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

DOI: 10.1111/j.1365-3180.2011.00870.x

The cure model: an improved way to describe seed germination?

A ONOFRI*, M B MESGARAN†[1], F TEI* & R D COUSENS‡

*Dipartimento di Scienze Agrarie ed Ambientali, Università degli Studi di Perugia, Perugia, Italy, †Department of Agronomy and Plant Breeding, University of Tehran, Tehran, Iran, and ‡Department of Resource Management & Geography, The University of Melbourne, Parkville, Victoria, Australia

Received 21 February 2011
Revised version accepted 15 April 2011
Subject Editor: José Gonzalez-Andujar, CSIC, Spain

Summary

Traditional methods of data analysis (ANOVA and linear/non-linear regression) may often not be appropriate for datasets resulting from seed germination/emergence assays. One major problem is that they take the form of 'time to event' data with censoring, i.e. the event is only known to have occurred within an interval of time. Parametric survival models have, therefore, been proposed as appropriate alternatives. These, in turn, have the disadvantage that they assume that all the seeds will germinate at some future time, conflicting with the fact that a proportion of ungerminated viable seeds is frequently observed at the end of an experiment. The 'cure' model, which we present here, has been used in cancer research as an extension to parametric survival models, to account for a final fraction of individuals that is cured from the disease under study and will not die because of that cause. We show that the 'cure' model holds the potential to be used successfully in germination assays, as an extension to non-linear regression and conventional accelerated failure time (AFT) models, with several logical improvements in terms of distributional assumptions and censoring. By way of selected examples, it is shown that the 'cure' model can separate the effect of factors/covariates on germination capacity (final germination percentage) from that on germination velocity (germination rate) and uniformity (synchrony of germination), which may represent an advantage from a biological point of view.

Keywords: survival analysis, cure model, weeds, germination, dormancy, censoring.

ONOFRI A, MESGARAN MB, TEI F & COUSENS RD (2011). The cure model: an improved way to describe seed germination? *Weed Research* **51**, 516–524.

Introduction

Although computer packages allow researchers to fit curves easily to data, this does not necessarily mean that the analysis is correct. The packages cannot determine the plausibility of the assumptions of a particular analysis from the way in which the data were collected. Describing the time course of germination/emergence is not a trivial task, as a number of notable issues often conflict with the assumptions of traditional methods of data analysis (e.g. ANOVA and linear/non-linear regression). In particular, (i) observed data are discrete, not continuous and take the form of counts; (ii) observational units (seeds) are not independent, as they are usually clustered in replicated randomisation units (petri dishes/pots/plots) (iii) the same randomisation unit is repeatedly inspected at successive times (longitudinal data), and thus, data from different times are not independent (iv) the exact germination/emergence time may not be known, and thus, data are 'censored'. This

Correspondence: Andrea Onofri, Dipartimento di Scienze Agrarie ed Ambientali, Università degli Studi di Perugia, Borgo XX Giugno 74, 06121 Perugia, Italy. Tel: (+39) 075 5856324; Fax: (+39) 075 5856344; E-mail: onofri@unipg.it
[1]Present address: Department of Resource Management and Geography, The University of Melbourne, Parkville, Vic. 3010, Australia.

latter aspect can be highly relevant, but is almost always neglected. Taking into account the biology of seed germination/emergence, as well as the monitoring scheme, the following issues need to be considered carefully:

- Assays are often terminated before germination/emergence is complete. If ungerminated/unemerged seeds are not viable and if it is reasonable to presume that they were already non-viable at the beginning of the experiment, it may be acceptable to erase them from statistical analysis (Scott & Jones, 1990). On the contrary, it can be argued that if they are still viable, their germination/emergence time, if any, is simply longer than the duration of the experiment, and they should therefore be included in the statistical analysis. These are referred to as 'right censored' observations;
- At the first assessment time, some seeds may have already germinated/emerged, so that we only know that their time of onset of germination/emergence is earlier than the first assessment time. These are 'left censored' observations;
- Furthermore, the exact germination/emergence time is never known precisely, but it is somewhere between two successive monitoring dates. These are 'interval censored' observations.

The above issues are not always equally relevant. For example, in seed germination assays with very regular and frequent (e.g. daily) assessments, it may be that only right censoring is of practical relevance. On the other hand, in case of loose and irregular monitoring schemes, which are very frequent in seed emergence studies, left and interval censoring may also be a major concern. In any case, completely neglecting the existence of censored data may lead to biased results, as a given germination percentage is considered implicitly in most analyses to be attained at a precise moment, when in fact it will usually have occurred earlier (but never later) than the time that counts are made (Fig. 1; see also Ritz *et al*., 2010).

Survival analysis methodologies, such as accelerated failure time (AFT) and Cox regression (see Bradburn *et al*., 2003a,b for a brief introduction), have been proposed as appropriate methods of statistical modelling in seed germination/emergence assays, to account for all the above issues (Scott & Jones, 1982, 1990; Scott *et al*., 1984). In particular, AFT and other parametric survival regression procedures appear to be particularly suitable, as they directly model the probability distribution of germination/emergence times, by way of some specific function (Weibull, log normal or log logistic, for example), while the effects of explanatory variables (temperature, light, osmotic potential, etc.) are modelled by considering how they

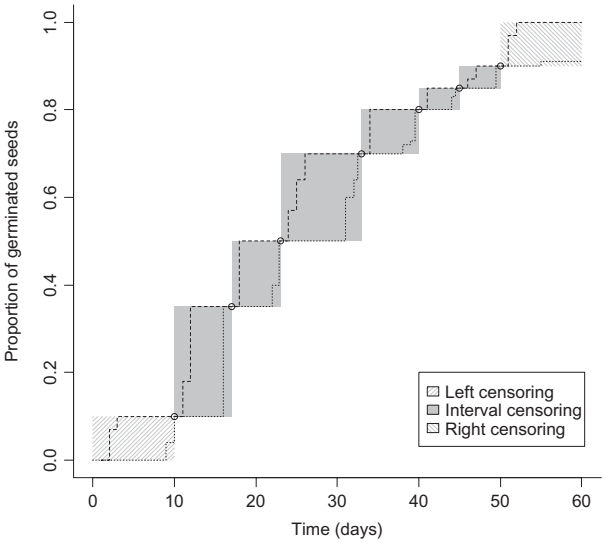

Fig. 1 Illustration of uncertainty in germination times because of the different types of censoring (see legend). Dotted lines represent two possible germination patterns, both compatible with the observed data (open circles). Note that this uncertainty is additional to that caused by the more familiar sources of experimental error.

modify the parameters (central tendency or shape, for example) of the distributions.

A recent paper has demonstrated some of the advantages of AFT regression, as compared with traditional non-linear regression models, for seed germination (Onofri *et al*., 2010), i.e. (i) the ability to cope with all the requirements in terms of error distribution and independence, without the need for stabilising transformations, (ii) a higher degree of parsimony (lower number of estimated parameters), (iii) a more efficient use of all the available experimental information, (iv) the ability to account for the existence of subpopulations within the seed lot, which may impose steps in the germination time course and (v) the ability to deal with 'censoring'.

One strong disadvantage of AFT models, as implemented in the most common statistical packages, is that they assume that all individuals will experience the event at some future time, which means that the cumulative proportion of germinated seeds must be 0 at the beginning of the assay and increase progressively afterwards, approaching 1 as time goes to infinity. Such an assumption is not generally tenable, as, especially in case of assays carried out under stressful conditions, seeds can acquire conditional dormancy and lose the ability to germinate under the imposed experimental conditions (Scott & Jones, 1990). As a consequence, the time cumulative germination/emergence curve may have an upper asymptote less then 100%, which cannot be properly described by an AFT model (Onofri *et al*., 2010). Indeed, rather

than being merely a peripheral issue, estimating the final germination proportion may be a primary interest of some germination/emergence assays.

In cancer research, the 'cure' model is often used as an extension to parametric survival models, to account for the fraction of diagnosed individuals who will not experience the event of interest (i.e. they will survive). This fraction is assumed to be cured from the disease and thus will not die because of that particular cause (Lambert *et al.*, 2007). Apart from cancer and medical research, the 'cure' model has been used also in social studies relating to marriage, motherhood, changing jobs and other events that are not certain to occur (Yamaguchi, 1992). Thus, the 'cure' model holds the potential to be flexible enough to describe the germination/emergence process; though to our knowledge, it has not been used previously in plant biology and, more specifically, in seed germination studies. In the latter case, the 'cured' fraction corresponds to the final ungerminated (but still viable) fraction of seeds at the end of an assay.

The aim of this paper is to extend the concepts presented by Onofri *et al.* (2010) and to reframe the 'cure' model within germination/emergence studies, giving an insight into its potential usefulness by applying it to some selected examples.

Definition of the 'cure' model

'Cure' models are generally implemented as 'mixture models', wherein two probability distributions are combined, i.e. one for the time to the event (for those who will experience it) and one for the surviving fraction. In other disciplines, these two components are, respectively, named 'latency' and 'incidence' (Li & Taylor, 2002); in germination studies, we will refer instead of the two main features of seed germination, i.e. germination velocity and germination capacity, respectively.

For simplicity, in the following sections, we will refer only to seed germination, even though the concepts also apply to seedling emergence. Let us assume that a germination assay has been carried out on more N seeds, by repeatedly inspecting petri dishes at successive times and collecting the number of individuals that germinated within each time interval. The assay has lasted for sufficient time to conclude that no further germination is likely to occur and a final fraction of ungerminated viable seeds (dormant) has been recorded. A 'cure' model for seed germination may be expressed as:

$$S_c(t) = \Pr(T > t) = \pi + (1 - \pi)S(t) \qquad (1)$$

where $S_c(t)$ is the probability that a given seed germinates after time t and corresponds to the cumulative proportion of ungerminated seeds, $S(t)$ is the same probability for non-dormant seeds (a conditional probability, given that the event of interest has occurred), π is the final ungerminated (dormant) fraction. Clearly, Eqn (1) reduces to a standard survival model when π is 0.

$S(t)$ is assumed to follow a specific parametric form, with unknown parameter values:

$$S(t) = 1 - \varphi(t, \mu, \sigma) \qquad (2)$$

where φ may commonly be either a normal, log-normal, log-logistic or Weibull cumulative distribution, parameterised using one location (μ) and one shape (σ) parameter. In case of the normal distribution, μ and σ represent the mean germination time ($\bar{t}$, equal in such symmetrical distributions to the median germination time t_{50}) and the standard deviation of t, respectively. Other distributional forms (probability models) are possible: for example, Peng *et al.* (1998) have proposed a very flexible F distribution that contains most of the other distributions as special cases.

The parameters μ, σ and π, respectively, quantify three different features of seed germination, i.e. speed, uniformity and capacity (Bewley & Black, 1985) and they can be easily made to depend on explanatory variables. For example, if a covariate vector $\mathbf{Z} = (z_1, z_2, \ldots z_p)$ is given, an appropriate function g may be found to model π as a function of $\mathbf{Z}$, by way of a parameter vector $\boldsymbol{\alpha}$ (germination capacity submodel):

$$\pi = g(\mathbf{Z}, \boldsymbol{\alpha}) \qquad (3)$$

Likewise, the location parameter μ can be assumed to depend on the covariate vector $\mathbf{Z}$, by way of a parameter vector $\boldsymbol{\beta}$ and a function h (germination speed submodel):

$$\mu = h(\mathbf{Z}, \boldsymbol{\beta}) \qquad (4)$$

and σ might be made to depend on covariates using a parameter vector $\boldsymbol{\gamma}$ (germination uniformity submodel) (though the shape parameter in germination assays seems to be less likely and consistently influenced by covariates: Keshtkar *et al.*, 2009; Ohadi *et al.*, 2010). The final cure model is, therefore, defined as:

$$S_c(t) = g(\mathbf{Z}, \boldsymbol{\alpha}) + (1 - g(\mathbf{Z}, \boldsymbol{\alpha})) \cdot (1 - \varphi(t, h(\mathbf{Z}, \boldsymbol{\beta}), \sigma)) \qquad (5)$$

where the proportion of ungerminated seeds is obtained by separately modelling the three main features of seed germination as functions of external covariates. The above approach can be extended to incorporate any kind of effects (linear/non-linear) of covariates/factors on μ, σ and π. Also, the three submodels do not need to be based on the same set of explanatory variables. The only limit to this is in the ability to define a meaningful 'dummy' coding for each specific experimental design and hypothesis to be tested.

Estimation of parameters and standard errors

Once the functions φ, g, h have been defined, the estimation of the unknown parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, plus the shape parameter σ, should be performed by taking into appropriate account the problem of censoring. This is clearly not possible using estimation methods based on least squares (as in traditional linear/non-linear regression). Conversely, the estimation of parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and σ (or γ, if appropriate) should be performed using maximum likelihood, based on a convenient form for the likelihood function that appropriately accounts for censoring (see for example Corbiere & Joly, 2007). An example of how a likelihood function can be coded is given in the Appendix.

Particular care needs to be taken in calculating standard errors for the estimated parameters. In non-linear regression, these are usually obtained by taking the diagonal elements of the variance-covariance matrix of parameter estimates. However, the algorithms used to approximate this matrix rely on an independence assumption, which does not hold in seed germination assays, as observational units (seeds) are generally clustered within randomisation units (petri dishes/pots/plots). This problem may be tackled in two ways. One approach is based on introducing random effects into the model, to account for the clustering units (Pinheiro & Bates, 2000). This approach has also been advocated for the 'cure' model (Price & Manatunga, 2001; Yu, 2008) and requires an appropriate redefinition of the likelihood function. Another possible approach (which was used in the following examples) is based on robust estimates of standard errors, ('sandwich estimators' – Lipsitz *et al*., 1994), which can be approximated by way of a fully iterated jackknife variance estimator, analogously to that shown by Yu and Peng (2008).

Examples

The following examples relate to the so-called hydrotime model of seed germination. The first dataset (Young, 2001) consists of germination assays with one weed species (*Raphanus raphanistrum* L., wild radish). Three replicates of 50 seeds were placed on two Whatman 1001 filter paper discs in 10 cm-diameter dishes. Each dish contained 12 mL of four PEG (8000) concentrations (Michel, 1983) to establish osmotic potentials of 0, −0.1, −0.2, −0.4 or −0.8 MPa. Dishes were put into an incubator at a 15–25°C alternating temperature and dark/light regime (12:12 h). Germination counts were made daily for the first 10 days, then every 2–3 days for 31 days and then after 46, 52 and 57 days from the start of the experiment.

It is clear from the uneven monitoring scheme that it is not possible to ensure that the exact germination time of each seed is known. This kind of interval censoring must be appropriately accounted for in data analysis.

After a preliminary selection based on Akaike's Information Criterion (AIC; Akaike, 1974), a log-normal cumulative distribution was chosen in the germination submodel (for viable non-dormant seeds) for φ (Eqn 5), with constant σ (shape parameter, corresponding to the standard deviation of germination log time) and independent of the osmotic potential. The location parameter (on a log-time scale) was modelled as a linear function of osmotic potential (Z) as an absolute value (MPa):

$$\mu = b_0 + b_1 Z \tag{6}$$

where μ on a log-time scale corresponds to the logarithm of t_{50} (time to 50% germination), b_0 is the logarithm of t_{50} at 0 MPa, while b_1 is the change in $\log(t_{50})$ when osmotic potential increases in absolute value.

For the submodel for dormant seeds, π (ungerminated fraction) was modelled using a log–log link, which implies an S-shaped asymmetric increasing response to water potential (as an absolute value):

$$\pi = \exp[-\exp(a_0 - a_1 Z)] \tag{7}$$

The estimated parameters (robust standard errors are in brackets) were σ = 0.633 (0.0224), b_0 = 0.847 (0.0433), b_1 = 3.033 (0.2450), a_0 = 1.558 (0.1104) and a_1 = 6.470 (0.4812). In the absence of a widely accepted, easy-to-calculate and easy-to-interpret measure of goodness of fit (analogous to the R^2 in linear models) in case of censoring (Schemper & Stare, 1996), a graphical comparison of predicted and observed germinations shows that the fit of the model is fairly good (Fig. 2), even though it includes only five estimated parameters. A slight underestimation/overestimation of the final dormant fraction was noted, respectively, at −0.1 and −0.2 MPa, which could not be improved by adopting a different function for the dormancy submodel. However, the fit was slightly better than that of the conventional hydrotime model in all cases (Bradford, 1990 and Gummerson, 1986; estimated parameters with standard errors were θ_H = 0.71 ± 0.022, $\psi_{b(50)}$ = −0.33 ± 0.004 and $\sigma_{\psi b}$ = 0.16 ± 0.004).

The 'cure' model can be used to extract further information about the effect of osmotic potential on median germination time and final ungerminated fraction; for example, the final ungerminated fraction at

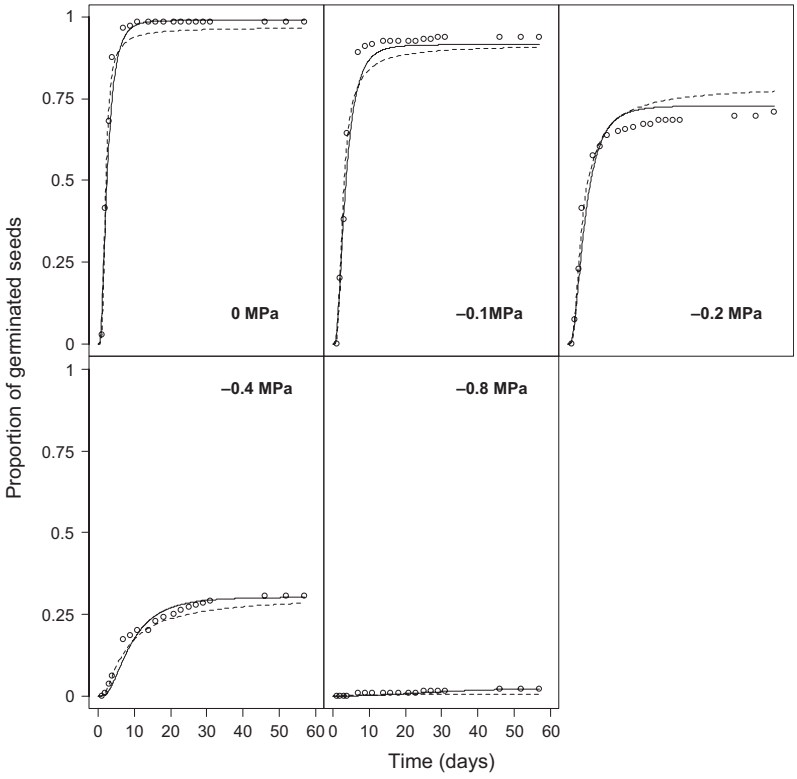

Fig. 2 'Cure' model (solid line) depicting the time course of seed germination for *Raphanus raphanistrum* at different osmotic potentials. Symbols show observed data, and the dotted line shows the conventional hydrotime model.

0 MPa is equal to exp(−exp(1.558)) = 0.0087, which is small but significantly different from 0 (P = 0.022). This final ungerminated fraction increases linearly via a log–log link with osmotic potential according to a slope value of 6.47, indicating a highly adverse effect of water stress on the final germination percentage (Fig. 3). For germinated seeds, the t_{50} at 0 MPa is exp(0.633) = 1.88 days and this value increases linearly on a logarithmic time scale (Fig. 3).

The second dataset (Tei *et al.*, 2001) was generated from an examination of *Festuca rubra* L. ssp. *commutata* Gaud. (cv. Bargreen) germination. Three replicates of 50 seeds were placed on two Whatman 1001 filter paper discs in 13 cm diameter petri dishes. Each dish contained 22 mL of three PEG (8000) concentrations to establish osmotic potentials of 0, −0.4, −0.6 or −0.8 MPa. Dishes were incubated at 15°C (alternating 14/10 h light/dark photoperiod). Germination counts were made daily for the first 10 days, then every 2–3 days for a total of 21 days from the start of the experiment.

In this second example, preliminary analyses based on the AIC showed that φ cumulative distribution of germination times was approximately log logistic with σ (shape parameter for germination uniformity) independent of osmotic potential. The germination submodel required a slightly more complex equation for μ (location parameter on a log-time scale):

$$\mu = b_0 + b_1 Z + b_2 Z^2 \tag{8}$$

where the log-t_{50} at 0 MPa (b_0) increased with osmotic potential according to a second-order polynomial. Based on AIC, a slightly more complex expression for π was also necessary to account for a very high fraction of dormant seeds at 0 MPa:

$$\pi = a_2 + (1 - a_2)\exp[-\exp(a_0 - a_1 Z)] \tag{9}$$

In this second example, the estimated parameters (robust standard errors in brackets) were σ = 0.134 (0.007), b_0 = 2.089 (0.0240), b_1 = 0.258 (0.181), b_2 = 1.148 (0.2831), a_0 = 5.057 (2.001), a_1 = 6.949 (2.729) and a_2 = 0.363 (0.029). The 'cure' model showed a much better fit than the conventional hydrotime model (estimated parameters with standard errors were θ_H = 9.74 ± 0.764, $\psi_{b(50)}$ = −0.95 ± 0.055 and $\sigma_{\psi b}$ = 0.53 ± 0.032), because of a much greater flexibility in modelling the final fraction of dormant seeds and to the possibility of using a curvilinear function for μ (Fig. 4). Also in this case, the estimated parameters may be used to derive some biologically relevant information; for example, at 0 MPa, the median germination time is exp(2.089) = 8.1 days, while the proportion of

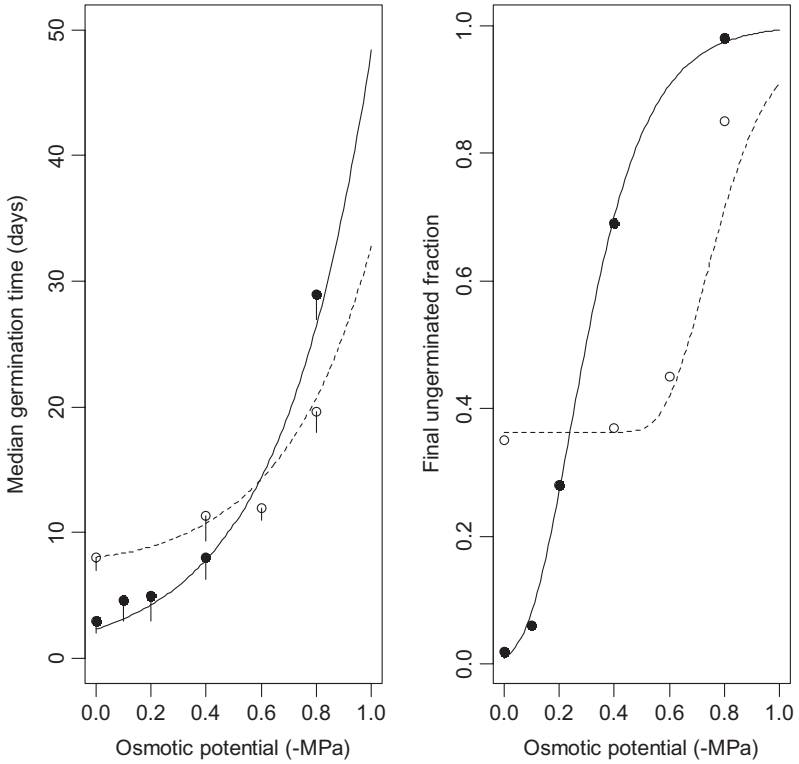

Fig. 3 Effect of osmotic potential on median germination time (for non-dormant seeds) and final ungerminated fraction for *Raphanus raphanistrum* (solid line) and *Festuca rubra* ssp. *commutata* (dotted line), as derived from the fitted models. Closed (*R. raphanistrum*) and open (*F. rubra*) circles show observed values, with vertical lines showing the uncertainty because of interval censoring (that is additional to the 'usual' sources of experimental errors).

ungerminated seeds (see Eqn 9) is 0.363. This latter quantity increases approaching 1, when the osmotic potential increases in absolute value (Fig. 4).

Concluding remarks

It is clear that much more work and the analysis of further datasets are necessary to fully assess the merits of this relatively new technique, not only for descriptive purposes, but also for predictive modelling. Therefore, any comparisons with traditional regression models would be far beyond the scope of this paper. However, the 'cure' model seems to be very promising, with particular reference to the following aspects.

Modelling approach

The 'cure' model considers separately the three main features of seed germination assays, i.e. germination velocity (μ, the reciprocal of the mean germination rate), germination capacity (π) and germination uniformity (σ). A similar approach has been advocated by Glen *et al*. (2000), even though their proposed solution is not within the framework of survival analysis. In case of osmotic potential or temperature (the latter has not been considered here), it is worth noting that none of the conventional hydrotime, thermal and hydrothermal models directly consider the final fraction of dormant seeds.

Flexibility

The wide array of functions and parameterisations for φ, g and h makes this model very flexible (at least potentially). For example, the existence of subpopulations within the seed lot might be detected using distributional forms that are able to account for bimodality, e.g. inverse Gaussian type distribution (Sanhueza *et al*., 2008) or bi-Weibull distribution. Of course, the parsimony of the final model should always be carefully considered.

Correct inferences

It may be argued that a carefully selected non-linear regression model may achieve the same modelling flexibility as the 'cure model' but with the advantage of being more straightforward to fit. However, ease of

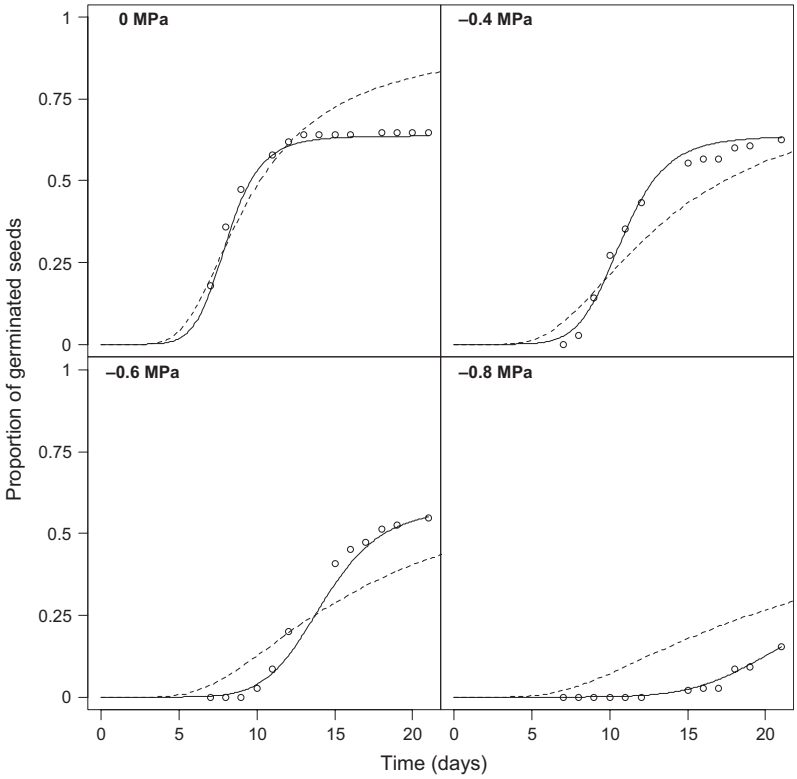

Fig. 4 'Cure' model (solid line) depicting the time course of seed germination for *F. rubra* at different osmotic potentials. Symbols show observed data, and the dotted line shows the conventional hydrotime model.

fitting and familiarity to researchers should not, on their own, determine the approach to be taken. The assumptions of the method need to be valid, and the resulting function should provide an acceptable description of the data. From a purely statistical point of view, fitting a 'cure' model to the observed germination times will always lead to correct inferences, as long as the clustering of seeds within randomisation units is appropriately taken into account. In contrast, the straightforward process of fitting a hydrotime model (or any other non-linear regression model) into the observed time course of cumulative percentages of germination may easily lead to incorrect inferences, because of heteroscedasticity and autocorrelation of errors over time. Furthermore, traditional non-linear regression will not be able to account for interval censoring of germination times, wherever this may represent a relevant issue.

Many weed scientists, lacking advanced statistical training, will find our method difficult to implement and thus will be inclined to continue using more familiar but less appropriate methods. We propose that a colleague with the appropriate expertise be sought, to ensure that good research is not reduced in value at the analysis stage. To encourage greater use of the more appropriate methods, friendlier computer implementations need to be developed. Furthermore, extending the 'cure' model to include random effects might be advised, to deal with more complex experimental designs used frequently in field experiments.

References

Akaike H (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control*. **AU-19,** 716–722.

Bewley JD & Black M (1985) *Seeds: Physiology of Development and Germination*. Plenum Press, New York.

Bradburn MJ, Clark TG, Love SB & Altman DG (2003a) Survival analysis part II: multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer* **89,** 431–436.

Bradburn MJ, Clark TG, Love SB & Altman DG (2003b) Survival analysis part III: multivariate data analysis – choosing a model and assessing its adequacy and fit. *British Journal of Cancer* **89,** 605–611.

Bradford KJ (1990) A water relations analysis of seed germination rates. *Plant Physiology* **94,** 840–849.

Corbiere F & Joly P (2007) A SAS macro for parametric and semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine* **85,** 173–180.

Glen D, Wilson M, Brain P & Stroud G (2000) Feeding activity and survival of slugs, *Deroceras reticulatum*, exposed to the rhabditid nematode, *Phasmarhabditis hermaphrodita*: a model of dose response. *Biological Control* **17,** 73–81.

Gummerson R (1986) The effect of constant temperatures and osmotic potentials on the germination of sugar beet. *Journal of Experimental Botany* **37,** 729–741.

Keshtkar E, Kordbacheh F, Mesgaran M, Mashhadi H & Alizadeh H (2009) Effects of the sowing depth and temperature on the seedling emergence and early growth of wild barley (*Hordeum spontaneum*) and wheat. *Weed Biology and Management* **9,** 10–19.

Kleinbaum DG & Klein M (2005) *Survival Analysis*. Springer Science Inc., New York.

Lambert P (2007) Modeling the cure fraction in survival studies. *The Stata Journal* **7,** 1–25.

Lambert P, Thompson J, Weston C & Dickman P (2007) Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* **8,** 576–594.

Li C & Taylor JMG (2002) A semi-parametric accelerated failure time cure model. *Statistics in Medicine* **21,** 3235–3247.

Lipsitz SR, Dear KBG & Zhao L (1994) Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics* **50,** 842–846.

Michel B (1983) Evaluation of the water potentials of solutions of polyethylene glycol 8000 both in the absence and presence of other solutes. *Plant Physiology* **72,** 66–70.

Ohadi S, Mashhadi H, Tavakkol-Afshari R & Mesgaran M (2010) Modelling the effect of light intensity and duration of exposure on seed germination of *Phalaris minor* and *Poa annua*. *Weed Research* **50,** 209–217.

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

Peng Y, Dear K & Denham J (1998) A generalized F mixture model for cure rate estimation. *Statistics in Medicine* **17,** 813–830.

Pinheiro J & Bates D (2000) *Mixed-Effects Models in S and S-Plus*. Springer-Verlag Inc., New York.

Price DL & Manatunga AK (2001) Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine* **20,** 1515–1527.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R Foundation for statistical Computing. URL: http://www.R-project.org (ISBN 3-900051-00-3), Vienna, Austria.

Ritz C, Pipper C, Yndgaard F, Fredlund K & Steinrucken G (2010) Modelling flowering of plants using time-to-event methods. *European Journal of Agronomy* **32,** 155–161.

Sanhueza A, Leiva V & Balakrishnan N (2008) A new class of inverse Gaussian type distributions. *Metrika* **68,** 31–49.

SAS Institute Inc (2000) The NLMIXED procedure. In: *SAS/STAT User's Guide*, Version 8, 2419–2504. SAS Inc., Cary, NC.

Schemper M & Stare J (1996) Explained variation in survival analysis. *Statistics in medicine* **15,** 1999–2012.

Scott S & Jones R (1982) Low temperature seed germination of *Lycopersicon* species evaluated by survival analysis. *Euphytica* **31,** 869–883.

Scott S & Jones R (1990) Generation means analysis of right-censored response-time traits: low temperature seed germination in tomato. *Euphytica* **48,** 239–244.

Scott S, Jones R & Williams W (1984) Review of data analysis methods for seed germination. *Crop Science* **24,** 1192–1199.

Tei F, Benincasa P & Ciriciofolo E (2001) Effetto del potenziale idrico e della temperatura sulla germinazione di alcune specie graminacee da tappeto erboso. In: Atti XXXIV Convegno della Società Italiana di Agronomia, Pisa, Italy, 200–201.

Yamaguchi K (1992) Accelerated failure-time regression models with a regression model of Surviving Fraction: an application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association* **87,** 284–292.

Young K (2001) *Germination and emergence of wild radish* (*Raphanus raphanistrum* L.), 205. PhD Thesis, University of Melbourne.

Yu B (2008) A frailty mixture cure model with application to hospital readmission data. *Biometrical Journal* **50,** 386–394.

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

Appendix

The likelihood function for parametric survival models has been motivated by Kleinbaum and Klein (2005), and the extension to the 'cure model' can be found, for example, in Corbiere and Joly (2007). We will hereby give the application to the selected examples that we studied. Assume that we have N seeds, $f(t)$ is the probability density function corresponding to the selected form for φ (in this case log normal or log logistic) and that the data are of the form $Y_i=(\delta_i,t_i,t_{2i},Z_i)$ where δ is a censoring indicator for the seed i (with i from 1 to N and δ equal to 0, 1, 2 and 3, respectively, for right censored, uncensored, left censored and interval censored seeds), t is the germination/censoring time, t_2 is the end of the interval for interval censored observations, Z_i is the osmotic potential for the seed i, g and h are the functional forms to model the effect of osmotic potential on π (via the parameter vector $\boldsymbol{\alpha}$) and μ (via $\boldsymbol{\beta}$). Consequently, the likelihood contribution for each individual in the above examples is:

$$L_i=\begin{cases}[1-g(Z_i,\boldsymbol{\alpha})]f(t_i,h(\mathbf{Z}_i,\boldsymbol{\beta}),\sigma); & \text{if } \delta=1\\ g(Z_i,\boldsymbol{\alpha})+[1-g(Z_i,\boldsymbol{\alpha})][S(t_i,h(\mathbf{Z}_i,\boldsymbol{\beta}),\sigma)]; & \text{if } \delta=0\\ g(Z_i,\boldsymbol{\alpha})+[1-g(Z_i,\boldsymbol{\alpha})][1-S(t_i,h(\mathbf{Z}_i,\boldsymbol{\beta}),\sigma)]; & \text{if } \delta=2\\ \{g(Z_i,\boldsymbol{\alpha})+[1-g(Z_i,\boldsymbol{\alpha})][S(t_i,h(\mathbf{Z}_i,\boldsymbol{\beta}),\sigma)]\}- & \\ \{g(Z_i,\boldsymbol{\alpha})+[1-g(Z_i,\boldsymbol{\alpha})][S(t_{2i},h(\mathbf{Z}_i,\boldsymbol{\beta}),\sigma)]\}; & \text{if } \delta=3\end{cases} \tag{10}$$

The estimate of the unknown parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and σ) is obtained by maximising the log-likelihood function:

$$\mathrm{LL}=\sum_{i=1}^{N}\log(L_i) \tag{6}$$

If all the observations are uncensored and f is normal or log normal, the model reduces to a traditional linear/non-linear model. This supports the idea that the above method can be seen as an extension to either

parametric survival regression (to incorporate a final surviving fraction) or linear/non-linear regression (to incorporate censored observations).

The maximisation of the likelihood function can be performed using one of the available packages; in the above examples, the minimisation of the negative log likelihood was performed using the optim function in the R statistical environment (version 2.11.0; R Development Core Team, 2010). Readers may also refer to the work performed by Corbiere and Joly (2007) with the NLMIXED procedure of SAS (SAS Institute Inc., 2000) or by Peng (http://post.queensu.ca/~pengp/software.html: date of last access 6 June 2011) with R/S-plus and Lambert (2007) with STATA.