**Title:** Standardised measurement of seed functional traits: automated calculation of germination cardinal temperatures and thermal time using R

**Running Head:** Thermal Time & Cardinal Temperatures in R

Emma Ladouceur1,2, Hugh W. Pritchard3, Eduardo Fernández-Pascual3\*

1 Museo Delle Scienze (Muse), Corso del Lavoro e Scienze, 3, Trento, Italy

2 University of Pavia, Department of Earth and Environmental Science, Via S. Epifanio 14, 27100 Pavia (Italy)

3 Royal Botanic Gardens, Kew, Wellcome Trust Millennium Building, Wakehurst Place, West Sussex RH17 6TN, England

**\*Corresponding author:** [eduardofp.indurot@uniovi.es](mailto:eduardofp.indurot@uniovi.es); Tel.: +44(0)1444894184

**Keywords (7 max):** base water potential, physiological thermal thresholds, piecewise regression, regeneration traits, seed germination traits, segmented model, thermal time models

**Abstract (Currently 241--250 max)**

Seed germination traits depend on both environmental and biological factors, and are emerging as an underappreciated and powerful mechanism to understand plant regeneration in a changing world. Identifying standardized methodologies to measure germination traits is of prime interest to many fields of the plant sciences, from population ecology to vegetation science and biogeography. A good comparative framework is provided by physiological thermal time models, in which germination is quantified by the three cardinal temperatures (base,Tb; optimum, To; and ceiling, Tc), and thermal time. Traditionally, calculating these traits has required user judgement, which can introduce bias into the analysis, and is a lengthy process. Here, we present an easy and automated method to calculate these traits with R which has three main benefits; 1) it identifies the breaking point in the data systematically (i.e. the separation between the sub-optimal and supra-optimal germination temperature ranges), avoiding personal bias; 2) it is computed in seconds, in contrast to the hours generally taken using traditional methods; and 3) it does not require the purchase of commercial statistical software. We provide an example of the application of this method and a help annex to guide users who are unfamiliar with R. Finally, we briefly discuss the best policy to store and share the results of germination studies, in order to encourage co-operative meta-analyses of seed germination, and link seed biology with broader fields of plant science.

**Introduction**

There is a high demand for standardized seed germination traits in the plant sciences: from population ecology (Huang et al 2016), to vegetation science (Larson and Funk 2016), and biogeography (Bykova et al 2012). Of prime interest is the contribution that large germination datasets can make to the development of mechanistic trait-based models that predict how plant communities respond to global environmental changes (Funk et al 2016). Germination traits are however underrepresented in public trait databases (Jiménez-Alfaro et al 2016), and researchers in plant and seed biology urgently need to fill this gap. This task is best advanced through comparative studies, but the value of these will only be maximized if there is a consistent approach to the generation of data on germination phenotypes and robust means of analyzing the outputs. A solid comparative framework for germination studies is offered by the application of physiological thermal time models (Donohue et al 2015, Dürr et al 2015).

Germination occurs in response to a series of environmental cues including temperature, moisture, light and chemical signals (Bewley et al 2013). For each of these cues there is an optimal value (or sometimes a narrow range) at which the rate of germination is maximal. Below and above this optimum (To), the rate progressively decreases until the cue reaches estimated base (Tb) and ceiling (Tc) thresholds beyond which the progression of germination is predicted to stop. Thus, the rate of germination as a function of temperature can be described in non-dormant seeds as an accumulation of degrees-day above or below these thresholds (Garcia-Huidobro et al 1982). When a certain amount of degrees-day has been accumulated, the seed germinates. Every seed will require its own amount, but since experiments necessarily involve seed populations, thermal time ( θ ) is expressed as the degrees-day needed to produce germination in a percentage of this population, usually the 50 %, i.e., °C d G50. Therefore, the thermal control of germination can be described in a mechanistic way using a handful of seed traits: firstly, relating to the three cardinal temperatures of Tb, To and Tc; and secondly, the sub-optimal and supra-optimal temperature range thermal times. Thermal time traits have two powerful applications. Firstly, parameters calculated in a restricted set of experimental treatments can give good estimations of germination in more complex thermal environments, such as climate change scenarios (Orru et al 2012). Secondly, thermal time parameters can be compared in a standard way across seed individuals, populations and species (Trudgill et al 2000)(Dürr et al 2015)(Arène F, Affre L, Doxa A, Saatkamp A (2017) Temperature but not moisture response of germination shows phylogenetic constraints while both interact with seed mass and life span. Seed Science Research, in press).

Here, we present a standardized method to calculate germination cardinal temperatures and thermal time using segmented regression in the open-source software *‘R statistical computing language and platform’* (R Core Development Team 2016). This method has three advantages over methods currently in use: 1). It identifies the breaking point in the data systematically (i.e., the separation between the sub-optimal and supra-optimal germination temperature ranges), rather than the user estimating breaking points visually and on a case-by-case basis, which can lead to personal bias; 2). It is computed in seconds, in contrast to the hours it may take using alternative methods; and 3). It does not require the purchase of commercial statistical software. We test this method calculating cardinal temperatures in two examples: one in which the three cardinal temperatures can be identified relatively easily, and one in which experimental results did not lead to the successful identification of all the cardinal temperatures. We present here the format in which the data should be arranged (**Supplementary Information 1**), the script to analyse the data (**Supplementary Information 2**) and basic instructions for users new to *R* (**Supplementary Information 3**). Our aim is to make this method accessible, completely transparent and rigorously reproducible by all researchers (Rocchini and Neteler 2012).

**Description of the method**

*Data preparation*

We provide an example dataset (**Supplementary Information 1**) with germination results from a previous article (Fernandez-Pascual et al 2015). Users should format their data in the same way, keeping the columns, their order, and their column names. The first column, *Grouping*, represents an experimental factor other than temperature. In the example this factor is species, with two levels *species A* and *species B*. Groupingcan also represent different individuals, populations or experimental treatments, depending on the data being analysed. When users have more than one grouping factor in their data (e.g., ‘population’ and ‘stratification’), the two factors should be combined in one Grouping column (e.g., ‘population x stratification’). Although the example is presented with two species, the script is prepared to analyse larger numbers of factor levels. The second column, *Treatment*, records the range of temperatures each species was tested by. The third column, *Dish*, indicates the Petri dish (or other container) number. In this example, there is only one Dish for each Treatment. In other cases, where an experiment may have several Dishes per treatment, the script should work in the same way. The fourth column, *Time*, is the time in which each data point was recorded (in days in this example, but it could be hours or any other unit of time). The fifth column, *G*, indicates the cumulative germination count at that scoring date. The sixth column, *PG*, indicates the total sample size of each Petri dish, which in this case is 25 seeds. If you have a proportion of seed that do not germinate by the end of the test, they may be empty, dead, or simply dormant. These categories should be distinguished through cutting or other test. The number in PG should always be the number of germinable seeds in the dish including dormant but not empty or dead seeds.

*R and R packages needed*

All analyses are performed in R(R Core Development Team 2016). For users new to *R*, we have prepared the essential information to get started in **Supplementary Information 3***.* The script (**Supplementary Information 2**) presents the method and, when used with the example dataset (**Supplementary Information 1**), produces four tables and four figures that we will describe below. Before starting the analyses, six specialist R packages must be installed: ‘plyr’and *‘*dplyr’ (Wickham et al 2016) are used for data manipulation;‘binom’ (Dorai-Raj (n.d.)) is used to calculate binomial confidence intervals on the germination proportions; ‘drc’ (Ritz et al 2015) is used to fit dose-response models to cumulative germination data, in order to calculate the germination times and rates; ‘segmented’ (Vito 2008) is used to fit a segmented regression and compute the cardinal temperatures and thermal times; and ‘ggplot2’ is used to create figures (Wickham 2009).

*Step 1: Checking whether the data represents the full germination temperature range*

A preliminary analysis is done with the results of the final scoring date (i.e., the final germination proportions). The mean final germination proportions and their 95 % binomial confidence intervals are calculated for each combination of Grouping and Temperature. This information is exported as **Table S1**. The final germination proportions are also plotted as **Figure 1**, which needs to be inspected visually for evidence that the temperature treatments used in the experiment do represent the full germination temperature range of the study species. If this is the case, the segmented model can be fitted and used to calculate the three cardinal temperatures. This would be the case in species A, for which we generated data representing the full range of germination temperatures (**Figure 1A**). However, when the experimental treatments only give results in either the sub- or the supra-optimal germination temperature range, the segmented model cannot be fitted and only the base or ceiling temperatures can be calculated. This is the case in species B*,* where there was no germination above the treatment with the highest germination, ‘23.75 ºC’ (**Figure 1B**). Thus, for species B we only have data in the sub-optimal germination temperature range.

*Step 2: Estimating germination rates from the cumulative germination curves*

A dose-response model is fitted to the cumulative germination data, in which time is the dose and germination is the response. This model is used to estimate the time required to reach successive germination percentages. This is the most sensitive part of the analysis, as poor estimations of the germination times will lead to less accurate cardinal temperatures and thermal times. There are several dose-response functions that can be fitted to this type of data: Weibull, logistic or Boltzmann, and log-logistic or Hill (Ritz et al 2015). For each combination of grouping and treatment, the script selects the function with the best fit to the data per Akaike’s Information Criterion. This is expected to give more accurate estimations of the germination times than applying the same function across treatments. Nonetheless, the function fit needs to be checked visually in **Figure 2**. It is possible that no good fit can be found; for example, if final germination is very low or if the scoring times were inadequate (e.g., if germination is scored every 24 h but all seeds germinate within the first 24 h). If this is the case, cardinal temperatures and thermal time cannot be calculated, and we suggest to repeat the experiment with new treatments or a move-along design (C C Baskin and Baskin 2003) to learn more about the species before attempting another trial aimed at the specifics of cardinal temperatures.

If the models can be fitted, they are used to calculate the time to reach successive deciles of germination (from 10 to 90 %) in each treatment. The inverse of the time is then calculated to obtain the germination rates of each temperature treatment, and this information is exported as **Table S2**.

*Step 3a: Fitting a segmented model to the full germination temperature range*

Where sub-optimal and supra-optimal temperatures are available in the data set (**Figure 1A**), the script can calculate the three cardinal temperatures (Tb, To and Tc) and the thermal time (θ-sub, θ-supra). In this scenario, a segmented regression model is fitted to the data (**Figure 3A**). A segmented model identifies breaking points in the data (i.e., the optimal germination temperature at which the germination rate is maximal). Then the model fits linear regressions separately to the two segments defined by this breaking point (*i.e*., the sub-optimal and supra-optimal germination temperature ranges) (**Figure 3A**). Once the models are fitted, the base temperature (Tb) is calculated by solving the sub-optimal linear regression for the x-intercept (i.e., the temperature in which the value of the germination rate equals zero). An analogous procedure calculates the Tc using the supra-optimal linear regression. The intercept of the sub-optimal and supra-optimal lines gives the To. The inverse of the slopes of each regression gives the estimated thermal time for germination in degrees-day (or the time unit used in the data). The script calculates the cardinal temperatures and thermal times for each of the ten decile germination rates (10 – 90 %) and exports this information as **Table S3**. The plots with the germination rate versus temperature and the fitted models for every decile are exported as **Figure 3**.This figure needs to be carefully inspected to detect weird fits and outlier temperatures, which should be removed from the dataset in order to calculate accurate cardinal temperatures and thermal time. It must be considered that, when dealing with wild species with a certain degree of dormancy, the cardinal temperatures calculated for the lower and higher germination deciles can give extreme results. Because of this, the parameters estimated for the t50 are probably the most robust trait to use in comparative analyses.

It may be the case that the temperatures used in a given experiment fall only on the sub-optimal or supra-optimal germination temperature range, and do not succeed to identify the full range of temperatures (**Figure 1B**). In this case, the dataset **Table S3** cannot be produced. In the example of species B, it is impossible to determine neither the Tc nor the To, and the segmented model cannot be fitted to the data (**Figure 3B**). In this scenario, it would be necessary to skip forward to the next step in the script.

*Step 3b: Fitting a linear model to the sub- or supra-optimal germination temperature range*

In the case of species B, or any experiment which fails to identify the full range of temperatures (**Figure 3B**), a standard linear regression can be used to identify either the Tb or the Tc. The inverse of the slope of this regression gives the thermal time for germination in degrees-day (or the time unit used in the data). This is exported as **Table S4** which will automatically include information for whichever single segment was identified in each unique dataset, either the sub-optimal or the supra-optimal. The plot of these linear fits (**Figure 4**) must be inspected to detect outliers. Particularly, it can happen that a single data point above or below the optimal temperature exists. This data point is enough to prevent fitting the segmented model, but can affect the slope of the linear model and should be removed.

It must be noted that this part of the script can be used to calculate the germination base water potential. In this case, the dataset that is being analysed would represent an experiment of germination in different water potential solutions. The values of the column Treatment would be the water potential instead of the temperatures.

Lastly, we attach a short script to produce a smooth curve in *SMTHfig*. This can be useful to observe irregular data, which cannot be fitted to a segmented model or a linear model, simply as an exploratory exercise to best inform new approaches to the next experiment. But this should not be used as a rigorous test for publication.

**Discussion**

*How to use the script*

The method we have presented above is appropriate for the analysis of germination data to identify cardinal temperatures and thermal time, and can also be used to calculate base water potential. Once the user has become familiar with the script, running the method is fast and straightforward. The user just needs to place the script and his formatted dataset in his R working directory, and run the script within R. The script will perform all calculations and export the results as four tables and four figures ready for publication.

*Experimental designs for standardized thermal time*

An experimental design to identify the cardinal temperatures and thermal time should be tailored to include the breadth of the germination temperature range. The aim should be to generate data points purposefully below the optimal (sub-optimal) and above the optimal germination temperature (supra-optimal), pushing each species to its thermal thresholds. Fitting a segmented models requires a minimum of three sub-optimal and three supra-optimal temperatures. It is thus recommended that experimental designs include a minimum of 5 temperatures for each temperature range to effectively explore the limits, in which case some temperatures may not yield responses. The less that is known about a given species from the outset, the more temperatures that should be tested.

Constant temperatures are known to produce promising results for testing cardinal temperatures, particularly for agricultural varieties which have been bred for uniformity and stable germination temperatures (Covell et al 1986). However many wild species may require alternating temperatures, or alternating temperatures may better represent real ambient conditions found in nature (Fernandez-Pascual et al 2015, Galíndez et al 2017). When testing with alternating temperatures, it is recommended that the average temperature is plotted when the alternating temperatures remain within either the sub- or supra-optimal range (Ellis and Barret 1994). Thus, alternating temperature regimes should be carefully planned to avoid exceeding the cardinal temperatures. Constant temperatures should never be compared to alternating temperatures within the same analysis, but if testing both, the two temperature regimes should be set up as two experiments or *‘*Groupings’, to compare against each other. Ideally, in setting up an experimental design for wild species, a regime of both constant and alternating temperatures could be tested separately, and the results compared.

Experiments are recommended to be run until cumulative germination reaches a plateau. Scoring dates should be adjusted to the speed of germination of the study species, so cumulative germination curves with good resolution are produced. Finally, it is important to consider that the cardinal temperatures are not fixed values but depend on the dormancy state of the seed lot, as affected by any pre-treatment or variation if developmental time (Daws et al 2004, Pritchard et al 1999). Generally, non-dormant seeds with high germination percentages are needed to produce robust estimations of the cardinal temperatures and the thermal time.

*Conclusions*

We have found and confirmed that this method is useful to calculate thermal time and cardinal temperatures in germination experiments which have tested a full range of temperatures across a given species germination gradient. We believe that this method will be of great value to generate cardinal temperature data for an ever-increasing number of species, setting a standardized methodology for the measurement of species germination traits. Physiological thresholds such as the cardinal temperatures are highly informative plant traits, and a key tool to integrate plant regeneration into multi-species community studies (Jiménez-Alfaro et al 2016, Larson and Funk 2016). But regardless of the data analyses that they are pursuing, we strongly encourage seed researchers to adopt the format proposed here (**Supplementary Information 1**) when storing and sharing their original germination data. This format - in which each data record is a row, and each variable is a column – is the standard used by most statistical languages, including R. Using this standard for record keeping would improve the communication within seed research and with other scientific fields. Moreover, we propose the policy of making these standardized records available to the public as supplementary material of the publications in which they are first reported. This would ensure that data from single species studies can be incorporated into meta-analyses of seed germination, to the common benefit of seed science.

**Acknowledgements**

Special thanks to Maria Tudela Isanta of the NAtive Seed Science TEchnology and Conservation (NASSTEC) Initial Training Network (ITN). The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n°607785. E.F.P. had the financial support of the Government of Asturias and the FP7 – Marie Curie - COFUND programme of the European Commission (Grant ‘Clarín’ ACA14-19) to work at the Royal Botanic Gardens, Kew, which also receives grant-in-aid from Defra.

**References**

Baskin CC and Baskin JM (2003) When breaking seed dormancy is a problem try a move-along experiment. *Native Plants Journal* 4(1): 17–21.

Bewley DJ, Bradford KJ, Hilhorst HWM and Nonogaki H (2013) *Seeds Physiology of Development Germination and Dormancy* (Third.). New York: Springer, 1–407.

Bykova O, Chuine I, Morin X and Higgins SI (2012) Temperature dependence of the reproduction niche and its relevance for plant species distributions. *Journal of Biogeography* 39(12): 2191–2200.

Covell S, Ellis RH, Roberts EH and Summerfield RJ (1986) The influence of temperature of seed germination rate in grain legumes. *Journal of Experimental Botany* 37(178): 705–715.

Daws MI, Lydall E, Chmielarz P, Leprince O, Matthews S, Thanos CA and Pritchard HW (2004) Developmental heat sum influences recalcitrant seed traits in Aesculus hippocastanum across Europe. *New Phytologist* 162(1): 157–166.

Donohue K, Burghardt LT, Runcie D, Bradford KJ and Schmitt J (2015) Applying developmental threshold models to evolutionary ecology. *Trends in Ecology & Evolution* 30(2): 66–77.

Dorai-Raj S ((n.d.)) binom: Binomial Confidence Intervals For Several Parameterizations (1st edition). Comprehensive R Archive Network (CRAN). Available at: http://CRAN.R-project.org/package=binom.

Dürr C, Dickie JB, Yang XY and Pritchard HW (2015) Ranges of critical temperature and water potential values for the germination of species worldwide: Contribution to a seed trait database. *Agricultural and Forest Meteorology*. Elsevier B.V. 200: 222–232.

Ellis RH and Barret S (1994) Alternating temperatures and rate of seed germination in lentil. *Annals of Botany* 74: 519–524.

Fernandez-Pascual E, Seal CE and Pritchard HW (2015) Simulating the germination response to diurnally alternating temperatures under climate change scenarios: comparative studies on Carex diandra seeds. *Annals of Botany*. Oxford University Press 115(2): 201–209.

Funk JL, Larson JE, Ames GM, Butterfield BJ, Cavender-Bares J, Firn J, Laughlin DC, Sutton-Grier AE, Williams L and Wright J (2016) Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biological Reviews* (Early View).

Galíndez G, Seal CE, Daws MI and Lindow L (2017) Alternating temperature combined with darkness resets base temperature for germination (Tb) in photoblastic seeds of Lippia and Aloysia (Verbenaceae). *Plant Biology* (19): 1–5.

Garcia-Huidobro J, L MJ and Squires GR (1982) Time, Temperature and Germination of Pearl Millet (Pennisetum typhoides S. & H.). *Journal of Experimental Botany* 33(133): 288–296.

Huang Z, Liu S, Bradford KJ, Huxman TE and Venable LD (2016) The contribution of germination functional traits to population dynamics of a desert plant community. *Ecology* 97(1): 250–261.

Jiménez-Alfaro B, Silveira FAO, Fidelis A, Poschlod P and Commander LE (2016) Seed germination traits can contribute better to plant community ecology. *Journal of Vegetation Science* 27(3): 637–645.

Larson JE and Funk JL (2016) Regeneration: an overlooked aspect of trait-based plant community assembly models. *Journal of Ecology* 104(5): 1284–1298.

Orru M, Mattana E, Pritchard HW and Bacchetta G (2012) Thermal thresholds as predictors of seed dormancy release and germination timing: altitude-related risks from climate warming for the wild grapevine Vitis vinifera subsp. sylvestris. *Annals of Botany* 110(8): 1651–1660.

Pritchard HW, Steadman KJ, Nash JV and Jones C (1999) Kinetics of dormancy release and the high temperature germination response in Aesculus hippocastanum seeds. *Journal of Experimental Botany* 50(338): 1507–1514.

R Core Development Team (2016) R: language and environment for statistical computing (3rd edition). Vienna, Austria: Comprehensive R Archive Network (CRAN). Available at: http://www.R-project.org/.

Ritz C, Baty F, Streibig JC and Gerhard D (2015) Dose-Response Analysis Using R. *PLoS ONE*. Public Library of Science 10(12): 1–13.

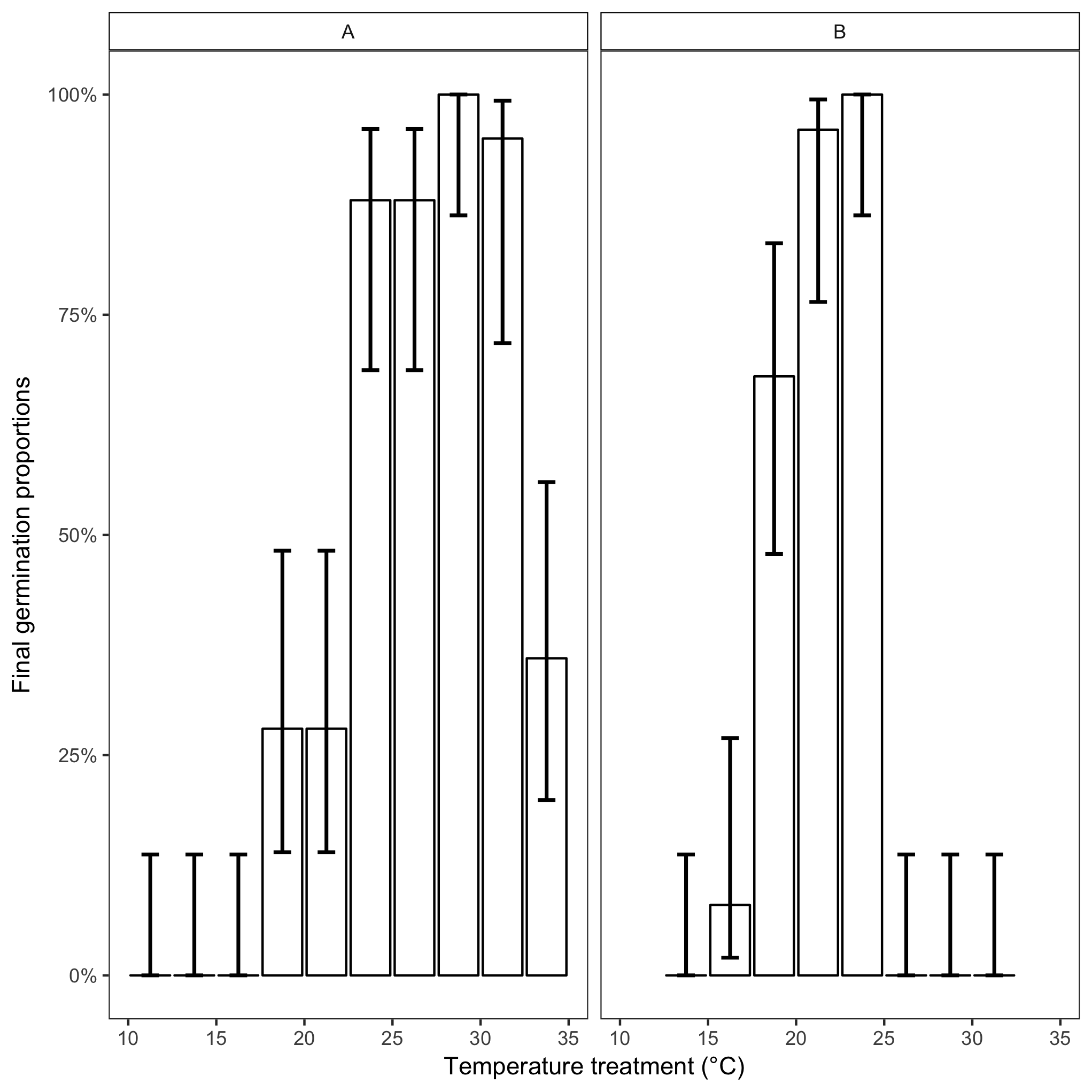
Rocchini D and Neteler M (2012) Let the four freedoms paradigm apply to ecology. *Trends in Ecology & Evolution*. Elsevier Ltd 27(6): 310–311.

Trudgill DL, Squire GR and Thompson K (2000) A thermal time basis for comparing the germination requirements of some British herbaceous plants. *New Phytologist*. Cambridge University Press 145(1): 107–114.

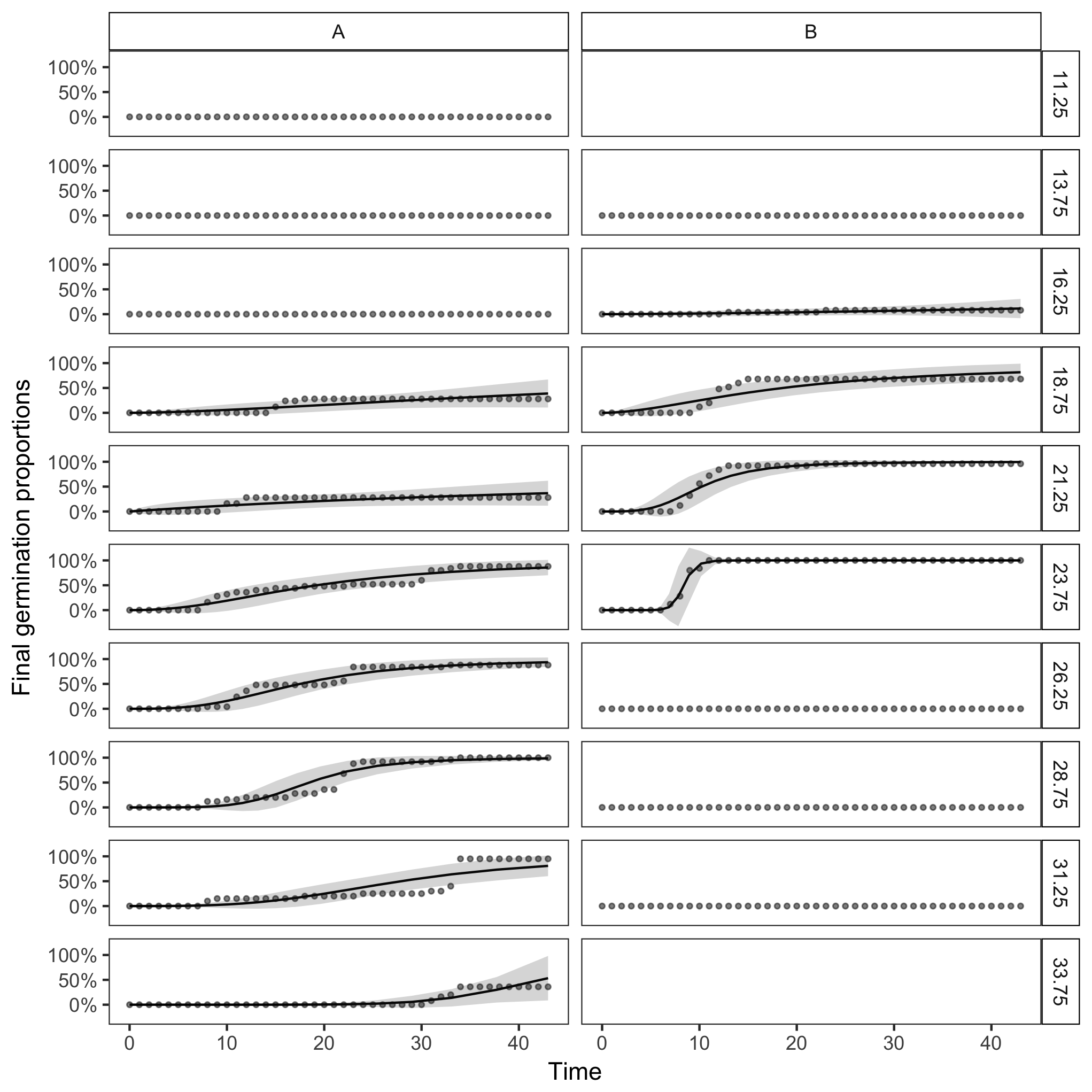
Vito MR (2008) Segmented: An R Package to fit regression models with broken-line relationships. *R News* 8(1): 20–25. Available at: http://cran.r-project.org/doc/Rnews/.

Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham H, Francois RFrancois (2016) dplyr: A grammar of data manipulation (0 edition). Comprehensive R Archive Network (CRAN). Available at: http://CRAN.R-project.org/package=dplyr.



**Figure 1:** Final germination percentages across all temperature treatments for species A and species B. Bars represent the final germination percentage of each temperature treatment, brackets represent the confidence interval of every treatment.



**Figure 2:** Cumulative germination curves across all temperature treatments for species A and species B. Points represent germination percentage for each scoring time, and lines represent the best fit dose-response model to the scoring days.



**Figure 3:** Time to germination across each decile (%) of total germination, across each treatment using a segmented model for species A and species B. Points represent the germination rate for every temperature treatment, and lines represent the fit of the segmented model to the points for every decile.



**Figure 4:** Time to germination across each decile (%) of total germination across each temperature treatment, using a smooth linear model for species A and species B. Points represent the germination rate for every temperature treatment, and lines represent the best fit of a linear smooth line to the data points.