WEED RESEARCH An International Journal of Weed Biology, Ecology and Vegetation Management

DOI: 10.1111/wre.12095

METHODS

Experimental design and parameter estimation for threshold models in seed germination

A ONOFRI*, M B MESGARAN†, P NEVE‡§ & R D COUSENS†

*Dipartimento di Scienze Agrarie Alimentari ed Ambientali, Università degli Studi di Perugia, Perugia, Italy, †Department of Resource Management & Geography, The University of Melbourne, Victoria, Australia, and ‡School of Life Sciences, University of Warwick, Coventry, UK

Received 6 March 2014
Revised version accepted 22 April 2014
Subject Editor: Lammert Bastiaans, WUR, the Netherlands

Summary

Hydrotime threshold models are used to describe the dynamics of seed germination in response to reduced water availability. Although these models provide several biologically relevant parameters, it is unclear which statistical technique is best suited to their estimation. Most commonly, these models are fitted to the observed cumulative proportions of germinated seeds, using nonlinear regression. However, this approach has been questioned, due to its inability to account for some characteristics of data sets obtained from germination assays, such as interval censoring and correlated observations. We used Monte Carlo simulations to determine the bias and precision of nonlinear regression estimators for a wide range of experimental designs and hypothetical plant species. Results showed that point estimates of model parameters were almost unbiased, while standard errors obtained from nonlinear regression were on average 3–4 times smaller than the Monte Carlo precision. Standard errors obtained by nonparametric resampling methods were comparable to Monte Carlo precision and provided good coverage (very close to the nominal 95% value), with at least 4–8 treatments by four replicates and 50 seeds per Petri dish. With 10 seeds per Petri dish, a higher number of replicates were necessary to achieve good coverage. In particular, good results were obtained with the grouped jackknife (delete-a-Petri-dish), which accounts for repeated observations on the same Petri dish. It is suggested that nonlinear regression may be used to fit the hydrotime model, in association with resampling methods, particularly when the purpose is to compare 'hydrotime' parameters across treatments or plant species.

Keywords: seed germination, hydrotime model, Monte Carlo simulation, bootstrap, jackknife.

ONOFRI A, MESGARAN MB, NEVE P & COUSENS RD (2014). Experimental design and parameter estimation for threshold models in seed germination. *Weed Research* **54**, 425–435.

Introduction

Hydrotime threshold models have been used to describe the response of a seed population to reduced water availability, in terms of progress towards germination. In their most widespread form, these models assume that the germination rate (GR, day^{-1} or h^{-1}) for the i-th seed of a population is linearly related to the environmental water potential (ψ), above a certain threshold (base) level (Bradford, 2002):

§Present address: Rothamsted Research, Department of Agroecology, West Common, Harpenden, Hertfordshire, UK.
Correspondence: Dr. A Onofri, Dipartimento di Scienze Agrarie, Alimentari e Ambientali, Università degli Studi di Perugia, Borgo XX Giugno 74, 06121, Perugia, Italy. Tel: (+39) 075 5856324; Fax: (+39) 075 5856344; E-mail: andrea.onofri@unipg.it

$$\mathrm{GR}_i(\psi) = \frac{1}{t_i(\psi)} = \frac{\psi - \psi_{\mathrm{b}(50)}}{\theta_{\mathrm{H}}} + \varepsilon_i \qquad (1)$$

where $\psi_{\mathrm{b}(50)}$ is the mean base water potential at the population level, θ_{H} is the hydrotime constant (in MPa day or MPa hour), and ε_{i} is the error component, representing the seed-to-seed variability within the seed population, which is assumed to follow a normal distribution, with a mean of 0 and standard deviation σ_{GR}: the higher the standard deviation, the greater the variation in germination timing.

Eqn (1) is presented in the usual linear form, and it is equivalent to the original formulation of Bradford (2002), who related the seed-to-seed variability in germination rates to the variability in base water potential within the seed population. If base water potential is normally distributed with mean equal to $\psi_{\mathrm{b}(50)}$ and standard deviation equal to $\sigma_{\psi\mathrm{b}}$, germination rates are also normally distributed with mean GR(50) and standard deviation $\sigma_{\mathrm{GR}} = \sigma_{\psi\mathrm{b}}/\theta_{\mathrm{H}}$. It is important to note that, in practice, there is no way to separate the variability in germination rates induced by the stochastic distribution of base water potential from that attributable to other potential sources of experimental variability (such as seed size, inaccuracies in manipulations, environmental variability within the incubator).

Even though Eqn (1) is fundamentally linear, a survey of literature shows that it is rarely fit to data in such a form. Indeed, individual germination rates are truncated at 0, as negative values are not possible. Furthermore, individual germination rates are not directly observed during germination assays, and they are often derived by interpolation from germination curves (see for example Finch-Savage *et al.*, 1998). Ritz *et al.* (2013) noted that such a 'two-step' fitting procedure is inefficient, because the uncertainty in germination rates is neglected in the second step.

Eqn (1) is therefore reparameterised as:

$$G(t,\psi) = \phi\left\{\frac{\psi - \frac{\theta_{\mathrm{H}}}{t} - \psi_{\mathrm{b}(50)}}{\sigma_{\psi_{\mathrm{b}}}}\right\} \qquad (2)$$

where G is the proportion of germinated seeds at time t, and ϕ is the cumulative normal distribution.

Note that this model explicitly includes the standard deviation of ψ_{b} within the population ($\sigma_{\psi\mathrm{b}}$). The basic hydrotime model given by Eqn (2) can be modified to accommodate the effects of both suboptimal and super-optimal temperatures (Alvarado & Bradford, 2002; Rowse & Finch-Savage, 2003; Watt & Bloomberg, 2012) and has also been extended to account for dormancy (Finch-Savage & Leubner-Metzger, 2006; Chantre *et al.*, 2009; Wang *et al.*, 2009). Furthermore, Watt *et al.* (2010) and Mesgaran *et al.* (2013) have shown that other distributions can be used to replace ϕ, to more flexibly describe the asymmetric germination pattern often observed. Several reviews have concluded that Eqn (2) and other models of the same type represent a sound modelling platform, providing good insight into the germination process (Allen, 2003; Allen *et al.*, 2007).

Estimation of θ_{H}, $\psi_{\mathrm{b}(50)}$ and $\sigma_{\psi\mathrm{b}}$ (as well as other biologically meaningful parameters, such as the cardinal temperatures within a hydrothermal modelling context) is thus of benefit in comparing the dynamics of seed germination between species and between populations of the same species. To obtain reliable and accurate estimates with efficient experiments, two issues become critical: the use of an appropriate statistical method and the use of an adequate experimental design. Both the statistical method and the experimental design should be evaluated in terms of their capability to provide reliable measures of uncertainty (standard errors). Reporting standard errors is commonly considered as a basic requirement for published papers (see for example Onofri *et al.*, 2010a), even though we noted that they are missing from most of the above-mentioned papers dealing with parameter estimation for the hydrotime/hydrothermal model.

The most straightforward method for data analysis is to fit Eqn (2) to the observed cumulative proportion of germinated seeds, using nonlinear regression techniques or similar methods (Finch-Savage *et al.*, 1998; Alvarado & Bradford, 2002; Rowse & Finch-Savage, 2003; Hardegree, 2006; Pace & Benincasa, 2010; Watt *et al.*, 2010). Biologists are generally familiar with nonlinear least squares regression, even though it has been highlighted that the use of such a method in germination assays is incorrect, and other techniques should be preferred (see for example Hunter *et al.*, 1984; O'Neill *et al.*, 2004; Onofri *et al.*, 2010b, 2011; McNair *et al.*, 2012; Ritz *et al.*, 2013 and references therein). In particular, it has been highlighted that nonlinear least squares regression does not account for non-normal, heteroscedastic responses (counts), repeated observations on the same experimental units over time (lack of independence) and the fact that the exact germination moment is never directly observed, as this event takes place between two successive assessment times (interval censoring).

Unless biologists abandon the hydrotime model and other types of threshold models or change totally modelling platform (as more or less explicitly suggested in the above-mentioned papers), it is fundamental to test whether and how nonlinear least squares regression can be reliably used to fit models based on the proportion of germinated seeds. In this respect, Ritz *et al.*

(2013) showed that the main problem with nonlinear regression in germination assays is that the standard errors of model parameters are strongly underestimated. One possible solution may be to amalgamate nonlinear least squares regression with a more robust (less parametric) approach to estimate standard errors. Resampling methods (in particular the jackknife and bootstrap) could be of benefit in the context of seed germination assays (Shafii & Price, 2001), but this remains to be demonstrated.

Various aspects of assay design will contribute to the accuracy and precision of estimators. One of the most influential is sample size, composed of the number of treatment levels (in this case the water potentials), the number of replicates and the number of seeds per Petri dish. As a general principle, larger sample sizes should result in more accurate parameter estimates. Another important issue relates to the selection of a monitoring programme: infrequent monitoring or the premature cessation of monitoring is likely to decrease the accuracy and precision of estimators. In this respect, a species that germinates rapidly will allow fewer counts to be made before germination is complete, while with a species that germinates slowly, we might stop the assay before germination is complete.

Even though some consideration has been given to the recording schedule (see Mesgaran *et al.*, 2013), the influence of assay design on the estimation of germination parameters has not so far been explored formally.

The objective of this study was to investigate the combined effects of fitting method and assay design on parameter estimates. We used Monte Carlo simulation to answer the following questions: (i) Are point estimates, variances and interval estimates for θ_H, $\psi_{b(50)}$ and $\sigma_{\psi b}$ reliable when they are obtained using nonlinear least squares regression based on Eqn (2)? (ii) Are resampling methods (jackknife and bootstrap) more appropriate alternatives to derive standard errors for model parameters? (iii) To what extent do the potential answers depend on the experimental design, plant species and monitoring schedule?

Materials and methods

Monte Carlo simulations

The rationale behind using simulated data sets instead of real data is that with Monte Carlo simulation, we can start from some parameter values, incorporate random variability (noise) and generate experimental observations. If we repeat the simulations a number of times and fit the same model to each generated data set, we are able to obtain a clear picture about the sampling distribution of estimates, that is, their interassay variability.

In our case, the procedure was as follows:

1 We fixed $\theta_H = 81.268$ MPa °C h, $\psi_{b(50)} = -0.434$ MPa and $\sigma_{\psi b} = 0.365$ Mpa. These values are realistic, because they were taken from a real data set, collected from a germination assay carried out at the School of Life Sciences, University of Warwick (Wellesbourne, UK), with *Tripleurospermum inodorum* (Neve P, unpublished data; see Fig. 1). These parameters were assumed to represent the seed population of interest and were used as the starting point for the subsequent simulation work.

2 We planned a Monte Carlo experiment at constant temperature, with eight water potential levels (0, −0.2, −0.4, −0.6, −0.8, −1, −1.2 and −1.4 MPa), six replicates of 50 seeds per Petri dish and 21 inspection times (53, 74, 96, 122, 147, 172, 191, 241, 287, 314, 363, 408, 482, 531, 625, 700, 794, 900, 1054, 1221, 1370 h from the beginning of the assay). This design closely resembles the real germination assay mentioned at step 1 and may be regarded as an optimal design for this type of assays.

3 We used Eqn (1) and the above-mentioned values for θ_H and $\psi_{b(50)}$ to generate the average germination rate (GR_{50}) for each dish, according to the environmental water potential level. The use of Eqn (1) is motivated by the fact that this model is the 'core' of population-based threshold models, which are all built on the very same biological background, that is, the interseed variability in base water potential (Bradford, 2002);

4 We fixed $\sigma_{\psi b}$ at its selected value (see above) and used the random number generator function rnorm() in R (R Core Team, 2013) to add seed-to-seed variability within each Petri dish, using a normal distribution with mean 0 and standard deviation $\sigma_{GR} = \sigma_{\psi b}/\theta_H$. According to the hydrotime model theory (Bradford, 2002), we truncated this distribution at 0, to represent those seeds with a base water potential value lower than the environmental water potential, which are not able to germinate;

5 The real (but unobservable in real experiments) time of germination (t) of each seed was calculated as the reciprocal of germination rate (1/GR). These real germination times were transformed into the number of seeds germinated within each monitoring interval, according to the above monitoring schedule. Seeds whose germination times were longer than the assessment duration were considered as ungerminated (right censoring);

6 As is common practice in germination assays, counts were transformed into cumulative proportions

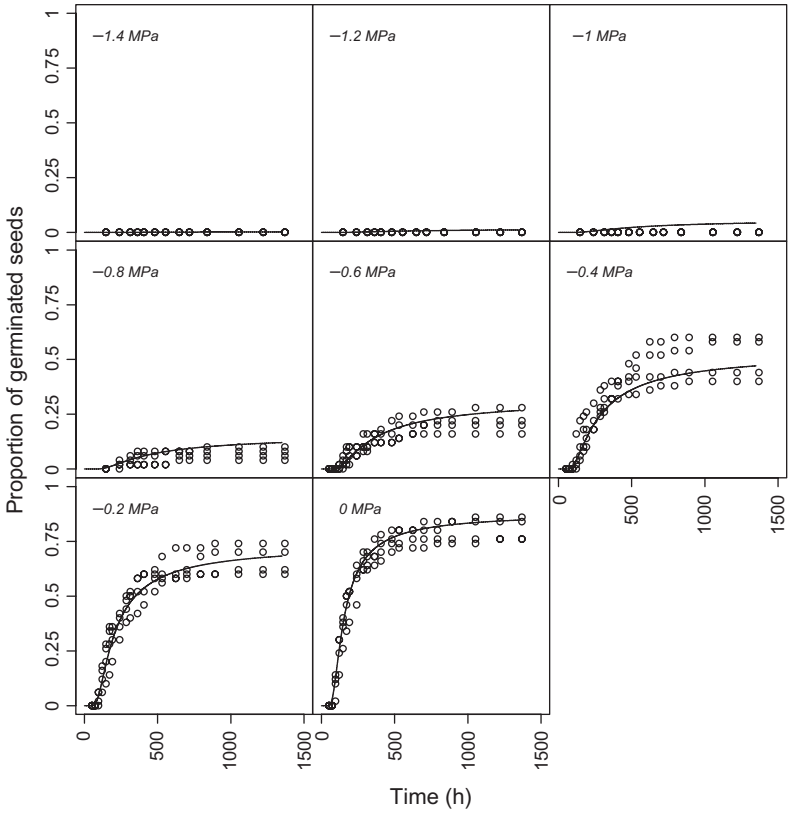

Fig. 1 Germination assay with *Tripleurospermum inodorum*, at 18°C, eight water potential levels (0, −0.2, −0.4, −0.6, −0.8, −1, −1.2 and −1.4 MPa), four replicates, 50 seeds per Petri dish (Paul Neve, unpublished data). Solid lines represent the fitted models (Eqn 2).

of germinated seeds at each assessment time, thus reproducing the potential error due to interval censoring;

7 Apart from the above sources of noise (seed-to-seed variability and censoring), we also ran simulations attaching additional random variability at the Petri dish level, by assuming that the nominal value of water potential in each dish was subject to experimental error. This aspect is rather crucial, because the presence of errors in the explanatory variables represents a frequently overlooked problem, which would itself prevent the use of several types of regression models. Based on experience, such random variability was simulated using a normal distribution with mean 0 and standard deviation 0.02. In all cases, the normal distribution was truncated to avoid positive values for water potential.

The above procedure generates the data that might be expected from a real experiment, while the bounce is that we know the real underlying parameter values (i.e. parameters set in step 1).

Each simulated data set was analysed with several procedures (estimators: see below) to obtain estimates of θ_H, $\psi_{b(50)}$ and $\sigma_{\psi b}$. Obviously, these estimates were different from the 'real' values, because of the experimental noises. However, simulations were repeated a high number of times (i.e. 1000) using the same set of parameters and the same model (thus only the random variability changed). Therefore, we ended up with 1000 data sets and, consequently, 1000 point estimates for each parameter, accompanied by their respective standard errors and 95% confidence intervals. From this information, we could derive:

1 The average of the 1000 estimates. This should be close to the 'true' value used to generate the samples, and the amount of the difference may be used to assess the bias of estimators (as a percentage of the 'true' value);
2 The standard deviation of the 1000 estimates (Monte Carlo precision), which may be regarded as a reasonable estimate of the true standard error for each estimated parameter (assay-to-assay variability of parameter estimates: a lower value indicates a more precise experiment);
3 The average of the 1000 standard errors for each parameter (obtained from refitting the same selected model to each simulated data sets), which may be compared to the Monte Carlo precision, to check the bias in the estimated standard errors;
4 The proportion of the 1000 confidence intervals that actually captured the 'true' parameter value. This proportion represents the observed 'coverage' which should be as close as possible to the nominal 95% value.

Estimation of parameters and standard errors

For each generated data set, we used four estimation methods. First of all, model parameters and standard errors were obtained using standard nonlinear regression (Bates & Watts, 1988) and fitting the Eqn (2) to the cumulative proportion of germinated seeds. In addition to this reference method, we used two resampling approaches to estimate standard errors, that is, the jackknife and bootstrap. We also used an alternative hydrotime model that makes use of 'non-cumulative' germination counts.

In more detail, fully iterated 'delete-a-group' jackknife estimates of standard errors (Wu, 1986) were obtained as follows:

1 The data set was resampled by removing one Petri dish at a time, which gave as many resamples as the number of Petri dishes;
2 Nonlinear least squares regression was performed on each resample (jackknife values);
3 Jackknife standard errors for each parameter were obtained using the following equation (Yu & Peng, 2008):

$$se_{\text{jack}} = \sqrt{\left[\frac{n-p}{n}\sum_{i=1}^{n}(\theta_{-1,i}-\bar{\theta})^2\right]} \qquad (3)$$

where n is the number of Petri dishes, p is the number of estimated parameters, $\theta_{1,i}$ is the i-th jackknife value (with i ranging from 1 to n) for the specified parameter, and $\overline{\theta}$ is the whole-sample least squares estimate for that parameter.

It has been shown (Wu, 1986) that jackknife estimates of standard errors are robust against non-normality and heteroscedasticity and, in their grouped form, they can account for repeated observations on the same experimental unit over time.

Bootstrap standard errors are similar to jackknife standard errors, with the only difference that the jackknife resamples are created by removing one Petri dish at a time, whereas bootstrap resamples are created by randomly drawing n Petri dishes with replacement. For each simulated experiment, 1000 bootstrap resamples were drawn, and, for each resample, hydrotime model's parameters were estimated by nonlinear least squares regression. Bootstrap estimates of standard errors were obtained as the standard deviations of the above values (Davison & Hinkley, 1997). Similarly to the jackknife, the nonparametric bootstrap in its grouped form only assumes independence among groups (Petri dishes), which should be ensured by randomisation theory (Davison & Hinkley, 1997).

Finally, non-cumulative counts C (i.e. the number of seeds germinated between two successive monitoring times t_1 and t_2) were used to parameterise (also using nonlinear least squares) the following hydrotime model:

$$C(t_1, t_2, \psi) = [G(t_2, \psi) - G(t_1, \psi)] \times N_{\text{tot}} \qquad (4)$$

where $G(t_1, \psi)$ and $G(t_2, \psi)$ are the proportions of seeds germinated at times t_1 and t_2 (Eqn 2), and N_{tot} is the total number of seeds at the beginning of the assay. The use of 'non-cumulative' data has been suggested as a way to alleviate problems related to correlated and non-homogenous errors (Mandel, 1957) in seed germination assays (Mesgaran *et al.*, 2013).

It should be noted that, in principle, all the above calculations are feasible using a nonlinear least squares solver, which is something that many biologists are familiar with. In this work, we only made use of the nls() function within the statistical environment of R (R Core Team, 2013).

Effect of experimental design

To explore the role of experimental design on the bias and precision of estimators, several other Monte Carlo simulations were performed, by considering smaller sample sizes as compared with that of the optimal design described above. In detail, we considered: (i) six, four and two replicates, (ii) eight and four water potential levels and (iii) 50 and 10 seeds per Petri dish. Using a factorial combination of the above elements, a total of 12 experimental designs (each with 1000 simulation runs) were evaluated. For all simulations, the θ_H, $\psi_{b(50)}$ and $\sigma_{\psi b}$ were set, respectively, at 81.268 MPa °C h, −0.434 MPa and 0.365 MPa (see above), and inspections were made at 53, 74, 96, 122, 147, 172, 191, 241, 287, 314, 363, 408, 482, 531, 625, 700, 794, 900, 1054, 1221 and 1370 h from the beginning of the assay.

Effect of plant species and monitoring schedule

In a separate Monte Carlo experiment, we investigated the combined effects of plant species and monitoring schedule on the accuracy and precision of parameter estimates. We considered four hypothetical species with different germination behaviour, in terms of germination speed and capacity (final proportion of germinated seeds): (i) high germination speed and capacity (θ_H = 40.5 MPa °C h, $\psi_{b(50)}$ = −0.63 MPa and $\sigma_{\psi b}$ = 0.37 MPa); (ii) low germination speed and high capacity (θ_H = 486 MPa °C h, $\psi_{b(50)}$ = −0.63 MPa and $\sigma_{\psi b}$ = 0.37 MPa); (iii) high germination speed and low capacity (θ_H = 40.5 MPa °C h, $\psi_{b(50)}$ = −0.15 MPa and $\sigma_{\psi b}$ = 0.555 MPa); and (iv) low germination speed and capacity (θ_H = 486 MPa °C h,

$\psi_{b(50)} = -0.15$ MPa and $\sigma_{\psi b} = 0.555$). The germination capacities obtained from the these parameterisations at 0 MPa were, respectively, 0.96, 0.96, 0.61, 0.61, while the times to 50% germination (absolute level) were 64, 771, 270 and 3240 h respectively. The overall germination patterns are depicted in Fig. 2.

Three monitoring schedules were imposed on the species: (i) high frequency (i.e. every 12 h for the first 4 days, daily from the 5th to the 15th day, every 2nd day from 17th to the 31st day, every 4th day from the 35th day to the 67th day and weekly from the 74th to the 130th day); (ii) half frequency (i.e. daily for the first 4 days, every 2nd day from the 6th to the 16th day, every 4th day from the 20th day to the 40th day and weekly from the 47th to the 124th day; and (iii) high frequency but short duration (i.e. the same as 1, but ending at the 67th day). For all 12 combinations (4 species × 3 monitoring schemes), the experimental design was kept constant with eight osmotic levels, 4 replicates and 50 seeds per Petri dish, and each simulation was run 1000 times.

It should be noted that these bioassays are very long lasting, which is motivated by the presence of one very slow germinating species. Such a hypothetical species was included as the worst-case scenario. Based on experience, when germination assays include both fast germinating and slow germinating seed lots, one can be tempted to interrupt the inspections too early, which might increase the uncertainty of estimates for slow germinating lots, due to right censoring.

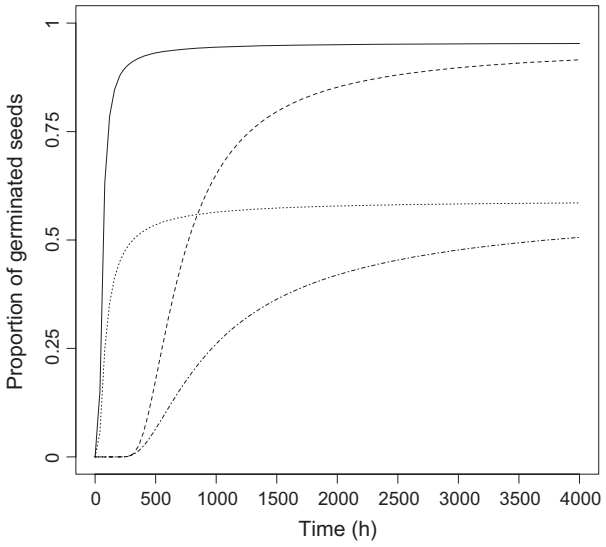

Fig. 2 Simulated germination patterns for four hypothetical plant species varying in germination rate (rapidity) and extent (capacity) under no water stress ($\psi = 0$ MPa). Solid line: fast with high germination capacity (species A); dashed line: slow with high germination capacity (species B); dotted line: fast with low germination capacity (species C); dot-dashed line: slow with low germination capacity (species D). See text for details on model parameters.

Results

The results of 1000 Monte Carlo experiments with eight water potentials, six replicates and 50 seeds per Petri dish showed that the sampling distribution was 'close to normal' for all the three parameters (Fig. 3).

With this experimental design and no random variability relating to the water potential level in each Petri dish, nonlinear regression estimators based on Eqn (2) were always very close to the real value (i.e. the initial parameters used to generate the simulated data), with almost no bias (Table 1). When random variability was added to the water potential level in each Petri dish, the bias increased slightly, but it was still negligible (lower than 0.9%). Results obtained with 'non-cumulative' counts (Eq. 4) were practically equivalent, apart from a slightly higher bias for θ_H and $\psi_{b(50)}$ (Table 1).

Below, in our consideration of the impacts of different fitting methods on the estimation of standard errors, we will only consider simulations with variation in water potential at the Petri dish level. This additional variability caused only a slight increase in the estimated standard errors, and this scenario is closer to a real experimental situation.

Considering the simulation runs with eight water potentials, six replicates and 50 seeds per Petri dish, standard errors from nonlinear regression with Eqn (2) were unacceptably low (Fig. 4), ranging on average from 1/3 to 1/4 of the Monte Carlo precision. When these standard errors were used to build 95% confidence intervals, the observed coverage (i.e. the percentage of the 1000 runs wherein the estimated interval actually captured the true value) was lower than 44%.

When using nonlinear regression and Eqn (4) (i.e. the non-cumulative model), standard errors were closer to, but still lower than, the Monte Carlo precision. The observed coverage of confidence intervals was 86%, 78% and 75%, respectively, for $\psi_{b(50)}$, $\sigma_{\psi b}$ and θ_H.

On the contrary, resampling methods (with very small difference between the jackknife and bootstrap) gave good estimates of standard errors, all being very close to the Monte Carlo precision of estimators. The observed coverage was always above 93.4%, which is not significantly different from the nominal 95% value.

Effect of experimental design

The above results were obtained from an optimal design where simulations were run with a very high sample size (eight water potentials, six replicates and

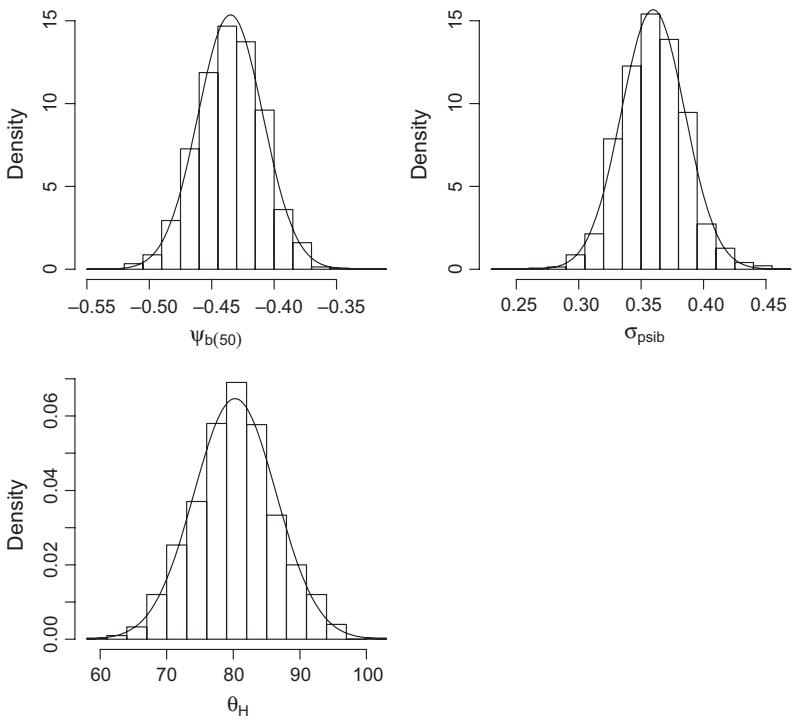

Fig. 3 Sampling distribution of parameter estimates across 1000 Monte Carlo simulations. The figure refers to the situation with additional variability at the Petri dish level (see text for more detail).

50 seeds per Petri dish); however, it is interesting to assess whether a simpler design (thus cheaper to undertake) would affect the estimation process. Indeed, the bias of estimators increased up to 9% for very small experiments, that is, with four water potentials, two replicates and 10 seeds per Petri dish (Table 1).

In general, $\sigma_{\psi b}$ and θ_H were slightly more biased than $\psi_{b(50)}$. In the case of 50 seeds per Petri dish, it was necessary to maintain the number of water potential levels and replicates equal to or above eight and four, respectively, in order to keep the bias below 1.5% for these two parameters. With 10 seeds per Petri dish, however, at least 6 replicates were necessary to keep the same low level of bias.

The value for Monte Carlo precision was doubled (less precise experiment) when the number of seeds decreased from 50 to 10 (Fig. 4). In this respect, an experiment with two replicates, four doses and 50 seeds per Petri dish showed almost the same precision as an experiment with six replicates, eight doses and 10 seeds per Petri dish. Regardless of the number of seeds, experiments with two replicates and eight doses showed the same precision as experiments with four replicates and four doses.

Resampling methods gave good estimates of standard errors, apart from the case of two replicates. Excluding these small sample sizes, standard errors were very close to the Monte Carlo precision of estimators, and the observed coverage was never significantly different from the nominal 95% value.

Effect of plant species and monitoring schedule

Plant species and monitoring schedule had no marked effects on the bias for both θ_H and $\sigma_{\psi b}$ parameters. The biases for these parameters were on average −0.7% and −0.8%, respectively, in the presence of random errors on the water potential levels, but dropped to 0.03% and 0.32% without this latter source of variability. For $\psi_{b(50)}$, monitoring schedule had no marked effects, while the average bias with fast germinating species ($\psi_{b(50)} = -0.63$) was −0.14%, it increased to 2.3% for slow germinating species ($\psi_{b(50)} = -0.15$). It should be pointed out that this bias was still rather low in absolute terms, and it became even lower (−0.73%) in the absence of random errors in water potential. The precision of estimators was relatively unaffected by plant species and monitoring schedule, although the earlier termination of monitoring led to an increase in the standard deviation of the sampling distribution by 31.8%, 24.9% and 22.6%, respectively, for $\psi_{b(50)}$, $\sigma_{\psi b}$ and θ_H.

With reference to standard errors, this experiment confirmed that only the jackknife and bootstrap methods gave reliable estimates (see Table 2). In particular, the difference between the Monte Carlo precision and jackknife SEs was on average −0.3% for $\psi_{b(50)}$, −0.59% for $\sigma_{\psi b}$ and 0.24% for θ_H and increased to −1.8, −0.67 −1.0%, respectively, for these three parameters in the presence of random errors in water potential level.

Table 1 Mean parameter estimates from 1000 Monte Carlo simulations, as obtained with two nonlinear regression estimators, with several types of experimental designs

Number of water potentials	Number of replicates	Number of seeds per dish	Error on water potential	Nonlinear regression estimates					
				Eqn 2			Eqn 4		
				θ_H	$\sigma_{\psi b}$	$\psi_{b(50)}$	θ_H	$\sigma_{\psi b}$	$\psi_{b(50)}$
8	6	50	No	81.44	0.365	−0.434	81.51	0.365	−0.434
			Yes	80.52	0.362	−0.434	80.12	0.359	−0.435
		10	No	81.45	0.363	−0.432	81.01	0.36	−0.432
			Yes	80.54	0.362	−0.434	80.27	0.355	−0.435
	4	50	No	81.48	0.365	−0.434	81.42	0.364	−0.434
			Yes	80.87	0.362	−0.434	80.19	0.359	−0.435
		10	No	81.79	0.364	−0.434	81.95	0.36	−0.435
			Yes	80.45	0.36	−0.433	80.22	0.353	−0.433
	2	50	No	81.33	0.364	−0.435	81.4	0.361	−0.436
			Yes	81.07	0.36	−0.434	80	0.356	−0.434
		10	No	82.47	0.362	−0.432	81.79	0.354	−0.431
			Yes	80.31	0.357	−0.435	80.43	0.346	−0.436
4	6	50	No	81.47	0.365	−0.435	81.6	0.364	−0.435
			Yes	81.06	0.36	−0.435	80.06	0.358	−0.436
		10	No	82.22	0.362	−0.435	81.95	0.357	−0.435
			Yes	80.19	0.36	−0.435	80.63	0.356	−0.435
	4	50	No	81.63	0.365	−0.434	81.66	0.364	−0.435
			Yes	81.83	0.36	−0.435	80.28	0.358	−0.436
		10	No	81.67	0.361	−0.433	81.48	0.353	−0.435
			Yes	80.27	0.36	−0.437	81.42	0.352	−0.439
	2	50	No	81.85	0.363	−0.436	81.71	0.361	−0.436
			Yes	80.53	0.359	−0.434	80.24	0.357	−0.434
		10	No	82.23	0.355	−0.423	80.54	0.332	−0.425
			Yes	82.62	0.354	−0.441	81.2	0.342	−0.437

Simulations were performed both with and without an additional source of random variability relating to water potential level in each Petri dish. See text for more detail about the simulation process.
The true values of parameters used in simulations were: θ_H = 81.268 MPa, $\psi_{b(50)}$ = −0.434 MPa and $\sigma_{\psi b}$ = 0.365 MPa.

Discussion and conclusions

In this paper, we explored some aspects of experimental design and statistical methods for parameter estimation in 'threshold' models of germination. As a working example, we used the hydrotime model, a type of threshold model that relates the time to germination to the magnitude of the difference between the water potential of the substrate and a physiological threshold water potential (ψ_b) for radicle protrusion. However, our approach and findings are in general applicable to other types of germination models, based on the cumulative proportion of germinated seeds (e.g. hydrothermal, thermal, etc.; see Data S1), which rely on the same assumptions and share the same problems (e.g. heteroscedastic responses, lack of independence and censoring).

This simulation study shows that the use of nonlinear least squares regression to fit Eqn (2) to the observed cumulative proportion of germinated seeds is relatively reliable for point estimation and, with this respect, the consequences of interval censoring seemed to be rather low. However, it became also clear that we should never trust the 'default' standard errors obtained from standard nonlinear regression, because they are highly underestimated. Clearly, the variability in germination responses among Petri dishes treated alike does not reflect the seed-to-seed variability in terms of base water potential. Furthermore, repeated measurements are taken on the same Petri dish at different times and, therefore, each successive datum has an additional error component accruing from the proceeding observations, which is neglected by nonlinear regression (Mandel, 1957).

Ritz *et al.* (2013) have shown that nonlinear regression gives standard errors that are always smaller than those obtained using time-to-event methods in seed germination and suggested that biologists shift to the time-to-event modelling platform. Although this suggestion is appropriate for general data description in seed germination assays, at present, there are no literature examples of time-to-event models based on the hydrotime/hydrothermal time concept and no current time-to-event models have so far proven useful for predictions in natural environments. Indeed, although time-to-event models provide biologically meaningful

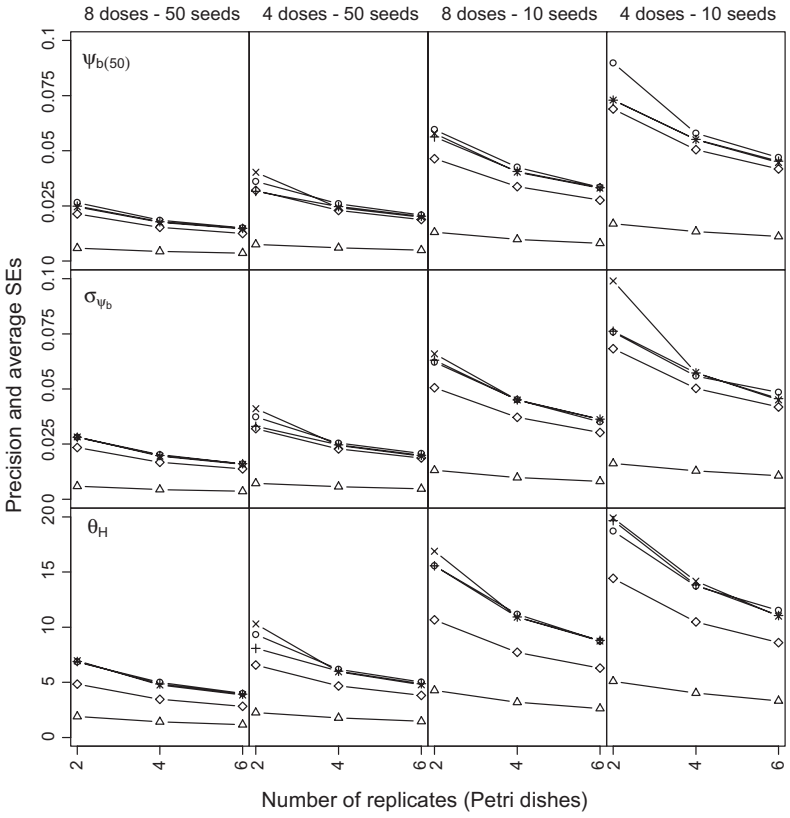

Fig. 4 Effect of experimental design on the variability of estimates across assays (Monte Carlo precision: *O*) and on the average standard errors obtained with four different estimation methods: △: nonlinear least squares regression; ◇: nonlinear regression on non-cumulative counts; +: jackknife; x: bootstrap. The figure refers to the situation with additional variability at the Petri dish level (see text for more detail).

parameters, they are mainly empirical, with no or little foundations on the physiology of seed germination. In contrast, the hydrotime/hydrothermal models are physiological-based and hold the potential to be used for dynamic simulation in fluctuating conditions of water availability and temperature (Finch-Savage *et al*., 1998).

We feel that there are strong biological reasons not to abandon threshold models based on the proportion of germinated seeds. However, in order to be able to fit these models by standard nonlinear regression, it is necessary to reinforce the accuracy of standard errors estimation. In this respect, the use of nonlinear regression based on non-cumulative counts, as suggested by Mandel (1957), brought improved standard errors (as shown in Mesgaran *et al*., 2013), but such an improvement was still inadequate, especially in the case of θ_H.

Table 2 Jackknife standard errors (average of 1000 Monte Carlo simulations) for hypothetical plant species characterised by different germination behaviour (A: high germination speed and capacity; B: high germination speed and low capacity; C: low germination speed and high capacity; D: low germination speed and capacity) and three types of monitoring schedules (F: high frequency; H: half frequency; S: high frequency, short duration)

		Plant species			
Parameter	Monitoring schedule	A	B	C	D
$\psi_{b(50)}$	F	0.016 (0.016)	0.021 (0.022)	0.029 (0.028)	0.035 (0.036)
	H	0.016 (0.016)	0.021 (0.022)	0.029 (0.029)	0.035 (0.035)
	S	0.016 (0.017)	0.030 (0.031)	0.030 (0.03)	0.044 (0.045)
$\sigma_{\psi b}$	F	0.016 (0.016)	0.018 (0.018)	0.041 (0.039)	0.052 (0.052)
	H	0.016 (0.016)	0.018 (0.018)	0.040 (0.041)	0.050 (0.050)
	S	0.016 (0.016)	0.022 (0.022)	0.042 (0.042)	0.063 (0.062)
θ_H	F	1.694 (1.740)	24.383 (24.790)	3.381 (3.360)	50.93 (50.85)
	H	1.685 (1.669)	24.216 (25.196)	3.354 (3.373)	49.65 (49.91)
	S	1.696 (1.713)	31.283 (31.572)	3.444 (3.561)	62.78 (62.78)

Monte Carlo precisions are shown in brackets. The data refer to the situation with additional variability at the Petri dish level (see text for more detail).

This paper provided the first empirical demonstration of how more reliable estimates of standard errors can be obtained using resampling methods. In particular, the grouped version of the jackknife resulted in standard error estimates very close to the Monte Carlo precision. Jackknife and bootstrap standard errors can be used to build 'naive' confidence limits using the normal approximation, which, in this work, led to an observed coverage very close to the nominal value.

In any case, the experimental design and the accuracy with which experimental units are manipulated can affect the accuracy and precision of estimators, and it is fundamental to ensure that the manipulation of moisture levels within Petri dishes is as accurate as possible, to avoid the bias resulting from this type of experimental error. The number of seeds per Petri dish was the most influential factor on the bias and precision of estimates. It is therefore important to ensure that a sufficiently high number of seeds has been used, while avoiding seed-to-seed contact by selecting dishes with the proper diameter. In practice, to obtain unbiased estimates with nonlinear regression, 16 Petri dishes (four doses × four replicates or eight doses × two replicates) will suffice when the number of seeds is high (50 or more), while 32–48 Petri dishes (at least 8 doses and 4–6 replicates) appear to be a more reasonable minimum requirement when the number of seeds within each dish is as low as 10. With these sample sizes, standard errors obtained with both the jackknife and the bootstrap were reliable.

With adequate sample size and good laboratory practices, germination behaviour (plant species) and monitoring schedule may not be very influential on the bias of estimates, although they may affect their precision. This supports the suggestion made by Mesgaran *et al*. (2013) that the final zeroes that are recorded after the germination curve levels off should be included in the process of data analysis.

In conclusion, when working with the hydrotime model or other types of threshold models based on the proportion of germinated seeds, we suggest that nonlinear least squares regression is always used in association with resampling methods, particularly when the purpose of model fitting is to compare model parameters across treatments or plant species. The jackknife method and bootstrap method (in 'grouped' form) proved to be equivalent in terms of estimation of standard errors. The latter may be more flexible in terms of hypothesis testing, but the price to pay is that it is more intensive in terms of computing power and time compared with the jackknife.

Calculating jackknife standard errors is not difficult *per se*, because it only requires fitting as many regressions as the number of Petri dishes in the experiment. Therefore, whoever is familiar with nonlinear regression can implement the required calculations, even though this may become a tedious task when the number of Petri dishes is high. The availability of some sort of scripting/programming language may be very useful. As an aid, we included a running example, together with the R script that we used for the analysis (see Data S1).

Acknowledgements

Mohsen B. Mesgaran was supported by The University of Melbourne John McKenzie Postdoctoral Fellowship. The authors wish to thank Andrew Mead for useful discussions on the hydrotime model.

References

Allen P (2003) When and how many? Hydrothermal models and the prediction of seed germination *New Phytologist* **158**, 1–3.

Allen P, Benech-Arnold R, Batlla D & Bradford K (2007) Modeling of seed dormancy. In: *Seed Development, Dormancy and Germination*, 72–106. Blackwell Publishing, Oxford (UK).

Alvarado V & Bradford K (2002) A hydrothermal time model explains the cardinal temperatures for seed germination. *Plant, Cell and Environment* **25**, 1061–1069.

Bates D & Watts D (1988) *Nonlinear Regression and its Applications*. John Wiley and Sons Inc., New York.

Bradford K (2002) Applications of hydrothermal time to quantifying and modeling seed germination and dormancy. *Weed Science* **50**, 248–260.

Chantre GR, Batlla D, Sabbatini MR & Orioli G (2009) Germination parameterization and development of an after-ripening thermal-time model for primary dormancy release of *Lithospermum arvense* seeds. *Annals of Botany* **103**, 1291–1301.

Davison A & Hinkley D (1997) *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge (UK).

Finch-Savage WE & Leubner-Metzger G (2006) Seed dormancy and the control of germination. *New Phytologist* **171**, 501.523

Finch-Savage WE, Steckel JRA & Phelps K (1998) Germination and post-germination growth to carrot seedling emergence: predictive threshold models and sources of variation between sowing occasions. *New Phytologist* **139**, 505–516.

Hardegree S (2006) Predicting germination response to temperature. 1. Cardinal-temperature models and subpopulation-specific regression. *Annals of Botany* **97**, 1115–1125.

Hunter E, Glasbey C & Naylor R (1984) The analysis of data from germination tests. *Journal of Agricultural Science* **102**, 207–213.

Mandel J (1957) Fitting a straight line to certain types of cumulative data. *Journal of the American Statistical Association* **52**, 552–556.

McNair JN, Sunkara A & Frobish D (2012) How to analyse seed germination data using statistical time-to-event analysis: non-parametric and semi-parametric methods. *Seed Science Research* **22**, 77–95.

Mesgaran MB, Mashhadi HR, Alizadeh H, Hunt J, Young KR & Cousens RD (2013) Importance of distribution function selection for hydrothermal time models of seed germination. *Weed Research* **53**, 89–101.

O'Neill ME, Thomson PC, Jacobs BC *et al*. (2004) Fitting and comparing seed germination models with a focus on the inverse normal distribution. *Australian and New Zealand Journal of Statistics* **46**, 349–366.

Onofri A, Carbonell E, Piepho H, Mortimer AM & Cousens RD (2010a) Current statistical issues in Weed Research. *Weed Research* **50**, 5–24.

Onofri A, Gresta F & Tei F (2010b) A new method for the analysis of germination and emergence data of weed species. *Weed Research* **50**, 187–198.

Onofri A, Mesgaran MB, Tei F & Cousens RD (2011) The cure model: an improved way to describe seed germination? *Weed Research* **51**, 516–524.

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. *Italian Journal of Agronomy* **5**, 69–77.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ritz C, Pipper CB & Streibig JC (2013) Analysis of germination data from agricultural experiments. *European Journal of Agronomy* **45**, 1–6.

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 Phytologist* **158**, 101–108.

Shafii B & Price WJ (2001) Estimation of cardinal temperatures in germination data analysis. *Journal of Agricultural, Biological, and Environmental Statistics* **6**, 356–366.

Wang W, Song S, Li SH, Gan Y, Wu J & Cheng H (2009) Quantitative description of the effect of stratification on dormancy release of grape seeds in response to various temperatures and water contents. *Journal of Experimental Botany* **60**, 3397–3406.

Watt MS & Bloomberg M (2012) Key features of the seed germination response to high temperatures. *New Phytologist* **196**, 332–336.

Watt MS, Xub V & Bloombergb M (2010) Development of a hydrothermal time seed germination model which uses the Weibull distribution to describe base water potential. *Ecological Modelling* **221**, 1267–1272.

Wu C (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* **14**, 1261–1295.

Yu B & Peng Y (2008) Mixture cure models for multivariate survival data. *Computational Statistics and Data Analysis* **52**, 1524–1532.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1 Running example and the R script used for the analysis.