**Title:** An easy and automated calculation of the 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\*

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

2University of Pavia, Pavia, Italy

3Royal 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):** broken-stick regression, piecewise regression, seed traits, segmented model, thermal thresholds, thermal time models, functional traits

**Abstract (Currently 176--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 a 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 (minimum Tb, optimum To and maximum Tc) and thermal time. However, in the past calculating these traits has been problematic as it requires 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 systematically the breaking point in the data (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 or days it may take using alternative methods; and 3. It does not require the purchase of commercial statistical software. We provide an example of the application of the method and a help annex to guide users who are unfamiliar with R. Finally, we discuss the best policy to store and share the results of germination studies, in order to encourage cooperative 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, Ecology 97, 250-261) to vegetation science (Larson and Funk 2016) and biogeography (Bykova et al 2012, Journal of Biogeography 39, 2191-2200). (Fenner 2000, Grubb 1977, Jiménez-Alfaro et al 2016, Larson and Funk 2016, Walck et al 2011). 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, Biological Reviews). Germination traits are however underrepresented in public trait databases (Jiménez-Alfaro et al. 2016) and it is up to seed researchers 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 the germination phenotype 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 201, Trends in ecology & evolution, 30, 66-77).

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 at which the rate of germination is maximal. Below and above this optimum, the rate progressively decreases until the cue reaches base and ceiling thresholds beyond which germination stops. 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 need to be done with a seed population, thermal time is expressed as the degrees-day needed to produce germination in a percentage of this population, usually the 50 %. (Romo and Eddelman 1995). Therefore, the thermal control of germination can be described in a mechanistic way using a handful of seed traits: the three cardinal temperatures, sub-optimal (base temperature Tb), optimal (To) and supra-optimal (ceiling temperature Tc); and the thermal time. Thermal time traits have two powerful applications. First, 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 (Hardegree et al 1999). Second, thermal time parameters can be compared in a standard way across seed individuals, populations and species (Trudgill et al 2000). As such, thermal time models provide the broadest base to conduct comparative germination studies (Dürr et al 2015).

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 (n.d.)). This method has three advantages over methods currently in use: 1. It identifies systematically the breaking point in the data (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 or days it may take using alternative methods; and 3. It does not require the purchase of commercial statistical software. We test the method calculating cardinal temperatures in two examples; one in which the three cardinal temperatures can be identified, 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); so that out method can be completely transparent and rigorously reproducible by all researchers (Rocchini and Neteler 2012).

**Description of the method**

*Data preparation*

The dataset used here was assembled from the data presented by Fernández-Pascual et al. (Fernandez-Pascual et al 2015) and is given as a working example in ‘Supplementary Information 1’. 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 the two levels ‘*species A’* and ‘*species ‘B’*. The *Grouping* factor can be substituted for 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. populationxstratification). Although the example is presented with two species, the script is prepared to analyse larger numbers of factor levels. The second column, *‘Treatment’*, indicates the range of temperatures each species was tested by. The third column, *‘Dish’*, indicates the petri dish number. In this example dataset, there is only one *Dish* for each *Treatment*. In other cases, where an experiment may have several *Dishes* for each 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 the example, but can be in any other unit). 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. The sample seeds should always be the number of germinable seeds in the dish, and not the total number of seeds that may include empty or dead seeds.

*R and R packages needed*

All analyses are performed in the *R language and environment for statistical computing* (R Core Development Team (n.d.)) (Supplementary Information 2). The script we have prepared (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 packages must be installed into the *R* library. The packages *plyr* and *dplyr* (Wickham and Francois 2016) are used for all activities related to data manipulation and filtering.The package *binom* (Dorai-Raj (n.d.)) is used to manage and analyse binomial data which is a particularity of germination experiments. The package *drc* (Ritz et al 2015) is used to fit a dose-response model to the cumulative germination data, in order to calculate the germination times and rates. The package *segmented* (Vito (n.d.)) is used to fit a segmented regression and compute the cardinal temperatures and thermal times. Finally, the *ggplot2* package is used to create the figures (Wickham 2009)

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

The data is first grouped by germination treatment and dish, and then filtered into a new file to represent only the results of the final scoring date (i.e., the final germination proportions). Then, a function is created to estimate the mean final germination proportions and binomial confidence intervals and this function is applied across treatments to produce the first new dataset, *Table 1:‘FGP*’ (Final germination proportions). This new dataset is plotted in the first plot, *Figure 1:* *FGPfig*, which represents the total mean germination of each temperature treatment (Figure 1). At this point, *Figure 1* needs to be inspected visually for evidence that the temperature treatments used in the experiment resulted in representing the full germination temperature range of the study species. If this is the case, the segmented model we present here 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 fastest germination rate, ‘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 the response. This model is used to estimate the time required to reach successive germination percentages. probably the sensitivee, as poor estimations of the germination times will lead to inaccurate cardinal temperatures and thermal timesThere are several dose-response functions that can be fitted to the data (for more information see Ritz et al 2015). We have chosen as a default the log-logistic, because it gave the best fit in previous experiences with several datasets. Nonetheless users should check at this stage how well the log-logistic fits their own data, and compare it to alternative functions. The script tests automatically the fit of different functions, and suggests the model that best fits each treatment. This information is printed within the R workspace as the ‘FSfit’ table. The user needs to read this table and change the fitted function manually, if necessary (see comments within Supplementary Information 2). It is important to note that different groupings or treatments can have different functions that fit them best. However, it is important that the same model is used across all treatments for a balanced comparison. If there are discrepancies among groupings or treatments, we suggest to use the model that fits the majority of cases.

Once the best function is chosen, the model with the best fit is used to plot the cumulative germination of each treatment against time, and the function fit needs to be checked visually in our second figure *Figure 2:* *CGfig* (Figure 2). Users must decide if they can trust the fit of their curve to the data using both the information in table *‘FSfit’* and the curves produced in *Figure 2:* *‘CGfig’*. It can be possible that no good fit can be found, for example, if final germination is very low or if the scoring dates were inadequate (e.g., 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. 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, tyThe inverse of the time is then calculated to obtain the germination rate of each temperature treatment, and we create a second new dataset, *Table 2: GR* (Germination rates).

*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 (theta-sub, theta-supra). In this scenario, a segmented regression model is fitted to the data (Figure 3A). A segmented model allows the identification of breaking points in the data (i.e., the optimal germination temperature in 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). A function is set to fit a segmented regression and compute the intercept and the slope for each segmented relationship in the model. The function requires users to give an estimation of the breaking point or optimal temperature (‘psi=’), which can be based on the visual inspection of *Figures 1 & 2*. In this example we estimate the breaking point to be ‘psi=25’. Once the model is 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. The script calculates the cardinal temperatures and thermal times for each of the ten decile germination rates (10-90%). The script exports this in a third new dataset, *Table 3: CT* which compiles the estimated cardinal temperatures and thermal times. The plots with the germination rate versus temperature and the fitted models for every decile are exported in the third figure we create *Figure 3:* *BLfig*. *Figure 3A* allows to users to visually check the fit of the model to the data.

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 3:* *CT* cannot be produced. In the example of *species B* it is impossible to determine the Tc nor the To, and the segmented model cannot be fit to the data (Figure 3B), and it is 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. Here, we can produce a fourth new dataset, which indicates the Tb and thermal time calculated for *species B; Table 4: LM*. This will automatically produce information for whichever single segment was identified in each unique dataset, either the sub-optimal or the supra-optimal. We then plot the germination rate against each temperature treatment using a smooth linear model for each decile in *Figure 4B* (*LMfig*). Furthermore, it must be noted that this script can be used to calculate the germination base water potential, if the data set that is analysed represents an experiment of germination in different water potential solutions. In this case, the values of the column *Treatment* would be the water potentials instead of the temperatures. Lastly, we attach a short script to produce a smooth curve in the case of irregular data exploration in *SMTHfig*.

**Discussion**

The method we present here is appropriate for the analysis of germination data to identify cardinal temperatures and thermal time. This method offers an opportunity to calculate the analyses rapidly, and to select the breaking point in the data without user bias. When using this script we recommend that users become familiar with the basics of using R, and seek out one of the many introductory tutorials that exist online. We provide some guidance here (Supplementary Information 2 & 3). We suggest to ensure all packages are installed, and libraries uploaded, and follow the notes in the script using the example data sets in the supplementary information, using this paper as a complimentary resource to the script.

To engage in an experiment to identify cardinal temperatures, one must plan an experimental design which will purposefully identify all the germination temperature range, with data points below the optimal temperature (sub-optimal) and above the optimal germination temperature (supra-optimal), pushing each species to its thermal thresholds. In order to fit a segmented curve, there must be a minimum of three sub-optimal temperatures, and a minimum of three supra-optimal temperatures. It is thus recommended that experimental designs include a minimum of 5-6 temperatures for each to effectively explore potential germination temperature ranges. The less that is known about a given species, 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, Ellis et al 1986). However many wild species may require alternating temperatures, or alternating temperatures may better represent real ambient conditions found in nature (Carol C Baskin and Baskin 2014). When testing with alternating temperatures, it is recommended for plotting that the average temperature be taken of the two alternating temperatures for each treatment (Ellis and Barret 1994). Thus, alternating temperature regimes should be carefully planned to still result in temperature gradients when averaged, and may lead to odd results when one of the two alternating steps is outside the germination temperature range. 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 stops or 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 seeds (Pritchard et al 1999).

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 the NAtive Seed Science TEchnology and Conservation (NASSTEC) Initial Training Network (ITN) consortium. 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). The Royal Botanic Gardens, Kew, receive 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*.

Baskin CC and Baskin JM (2014) *Seeds: Ecology, Biogeography, and Evolution of Dormancy and Germination* (Second Edition.). Elsevier Inc., 1–1602.

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

Covell S, Ellis RH, Roberts EH and Summerfield RJ (1986) The Influence of Tmperature of Seed Germination Rate in Grain Legumes. *Journal of Experimental Botany* 37(178): 705–715.

Dorai-Raj S ((n.d.)) binom: Binomial Confidence Intervals For Several Parameterizations (1st edition). 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.

Ellis RH, Covell S, Roberts EH and Summerfield RJ (1986) The Influence of Temperature on Seed Germination Rate in Grain Legumes. *Journal of Experimental Botany* 37(183): 1503–1515.

Fenner M (2000) *Seeds: The Ecology of Regeneration in Plant Communities* (2nd edition). 1–423.

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* 115(2): 201–209.

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.

Grubb PJ (1977) The maintenance of species-richness in plant communities: the importance of the regeneration niche. *Biological Reviews* 52: 107–145.

Hardegree SP, Van Vactor SS, Pierson FB and Palmquist DE (1999) Predicting variable-temperature response of non-dormant seeds from constant-temperature germination data. *Journal of Range Management* 52(1): 83–91.

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* 1: 1–9.

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

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 ((n.d.)) R: language and environment for statistical computing. Vienna, Austria. Available at: http://www.R-project.org/.

Ritz C, Baty F, Sreibig JC and Gerhard D (2015) Dose-Response Analysis Using R. *PLoS ONE*.

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

Romo JT and Eddelman LE (1995) Use of Degree-Days in Multiple-Temperature Experiments. *Journal of Range Management* 48(5): 410–416.

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 ((n.d.)) 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/.

Walck JL, Hidayati SN, Dixon KW, Thompson K and Poschlod P (2011) Climate change and plant regeneration from seed. *Global Change Biology* 17(6): 2145–2161.

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

Wickham H and Francois R (2016) dplyr: A Grammar of Data Manipulation. Available at: http://CRAN.R-project.org/package=dplyr.



**Figure 1:** Final Germination proportions across all temperature treatments for species A and species B



**Figure 2:** Cumulative Germination Curves across all temperature treatments for species A and species B



**Figure 3:** Time to germination across each decile (%) of total germination, across each treatment using a segmented model for species A and species B.



**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