Evaluating Thermodynamic Models Performance for Predicting Microbial Responses to Temperature

Yutao ZHOU

Affiliation: Imperial College London

Email: lz723@ic.ac.uk

Supervisor

Sammraat Pawar, Imperial College London

Declaration: All the data and model function used in these thesis are based on the previous work of my supervisor and his collaborators in Pawarlab. I adapted these models to suit the specific requirements of this study and conducted all the analyses independently. I was solely responsible for the data processing and cleaning required for this project. The analytical framework presented in this project was primarily developed by me. My supervisor played a supportive role, offering advice and feedback on the methods and interpretation of results. The final analysis, however, was conducted and refined independently by me.

A thesis submitted in partial fulfilment of th e requirements for the degree of Master of Science/Research at Imperial College London submitted for the Msc Computational Methods in Ecology and evolution

Date: 22 Aug 2024

Abstract

Temperature constitutes a primary determinant influencing microorganism metabolic pathways, including bacteria, fungi, and phytoplankton, thus modulating their functions in the cycle of nutrients and the dynamics of the ecosystem. Understanding how these organisms respond to temperature changes is essential to predict the impacts of climate change on microbial communities and broader ecological processes. In this study, thirteen thermodynamic models were applied to 987 unique thermal performance curves (TPCs) derived from an exceptionally comprehensive bacterial dataset. This dataset, distinguished by its comprehensive scope and diversity, encompasses a wide range of biological processes including bacterial growth rate, enzymatic activity, gross photosynthesis rate, respiration rate, and fungal canopy growth rate. The principal aim was to ascertain the thermodynamic model that most precisely characterises the TPCs across these diverse of biological functionalities. The results show that models predicated on the temperature-induced denaturation of enzymes consistently outperformed those based on alternative molecular dynamic theories. One such model exhibited the best-fitting performance across all datasets, with the exception of the gross photosynthesis rate. For the bacterial growth rate data, two other models demonstrated fitting probabilities comparable to the top-performing model. Additionally, a different set of models provided the best fit for the gross photosynthesis data, outperforming others in this specific category. These findings suggest that model selection should be tailored to specific microbial processes, as these could improve the accuracy of predictions in the dynamics of the microbial community and improve our understanding of ecosystem functions and stability.

Introduction

Microorganisms are integral components of ecosystems, playing a crucial role in cycling of nutrients, the flow of energy, and the maintenance of environmental balance (Falkowski et al., 2008). Traits are distinct and measurable attributes of microorganisms that include morphological, anatomical, physiological, biochemical, and phenological characteristics (Nock et al., 2016). These properties are typically assessed at the individual level and used for comparative analysis between different species. The traits of bacteria, fungi, and phytoplankton are sensitive to temperature because these organisms are highly adapted to their specific thermal environments, and temperature fluctuations can directly affect their physiological processes and ecological functions (Cébron et al., 2021; Thomas et al., 2016). For example, an increase in temperature can enhance the metabolic rates of *Escherichia coli* up to a certain threshold, beyond which their activity could decline sharply, leading to reduced efficiency in cycling of nutrients (Blaustein et al., 2013). However, thermophilic bacteria like *Thermus aquaticus* can maintain or even increase their metabolic activity at higher temperatures, thriving in conditions that would inhibit other species (Ray et al., 1971). Studies on the temperature sensitivity of these traits are crucial for predicting impacts of climate change on microbial communities, ecosystem functions, and global biogeochemical cycles (Thomas et al., 2016). Recognising these sensitivities also offers valuable insights into the evolutionary pressures and adaptations that influence the distribution and resilience of these organisms across various environments.

The Temperature Performance Curve (TPC) is a prominent concept in the study of biological traits, illustrating how an organism's performance metrics, such as growth rate, enzymatic activity, or reproductive success, vary with temperature (Figure 1). These curves typically show a distinct shape, reaching a maximum at an optimal temperature (E) and declining at temperatures above and below this point (E_{D}) (Pawar et al., 2016). Modelling TPCs allows researchers to predict how organisms respond to temperature variations, which is essential for understanding ecological and evolutionary dynamics. Numerous models incorporating thermodynamic parameters have been developed to describe TPCs. For example, some of these models incorporate parameters like activation energy, enthalpy, and entropy change, which serving as physiological constraints on biological rates. To achieve more biologically realistic parameter estimates, it is essential to impose distinct constraints on parameters including the activation energy of enzymatic reactions and protein denaturation, ensuring that they do not converge to the same value when fitting the models to datasets (Noll et al., 2020).

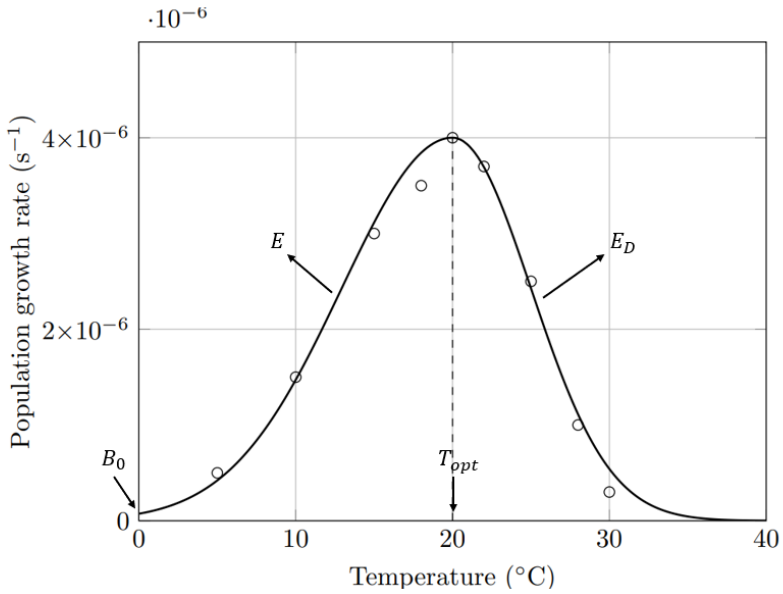

Figure 1: ***An example of the application of the four-parameter Sharpe-Schoolfield model to the TPC of Salmonella enterica.*** *T_{opt} is the optimal performance temperature, which is performance peak. B_0 is the rate performance at reference temperature (normally $< T_{opt}$), E is the increasing phase, E_D is the decling phase.*

Thermodynamic models are mathematical frameworks that describe and predict the behaviour of biological systems by quantifying the relationships between energy transformations, entropy changes, and the equilibrium states of physical and chemical processes (Vanchurin et al., 2022). These models are grounded in the fundamental laws of thermodynamics and are particularly useful in elucidating the influences of temperature on biological rates and processes, such as metabolic reactions, enzyme kinetics, and organismal performance across different environmental conditions (Steane, 2016). In biology, thermodynamic modeling has played a crucial role in indicating how living organisms manage their energy to maintain life processes. This approach traces its roots back to Erwin S. Bauer's seminal work in 1935, where he proposed the "Universal Law of Biology," signifying that living systems exist in a perpetual state of non-equilibrium, performing continuous work contrary to the equilibrium dictated by physical and chemical laws (Elek and Müller, 2024). Non-equilibrium thermodynamic models enable us to understand how microbes dynamically allocate energy and resources, adapt to environmental stressors, and maintain stability under varying conditions. These models are particularly useful in predicting how microorganisms adjust their metabolic processes in response to temperature fluctuations, shifts in nutrient availability, and other environmental changes.

In this paper, I have selected thirteen non-equilibrium thermodynamic models to evaluate their applicability across a diverse range of microbial datasets, including those from bacteria, fungi, and phytoplankton. The inclusion of fungi and phytoplankton data is crucial for a more comprehensive understanding of microbial thermodynamics (Brown et al., 2004). Although early models by Hinshelwood (1946), Johnson and

Lewin (1946), Schoolfield et al. (1981), and Ratkowsky et al. (2005) have focused primarily on bacterial growth rate data, their applicability to other microbial groups remains underexplored. As principal decomposers and primary producers, fungi and phytoplankton hold essential roles in ecosystems. Their metabolic reactions to temperature changes are crucial to understanding larger ecological processes, particularly in the context of global climate change (Falkowski et al., 2008). The models developed by Hobbs et al. (2013) and Ritchie (2018), which incorporate enzyme kinetic data, offer a more detailed perspective at the molecular level, potentially applicable to these diverse organisms. Incorporating data sets from fungi and phytoplankton, this research seeks to evaluate the robustness and flexibility of these models in various microbial groups.

There is a scarcity of publications that have compared different microorganism datasets. Hence, extending the use of thermodynamic models to include fungi and phytoplankton is crucial (Popovic, 2019). These organisms have distinct ecological roles, differing significantly from bacteria in their metabolic pathways and reactions to environmental changes. By incorporating data from a wider array of microbial life, we can better assess the adaptability and precision of these models across various scenarios, preventing them from being overly tailored to bacterial systems. To guide this comprehensive analysis, several key questions will frame the research:

1. Which model demonstrated the best-fitting performance across all microorganism datasets (including bacteria, fungi and phytoplankton)?
2. How did models performed in the specific categories of datasets?
3. How do models based on temperature-induced enzyme denaturation compare in performance to those based on molecular dynamic theories?

Methods

To address these questions, we assembled a comprehensive dataset that includes key biological traits from bacteria, fungi, and phytoplankton. The model fitting and results generation were conducted using RStudio (version 2024.04.0). Subsequently, the performance of the model was compared by evaluating the AIC weights and its heat map.

Data compilation

The data was sourced from a microorganism trait dataset developed by Kontopoulos et al. (2023). Data on bacterial population growth rate, enzyme activity levels, fungal canopy growth rate, and respiration and photosynthetic rates in phytoplankton were extracted for model fitting (Figure 2). The models used in this study are based mainly on bacterial growth and enzyme activity data sets. To provide a more comprehensive evaluation, I have expanded the scope by incorporating additional data on fungal canopy growth rate,

respiration rates, and gross photosynthetic rates—an approach that has not been previously undertaken. This expanded dataset allows for a more thorough comparison of the models, offering broader insights into their applicability and performance across different biological contexts. By including these diverse datasets, the study seeks to identify the best-fitting models and explore how these models can be specifically applied to explain the biological processes represented in each dataset.

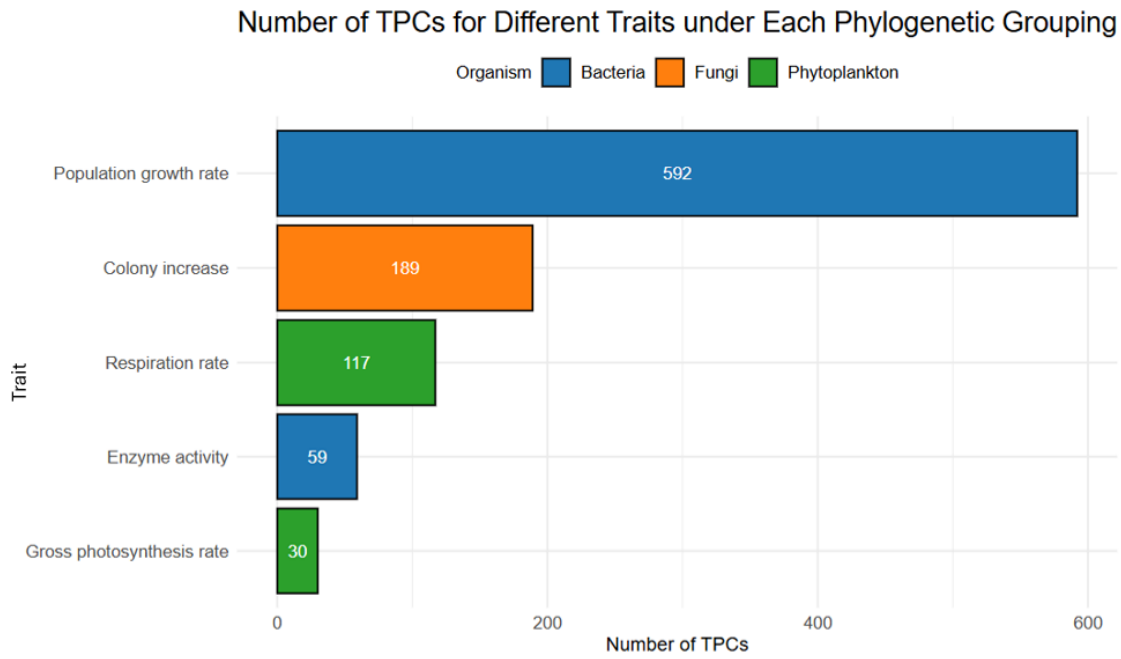

Figure 2: ***Number of TPCs through each microorganism group*** *The bar chart displays the number of TPCs for various traits under different phylogenetic groupings. Population Growth Rate, represented by bacteria, has the highest number of TPCs at 592, followed by Colony Increase in fungi with 189 TPCs, Respiration Rate in phytoplankton with 117 TPCs, Enzyme Activity in bacteria with 59 TPCs, and Gross Photosynthesis Rate in phytoplankton with 30 TPCs.*

Models use for fitting

Despite the model mentioned in introduction, these study also added the extended, simplified and simplified-extended version of Johnson-Lewin and Sharpe-Schoolfield. For instance, the Sharpe-Schoolfield model and its variations differ in the number and type of parameters they include, which influence how comprehensively they model thermal performance across temperature ranges. The basic Sharpe-Schoolfield model, developed by Schoolfield et al. (1981), includes six parameters: B_0 (baseline rate at a reference temperature), ΔH_A (activation energy), ΔH_L (low temperature deactivation energy), T_{L50} (temperature at which the rate is half-maximal at low temperatures), ΔH_H (high temperature deactivation energy) and T_{H50} (temperature at which the rate is half-maximal at high temperatures). The extended version adds a seventh parameter, a, which further refines the model by adjusting the asymmetry between the low- and high-temperature effects. The simplified version retains the six original parameters from Schoolfield et al. (1981), simplifying the model for situations where the additional parameter is unnecessary. The simplified-extended model also includes seven parameters, maintaining the comprehensive nature of the extended model while aiming

for computational simplicity when applying it to datasets that span a wide range of temperatures. The relationship between Johnson-Lewin and its other version is similar to that of Sharpe-Schoolfield.

Models fitting

In each TPC fitting, thermodynamic models were applied using the `RTPC` package (Padfield, 2023). The model functions that are not included in this package were derived from the work of Kontopoulos et al. (2023). Initial parameter values were set arbitrarily based on the values proposed by the model authors. Each model was iterated up to 1000 times, with the initial parameter estimates adjusted within a range of 0.5 to 1.5 times the estimated values.The data was logarithmically transformed to align with the non-linear least squares (nls) model used in this analysis. This process stabilizes the variance and improves the distribution of the data, allowing the nls algorithm to converge more efficiently and accurately estimate the model parameters.

Assessing model performance

The model performance was assessed using Akaike Information Criterion (AIC) weights rather than the raw AIC values. AIC weights are particularly useful because they provide a relative measure of model quality, allowing for a straightforward comparison between models. While the AIC value itself quantifies the goodness-of-fit of a single model, it does not offer a direct interpretation of the model's likelihood relative to others in the candidate set. AIC weights, on the other hand, normalise the AIC values across all models, transforming them into probabilities that sum to 1. This makes it easier to interpret which model is most likely to be the best given the data, rather than simply relying on raw AIC values (Johnson and Omland, 2004). Thus, in this study, AIC weights were favored for their ability to offer a more intuitive and comparative assessment of model performance.

Results

Table 1 provides an overview of the structure of the data set used in the study. For bacteria, the dataset includes 592 TPCs for population growth rate and 59 TPCs for enzyme activity. The models performed well in these cases, with high fit rates of 0.65 and 0.70, respectively. In contrast, fungi, measured by the increase in colony diameter, had a lower fit rate of 0.50 in 189 TPCs, suggesting less consistent model accuracy. Phytoplankton showed strong model performance, with fit rates of 0.90 for gross photosynthesis rate and 0.83 for respiration rate, indicating reliable predictions.

The best-fitting model across all dataset

Among the models assessed, the Johnson-Lewin, Simplified Sharpe-Schoolfield, Sharpe-Schoolfield, Simplified Johnson-Lewin, and Ross-Ratkowsky models exhibited strong fitting capabilities across numerous

datasets (Figure 3). These models consistently performed well, as indicated by the proportion of their AIC weights exceeding 0.25 across various biological data types, with values of 0.32, 0.25, 0.25, 0.23, and 0.19 respectively. The John-Lewin model demonstrated a notably superior fit, as indicated by an AIC weight of 25%. This suggests that the John-Lewin model outperforms other models in approximately one-quarter of the dataset (Johnson and Omland, 2004). In comparison, if the probabilities were evenly distributed across the other twelve models, each model would only have a 6.25% chance of being the best model. Although a 25% probability may seem relatively low, it is, in fact, significantly higher than the likelihood of any of the other models being the best, especially considering the outcomes related to respiration rate, fungal activity, and enzyme activity (Figure 4), where the probabilities of other models being the best are far below 75%. It is also worth noting that when all data are considered, the combined probability of models exceeding 25% is actually greater than 1. This indicates that some of the Thermal Performance Curves (TPCs) models have two or more models that can be considered to explain more than 50% of the specific TPC, although this phenomenon was not fully captured in this study. This oversight could be regarded as a limitation of the current research, suggesting the need for future work, possibly involving the use of classifiers or machine learning techniques, to more accurately identify the best models in such scenarios.

Model performance in specific dataset

Besides the consistency noted in Figure 3, the heatmap of individual datasets highlights significant differences in model fitting across various datasets (Figure 4). In particular, for the respiration rate, population growth rate, fungal growth rate, and enzyme activity datasets, the Johnson-Lewin model exhibits high AIC weights over 25% rate of 0.31, 0.31, 0.37, and 0.27 respectively. This suggests that the Johnson-Lewin model is particularly well suited for capturing the dynamics of these biological processes, likely because its structure is more aligned with the specific characteristics of these datasets. Furthermore, in bacterial growth rate data, the Ross-Ratkowsky, Sharpe-Schoolfield and Simplified Sharpe-Schoolfield models show a similar good fit like Johnson-Lewin, with AIC weight proportions of 0.29, 0.30 and 0.31.

In contrast, the analysis of gross photosynthesis rate data indicates a decline in the fit quality for the Johnson-Lewin model, showing a reversal of its previously strong performance. In these cases, the Ross-Ratkowsky and Sharpe-Schoolfield models, along with their simplified versions, demonstrate superior fitting performance compared to the Johnson-Lewin models. This indicates that these models are more effective at modelling the temperature-dependent dynamics in photosynthesis, possibly because of their ability to capture the broader range of responses seen in these processes.

Although the models that consistently performed well were highlighted, the Hinshelwood, Hobbs, and Ritchie models showed particularly strong fits in specific thermal performance curves (TPCs), with AIC

weights exceeding 0.9. These models demonstrated superior explanatory power, with counts of 10, 3, and 3, respectively, indicating that they could explain 90% of the variation in specific TPCs. This is reflected by the distinct "extremely red" points in the heatmap, particularly within datasets related to respiration rate, fungal growth rate, and enzyme activity (Figure 4). Interestingly, despite the Enzyme-Assisted Arrhenius (EAAR) model having the highest fit rate in Table 2, it did not consistently perform well by comparing wtih other models across all datasets.

Comparative Performance of Models Based on Temperature-Induced Enzyme Denaturation Versus Molecular Dynamic Theories

According to the goodness-of-fit data presented in Table 2, the average values for models based on temperature-induced enzyme denaturation and molecular dynamic theories are 0.61 and 0.63, respectively, indicating a close similarity in their overall fit rate. However, when comparing the Top 5 AIC weights models, the molecular dynamic theories models do not perform well overall (Figure 3). While these models demonstrate an excellent fit for certain specific TPCs and even outperform some temperature dependence models (such as Hinshelwood, the extended version of Johnson-Lewin, and Sharpe-Schoolfield), their overall consistency across different datasets is lacking, leading to generally poor performance.

Organism	Parameter Description	Measurement Type	Number of TPCs	Fit Rate
Bacteria	Population growth rate	Rate	592	0.65
	Enzyme activity	Activity level	59	0.70
Fungi	Increase in diameter of a circular colony growing	Rate	189	0.50
Phytoplankton	Gross photosynthesis rate	Rate	30	0.90
	Respiration rate	Rate	117	0.82

Table 1: ***Overview of the dataset structure utilized in the study.*** *This table lists various organisms and the corresponding parameters measured, categorized by measurement type. The " Number of TPC " column indicates the total number of thermal performance curves (TPC) available for each parameter, while the "Fit Rate" column shows the goodness-of-fit for the models applied to these data, represented as a proportion.*

Model Number	Model Name	Number of Parameters	Type of model	Fit Rate
1	Ritchie	4	Molecular Dynamic Theories	0.43
2	Enzyme-Assisted Arrhenius	5	Molecular Dynamic Theories	0.86
3	Hobbs	4	Molecular Dynamic Theories	0.61
4	Ross-Ratkowsky	5	Temperature-Induced Enzyme Denaturation	0.71
5	Simplified Extended Sharpe-Schoolfield	7	Temperature-Induced Enzyme Denaturation	0.63
6	Simplified Sharpe-Schoolfield	6	Temperature-Induced Enzyme Denaturation	0.67
7	Extended Sharpe-Schoolfield	7	Temperature-Induced Enzyme Denaturation	0.63
8	Sharpe-Schoolfield	6	Temperature-Induced Enzyme Denaturation	0.66
9	Simplified Extended Johnson-Lewin	5	Temperature-Induced Enzyme Denaturation	0.62
10	Simplified Johnson-Lewin	4	Temperature-Induced Enzyme Denaturation	0.70
11	Extended Johnson-Lewin	5	Temperature-Induced Enzyme Denaturation	0.60
12	Johnson-Lewin	4	Temperature-Induced Enzyme Denaturation	0.72
13	Hinshelwood	4	Temperature-Induced Enzyme Denaturation	0.17

Table 2: ***Overview of the thermodynamic models used in the study.*** *including the number of parameters for each model and their respective fit rates. The models are listed by their model number and name, with the number of parameters indicating the complexity of each model in capturing the thermal performance data. The "Fit Rate" column reflects the goodness-of-fit for each model, providing an indication of how well each model predicts the experimental data.*

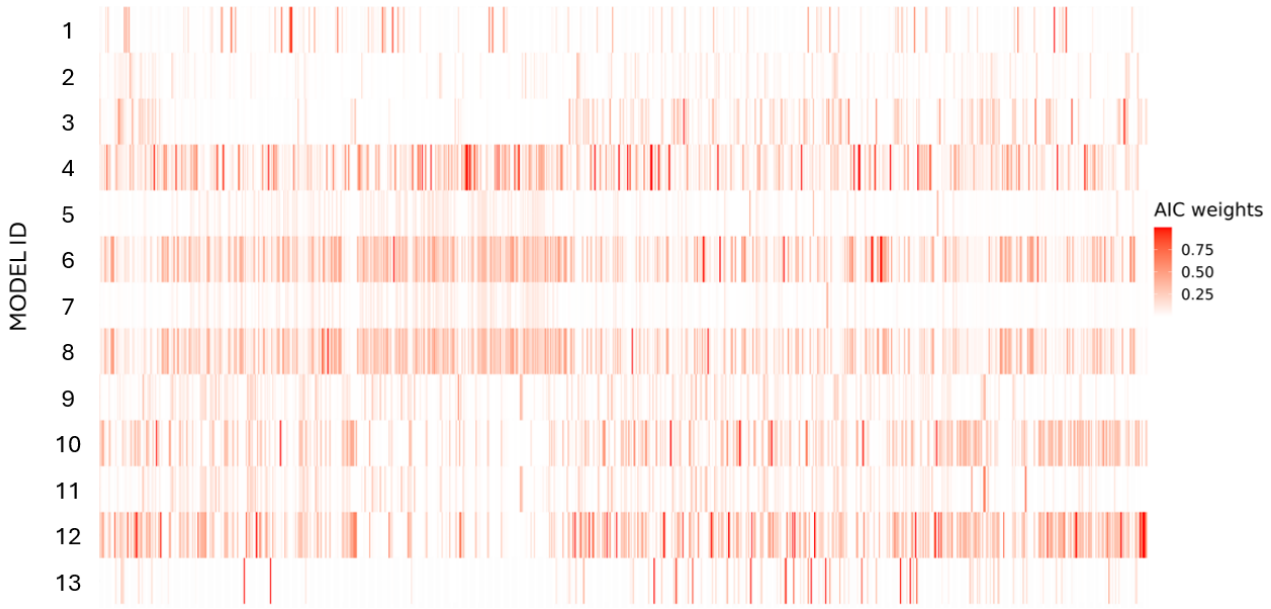

Figure 3: ***AIC Weight of Each Model in Fitting.*** *The color intensity represents the relative performance of each model, with darker shades indicating higher AIC weights and, consequently, a better fit to the data. The figure highlights models that consistently perform well across different biological data sets, such as Ratkowsky, Sharpe-Schoolfield, and Johnson-Lewin models, which demonstrate robust adaptability. In contrast, the Enzyme-Assisted Arrhenius (EAAR) model exhibits suboptimal performance.*

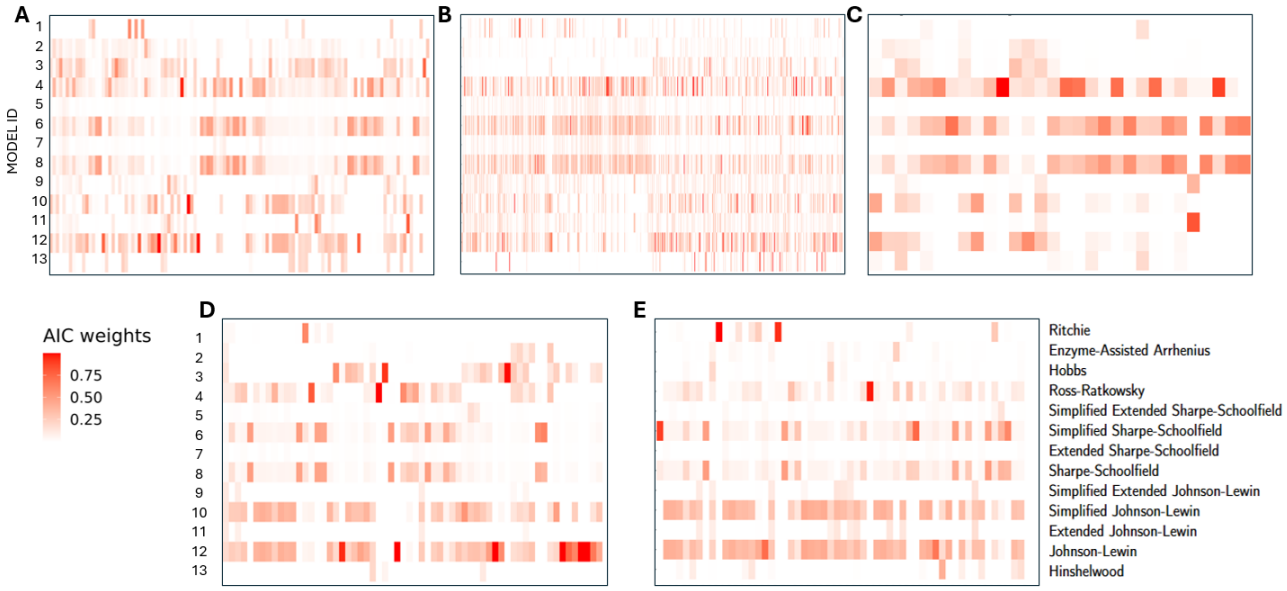

Figure 4: ***Combined AIC Weight Heatmaps.*** *This figure presents the AIC weight distributions for various thermodynamic models across different datasets, displayed as a series of heatmaps. The heatmaps are organized as follows: (a) Respiration rate data, (b) Bacterial growth rate data, (c) Gross photosynthesis rate data, (d) , Fungal growth data and (e) Enzyme activity data. The intensity of the color in each heatmap represents the relative AIC weight, reflecting the performance of each model across the respective datasets.*

Discussion

The overall fitting probability of this study is 63%, based on 987 TPC data points. Among these, the fitting performance for fungi is the lowest, with a value of only 0.50. This lower performance may be due to the use of the linear growth rate, which tracks the increase in colony canopy diameter over time. One limitation of using the diameter growth rate is that it does not capture the non-linear growth dynamics of fungi, which involve three-dimensional expansion. As the colony grows, the relationship between diameter and overall biomass or volume becomes increasingly non-linear, particularly in later stages of growth (Põlme et al., 2020). If the relative growth rate of the colony volume were used instead, the model's fitting performance might improve.

The John-Lewin model shows superior performance for several reasons. With the least number of parameters among all tested models, it bypasses the issue of overfitting during the fitting process. In contrast, models with more parameters, like the extended Sharpe-Schoolfield model with seven parameters, face a higher risk of overfitting, which can lead to the capture of noise instead of true data trends (Kontopoulos et al., 2023). For instance, although the Enzyme-Assisted Arrhenius model (5 parameters) exhibits a high overall fitting probability of 0.86, it does not perform aic weight over than 0.2 across all datasets. Notably, the Ross-Ratkowsky (5 parameters), Sharpe-Schoolfield model (6 parameters) and its simplified version (6 parameters) display performances comparable to the John-Lewin model in bacterial growth rate datasets. However, despite this similar performance, the prevalence of these models as the best fit differs substantially. These models exceed John-Lewin in the initial half of the TPCs (Figure 4B), possibly due to varied temperature distributions within large datasets. Unlike the Johnson-Lewin model, the additional parameter in the Sharpe-Schoolfield models allows them to account for low-temperature data, which helps explain the observed variance (Kontopoulos et al., 2018).

Additionally, in this study, the Sharpe-Schoolfield and Ross-Ratkowsky models were particularly notable for their precise ability to represent the thermal performance of the gross photosynthetic rate. This process involves a complex interplay between temperature and physiological mechanisms, with the Sharpe-Schoolfield model being highly regarded for its effectiveness in capturing these intricate dynamics (Kontopoulos et al., 2020). It is particularly well suited to describing the optimal temperature range and the eventual decline in activity at high temperatures. Their ability to simulate these processes across a wide temperature range makes them especially effective for analysing such datasets. The Ross-Ratkowsky model is less effective at fitting data at both low and high temperatures due to its constraints by enthalpy and entropy changes (Ratkowsky et al., 2005). However, this characteristic makes it well-suited for general predictions across a moderate temperature range in this study. Models like the Sharpe-Schoolfield and Simplified

Sharpe-Schoolfield offer a more comprehensive depiction of enzyme kinetics, particularly regarding thermal inactivation at elevated temperatures. This complementarity is further evidenced by their AIC weights in photosynthesis modelling, where they provide complementary insights (Figure 4C). In contrast, the John-Lewin model, while effective in biological processes such as enzyme activity or bacterial growth, may not be as well-suited for describing photosynthesis due to its limitations in capturing the non-linear and often asymmetric temperature responses observed in this process (Kontopoulos et al., 2018). The John-Lewin model assumes a more straightforward relationship between temperature and reaction rates, which does not adequately account for the complex thermal dynamics, including the sharp decline in efficiency at extreme temperatures. Therefore, while the John-Lewin model has its strengths in other contexts, it may fall short in accurately modeling the intricate thermal dependencies of photosynthesis, highlighting the need for more nuanced models like Sharpe-Schoolfield and Ross-Ratkowsky.

Models that incorporate the concept of temperature-induced enzyme denaturation tend to outperform those based on molecular dynamics because enzyme stability is a crucial factor determining biological activity across varying temperature conditions. This approach more accurately reflects the physiological limitations imposed by temperature on enzyme function, thereby providing a more reliable prediction of biological responses in different thermal environments. The Ritchie, Hobbs, and EAAR models attempt to incorporate molecular dynamics mechanisms as an alternative to thermal denaturation, but they were generally unsuccessful in this study, except for a few specific TPCs where they performed well. The Ritchie model, which suggests that reaction rates are primarily influenced by diffusion at lower temperatures, was found to be effective in fitting those that do not exhibit a prominent rising phase. This implies that the model is more suitable for situations where the temperature remains close to the optimum and does not significantly exceed it (Ritchie, 2018). This limitation arises because enzyme stability under thermal stress is a fundamental determinant of biological activity, and the models that fail to account for this are less effective in predicting biological responses across a broader range of temperature conditions (Hobbs et al., 2013; Kontopoulos et al., 2023). The failure of molecular dynamics models to fully describe this aspect highlights the critical importance of incorporating enzyme denaturation into predictive models to accurately reflect the physiological realities of temperature-dependent biological processes.

Limitation and further direction

One of the major limitations was the significant difference in the size of the datasets used. This discrepancy hindered the ability to make comprehensive cross-comparisons between different models. Smaller datasets (gross photosynthesis rate) may not provide sufficient data points to capture the full variability in thermal performance curves (TPCs), leading to potential biases in model selection. Future studies should aim to balance the dataset sizes across different biological groups or employ statistical methods that can account

for such discrepancies to enable more accurate and meaningful comparisons.

The study observed that multiple models were capable of fitting the same dataset well, but it was challenging to distinguish which model was truly superior. This overlap in model performance suggests that while some models may be flexible enough to explain certain data patterns, they may not necessarily reflect the underlying biological mechanisms accurately. To address this, future research could focus on developing criteria or additional parameters that better differentiate between models, potentially through cross-validation or model averaging techniques that can identify the most parsimonious model.

Another limitation was the difficulty in determining appropriate starting values for models, especially when dealing with datasets that exhibit multiple distributions. Incorrect starting values can lead to poor model convergence and biased parameter estimates, which affect the overall reliability of the model. Future research should explore advanced optimization techniques or adaptive algorithms that can more accurately estimate starting values, thereby improving the consistency and reliability of model fitting across diverse datasets.

Conclusion

In order to determine the most suitable model to elucidate the thermal dependence of the microorganism dataset, I applied 13 thermodynamic models to a dataset comprising 987 thermal performance curves (TPC). The comparative efficacy of each model in fitting the TPCs was assessed utilising Akaike Information Criterion (AIC) weights, leading to the following conclusions:

1. Johnson-Lewin model fits best across all dataset.
2. Sharpe-Schoolfield, Ross-Ratkowsky and simplified Sharpe-Schoolfield model fits best to gross photosynthesis data.
3. Sharpe-Schoolfielde, Simplified Sharpe-Schoolfield and Ross-Ratkowsky show same best performance with John-Lewin model in bacterial growth rate data.
4. Models predicated on the temperature-induced denaturation of enzymes outperformed those grounded in molecular dynamic theories.

Data and code Availability

All my work has post my own CMEE github repository: https://github.com/LeoZHOU22/Final-Project-for-CMEE/tree/main/FinalProject.

References

Blaustein, R., Pachepsky, Y., Hill, R., Shelton, D. and Whelan, G. (2013), 'Escherichia coli survival in waters: Temperature dependence', *Water Research* **47**(2), 569–578.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. and West, G. B. (2004), 'Toward a metabolic theory of ecology', *Ecology* **85**(7), 1771–1789.

Cébron, A., Zeghal, E., Usseglio-Polatera, P., Meyer, A., Bauda, P., Lemmel, F., Leyval, C. and Maunoury-Danger, F. (2021), 'Bactotraits – a functional trait database to evaluate how natural and man-induced changes influence the assembly of bacterial communities', *Ecological Indicators* **130**, 108047.

Elek, G. and Müller, M. (2024), 'Ervin bauer's concept of biological thermodynamics and its different evaluations', *Biosystems* **235**, 105090.

Falkowski, P. G., Fenchel, T. and Delong, E. F. (2008), 'The microbial engines that drive earth's biogeochemical cycles', *Science* **320**(5879), 1034–1039.

Hinshelwood, C. N. (1946), *The chemical kinetics of the bacterial cell*, Clarendon Pr, Oxford.

Hobbs, J. K., Jiao, W., Easter, A. D., Parker, E. J., Schipper, L. A. and Arcus, V. L. (2013), 'Change in heat capacity for enzyme catalysis determines temperature dependence of enzyme catalyzed rates', *ACS chemical biology* **12**(3), 868.

Johnson, F. H. and Lewin, I. (1946), 'The growth rate of e. coli in relation to temperature, quinine and coenzyme', *Journal of Cellular and Comparative Physiology* **28**(1), 47–75.

Johnson, J. B. and Omland, K. S. (2004), 'Model selection in ecology and evolution'.

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

Kontopoulos, D. G., van Sebille, E., Lange, M., Yvon-Durocher, G., Barraclough, T. G. and Pawar, S. (2020), 'Phytoplankton thermal responses adapt in the absence of hard thermodynamic constraints', *Evolution* **74**(4), 775–790.

Kontopoulos, D., Sentis, A., Daufresne, M., Dell, A. and Pawar, S. (2023), No model to rule them all: a systematic comparison of 83 thermal performance curve models across traits and taxonomic groups, Technical report.

Nock, C., Vogt, R. and Beisner, B. (2016), *Functional Traits*, In: eLS. John Wiley Sons,Ltd: Chichester.

Noll, P., Lilge, L., Hausmann, R., Henkel, M. and Henkel, M. (2020), 'Modeling and exploiting microbial temperature response'.

Padfield, D. (2023), 'nls.multstart: robust and reproducible non-linear regression in r'. R package version 1.3.0.

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. PMID: 26731029.

Popovic, M. (2019), 'Thermodynamic properties of microorganisms: determination and analysis of enthalpy, entropy, and gibbs free energy of biomass, cells and colonies of 32 microorganism species', *Heliyon* **5**(6), e01950.

Põlme, S., Abarenkov, K. and Henrik, N. (2020), 'Fungaltraits: a user-friendly traits database of fungi and fungus-like stramenopiles.', *Fungal Diversity 105, 1–16* .

Ratkowsky, D. A., Olley, J. and Ross, T. (2005), 'Unifying temperature effects on the growth rate of bacteria and the stability of globular proteins', *Journal of Theoretical Biology* **233**(3), 351–362.

Ray, P. H., White, D. C. and Brock, T. D. (1971), 'Effect of temperature on the fatty acid composition of thermus aquaticus', *Journal of Bacteriology* **106**(1), 25–30.

Ritchie, M. E. (2018), 'Reaction and diffusion thermodynamics explain optimal temperatures of biochemical reactions', *Scientific reports* **8**(1), 11105–10.

Schoolfield, R. M., Sharpe, P. J. H. and 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.

Steane, A. M. (2016), *Thermodynamics: A Complete Undergraduate Course*, online edn edn, Oxford University Press, Oxford. Accessed 25 July 2024.

Thomas, M. K., Kremer, C. T. and Litchman, E. (2016), 'Environment and evolutionary history determine the global biogeography of phytoplankton temperature traits', *Global Ecology and Biogeography* **25**(1), 75–86.

Vanchurin, V., Wolf, Y. I., Koonin, E. V. and Katsnelson, M. I. (2022), 'Thermodynamics of evolution and the origin of life', *Proceedings of the National Academy of Sciences* **119**(6), e2120042119.