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

A R T I C L E   I N F O

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 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

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 gth 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):

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

where Ψ is the water potential in the substrate, $\Psi_{b(g)}$ is the base osmotic potential for the gth 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 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_{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:

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

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.