The timing of seed germination is one of the key steps in the plant life cycle, determining when plant growth begins in natural or agricultural ecosystems. In nature, many seeds exhibit various types and levels of dormancy (Baskin and Baskin, 2014). In contrast, crop seeds often germinate as soon as water is imbibed, usually at planting time.

Germination is defined as the events that commence with the uptake of water by the quiescent dry seed and terminate with the elongation of the embryonic axis and the penetration of the radicle. The term germination is often used to indicate its completion; for example, "50% germination" indicates that 50% of a seed population has completed germination (Bewley et al., 2013). In contrast, seed dormancy blocks the completion of germination in an intact viable seed under favorable conditions (Weitbrecht et al., 2011; Baskin and Baskin, 2014). For germination to occur, quiescent seeds must be

**PREMISE:** Seed germination over time is characterized by a sigmoid curve, called a germination curve, in which the percentage (or absolute number) of seeds that have completed germination is plotted against time. A number of individual coefficients have been developed to characterize this germination curve. However, as germination is considered to be a qualitative developmental response of an individual seed that occurs at one time point, but individual seeds within a given treatment respond at different time points, it has proven difficult to develop a single index that satisfactorily incorporates both percentage and rate. The aim of this paper is to develop a new coefficient, the *continuous germination index* (CGI), which quantifies seed germination as a continuous process, and to compare the CGI with other commonly used indexes.

**METHODS:** To create the new index, the germination curves were smoothed using nondecreasing splines and the CGI was derived as the area under the resulting spline. For the comparison of the CGI with other common indexes, a regression model with functional response was developed.

**RESULTS:** Using both an experimentally obtained wild pea (*Pisum sativum* subsp. *elatius*) seed data set and a hypothetical data set, we showed that the CGI is able to characterize the germination process better than most other indices. The CGI captures the local behavior of the germination curves particularly well.

**DISCUSSION:** The CGI can be used advantageously for the characterization of the germination process. Moreover, B-spline coefficients extracted by its construction can be employed for the further statistical processing of germination curves using functional data analysis methods.

**KEY WORDS:** continuous germination index; functional regression; germination curve; nondecreasing positive smoothing splines; seed germination.

The timing of seed germination is one of the key steps in the plant life cycle, determining when plant growth begins in natural or agricultural ecosystems. In nature, many seeds exhibit various types and levels of dormancy (Baskin and Baskin, 2014). In contrast, crop seeds often germinate as soon as water is imbibed, usually at planting time.

Germination is defined as the events that commence with the uptake of water by the quiescent dry seed and terminate with the elongation of the embryonic axis and the penetration of the radicle. The term germination is often used to indicate its completion; for example, “50% germination” indicates that 50% of a seed population has completed germination (Bewley et al., 2013). In contrast, seed dormancy blocks the completion of germination in an intact viable seed under favorable conditions (Weitbrecht et al., 2011; Baskin and Baskin, 2014). For germination to occur, quiescent seeds must be
hydrated at suitable temperatures and in the presence of oxygen. In some cases, specific germination triggers such as fire are required (reviewed by Baskin and Baskin, 2014).

Under natural conditions, seasonal changes play a major role in the plant life cycle. Temperature is a good indicator of the time of year (Burghardt et al., 2015), and thus often regulates seed dormancy and germination (Probert, 2000; Fenner and Thompson, 2005; Baskin and Baskin, 2014). Two distinct effects of temperature have been identified in these processes: first, temperature influences the dormancy per se, and second, temperature determines the germination speed of nondormant seeds (Bradford, 2002). Another environmental factor is water availability, as the water potential of both the soil and seeds varies with the seasons (Bradford, 2002).

Seed populations tend to germinate in a characteristic pattern over time, resulting in a sigmoid curve when the percentage (or absolute number) of seeds that have completed germination is plotted against time. These germination time courses can be used to elucidate the timing, uniformity, and extent of germination in seed populations. The uniformity of germination is indicated by the time between two germination percentiles, such as the period between the points when 10% and 90% of seeds have germinated, with smaller values indicating a greater uniformity. The nondormant viable seeds complete germination, but the dormant or nonviable seeds fail to do so, resulting in a time course that reaches a final percentage of less than 100%.

Seed germination time courses often exhibit complex patterns (Bewley et al., 2013), and most seed germination experiments present difficulties in interpreting and analyzing results (Kader, 2005; Ranal and Santana, 2006). Germination is a qualitative developmental response of seeds that occurs in time, but individual seeds within a given treatment respond within different time intervals. This results in a situation where the final germination percentage alone is not sufficient for a comparative analysis of data sets, a problem for which various germination measurement techniques and methods have been proposed as solutions (reviewed by Kader, 2005; Ranal and Santana, 2006). Time, rate, homogeneity, and synchrony are important aspects that can be measured, informing the dynamics of the germination process. Because they are relevant for crop establishment, these characteristics are important not only for physiologists, seed technologists, and ecologists, but also for agronomists.

A number of coefficients, discussed by Kader (2005), were developed to characterize the entire course of the seed germination process. The mean germination time (MGT) (Orchard, 1977) computes the day of average germination; accordingly, the lower the MGT, the faster a population of seeds reaches germination. The coefficient of velocity of germination (CVG) (Jones and Sanders, 1987) gives an indication of the rapidity of germination. Its value increases when the number of germinated seeds increases and the time required for germination decreases. Finally, the germination index (GI) (Benech Arnold et al., 1991) is defined as a weighted sum of the daily numbers of germinated seeds. The maximum weight is given to the seeds germinated on the first day, while lesser weights are given to the seeds germinated at later dates, with the lowest weight being assigned to seeds germinated on the last day. The GI therefore emphasizes both the percentage of germination and its speed. A higher GI value denotes a higher percentage and a higher rate of germination.

None of these commonly used indices are able to satisfactorily differentiate between germination patterns and capture all important aspects of the germination process. The same is true for other measures such as the cumulative percentage, which is the static expression of germination behavior at a given time point; LT50, which is the time taken to reach 50% of the final germination level; total germination (TG); and the coefficient of velocity of germination (CVG), which is used as a measure of the rate and time-spread of germination.

The aim of this paper is to compare the above coefficients with a new index that considers seed germination to be a continuous process. The comparison is performed using a functional regression model, which reflects how well the respective coefficients predict the course of germination. Moreover, the approximation of the germination process by splines, which is a necessary step for the construction of the new coefficient, enables the derivation of so-called B-spline coefficients, which can be used for the further statistical processing of germination curves using methods of functional data analysis (Ramsay and Silverman, 2005). This is not possible with any of the previous indices. In this study, the seed germination data are presented as a discrete version of a potentially continuous process and a non-decreasing monotone smoothing spline is developed to approximate them for further purpose of the analysis. The new coefficient, called the continuous germination index (CGI), is introduced, and a functional regression model is proposed to compare the CGI with the other germination coefficients using simulated and real seed germination data.

METHODS

Germination data can be considered from two basic perspectives: as truly discrete data and as a discrete version of a potentially continuous process. In the case of the former, the seed germination coefficients listed above are designed to address this assumption, at least indirectly. The latter approach, further developed here, considers seed germination data as the result of discretization of a (theoretically) continuous function, with the assumption that such a function is able to better capture the specific course of the germination process (Brown and Mayer, 1988a; b; El-Kassaby et al., 2008). When seed germination data are considered as a function, a specific approach for their analysis is needed. Functional data analysis (Ramsay and Silverman, 2005) is aimed at developing such methods for a relevant analysis of data where each observation corresponds to a function. Growing curves represent an important example in functional data analysis, as seen in the seminal book by Ramsay and Silverman (2005, p. 2), which presents an example of growing curves based on the heights of 10 girls measured at 31 ages. The discretization is clearly considered in the x direction (i.e., time), whereas heights are continuous by nature. Similarly, seed germination data can also be considered to be a type of growing curve. In this case, however, a reverse process needs to be considered in the y direction to obtain the resulting functional data; specifically, as there is always a finite number of seeds to germinate, the numbers of germinated seeds are discrete values. However, with an increasing total number of seeds (N*), the approximation by a continuous variable becomes fully reasonable. As a consequence, seed germination can be considered as a continuous process, represented by a positive non-decreasing curve. Of course, in practice only (discrete) germination data are available, which must first be smoothed to obtain an impression of the (theoretical) continuous germination curve.
Splines and the optimal smoothing problem

A mathematically elegant approach to obtain easily interpretable smoothed curves (functions) is to use splines, which are functions defined piecewise by polynomials with a range of possible applications, particularly in computer graphics. Splines can also be used to solve the smoothing problem (De Boor, 1978), i.e., to find a function \( f(x) \) that minimizes the functional

\[
J_f = (1-\alpha) \int \frac{[f^{(i)}(t)]^2}{\alpha} \, dt + \alpha \sum_{i=1}^{n} w_i \left( y_i - f(t_i) \right)^2,
\]

where parameter \( \alpha \in (0,1) \) and data \( (t_i,y_i), i=1,\ldots, n \) are given, and where number \( l \in \{0,1,\ldots,k-1\} \) stands for derivation and \( w_i \geq 0, i=1,\ldots, n \) are given weights. We note that this minimization problem is in fact a compromise between staying close to the given data and obtaining a smooth function. It was shown that the minimizing natural spline \( s_l(t) \), i.e., a spline of odd degree \( k = 2l-1 \), \( l \geq 2 \) that satisfies the additional constraints

\[
s_k^{(l+j)}(t_i) = s_k^{(l+j)}(t_{i-1}) , j = 0, 1, \ldots, l-2,
\]

having as knots the given data \( t_i \). Further details about the approach to identifying the optimal smoothing splines are provided in Appendix 1, along with functionals 2–16.

Nondecreasing optimal smoothing spline

The resulting smoothing spline can have any possible course, and so can also be negative and possibly decreasing on part(s) of its domain. Such spline functions were used for smoothing seed germination data by Hradilová et al. (2019). Although the resulting functions were mostly reasonable, from the methodological perspective it is better to completely avoid the occurrence of the abovementioned situations. Therefore, in the next step the aim was to find an optimal smoothing spline that is nondecreasing over the whole domain, that is, a spline that minimizes functional (2) and satisfies the condition

\[
s_k^l(t) \geq 0 \quad \forall t \in [a,b].
\]

Its detailed construction is available in Appendix 1.

Seed germination coefficient derived from splines

The nondecreasing optimal smoothing spline \( s_k^l(t) \) is a continuous representation of the seed germination data and can be used to analyze the course of the germination. This function reflects well the specific patterns related to germination of a given accession, including the final germination proportion and germination speed. Its advantage is that it assigns a (predicted) number of germinated seeds to each value from the domain \( I = [a,b] \), where \( a \) and \( b \) are real numbers (e.g., \( a < b \)). Accordingly, the number of seeds is, in general, a positive real value and corresponds to a continuous approximation of the original discrete values. For interpretation purposes, it is therefore preferred to consider relative numbers of germinated seeds (e.g., proportions of germinated seeds in given time points as used in the sequel), which are not related to any fixed number of planted seeds at the beginning of the experiment. Due to scale invariance of the spline functions (and the respective B-spline coefficients), this does not affect the quality of the resulting smoothing. It is important to note that the absolute values of the resulting function matter, because not all seeds necessarily germinate during the experiment. Accordingly, although the nonnegativity and monotonicity of \( s_k^l(t) \) might evoke the properties of a probability distribution function, it must not necessarily reach \( 1 \) in \( b \), the right endpoint of \( I \). Hence, Hradilová et al. (2019) called the germination function represented by \( s_k^l(t) \) the absolute germination distribution function (AGDF).

Because AGDF characterizes the complete information about the course of germination, it can also be used to derive the coefficient(s) that could sufficiently characterize the germination process. Hradilová et al. (2019) proposed that the area under AGDF, hereafter termed the CGI, should be considered; this is computed simply as

\[
\text{CGI} = \int_a^b s_k^l(t) \, dt
\]

using a proper numerical integration rule, e.g., the composed trapezoidal rule. Compared with any similar index derived directly from the original (discrete) seed germination data, the advantage of the CGI is that it is able to capture the germination process continuously within the whole domain and thus also better captures the local behavior of germination. Unlike the germination indices discussed above, the CGI assigns the same weight to the whole domain, but in principle any weighting is also possible.

Practical implementation of the CGI

In order to implement the CGI, it is necessary to first approximate the (theoretically continuous) germination curve using a nondecreasing optimal smoothing spline, in this context called the AGDF. The concrete procedure, which is described in detail in Appendix 1, also guarantees the nonnegativity of the resulting function. Moreover, a proper choice of the parameter \( a \in (0,1) \) in (1) enables the user to choose whether to stay close to the given data or rather to obtain a smooth function (here \( a = 0.5 \) represents a good general compromise). The next step is to determine the value of the CGI, which is a single number representing the course of the germination curve. This is achieved by computing the area between the spline function and the \( x \)-axis. Theoretically, this is performed using the definite integral, but in practice an effective numerical integration procedure is employed instead.

Functional regression for the comparison of seed germination coefficients

For a comparison of different seed germination coefficients, the functional regression model with functional response and scalar predictors in the standard space \( L^2(I) \) was utilized. The motivation for this model is very intuitive. The purpose of the seed germination coefficients is to explain the course of the germination curve. In other words, a function is explained by a single number (i.e., the germination coefficient), which could be considered as a real (scalar) predictor variable in a regression model. The standard least squares method cannot be used directly for the estimation of the regression parameters because it assumes a real response; if the response is instead represented by a function, a specific approach is required. For easy interpretability, it is also desirable to make the model as simple as possible; therefore, a linear trend is preferred.

The so-called *function-on-scalar regression* was introduced by Ramsay and Silverman (2005, Chapter 13) and subsequently used in
various contexts (see Talská et al., 2018, for the case of probability density functions). Here, the aim is to predict the AGDF, a theoretical function \( y(t) \) represented by the nondecreasing optimal smoothing spline \( s^*_2(t) \), using any of the germination coefficients introduced above. This spline thus plays the role of the functional response \( y(t) \), and each of the germination coefficients plays the role of the scalar covariate \( x \). The results (presented in the next section) show that the better the obtained fit, the better the respective coefficient characterizes the germination process. Using the functional regression model may represent an elegant way to evaluate the coefficients; although it necessarily considers germination as a continuous process, from its construction the model does not prefer a priori any of the coefficients, working on the basis of discrete germination data.

Specifically, a function-on-scalar regression model is considered, which relates a functional response \( y(t) \) to an independent scalar covariate \( x \) and the intercept. Consider an \( N \)-dimensional vector of functional observations \( y(t) \), a design matrix \( X \) of dimension \( N \times 2 \) (the first column is made of ones) and a two-dimensional vector of unknown functional regression parameters \( \beta(t) \). Let \( \epsilon(t) \) be an \( N \)-dimensional vector of independent identically distributed (functional) random errors with a zero-mean. The functional linear model for the \( i \)th observation \( y_i, i = 1, \ldots, N \), associated with the regressor \( x_i \), is expressed as:

\[
y_i(t) = \hat{\beta}_0(t) + x_i \hat{\beta}_1(t) + \epsilon_i(t), \quad i = 1, \ldots, N,
\]

or, in matrix notation, as \( y(t) = X \hat{\beta}(t) + \epsilon(t) \). The estimators \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) of the regression parameters \( \beta_j, j = 0, 1 \), can be obtained by minimizing the least-square fitting criterion,

\[
SSE(\beta) = \int \left[ y(t) - X \hat{\beta}(t) \right]^T \left[ y(t) - X \hat{\beta}(t) \right] dt.
\]

The smoothness of the resulting estimates is controlled automatically by the smoothness of the observations using the optimal smoothing splines (Talská et al., 2018).

To assess the goodness-of-fit of the model on the observed AGDFs, a pointwise version of the coefficient of determination \( R^2(t) \), \( t \in I \) was computed based on the pointwise comparison between the predicted response and the actual data. Additionally, a global coefficient of determination, denoted by \( R^2_{glob} \), was computed as

\[
R^2_{glob} = \frac{\sum_{i=1}^N \| \tilde{y}_i - \overline{y} \|^2_2}{\sum_{i=1}^N \| y_i - \overline{y} \|^2_2},
\]

where \( \| \cdot \|^2_2 \) stands for the \( L^2 \) norm defined for any function \( g \) from \( L^2 \) space as \( \| g \|_2 = (\int |g(t)|^2 dt)^{1/2} \), and \( \tilde{y}_i \) is the \( i \)th fitted response function and \( \overline{y} \) is the overall mean function \( \overline{y} = \frac{1}{N} \sum_{i=1}^N y_i(t), t \in I \). The latter measures the total sample variance of the \( y_i(t) \) explained by the model in a global sense.

Results

The proposed methodology for the seed germination analysis based on the germination measure CGI defined in (17) was evaluated using both a real germination data set from wild pea (Pisum sativum L. subsp. elatius (M. Bieb.) Asch. & Graebn.) seeds, previously analyzed by Hradilová et al. (2019), and using hypothetical data previously presented by Kader (2005). The comparison of the CGI with the other commonly used germination indices was based on the methodology using functional regression. All computations were performed using R (R Core Development Team, 2019) and its packages splines, nls, and lme4. The code is available upon request from the authors.

Analyzing the germination process using a wild pea seed data set

The data used in this example consisted of 115 wild pea accessions (N) originating from various geographical locations, spanning from the western Mediterranean, through southeastern Europe to the Middle East. A total of 25 seeds per accession were considered (N'), and the number of germinated seeds was scored daily for 28 days (n). These observed numbers of germinated seeds per day were converted into relative values by dividing each number by 25.

Because the absolute:relative values of cumulative seed germination are considered to be generated in a continuous process, the next step was to estimate the associated underlying (hypothetical) germination function for each of the pea accessions. For this purpose, we denote the relative cumulative number of germinated seeds of the \( i \)th accession by \( y_{ij}, i = 1, \ldots, N, j = 0, \ldots, n \) until the day \( t_j = j, j = 0, \ldots, n \). The collected data \( (t_j, y_{ij}) \) were subsequently smoothed using a system of smoothing splines with support \( I = [0,28] \), fulfilling positiveness and monotonicity, as described in the above sections ("Splines and the optimal smoothing problem" and "Nondecreasing optimal smoothing spline"). For all 115 observations, the same strategy was followed to set the values of the input parameters for the smoothing procedure. We employed smoothing splines of degree \( k = 5 \) (and the first derivative in the \( L^2 \) norm, i.e., \( I = 1 \)) with the five equally spaced knots in the days \( \Delta t = \{0,7,14,21,28\} \), the vectors of weights \( w \) for all data equal to vectors of ones, and the smoothing parameter \( \lambda \) was set to 0.5. Thus, when minimizing the functional (2) (see Appendix 1), the smoothness requirement of the resulting approximation of the germination function using smoothing splines has the same importance as the requirement for staying close to the observed data \( (t_j, y_{ij}) \). The resulting spline coefficients (Appendix S1) were then used to obtain the resulting spline functions, expressed as:

\[
s^*_2(t) = \sum_{j=5}^1 b^*_j B^*_j(t), \quad i = 1, \ldots, N, \quad t \in I = [0,28].
\]

Figure 1A displays an example of the cumulative germination data for a subset of five pea accessions together with the smoothed germination curves. The curves reflect the course of the germination process and the related patterns, such as the proportion of germinated seeds or the germination speed. It must be stressed that whereas the maximum germination proportion is attained for the germination process, represented by red, orange, and dark blue curves for which the underlying data represent proportions of germinated seeds in given time points, the speed of germination differs; the red curve represents the fastest germination process (perhaps the fastest possible process), and the orange and dark blue curves represent gradually slower
germination processes. The last two curves, denoted in light blue and green, represent the slower germination processes with the proportion of germinated seeds lower than 50%. This is also well reflected by the germination measure CGI, computed as the area under the germination curve. The CGI values for all 115 accessions are listed in Appendix S1. For the germination processes displayed in Fig. 1A, the following CGI values were calculated: 27.3 (red), 25.6 (orange), 19.5 (dark blue), 8.1 (light blue), and 2.2 (green). Note that, due to the construction of the smoothing spline and the range of values on the y-axis, the CGI cannot exceed the value 28 (see section “Nondecreasing optimal smoothing spline” [above] for more information). A relatively higher value of this germination measure indicates that at least one of the observed parameters, germination proportion or germination speed, is higher than for other accessions. See also Fig. 1B, which illustrates a comparison of the germination measure CGI between the germination processes given by the orange and dark blue curves.

Finally, in order to make a comparison of CGI with other germination measures such as the LT50, TG, MGT, CVG, and GI, we used a functional regression model (18). In particular, the dependency of estimated germination curves \( \gamma_i(t), i=1, \ldots, N \) on the aforementioned germination measures \( x_{im}, m=1, \ldots, M, i=1, \ldots, N \) was modeled by building a total of \( M=8 \) functional regression models of the following formula

\[
y_i(t) = \beta_0^m(t) + x_{im} \beta_1^m(t) + \epsilon_i^m(t), i=1, \ldots, N, m=1, \ldots, M, t \in [0, 28].
\]

Here, the same B-spline basis system for the response \( y(t) \), the regression parameters \( \beta_0^m(t) \) and \( \beta_1^m(t) \), and the error term \( \epsilon(t) \) was considered and functional regression models (22) were reduced to multivariate ones where the B-spline coefficients of the estimated germination curves play the role of the (observed) multivariate response variable. Consequently, the estimation of regression parameters \( \beta_0^m(t) \) and \( \beta_1^m(t) \) via the SSE criterion (19) reduces to an estimation of the matrix of their corresponding B-spline coefficients (Talšká et al., 2018). The output was then used to compute the fitted germination curves \( \hat{y}_i^m(t) \), that is

\[
\hat{y}_i^m(t) = \hat{\beta}_0^m(t) + \hat{x}_{im} \hat{\beta}_1^m(t), i=1, \ldots, N, m=1, \ldots, M, t \in [0, 28].
\]
Analyzing the germination process using a hypothetical data set

The hypothetical germination data set published by Kader (2005) consisted of four seed populations with eight different scenarios. Kader subjected these data to a germination analysis in order to evaluate the various germination measures. The design of the scenarios aimed to cover different time spreads of germination as well as the final proportion of germinated seeds. The germination data based on a germination period of 10 days are presented in Table 2. This study shows that the GI best reflects both the final germination proportion of germinated seeds and the germination speed, including the spread, duration, and “high/low” germination events (for more details, see Kader, 2005).

At this point, we will proceed with a calculation of the CGI. To this end, the relative values of cumulative germination were represented by germination curves. Although a small number of time points make the continuous approximation somewhat difficult, it is still within a range acceptable in a functional data analysis context (Ramsay and Silverman, 2005). Here, the smoothing procedure had the same design as the one applied in the case study dealing with wild pea seeds. It is based on quintic splines \(k = 5, l = 1\) on \(I = [0,10]\) with a smoothing parameter \(\alpha = 0.5\) and the five knots in the days \(\Delta \lambda = \{0, 2, 5, 8, 10\}\). The resulting spline approximation of germination curves, together with the raw relative values of cumulative germination, is plotted in Fig. 3 where, in each scenario A–H, the colors distinguish the individual groups of seed populations. The associated CGI is displayed on Fig. 3 and listed in Table 3, together with the same germination measures that were used in the previous wild pea example. As expected due to its nature as a continuous counterpart to the GI, the CGI describes the germination process in a very similar way. For instance, in scenarios B and E, where the four seed populations began germination on the same day and attained the same TG, the higher values of CGI reflect a higher speed of germination. In

**TABLE 1.** Global coefficient of determination for all germination data and all non-zero germination data.

| Germination measure | All germination data \((N = 115)\) | All non-zero germination data \((N = 94)\) |
|---------------------|----------------------------------|----------------------------------|
| LT50                | 0.994                            | 0.302                            |
| TG                  | 0.824                            | 0.849                            |
| MGT                 | 0.094                            | 0.414                            |
| CVG                 | 0.292                            | 0.297                            |
| GI                  | 0.938                            | 0.933                            |
| CGI                 | 0.956                            | 0.951                            |

Note: CGI = continuous germination index; CVG = coefficient of velocity of germination; GI = germination index; LT50 = time taken to reach 50% of the final germination level; MGT = mean germination time; TG = total germination.
FIGURE 3. Smoothed germination functions via optimal smoothing splines of hypothetical data using relative cumulative totals together with germination coefficient CGI values computed from the spline approximation. Color resolution corresponds to four seed populations for all eight scenarios A–H: red for set 1, green for set 2, blue for set 3, and yellow for set 4. A total of 100 seeds were included in the germination analysis.
TABLE 3. Germination measures for the hypothetical germination data sets from Table 2.

| Measure  | Set 1 | Set 2 | Set 3 | Set 4 | Set 1 | Set 2 | Set 3 | Set 4 | Set 1 | Set 2 | Set 3 | Set 4 | Set 1 | Set 2 | Set 3 | Set 4 | Set 1 | Set 2 | Set 3 | Set 4 | Set 1 | Set 2 | Set 3 | Set 4 |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| LT50     | 3    | 5    | 1    | 4    | 3    | 3    | 2    | 1    | 2    | 2    | 2    | 2    | 3    | 1    | 2    | 2    | 4    | 3    | 2    | 1    | 9    | 7    | 5    | 3    |
| TG       | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.75 | 0.55 | 0.35 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |
| MGT      | 4.8  | 1.0  | 2.5  | 4.1  | 3.0  | 2.4  | 1.9  | 1.5  | 1.9  | 1.9  | 1.9  | 1.9  | 2.6  | 1.3  | 2.0  | 1.9  | 3.9  | 3.0  | 1.9  | 1.9  | 2.6  | 2.6  | 2.6  |
| CVG      | 20.8 | 100.0| 38.7 | 23.1 | 13.3 | 40.0 | 50.0 | 66.6 | 50.0 | 50.0 | 50.0 | 50.0 | 31.6 | 73.0 | 48.7 | 87.0 | 71.0 | 97.1 | 80.0 | 96.0 | 86.0 | 96.0 |
| GI       | 5.447| 8.873| 7.483| 5.862| 7.076| 7.525| 7.979| 8.440| 7.979| 6.313| 4.621| 2.929| 7.437| 8.551| 7.955| 8.048| 6.109| 7.076| 7.979| 8.873| 1.512| 3.387| 5.229|

Note: CGI = continuous germination index; CVG = coefficient of velocity of germination; GI = germination index; LT50 = time taken to reach 50% of the final germination level; MGT = mean germination time; TG = total germination.

In the scenario F, with a uniform distribution of germinated seeds and the same TG, the CGI reflects the time of the first seed germination. A similar situation is represented by scenario H (having the same TG), but in addition to the germination speed, the CGI also accounts for a different distribution of germinated seeds. Table 4 presents a global coefficient of determination based on the functional regression, with missing values indicating a singularity of the respective regression model. Although CGI performed well across the scenarios, it should be noted that the coefficient $R^2$ is not of great value in this case due to the lower number of observations ($N = 4$).

**DISCUSSION**

We have introduced a new germination coefficient, CGI, which was developed by considering germination as a continuous process. As such, the CGI represents a continuous counterpart to the more widely used germination coefficient, the GI. The GI thus naturally forms the competitive alternative to CGI, unlike most other coefficients examined in this study, using both experimental and simulated germination data.

Seed germination is assessed in various biological disciplines in both theoretical and applied fields. Depending on the purpose of the experiments, the type of desired outcome varies. For the seed industry, the final germination percentage is usually a sufficient measure, whereas biological studies require more precise values capturing most aspects of the dynamic germination process (reviewed by Kader, 2005; Ranal and Santana, 2006). Because of these considerations, the use of any given germination data analysis method is prone to misinterpretation if all germination parameters (percentage, speed, and timing) are not taken into account.

Attempts have been made to simplify the characterization of seed germination performance by reducing various germination parameters into a single index (Czabator, 1962); however, reducing multiple germination parameters into one index provides an incomplete picture of germination behavior. Comparisons of the existing commonly used calculation formulae were conducted by Kader (2005), following a previous comparison performed by Brown and Mayer (1988a, b). These studies showed that the GI was the most suitable coefficient for describing the germination percentage and speed relationship because it magnifies the variation among the tested seed lots. Other approaches often use MGT, but this index was shown to be unsuitable for statistical tests (Soltani et al., 2015), which is why the LT50 is often used instead.

Curve-fitting methods using the Weibull, Gompertz, and probit functions have been used to characterize germination, compare seed pretreatments, and measure seed lot differences (reviewed by Brown and Mayer, 1988a, b, or applied further in the context of generalized linear models by Hay et al., 2014). Similarly, El-Kassaby et al. (2008) used a four-parameter Hill function as a curve-fitting method. Yuan et al. (2016) further conducted genetic mapping of seed dormancy/germination using the Hill function. In those approaches, the germination curve was therefore implicitly considered as the result of a continuous germination process.

Despite various assertions, the parameters in these functions do not, however, necessarily lend themselves to simple biological interpretation. Representations of splines with B-spline coefficients, extracted during the calculation of the CGI measure, are simply constructed to best characterize the spline function with respect to a given basis. The purpose of these curve-fitting models differs...
from the other coefficients: they do not aim primarily to yield one value that would characterize the course of the germination curve, but rather to provide its interpretable approximation through easily tractable parameterization.

In this study, we showed that the CGI provides better overall results with respect to the quality of the regression fit than the competing coefficients from Kader (2005), including the GI. The CGI is also able to capture the local effects of the germination process more precisely. This further confirms the applicability of the CGI, which was previously used to test wild pea seeds for dormancy using various experimental setups and conditions (Hradilová et al., 2019). Moreover, the spline approximation of the raw germination data used to compute the CGI also enables the derivation of a finite-dimensional representation of the germination curve using B-spline coefficients. These coefficients can be further used, directly or indirectly, for the statistical processing of a sample of germination curves using functional data analysis methods as an extension of processing single curves using, for example, time-to-event (survival) analyses (McNair et al., 2012). Hradilová et al. (2019) accordingly performed a cluster analysis, but other methods can also be considered, such as functional principal component analyses, regression analyses, or any classification method. The CGI thus not only represents a relevant coefficient for the evaluation of the course of the germination process, but also opens new perspectives for germination data analysis. Moreover, the analysis can also be extended to other biological experiments working with growth parameters, which further expands the potential of the approach presented here.

**ACKNOWLEDGMENTS**

This work was funded by the Grant Agency of the Czech Republic (grant number 16-21053S). The assistance of I. Hradilová with seed germination counting is gratefully acknowledged.

**AUTHOR CONTRIBUTIONS**

R.T. performed the analysis and wrote the manuscript; K.H. and J.M. supervised the data processing and edited the manuscript; and P.S. was involved in the study design, funding acquisition, and manuscript writing and editing. All authors contributed to the discussion and preparation of the final manuscript.

**DATA AVAILABILITY**

The data used in this study were taken from published work by Hradilová et al. (2019).

**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the supporting information tab for this article.

**APPENDIX S1. B-spline coefficients and germination measures.**

**LITERATURE CITED**

Baskin, C. C., and J. M. Baskin. 2014. Seeds: Ecology, biogeography, and evolution of dormancy and germination. Academic Press, San Diego, California, USA.

Benech-Arnold, R. L., M. Fenner, and P. Edwards. 1991. Changes in germinability, ABA content and ABA embryonic sensitivity in developing seeds of *Sorghum bicolor* (L.) Moench. induced by water stress during grain filling. *New Phytologist* 118(2): 339–347.

Bewley, J. D., K. J. Bradford, H. W. M. Hilhorst, and H. Nonogaki. 2013. Seeds: Physiology of development, germination and dormancy, 3rd ed. Springer, New York, New York, USA.

Bradford, K. J. 2002. Applications of hydrothermal time to quantifying and modelling seed germination and dormancy. *Weed Science* 50(2): 248–260.

Brown, R., and D. Mayer. 1988a. Representing cumulative germination. 1. A critical analysis of single-value germination indices. *Annals of Botany* 61: 117–125.

Brown, R., and D. Mayer. 1988b. Representing cumulative germination. 2. The use of Weibull and other empirically derived curves. *Annals of Botany* 61: 127–138.

Burghardt, L. T., B. R. Edwards, and K. Donohue. 2015. Multiple paths to similar germination behavior in *Arabidopsis thaliana*. *New Phytologist* 209(3): 1301–1312.

Czabator, F. 1962. Germination value: An index combining speed and completeness of pine seed germination. *Forest Science* 8: 386–396.

De Boor, C. 1978. A practical guide to splines, vol. 27. Springer-Verlag, New York, New York, USA.

Dierckx, P. 1993. Curve and surface fitting with splines. Oxford University Press, Oxford, United Kingdom.

El-Kassaby, Y., I. Moss, D. Kolotelo, and M. Stoehr. 2008. Seed germination: Mathematical representation and parameters extraction. *Forest Science* 54(2): 220–227.

Fenner, M., and K. Thompson. 2005. The ecology of seeds. Cambridge University Press, Cambridge, United Kingdom.

Hay, F., A. Mead, and M. Bloomberg. 2014. Modelling seed germination in response to continuous variables: Use and limitations of probit analysis and alternative approaches. *Seed Science Research* 24(3): 165–186.

Hradilová, I., M. Duchoslav, J. Brus, V. Pechanec, M. Hýbl, P. Kopecký, L. Smržová, et al. 2019. Variation in wild pea (*Pisum sativum* subsp. *elatius*) seed dormancy and its relationship to the environment and seed coat traits. *PeerJ* 7: e6263.

Jones, K. W., and D. Sanders. 1987. The influence of soaking pepper seed in water or potassium salt solutions on germination at three temperatures. *Journal of Seed Technology* 11(1): 97–102.

Kader, M. 2005. A comparison of seed germination calculation formulae and the associated interpretation of resulting data. *Journal & Proceedings of the Royal Society of New South Wales* 138: 65–75.
Lawson, C. L., and R. J. Hanson. 1995. Solving least squares problems, vol. 15. SIAM, Philadelphia, Pennsylvania, USA.

Machalová, J. 2002a. Optimal interpolating and optimal smoothing spline. Journal of Electrical Engineering 53:79–82.

Machalová, J. 2002b. Optimal interpolatory splines using B-spline representation. Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium, Mathematica 41(1):105–118.

Machalová, J., K. Hron, and G. S. Monti. 2016. Preprocessing of centred logratio transformed density functions using smoothing splines. Journal of Applied Statistics 43(8):1419–1435.

McNair, J., A. Sunkara, and D. Frobish. 2012. How to analyse seed germination data using statistical time-to-event analysis: Non-parametric and semi-parametric methods. Seed Science Research 22(2):77–95.

Orchard, T. 1977. Estimating the parameters of plant seedling emergence. Seed Science and Technology 5(1):61–69.

Probert, R. J. 2000. The role of environment in the regulation of seed dormancy and germination. In M. Fenner [ed.], Seeds: The ecology of regeneration in plant communities, 2nd ed., 261–292, CABI, Wallingford, Oxfordshire, United Kingdom.

Ramsay, J. O., and B. W. Silverman. 2005. Functional data analysis. In J. O. Ramsay and B. W. Silverman [eds.], Encyclopedia of statistics in behavioral science, 2nd ed. Springer, New York, New York, USA.

Ranal, M. A., and D. G. de Santana. 2006. How and why to measure the germination process? Brazilian Journal of Botany 29(1):1–11.

R Core Development Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website https://www.R-project.org/ [accessed 16 May 2020].

Soltani, E., F. Ghaderi-Far, C. C. Baskin, and J. M. Baskin. 2015. Problems with using mean germination time to calculate rate of seed germination. Australian Journal of Botany 63(8):631–635.

Talská, R., A. Menafoglio, J. Machalová, K. Hron, and E. Fišerová. 2018. Compositional regression with functional response. Computational Statistics & Data Analysis 123:66–85.

Weibrecht, K., K. Müller, and G. Leubner-Metzger. 2011. First off the mark: Approximation of germination curves using splines. Applications in Plant Sciences 2020 8(8): e11366 Talská et al.—Functional regression to describe seed germination • © 2020 Talská et al.

APPENDIX 1. Approximation of germination curves using splines.

Optimal smoothing splines
Machalová and colleagues (Machalová, 2002a, b; Machalová et al., 2016) studied optimal smoothing splines under different constraints. In a basic setting, finding an optimal smoothing spline means that for given data (t_i,y_i), a ≤ t_i ≤ b; weights w_i ≥ 0, i = 1, ..., m; sequence of knots Δz = x_0 < x_1 < ... < x_k = b = x_{k+1}, n ≥ g + 1; parameter α ∈ (0, 1) and an arbitrary l ∈ {1, ..., k − 1}, the task is to find a spline s_k(t) ∈ S^Δ_k [a,b] that minimizes the functional

\[ J_l(s_k) = \frac{b-a}{l^2} \left[ \int_{x_0}^{x_l} (s_k(t))^2 \ dt + a \sum_{i=0}^{n} w_i \left| y_i - s_k(t_i) \right| \right]^2 \]  \tag{2}

The choice of parameter l affects the smoothness of the resulting spline. Symbol S^Δ_k [a,b] denotes the vector space of polynomial splines of degree k > 0, defined on a finite interval [a,b] used here to time the interval for which the germination is observed, with the sequence of knots Δz. It is known that \( \dim(S^\Delta_k[a,b]) = g + k + 1 \). Every spline \( s_k(t) \in S^\Delta_k[a,b] \) has a unique representation

\[ s_k(t) = \sum_{i=k}^{g} b_i B_i^{k+1}(t), \]  \tag{3}

with vector \( b = (b_{-k}, ..., b_{-1})^T \) of the B-spline coefficients. Functions \( B_i^{k+1}(t) \), \( i = -k, ..., g \) are B-splines of degree k and form a basis in \( S^\Delta_k[a,b] \). Using matrix notation, the spline \( s_k(t) \) can be written as

\[ s_k(t) = C_{k+1}(t) b, \]  \tag{4}

where \( C_{k+1}(t) = (B_i^{k+1}(t), ..., B_g^{k+1}(t))^T \) is the so-called collocation matrix. If we denote \( t \equiv (t_1, ..., t_n)^T \), \( y = (y_1, ..., y_n)^T \), \( w = (w_1, ..., w_n)^T \), and \( W = \text{diag}(w) \), the functional \( J_l(s_k) \) can be written in a matrix form as

\[ J_l(b) = (1-a) b^T N_l a + b^T \left[ y - C_{k+1}(t) b \right]^T W \left[ y - C_{k+1}(t) b \right], \]  \tag{5}

where

\[ C_{k+1}(t) = \begin{pmatrix} B_{k+1}^{k+1}(t_1) & ... & B_{k+1}^{k+1}(t_1) \\ \vdots & \ddots & \vdots \\ B_{k+1}^{k+1}(t_n) & ... & B_{k+1}^{k+1}(t_n) \end{pmatrix} \in \mathbb{R}^{n \times g+k+1}, \]  \tag{6}

matrix \( N_l = S_l^T M_l S_l \) is positive semidefinite with

\[ M_l = \begin{pmatrix} (B_{k+1}^{k+1}(t), ..., B_{g+1}^{k+1}(t))^T & ... & (B_{k+1}^{k+1}(t), ..., B_{g+1}^{k+1}(t))^T \\ \vdots & \ddots & \vdots \\ (B_{k+1}^{k+1}(t), ..., B_{g+1}^{k+1}(t))^T & ... & (B_{k+1}^{k+1}(t), ..., B_{g+1}^{k+1}(t))^T \end{pmatrix} \in \mathbb{R}^{(g+k+1)^2} \]

and

\[ (B_{k+1}^{k+1}(t), ..., B_{g+1}^{k+1}(t))^T dt \]

stands for the scalar product of B-splines in \( L^2([a,b]) \) space. The matrix \( M_{l+1} \) is positive definite, because \( B_i^{k+1+l} \geq 0, i = -k-l, ..., g \) are basis functions. Furthermore, matrix \( S_l \equiv D_1 \ L_1 \ ... \ D_l \ L_l \in \mathbb{R}^{(g+k+1)^2} \) is an upper triangular matrix that has full row rank. \( D_j \in \mathbb{R}^{g+k+1-j \times g+k+1-j} \) is a diagonal matrix, such that

\[ D_j = (k+1-1) \ \text{diag}(d_{-k-j}, ..., d_{-1}) \]  \tag{7}

with

\[ d_j = \frac{1}{\lambda_{k+1-j} - \lambda_i} \]  \tag{8}

and

\[ L_j = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{g+1-j \times g+1-j} \]  \tag{8}

Because the derivative of the spline can be expressed as a spline of a lower degree (De Boor, 1978; Dierckx, 1993), we can easily
find sufficient conditions for the nondecreasing spline. In matrix notation, we have

$$s'_k(t) = C_k(t) D_1 L_1 b,$$

where $c_{i,n}(a_{i,1}, \ldots, a_{i,n})$. According to the positivity of the B-splines, $b^j(t) \geq 0$ (De Boor, 1978), formulae (7) and (8), it follows that if $L_1 b \geq 0$, $s_k(t)$ is a nondecreasing spline. Finally, the above considerations imply that a nondecreasing optimal smoothing spline is a spline whose vector of the B-spline coefficients $b$ minimizes function $J_f(b)$, (5), and satisfies the condition $L_1 b \geq 0$. The function $J_f(b)$ is in fact a quadratic function, which can be written as

$$J_f(b) = b^T H b - 2 a b^T C^T_{k+1}(t) W y + y^T W y.$$

The matrix

$$H := (1-a) N_{kl} + a C^T_{k+1}(t) W C_{k+1}(t)$$

is regular and positive definite; therefore, there exists an upper triangular matrix $T$ such that $T^T T = H$. The last term in $J_f(b)$ is constant, so it can be omitted for minimization. This leads to a problem of quadratic programming with the linear inequality constraints

$$\min b^T T^T T b - 2 a b^T C^T_{k+1}(t) W y$$

subject to

$$L_1 b \geq 0.$$  \hspace{1cm} (9)

Let

$$z := a (T^T)^{-1} C^T_{k+1}(t) W y,$$

then

$$||T b - z||_2^2 = b^T T^T T b - 2 a b^T C^T_{k+1}(t) W y + z^T z,$$

where $||a||_2 = a^T a$. The last term is constant, so it does not affect the minimization process. By denoting

$$u := T b - z$$

the so-called least distance programming problem (Lawson and Hanson, 1995) is obtained in the form of

$$\min ||u||_2^2,$$  \hspace{1cm} (11)

where

$$G := L_1 T^{-1}, \ h := -L_1 T^{-1} z$$

because from the definition of $u$ we have

$$u + z = T b \Rightarrow b = T^{-1} (u + z) \Rightarrow L_1 b = L_1 T^{-1} u + L_1 T^{-1} z. $$

Hence, it is clear that

$$L_1 b \geq 0 \Leftrightarrow L_1 T^{-1} u \geq -L_1 T^{-1} z.$$

Accordingly, if $u^*$ is the solution of the least distance programming problem (11), (12), then

$$b^* = T^{-1} (u^* + z)$$

is the solution of problem (9), (10). Consequently, the resulting non-decreasing optimal smoothing spline is obtained by the formula

$$s_k^*(t) = \sum_{i=-k}^{g} b^*_i B^{i+1}_k(t),$$

where the vector $b^* = (b^*_1, \ldots, b^*_g)^T$ is given in (13).

In order to obtain a positive nondecreasing optimal smoothing spline, a further condition

$$s_k^*(t) \geq 0$$

needs to be added. Due to the nonnegativity of B-splines, it is known that if $b \geq 0$ then spline $s_k(t) = C_{k+1}(t) b$ is nonnegative. Hence, a condition $b \geq 0$ is added to the optimization problem (9), (10) as an additional constraint. Finally, the optimization problem is given as

$$\min b^T T^T T b - 2 a b^T C^T_{k+1}(t) W y$$

subject to

$$A b \geq 0,$$  \hspace{1cm} (16)

with matrix

$$A = \begin{pmatrix} L_1 \\ I \end{pmatrix}, \ I \in \mathbb{R}^{k+1}.$$

The next steps are the same as for a nondecreasing optimal smoothing spline with the only difference that

$$G := AT^{-1}, \ h := -AT^{-1} z.$$