

rTPC and *nls.multstart*: A new pipeline to fit thermal performance curves in R

Daniel Padfield¹  | Hannah O'Sullivan² | Samraat Pawar² ¹College of Life and Environmental Sciences,
Environment and Sustainability Institute,
University of Exeter, Cornwall, UK²Department of Life Sciences, Imperial
College London, Berkshire, UK**Correspondence**Daniel Padfield
Email: d.padfield@exeter.ac.uk**Funding information**Natural Environment Research Council,
Grant/Award Number: NE/P001130/1,
NE/S000348/1 and NE/M020843/1**Handling Editor:** Laura Graham**Abstract**

1. Quantifying thermal performance curves (TPCs) for biological rates has many applications to important problems such as predicting responses of biological systems—from individuals to communities—to directional climate change or climatic fluctuations.
2. Current software tools for fitting TPC models to data are not adequate for dealing with the immense size of new datasets that are increasingly becoming available. We require tools capable of tackling this issue in a simple, reproducible and accessible way.
3. We present a new, reproducible pipeline in R that allows for relatively simple fitting of 24 different TPC models using nonlinear least squares (NLLS) regression. The pipeline consists of two packages—*rTPC* and *nls.multstart*—that provide functions which conveniently address common problems with NLLS fitting such as the NLLS parameter starting values problem. *rTPC* also includes functions to set starting values, estimate key TPC parameters and calculate uncertainty around parameter estimates as well as the fitted model as a whole.
4. We demonstrate how this pipeline can be combined with other packages in R to robustly and reproducibly fit multiple mathematical models to multiple TPC datasets at once. In addition, we show how model selection or averaging, weighted model fitting and bootstrapping can be easily implemented within the pipeline.
5. This new pipeline provides a flexible and reproducible approach that makes the challenging task of fitting multiple TPC models to data accessible to a wide range of users across ecology and evolution.

KEYWORDS

nonlinear least squares, reaction norms, regression, thermal performance curves, thermal tolerance curves, thermodynamic models

1 | INTRODUCTION

Thermal performance curves (TPCs) describe how biological rates such as growth, photosynthesis and respiration change with temperature. TPCs and the parameters that underpin them have been used widely in biology, from studying thermal adaptation

(Kontopoulou, Van Sebille, et al., 2020; Schaum et al., 2017), to predicting ectotherm range shifts (Sinclair et al., 2016; Sunday et al., 2012) and changes in disease dynamics (Molnár et al., 2013; Mordecai et al., 2019) under expected climate change. Studies looking across large spatial scales, or using a comparative approach, need to fit mathematical models to large datasets of

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

hundreds or even thousands of TPCs across multiple taxonomic groups (Kontopoulos et al., 2020). Increasingly large datasets require reproducible and user-friendly computational pipelines for fitting multiple, competing TPC models to data. A few R packages (e.g. *temperatureresponse* (Low-Décarie et al., 2017) and *devRate* (Rebaudo et al., 2018)) provide methods for fitting TPC models to data using nonlinear least squares regression (NLLS). While these packages are a significant advance to TPC model fitting, no single pipeline addresses three key requirements: (a) implementation of a representative number of mathematical models, (b) methods to overcome the well-known sensitivity of NLLS algorithms to parameter starting values and (c) calculation of uncertainty in parameter estimates and the model fit as a whole.

The first requirement—of fitting a sufficient number of mathematical models—is important because a large number of mathematical models have been proposed to quantify TPC data (DeLong et al., 2017; Krensek et al., 2011). This makes the challenge of determining the ‘best’ model for any given dataset particularly difficult. A few papers have evaluated the performance of TPC models (Angilletta Jr., 2006; Krensek et al., 2011; Shi & Ge, 2010; Shi et al., 2016, 2017). The most comprehensive analysis to date compared 12 models, and demonstrated how model choice alters the predicted species-level response to temperature (Low-Décarie et al., 2017). However, despite the wide uptake of model selection across ecology and evolution (Johnson & Omland, 2004), fitting multiple mathematical models to TPCs remains rare practice. Instead, a single model is typically used, either because of its mechanistic underpinnings or simply because it is ‘well known’ and provides ‘adequate’ estimates of desired parameters (e.g. optimum temperature). Indeed, there is likely no ‘best’ model to use for fitting TPCs, with different models proving the most appropriate for different biological processes, taxa and levels of data quality. Allowing users with different research questions and modelling requirements to fit TPCs in a reproducible manner requires a pipeline that is flexible, modular. The second requirement—finding starting values—is a well-known challenge with NLLS model fitting (Burnham & Anderson, 2002), and generally requires the development of bespoke methods that vary with the mathematical model. This issue is particularly challenging when it comes to mathematical TPC models because of the number and diversity of models that are available. Finally, the third requirement—calculating uncertainty—is especially important for TPCs when fitting models to multiple datasets from diverse taxonomic groups or traits as the data can vary widely in sampling replication, measurement accuracy and coverage of temperature range.

Here, we present *rTPC* and *nls.multstart*, two open-source R packages that provide the basis for a pipeline to robustly and reproducibly fit TPCs by addressing these three key requirements. The pipeline allows the fitting of 24 different TPC model formulations, and we demonstrate how multiple models can be fitted to the same curve, as well as how multiple datasets can be fitted. We also describe new helper functions within *rTPC* for the estimation of start parameters, upper and lower parameter limits and the calculation of commonly used parameters (e.g. optimum temperature, activation energy or Q_{10}). Finally, we illustrate how this pipeline can be used

for model selection and model averaging, as well as how weighted model fitting and bootstrapping implemented using *rTPC* can be used to account for parameter and model uncertainty.

2 | PIPELINE OVERVIEW

The goal of *rTPC* and the associated pipeline is to make fitting TPCs easier, repeatable and transparent. Tutorials can be found at <https://padpadpadpad.github.io/rTPC> where all vignettes are available. When developing *rTPC*, we made a conscious decision not to repeat code and methods that are already optimised and available in the R ecosystem. Instead, they are utilised and incorporated into the pipeline (see Table S1 for a list of R packages used). This modularity of design improves flexibility, allowing users to incorporate *rTPC* and *nls.multstart* into their own pipelines, but still benefit from the helper functions.

2.1 | Pre-processing of data before using *rTPC*

rTPC can fit TPCs to any biological rate or fitness proxy that shows a unimodal response to temperature. Data need to be stored in long format, where each row is one rate measurement per curve. This means that each TPC will have multiple rows in the dataframe, with extra treatment columns added to distinguish between curves (Figure 1a). Pre-processing to reformat data to long format can easily be done using, for example, `tidyr::pivot_longer()`.

2.2 | Models contained in *rTPC*

rTPC contains 24 mathematical TPC models (Figure 1b; Table S2). Models were chosen through an extensive search of the primary literature and review papers of TPC model performance. Most models are named after the author who first formulated the model and the year of its first use (e.g. *thomas_2012()*). A list of all models in *rTPC* can be accessed using `get_model_names()`. Models can be characterised by whether they appropriately model negative rates before and after the optimum temperature (Table S2). As the search was not exhaustive, some models—for example, the Logan model (Logan et al., 1976)—are not currently implemented. However, requests to add new models can be made on *rTPC*'s github repository: <https://github.com/padpadpadpad/rTPC/issues>.

2.3 | Reliable NLLS fitting using *nls.multstart*

The Gauss–Newton (implemented in *nls*) and the Levenberg–Marquardt (implemented in *minpack.lm::nlsLM*) NLLS fitting algorithms are sensitive to the choice of starting values for the model parameters. This sensitivity can result in differences in parameter estimates between separate fitting attempts for the same dataset, or a complete failure to fit the model (the optimisation does not converge). To address this, the R package

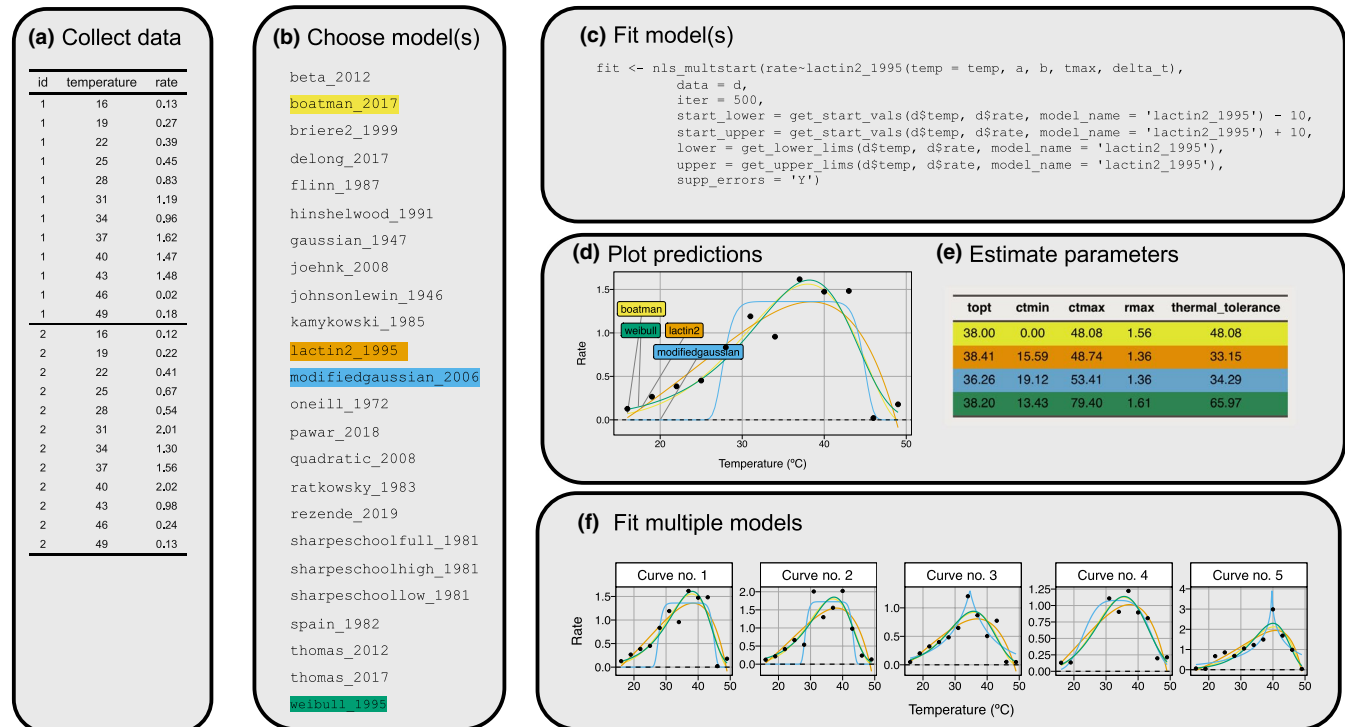


FIGURE 1 General pipeline for fitting thermal performance curves using *rTPC*. (a) Collect, check and present data in long format. (b) Choose which models from *rTPC* to be used. Here, a random assortment of four models were chosen. (c) Fit the models using *nls.multistart* and helper functions from *rTPC*. (d) Models can be visualised, and (e) common traits of TPCs can be estimated using *rTPC::calc_params()*. (f) This simple pipeline can be scaled up to be used on multiple curves

Function	Description
<code>get_model_names()</code>	Lists the models available in <i>rTPC</i>
<code>get_start_vals()</code>	Estimates start values given the temperature, rate values and model selected
<code>get_lower_lims()</code>	Sets lower limits given the data and the model selected
<code>get_upper_lims()</code>	Sets upper limits given the data and the model selected
<code>get_ctmax()</code>	Estimates the critical thermal maximum of the model fit. Where the predicted rate can never be 0, the temperature at which the rate is 5% of r_{max} is returned
<code>get_ctmin()</code>	Estimates the critical thermal minimum of the model fit. Where the predicted rate can never be 0, the temperature at which the rate is 5% of r_{max} is returned
<code>get_e()</code>	Estimates the activation energy of the model fit
<code>get_eh()</code>	Estimates the deactivation energy of a thermal performance curve
<code>get_q10()</code>	Estimates the Q_{10} value of a thermal performance curve
<code>get_topt()</code>	Estimates the optimum temperature
<code>get_rmax()</code>	Estimates the rate at optimum temperature
<code>get_skewness()</code>	Estimates skewness of a thermal performance curve
<code>get_thermalsafetymargin()</code>	Estimates the thermal safety margin of a thermal performance curve ($CT_{max} - T_{opt}$)
<code>get_thermaltolerance()</code>	Estimate the thermal tolerance of a thermal performance curve ($CT_{max} - CT_{min}$)
<code>get_breadth()</code>	Estimates thermal performance breadth of a thermal performance curve
<code>calc_params()</code>	Returns a table of all the estimated parameters

TABLE 1 Overview of helper functions included in *rTPC*

nls.multstart—and its only function *nls_multstart()*—automatically runs many iterations of the model fitting with different starting values using *minpack.lm::nlsLM* (Elzhov et al., 2016). Starting values can be generated in two ways. The first is a simple, constrained random search, where starting parameter values are sampled from a uniform distribution between pre-defined lower and upper start values for each parameter. The second is a grid search, where the start parameters are generated such that space between the given parameter bounds is evenly sampled. The best model is then picked and returned using Akaike's information criterion corrected for small sample size (AICc) (Padfield & Matheson, 2018). *nls.multstart* can also be used for fitting other nonlinear models used in biology, such as logistic growth curves (Padfield et al., 2020) and photosynthesis-irradiance curves.

2.4 | Estimating starting parameter values and limits for fitting TPCs using *rTPC*

rTPC helper functions *get_start_vals()*, *get_lower()* and *get_upper()* aid in the specification of sensible start values and limits that can be used by *nls_multstart()* (or *nls* and *nlsLM*; Figure 1c). These functions return values for the desired model, which is specified using the argument *model_name*. Where possible, the model's starting parameter values are estimated from the data. In all other instances, start values are the average fitted parameters from studies that used that equation. Upper and lower limits are set at biologically implausible

values. Users also have the option of setting values manually if these helper functions are not required.

2.5 | Calculating derived TPC parameters

One common motivation for fitting TPCs to mathematical models is to extract key TPC parameters, such as optimum temperature or Q_{10} , for further downstream analyses (e.g. to determine if optimum temperature correlates with local climate across taxa). However, the best-fitting model's parameters may not include the parameter of interest. *rTPC* includes helper functions, such as *get_topt()* and *get_rmax()*, that numerically calculate the parameters of interest from any fitted TPC (Table 1). These derived parameters are calculated from high resolution predictions (0.001°C intervals) of the fitted model. The function *calc_params()* returns values for 11 derived parameters in a dataframe (Table 1; Figure 1e). *calc_params()* does not return estimates of uncertainty in these derived parameters.

3 | USES FOR THE *rTPC* PIPELINE

Below we give examples of potential applications and extensions to the pipeline, why they are important, and guidance as to how they can be incorporated.

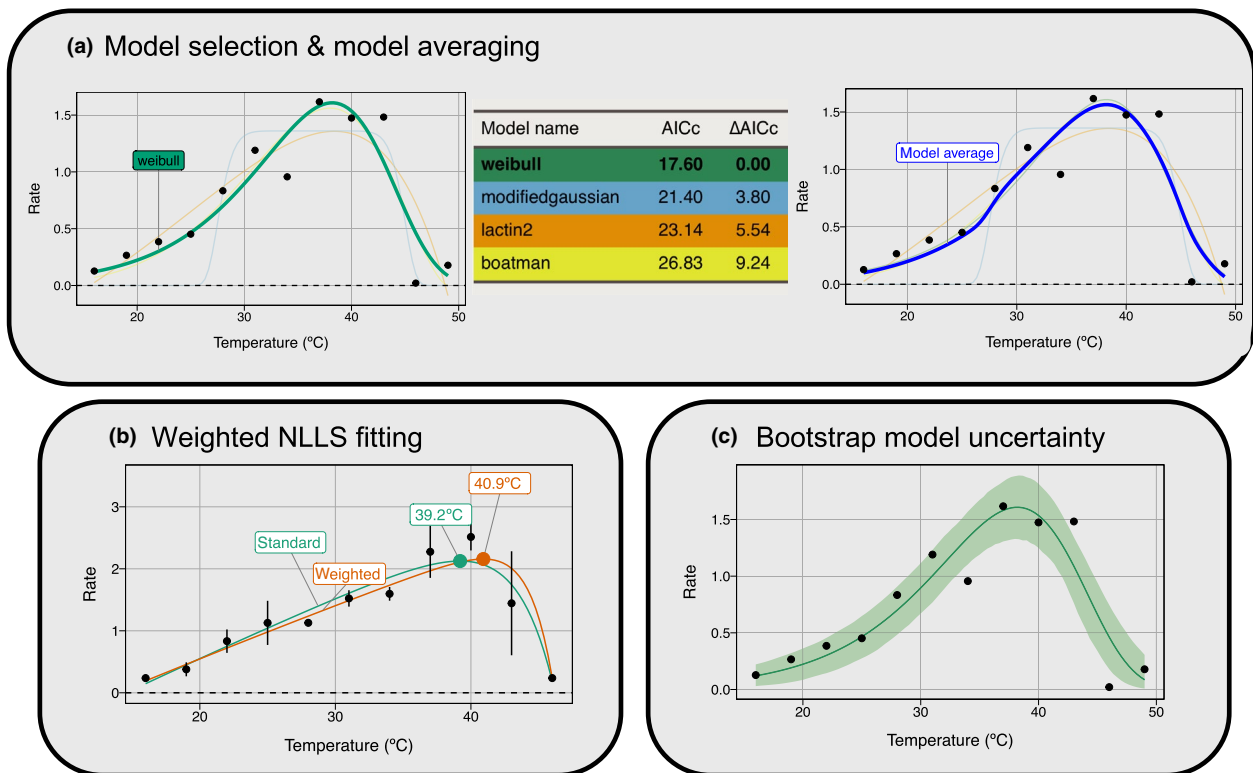


FIGURE 2 Potential extensions for fitting thermal performance curves using *rTPC*. (a) AICc scores of model fits can be calculated to help with model selection or model averaging. (b) If TPCs are fit to averages of replicates, weighted NLLS can be used to reduce parameter bias. (c) After the model has been fitted, nonparametric bootstrapping can estimate model uncertainty and confidence intervals for parameters

3.1 | Model selection and model averaging

The 'best' model for one dataset is not necessarily the best across other datasets. Fitting multiple models to TPCs allows the user to select the most suitable model for their research question. Our pipeline provides a flexible approach to help with model selection. For example, after fitting a number of potential models, AICc scores can be used to rank the models for each individual curve fit and pick the best overall model across all curves in a dataset. Alternatively, one may choose the best model specific to each TPC, or use model averaging to obtain an overall TPC curve and parameter estimates by weighting each model's fit by its AICc (Figure 2). `vignette("model_selection_averaging")` provides an example of how to implement model selection and model averaging. Some important extensions of the *rTPC* pipeline [data-weighted TPC model fitting (Figure 2), and bootstrapping to quantify uncertainty in the model fit and in parameter estimates (Figure 2)] are included in the Supplementary Information and the vignettes `vignette("model_weighting")`, `vignette("bootstrapping_models")` and `vignette("weighted_bootstrapping")`.

4 | KEY CONSIDERATIONS WHEN FITTING TPCS

Effective fitting of TPCs depends on decisions made during experimental design, data collection and model choice.

4.1 | Data considerations

For effective fitting of TPCs, the number of unique temperature values used, the level of replication at each temperature and the temperature range, all need to be considered. In the (common) scenario where all three cannot be maximised, the objective of the TPC fitting—and the parameters of particular interest—need to be considered. For example, in thermodynamic models, if the objective is to quantify the activation energy accurately, thermal range can be traded off for a finer degree of temperature resolution in the operational temperature range of the study organism (Pawar et al., 2016). It is particularly important to consider the level of replication at each temperature. Sampling multiple individuals at each temperature can give multiple individual TPCs of a population, which could be used to evaluate intraspecific variation in traits.

4.2 | Which models to fit

The decision on which TPC models to fit largely depends on the type and quality of data, and the questions being asked. In terms of data requirements, there must be at least $k + 1$ points to fit a given model, where k is the number of model parameters. However, in NLLS fitting, the minimum number of data points needed to reliably

fit a model to data can vary with the mathematical structure of the model (Burnham & Anderson, 2002), so in general, 'the more the merrier'. Carefully consider what model(s) you want to use before starting the analysis. If there are negative rate values, it is wise to fit models that can cross the x-axis both below and above the optimum temperature, such as `thomas_2012()`, `thomas_2017()` or `joehnk_2008()` (Table S2). In terms of the questions being asked, if the model output is being used for forecasting, a more mechanistic model (e.g. Sharpe-Schoolfield or Johnson-Lewin) might be preferred over a descriptive model like the Gaussian. Moreover, if there are specific traits of interest (e.g. optimum temperature), it may be beneficial to only consider models that explicitly include that parameter in their formulation. This may be especially pertinent for the activation energy, deactivation energy and Q_{10} , as they are sensitive to the calculation of the optimum temperature when calculated from model predictions. Finally, because NLLS is a numerical (inexact) model fitting method, the correlations between parameters should be given due consideration, as they that may result in spurious parameter estimates (see Kontopoulou et al., 2018 in the case of the Sharpe-Schoolfield model).

5 | CONCLUDING REMARKS

The pipeline presented here allows TPC data to be fitted to mathematical models in a simple, reproducible and flexible way. *rTPC* includes 24 mathematical models that represent the wide diversity of nonlinear TPC models available in the literature, and *nls.multstart* allows this set of models to be reliably fitted to data using NLLS by addressing the starting values problem. However, this pipeline does not accommodate non-independent (related) replicates, and clustered or stratified sampling (possibly with missing values). In such situations, nonlinear mixed effects model fitting (e.g. using the *nlme* R package; Oddi et al., 2019) or Bayesian approaches (e.g. using the *brms* R package; Bürkner, 2017) would be more appropriate. Nevertheless, for fitting massive TPC datasets to multiple mathematical models, *rTPC* offers a simple, reliable and reproducible computational pipeline with robust methods for calculation of model uncertainty, requiring minimal statistical and computational expertise, and suitable for a wide range of applications.

ACKNOWLEDGEMENTS

D.P. was funded by a NERC grant (NE/P001130/1) and S.P. was funded by NERC grants (NE/S000348/1 and NE/M020843/1).

AUTHORS' CONTRIBUTIONS

D.P. conceived the ideas and designed the pipeline; D.P. authored the R package and wrote the initial draft. All the authors contributed to developing the manuscript and gave final approval for publication.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13585>.

DATA AVAILABILITY STATEMENT

All data and code used in *rTPC* are archived at <https://doi.org/10.5281/zenodo.4561704>, and *nls.multstart* is available on CRAN.

ORCID

Daniel Padfield  <https://orcid.org/0000-0001-6799-9670>

Samraat Pawar  <https://orcid.org/0000-0001-8375-5684>

REFERENCES

- Angilletta Jr., M. J. (2006). Estimating and comparing thermal performance curves. *Journal of Thermal Biology*, 31(7), 541–545. <https://doi.org/10.1016/j.jtherbio.2006.06.002>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Burnham, K. P., & Anderson, D. R. (2002). A practical information-theoretic approach. In *Model selection and multimodel inference* (2nd ed.). Springer.
- 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. <https://doi.org/10.1002/ece3.2955>
- 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*. R package version 1.2-1.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19(2), 101–108. <https://doi.org/10.1016/j.tree.2003.10.013>
- Kontopoulos, D.-G., García-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. <https://doi.org/10.7717/peerj.4363>
- Kontopoulos, D.-G., Smith, T. P., Barraclough, T. G., & Pawar, S. (2020). Adaptive evolution shapes the present-day distribution of the thermal sensitivity of population growth rate. *PLoS Biology*, 18(10), e3000894. <https://doi.org/10.1371/journal.pbio.3000894>
- Kontopoulos, D.-G., Van Sebille, E., Lange, M., Yvon-Durocher, G., Barraclough, T. G., & Pawar, S. (2020). Phytoplankton thermal responses adapt in the absence of hard thermodynamic constraints. *Evolution*, 74(4), 775–790. <https://doi.org/10.1111/evo.13946>
- Krenek, S., Berendonk, T. U., & Petzoldt, T. (2011). Thermal performance curves of *Paramecium caudatum*: A model selection approach. *European Journal of Protistology*, 47(2), 124–137. <https://doi.org/10.1016/j.ejop.2010.12.001>
- Logan, J. A., Wollkind, D. J., Hoyt, S. C., & Tanigoshi, L. K. (1976). An analytic model for description of temperature dependent rate phenomena in arthropods. *Environmental Entomology*, 5(6), 1133–1140.
- 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. <https://doi.org/10.1002/ece3.3576>
- Molnár, P. K., Kutz, S. J., Hoar, B. M., & Dobson, A. P. (2013). Metabolic approaches to understanding climate change impacts on seasonal host-macroparasite dynamics. *Ecology Letters*, 16(1), 9–21. <https://doi.org/10.1111/ele.12022>
- Mordecia, E. A., Caldwell, J. M., Grossman, M. K., Lippi, C. A., Johnson, L. R., Neira, M., Rohr, J. R., Ryan, S. J., Savage, V., Shocket, M. S., Sippy, R., Stewart Ibarra, A. M., Thomas, M. B., & Villena, O. (2019). Thermal biology of mosquito-borne disease. *Ecology Letters*, 22(10), 1690–1708. <https://doi.org/10.1111/ele.13335>
- Oddi, F. J., Miguez, F. E., Ghermandi, L., Bianchi, L. O., & Garibaldi, L. A. (2019). A nonlinear mixed-effects modeling approach for ecological data: Using temporal dynamics of vegetation moisture as an example. *Ecology and Evolution*, 9(18), 10225–10240. <https://doi.org/10.1002/ece3.5543>
- Padfield, D., Castledine, M., & Buckling, A. (2020). Temperature-dependent changes to host-parasite interactions alter the thermal performance of a bacterial host. *The ISME Journal*, 14(2), 389–398. <https://doi.org/10.1038/s41396-019-0526-5>
- Padfield, D., & Matheson, G. (2018). *nls.multstart: Robust non-linear regression using AIC scores*. R package version 1.0.1.
- 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. <https://doi.org/10.1086/684590>
- Rebaudo, F., Struelens, Q., & Dangles, O. (2018). Modelling temperature-dependent development rate and phenology in arthropods: The devRate package for R. *Methods in Ecology and Evolution*, 9(4), 1144–1150. <https://doi.org/10.1111/2041-210X.12935>
- Schaum, C.-E., Barton, S., Bestion, E., Buckling, A., Garcia-Carreras, B., Lopez, P., Lowe, C., Pawar, S., Smirnov, N., Trimmer, M., & Yvon-Durocher, G. (2017). Adaptation of phytoplankton to a decade of experimental warming linked to increased photosynthesis. *Nature Ecology & Evolution*, 1(4). <https://doi.org/10.1038/s41559-017-0094>
- Shi, P., & Ge, F. (2010). A comparison of different thermal performance functions describing temperature-dependent development rates. *Journal of Thermal Biology*, 35(5), 225–231. <https://doi.org/10.1016/j.jtherbio.2010.05.005>
- Shi, P.-J., Reddy, G. V. P., Chen, L., & Ge, F. (2016). Comparison of thermal performance equations in describing temperature-dependent developmental rates of insects: (I) Empirical models. *Annals of the Entomological Society of America*, 109(2), 211–215. <https://doi.org/10.1093/aesa/sav121>
- Shi, P.-J., Reddy, G. V. P., Chen, L., & Ge, F. (2017). Comparison of thermal performance equations in describing temperature-dependent developmental rates of insects: (II) Two thermodynamic models. *Annals of the Entomological Society of America*, 110(1), 113–120. <https://doi.org/10.1093/aesa/saw067>
- Sinclair, B. J., Marshall, K. E., Sewell, M. A., Levesque, D. L., Willett, C. S., Slotsbo, S., Dong, Y., Harley, C. D. G., Marshall, D. J., Helmuth, B. S., & Huey, R. B. (2016). Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures? *Ecology Letters*, 19(11), 1372–1385. <https://doi.org/10.1111/ele.12686>
- Sunday, J. M., Bates, A. E., & Dulvy, N. K. (2012). Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, 2(9), 686–690. <https://doi.org/10.1038/nclimate1539>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Padfield D, O'Sullivan H, Pawar S. *rTPC* and *nls.multstart*: A new pipeline to fit thermal performance curves in R. *Methods Ecol Evol*. 2021;12:1138–1143. <https://doi.org/10.1111/2041-210X.13585>