Seed Science Research

cambridge.org/ssr

Seed longevity: analysing post-storage germination data in R to fit the viability equation

Dustin Wolkis[1,2], Angelino Carta[3], Shabnam Rezaei[4] and Fiona R. Hay[4]

[1]Department of Science and Conservation, National Tropical Botanical Garden, Kalāheo, HI 96741, USA; [2]Seed Conservation Specialist Group, Species Survival Commission, International Union for Conservation of Nature, Gland 281196, Switzerland; [3]Department of Biology, Botany Unit, University of Pisa, Pisa, Italy and [4]Department of Agroecology, Aarhus University, Slagelse 4200, Denmark

Methods Paper

Cite this article: Wolkis D, Carta A, Rezaei S, Hay FR (2025). Seed longevity: analysing post-storage germination data in R to fit the viability equation. *Seed Science Research* **35**, 33–40. https://doi.org/10.1017/S0960258524000291

Received: 28 March 2024
Revised: 11 July 2024
Accepted: 12 July 2024
First published online: 13 January 2025

Keywords:
germination data; moisture relations; probit analysis; R code; seed longevity; seed survival curve; viability equation

Corresponding authors:
Dustin Wolkis;
Email: dwolkis@ntbg.org
Fiona R. Hay;
Email: fiona.hay@agro.au.dk

Abstract

For many decades, seed germination data have been modelled by probit analysis. In particular, it is the basis of the seed viability equation used, in the first instance, to describe the decline in germination of seeds in storage, but then also the rate of the decline, depending on seed moisture content and the temperature of storage. The underlying assumption of a probit model is that the response follows a normal distribution, in this case, loss of the ability to germinate over time. Probit analysis also takes into account the binomial error associated with germination data. Many statistical packages have probit analysis as an option within the generalized linear modelling framework; here, we present code for applying probit analysis in the free software, R. Codes are provided for fitting a single survival curve, for a single seed lot stored in a constant storage environment; for fitting multiple survival curves and evaluating the effect of constraining parameters for the different seed lots; and lastly, to model the moisture relations of seed longevity. The code bases provided could also be used in pollen and fern/bryophyte spore longevity modelling.

Introduction

Probit analysis is often used in seed research to analyse germination data collected in response to a continuous variable that either promotes or prevents – for various reasons including seed death – germination (Hay et al., 2014). In particular, probit analysis is the basis of the seed viability equation that can be used to predict seed longevity depending on the storage environment (moisture content, temperature; Ellis and Roberts, 1980a, 1980b; Ellis, 2022). Probit analysis is also used to compare the longevity (shapes of the survival curves showing germination percentage vs. storage period) of different seed lots. Samples of seeds from the different lots are stored under the same conditions and subsamples are removed after different periods of time for a germination test (Hay et al., 2022). These different 'seed lots' may represent different species, genotypes within a species (Lee et al., 2019) or, for example, seeds that have experienced a different maturation environment or duration (Hay et al., 2015; White et al., 2023), seeds resulting from distinct mating (e.g., self- vs. cross-pollination; Carta et al., 2015), or which have been treated differently in some way after harvest as part of a designed experiment (e.g. Whitehouse et al., 2015).

Probit analysis belongs to the generalized linear family of models (McCullagh and Nelder, 1989). In a generalized linear model (GLM), rather than fitting a linear relationship between the dependent variable (in our case, germination percentage) and the explanatory variable (storage period) directly, there is a transformation of the dependent variable and a linear relationship is fitted between this 'linear predictor' and the explanatory variable. In probit analysis, the germination data are transformed to the linear predictor probit value using the cumulative normal distribution function, Φ^{-1}. Thus, it is assumed that there is a normal distribution in the response of the seeds to the explanatory variable. Most statistical tools use a maximum-likelihood estimation (MLE) procedure to fit a probit GLM while taking into account the binomial error distribution of germination data, which is determined by the number of seeds tested. The error distribution is binomial since each seed has two possible outcomes: it can either germinate or not germinate. The two parameters that are usually estimated are the *intercept*, which is the fitted or theoretical initial viability of the seed lot on the probit, linear predictor scale and the *slope*, the rate of change in the value of the linear predictor. Using the terminology of the viability equations, these parameters are referred to as K_i and $-1/\sigma$, respectively. Thus, the GLM describing the change in viability during storage can be written:

$$y = \phi(v) = \phi\left(K_i - \left(\frac{1}{\sigma}\right)p\right) \quad (1)$$

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

where y is the proportion of seeds that germinate, v is the linear predictor expressed in probits, K_i is the initial viability in probits, p is the storage period and σ is the time for viability to fall by one probit and hence, the standard deviation of the normal distribution of seed deaths over time. After applying probit analysis, the results are usually illustrated by plotting the cumulative normal distribution of germination over time which typically declines and is referred to as the seed 'survival curve'. By fitting Equation 1 using probit analysis, the time for viability to fall to certain levels may also be estimated. For example, p_{50} is the time for viability to fall to 50% which is often used as an index of seed longevity (Hay et al., 2022).

Despite recent efforts aimed at describing and enhancing the use of free software for germination modelling (Carta et al., 2022; Onofri et al., 2022), seed survival modelling in such an open-source framework is still under-applied (but see Carta et al., 2018). This may limit longevity studies – or the interpretation of such studies – among the seed science community. Here, we show how probit analysis can be applied in the free statistical software environment, R (R Core Team, 2023), to model seed survival data. In addition to showing how to fit the basic GLM to data for one seed lot (Box 1), thereby fitting Equation 1, we also show how to test hypotheses when survival curves for multiple seed lots or treatments are being compared (Supplementary Material 1). Lastly, we present the probit analysis code in which σ (Equation 1) is modelled as a function of the storage environment (Supplementary Material 2).

Fitting a simple single survival curve (data for one seed lot)

It may be appropriate to analyse the data for multiple stored seed lots completely independently, i.e., without any hypothesis testing (see below). This might occur for example, in a study where the aim is to screen diverse taxa for seed longevity (e.g., Probert et al., 2009). In this case, the following code (Box 1) could be run for the data for each seed lot. The data should be arranged as quantity or 'dose' of the explanatory variable (in our case, storage period), number of seeds that germinated and number of seeds that were sown. Strings of observations where germination is equal to ~100%, at the start of storage, or to 0% at the end of storage should be excluded from the analysis; this is usually done manually. Although these values have less influence in the MLE fitting process, including a run of 100 or 0% values can be likened to trying to fit a linear relationship to data which has a 'broken stick' response, i.e. with a high or low plateau, before or after the slope. This means only including data from the last highest percentage germination value (but see below) to the first zero percentage germination value.

Fitting multiple survival curves (data for multiple seed lots) and hypothesis testing

Often when we study seed longevity, we are interested in the impact of some sort of treatment on subsequent seed longevity. Hence, we conduct storage experiments on multiple seed lots and want to test whether the survival curves are significantly different or not. This hypothesis testing can be done using approximate F-tests. The data from such designed experiments should be arranged as before, but with an additional factor variable which is the treatment code. Different models are then fitted to the data:

(i) the 'independent model' with different estimates for both the slope and intercept for each treatment (or species, seed lot, or genotype) (Fig. 1A);
(ii) the number of parameters in the model is reduced by estimating a single, common intercept estimate for all the seed lots ('common intercept'; Fig. 1B);
(iii) instead of estimating a common intercept for the seed lots, we reduce the number of parameters compared with the independent model, by estimating a common value for the slope parameter for all the seed lots ('common slope'; Fig. 1C);
(iv) lastly, we fit a model with common estimates for both the slope and intercept among all the seed lots ('one line'; Fig. 1D).

The F-test is used to test whether there is a significant increase in the residual deviance after fitting a model with fewer parameters (the independent model has the most parameters: an intercept, and slope for each seed lot/treatment). Thus, the residual deviance after fitting the common slope or common intercept models is compared against the residual deviance after fitting the independent model; if there has not been a significant increase (F-probability >0.05), the residual deviance after fitting the one-line model is compared with the residual deviance from the common slope and/or common intercept models. If there is a significant increase in residual deviance comparing the one-line model against the common slope or common intercept models (F-probability ≤0.05), then the choice of accepting one of these models can be based on the relative F-probability values for these two models when compared against the independent model (choosing the one with the highest F-probability), unless there is a scientific (i.e., biological) reason why the other is more appropriate. In this way, by the use of F-tests, we determine the model with the fewest number of parameters that adequately describes the data.

In some experiments, there may be a sensible reason why seed lots do not have the same intercept, the most obvious being if seeds from the same bulk seed lot are stored under different storage conditions, in which case the treatment is the different storage environments (e.g., storage moisture content, as described below). Otherwise, this hypothesis is perhaps in general, less relevant to test. In the initial model development that led to the viability equations, it was concluded that the slopes of the survival curves are the same for all seed lots within a species if stored under identical conditions of moisture content and temperature (Ellis and Roberts, 1980a, 1980b). Although it is now generally accepted that this is not necessarily the case (Whitehouse et al., 2018; Lee et al., 2019; Ellis, 2022), this is still an obvious simplification of the full (independent) model and it is important to test this assumption. If the seed lots (or treatments) being compared are not from the same or closely related species, there is no underlying reason to expect them to respond similarly to a particular storage environment. Having established that it is possible to constrain the survival curves from the different seed lots (or treatments) to either a common intercept or a common slope, the next step is to examine whether it is possible to accept a model where both these parameters are constrained among different seed lots, i.e. whether a single survival curve adequately explains the data obtained for seeds subjected to or derived from different treatments. Testing whether a single curve can be fitted is in fact testing whether there has been a significant effect of the

https://doi.org/10.1017/S0960258524000291 Published online by Cambridge University Press

BOX 1. R script for fitting a simple single survival curve.

This is sample code for fitting a simple single survival curve (data for one seed lot). Original data are from Rezaei et al. (2023) and are supplied as Supplementary Material 3.

```
# load libraries
library(tidyverse) # useful for data manipulation and
visualization

## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.5.0     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

# set working directory
setwd("my/file/path")

# Read in data
# brna short for the study organism, Brassica napus
brna <- read.csv("Rezaei_etal_2023_SSR.csv", header = TRUE)
# NB: MC = %f.wt.

# citation for data set;
# Rezaei, S., Buitink, J. & Hay, F.R. (2023) Contrasting seed moisture
sorption behaviour between two species and the implication for seed
longevity. Seed Science Research, 1–9. https://doi.org/10.1017/
S0960258523000156

# Tell R which variables are factors
brna <- brna |>
  mutate(Treatment = as.factor(Treatment),
         sorption = as.factor(sorption),
         chamber.RH = as.factor(chamber.RH),
         cycle = as.factor(cycle))

# Subset a single treatment
# We will use desorption, cycle 1, and 30% RH
d130<-droplevels(subset(brna, sorption == "Desorption" &  cycle == "1" &
chamber.RH == "30"))
glimpse(d130)

## Rows: 6
## Columns: 10
## $ Treatment    <fct> Desorption 130, Desorption 130, Desorption 130,
```

https://doi.org/10.1017/S0960258524000291 Published online by Cambridge University Press

```
Desorptio…
## $ MC           <dbl> 5.212099, 5.212099, 5.212099, 5.212099, 5.212099,
5.212099
## $ storage.RH   <dbl> 32.43, 32.43, 32.43, 32.43, 32.43, 32.43
## $ period       <int> 72, 84, 85, 99, 128, 183
## $ sown         <int> 90, 90, 90, 90, 90, 90
## $ germinated   <int> 83, 86, 82, 81, 58, 0
## $ germ.percent <dbl> 92.22222, 95.55556, 91.11111, 90.00000, 64.44444,
0.00000
## $ sorption     <fct> Desorption, Desorption, Desorption, Desorption,
Desorptio…
## $ chamber.RH   <fct> 30, 30, 30, 30, 30, 30
## $ cycle        <fct> 1, 1, 1, 1, 1, 1
```

```
### Fit simple single survival curve
#### "germinated" is the number of seeds that germinated. "sown" are the
number of seeds that were sown.

simple<-glm(cbind(germinated, sown-germinated) ~ period, family =
binomial(link = "probit"),  data = d130)
summary(simple)

##
## Call:
## glm(formula = cbind(germinated, sown - germinated) ~ period,
##     family = binomial(link = "probit"), data = d130)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.315944   0.281077   15.36   <2e-16 ***
## period      -0.032867   0.002419  -13.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 326.532  on 5  degrees of freedom
## Residual deviance:  21.241  on 4  degrees of freedom
## AIC: 44.852
##
## Number of Fisher Scoring iterations: 6
```

View the model coefficients

```
simple$coefficients

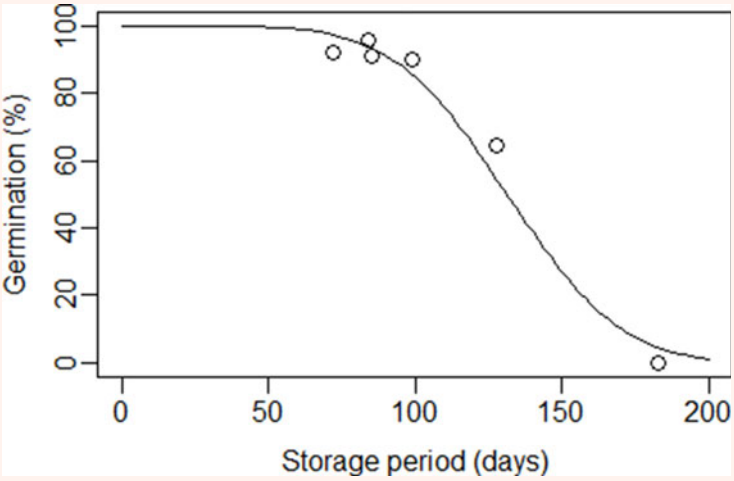
## (Intercept)      period
##  4.31594357 -0.03286691
```

Plot the simple model

```
plot(d130$period, d130$germinated/d130$sown*100, xlab="Storage period
(days)", ylab="Germination (%)", xlim=c(0,200), ylim=c(0,100), mgp=c(2,.5,0),
```

https://doi.org/10.1017/S0960258524000291 Published online by Cambridge University Press

```
pch=1, cex=1.2)
curve(pnorm(coef(simple)[1] + coef(simple)[2]*x)*100, add=TRUE, lty=1, lwd=1)
```



Construct a table of coefficients

Load package 'MASS' to estimate p_{50} and p_{85}

```
library(MASS)
```

Contruct the acual table

```
table.simple<- signif(as.data.frame(summary(simple)$coefficients[,c(1,2,4)]),
4)
table.simple[2,1]<- signif(-1/table.simple[2,1], 4)

# dose.p
table.simple[3,1]<- signif(dose.p(simple, cf = 1:2, p = 0.5), 4)
table.simple[3,2]<- signif(unname(attributes(dose.p(simple, cf = 1:2, p =
0.5))$SE[, 1]), 4)
table.simple[4,1]<- signif(dose.p(simple, cf = 1:2, p = 0.85), 4)
table.simple[4,2]<- signif(unname(attributes(dose.p(simple, cf = 1:2, p =
0.85))$SE[, 1]), 4)

row.names(table.simple)<-c("Ki", "sigma", "p50", "p85")
table.simple$`Std. Error` <- round(table.simple$`Std. Error`, 4)

table.simple[3,3] <- "-"
table.simple[4,3] <- "-"
table.simple$AIC<- "-"
table.simple$AIC[1]<- round(summary(simple)$aic, 4)

table.simple
```

```
##       Estimate Std. Error  Pr(>|z|)     AIC
## Ki       4.316     0.2811 3.277e-53 44.8524
## sigma   30.420     0.0024 4.937e-42       -
## p50    131.300     2.7250         -       -
## p85     99.780     2.4810         -       -
```

https://doi.org/10.1017/S0960258524000291 Published online by Cambridge University Press

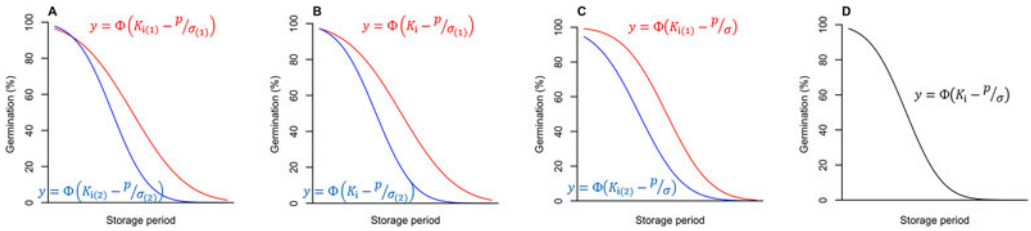


Figure 1. Examples of fitting multiple survival curves and different models. (A) 'Independent model' with different estimates for both the slope and intercept for each treatment (or species, seed lot, or genotype); (B) the number of parameters in the model is reduced by estimating a single, common intercept estimate for all the seed lots ('common intercept'); (C) instead of estimating a common intercept for the seed lots, we reduce the number of parameters compared with the independent model, by estimating a common value for the slope parameter for all the seed lots ('common slope'); (D) lastly, we fit a model with common estimates for both the slope and intercept among all the seed lots ('one line').

treatment on the subsequent response to seed storage. Often, we summarize this as 'seed longevity' although it is not a direct comparison of what we actually normally measure as seed longevity (e.g. p_{50}).

The R code for fitting this set of models and for the corresponding *F*-tests is provided in Supplementary Material 1. In the table showing the model comparisons from the *F*-tests, all the comparisons are significant at the 5% level (*F*-probability <0.05) except the comparison of the common intercept model vs. the independent model. Hence, in this case, the common intercept model can be accepted.

One-step fitting of the viability equation

The effect of storage environment on seed longevity was originally determined by storing seeds at different moisture contents and temperatures according to a factorial design (Ellis and Roberts, 1980b). Then, individual survival curves were fitted to get estimates of the intercept and slope for each combination of storage moisture content and temperature using probit analysis. Lastly, regression analysis was used to model the effect of moisture content or temperature on σ (i.e., −1/slope from the probit analysis). Separately, these two relations can be written as:

at constant temperature, *t*:

$$\log \sigma = K_t - C_W \log m \tag{2}$$

and at constant moisture content, *m*:

$$\log \sigma = K_m - C_H t - C_Q t^2. \tag{3}$$

Combined, they become

$$\log \sigma = K_E - C_W \log m - C_H t - C_Q t^2, \tag{4}$$

where m is seed moisture content during storage (f.wt. basis) and t is the temperature of storage; K_E, C_W, C_H and C_Q are 'species constants' which, it was assumed, do not vary among seed lots within the same species, thereby resulting in the same value for σ when seeds of the same species are stored at the same moisture content and temperature (see above).

However, this is a two-step process which does not take into account the error in the germination data. It also requires a factorial design for the storage experiments, with seeds stored at fixed moisture contents at one or more fixed temperatures, though as a 'short cut', temperature can be disregarded since we expect all seed lots to respond similarly to changes in temperature (Dickie et al., 1990). As an alternative approach, and given the advances in statistical software, it was recommended that a 'one-step' approach be used (Hay et al., 2003), to directly fit the full model (i.e., Equation 4 substituted into Equation 1) while taking into account all of the error in the data, including the binomial sampling error. This approach also means that the storage moisture content and temperature are entered as variables and they can vary. So, for example, the moisture content does not need to be exactly the same for samples of seeds stored at different temperatures. The model that is therefore fitted using the one-step approach is:

$$y = \phi(v) = \phi\left(K_i - \left(\frac{1}{10^{K_E - C_W \log m - C_H t - C_Q t^2}}\right)p\right). \tag{5}$$

This can be reduced if seeds are stored at one temperature or moisture content (i.e., fitting Equations 2 or 3, respectively).

$$y = \phi(v) = \phi\left(K_i - \left(\frac{1}{10^{K_t - C_W \log m}}\right)p\right); \tag{6}$$

$$y = \phi(v) = \phi\left(K_i - \left(\frac{1}{10^{K_m - C_H t - C_Q t^2}}\right)p\right). \tag{7}$$

Hypothesis testing can also be carried out, to test whether parameters of the equation vary for different seed lots or following different treatments, as for example, done by Hay et al. (2003) for different ecotypes of *Arabidopsis thaliana*. The process for this is similar to that for comparing the survival curves for multiple seed lots, above. We provide the code for fitting Equation 2 in Supplementary Material 2, which can easily be modified to fit equations 3 or 5. In this code, we test whether K_t and C_W are the same depending on whether seeds are losing (desorbing) or gaining (adsorbing) moisture (Rezaei et al., 2023). In the model comparison table, the *F*-probabilities are all greater than 0.05, hence in this case, it can be concluded that the one-line relationship adequately explains the data. That is, there are not significant differences in K_t (and hence, assuming the effect of temperature on seed longevity is also the same, K_E) and C_W depending on whether the seeds are de- or adsorbing moisture.

Fitting a modified survival curve

There are two main modifications that might be appropriate to consider, when the data are indicating that either (a) initial viability increases during the first period of storage due to the breaking of dormancy due to 'after-ripening' (Baskin and Baskin, 2022); or (b) there

https://doi.org/10.1017/S0960258524000291 Published online by Cambridge University Press

is a 'lag' phase before percentage viability declines, but the viability during this phase is hovering at a level that is significantly below 100% (Mead and Gray, 1999). The ability to model the after-ripening effect is greater when a significant proportion of the seed lot is dormant when the seeds are first put into the storage environment and if there are a number of data points collected during the breaking of dormancy (Whitehouse et al., 2015, 2018; Ellis, 2022). This may mean that it is helpful to sample more frequently at the start of the storage experiment, if the study of after-ripening is an objective. We have not provided these options here, but are working on the R code, depending on the needs of the seed research community.

Conclusions

Here, we have shown how to apply probit analysis in R to model germination data for stored seeds, either from designed experiments or data that is collected for example, as part of routine viability monitoring of seed/gene banks. This should help advance our understanding of seed longevity: across different taxa, in response to different harvesting and/or post-harvest treatments, and in response to different storage environments. This will contribute to overarching efforts in relation to the conservation of (agro-)biodiversity, as well as improving our understanding of seed physiological quality and the factors that influence the natural capital value of seeds (Mattana et al., 2022). As well as generating survival curves from relatively simple experiments, we also provide the code to proceed with one-step fitting of the viability equation describing the effect of change in storage moisture content on the slope of the survival curve. This can be adapted to also (or perhaps separately) model the effect of change in storage temperature on the survival curve or for example, see how non-storage variables are related to seed longevity (White et al., 2023). This is of particularly importance when examining seed longevity variation in natural systems. For instance, if seed longevity was estimated across multiple seed lots and taxa or if storage conditions were not properly known for all seed lots (e.g. seed 'storage' in the soil; Moravcová et al., 2022), our codes could be adapted to model seed longevity while accounting for the statistical non-independence of data points due to phylogenetic relatedness of taxa or shared storage conditions among groups of seed lots (see Carta et al. 2022 for a detailed description of this methodological approach). Within seed science, however, it is not just germination data for stored seeds that are modelled using probit analysis. Population-based threshold models similarly assume that there are underlying normal distributions in the germination response of seeds to variables such as time in the germination test, water potential, salinity, etc. (Bradford and Bello, 2022). Hence, our codes could also be adapted to analyse data from other experiments where the objective is to understand the seed germination response to a single or multiple continuous variables (Hay et al., 2014). Additionally, the demonstration we have provided could also be used to model longevity and fit survival curves for pollen as well as fern and bryophyte spores.

Data availability. The data used in this demonstration were originally presented in Rezaei et al. (2023). We provided a subset of those data as well as our R script in the supplementary materials.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0960258524000291.

Acknowledgements. The authors wish to thank two anonymous reviewers and the handling editor, Olivier Leprince, whose comments greatly improved the manuscript and the R code. Author institutions provided institutional support.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflicts of interest. The authors declare none.

Author contributions. FRH and DW conceived the idea, AC, FRH and DW designed it, AC and DW compiled the codes, SR and FRH provided the data and Genstat codes, FRH led the writing with contributions from all authors.

References

Baskin CC and Baskin JM (2022) Mimicking the natural thermal environments experienced by seeds to break physiological dormancy to enhance seed testing and seedling production. *Seed Science and Technology* **50**, 21–29.

Bradford KJ and Bello P (2022) Applying population-based threshold models to quantify and improve seed quality attributes. In Buitink J and Leprince O (eds), *Advances in Seed Science and Technology for more Sustainable Crop Production*, Cambridge: Burleigh Dodds Science Publishing Limited, pp. 67–154.

Carta A, Bedini G, Giannotti A, Savio L and Peruzzi L (2015) Mating system modulates degree of seed dormancy in *Hypericum elodes* L.(Hypericaceae). *Seed Science Research* **25**, 299–305.

Carta A, Bottega S and Spanò C (2018) Aerobic environment ensures viability and anti-oxidant capacity when seeds are wet with negative effect when moist: implications for persistence in the soil. *Seed Science Research* **28**, 16–23.

Carta A, Fernández-Pascual E, Gioria M, Müller JV, Rivière S, Rosbakh S, Saatkamp A, Vandelook F and Mattana E (2022) Climate shapes the seed germination niche of temperate flowering plants: a meta-analysis of European seed conservation data. *Annals of Botany* **129**, 775–786.

Dickie JB, Ellis RH, Kraakj HL, Ryder K and Tompsett PB (1990) Temperature and seed storage longevity. *Annals of Botany* **65**, 197–204.

Ellis RH (2022) Seed ageing, survival and the improved seed viability equation; forty years on. *Seed Science and Technology* **50**, 1–20.

Ellis RH and Roberts EH (1980a) Improved equations for the prediction of seed longevity. *Annals of Botany* **45**, 13–30.

Ellis RH and Roberts EH (1980b) The influence of temperature and moisture on seed viability period in barley (*Hordeum distichum* L.). *Annals of Botany* **45**, 31–37.

Hay FR, Mead A, Manger K and Wilson FJ (2003) One-step analysis of seed storage data and the longevity of *Arabidopsis thaliana* seeds. *Journal of Experimental Botany* **54**, 993–1011.

Hay FR, Mead A and Bloomberg M (2014) Modelling seed germination in response to continuous variables: use and limitations of probit analysis and alternative approaches. *Seed Science Research* **24**, 165–186.

Hay FR, Timple S and van Duijn B (2015) Can chlorophyll fluorescence be used to determine the optimal time to harvest rice seeds for long-term genebank storage. *Seed Science Research* **25**, 321–334.

Hay FR, Davies RM, Dickie JB, Merritt DJ and Wolkis DM (2022) More on seed longevity phenotyping. *Seed Science Research* **32**, 144–149.

Lee J-S, Velasco-Punzalan M, Pacleb M, Valdez R, Kretzschmar T, McNally KL, Ismail AM, Sta. Cruz PC, Sackville Hamilton NR and Hay FR (2019) Variation in seed longevity among diverse *Indica* rice varieties. *Annals of Botany* **124**, 447–460.

Mattana E, Ulian T and Pritchard HW (2022) Seeds as natural capital. *Trends in Plant Science* **27**, 139–146.

McCullagh P and Nelder JA (1989) *Generalized Linear Models*, 2nd Edn. Boca Raton, FL, Chapman and Hall/CRC.

Mead A and Gray D (1999) Prediction of seed longevity: a modification of the shape of the Ellis and Roberts seed survival curves. *Seed Science Research* **9**, 63–73.

Moravcová L, Carta A, Pyšek P, Skálová H and Gioria M (2022) Long-term seed burial reveals differences in the seed-banking strategies of naturalized and invasive alien herbs. *Scientific Reports* **12**, article 8859.

Onofri A, Mesgaran M and Ritz C (2022) A unified framework for the analysis of germination, emergence, and other time-to-event data in weed science. *Weed Science* **70**, 259–271.

Probert RJ, Daws MI and Hay FR (2009) Ecological correlates of *ex situ* seed longevity: a comparative study on 195 species. *Annals of Botany* **104**, 57–69.

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Rezaei S, Buitink J and Hay FR (2023) Contrasting seed moisture sorption behaviour between two species and the implication for seed longevity. *Seed Science Research* **33**, 126–134.

White FJ, Hay FR, Abeli T and Mondoni A (2023) Two decades of climate change alters seed longevity in an alpine herb: implications for ex situ seed conservation. *Alpine Botany* **133**, 11–20.

Whitehouse KJ, Hay FR and Ellis RH (2015) Increases in the longevity of desiccation-phase developing rice seeds: response to high-temperature drying depends on harvest moisture content. *Annals of Botany* **116**, 247–259.

Whitehouse KJ, Hay FR and Ellis RH (2018) Improvement in rice seed storage longevity from high-temperature drying is a consistent positive function of harvest moisture content above a critical value. *Seed Science Research* **28**, 332–339.

https://doi.org/10.1017/S0960258524000291 Published online by Cambridge University Press