models on the same data but penalises number of parameters used as this can sometimes lead to overfitting. However, despite this penalty, AIC can still be prone to favouring models with more parameters if the sample size is small. This can be circumvented by using AICc (Hurvich & Tsai 1989), an extension of AIC given by:

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1} \tag{6}$$

Where n is the sample size and k is the number of parameters as before. It should be noted that as $n \to \infty$, the additional parameter penalty tends to zero and thus AICc tends to AIC, making it suitable for large samples too. As some of the datasets used in this analysis had only a handful of datapoints, AICc was used to compare models instead of AIC.

In addition to AICc, adjusted R^2, or $\bar{R}^2$, was used as an alternative comparison tool. Generally attributed to Wright (1921), R^2 is purely a measure of goodness of fit and is given by:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{7}$$

Where SS_{res} is the sum of the squared residuals of the model and SS_{tot} is the total sum of squares. A score of 1 is a 'perfect' fit and a negative score is considered a worse first than a straight, horizontal line through the mean as a model. Similar to AIC, R^2 is susceptible to overfitting as the addition of a new parameter will always improve the score. Fortunately, adjusted R^2, $\bar{R}^2$:

$$R^2 = 1 - \frac{VAR_{res}}{VAR_{tot}} \tag{8}$$

Where $VAR_{res} = SS_{res}/n$ and $VAR_{tot} = SS_{tot}/n$, only improves its score if an additional parameter improves the model more than would be expected by chance, making

it less susceptible to overfitting and a better comparative tool for this study.

III. Results

The cubic polynomial may have no biological underpinning but is sacrificing realism much different to sacrificing precision (Levins 1966)?

IV. Discussion

References

A, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Briere, J. F., Pracros, P., Le Roux, A. Y. & Pierre, J. S. (1999), 'A novel rate model of temperature-dependent development for arthropods', *Environmental Entomology* **28**(1), 22–29.

Dell, A. I., Pawar, S. & Savage, V. M. (2011), 'Systematic variation in the temperature dependence of physiological and ecological traits', *Proceedings of the National Academy of Sciences* **108**(26), 10591–10596.

DeLong, J. P., Gibert, J. P., Luhring, T. M., Bachman, G., Reed, B., Neyer, A. & Montooth, K. L. (2017), 'The combined effects of reactant kinetics and enzyme stability explain the temperature dependence of metabolic rates', *Ecology and Evolution* **7**(11), 3940–3950.

Hurvich, C. M. & Tsai, C. L. (1989), 'Regression and time series model selection in small samples', *Biometrika* **76**(2), 297–307.

Levins, R. (1966), 'The Strategy of Model Building in Population Biology', *American Sci* **54**(4), 421–431.

Montoya, J. M., Trimmer, M., Pumpanen, J., del Giorgio, P., Gasol, J. M., Cescatti, A., Dossena, M., Woodward, G., Caffrey, J. M., Staehr, P. A., Yvon-Durocher, G. & Allen, A. P. (2012), 'Reconciling the temperature dependence of respiration across timescales and ecosystem types', *Nature* **487**(7408), 472–476.

Sal, S., Smith, T. P., García-Carreras, B., Pawar, S. & Kontopoulos, D. G. (2018), 'Use and misuse of temperature normalisation in meta-analyses of thermal responses of biological traits', *PeerJ* **6**(February), e4363.

Schoolfield, R. M., Sharpe, P. J. & Magnuson, C. E. (1981), 'Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory.', *Journal of Theoretical Biology* **88**(4), 719–731.

Tüzün, N. & Stoks, R. (2018), 'Evolution of geographic variation in thermal performance curves in the face of climate change and implications for biotic interactions', *Current Opinion in Insect Science* **29**(Figure 1), 78–84.

Wright, S. (1921), 'Correlation and Causation', *Journal of Agricultural Research* **XX**(7), 557–585.