Title
Hydrothermal-time-to-event models for seed germination

Authors
Onofri, Andrea
Benincasa, Paolo
Mesgaran, Mohsen B
et al.

Publication Date
2018-11-01

DOI
10.1016/j.eja.2018.08.011

Peer reviewed

European Journal of Agronomy 101 (2018) 129–139

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Hydrothermal-time-to-event models for seed germination

Andrea Onofri[a,*], Paolo Benincasa[a], Mohsen B. Mesgaran[b], Christian Ritz[c]

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

ARTICLE INFO

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

ABSTRACT

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 inherited 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.

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

* Corresponding author at: Department of Agricultural, Food and Environmental Sciences, University of Perugia, Borgo XX Giugno 74, 06121, Perugia, Italy.
E-mail address: andrea.onofri@unipg.it (A. Onofri).

https://doi.org/10.1016/j.eja.2018.08.011
Received 12 June 2018; Received in revised form 20 July 2018; Accepted 31 August 2018

water potential threshold (base water potential: ψ_b). 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^{th} percentile of a population can be described as a linear or nonlinear function of water potential (ψ) 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):

$$GR_g = \frac{\Psi - \Psi_{b(g)}}{\theta_H} \quad (1)$$

where Ψ is the water potential in the substrate, $\Psi_{b(g)}$ is the base osmotic potential for the g^{th} percentile within the population and θ_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; Rowse 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 Rowse 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 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 estimatedGR_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_{50}, GR_{30} or GR_{10}).

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:

$$p(t, \psi) = \phi\left\{\frac{\left[\Psi - \frac{\theta_H}{t} - \Psi_{b(50)}\right]}{\sigma_{\Psi_b}}\right\} \quad (2)$$

where p is the proportion of germinated seeds at time t and ϕ 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. develope 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, −1.5 MPa) 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):

$$P(t) = \frac{P_{MAX}}{1 + exp\{b[log(t) - log(t_{50})]\}} \quad (3)$$

In the above equation, t_{50} 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 ∞; see Table 1).

The most common option for modelling the data presented in Table 1 is to fit Eq. 3 to the observed cumulative proportion of germinated seeds, by using the method of nonlinear least squares. In this case, the time point when an observation is made (second column in Table 1) is erroneously assumed to be the same as the timing of the germination event, while the exact germination time is unknown and we only know that the event has happened at some time point between two consecutive inspections (interval censoring; see below).

Table 1
A hypothetical example showing germination counts recorded between consecutive monitoring events.

Beginning of interval	End of interval	No. of germinated seeds	Cumulative proportion
0	3	1	0.05
3	6	3	0.20
6	12	12	0.80
12	16	1	0.85
16	Inf	8	–

The alternative (correct) option is to use directly the combined information about the counts of germinated seeds and the length and position (on the time axis) of the corresponding time interval. In this way, it is recognised that the seeds have germinated in an unknown moment between the present and the previous scoring time. Obviously, the variation of counts over time should follow the underlying distribution of germination times. Therefore, the best-fitting version of Eq. (3) is obtained by selecting the values for parameters (t_{50}, b, and P_{MAX}), so that the observed counts are as likely as possible, i.e., by means of maximum likelihood estimation. General (logarithm-transformed) likelihood expressions for time-to-event data have been provided by Onofri et al. (2011) and Ritz et al. (2013).

For the data presented in Table 1, if we assume that the cumulative proportion of germinated seeds at time t [i.e., (t)] is described by the log-logistic model (Eq. (3)), the probability of having one seed germinated e.g., in the second interval (from 3 to 6 DAB) is equal to $P(6)-P(3)$. Likewise, the probability of having one seed ungerminated at the end of the assay would be the sum of two probabilities: that is, the probability that this seed has a germination time longer than the last assessment time [i.e., $P_{MAX}-P(16)$] plus the probability that the seed is in the ungerminated fraction [i.e., $1-P_{MAX}$]. Considering all the seeds in the lot, the logarithm of the likelihood (LL) for the data in Table 1 is therefore:

$$\begin{aligned} LL = {} & 1 \cdot log[P(3) - P(0)] + 3 \cdot log[P(6) - P(3)] + \\ & + 12 \cdot log[P(12) - P(6)] + 1 \cdot log[P(16) - P(12)] + \\ & + 8 \cdot log[1 - P(16)] \end{aligned}$$

By repeating the above calculations, we can see that the log-likelihood is, e.g., higher with parameter values fixed at $t_{50} = 7$, $b = 4$ and $P_{MAX} = 0.95$ (LL = −37.5), than with parameter values fixed at $t_{50} = 3$, $b = 4$ 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) $t_{50} = 7.72$ (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 $t_{50} = 7.63$ (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:

$$P(t, \Psi, T) = P_{MAX}\Phi\{b[log(t) - log(t_a)]\} \quad (4)$$

where t_a is the 50^{th} percentile (t_{50}) if Φ is e.g., log-normal or log-logistic or some other specific percentiles in other cumulative distribution functions (e.g., 63^{th} 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 $t_a = 1/GR_a$, 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., Φ, 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 TT-to-event (TTE) model. Relating to the GR_{50} 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 Φ, 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_{50} 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_{50} for the whole seed lot can also be retrieved from the time-to-event fit (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.

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.

Model	Sub-model	Equation	Eq. No.
HTE	GR_{50}	$f_1(\Psi) = \frac{\Psi - \Psi_b}{\Theta_H}$	(6)
	P_{MAX}	$h_1(\Psi) = G\left[1 - exp\left(\frac{\Psi - \Psi_b}{\sigma_{\Psi_b}}\right)\right]$	(7)
	Distribution	$P_1\left(t, \Psi\right) = \frac{h_1(\Psi)}{1 + exp\{b[log(t) - log(1/f_1(\Psi))]\}}$	(8)
HTTE	GR_{50}	$f_2(\Psi, T) = \frac{T - T_b}{\Theta_{HT}}[\Psi - \Psi_b - k(T - T_b)]$	(9)
	P_{MAX}	$h_2(\Psi, T) = G\left[1 - exp\left(\frac{\Psi - \Psi_b - k(T - T_b)}{\sigma_{\Psi_b}}\right)\right]$	(10)
	Distribution	$P_2\left(t, \Psi, T\right) = \frac{h_2(\Psi, T)}{1 + exp\{b[log(t) - log(1/f_2(\Psi, T))]\}}$	(11)
TTE	GR_{50}	$f_3(T) = \frac{T - T_b}{\Theta_T}\left[1 - \frac{(T - T_b)}{(T_c - T_b)}\right]$	(12)
	P_{MAX}	$h_3(T) = G\left\{1 - exp\left[-\frac{(T_c - T)}{\sigma_{Tc}}\right]\right\}$	(13)
	Distribution	$P_3\left(t, T\right) = \frac{h_3(T)}{1 + exp\{b[log(t) - log(1/f_3(T))]\}}$	(14)

Ψ: water potential (MPa); T: temperature (C); t: time (days or hours); Ψ_b: base water potential (MPa); Θ_H: hydrotime constant (MPa day or MPa hour); G: maximum proportion of germinated seeds; b : slope of germination curve; T_b: base temperature (°C); T_c: ceiling temperature (°C); k, σ_{Ψ_b} 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.

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_d)

$$HTT_d = (T - T_b)[\Psi - \Psi_b - k(T - T_b)] \quad (15)$$

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

$$P(t, \Psi, T) = \frac{h_2(\Psi, T)}{1 + exp\left\{b\left[log\left(\frac{\sum_{i=1}^{n} HTT_d(i)}{\Theta_{HT}}\right)\right]\right\}} \quad (16)$$

When the daily accumulation of HTT reaches the value of Θ_{HT}, 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:

$$GR_g = exp\left[-\left(\frac{1}{b}\right)log\Delta + log\left(\frac{1}{f_2(T, \Psi)}\right)\right] \quad (17)$$

where Δ is equal to $(1 - g)/g$ or $[h_2(T, \Psi) - g]/g$, depending on whether the percentile g refers to the germinated fraction or to the whole population. Predicting the GR_g levels for the whole population is fundamental whenever we intend to compare seed populations with a different P_{MAX}.

Table 3
Parameter estimates and standard errors for the HTE model (Eq. 8) fitted to the rapeseed data.

Parameter	Estimate	Naive SE	Sandwich SE
G	0.958	0.006	0.008
Ψ_b	−1.040	0.005	0.005
$\sigma_{\Psi b}$	0.111	0.009	0.012
θ_H	0.906	0.030	0.041
b	4.027	0.196	0.193

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_g 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_{50} stands for the

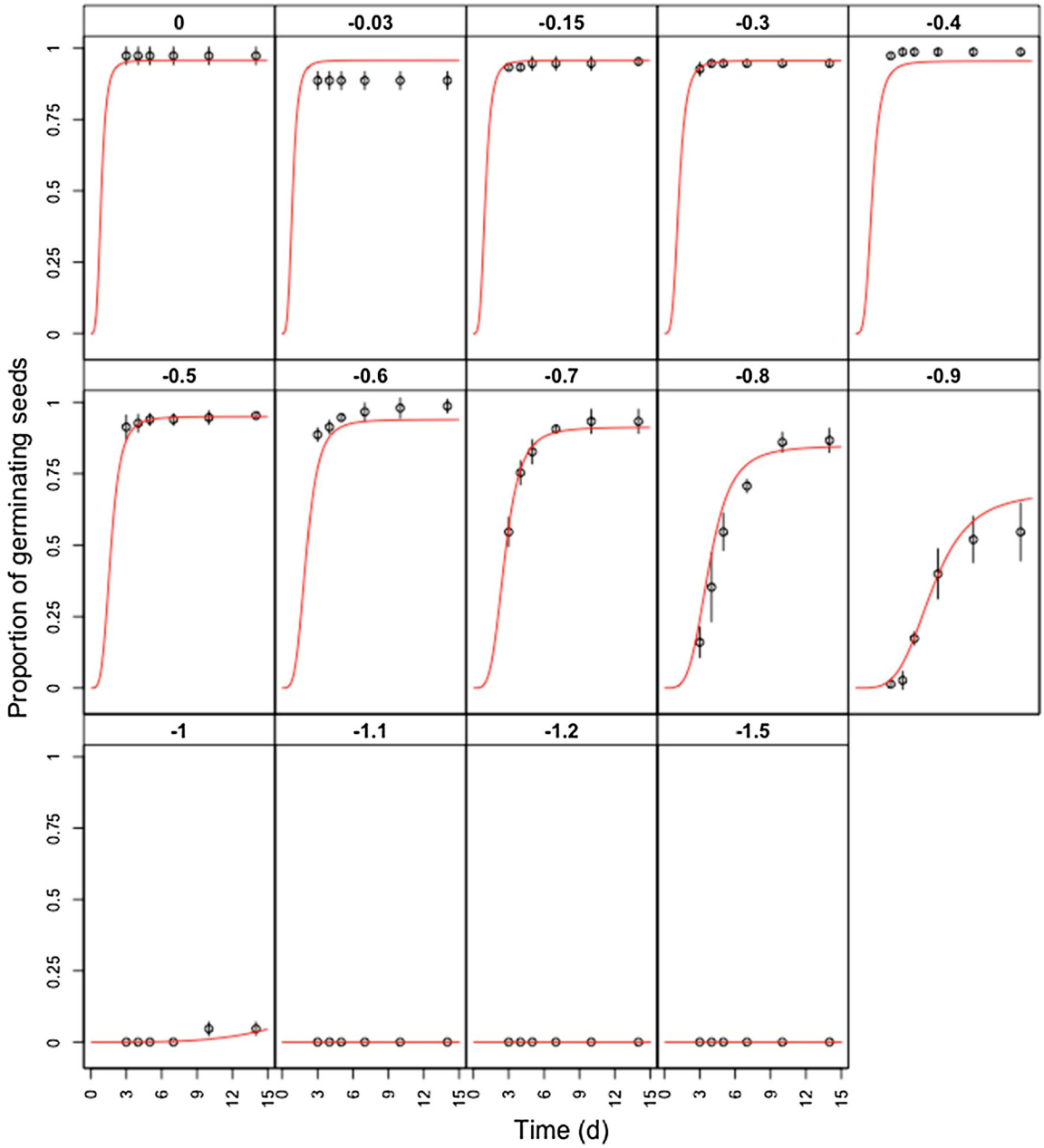

Fig. 1. Effect of water potential on the time-course of germinations in rapeseed (*Brassica napus* var. Excelsior; Example 1). Symbols denote the observed data and lines show the HTE model fits (Eq. (8); model parameters are reported in Table 3) with vertical bars representing the 95% confidence limits.

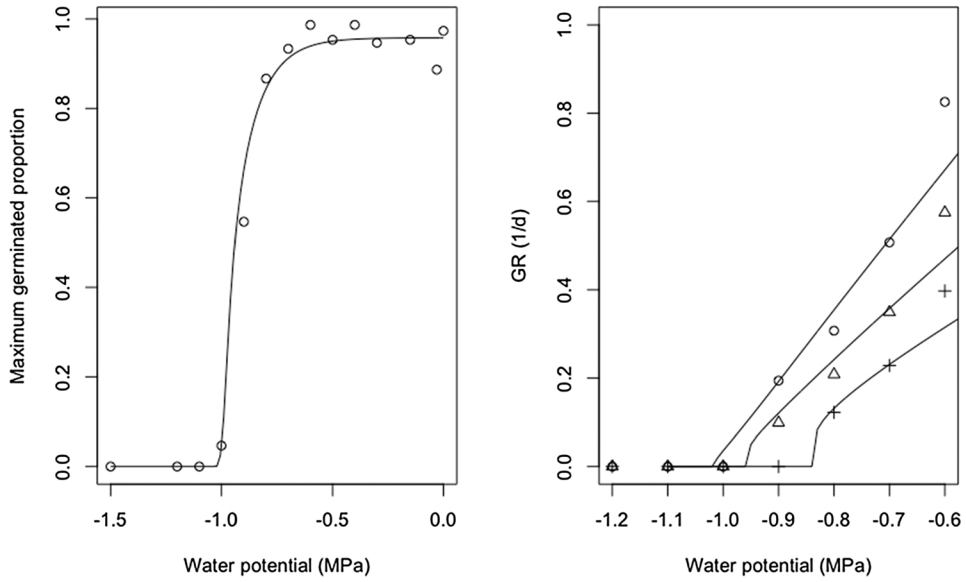

Fig. 2. Effect of water potential on the germination of rapeseed (*Brassica napus* L. var. Excelsior; Example 1). Symbols denote the observed data and lines show the HTE model fit (Eq. (8); model parameters are reported in Table 3). Left: maximum germinated proportion (P_{MAX}); Right: germination rate (GR) for the 20^{th}, 50^{th} and 80^{th} percentiles for the whole population.

Table 4
Parameter estimates and standard errors for the HTTE model (Eq. (11)), fitted to the *H. spontaneum* data.

Parameter	Estimate	SE
G	0.99	0.012
Ψ_b	−2.91	0.035
k	0.07	0.001
T_b	−0.75	0.353
σ_{Ψ_b}	0.55	0.029
Θ_{HT}	1309.06	40.638
b	4.17	0.113

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_g values across populations with a different P_{MAX}. However, the function "ED()" can be used in "drc" to calculate GR_g values both for the germinated fraction and for the whole population. Examples are given in the Appendix C.

If we consider the GR_g values for the whole population (Fig. 2 right), the relationship between water potential and GRs 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 log-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}, $GR_{50} = 1/t_{50}$ 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 $T = T_d > T_b$, where T_d is a temperature threshold close to, but not equal to the optimal temperature, T_o:

$$\mathrm{GR}_{50} = f_4(T) = \begin{cases} \frac{T - T_b}{\Theta_T} & \text{if } T_b < T < T_d \\ \frac{T - T_b}{\Theta_T}\left[1 - \frac{T - T_d}{T_c - T_d}\right] & \text{if } T_d < T < T_c \\ 0 & \text{if } T \le T_b \text{ or } T \ge T_c \end{cases} \quad (18)$$

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:

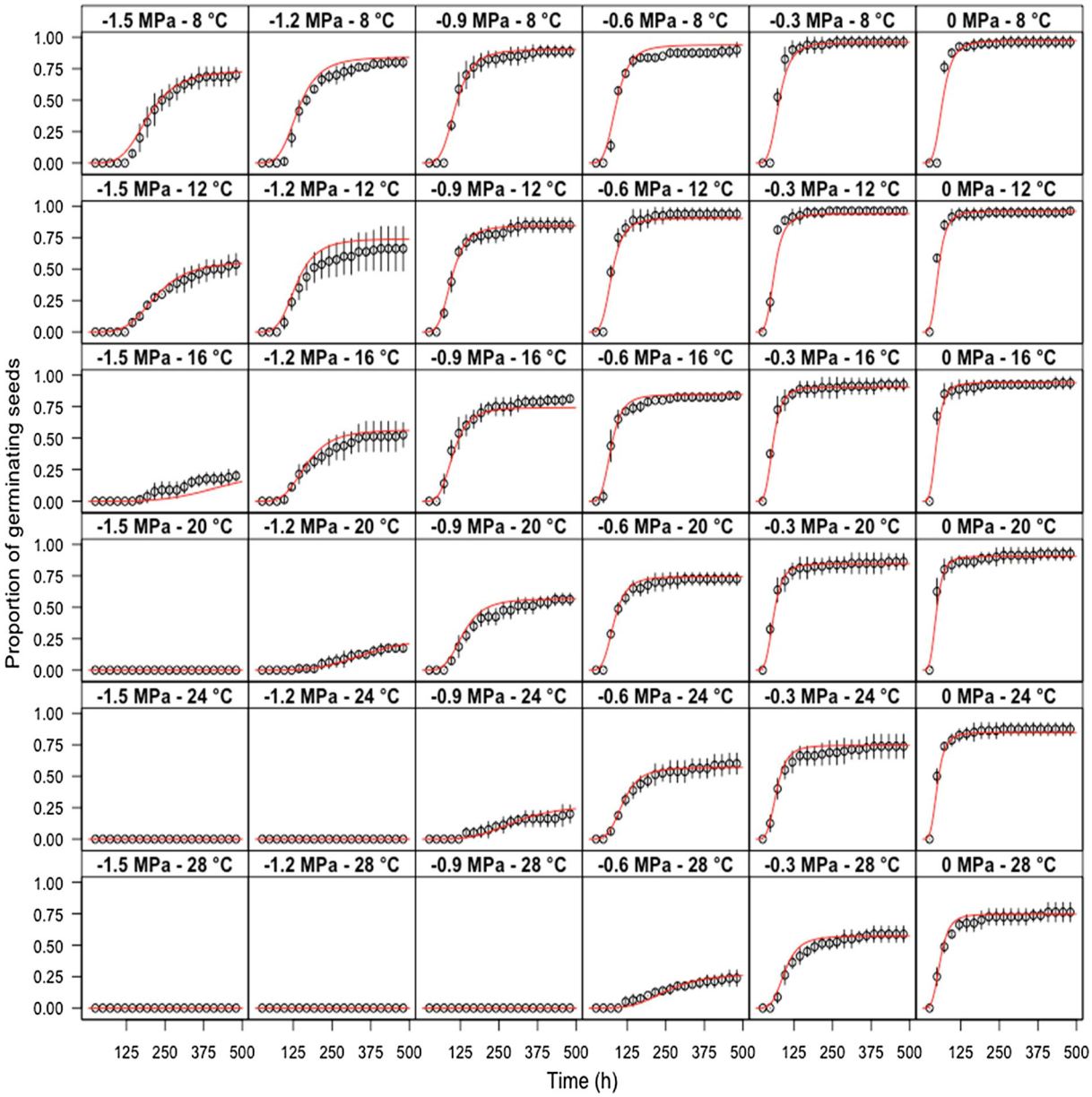

Fig. 3. Effect of water potential and temperature on the time-course of germinations in *Hordeum spontaneum* [C. Koch] Thell. (Example 2). Symbols denote the observed data and lines show the HTTE model fits (Eq. (11); model parameters are reported in Table 4), with vertical bars representing the 95% confidence limits.

$$P(t) = \frac{h_3(T)}{1 + exp\{b[log(t) - log(1/f_4(T))]\}} \quad (19)$$

The above model is defined for temperature levels between T_b and T_c, while it is 0 for temperatures $T \leq T_b$ or $T \geq T_c$. 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_{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

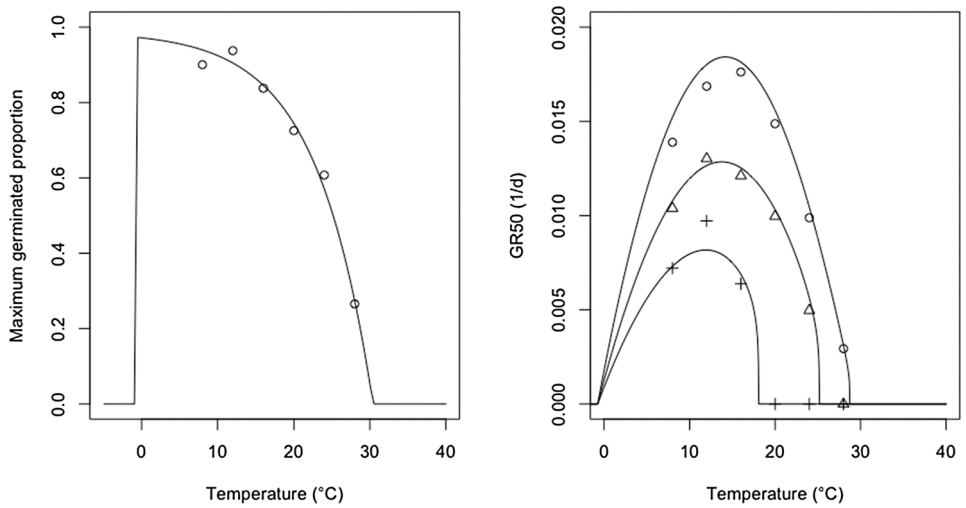

Fig. 4. Effect of temperature on the germination characteristics of *Hordeum spontaneum* [C. Koch] Thell. (Example 2). Symbols denote the observed data and lines show the HTTE model fit (Eq. (11); model parameters are reported in Table 4). Left: maximum germinated proportion (P_{MAX}); Right: germination rate (GR) for the 20^{th}, 50^{th} and 80^{th} 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.

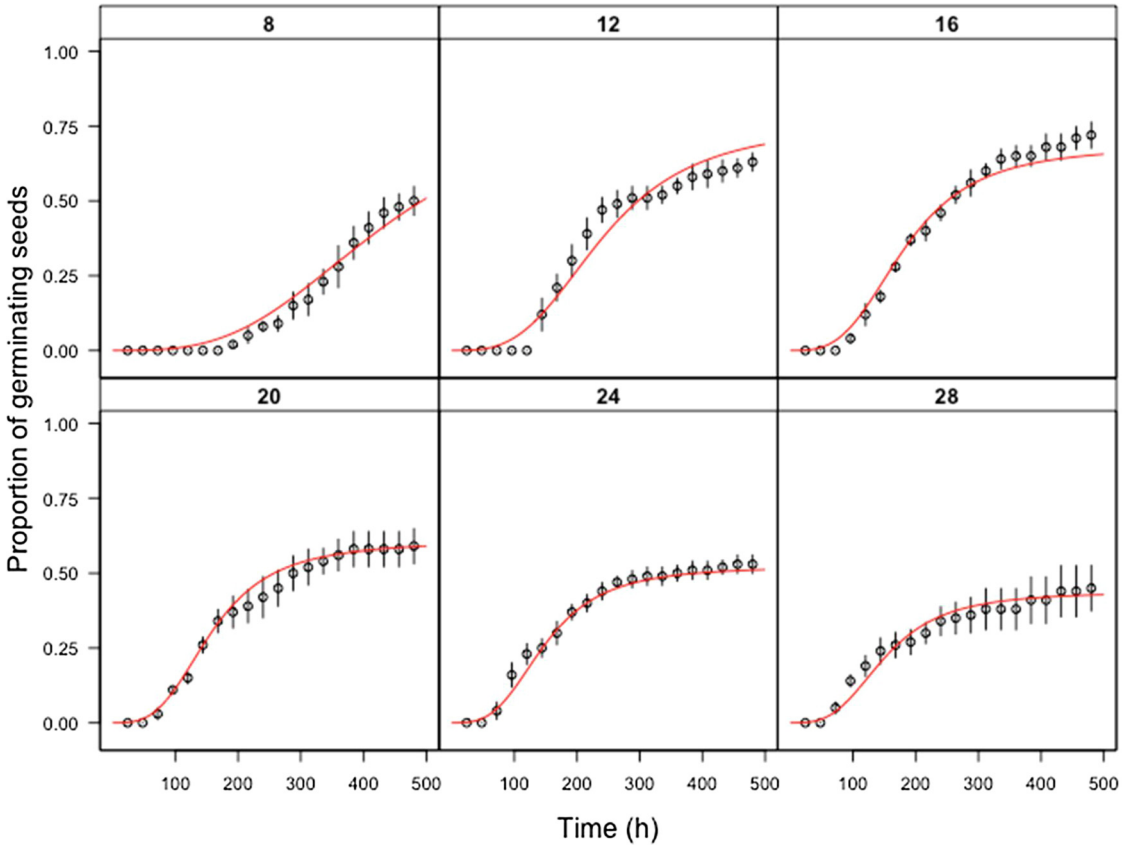

Fig. 5. Effect of temperature on the time-course of germinations in *Phalaris minor* (Example 3). Symbols denote the observed data and lines show the TTE model fits (Eq. (14); parameter estimates are reported in Table 5), with vertical bars representing 95% confidence limits.

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

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

Parameter	Estimate	SE
G	1.00	0.081
T_c	48.80	3.933
σ_{Tc}	33.37	5.557
T_b	3.35	0.525
Θ_T	1655.56	128.040
b	3.16	0.081

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

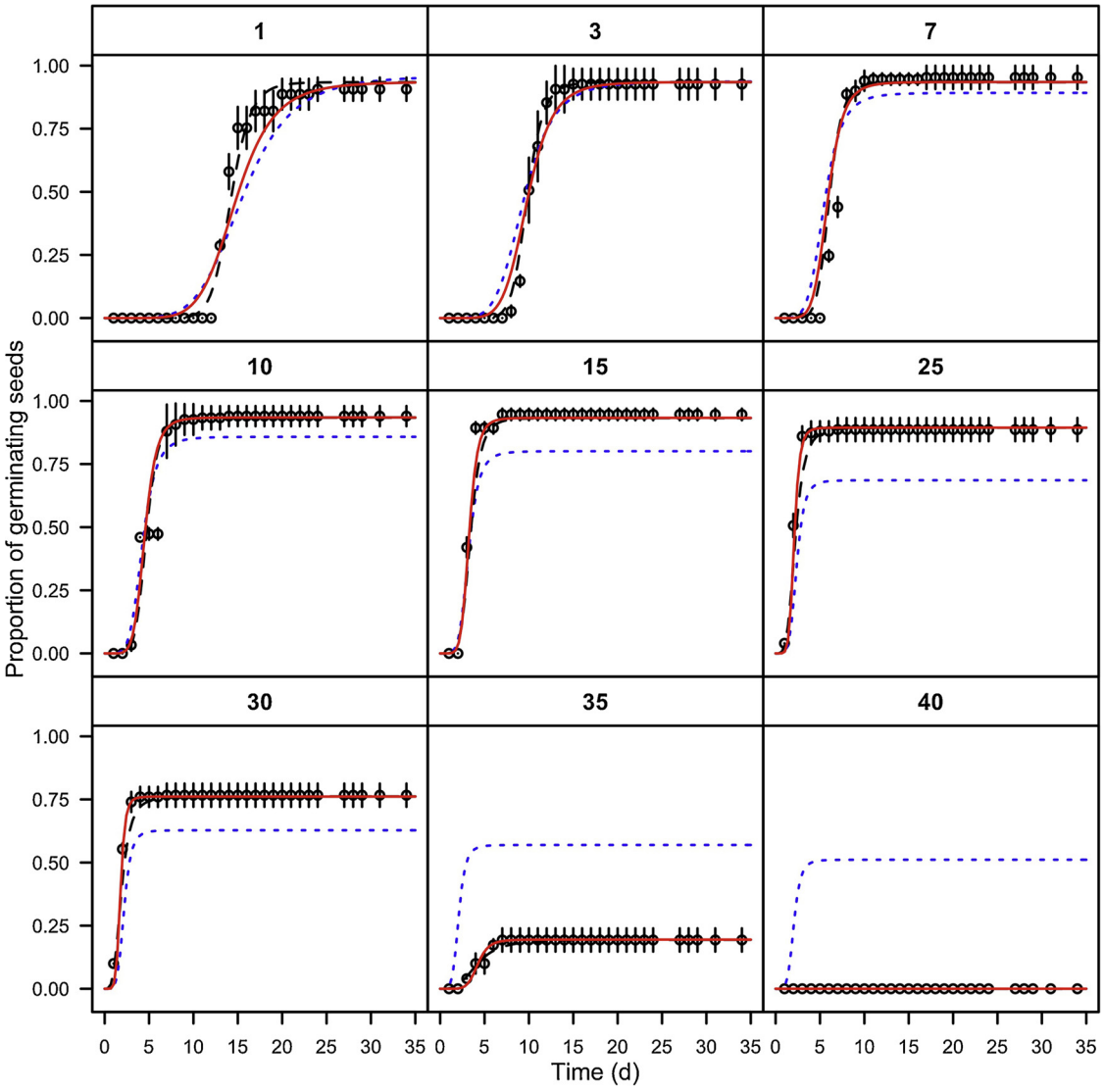

Fig. 6. Effect of temperature on the time-course of germinations in barley (Example 4). Symbols denote the observed data; the solid, red lines show the TTE model fits (Eq. 19; Rowse and Finch-Savage, 2003; parameter estimates are reported in Table 6). The dotted, blue lines show the fit from a simpler 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).

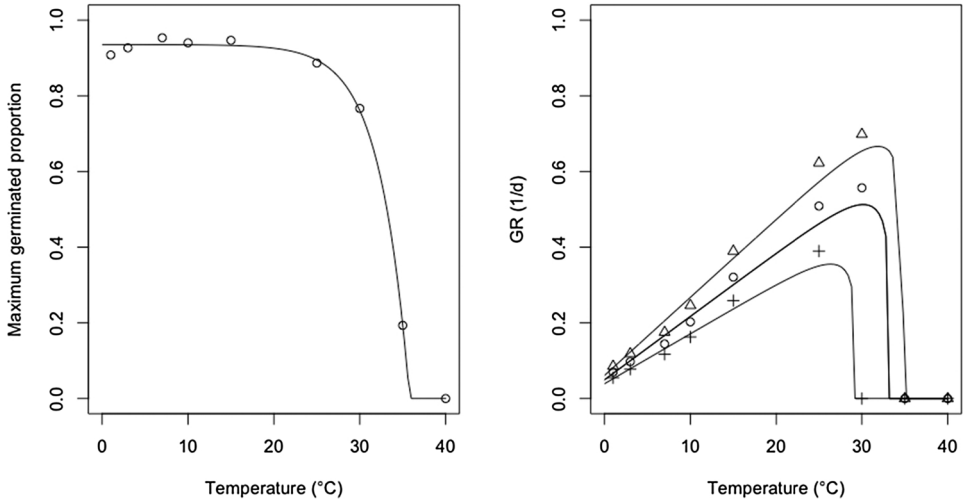

Fig. 7. Effect of temperature on the germination characteristics of barley (Example 4). Symbols denote the observed data and lines show the TTE fit (Eq. 19; parameter estimates are shown in Table 6). Left: maximum germinated proportion (P_{MAX}); Right: germination rate (GR) for the 20^{th}, 50^{th} and 80^{th} percentiles for the whole population.

Table 6
Parameter estimates and standard errors for the TTE model (Eq. (19)), fitted to the barley data.

Parameter	Estimate	SE
G	0.94	0.007
T_c	35.81	0.071
σ_{Tc}	3.46	0.186
T_d	33.59	0.221
T_b	−2.96	0.209
Θ_T	58.52	2.128
b	6.89	0.680

estimates of parameters and standard errors. This is not necessarily true for nonlinear regression models and proposed improvements, such as the "deacumulation" of counts (Mesgaran et al., 2013) or the use of resampling methods (i.e., the jackknife or bootstrap; Onofri et al., 2014), are not immediately available in statistical packages and may require some tedious coding, which is beyond the interests of most biologists. In our opinion, this is a strong argument in favour of time-to-event models: reliable standard errors are a fundamental requirement of inference and hypothesis testing, which are the basis of scientific progress.

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_{50}, 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

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat.

Contr. 19, 716–723.
Alvarado, V., Bradford, K., 2002. A hydrothermal time model explains the cardinal temperatures for seed germination. Plant Cell Environ. 25, 1061–1069.
Bello, P., Bradford, K.J., 2016. Single-seed oxygen consumption measurements and population-based threshold models link respiration and germination rates under diverse conditions. Seed Sci. Res. 26, 199–221.
Berger, S., Graham, N., Zeileis, A., 2017. Various versatile variances: an object-oriented implementation of clustered covariances. R. Working Paper 2017-12, Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics. Universität Innsbruck.
Bradford, K., 2002. Applications of hydrothermal time to quantifying and modeling seed germination and dormancy. Weed Sci. 50, 248–260.
Carroll, R.J., Wang, S., Simpson, D.G., Stromberg, A.J., Ruppert, D., 1998. The sandwich (robust covariance matrix) estimator. Technical Report, Department of Statistics. A& M University, Texas.
Catara, S., Cristaudo, A., Gualtieri, A., Galesi, R., Impelluso, C., Onofri, A., 2016. Threshold temperatures for seed germination in nine species of *Verbascum* (*Scrophulariaceae*). Seed Sci. Res. 26, 30–46.
Finch-Savage, W.E., Steckel, J.R.A., Phelps, K., 1998. Germination and post-germination growth to carrot seedling emergence: predictive threshold models and sources of variation between sowing occasions. New Phytol. 139, 505–516.
Gonzalez-Andujar, J.L., Francisco-Fernandez, M., Cao, R., Reyes, M., Urbano, J.M., Forcella, F., Bastida, F., 2016. A comparative study between nonlinear regression and nonparametric approaches for modelling *Phalaris paradoxa* seedling emergence. Weed Res. 56, 367–376.
Grundy, A., Phelps, K., Reader, R., Burston, S., 2000. Modelling the germination of *Stellaria media* using the concept of hydrothermal time. New Phytol. 148, 433–444.
Hardegree, S.P., Winstral, A.H., 2006. Predicting germination response to temperature. II. Three-dimensional regression, statistical gridding and iterative-probit optimization using measured and interpolated-subpopulation data. Ann. Bot. 98, 403–410.
Hardegree, S.P., Walters, C.T., Boehm, A.R., Olsoy, P.J., Clark, P.E., Pierson, F.B., 2015. Hydrothermal Germination Models: comparison of two data-fitting approaches with probit optimization. Crop Sci. 55, 2276.
Hay, F.R., Mead, A., Bloomberg, M., 2014. Modelling seed germination in response to continuous variables: use and limitations of probit analysis and alternative approaches. Seed Sci. Res. 24, 165–186.
Jensen, S.M., Andreasen, C., Streibig, J.C., Keshtkar, E., Ritz, C., 2017. A note on the analysis of germination data from complex experimental designs. Seed Sci. Res. 27, 321–327.
Masin, R., Onofri, A., Gasparini, V., Zanin, G., 2017. Can alternating temperatures be used to estimate base temperature for seed germination? Weed Res. 57, 390–398.
McNair, J.N., Sunkara, A., Frobish, D., 2012. How to analyse seed germination data using statistical time-to-event analysis: non-parametric and semi-parametric methods. Seed Sci. Res. 22, 77–95.
Mesgaran, M.B., Mashhadi, H.R., Alizadeh, H., Hunt, J., Young, K.R., Cousens, R.D., 2013. Importance of distribution function selection for hydrothermal time models of seed germination. Weed Res. 53, 89–101.
Mesgaran, M.B., Onofri, A., Mashhadi, H.R., Cousens, R.D., 2017. Water availability shifts the optimal temperatures for seed germination: a modelling approach. Ecol. Modell. 351, 87–95.
Onofri, A., Gresta, F., Tei, F., 2010. A new method for the analysis of germination and emergence data of weed species. Weed Res. 50, 187–198.
Onofri, A., Mesgaran, M.B., Tei, F., Cousens, R.D., 2011. The cure model: an improved way to describe seed germination? Weed Res. 51, 516–524.
Onofri, A., Mesgaran, M., Neve, P., Cousens, R., 2014. Experimental design and parameter estimation for threshold models in seed germination. Weed Res. 54, 425–435.
Pace, R., Benincasa, P., 2010. Effect of salinity and low osmotic potential on the germination and seedling growth of rapeseed cultivars with different stress tolerance. Ital. J. Agron. 5, 69–77.
Pipper, C.B., Adolf, V.I., Jacobsen, S.-E., 2013. A model for quantification of temperature profiles via germination times. J. Agric. Biol. Environ. Stat. 18, 87–101.
R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Ritz, C., Pipper, C., Yndgaard, F., Fredlund, K., Steinrucken, G., 2010. Modelling flowering of plants using time-to-event methods. Eur. J. Agron. 32, 155–161.
Ritz, C., Pipper, C.B., Streibig, J.C., 2013. Analysis of germination data from agricultural experiments. Eur. J. Agron. 45, 1–6.
Ritz, C., Baty, F., Streibig, J.C., Gerhard, D., 2015. Dose-response analysis using R. PLoS One 10.
Rowse, H., Finch-Savage, W., 2003. Hydrothermal threshold models can describe the germination response of carrot (*Daucus carota*) and onion (*Allium cepa*) seed populations across both sub- and supra-optimal temperatures. New Phytol. 158, 101–108.
Therneau, T., 2012. A Package for Survival Analysis in S. R Package Version 2.36-14. URL: https://CRAN.R-project.org/package=survival (Date of Last access: 12/09/18).
Watt, M.S., Xub, V., Bloombergb, M., 2010. Development of a hydrothermal time seed germination model which uses the weibull distribution to describe base water potential. Ecol. Modell. 221, 1267–1272.
Yu, B., Peng, Y., 2008. Mixture cure models for multivariate survival data. Comput. Stat. Data Anal. 52, 1524–1532.
Zeileis, A., Hothorn, T., 2002. Diagnostic checking in regression relationships. R News 2, 7–10.