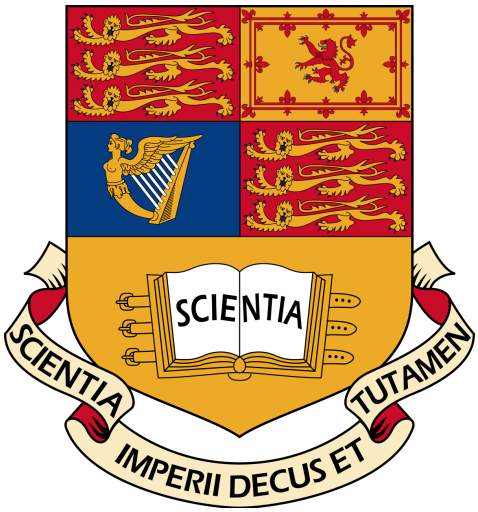
INVESTIGATING THE USE OF PHENOMENOLOGICAL AND MECHANISTIC MODELS TO FIT THE THERMAL RESPONSES OF METABOLIC TRAITS

Eva Linehan



Computational Methods in Ecology and Evolution

Miniproject

Imperial College London

March 2019

WC 3091

Contents

1 Abstract

The effect of temperature on biological processes permeates across all systems and levels of organization. To understand this effect, the thermal response of metabolic traits is often described using temperature performance curves (TPCs). In this study, 1,577 TPCs were analyzed across a large dataset of 397 species within 8 Kingdoms, demonstrating 18 different standardized trait values to describe respiration, growth and photosynthesis. Mechanistic and phenomenological models were fit to the data and compared to investigate which model best described the observed effect of temperature on trait performance. Overall the cubic model was identified as having the best fit, converging to all data and most frequently being selected as the best model for each TPC. The high temperature deactivation Schoolfield model was the best performing mechanistic model with the highest parsimony among the selected set of model. This facilitated the biological interpretation of results.

2 Introduction

It has been argued that life is a response to the thermodynamic requirements of dissipating systems, providing criteria for evaluating growth and development across a range of biological systems (Schneider & Kay, 2007). The importance of thermodynamic imperatives in fluctuating environments and in ecosystem maintenance cannot be understated in this present climate. The most recent IPCC 2018 report (O.Hoegh-Guldberg et al., 2018) identified the investigation of thermal responses across terrestrial, coastal and oceanic ecosystems as a crucial gap in our knowledge. To understand these thermal responses would be to understand the underlying mechanistic processes by which organisms react to Earth's rapidly changing thermal landscape. To observe the profound effects of temperature on biological functions at all levels of organization, temperature performance curves are commonly used.

Temperature performance curves (TPCs) incorporate temperature tolerance and temperature-dependent effects on performance from the whole-organism level, including growth and metabolic rate, to the underlying physiological level which include functions such as enzyme activity (Fangue, Healy, & Schulte, 2011)(Dell, Pawar, & Savage, 2013). The effects of tempera-

ture on performance traits, illustrated by TPCs, follow a general trend with three distinct phases: (1) a growth phase with trait performance increasing with temperature; (2) a peak or thermal optimum at the highest trait value; followed by (3) a sharp decline at higher temperatures (Schulte, 2015a). This unimodal, left skewed distribution has been described by both phenomenological and mechanistic models in order to capture the general features observed for trait values across a range of temperatures. There is currently no one general model favoured over another which may be explained by the fundamental differences between biological responses among taxa (Dell, Pawar, & Savage, 2011)(Low-Décarie et al., 2017). An ideal generalized model may be hard to derive as although complex models often perform poorly in relation to a more simple counterpart, they can be improved when wider temperature ranges are applied (Quinn, 2017). Phenomenological models, often used to predict thermal responses, lack any meaningful parameter interpretation and as a result are flexible tools that can be used to model any species or fitness component. The unknown suitability of such models in different conditions may lead to underestimation (Martin et al., 2017). In contrast, mechanistic models are used to explain the processes underlying phenomena in empirical data and are established on a theoretical basis. The mechanistic foundation for model predictions provides a useful criterion in a range of applications, including conservation assessments (Bernardo & Spotila, 2006).

Both phenomenological and mechanistic models were compared in this study to analyze the effect of temperature on trait performance. Phenomenological models included; a cubic polynomial which is unimodal and asymmetrical in shape as well as Briere's model that accounts for upper and lower temperature thresholds (Pracros, Briere, Le Roux, & Pierre, 1999). The mechanistic models chosen were variations of the Schoolfield-Sharpe model that consider reversible enzyme denaturation at high temperatures (S1), low temperatures (S2) or both (S3). Formulated from the Sharpe and DeMichele model which was unsuitable for non-linear regression, three new thermodynamic parameters were introduced which have a more intuitive biological interpretation: (1) a development rate at a reference temperature that assumes no enzyme activity; (2) a temperature at which enzyme activity is half low temperature inactive; and (3) a temperature at which enzyme activity is half high temperature

inactive. (Schoolfield, Sharpe, & Magnuson, 1981a).

Relatively few studies have measured performance at a large scale to distinguish between underlying shpes and phenomena (Dell et al., 2011) (Dowd, King, & Denny, 2015). The purpose of this study was to conduct a broad-scale comparative analysis of the performance of phenomenological and mechanistic models on thermal response data.

3 Methods

3.1 Data

Metabolic trait data was derived from the BioTraits dataset accumulated by Dell *et al.,* (2013) . This included intraspecific temperature responses for growth, respiration and photosynthesis rates in plants and bacteria from terrestrial and aquatic environments. TPCs with positive and non zero trait values were selected and those with 6 or more observations were included in the analysis. A minimum of 6 data points was chosen in order to successfully fit the full Schoolfield model (S3) which requires at least 6 observations. A total of 1,577 TPCs were fitted with both phenomenological and mechanistic models to assess the effect of temperature on biological trait values. The data for analysis were filtered and exported using Python 3.6.7.

3.2 Model fitting

3.2.1 Phenomenological Models

Two different phenomenological models were used in this study. A generic cubic polynomial model was fit to each TPC using Ordinary Least Squares as it is often used to describe unimodal data.

$$\beta = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta_3 T^3 \tag{1}$$

β, is the trait of interest (respiration, growth or photosynthesis), T is temperature (°C) and β_0, β_1 , β_2 and β_3 are coefficients of the function lacking any mechanistic interpretation.

Briere's model is a 3 parameter model that describes the non-linear relationship of developmental rates for insect species (Pracros et al., 1999). This empirical function accounts

for upper and lower temperature thresholds, asymmetry about the optimum temperature, presence of an inflection point and a sharp decline in developmental rate above optimum temperature.

$$\beta = \beta_0 T(T - T_0)\sqrt{T_m - T} \tag{2}$$

β represents the trait value where T_0 is the minimum feasible temperature (°C) and T_m is the maximum temperature (°C) that a trait can withstand before going to 0. β_0 is the normalization constant and was initialized as 1 for TPC fitting. T_0 was estimated as the minimum temperature and T_m as the maximum temperature recorded for each TPC.

3.2.2 Mechanistic Models

The mechanistic models fitted were variations of the Schoolfield model based on thermodynamics and enzyme kinetics (Schoolfield et al., 1981a). The full model (S3) contains 6 parameters including both high and low deactivation energy.

$$\beta = \frac{\beta_0 e^{\frac{-E}{k}(\frac{1}{T} - \frac{1}{278.15})}}{1 + e^{\frac{E_l}{k}(\frac{1}{T_l} - \frac{1}{T})} + e^{\frac{E_h}{k}(\frac{1}{T_h} - \frac{1}{T})}} \tag{3}$$

k is the Boltzmann constant (8.617 x 10^-5 eV x K - 1). β represents the value of the trait at a given temperature (T) in Kelvin (K = °C + 273.15). β_0 is the trait value that corresponded to the temperature closest to 278.15 K (5°C), controlling the vertical offset of the curve. E is the activation energy (eV) and controls the rise of the curve up to the peak in the "normal operating range" for the enzyme. E_l is the enzyme's low-temperature de-activation energy (eV) which controls the behavior of the enzyme (and the curve) at very low temperatures while E_h is the enzyme's high-temperature de-activation energy (eV), controlling the behavior of the enzyme at very high temperatures. T_l is the temperature at which the enzyme is 50% low-temperature deactivated and T_h is the temperature at which the enzyme is 50% high-temperature deactivated. The simplified models are similar to the full model but incorporate either high temperature deactivation (S1, equation 4) or low temperature deactivation (S2,

equation 5):

$$\beta = \frac{\beta_0 e^{\frac{-E}{k}(\frac{1}{T} - \frac{1}{283.15})}}{1 + e^{\frac{E_l}{k}(\frac{1}{T_l} - \frac{1}{T})}} \tag{4}$$

$$\beta = \frac{\beta_0 e^{\frac{-E}{k}(\frac{1}{T} - \frac{1}{283.15})}}{1 + e^{\frac{E_h}{k}(\frac{1}{T_h} - \frac{1}{T})}} \tag{5}$$

β_0 was initialized as the untransformed trait value corresponding to the temperature closest to 278.15 K (5°C). From here, each TPC was divided into two sections either side of the peak. The peak was calculated as the highest trait value and it's corresponding temperature. The first section, or left hand side of the curve, comprised of temperatures and their corresponding trait values below the peak while the second section, or right hand side of the curve, contained values above the peak. For each section, temperature data were multiplied by 1/k and trait values logged. A linear regression was then fit to the data and starting parameter estimates were calculated. For the left hand side of the curve, below the peak, E was calculated as the absolute value of the slope of the line. T_l was recorded as the temperature (transformed from 1/kt to K) from which the mean logged trait value was taken and the 1/kt value calculated from the regression line. E_l was assumed to be half the value of E. For the right hand side of the curve, above the peak, E_h was recorded as the slope of the line and T_h was estimated similar to T_l. For cases in which above or below peak data contained 1 point or were absent, E was initialized as the reported mean value of 0.65 (Dell et al., 2011) and E_h as 3 times E to reflect a sharper slope as trait values decline with increasing temperature. T_l was estimated as the lowest tempertaure and T_h as the maximum temperature.

Apart from the cubic model, all models were fitted using non-linear least squares methods. The nlsLM function was used to incorporate the Levenberg-Marquardt fitting algorithm, returning a vector of weighted residuals whose sum of square was minimized (Elzhov, Mullen, Spiess, & Bolker, 2016). Parameters were bounded and optimized within the function. Estimated starting parameters for all 3 Schoolfield models were fitted and then randomized with a gaussian fluctuation before re-fitting. Each Schoolfield model was fitted to the TPC 20 times, once with the original estimated parameters and 19 times with starting values ran-

domly sampled from a gaussian distribution with a mean of the calculated parameter and standard deviation of 0.05. The convergence rate was recorded as the number of successful fits achieved out of a total of 20 attempts.

3.3 Model Selection

Models with the best fit for each individual TPC were defined as those with the lowest Akaike Information Criterion, ΔAIC_c score. AIC_c, a second order derivative of the original AIC, contains a bias correction term for small sample size and is suggested as an appropriate selection tool when the number of parameters exceed n/40, n referring to sample size (B Johnson & Omland, 2004). AIC or AIC_c is often used to find the best approximating model to the unknown data as it accounts for the sum of squares, goodness-of-fit measure, and varying numbers of parameters (Burnham & Anderson, 2002). The original AIC equation, used in calculating AIC_c is as follows;

$$AIC = -2log(\mathcal{L}(\hat{\theta} \mid y)) + 2k \tag{6}$$

where $\mathcal{L}(\hat{\theta} \mid y)$ is the log-likelihood at it's maximum point, corresponding to the probability of the data given a model. k is defined as the number of free parameters in the model. An alternative formula to calculate AIC_c , proposed by Hurvich & Tsai (1989), was used to prevent values tending to infinity;

$$AIC_c = AIC + \frac{2k(k+1)(k+2)}{max(n, k+3) - k - 2} \tag{7}$$

k is the number of parameters in the fitted model and n the number of observations. For phenomenological models which were not fit with varying starting parameters, the best model was taken as that with the lowest AIC_c. When fitting the mechanistic models AIC_c, was rescaled to ΔAIC_c using the following equation;

$$\Delta_i = AIC_c i - AIC_c min \tag{8}$$

AIC_c i is the AIC_c for the i^{th} model and $AIC_c min$ is the minimum AIC_c among all

models. The larger the Δ_i, the weaker the model. The best model was considered to be the model in which ΔAIC_c was equal to 0 (minimum ΔAIC score in the TPC) (P Burnham & R Anderson, 2004). Each TPC was then plotted with the best selected fit per model and analyzed visually. An Akaike weight, Wi (AIC_c), was used for model averaging, representing the probability that the model chosen is the best model for the observed data.

$$W_{\mathrm{i}} = \frac{exp(-\Delta_i/2)}{\sum_{r=1}^{R} exp(-\Delta_i/2))} \tag{9}$$

Wi is the weight of evidence in favour of the model. These depend on the full set of models and must sum to 1. R in the denominator incorporates the summation of all AIC_c weights to generate this ratio.

3.4 Computing Languages

Python, R and bash were used to facilitate data wrangling, analysis, plotting and project compilation. The original dataset was cleaned and prepared for model fitting using Python 3.6.7 which included the NumPy (Travis E, 2016) and pandas (McKinney, 2011) libraries.

R 3.4.4 (R Core Team, 2018) was used for model fitting. Standard polynomial regression facilitated ordinary least squares fitting while non-linear least squares fitting was conducted using nlsLM from the minpack.lm package (Elzhov et al., 2016). Figure plotting packages included ggplot2 (Wickham, 2016), xtable (Dahl, Scott, Roosen, Magnusson, & Swinton, 2018), dplyr (Wickham, François, Henry, & Müller, 2019), plyr (Wickham, 2011), gridExtra (Auguie, 2017) and reshape2 (Wickham, 2007).

Bash 4.4.19 was used to compile the project into a reproducible workflow with the final report converted from LaTeXto pdf.

4 Results

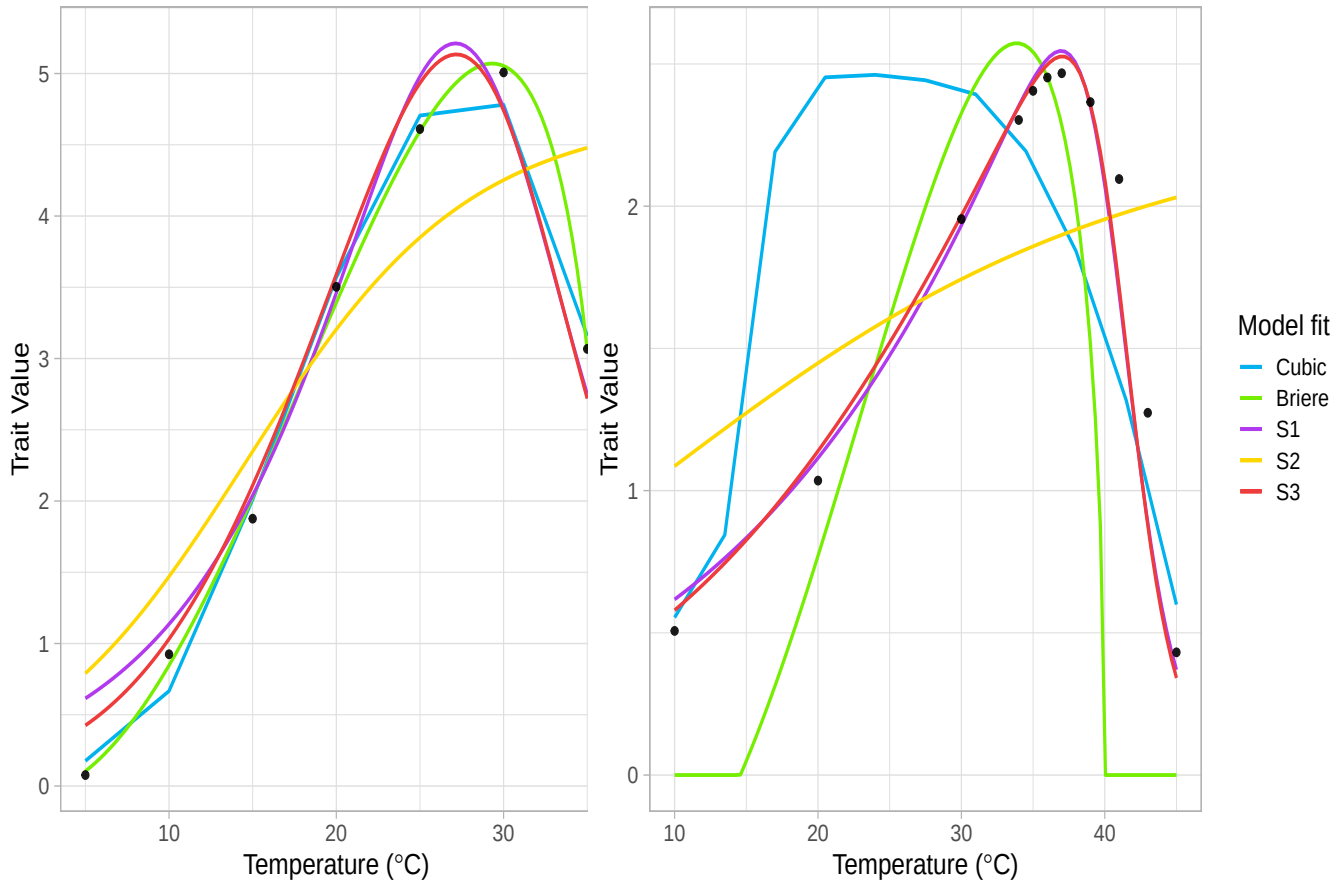
A total of 1,577 TPCs were analyzed following data manipulation. The sample size per TPC did not follow a normal distrbution, ranging from 6 to 637 samples with a median value of 8. 75% of curves contained 13 values or less (see Appendices; Figure 1). Both phenomenological

and mechanistic models fit the majority of the data, as seen in table 1, with the Cubic fitting more curves overall. Among the mechanistic models, those simplified with fewer parameters (S1 and S2) achieved a greater number of fits. Convergence was defined as the proportion of fits out of 20 attempts for each Schoolfield model. The high energy deactivation Schoolfield model (S1) converged more on average compared to the full Schoolfield model which had the lowest convergence rate.

Table 1: Number of successful model fits and proportion per model type for each Temperature Performance Curve across the entire dataset. Convergence rate, average number of successful fits out of 20 attempts, also included for all Schoolfield models (n = 1577).

	Cubic	Briere	S1	S2	S3
Number	1577	1368	1422	1431	1275
Proportion	100%	87%	90%	91%	81%
Convergence	-	-	77%	75%	68%

Upon visual inspection of TPCs, mechanistic models S1 and S3, often demonstrated a good fit and successfully captured the gradual rise and sharp fall of full TPC curves. The low temperature deactivation Schoolfield model (S2) was generally insufficient at fitting to the observed data. Briere was often suitable for full TPCs but did not always fit curves outside of this trend. Both Briere and the Schoolfield models were not as flexible as the Cubic model which fit to more TPCs, often demonstrating a roughly good fit. Figure 1, illustrates the range of fits achieved for a full TPC. Table 2 confirms that Briere provided the best fit for subfigure (a) followed by the Cubic model and S1 which had the lowest ΔAIC_c out of the mechanistic models. S1 was the best fitting model for subfigure (b) also reflected in the accompanying ΔAIC_c. According the table 2, Cubic would be the next best model which visually, is inaccurate.



(a) Green Seaweed *Caulerpa serrulata* (b) Proteobacteria *Escherichia colia*

Figure 1: Two typical TPCs demonstrating all 5 models fit to observed data. a) All models, except for S2, captured the data well, particularly Briere. b) Only the mechanistic models, particularly S1 and S3, come close to an optimal fit.

Table 2: ΔAIC_c results for all models fitted to the TPCs in Figure 1.

TPC	Cubic	Briere	S1	S2	S3
a) Green Seaweed	21.73	-0.59	27.26	44.19	65.78
b) Proteobacteria	10.75	38.59	-4.62	43.68	23.11

Model performance was assessed by evaluating the proportion of best fitting models, according to their ΔAIC_c value, across the entire dataset and between groups. This information criterion, chosen for it's bias correction term towards sample size, demonstrated a relatively

consistent range of scores with increasing sample size (see Appendices; Figure 2). In figure 2 below, it is evident that the Cubic model was selected most frequently as the best fitting model, followed by the high temperature deactivation Schoolfield model (S1), Briere and the remainder of the Schoolfield models. The low temperature deactivation Schoolfield model had the lowest number of best fits within the overall dataset.

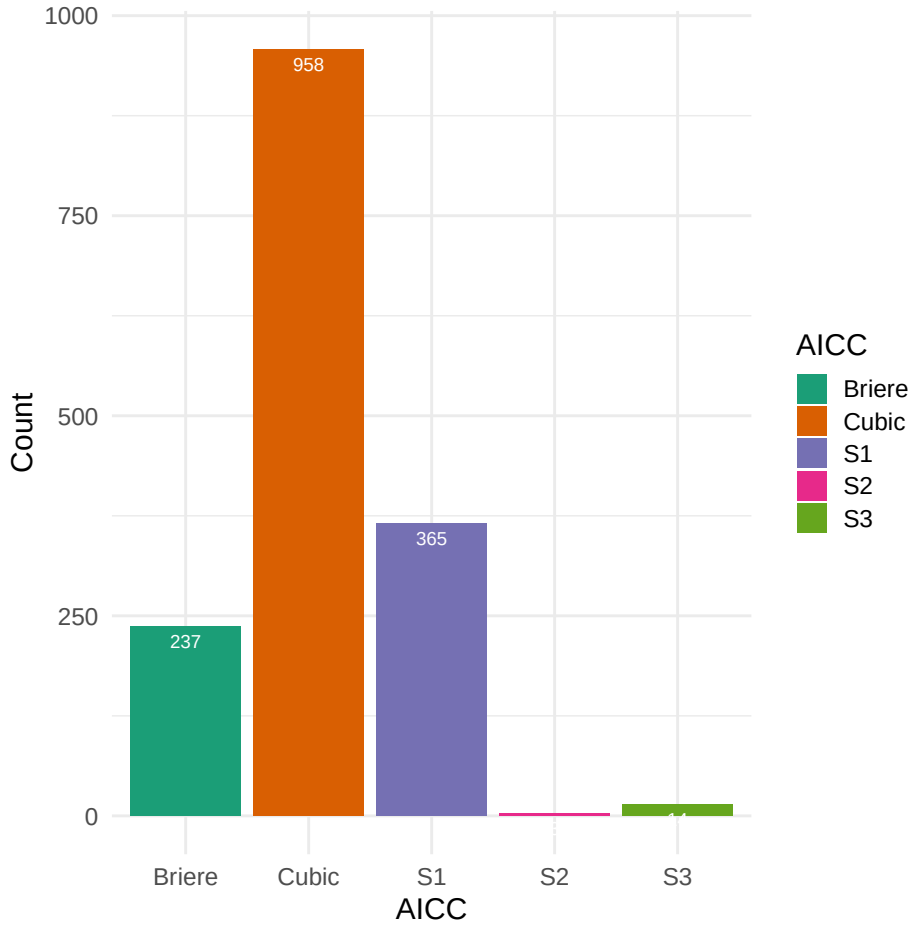


Figure 2: Number of times each model was selected as the best model per TPC, according to ΔAIC_c score, within the entire dataset (n = 1577).

Model performance was also assessed between different trait values which were grouped under the following categories; Photosynthesis, Respiration and Growth. From figure 3, it is evident that the cubic model comprised of the largest proportion of best fits. The poorest performing models, S2 and S3, were only fit to growth and respiration data.

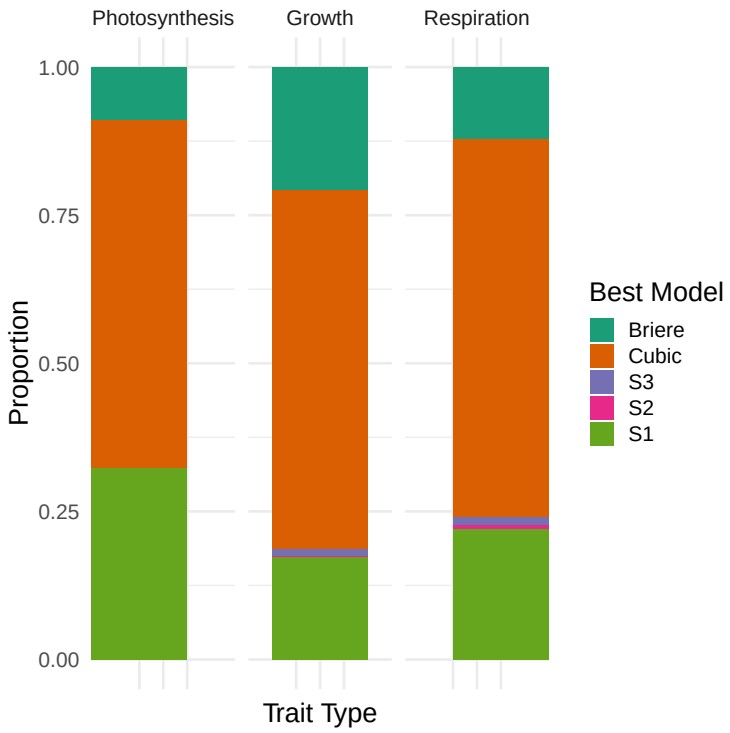


Figure 3: Proportion of best models selected for each trait type summarized into relative categories; Photosynthesis, Growth and Respiration. Best model was selected according to ΔAIC_c score, within the entire dataset (n = 1577).

In figure 4, the general pattern of model performance between kingdoms of consumers was observed. The kingdoms Archae and Protozoa strongly favoured phenomenological models with Cubic and Briere selected as the best models respectively. Protozoa was the only kingom in which no mechanistic models provided the best fit. Apart from the high temperature deactivation Schoolfield model across all other kingdoms, the best fits for the low temperature deactivation and full Schoolfield model were present only in Bacteria and Plantae.

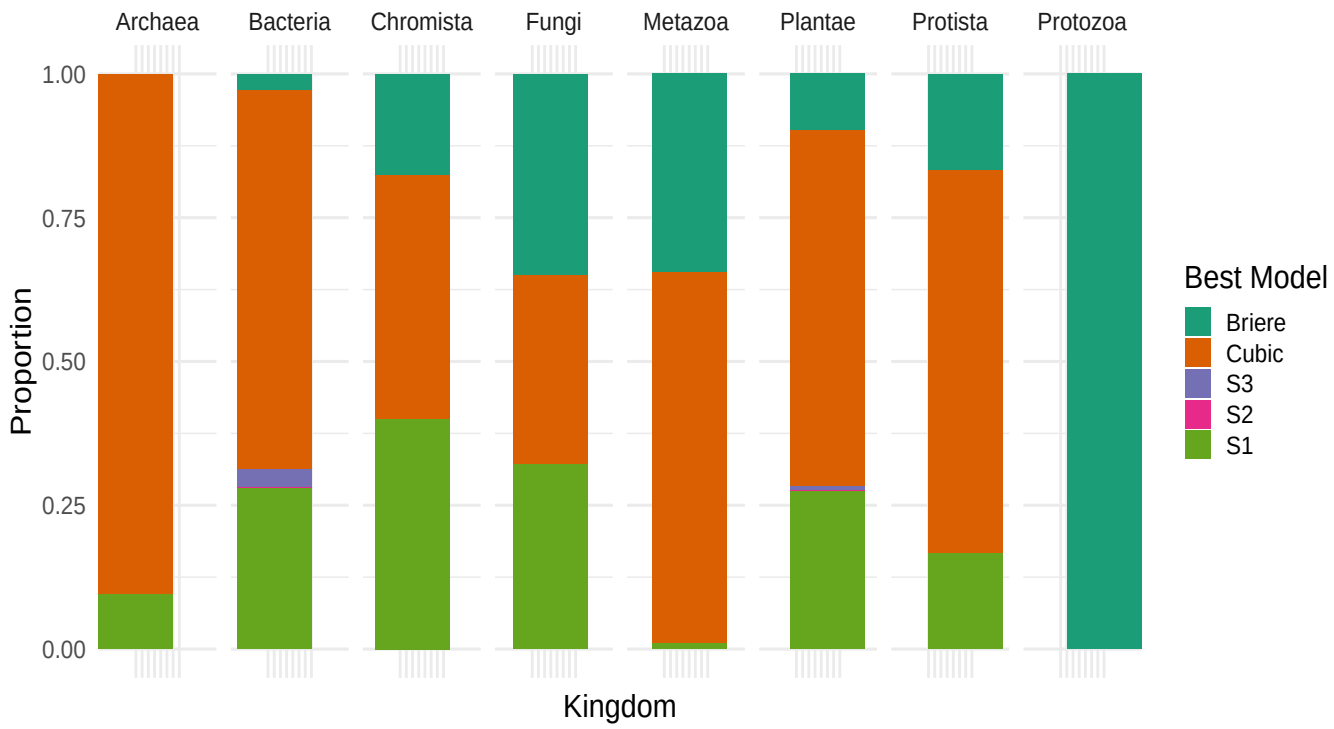


Figure 4: Proportion of best models selected for each Kingdom of consumers. Best model was selected according to ΔAIC_c score, among consumers that had a Kingdom specified (n = 1576).

The evidence for the strength of fit for each model was calculated by categorizing ΔAIC_c scores according to guidelines adapted by Burnham & Anderson (2004); models with $\Delta AIC_c \leq 2$ show substantial support; those in which $4 \leq \Delta AIC_c \leq 7$ have considerably less support; and those with ΔAIC_c >10 have no support. In table 3, strength of fit was investigated for TPCs in which all models fit the observed data. It is clear that the Cubic model out-performed other models two to three times over as it had the strongest supporting evidence. The high temperature deactivation Schoolfield model comprised of a relatively large proportion of strongly supported models in comparison to the remaining models. The low temperature deactivation energy and full Schoolfield models (S2 and S3) had little to no support. This is reflected in AIC_c weights (figure 5) with the cubic model possessing a greater proportion of higher weights, followed by the simplified Schoolfield high energy deactivation model (S1). Both models demonstrate a higher probability of being the model that best describes the data.

Table 3: ΔAIC_c scores for TPCs in which all models converged. Scores fall into respective categories as per the recommended guidelines. Categories represent strength of fit, with lower ΔAIC_c indicative of supportive evidence in favour of the model (n = 1156).

Model	$\Delta < 2$	$2 < \Delta \leq 4$	$4 < \Delta \leq 7$	$7 < \Delta \leq 10$	$\Delta > 10$
Cubic	1099	88	119	83	188
Briere	275	48	78	60	907
S1	520	125	200	164	413
S2	49	51	91	110	1130
S3	14	4	5	8	1244

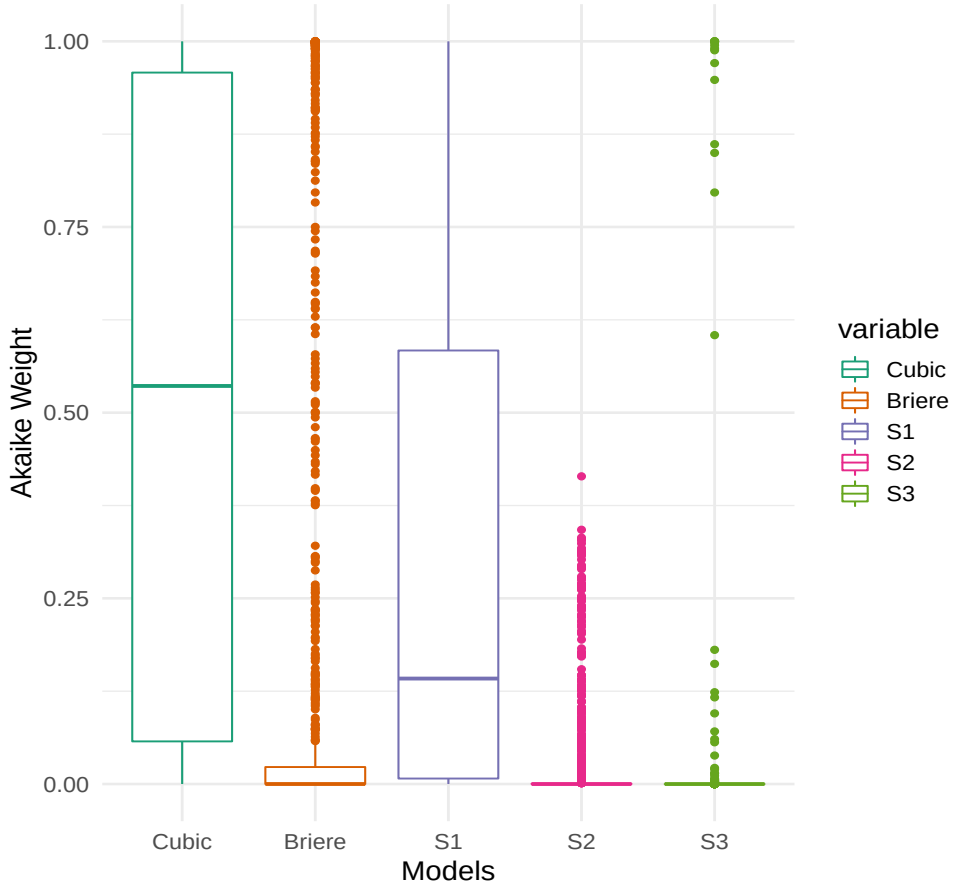


Figure 5: Distribution of Akaike weights for each model. Weightages were calculated for TPCs in which all models converged (n = 1156).

5 Discussion

Which model is 'best'? At first, the answer may seem straightforward but on further inspection this is not the case, comparing alternative models that make different simplifications with a common biological assumption (Levins, 1966). The goal of non-linear regression is to minimize the sum-of-squares and so it appears that the model with the smallest value is best. However, this is often not applicable when models have varying numbers of parameters, and therefore different inflection points. Any method to compare a simple model with a more complicated one must balance the decrease in sum-of-squares with the increase in the number of parameters. From analyzing the data, it is clear that the Cubic model was generally the better fitting model within the dataset and between Trait types and Kingdoms. This is possibly due to model flexibility as phenomenological models are not constrained by parameter assumptions. However, this provides little information about the variation in thermal physiology among organisms and so prevents any biological interpretation.

Mechanistic models are preferable in order to understand the thermal ecology and metabolic adaptations of organisms as well as generate testable predictons for future studies (Peek, Russek-Cohen, D Wait, & N Forseth, 2002)(Schulte, 2015b) (Martin et al., 2017) (DeLong et al., 2017). The best fitting mechanistic model was the Schoolfield high temperature deactivation model (S1) which may be a consequence of experimental limitations (Schoolfield, Sharpe, & Magnuson, 1981b), for example if organisms have only been studied over part of the temperature spectrum. Compared to S2 and S3, which both contain low temperature deactivation parameters, S1 performed significantly better. This may be due TPCs containing higher temperature values and/or to the fact that low temperature deactivation is difficult to detect (Pawar, Dell, Savage, & Knies, 2016).

Another plausible explanation for observed model performance lies in the calculated parameter estimates. General assumptions applied to the entire dataset can lead to over or underestimated values for individual TPCs. For example, the normalization constant, β_0, which standardizes rate performance across groups, is particularly susceptible to overestimating trait values at the selected reference temperature (Kontopoulos, Garcia-Carreras, Sal, Smith, & Pawar, 2018). Parameters such as T_h and T_l were also observed to be unrealistic in

certain datasets. The mean activation energy for each model was close to the reported activation energy of 0.65 eV with S1 having the closest mean value. Parameters were bound and optimized using the Levenberg-Marquardt algorithm as part of nlsLM function which is the virtual standard in optimization. However this approach is not immune to faults with a generally slow convergence rate and probability of getting lost in parameter space (K. Transtrum & Sethna, 2012).

Model selection was chosen on the basis of fit and complexity to compare all models simultaneously using AIC. Instead of the hypothesis testing approach, associated with the likelihood ratio test, or R^2 measure of fit which is inappropriate for non-liner regression (Spiess & Neumeyer, 2010), AIC combines the Kullback-Leiber distance, measure of discrepancy, and Fisher's maximized log-likelihood to select a parsimonious model to analyze empirical data (Akaike, 1998). Even in moderate sample sizes, the second order derivative AIC_c provides substantially better model selections than AIC (Hurvich & Tsai, 1991). One major disadvantage to this approach was that values tended to inifinity when the sample size, n, was equal to the number of parameters, k + 1, which occured for the full Schoolfield model (k = 6). An alternative formula was used to avoid omitting data but is not well explored in the literature.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Auguie, B. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics [Computer software manual]. Retrieved from `https://cran.r-project.org/package=gridExtra`

B Johnson, J., & Omland, K. (2004). Johnson JB, Omland KS.. Model selection in ecology and evolution. Trends Ecol Evol 19: 101-108. *Trends in ecology & evolution*, *19*, 101–108. doi: 10.1016/j.tree.2003.10.013

Bernardo, J., & Spotila, J. R. (2006). Physiological constraints on organismal response to global warming: mechanistic insights from clinally varying populations and implications for assessing endangerment. *Biology Letters*, *2*(1), 135–139. doi: 10.1098/rsbl.2005.0417

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). Springer.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2018). xtable: Export Tables to LaTeX or HTML [Computer software manual]. Retrieved from https://cran.r-project.org/package=xtable

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. Retrieved from https://app.dimensions.ai/details/publication/pub.1026914980 and http://www.pnas.org/content/108/26/10591.full.pdf doi: 10.1073/pnas.1015178108

Dell, A. I., Pawar, S., & Savage, V. M. (2013). The thermal dependence of biological traits. *Ecology*, *94*(5), 1205–1206. Retrieved from https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/12-2060.1 doi: 10.1890/12-2060.1

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.

Dowd, W. W., King, F. A., & Denny, M. W. (2015). Thermal variation, thermal extremes and the physiological performance of individuals. *Journal of Experimental Biology*, *218*(12), 1956–1967. Retrieved from http://jeb.biologists.org/content/218/12/1956 doi: 10.1242/jeb.114926

Elzhov, T. V., Mullen, K. M., Spiess, A.-N., & Bolker, B. (2016). minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds [Computer software manual]. Retrieved from https://cran.r-project.org/package=minpack.lm

Fangue, N. A., Healy, T. M., & Schulte, P. M. (2011). Thermal Performance Curves, Phenotypic Plasticity, and the Time Scales of Temperature Exposure. *Integrative and Comparative Biology*, *51*(5), 691–702. Retrieved from

https://dx.doi.org/10.1093/icb/icr097 doi: 10.1093/icb/icr097

Hurvich, C. M., & Tsai, C.-L. (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, *78*(3), 499–509. Retrieved from https://dx.doi.org/10.1093/biomet/78.3.499 doi: 10.1093/biomet/78.3.499

K. Transtrum, M., & Sethna, J. (2012). Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization.

Kontopoulos, D.-G., Garcia-Carreras, B., Sal, S., Smith, T. P., & Pawar, S. (2018). Use and misuse of temperature normalisation in meta-analyses of thermal responses of biological traits. *PeerJ*, *6*, e4363.

Levins, R. (1966). The strategy of model building in population biology. *Am. Sci.*, *54*, 421–431.

Low-Décarie, E., Boatman, T. G., Bennett, N., Passfield, W., Gavalás-Olea, A., Siegel, P., & Geider, R. J. (2017). Predictions of response to temperature are contingent on model choice and data quality. *Ecology and Evolution*, *7*(23), 10467–10481. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.3576 doi: 10.1002/ece3.3576

Martin, B. T., Pike, A., John, S. N., Hamda, N., Roberts, J., Lindley, S. T., & Danner, E. M. (2017). Phenomenological vs. biophysical models of thermal stress in aquatic eggs. *Ecology letters*, *20 1*, 50–59.

McKinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics. *Proceedings of the 9th Python in Science Conference*.

O.Hoegh-Guldberg, Jacob, Taylor, Bindi, Brown, Camilloni, ... Zhou (2018). Impacts of 1.5°C global warming on natural and human systems. An IPCC Special Report on the impacts of global warming of 1.5ÂřC above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty..

P Burnham, K., & R Anderson, D. (2004). Multimodel Inference: understanding AIC and BIC in Model Selection. *Sociological Methods Research*, *33*, 261–304.

Pawar, S., Dell, A. I., Savage, V. M., & Knies, J. L. (2016). Real versus artificial variation

in the thermal sensitivity of biological traits. *The American Naturalist*, *187*(2), E41—-E52.

Peek, M., Russek-Cohen, E., D Wait, A., & N Forseth, I. (2002). Physiological response curve analysis using nonlinear mixed models. *Oecologia*, *132*, 175–180. doi: 10.1007/s00442-002-0954-0

Pracros, P., Briere, J.-F., Le Roux, A.-Y., & Pierre, J.-S. (1999). A Novel Rate Model of Temperature-Dependent Development for Arthropods. *Environmental Entomology*, *28*(1), 22–29. Retrieved from `https://dx.doi.org/10.1093/ee/28.1.22` doi: 10.1093/ee/28.1.22

Quinn, B. K. (2017). A critical review of the use and performance of different function types for modeling temperature-dependent development of arthropod larvae. *Journal of thermal biology*, *63*, 65–77.

R Core Team. (2018). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.r-project.org/`

Schneider, E. D., & Kay, J. J. (2007). Order from Disorder : The Thermodynamics of Complexity in Biology..

Schoolfield, R. M., Sharpe, P. J., & Magnuson, C. E. (1981a). Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. *Journal of Theoretical Biology*, *88*(4), 719–731. doi: 10.1016/0022-5193(81)90246-0

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

Schulte, P. M. (2015a). The effects of temperature on aerobic metabolism: towards a mechanistic understanding of the responses of ectotherms to a changing environment. *Journal of Experimental Biology*, *218*(12), 1856–1866. Retrieved from `http://jeb.biologists.org/cgi/doi/10.1242/jeb.118851` doi: 10.1242/jeb.118851

Schulte, P. M. (2015b). The effects of temperature on aerobic metabolism: towards a mechanistic understanding of the responses of ectotherms to a changing environment. *The*

Journal of experimental biology, *218 Pt 12*, 1856–1866.

Spiess, A.-N., & Neumeyer, N. (2010). An evaluation of R 2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC pharmacology*, *10*(1), 6.

Travis E, O. (2016). *A guide to NumPy.* Trelgol Publishing.

Wickham, H. (2007). Reshaping Data with the {reshape} Package. *Journal of Statistical Software*, *21*(12), 1–20. Retrieved from `http://www.jstatsoft.org/v21/i12/`

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from `http://www.jstatsoft.org/v40/i01/`

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. Retrieved from `http://ggplot2.org`

Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation [Computer software manual]. Retrieved from `https://cran.r-project.org/package=dplyr`

6 Appendices

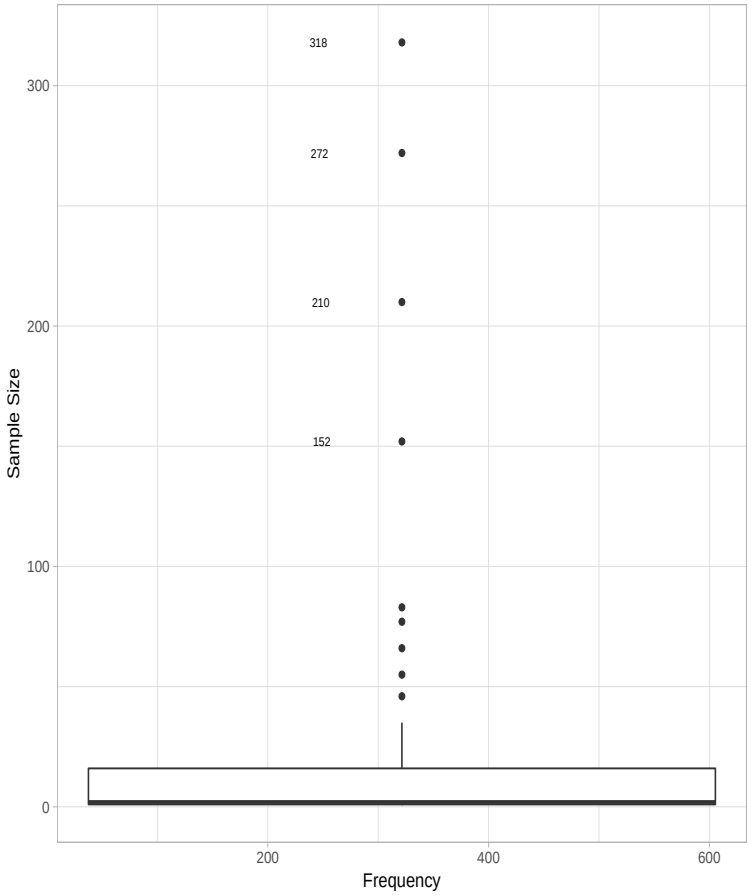


Figure 1: Boxlplot illustrating the range of observations per Thermal Performance Curve. Outliers are labelled

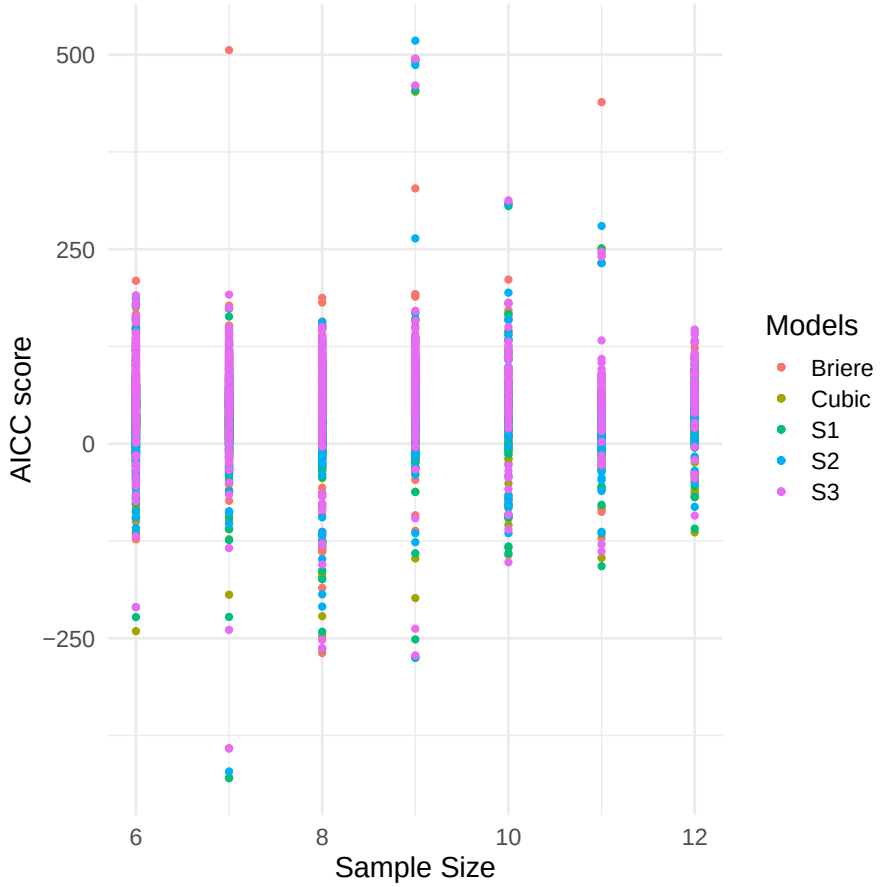


Figure 2: Plot of AIC_c scores with sample size, coloured according to model type

List of Figures

List of Tables