Hydrothermal-time-to-event models for seed germination

Title

Permalink
https://escholarship.org/uc/item/6qt6x8k8

Journal
European Journal of Agronomy, 101

ISSN
1161-0301

Authors
Onofri, A
Benincasa, P
Mesqaran, MB
et al.

Publication Date
2018-11-01

DOI
10.1016/j.eja.2018.08.011

Peer reviewed
Hydrothermal-time-to-event models for seed germination

Andrea Onofri\textsuperscript{a,b}, Paolo Benincasa\textsuperscript{b}, Mohsen B. Mesgaran\textsuperscript{b}, Christian Ritz\textsuperscript{c}

\textsuperscript{a} Department of Agricultural, Food and Environmental Sciences, University of Perugia, Perugia, Italy
\textsuperscript{b} Department of Plant Sciences, University of California, Davis, USA
\textsuperscript{c} Department of Nutrition, Exercise and Sports, Faculty of Science, University of Copenhagen, Copenhagen, Denmark

\textbf{A R T I C L E I N F O}

Keywords: Time-to-event data, Parametric models, Water potential, Rapeseed, Barley, Hordeum spontaneum, Phalaris minor

\textbf{A B S T R A C T}

Time-to-event methods have been proposed in the agricultural sciences, as one of the most suitable options for the analysis of seed germination data. In contrast to traditional linear/nonlinear regression, time-to-event methods can easily account for all statistical peculiarities inherent in germination assays, such as censoring, and they can produce unbiased estimates of model parameters and their standard errors. So far, these methods have only been used in combination with empirical models of germination, which are lacking biological underpinnings. We bridge the gap between statistical requirements and biological understanding by developing a general method that formulates biologically-oriented hydro time (HT), thermal time (TT) and hydrothermal time (HTT) models into a time-to-event framework. HT, TT, and HTT models are widely used for describing seed germination and emergence of plants as affected by the environmental temperature and/or water potential. Owing to their simplicity and the direct biological interpretation of model parameters, these models have become one of the most common tools for both predicting germination as well as understanding the physiology of germination responses to environmental factors. However, these models are usually fitted by using nonlinear regression and, therefore, they fall short of statistical rigor when inference about model parameters is of interest.

In this study, we develop HT-to-event, TT-to-event and HTT-to-event models and provide a readily available implementation relying on the package “drc” in the R statistical environment. Examples of usage are also provided and we highlight the possible advantages of this procedure, such as efficiency and flexibility.

\textbf{1. Introduction}

Time-to-event methods have been widely used to model the time until an event of interest occurs. Most frequently, these models have been used in medical sciences, to model the time to e.g., death (survival analysis), go out of remission, develop a certain pathology or other types of events. More recently, time-to-event methods have also appeared in the agricultural or crop sciences, e.g., to model the time-to-flowering (Ritz et al., 2010), the time-to-emergence (Onofri et al., 2010) or the time-to-germination (McNair et al., 2012; Onofri et al., 2011). In spite of few examples, however, time-to-event methods remain highly under-utilized in all disciplines relating to agriculture.

Several recent studies have shown that time-to-event methods provide a very general platform for the analyses of data from many types of germination experiments, leading to valid inferences and reliable hypotheses testing (Hay et al., 2014; Ritz et al., 2013). Indeed, germination assays naturally produce grouped time-to-event data (interval censoring): when we find \( n \) seeds germinated at a certain assessment time \( t_i \), we should only conclude that their germination timing must have occurred between \( t_{i-1} \) and \( t_i \). Grouping leads to loss of information or, in other words, added uncertainty; if this is neglected, standard errors will be underestimated and inferences will be unreliable (Ritz et al., 2013). In this respect, time-to-event methods are specifically devised to deal with all forms of censoring, as well as with the usual forms of experimental error, supporting the idea that they should always be preferred over linear and nonlinear regression to describe the progress to germination.

So far, time-to-event methods have only been used to empirically model cumulative seed germination curves, with little biological underpinnings. It is therefore relevant to use the time-to-event framework to build models that are both biologically meaningful and of good statistical quality. Specifically, we will focus on the use of time-to-event methods to describe the germination progress, as affected by environmental temperature and/or water potential.

The theoretical underpinning of hydro time (HT), thermal time (TT) or hydrothermal time (HTT) models is that germination does not take place below/above certain threshold temperature levels (base temperature: \( T_b \) or ceiling temperature: \( T_c \), respectively), or below a certain
water potential threshold (base water potential: \(\psi_0\)). When the ambient temperature or moisture level do not exceed these thresholds, the germination rate (GR: rapidity of germination in 1/days or 1/hours unit) for the \(g^\text{th}\) percentile of a population can be described as a linear or nonlinear function of water potential (\(\psi\)) and/or temperature (\(T\)). Due to presence of temperature and water potential thresholds, these models are also known as threshold models.

As an example, in an HT model, the germination rate of a given percentile \(g\) (\(GR_g\)) in response to water potential is described by (Bradford, 2002):

\[
GR_g = \frac{\psi - \psi_{bg}(g)}{\delta_h}
\]

where \(\psi\) is the water potential in the substrate, \(\psi_{bg}(g)\) is the base osmotic potential for the \(g^\text{th}\) percentile within the population and \(\delta_h\) (hydro-time constant) is the hydro-time to germination (in MPa h or MPa d unit) for the whole population.

In the same paper, Bradford (2002) also presents a TT model (see Eqs. (2) and (4) in his paper) and an HTT model (see Eqs. (9) and (10) in his paper; see also Alvarado and Bradford, 2002) where \(GR_g\) values linearly increase at sub-optimal temperatures and linearly decrease at super-optimal temperatures, with a sharp change at the optimal temperature level (\(T_o\)). Alternative models have been proposed to describe a curved relationship between \(GR_g\) and temperature around \(T_o\) (e.g. Grundy et al., 2000; Rowe and Finch-Savage, 2003; Mesgaran et al., 2017). More recently, the scope of threshold models has become more general including the effect of other environmental or endogenous factors on germination rates, such as hormones, ageing and oxygen (Bello and Bradford, 2016).

In general, threshold models for seed germination are well grounded in plant physiology. Their key aspect is that the \(GR_g\) for a given fraction of the population is expressed as a function of environmental variables, which is in contrast to what we really measure in a germination assay, that is the number of germinated seeds in different times after the beginning of the experiment. This raises the question as to how we should fit these GR-based models to the actual observed counts.

Thus far, two different approaches have been used: (i) fitting as a ‘two-steps’ procedure or (ii) re-parameterising the model. The first approach has been widely used, e.g., in Finch-Savage et al. (1998); Catara et al. (2016); Masin et al. (2017); Pace and Benincasa (2010) and Rowe and Finch-Savage (2003). In the first step, the observed counts are transformed into cumulative proportions and a sigmoidal model is fitted to these cumulative data using nonlinear least squares estimation. In the second step, the fitted sigmoidal model is used to derive the \(GR_g\) for the desired percentile \(g\) and these \(GR_g\) values are used to parameterise the selected HT, TT or HTT model. This two-steps approach may not be very efficient; first of all, nonlinear regression is used in the first step, which does not account for censoring. Secondly, some information from the first step will not be propagated to the second, i.e., uncertainty on estimated \(GR_g\) values is not carried forward. Third, it is also a limitation that only one subpopulation percentile can be considered at a time (e.g., \(GR_{R_0}\), \(GR_{R_0}\) or \(GR_{R_0}\)).

In the second approach, the dependent variable is the proportion/percentage of germinated seeds, instead of \(GR_g\), and threshold models are re-parameterised based on the assumption that one or several threshold parameters (e.g., base water potential) vary between individuals within the population, following a specific probability distribution (e.g., Bradford, 2002; Mesgaran et al., 2013 and Watt et al., 2010). For instance, if the distribution of base water potential is assumed to be normal, it is easy to show that Eq. (1) can be re-parameterised as follows:

\[
\psi(t, \psi) = \frac{\Psi - \Psi(t, \psi_{50})}{\sigma_{\psi}}
\]

where \(\psi\) is the proportion of germinated seeds at time \(t\) and \(\phi\) is the cumulative normal distribution (Bradford, 2002). Mesgaran et al. (2013) and Watt et al. (2010) have followed the same approach, but using different distributions i.e., log-logistic and Weibull probability distributions, respectively. In all cases, the re-parameterised model is fitted to the observed proportions by using some optimisation algorithm, such as nonlinear least squares (perhaps the most common choice in the literature), repeated probit analysis, or similar procedures (see Hardegree et al., 2015 for more details). These procedures either do not produce reliable standard errors or do not produce them at all.

Although HT, TT, and HTT models are well grounded in seed biology, they are often fitted by using inefficient or even questionable methods, not respecting the actual manner in which data are acquired from germination assays. Time-to-event methods can easily account for all statistical peculiarities inherited in germination assays, but no systematic effort has been so far made to build HT, TT, and HTT models in this framework, apart from a preliminary attempt of Pipper et al. (2013) that only included a TT model.

To overcome this dichotomy in modeling approaches, the objectives of this study were to:

1. develop a general method to re-formulate the commonly used HT, TT, and HTT models within a fully parametric time-to-event framework;
2. implement our models within the R statistical environment and, in particular, the package “drc”, which is commonly used for dose-response analysis in various other areas of agricultural research (Ritz et al., 2015);
3. examine the performance of these time-to-event models through a number of exemplary datasets;
4. highlight the advantages and limitations of the time-to-event approach against the other approaches that are currently used for modelling germination in response to temperature and water availability.

2. Materials and methods

We collated published and unpublished data on seed germination of four plant species from independent experiments, as described below.

2.1. Example 1: germination of rapeseed at different water potentials

This dataset was taken from previously published work (Pace and Benincasa, 2010) with rapeseed (Brassica napus L. var. oleifera, cv. Excalibur). Thirteen different osmotic potentials (−0.03, −0.15, −0.3, −0.4, −0.5, −0.6, −0.7, −0.8, −0.9, −1, −1.1, −1.2, −1.5MPa) were created by using a polyethylene glycol solution (PEG 6000). For each water potential level, three replicated Petri dishes with 50 seeds each were incubated at 20 °C. Germinated seeds were counted and removed every 2–3 days for 14 days.

2.2. Example 2: germination of Hordeum spontaneum [C. Koch] Thell. at different temperatures and water potentials

The second dataset was obtained from previously published work (Mesgaran et al., 2017) with Hordeum spontaneum [C. Koch] Thell. The germination assay was conducted using four replicates of 20 seeds tested at six different water potential levels (0, −0.3, −0.6, −0.9, −1.2 and −1.5 MPa). Osmotic potentials were produced using variable amount of polyethylene glycol (PEG, molecular weight 8000) adjusted for the temperature level. Petri dishes were incubated at six constant temperature levels (8, 12, 16, 20, 24 and 28 °C), under a photoperiod of 12 h. Germinated seeds (radicle protrusion >3 mm) were counted and removed daily for 20 days.
2.3. Example 3: germination of Phalaris minor L. at different temperatures

The third dataset was also taken from Mesgaran et al. (2017). Four replicates of 25 seeds of *P. minor* were placed in 9-cm-diameter Petri dishes. Originally, this experiment consisted of six different water potential levels, although here we will only use the data observed at −0.6 MPa, because at this water potential level the germination time-course was well depicted at many temperature levels. Seeds were incubated at six constant temperature levels (8, 12, 16, 20, 24 and 28 °C) under a photoperiod of 12 h. Germinated seeds (radicle protrusion >3 mm) were counted and removed daily for 20 days.

2.4. Example 4: germination of barley at different temperatures

This unpublished dataset was obtained from a germination assays performed at the Department of Agricultural, Food and Environmental Sciences (University of Perugia, Italy), by using *Hordeum vulgare* L., cv. Quenc. Three replicates of 50 seeds were placed over filter paper (Watman #1) in 13-cm-diameter Petri dishes and moistened with 9 ml of distilled water i.e., 0 MPa. Petri dishes were sealed by plastic bags to prevent water evaporation and incubated in climatic chambers at 9 constant temperature levels (1, 3, 7, 10, 15, 20, 25, 30, 35, 40 °C). Germinated seeds were counted and removed daily for 10 days.

2.5. Principles of time-to-event models

In time-to-event modeling, timing of germination within the seed population is considered to follow some types of probability density functions, such as log-normal, log-logistic or Weibull. As a result, the cumulative proportion of germinated seeds (*P*) increases over time according to a S-shaped curve, such as a log-logistic cumulative probability function (if a log-logistic distribution is assumed):

\[ P(t) = \frac{P_{\text{MAX}}}{1 + \exp\left[b\left(\log(t) - \log(t_0)\right)\right]} \]  

(3)

In the above equation, *t₀* is the median germination time, *P*_{MAX} is the maximum germinated proportion, and *b* is the slope at the inflection point, which is related to the standard deviation of the log-logistic distribution. Indeed, the three parameters describe the main features of germination, i.e., germination speed, capability, and uniformity, respectively.

The above curve (Eq. 3) can be fitted to germination data in two ways. For better clarity, let us consider a hypothetical example of a germination assay where one seed germinated at some (unknown) time point between 0 and 3 Days After the Beginning of assay (DAB), six seeds germinated between 3 and 6 DAB, and so on (Table 1). At the end of the assay, 8 seeds were still ungerminated and, therefore, their germination times, if any, would be larger than 16 days (i.e., from 16 to 37). Considering all the seeds in the lot, the log-likelihood of the likelihood (LL) for the data in Table 1 is therefore:

\[ LL = 1 - \log[P(3) - P(0)] + 3 - \log[P(6) - P(3)] + + 12 - \log[P(12) - P(6)] + 1 - \log[P(16) - P(12)] + + 8 - \log[1 - P(16)] \]

By repeating the above calculations, we can see that the log-likelihood is, e.g., higher with parameter values fixed at *t₀* = 7, *b* = 4 and *P*_{MAX} = 0.95 (LL = −37.5), than with parameter values fixed at *t₀* = 3, *b* = 4 and *P*_{MAX} = 0.8 (LL = −43.3), suggesting that former set of parameter values is more likely given the observed counts. By using some sort of optimisation algorithm, we can retrieve the optimal combination of parameters that results in the maximum log-likelihood. For this example data, the maximum log-likelihood estimates were (standard errors in brackets) *t₀* = 7.72 (SE = 1.07), *b* = 4.02 (SE = 1.28) and *P*_{MAX} = 0.72 (0.11), while nonlinear regression using the least squares method gave parameter estimates of *t₀* = 7.63 (SE = 0.75), *b* = 4.95 (SE = 1.68) and *P*_{MAX} = 0.70 (0.06). Point estimates are very similar, but, with the exception of *b*, standard errors obtained from the time-to-event fit are larger than those obtained from nonlinear regression, as the uncertainty relating to censoring has been correctly incorporated into the time-to-event fit. It is worth noting that, with both methods, we are fitting the same equation (Eq. (3)), although in two totally different ways (using two different metrics).

2.6. Building a general hydrothermal-time-to-event model

The time-to-event model in Eq. (3) can be made more general, by using any types of cumulative distribution function *φ*, as follows:

\[ P(t, \Psi, T) = P_{\text{MAX}} \Phi[b\left(\log(t) - \log(t_0)\right)] \]  

(4)

where *tᵢ* is the 50th percentile (*t₀*) if *φ* is e.g., log-normal or log-logistic or some other specific percentiles in other cumulative distribution functions (e.g., 63rd in Weibull distribution). Using the above general formulation, we can recast HT, TT, and HTT models as time-to-event models by:

1. selecting an appropriate distribution for germination times (*φ*) and, 2. expressing *tᵢ* = 1/GRₜ₀, *b* and *P*_{MAX} as linear/nonlinear functions of temperature (T in °C) and/or water potential (*Ψ* in MPa) in the substrate.
If we write $GR_a = f(\Psi, T)$, $P_{MAX} = h(\Psi, T)$ and $b = z(\Psi, T)$, a general HTT-to-event model is:

$$P(t, \Psi, T) = h(\Psi, T)\Phi[z(\Psi, T)[\log(t) - \log(1/f(\Psi, T))]] \quad (5)$$

This modelling approach attains a very high level of flexibility: models can be easily built by appropriately changing one or more of the above model components, i.e., $\Phi$, the functions $h$ ($P_{MAX}$ sub-model), $f$ ($GR_a$ sub-model), and $z$ ($b$ sub-model). In this work, we initially built one HT-to-event (HTE), one HTT-to-event (HTTE), and one TTE-to-event (TTE) model. Relating to the $GR_a$ component, we used the equations provided by Bradford (2002) and Mesgaran et al. (2017) (Eqs. (6), (9) and (12) in Table 2). Several forms of a shifted exponential distribution function were used for $P_{MAX}$ (Eqs. (7), (10) and (13) in Table 2), while $b$ was always found to be largely independent of temperature and water potential. As a good candidate for $\Phi$, throughout this paper we will use the log-logistic distribution (Eqs. (8), (11) and (14) in Table 2), though any other distribution can be used without loss of generality.

The whole model building process is detailed in the Appendix A. We would like to emphasize that, for all time-to-event models, $GR_a$ refers to the 50th percentile of the germinated fraction and not to the whole seed lot. This is an important difference with respect to the original HT or HTT models, as formulated in Bradford (2002), even though the $GR_a$ for the whole seed lot can also be retrieved from the time-to-event fitting (see later). Furthermore, in every time-to-event model, germination velocity and capability are described by two different sub-models, although they are not totally independent, as they partly share the same parameters (Table 2).

### 2.7. Model implementation

All the models and their sub-models, including the appropriate self-starting routines, were incorporated into the package “drc” (Ritz et al., 2015) within the R statistical environment (R Core Team, 2016). Maximum likelihood estimation requires initial guesses for all model parameters, which, in our experience, is often a troublesome step. The self-starting routines implemented in our modeling platform can, to large extent, alleviate this problem.

User-defined models can also be inserted into the “drc” package as shown in the example related to a TT time-to-event model (Appendix B). Furthermore, all functions and datasets (“Rape”, “Hordeum”, “Phalaris” and “Barley”) are available within the package “drcSeedGerm”, which can be installed and loaded to repeat the analyses. The codes to reproduce all the analyses presented in this paper are given in the Appendix C.

### 2.8. Model fitting

In order to perform time-to-event analyses in “drc”, the dataset needs to be organised as outlined in Table 1. After loading the data, the equations can be fit to the whole dataset, by using the function “drm()” in the “drc” package and by setting the argument “type” to “event” in the function call. The estimated parameters and standard errors can be retrieved by using the function “summary()” and passing the model object as an argument (See Appendix C).

It should be noted that in most germination assays, the observational units are clustered within Petri dishes and thus they are not independent. It has been shown that random variability at the Petri dish level may be very small and negligible from a practical point of view (Onofri et al., 2014). As suggested for clustered survival data (Yu and Peng, 2008), we decided to calculate cluster robust sandwich standard errors (Carroll et al., 1998), which can be easily done by using the facilities included in the “sandwich” package (Berger et al., 2017), together with the function “coeftest” in the “lmtest” package (Zeileis and Hothorn, 2002). Code snippets are given in the Appendix C.

The goodness of fit was checked graphically, while the Akaike’s Information Criterion (AIC; Akaike, 1974) was used to compare models, wherever necessary.

### 2.9. Making predictions

Our time-to-event modeling approach retains the same capacity as the conventional GDD and HTT models in predicting seed germination or seedling emergence (e.g., in the field) based on the daily soil temperature and moisture data. Using Eq. (9) (Table 2), as an example, we can derive an expression for the daily HTT accumulation ($HTT$)

$$HTT_i = (T_i - T_{0})[\Psi - \Psi_t - k(T - T_i)]$$

and incorporate it into Eq. (11) to obtain a prediction of the proportion of germinated seeds at day $n$:

$$P(t, \Psi, T) = \frac{h_t(\Psi, T)}{1 + \exp\left[-\left(\frac{\sum_{i=1}^{n} HTT_i}{\theta_{HTT}} \right)\right]} \quad (15)$$

When the daily accumulation of HTT reaches the value of $\theta_{HTT}$, the proportion of germinated seeds will be 50% of the maximum germinable fraction i.e., $P_{MAX}$, where $P_{MAX}$ itself can be predicted using Eq. (10).

Predicting $GR$ for any percentile $g$ can be achieved by using the following equation:

$$GR_g = \exp\left(-\frac{1}{b} \log\Delta + \log\left(\frac{1}{f(T, \Psi)}\right)\right) \quad (17)$$

where $\Delta$ is equal to $(1-g)/g$ or $[h(t, T, \Psi) - g]/g$, depending on whether the percentile $g$ refers to the germinated fraction or to the whole population. Predicting the $GR_a$ levels for the whole population is fundamental whenever we intend to compare seed populations with a different $P_{MAX}$.

---

**Table 2:** List of sub-models used for hydro-time-to-event (HTE), hydrothermal-time-to-event (HTTE), and thermal-time-to-event (TTE) models of seed germination.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sub-model</th>
<th>Equation</th>
<th>Eq. No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTE</td>
<td>$GR_a$</td>
<td>$f_1(\Psi) = \frac{\Psi - \Psi_t}{\Psi}$</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>$P_{MAX}$</td>
<td>$h_1(\Psi) = G\left[1 - \exp\left(-\frac{\Psi - \Psi_t}{\sigma_\Psi}\right)\right]$</td>
<td>(7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Distribution $P_1(t, \Psi, T) = \frac{h_1(\Psi)}{1 + \exp[\log(t) - \log(1/f(\Psi, T))]}$</td>
<td>(8)</td>
</tr>
<tr>
<td>HTTE</td>
<td>$GR_a$</td>
<td>$f_2(\Psi, T) = \frac{T - T_0}{T_0 - T_{0}}[\Psi - \Psi_t - k(T - T_0)]$</td>
<td>(9)</td>
</tr>
<tr>
<td></td>
<td>$P_{MAX}$</td>
<td>$h_2(\Psi, T) = G\left[1 - \exp\left(-\frac{\Psi - \Psi_t - k(T - T_0)}{\sigma_\Psi}\right)\right]$</td>
<td>(10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Distribution $P_2(t, \Psi, T) = \frac{h_2(\Psi, T)}{1 + \exp[\log(t) - \log(1/f(\Psi, T))]}$</td>
<td>(11)</td>
</tr>
<tr>
<td>TTE</td>
<td>$GR_a$</td>
<td>$f_3(T) = \frac{T - T_0}{T_0 - T_{0}}\left[1 - \frac{(T - T_0)}{(K - T_0)}\right]$</td>
<td>(12)</td>
</tr>
<tr>
<td></td>
<td>$P_{Hb}$</td>
<td>$h_3(T) = G\left[1 - \exp\left(-\frac{\Psi - \Psi_t}{\sigma_\Psi}\right)\right]$</td>
<td>(13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Distribution $P_3(t, T) = \frac{h_3(T)}{1 + \exp[\log(t) - \log(1/f(\Psi, T))]}$</td>
<td>(14)</td>
</tr>
</tbody>
</table>

$\Psi$: water potential (MPa); $T$: temperature (°C); $t$: time (days or hours); $\Psi_t$: base water potential (MPa); $\Theta_H$: hydrotiming constant (MPa day or MPa hour); $G$: maximum proportion of germinated seeds; $b$: slope of germination curve; $T_0$: base temperature (°C); $T_i$: ceiling temperature (°C); $k$, $\sigma_\Psi$, and $b$ are regression coefficients; see Appendix A for more detail. Eq. (6) is taken from Bradford (2002), Eq. (9) is taken from Mesgaran et al (2017), while Eq. (12) is derived from Eq. 9.
The function "drm()", embedded in the "drc" package, includes a "predict()" method to obtain the predicted proportions of germinated seeds as a function of time, temperature and water potential, while we modified the function "ED()" of the "drc" package to accommodate prediction of GR values as a function of temperature and water potential.

### 3. Results

#### 3.1. Example 1

The estimated parameters are given in Table 3, together with two sets of standard errors: the first set is obtained from the simple time-to-event model fit, assuming independence between all germination times, while the second set (sandwich standard errors) accounts for the fact that seeds are clustered within Petri dishes. We can see that the two sets of standard errors are not much different from each other.

As shown in Fig. 1, the model fitted the rapeseed germination data reasonably.

The HTE model matched closely the observed $P_{MAX}$ (Fig. 2, left) and the observed median germination rates (Fig. 2 right). For this latter variable, the figure reports only the water potential levels for which the $GR$ could be derived with enough precision (see also Fig. 1).

As mentioned above, in time-to-event models the $GR$ stands for the

![Graph showing germination data and model fits](image)

**Table 3**

Parameter estimates and standard errors for the HTE model (Eq. 8) fitted to the rapeseed data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Naive SE</th>
<th>Sandwich SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G$</td>
<td>0.958</td>
<td>0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>$\psi_b$</td>
<td>-1.040</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>$\sigma_{\psi}$</td>
<td>0.111</td>
<td>0.009</td>
<td>0.012</td>
</tr>
<tr>
<td>$\theta_H$</td>
<td>0.906</td>
<td>0.030</td>
<td>0.041</td>
</tr>
<tr>
<td>$\beta$</td>
<td>4.027</td>
<td>0.196</td>
<td>0.193</td>
</tr>
</tbody>
</table>

**Fig. 1.** Effect of water potential on the time-course of germinations in rapeseed (*B. napus* var. *Excelsior*; Example 1). Symbols denote the observed data and lines show the HTE model fits (Eq. (8); model parameters are reported in Table 3) with vertical bars representing the 95% confidence limits.
Table 4
Parameter estimates and standard errors for the HTTE model (Eq. (11)), fitted to the H. spontaneum data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G$</td>
<td>0.99</td>
<td>0.012</td>
</tr>
<tr>
<td>$\psi_b$</td>
<td>-2.91</td>
<td>0.035</td>
</tr>
<tr>
<td>$k$</td>
<td>0.07</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_k$</td>
<td>-0.75</td>
<td>0.353</td>
</tr>
<tr>
<td>$\sigma_b$</td>
<td>0.55</td>
<td>0.029</td>
</tr>
<tr>
<td>$\theta$</td>
<td>1309.06</td>
<td>40.638</td>
</tr>
<tr>
<td>$b$</td>
<td>4.17</td>
<td>0.113</td>
</tr>
</tbody>
</table>

germinated fraction. For instance, if $P_{MAX}$ is 0.9, the $GR_{50}$ corresponds to the 50th percentile of the germinated fraction, that is the 45th percentile of the whole population. Owing to this, it would not be appropriate to compare $GR_s$ values across populations with a different $P_{MAX}$. However, the function “ED()” can be used in “drc” to calculate $GR_s$ values both for the germinated fraction and for the whole population. Examples are given in the Appendix C.

If we consider the $GR_s$ values for the whole population (Fig. 2 right), the relationship between water potential and $GR_s$ is no longer linear, and base water potential varies among subpopulations, being more negative for the quicker percentiles. We see that this time-to-event model retains at least part of the biological assumptions made in Bradford (2002).

3.2. Example 2

Fitting the hydro-thermal-time-to-event model (Eq. (11)) is also straightforward and only required small changes in the R codes, as shown with the dataset from H. spontaneum (see Appendix C). The parameter estimates (cluster-robust standard errors in brackets) are reported in Table 4.

As shown in Fig. 3, the HTT time-to-event model provided good fits to cumulative germination data from all temperature by moisture combinations. Changes in germination capacity (i.e., $P_{MAX}$) and rapidity (i.e., $GR$) in responses to temperature were also well described by the sub-models (Fig. 4). Considering the $GR_s$ based on the whole population, we see that $T_c$ varies among germination percentiles and is warmer for early germinating seeds (Fig. 4, right).

3.3. Example 3

To construct a thermal-time-to-event model, we simply removed the parameters related to the effects of water potential from Eq. (11). The modified model (Eq. (14)) fitted well to the germination data of P. minor (Example 3), with some deviations observed at 12 °C (Fig. 5), probably due to the existence of distinct subpopulations in the seed lot. For this species, some parameters could not be estimated with high precision (Table 5).

3.4. Example 4: a more complex analysis

For the fourth example, the TTE model, which well fitted to the P. minor data (Fig. 5), failed to adequately describe the germination pattern in barley (Fig. 6, dashed, blue line). This example, therefore, represents a case where one needs to resort to the ‘empirical’ process of model building from the scratch. As the first step, we fitted the logistic distribution function to germination time courses of each temperature level separately, to obtain separate estimates of the three main features of seed germination ($P_{MAX}$, $GR_{50}$ = 1/$\theta_b$ and $b$). The code to accomplish this is reported in the Appendix C. We then investigated the shape of responses to temperature for these three germination parameters. For example, plotting the estimated $GR_{50}$ values against temperature showed a slow rise with temperature up until 30 °C, followed by a quick drop afterwards, down to 0 at 40 °C (Fig. 7).

In order to account for this behavior, we used a model derived from Rowse and Finch-Savage (2003). In this model, the negative effect of temperature on $GR$ takes place only at $T = T_d > T_r$, where $T_d$ is a temperature threshold close to, but not equal to the optimal temperature, $T_r$:

$$GR_{50} = f_d(T) = \begin{cases} \frac{T - T_d}{\sigma_d} & \text{if } T_d < T < T_r \\ \frac{T - T_r}{\sigma_d} & \text{if } T_d < T < T_r \\ 0 & \text{if } T \leq T_d \text{ or } T \geq T_r \end{cases}$$

(18)

In contrast to Eq. 12, this model predicts a curvilinear and asymmetric trend for the response of $GR$ to temperature. Accordingly, the new TTE model is:
The above model is defined for temperature levels between \( T_b \) and \( T_c \), while it is 0 for temperatures \( \leq T_b \) or \( \geq T_c \). Using above equation (Eq. (19)) with its modified sub-model for GR, we achieved a sensibly improved fit (Fig. 6, solid, red line) compared with Eq. (14) (Fig. 6, dashed, blue line), as evidenced by a decrease in the AIC value (6,479 vs 5,731). The final parameter estimates are reported in Table 6.

Although we could achieve further improvement in fit by imposing a linear decrease in \( 1/b \) with temperature (see the black line in Fig. 6), the additional parameter and increased complexity may not be warranted from exploration of a single dataset. We therefore used the more parsimonious model, with a constant parameter \( b \), as this model also gave a good description of the temperature-dependent variations in \( P_{\text{MAX}} \) and \( GR_{50} \) (Fig. 7).

### 4. Discussion and conclusions

Several excellent population-based threshold models exist to assess the effect of temperature, moisture content and other environmental variables on seed germination. These models incorporate physiologically relevant threshold parameters that are fundamental for understanding the responses of plants to their environments. In this paper, we used four such models, i.e., the HT model by Bradford (2002), the HTT model by Mesgaran et al. (2017) and two TT models, derived by Mesgaran et al. (2017) and Rowse and Finch-Savage (2003), respectively. These models are based on the GR for a certain population fraction as the response variable and, therefore, they do not immediately comply with the results of germination assays, which are based on the counts of germinated seeds over time. For this reason, fitting these models has usually been performed either: (i) by a two-steps fitting procedure (first-step: use the counts to derive \( GR_g \) values; second-step: fit the threshold model), or (ii) by re-formulating threshold models as nonlinear regression models, based on the time-course of the...
proportion/percentage of germinated seeds.

We showed, by way of several examples, that there is a third possible option, that has been, so far, largely neglected. In detail, we showed how the HT, HTT, and TT models can be formulated into a time-to-event framework. We also showed a general method to accomplish this task and defined a general hydrothermal-time-to-event model.

To make our modeling approach readily available, we implemented these models within the “drc” package in the R environment. We chose the R environment and the “drc” package because they are free and the function “drm()” can be used for various parametric time-to-event model and, moreover, it can easily account for the final fraction of ungerminated seeds, which is not possible in other time-to-event frameworks and packages, such as the “survival” package (Therneau, 2012). Besides R, the time-to-event models described in this paper can also be fitted by using the NLMIXED procedure of SAS, although this

---

**Fig. 4.** Effect of temperature on the germination characteristics of *Hordeum spontaneum* [C. Koch] Thell. (Example 2). Symbols denote the observed data and lines show the HTTE model fit (Eq. (11); model parameters are reported in Table 4). Left: maximum germinated proportion ($P_{MAX}$); Right: germination rate ($GR$) for the 20th, 50th and 80th percentiles for the whole population. The example data shown above are from seeds tested at $\Psi = -0.6$ where the differences between germination percentiles were more visible than with other water potentials.

**Fig. 5.** Effect of temperature on the time-course of germinations in *Phalaris minor* (Example 3). Symbols denote the observed data and lines show the TTE model fits (Eq. (14); parameter estimates are reported in Table 5), with vertical bars representing 95% confidence limits.
requires some manual work.

Time-to-event models may have several advantages over other modeling approaches, but, at present, they remain underutilized in seed biology studies. Compared with the two-steps fitting, time-to-event models can better represent the germination performance of the seed population, as they make use of the whole dataset to describe the time-course of germination and to estimate model parameters. On the contrary, with the two-steps fitting approach much information in the first step will not be propagated to the second, as we need to work only with one subpopulation at a time (e.g., $GR_{50}, GR_{30}$, or $GR_{10}$). Of course, we do not intend to totally rule out the two-steps approach: very recently Jensen et al. (2017) proposed a meta-analytic model for seed germination assays, where the information available in the first step can be re-used in the second one. This approach may be useful when dealing with complex experimental designs, although further work is needed to test whether this is also applicable to the parameterisation of HT, TT, and HTT models.

Time-to-event models can be used to predict the time course of germination. In this respect, they are very similar to nonlinear regression models, such as Eq. (2), or similar equations (e.g., Mesgaran et al., 2013, 2017, and Rowse and Finch-Savage, 2003). However, there are several important conceptual differences between these two modeling platforms. For example, Eq. (2) and other similar nonlinear regression models are fitted to the cumulative proportions of germinated seeds, while time-to-event models are parameterised by using the observed counts. We have seen that this is an advantage of time-to-event models, because the fitting method fully respects the actual manner in which data are acquired from germination assays and produces reliable predictions.

### Table 5
Parameter estimates and standard errors for the TTE model (Eq. (14)), fitted to the *P. minor* data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U$</td>
<td>1.00</td>
<td>0.081</td>
</tr>
<tr>
<td>$\tau_C$</td>
<td>48.80</td>
<td>3.933</td>
</tr>
<tr>
<td>$\sigma_T$</td>
<td>33.37</td>
<td>5.557</td>
</tr>
<tr>
<td>$\theta_T$</td>
<td>3.35</td>
<td>0.525</td>
</tr>
<tr>
<td>$\theta_T$</td>
<td>1655.56</td>
<td>128.040</td>
</tr>
<tr>
<td>$b$</td>
<td>3.16</td>
<td>0.081</td>
</tr>
</tbody>
</table>

**Fig. 6.** Effect of temperature on the time-course of germinations in barley (Example 4). Symbols denote the observed data; the solid, red lines show the TTE model fits (Eq. 19; Rowse and Finch-Savage, 2003; parameter estimates are reported in Table 6). The dotted, blue lines show the fit from a simpler TTE model (Eq. (12); Mesgaran et al., 2017), while the black, dashed lines represent the improvement in fit when the germination uniformity parameter (i.e., $b$) was allowed to vary as a function of temperature (model parameters for these latter two models are not shown) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
percentiles for the whole population. Models that make fewer a priori assumptions about the shape of response might therefore describe the germination pattern of all plant species. Models that make fewer a priori assumptions about the shape of response might therefore describe the germination pattern of all plant species. Thus, flexible and capacity. This is usually a less parsimonious approach, but it may be more flexible or more realistic. Furthermore, the two sub-models are not totally independent, as they are partly based on the same parameters.

Another important aspect of the Eq. 2 is that it considers the percentiles in relation to the whole seed population, including the ungerminated fraction. Therefore, germination velocity and capacity are modelled altogether, based on the distribution of base water potential within the population. The advantage of such an approach is that the resultant model becomes very parsimonious, although it may be “rigid” in some instances. On the contrary, time-to-event models decouple the germinated and the ungerminated fractions and, accordingly, two different sub-models are used for characterisation of germination velocity and capacity. This is usually a less parsimonious approach, but it may be more flexible or more realistic. Furthermore, the two sub-models are not totally independent, as they are partly based on the same parameters.

Flexibility can be a crucial aspect for further development and usability of HTT models, as no single equation has so far proven able to describe the germination pattern of all plant species. Models that make fewer a priori assumptions about the shape of response might therefore be preferred (Hardegree and Winstral, 2006), as they can provide more flexibility as well as better prediction power. In this vein, some authors have recently attempted a totally non-parametric approach to seed germination modeling (Gonzalez-Andujar et al., 2016). Although flexible, the non-parametric methods resemble a black box with little biological underpinnings and are prone to overfitting. Building parametric models with high flexibility is therefore needed to both improve the model predictive power and gain a better understanding of the biology of seed germination. Reformulating HTT, TT, or HT models into their time-to-event equivalents is a stride to address the above need: we can represent virtually every type of relationships between germination and temperature or water potential by simply changing the shape of the cumulative distribution of germination times and the sub-models for $GR_{00}$, $P_{MAX}$ and $b$, while maintaining the same biological interpretation of germination parameters.

Despite their vast usefulness to accommodate almost all types of data analysis in seed germination/emergence studies, time-to-event models have largely been overlooked, perhaps because, in their formal format, they are rather statistical than biological models. In this study, by incorporating threshold models into a time-to-event platform, we attempted to overcome this apparent gap in modeling approaches and devised germination models that are both statistically robust and biologically comprehensible. Future work based on a comprehensive meta-analysis should say whether the biological assumptions behind time-to-event models are realistic and can address a wide range of germination patterns. In this respect, in our “drcSeedGerm” package, we have also implemented most GR-based threshold models (e.g. Eq. (1) and other similar equations) and re-parameterised nonlinear regression models (e.g. Eq. (2) and other similar equations). We are confident that our R codes should enable biologists, including those who are less skilled in statistics, to use and experiment with these time-to-event methods and compare them with other more traditional germination models.

**Acknowledgment**

The research was partially supported by the funds for the Basic Research 2015 of the University of Perugia (Italy).

**Appendix A. Supplementary data**

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.eja.2018.08.011.

**References**

Contr. 19, 716–723.
URL: https://CRAN.R-project.org/package=survival (Date of Last access: 12/09/18).