# Imperial College London

Imperial College London

Department of Life Sciences, Faculty of Natural Sciences

# Model Analysis of Biological Temperature-Responsive Rates

*Author:*

Jingkai SUN

*CID:*

01991822

# Model Analysis of Biological Temperature-Responsive Rates

Jingkai Sun (CID: 01991822)

July 17, 2021

**Abstract**

The photosynthesis and respiration rate are essential biological traits of organisms, they are all responsed to their body temperature, indicating that creating effective mathematical model to describe them is becoming increasingly crucial. This project has used four candidate models (Quadratic, Cubic, Briere and simplified Schoolfield models), where quadratic, cubic and briere are phenomenological models and simplified Schoolfield is mechanistic model, to fit 378 datasets of photosynthesis rate and 249 datasets of respiration rate grouped by different IDs of organisms. The result shows that simplified Schoolfield model is effective for fitting datasets of photosynthesis rate, although cubic may be better than it in some cases.

## 1 Introduction

Thermal performance curve (TPC) describes the relationship between biological traits (e.g. photosynthesis, respiration and growth rate) and temperature, which is an essential factor in ecology, biology and physiology [7]. Due to the importance of TPC, numerous biological temperature-dependent rate models have been implemented, it has been increasingly important that using computer tools to fit the model [1], many authors have taken advantages of programming language for model fitting in their paper. However, the paper concerning the comparison between published models is scarce. Therefore, this project aims at giving a comparison between distinct models fitting the thermal performance curve. The remaining sections in this report will be explained as follow. In section 2, a brief introduction of candidate models and an explanation of data preprocessing will be given, then I will explain the method and assessment tool used. In section 3, the results of the fitted model will be explained. Next section will give a conclusion for the results and some future works that can be done.

# 2 Experimental Data and Fitting Methods

## 2.1 Experimental Data

The part of "BioTraits" database, containing 903 fitted subsets, was selected as the main research object for the model analysis with columns ("ID", "ConTemp(℃)", "OriginalTraitValue", "StandardisedTraitName"). Some preprocess had been done to the original dataset. Firstly, all observations where the Original Trait Value is less than or equal to zero had been removed for logarithm purpose. Then, to obtain more reasonable curves, all datasets in which data points are less than or equal to 5 (276 out of 903) were deleted. Afterwards, the datasets were divided into two subsets based on the "StandardisedTraitName", which is "photosynthesis rate" and "respiration rate". Then, there were 378 and 249 (627 subsets in total) subsets in two datasets with different biological traits respectively.

## 2.2 Candidate Models

Four models were chosen in this project, including two linear models (Quadratic and Cubic) and two non-linear models (Briere and Schoolfield Model with High-Temperature Term). For two linear and Briere models, they are phenomenological models lacking some biological interpretation, while Schoolfield is a mechanistic model that has some biological meaning behind that. Briere (1) and Schoolfield (2) model have mathematical expressions as follow:

$$B = \begin{cases} 0, & T \leq T_0 \\ B_0 T(T - T_0)\sqrt{T_m - T}, & T_0 \leq T \leq T_m \\ 0, & T \geq T_m \end{cases} \tag{1}$$

where B is biological trait, which is a function of temperature T. $T_l$ is the maximum temperature that organisms can bear. $T_0$ is the lowest temperature threshold, and B0 is the empirical constant. There are some assumptions for fitting the data well [2].

- It is possible to adjust the muximum or minimum thresholds of temperature [2].

- Asymmetric shape around peak temperature ($T_{opt}$) [2].

- There is a sharp drop after passing the $T_{opt}$ [2].

2

46    • There is an inflection point in the datasets [2].

47  The full Schoolfield model is expressed as follows [10]:

$$B = B_0 \frac{e^{\frac{-E_a}{k}(\frac{1}{T} - \frac{1}{283.5})}}{1 + e^{\frac{E_l}{k}(\frac{1}{T_l} - \frac{1}{T})} + e^{\frac{E_h}{k}(\frac{1}{T_h} - \frac{1}{T})}} \tag{2}$$

48  where k is Boltzmann constant, B is the biological rate responsed by temperature T in Kelvin
49  unit. $B_0$ is the biological rate at low temperature. $E_a$ is the activation energy that determine the
50  increase in temperature up to the optimum. $E_l$ is the de-activation energy at low temperature. $T_l$
51  is the low temperature that enzyme is 50% de-activated. $E_h$ is the de-activation energy at high
52  temperature, while $T_h$ is temperature at which the enzyme is 50% high-temperature deactivated.
53  However, it is difficult to detect the low-temperature deactivation energy [9]. Thus, the schoolfield
54  with only high temperature term (eq.3) was used in this porject:

$$B = B_0 \frac{e^{\frac{-E_a}{k}(\frac{1}{T} - \frac{1}{283.5})}}{1 + e^{\frac{E_h}{k}(\frac{1}{T_h} - \frac{1}{T})}} \tag{3}$$

55  To capture the data pattern more precisely, the Schoolfield equation is rewritten as logarithmic
56  form:

$$ln(B) = ln(B_0) - \frac{E_a}{k} \times (\frac{1}{T} - \frac{1}{283.5}) - ln(1 + e^{\frac{E_h}{k}(\frac{1}{T_h} - \frac{1}{T})}) \tag{4}$$

57  With simplified Schoolfield model, some assumptions are also needed to capture the data pattern
58  well [5]:

59    • All traits (biological rates) is controlled by single enzyme [5]

60    • enzymes need to be deactivated at high temperature [5]

61    • the activation energy ($E_a$) corresponds to the optimal enzyme activity level [5]

62  In addition, this project also assumes that there is no interation terms between linear models.

## 2.3 Model Assessment

All four candidate models mentioned above were assessed. The starting values of two non-linear models were found as follow: For Briere model, the maximum and minimum temperature is given by parameters $T_m$ and $T_0$ respectively. Initial $B_0$ is chosen to be 0.01 manually. For simplified Schoolfield model, initial $B_0$ is just the biological rate at the lowest temperature. $E_a$ and $E_h$ are selected to be the absolute values of slope and slope times ten by running the logarithmic Arrhenius model respectively. Finally, the temperature (or the average of the temperature if there more than two values) that has the highest rate is given to the initial $T_h$. After assigning initial values, the starting value will be chosen randomly using normal distribution method with mean as the value of parameters as well as appropriate bounding and standard deviation values, the fitted data such as TSS, RSS, $R^2$, AIC and BIC are calculated. The whole fitting process will be repeated by a minimum 200 and maximum of 300 rounds in order to find the best fit with the lowest AIC value.

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were calculated for model analysis. These metrics allow us to choose which model performs the best. The AIC approach is an information-theoretic selection by calculating Kullback-Leibler (K-L) distance [3] [4], while the BIC method is based on Bayes factors. All those values are expected to be the lower, the better. However, these metrics will cause the issue if they are used independently since they are used to compare with other values, there is no point computing AIC vlaue for single model. [6]. Thus, the AIC differences were also calculated in order to rank models and include more reasonable best-fit models. The rule of thumb of AIC differences allows us to consider more models as the best-fit model for each trial if the value differences between two models are less than 2, expressed as $\Delta AIC_i / BIC_i < 2$ [4].

## 2.4 Computing Tools

In this project, Python 3 with packages "lmfit", "numpy", "pandas" and "matplotlib" was used as the main tool of creating model fitting program. R (4.0.3) with packages "dplyr" and "ggplot2" was used to analyse and visualise the fitted information of each model obtained after the model fitting program is done. Python 3 with package "subprocess" and "os" was used to integrate all project scripts.

Figure 1 gives the best-fit occurrence for each models using AIC or BIC difference rule. Surprisingly, cubic has the most outstanding performance compared with all other models. However, simplified Schoolfield, the only mechanistic model in candidate ones, has the second most times becoming the

4

best fit model. Cubic is the best model of 387 datasets (based on AIC assessment) and 337 datasets (based on BIC assessment) out of 627 datasets, accounting for approximately 67% and 58% of total datasets respectively, while simplified Schoolfield is the best model of 245 and 237 datasets based on different measurement metrics, accounting for 42% and 41% of the total. Also, 108 datasets have both cubic and Schoolfield models as best fit models based on the rule of AIC/BIC differences. For the remaining two models, quadratic became the best-fit model of 198 and 201 datasets according to AIC and BIC values, which is 34 and 33 datasets more than Briere.

Figure 2 illustrates the number of best-fit models in datasets of photosynthesis and respiration rate using AIC or BIC difference and the method of single model selection with the lowest AIC/BIC values. First of all, with using the AIC/BIC difference approach to count best models, two datasets showed very different results. For the dataset of photosynthesis, furthermore, cubic and simplified Schoolfield are also the top 2 models that have the highest counts of the best fit. Cubic is the best-fit model of 191 datasets of fitting out of 335 datasets, which is 29 datasets more than the simplified Schoolfield using AIC approach but only 7 more by BIC approach, while there is no significant difference between quadratic and Briere models. In respiration datasets, cubic is still the winner compared other models, becoming the best-fit model in 81% or 70% of the total by distinct model assessment approaches. However, quadratic became the second best-fit model, taking the place of the simplified Schoolfield model. By choosing the single model with lowest AIC/BIC values in each trial, interestingly, the simplified Schoolfield model beats all other three models, being the best model in 35% - 36% of total photosynthesis data, while cubic and quadratic models still outperform other ones in the dataset of respiration rate.

# 3    Discussion

From the results shown above, there is no significant difference in model assessment between the AIC approach and BIC approach. The rank of the best-fit model is not affected, although a little gap can be seen for each model as different approaches of model assessment apply. Also, the trend is similar to the most datasets when distinct rules apply except for the dataset of photosynthesis, where a significant differences between model performance of cubic model can be seen as applying different rules.

Due to the interpretability of the mechanistic model, it is expected to be have better performance than some phenomenological or statistical models. However, it is not always true in the project, where the simplified Schoolfield model was the best model only in the photosynthesis dataset when

5

applying rule of single model selection. The reason for this may the mechanistic models have more restrict hypotheses than phenomenological ones, which results in the fitting of mechanistic model is invalid. The figure 3 showed some examples of failed or poor fittings of individual datasets.

By comparison, AIC and BIC both provided similar outcomes and model selection, although there may be a little differences between results of them. Thus, any one of them can be selected as a main model assessment method in my project. However, AIC was preferred to use in this project. The reasion is (1) AIC is a suitable approach for sparse datasets, given that it is an estimator derived from basic thoery called K-L information theory, which is possible for further correction.Thus, this leads AIC to have second-order derived version that can deliminate estimated bias of small sample size [3]. (2) BIC expects the "true model" to be in the model set while AIC does not [8].

The motivation of the project is to analyse four candidate models and provide a reasonable comparison between them. However, there are also more works need to be done. Firstly, the starting values of two non-linear models are selected by normal distribution sampling method, which may be not enough for finding the precise starting values. Thus, this may be also the reason why linear model outperformed non-linear model in respiration dataset. A better approach to find the starting value of parameters for non-linear models should be implemented. Secondly, only one mechanistic model are used to make a model comparison here, may causing bias of the results. Thus, more mechanistic models are worthy of consideration in order to improve the accuracy of the results. Third, due to limited computational power, only approximately maximum 300 repeats of fitting for each indivial dataset is available. Hence, more repeats are suppose to be attempted for obtaining more accurate outcome if the computer with stronger computational power is available such as high-performance computer or other supercomputers.

# 4    Conclusion

In conclusion, the mechanistic model, simplified Schoolfield, is significantly useful when fitting the data of biological rate since the data pattern of approximately 42% of dataset was captured, although cubic model outperformed it in some cases. However, the simplified Schoolfield model does not show an obvious advantage when fitting the datasets of respiration rate, due to the data specificity and strictness of assumptions of mechanistic models. Therefore, there is no sufficient evidence that can conclude that mechanistic model is better than phenomenological models. To select the best models between them, if possible, all datasets should be considered case by case.

# References

[1] Michael J. Angilletta, Peter H. Niewiarowski, and Carlos A. Navas. The evolution of thermal physiology in ectotherms. *Journal of Thermal Biology*, 27(4):249–268, 2002.

[2] Jean-Francois Briere, Pascale Pracros, Alain-Yves Le Roux, and Jean-Sebastien Pierre. A novel rate model of temperature-dependent development for arthropods. *Environmental Entomology*, 28(1):22–29, 1999.

[3] Kenneth P. Burnham and David R. Anderson. Multimodel inference. *Sociological Methods & Research*, 33(2):261–304, 2004.

[4] Kenneth P. Burnham, David R. Anderson, and Kathryn P. Huyvaert. Aic model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1):23–35, 2010.

[5] J. P. DeLong, J. P. Gibert, T. M. Luhring, G. Bachman, B. Reed, A. Neyer, and K. L. Montooth. The combined effects of reactant kinetics and enzyme stability explain the temperature dependence of metabolic rates. *Ecology and Evolution*, 7(11):3940–3950, 2017.

[6] Fred S. Guthery, Kenneth P. Burnham, and David R. Anderson. Model selection and multimodel inference: A practical information-theoretic approach. *The Journal of Wildlife Management*, 67(3):655, 2003.

[7] Raymond B. Huey and Joel G. Kingsolver. Evolution of thermal sensitivity of ectotherm performance, 1989.

[8] Jerald B. Johnson and Kristian S. Omland. Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19(2):101–108, 2004.

[9] Dimitrios Georgios Kontopoulos, Bernardo García-Carreras, Sofía Sal, Thomas P. Smith, and Samraat Pawar. Use and misuse of temperature normalisation in meta-analyses of thermal responses of biological traits. *PeerJ*, 6:e4363, 2018.

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