Modeling seed dormancy release and germination for predicting *Avena fatua* L. field emergence: A genetic algorithm approach

Aníbal M. Blanco [a,*], Guillermo R. Chantre [b], Mariela V. Lodovichi [b], J. Alberto Bandoni [a], Ricardo L. López [c], Mario R. Vigna [c], Ramón Gigón [d], Mario R. Sabbatini [b]

[a] *Planta Piloto de Ingeniería Química, Universidad Nacional del Sur/CONICET, Bahía Blanca, Buenos Aires 8000, Argentina*
[b] *Departamento de Agronomía/CERZOS, Universidad Nacional del Sur/CONICET, Bahía Blanca, Buenos Aires 8000, Argentina*
[c] *EEA INTA Bordenave, Bordenave, Buenos Aires 8187, Argentina*
[d] *CEI-INTA Barrow (B7500WAA), Tres Arroyos, Buenos Aires, Argentina*

ABSTRACT

Avena fatua is a cosmopolite weed species which produces severe yield losses in small-grain production systems in temperate and semiarid climates. In the semiarid region of Argentina, *A. fatua* field emergence patterns show great year-to-year variability mainly due to the effect of highly unpredictable precipitation regimes as well as a complex seedbank dormancy behavior regulated by both, genetic and environmental factors. Previously developed models for the same agroecological system based on Non-Linear Regression techniques (NLR) and Artificial Neural Networks (ANN) were either unable to accurately predict field emergence or lacked explanatory power. The main objective of the present work is to develop a simple (i.e. parsimonious) model for *A. fatua* field emergence prediction for the semiarid region of Argentina based on the disaggregation of the dormancy release phase from the germination/pre-emergence growth processes, using easy accessible soil microclimate derived indices as input variables and observed cumulative field emergence data as output variable. The parsimony and predictive capability of the newly developed model were compared with NLR and ANN approaches developed by the same authors for the same agroecological system. Specifically, dormancy release was modeled as a logistic function of an after-ripening thermal-time index while germination/pre-emergence growth was represented by a logistic distribution of hydrothermal-time accumulation. A total of 528 input/output data pairs corresponding to 11 years of data collection were used in this study. Due to its implementation simplicity and good convergence features, a Genetic Algorithm (GA) was adopted to solve the resulting optimization problem consisting on the minimization of the Mean Square Error (*MSE*) between training data and experimentally obtained field emergence data. The newly developed GA based approach resulted in a significantly more parsimonious model (*BIC* = −1.54) compared to ANN (*BIC* = −1.36) and NLR (*BIC* = −1.32) models. Model evaluation with independent data also showed a better predictive capacity of the GA approach (*RMSE* = 0.07) compared to NLR (*RMSE* = 0.19) and ANN (*RMSE* = 0.11) alternatives. These outcomes suggest the potential applicability of the proposed predictive tool in weed management decision support systems design.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Avena fatua L. (wild oat) is a cosmopolite weed species which produces severe yield losses in small-grain production systems in temperate and semiarid climates (Li and Foley, 1997; Scursoni et al., 2011). *A. fatua* field emergence patterns observed around the world are particularly interesting for weed scientists due to their challenging features from a forecasting perspective. For example, under highly variable soil thermal and moisture conditions, *A. fatua* shows multiple annual cohorts generating very irregular and time-distributed field emergence patterns, as those observed in semiarid southern regions of Argentina (Chantre et al., 2012). Conversely, in temperate environments with non-severe soil moisture limitations during the emergence period, regular and concentrated cumulative emergence curves are observed (Chauhan et al., 2006; Page et al., 2006; Martinson et al., 2007; Chantre et al., 2013).

Such a behavior suggests that wild oat might have developed specific ecological adaptations to different regional climatic conditions basically modulated by the seed dormancy phenomenon (Chantre et al., 2013). Naylor and Fedec (1978) suggested that *Avena fatua* genotype-by-germination temperature interaction has an adaptative ecological role that allows wild oat to colonize diverse ecosystems. Genetic adaptations of natural populations of *A. fatua* affecting duration of primary dormancy were previously

* Corresponding author. Tel.: +54 291 4861700; fax: +54 291 4861600.
E-mail address: ablanco@plapiqui.edu.ar (A.M. Blanco).

documented in the literature (Naylor and Jana, 1976; Jana and Naylor, 1980; Adkins et al., 1986). In this regard, Fennimore et al. (1999) proposed a genetic model to explain seed dormancy variability of *A. fatua* via selection of genes at three loci.

Seed dormancy reveals itself as a crucial mechanism to avoid fatal germination assuring seedling survival under highly unpredictable climatic environments. As stated by Grundy (2003), dormancy is the single most important feature of weed seedbank dynamics and periodicity. Soil temperature is the major environmental factor governing changes in the degree of dormancy in temperate environments (Benech-Arnold et al., 2000). For the case of *Avena fatua*, as for most winter annual weed species, a given after-ripening thermal-time requirement must be fulfilled for primary dormancy loss to occur (Bewley and Black, 1994; Baskin and Baskin, 1998). High temperatures during seed development in the mother plants have also proven to increase germinability of recently harvested wild oat seeds (Sawhney and Naylor, 1980; Peters, 1982). Conversely, for seeds induced into secondary dormancy by anaerobic conditions, increments in dormancy levels with temperature were observed (Symons et al., 1986, 1987; Foley and Fennimore, 1998).

Due to the dormancy characteristics of *A. fatua* described above, the estimation of field emergence is a challenging task. Both, mechanistic and empirical based models have been proposed by weed scientists in order to forecast emergence of annual species (Grundy, 2003).

Reductionist modeling approaches aim to provide a close description of the basic ecophysiological processes underlying weed emergence (i.e. dormancy, germination and pre-emergence growth). Many mechanistic models based on the description of the biophysical processes have been proposed (Vleeshouwers and Kropff, 2000; Colbach et al., 2005; Gardarin et al., 2012a). Although these models are desirable from a process-based knowledge acquisition perspective, they require a large amount of often unavailable or difficult to gather experimental information to be developed and validated. As remarked by Grundy (2003), the complexity of reductionist models is associated to over-parameterization and the reliance on large and complicated input data sets.

Empirical weather-based models seek to identify correlations between environmental variables and observed emergence patterns. Such models provide a simpler and less complex modeling framework compared to reductionist models, thus becoming "easily adoptable" to support decisions such as spraying timing and type of herbicide to use. However, an important weakness of empirical models is associated to the complexity of the interactions between weather variables that sometimes can produce considerable under or over-estimation of weed emergence, as highlighted by Grundy (2003).

Soil microclimate derived indices such as hydrothermal-time (θ_{HT}) and thermal-time (θ_T) have been used to quantify the effect of both, soil temperature (T) and soil water potential (Ψ) on weed field emergence. They assume that emergence rates are proportional to the amount by which soil temperature and soil water potential exceed a given threshold value (Bradford, 2002). These thresholds are basically the base or minimum temperature (T_b) and the base water potential (Ψ_b) for germination. Generally, these values are obtained from germination laboratory studies and in many cases without explicitly considering the dormancy status of seed population, which is expected to alter such cardinal values. However, as stated by Bradford (2002), the implementation of the hydrothermal model remains a valid approach for integrating the effects of both T and Ψ toward germination under the assumption of T_b and Ψ_b to be constant.

Non-Linear Regression (NLR) sigmoidal shape models (Weibull, Gompertz, Logistic, etc.), based on θ_{HT} and θ_T indices are by far the most widely adopted empirical approach for weed emergence

prediction in the field (Forcella, 1998; Schutte et al., 2008; Haj Seyed-Hadi and González-Andújar, 2009; Bullied et al., 2012; Zambrano-Navea et al., 2013). Specifically for *A. fatua*, different NLR models have been developed (Page, 2004; Page et al., 2006; Martinson et al., 2007; Moschini et al., 2009). The major limitation of NLR models is that they are sometimes not flexible enough to capture complex features in the hydrothermal-time distribution, as pointed out in Cao et al. (2011). This situation stems from their univariate nature and fixed underlying non-linearity.

In order to cope with such limitations, Artificial Neural Networks (ANN) have been recently developed by our work team for estimating weed emergence under field conditions (Chantre et al., 2012, 2013). ANN are flexible frameworks based on functional relations which admit any number of explanatory variables and are trained with a limited amount of data. Such empirical models are known for their capacity to describe highly non-linear relationships among variables thus showing a high potential applicability in ecological systems (Lek and Guégan, 1999). However, despite the acceptable predictive outcome of the developed ANN models, the approach remains a "black-box" with limited explanatory power (Olden and Jackson, 2002).

Regardless of the implemented approach, most of the proposed empirical models share the same major simplification which consists in visualizing the emergence process as "a-whole", and regressing observed cumulative emergence data as a function of a single hydrothermal-time index (NLR models) or as a combination of thermal-time and hydro-time indices (ANN models).

Colbach et al. (2005) highlighted the importance of avoiding "black-box" models by disaggregating the ecophysiological components of the emergence process. Thus, based on the importance of seed dormancy as an adaptative ecological feature in *A. fatua*, we hypothesize that the disaggregation of the dormancy release phase from the germination/pre-emergence growth processes would allow increasing the parsimony as well as the prediction capability of the model compared to previously developed NLR and ANN approaches for the same agroecological system (see Chantre et al., 2012).

In order to quantify the effect of temperature and after-ripening time on wild oat seed population dormancy status, an after-ripening thermal-time index is implemented within a distribution function. Index θ_{AT} allows quantifying the thermal-time accumulation above a 'base' temperature required by a seedbank fraction to lose its dormant condition. Many examples of this approach for seed dormancy release quantification in winter annual species can be found in the literature (Bauer et al., 1998; Bair et al., 2006; Gianinetti and Cohn, 2007; Chantre et al., 2009, 2010). Similarly, the germination/post-germination phase is modeled according to another distribution which adopts a hydrothermal-time index as explanatory variable, since soil water availability is a strong conditional for germination (Bradford, 1990, 2002). Once a given seedbank fraction accumulates enough after-ripening thermal-time followed by the required hydrothermal-time accumulation for germination and pre-emergence growth, seedlings emergence occurs.

In order to estimate the involved parameters, a parameter estimation problem has to be solved. Due to its implementation simplicity and good convergence features, Genetic Algorithms (GA) have been adopted lately to solve mean squares optimization problems. Genetic algorithms are stochastic optimization techniques based on the evolution of sets of potential solutions according to certain rules inspired in the natural selection process (Michalewicz, 1996). These techniques have demonstrated good global convergence properties in non-linear, non-constrained models with multiple local optima in many fields of science and engineering (Rangaiah, 2010). They have also been used for weed seedling emergence data fitting in Haj Seyed-Hadi and González-Andújar (2009).

The objectives of the present work were to: (i) develop a simple (i.e. parsimonious) model for *A. fatua* field emergence prediction in the semiarid region of Argentina based on the disaggregation of the dormancy release phase from the germination/pre-emergence growth processes, using easy accessible soil microclimate derived indices as input variables and observed cumulative field emergence data as output; (ii) investigate the applicability of GA as optimization tool to solve the proposed parameter estimation problem; (iii) compare the parsimony and predictive capability of the newly developed model with NLR and ANN approaches developed by the same authors for the same agroecological system.

2. Materials and methods

2.1. Field experimental data

Avena fatua emergence data were collected from years 2000 to 2010 at the experimental field of EEA INTA Bordenave (37°50′ S; 63°01′ W), located in the south-west area of the semiarid region of Argentina. The experiment was conducted on an undisturbed field with a high natural infestation of wild oat without crop presence. Seedling counting (destructive method) was performed at weekly intervals on three quadrats (1 m^2 each) randomly distributed on the field. At the end of each year of experimentation, the quadrats were removed and newly distributed at random into the remaining area of the 5 ha. experimental field. This way, the seedbank of a particular year was not disturbed by the elimination of seedlings during the counting process of the previous years.

2.2. Training and test sets

From the 11 years of experimentation, obtained data representing cumulative annual field emergence figures ($N = 528$) were divided into a train set (9 years, 82%) and a test set (2 years, 18%). In this way, the independently generated yearly data sets were used for parameter estimation and evaluation purposes. It should be mentioned that, although the environmental conditions of the different years were quite diverse, most of the data pool were characterized by moderate to extreme soil moisture availability limitations for wild oat germination and emergence. Thus, the train set was chosen to include a wide range of precipitation scenarios. Model validation was performed on years 2006 and 2008 (test sets) representing severe and moderate field soil moisture conditions during the emergence period, respectively.

2.3. Estimation of soil temperature and soil water potential

The Soil Temperature and Moisture Model (STM^2) developed by USDA-ARS was used to estimate soil microclimate conditions (Spokas and Forcella, 2009). STM^2 calculates soil moisture and temperature based on soil composition and daily minimum and maximum air temperature and precipitation. The model was originally tested for many global sites in Spokas and Forcella (2009) and further validated in both temperate and Mediterranean type climates (Schutte et al., 2008; Masin et al., 2012). Specifically, for the region under study, STM^2 outcomes were validated, showing a satisfactory agreement between predicted and observed experimental data (Damiano et al., 2010).

STM^2 was calibrated using soil site-specific parameters: soil texture (sandy loam = 53% sand, 31% silt, 16% clay), organic matter content (3.1%) and bulk density (1.2 Mg/m^3). Daily mean soil temperature (T) and water potential (Ψ) at 1 cm burial depth were estimated using weather data registered at a meteorological station located in the experimental field. In this work, 1 cm was considered to be a representative seed burial depth for a non-tillage field condition (direct-drilled) since previous studies indicate that most

of the seedbank is located within the first 0–5 cm of the soil profile for untilled *A. fatua* field systems (Ghersa and Martínez-Ghersa, 2000). After natural seed dispersal, wild oat plants were mechanically removed and the stubble was used as soil surface coverage. Residue cover represented by '% shade' in STM^2 was set as 100%.

2.4. Model elements

2.4.1. Dormancy release

In order to quantify the effect of after-ripening time and temperature on *A. fatua* seedbank dormancy release process, an after-ripening thermal-time index (θ_{AT}) was implemented. The accumulation of after-ripening thermal-time units was defined according to the following equation:

$$\theta_{AT_n} = \sum_{i=1,n} \theta_{A_i} \tag{1a}$$

where

$$\theta_{Ai} = \begin{cases} T_i - T_l & \text{if } T_l < T_i \\ 0 & \text{otherwise} \end{cases} \tag{1b}$$

θ_{ATn} is the after-ripening thermal-time at day n, T_i is the estimated mean daily soil temperature during the after-ripening time period, and T_l is the base temperature below which after-ripening does not occur.

2.4.2. Germination/pre-emergence growth

Most univariate weed emergence models have adopted hydrothermal-time as explanatory variable since both T and Ψ have proven to be critical variables for seedling emergence. Hydrothermal-time based models have shown to be more accurate than thermal-time approaches under restrictive soil moisture availability conditions during the emergence period of *A. fatua* (Martinson et al., 2007). Therefore, an index that combines both T and Ψ is necessary for the development of univariate models under the environmental conditions which prevail in semiarid regions. In this work, the following hydrothermal-time definition was adopted (Bradford, 2002; Martinson et al., 2007) to model germination in the proposed emergence model:

$$\theta_{HTG_n} = \sum_{i=1,n} \theta_i H_i \tag{2a}$$

where

$$\theta_i = \begin{cases} T_i - T_b & \text{if } T_i > T_b \\ 0 & \text{otherwise} \end{cases} \tag{2b}$$

$$H_i = \begin{cases} 1 & \text{if } \Psi_i > \Psi_b \\ 0 & \text{otherwise} \end{cases} \tag{2c}$$

T_i is the estimated mean daily soil temperature and T_b is the corresponding base temperature for germination. Similarly Ψ_i is the daily soil water potential and Ψ_b is the corresponding base water potential for germination. Notice that the definitions of thermal-time, hydro-time and its combined expression (i.e. hydrothermal-time) are functions of the cardinal parameters (T_l, T_b and Ψ_b). In this contribution, such parameters were assumed to be constant and fixed in values obtained from the literature. Specifically, $T_b = 1$ °C (Cousens et al., 1992) and $\Psi_b = -1.2$ MPa (Page, 2004). The calculation of after-ripening thermal-time accumulation was based on $T_l = 5$ °C (Lodovichi, unpublished).

It should be acknowledged that in theory, pre-emergence growth could have been modeled independently as an additional distribution function. However, due to the superficial location of

the seedbank, the post-germination phase is not considered a bottleneck for emergence in the system under study. Therefore, we opted not to model pre-emergence growth as an additional logistic function depending on some thermal accumulation index, since this would have introduced several additional adjustable parameters leading to a less parsimonious model. Rather, we preferred to integrate both: the germination and the post-germination subprocesses into one logistic function in terms of hydrothermal-time as explanatory variable.

2.4.3. Distribution functions

Logistic functions (Eq. (3)) were adopted to model separately seedbank dormancy release and germination/pre-emergence growth

$$Y = \frac{\gamma}{1 + \exp(-\delta(x - \lambda))} \tag{3}$$

Y is the accumulated quantity (in proportion), x is the explanatory variable (θ_{AT} for dormancy release; θ_{HTG} for germination and pre-emergence growth) and γ, δ and λ are model parameters.

It should be mentioned that many other distribution functions might be used in modeling these processes, as suggested by Mesgaran et al. (2013). Logistic distributions have the advantage of providing a simple explicit expression for the accumulated quantity (Eq. (3)), compared for example with the normal distribution which involves the calculation of an integral with the associated computational burden. However, any distribution function could be implemented within the proposed approach with minimum programming modifications.

2.5. Model structure

As previously mentioned, the proposed model disaggregates wild oat seedling emergence into seed dormancy release (i.e. after-ripening process) and germination/pre-emergence growth processes, considering them as independent phenomena that take place in series. Model assumptions are that each fraction of the seedbank requires a certain amount of after-ripening thermal-time accumulation for dormancy release to occur. Thereafter, the different fractions of the population must accumulate a given amount of hydrothermal-time to allow germination and pre-emergence growth, finally leading to emergence.

Specifically, each fraction of the seedbank ($AcDr_n$) accumulates a certain amount of after-ripening thermal-time for dormancy release (θ_{AT_n}) according to a logistic distribution (Eq. (4)). Following, each 'non-dormant' fraction of the seedbank has to accumulate a certain amount of hydrothermal-time (θ_{HTG_m}) distributed according to a logistic law (Eq. (5)) to achieve germination and pre-emergence growth (AcG_m) in order to finally emerge.

$$AcDr_n = \frac{1}{1 + \exp(-\delta_{dr}(\theta_{AT_n} - \lambda_{dr}))} \tag{4}$$

$$AcG_m = \frac{1}{1 + \exp(-\delta_g(\theta_{HTG_m} - \lambda_g))} \tag{5}$$

January 1st is arbitrarily chosen for after-ripening thermal time accumulation. There is experimental evidence that in this date the seedbank is established with the contribution of the plants of the previous season (seed rain), which in the southern hemisphere takes place during the last months of the year (from mid-November to late December).

In Fig. 1 the described logic is schematically depicted. Each day, a specific seedbank fraction loses its dormant condition due to after-ripening thermal-time accumulation. The accumulated dormancy release distribution (described by Eq. (4)) is obtained by integrating such fractions along the year. Moreover, each specific non-dormant
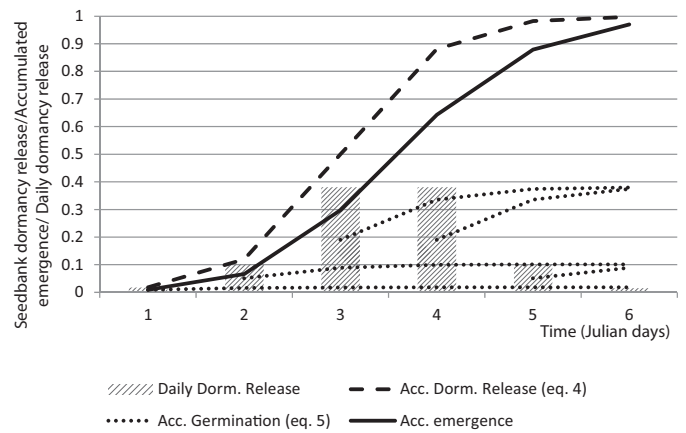

Fig. 1. Schematic representation of the proposed emergence model. By integrating daily dormancy release (bars) accumulated dormancy release is obtained (dashed line). Dotted lines represent accumulated germination/pre-emergence growth of each non-dormant fraction. Note that as time increases, the dotted line reaches the height of the bar, meaning that each non-dormant seedbank fraction germinates along the following days according a specific distribution. Accumulated emergence (solid line) is obtained by integrating each day the germinated/post-germinated fraction corresponding to each non-dormant seedbank fraction.

fraction undergoes the germination/post-germination process by accumulating hydrothermal-time according to Eq. (5). Finally, accumulated emergence is obtained by integrating each day the germinated/post-germinated fractions corresponding to the different portions of the seedbank that lost dormancy during the previous days.

Parameters γ_{dr} and γ_g were set to a value of 1 in Eqs. (4) and (5), respectively, since they represent the asymptotic value of the distribution which in both cases is 100%. The remaining parameters (δ_{dr}, λ_{dr}, δ_g and λ_g) were estimated in the parameter estimation problem by allowing them to move within the following intervals: $\delta = [0, 0.1]$, $\lambda = [50, 2000]$. These intervals correspond to broad bounds around promising solutions (low mean square errors), identified in previous computational experiments.

2.6. Genetic algorithms

Parameter estimation problems are special cases of optimization problems. In parameter estimation problems the model parameters are the optimization variables and Mean Square Error type functions are usually adopted as objective functions. A distinctive feature of parameter estimation problems is the existence of a large number of local minima. Stochastic optimization algorithms have shown a good performance in parameter estimation problems over their deterministic counterparts (Moles et al., 2003). Their most attractive feature is a good balance between exploration and exploitation of the search space which increases the chance of convergence to global optima. Moreover, opposite to deterministic algorithms where derivatives have to be provided, stochastic techniques only use objective function values in the calculation, which facilitates their computational implementation and broadens their scope of application. Therefore, only a function that returns the simulation results of the process to be optimized is required. This feature provides a very flexible framework to experiment with arbitrary descriptive functions (nonlinear, discontinuous, etc.) with any number of variables and parameters.

In particular, genetic algorithms (GA) are stochastic optimization techniques based on the evolution of populations of solutions according to operations inspired in the natural selection process (Michalewicz, 1996). The involved operations in GAs are basically selection, crossover and mutation. From an initial set of randomly generated individuals (i.e. vectors of optimization variables), a
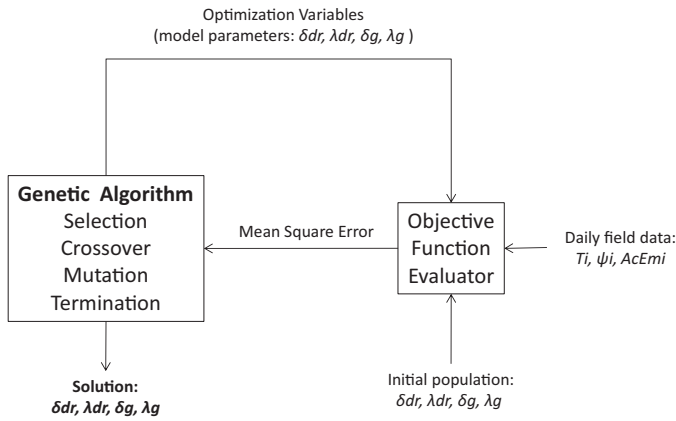
Fig. 2. Genetic algorithm based parameter estimation scheme.

1. Read population of optimization variables (δ_{dr}, λ_{dr}, δ_g and λ_g)

2. Read train sets (T_i, Ψ_i, $AcEm_i$)

3. For each individual of the population (i.e. vectors of optimization variables):

    a. Calculate after-ripening thermal-time and hydrothermal-time indices (θ_{ATn} and θ_{HTm})

    b. Calculate daily (n=2 to 365) seedbank dormancy release from Eq. (4) ($DR_n = AcDr_n - AcDr_{n-1}$)

    c. Calculate daily (m=n to 365) germination (Eq. 5) for each non-dormant seedbank fraction as: $DG_{n,m} = DR_n(AcG_m - AcG_{m-1})$

    d. Calculate daily (i=1 to 365) emergence as: $AcEm_i = \sum_{j \leq i} \sum_{j \leq k \leq i} DG_{j,k}$

    e. Calculate MSE of the training set

4. Return MSE of each individual to the GA

Fig. 3. Objective function evaluation.

subset of the best is selected according to their fitness value (objective function). Then, randomly chosen pairs of such selected individuals are mathematically combined (crossed) in order to generate a new population of solutions (offspring). The crossover operation emulates the exchange of genetic material between homologous chromosomes that results in recombinant chromosomes. Finally, random modifications are introduced on some individuals (mutations) in order to provide diversity. This procedure is applied repeatedly on each population until some termination criterion is fulfilled, for example, a maximum number of iterations (generations) is met or the difference between the best individuals of two successive iterations reaches a certain threshold value. There exist many different alternatives for each genetic operator and each one has a specific set of tunable options. A complete description of the different versions of genetic algorithms is beyond the scope of this section. The interested reader is referred, for example, to Michalewicz (1996), Rangaiah (2010) and Matlab (2013) for details on the methodology and implementation issues.

2.7. Parameter estimation problem

In Fig. 2 a schematic representation of the implemented parameter estimation problem based on a GA is presented. The Genetic Algorithm block implements the operations performed on the populations of optimization variables (selection, crossover and mutation) and also checks for the termination criterion. The Objective Function Evaluator block predicts accumulated emergence for each individual and computes the corresponding MSE value.

The GA option of the optimization tool of the MATLAB platform (Matlab, 2013), was used to solve the optimization problem. The termination criterion corresponds to whatever happens first: (i) maximum number of generations (250) reached or (ii) weighted average change in the fitness function value over stall generations (50) less than function tolerance (1E−6).

The Objective Function Evaluator was programmed as a function described by the pseudo-code of Fig. 3. In step 1 the population of parameter values is read. The initial population is created by randomly initializing each individual between their corresponding bounds. In the subsequent iterations the population is generated by the genetic algorithm. The training data sets are passed in step 2 as a 27-column matrix which provides daily mean temperature, daily water potential and daily accumulated emergence (T_i, Ψ_i and $AcEm_i$) for each one of the nine training years. This matrix is read by the provided function to calculate the thermal and hydrothermal-time indices in step 3a (Eqs. (1) and (2)). In step 3b daily dormancy release is evaluated from Eq. (4) by subtracting to the 'non-dormant' seedbank fraction of day n, the fraction corresponding to day $n-1$. This can be done since the parameters of

Eq. (4) and the daily thermal accumulation index are available from steps 1 and 3a, respectively. In a similar fashion, daily germination is evaluated for each fraction of non-dormant seedbank from Eq. (5) in step 3c. It should be stressed that germination is calculated for each non-dormant fraction of the seedbank as a distribution over the hydrothermal-time index, starting the day in which such fraction loses its dormant condition. Following (step 3d), accumulated emergence is calculated each day by integrating the seed fractions germinated that day. Such germinated seedlings correspond to different fractions of the seedbank that lost dormancy in the previous days. Finally, the objective function (MSE) is calculated for each individual from the estimated emergence and the experimental data from step 2 and returned to the GA block to proceed with the optimization (step 4).

2.8. Model analysis

Goodness-of-fit measures were based on the Root Mean Square Error (RMSE) of the training set. The predictive capability of the developed models was based on the RMSE of the test set. Comparisons between models were based on the Bayesian Information Criterion (BIC) adopting the definition proposed by De Gooijer and Kumar (1992):

$$BIC = \log(MSE) + \frac{m^d \log(N)}{N} \qquad (6)$$

where m is the number of parameters, N is the number of observations, d is a user defined constant which allows the tuning of the penalty term. Granger (1993) proposes $d > 1$ for nonlinear models. We selected 1 and 1.5 for evaluating purposes in this study.

In order to evaluate the performance of the proposed emergence model, a comparison with previous developed models is provided. Specifically, in Chantre et al. (2012), several univariate and bivariate models were developed with the same experimental data utilized in this work. In all cases, accumulated emergence (AcEm) was adopted as output variable. Alternative definitions for the explanatory indices than those used here were considered. Specifically, the following univariate models were developed in Chantre et al. (2012): $AcEm = \text{Weibull}(\theta_{HT})$, $AcEm = \text{Logistic}(\theta_{HT})$ and $AcEm = \text{ANN}_{hn=1,5}(\theta_{HT})$. ANN stands for Artificial Neural Network and hn represents the number of neurons in the hidden layer. Additionally, several bivariate models: $AcEm = \text{ANN}_{hn}(\theta_T, \theta_H)$ ($hn = 1, 2, 3, 5, 6, 7$) were also developed. In all cases θ_{HT} stand for hydro-thermal time and θ_T and θ_H represent thermal-time and hydrothermal-time, respectively. For details on the specifics see Chantre et al. (2012).

Table 1

Optimal parameters of the proposed emergence GA model based on both dormancy release (AcDr) and germination/pre-emergence growth (AcG) logistic distributions. Mean and variance values of the corresponding θ_{AT} (°Cd) and θ_{HTG} (MPa°Cd) distributions are reported.

Logistic models	Optimization parameters		Mean (λ)	Variance $(\pi/\delta)^2/3$
	δ	λ		
$AcDr(\theta_{AT})$	0.003	1770.03	1770.03 °Cd	365,540.4 °Cd
$AcG(\theta_{HTG})$	0.021	60.42	60.42 MPa °Cd	7460 MPa °Cd

3. Results

In Table 1, the parameters of the newly developed GA model are provided. By inspecting the variance parameter, it can be observed that the logistic function of the dormancy release process has a much broader distribution than the germination/pre-emergence growth phase, thus, indicating significant differences in the mean after-ripening thermal time requirement of the population among years.

In Fig. 4, simulation results for the two test years are presented. As observed in Fig. 4A, simulated emergence satisfactorily represented observed data for a year of severe soil availability limitations, while for 2008 (Fig. 4B) the first cohort was closely represented while the second was rather averaged.

In Table 2, results corresponding to the proposed GA emergence modeling approach as well as selected models from Chantre et al. (2012) are reported for comparison purposes. According to the BIC method, the best univariate (single-input) modeling alternative would be the NLR Weibull(θ_{HT}) model, while the ANN$_1$(θ_{HT}) model showed the best predictive performance based on the RMSE test. The bivariate ANN$_6$ model presented the lower BIC value for $d = 1$, showing the lowest training error of all compared models. However, by imposing a heavier weight on the penalty term ($d = 1.5$), BIC criteria selects for model GA(θ_{AT}, θ_{HTG}). According to this index, GA(θ_{AT}, θ_{HTG}) model is clearly more parsimonious than the ANN$_6$ alternative. In addition, it showed a superior predictive performance as indicated by the RMSE test value (Table 2).

4. Discussion

The newly developed model based on the genetic algorithms optimization method proved to be a simple modeling approach conceptualized as a serial combination of logistic distributions to describe both seedbank dormancy release and germination/pre-emergence growth, based on after-ripening thermal-time and

hydrothermal-time accumulation as explanatory variables, respectively. The GA model provided enough flexibility to closely represent the complex emergence patterns of the analyzed test sets, thus confirming our work hypothesis that the disaggregation of the after-ripening process from the germination/pre-emergence growth phases allowed increasing the parsimony as well as the prediction capability of the wild oat emergence model compared to previously developed NLR and ANN approaches for the same agroecological system. As previously reported by our work team, the implementation of bivariate ANN models based on thermal-time and hydro-time as independent explanatory variables demonstrated to be a superior approach compared to univariate (θ_{HT} based) models for A. fatua field emergence prediction under semiarid environmental conditions (Chantre et al., 2012). However, the ANN6 model selected as the best modeling alternative in such work still failed to accurately predict wild oat emergence cohorts despite its complexity in regard to the number of involved adjustable parameters.

Observed differences in the mean after-ripening thermal time requirement among years suggest that the seedbank dormancy release phase is actually the dominating process under the irregular climatic conditions of the system under study. As suggested by Sawhney and Naylor (1980) and Peters (1982), the effect of a heterogeneous maternal environment during A. fatua seed development and maturation in the mother plant may greatly influence seed dormancy expression. Foley and Fennimore (1998) stated that variability in seed dormancy among wild oat populations is about 50% genetic and 50% due to a combination of the influence of the environment and genotype-environment interactions, a crucial aspect that should be accounted for in the development of predictive models.

An important feature of the present modeling alternative is that input variables (i.e. soil mean daily temperature and soil water potential) are easily obtained soil microclimatic information which could be gathered from series of
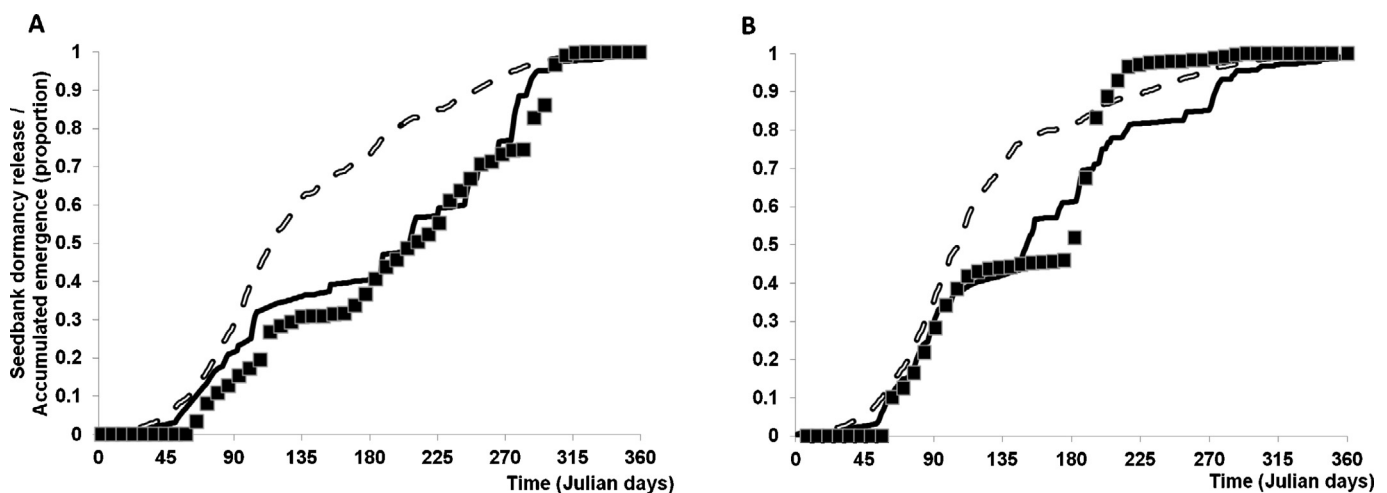

Fig. 4. Observed vs. predicted A. fatua cumulative emergence curves for the GA(θ_{AT}, θ_{HTG}) model for both test years (independent data sets): 2006 (A) and 2008 (B). Observed field emergence data (close symbols) as well as models predictions for both seedbank dormancy release process (dashed line) and germination/pre-emergence growth (solid line) are presented. Calendar time (Julian days) was used in the abscissas instead of soil microclimatic indices to simultaneously present both explanatory variables.

Table 2
Statistical results for the comparison of the GA model with previously developed NLR and ANN models for the same agroecological system.

Model	Reference	m	Train set				Test set
			BIC			RMSE	
			d = 1	d = 1.5			
$AcEm = $ Weibull(θ_{HT})	Chantre	2	−1.32	−1.32		0.215	0.187
$AcEm = $ ANN$_1(\theta_{HT})$	et al.	4	−1.29	−1.27		0.220	0.168
$AcEm = $ ANN$_6(\theta_T, \theta_H)$	(2012)	25	−1.93	−1.36		0.092	0.106
$AcEm = $ GA$(\theta_{AT}, \theta_{HTG})$	This work	4	−1.56	−1.54		0.162	0.068

m, total number of model parameters; BIC, Bayesian Information Criterion with different weights on the penalty term (d = 1 or d = 1.5); RMSE, Root Mean Square Error.

meteorological data available from official weather stations or obtained from web-based weather records (see for example http://inta.gob.ar/documentos/informacion-agrometeorologica-1). In addition, the modeling approach does not require the estimation of the initial seedbank which is a highly complex and time-consuming input variable of most of mechanistic-based models (Colbach et al., 2005).

Future work involves the extension of the proposed approach to study other systems. For example, in temperate environments with non-severe soil moisture limitations where seedbank dormancy does not play a dominating role on the emergence dynamics, the superiority of bivariate models was also demonstrated (Chantre et al., 2013). In such case, the improved performance is considered to be based on the fact that a more accurate discrimination of germination and post-germination sub-processes was obtained, mainly associated to the fact that germination and shoot growth are different hydro-thermal driven process (Bradford, 1990, 2002; Gardarin et al., 2012b). In this regard, the proposed modeling approach could also be applied to such systems, by representing the underlying phases (i.e. germination and pre-emergence growth) as different distributions of appropriate accumulated hydrothermal-time indices. Moreover, since the cardinal parameters of the accumulation indices (Eqs. (1) and (2)) are usually associated to a normal or log-normal distribution within the population, specific functional relationships between the level of dormancy of a given weed population and the above mentioned parameters might be established (Chantre et al., 2009, 2010). The proposed parameter estimation procedure, based on the genetic algorithm optimization technique is considered to be flexible enough to address the identification of the seed population distributions of each phase of the emergence process while simultaneously considering the distributions of the cardinal parameters of the population as functions of appropriate descriptive variables.

Obtained results suggest that the development of this type of models offer an enormous potential to be implemented as weed emergence predictors within management decision support tools in order to assist farmers and agricultural professionals to determine the best time to apply a control measure. For example, as recently shown in Lodovichi et al. (2013), the technical decision regarding the selection of herbicides to use and the optimal time of intervention generated by an operational planning model highly depends on an accurate estimation of the moment and magnitude of weed emergence flushes.

Finally, the proposed modeling framework is considered general for wild oat biotypes showing seed dormancy-based ecological adaptations. However, it should be stressed, that obtained parameters are considered to be representative only for the wild oat system of the region under study. Careful extrapolation might be attempted to other agroecological regions. If data series from other species and locations are available, specific models could be generated following a similar procedure. Model parameters, of course, should be evaluated for each specific case with the appropriate experimental data.

References

Adkins, S.W., Loewen, M., Symons, S.J., 1986. Variation within pure lines of wild oat (Avena fatua) in relation to degree of primary dormancy. Weed Science 34, 859–864.

Bair, N.B., Meyer, S.E., Allen, P.S., 2006. A hydrothermal after-ripening time model for seed dormancy loss in Bromus tectorum L. Seed Science Research 16, 17–28.

Bauer, M.C., Meyer, S.E., Allen, P.S., 1998. A simulation model to predict seed dormancy loss in the field for Bromus tectorum L. Journal of Experimental Botany 49, 1235–1244.

Benech-Arnold, R.L., Sánchez, R.A., Forcella, F., Kruk, B.C., Ghersa, C.M., 2000. Environmental control of dormancy in weed seed banks in soil. Field Crops Research 67, 105–122.

Bewley, J.D., Black, M., 1994. Seeds: Physiology of Development and Germination. Plenum Press, New York.

Bradford, K.J., 1990. A water relations analysis of seed germination rates. Plant Physiology 94, 840–849.

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

Baskin, C.C., Baskin, J.M., 1998. Seeds: Ecology, Biogeography, and Evolution of Dormancy and Germination. Academic Press, San Diego.

Bullied, W.J., Van Acker, R.C., Bullock, P.R., 2012. Hydrothermal modeling of seedling emergence timing across topography and soil depth. Agronomy Journal 104, 423–436.

Cao, R., Francisco-Fernández, M., Anand, A., Bastida, F., González-Andújar, J.L., 2011. Computing statistical indices for hydrothermal times using weed emergence data. Journal of Agricultural Science 149, 701–712.

Chantre, G.R., Batlla, D., Sabbatini, M.R., Orioli, G.A., 2009. Germination parameterization and development of an after-ripening thermal-time model for primary dormancy release of Lithospermum arvense seeds. Annals of Botany 103, 1291–1301.

Chantre, G.R., Sabbatini, M.R., Orioli, G.A., 2010. An after-ripening thermal-time model for Lithospermum arvense seeds based on changes in population hydro-time parameters. Weed Research 50, 218–227.

Chantre, G.R., Blanco, A.M., Lodovichi, M.V., Bandoni, J.A., Sabbatini, M.R., López, R., Vigna, M., Gigón, R., 2012. Modeling Avena fatua seedling emergence dynamics: an artificial neural network approach. Computers and Electronics in Agriculture 88, 95–102.

Chantre, G.R., Blanco, A.M., Forcella, F., Van Acker, R.C., Sabbatini, M.R., Gonzalez-Andujar, J.L., 2013. A comparative study between nonlinear regression and artificial neural network approaches for modeling wild oat (Avena fatua) field emergence. Journal of Agricultural Science (Cambridge), 1–9, http://dx.doi.org/10.1017/S002185961200109.

Chauhan, B.S., Gill, G., Preston, C., 2006. Seedling recruitment pattern and depth of recruitment of 10 weed species in minimum tillage and no-till seeding systems. Weed Science 54, 658–668.

Colbach, N., Dürr, C., Roger-Estrade, J., Caneill, J., 2005. How to model the effects of farming practices on weed emergence. Weed Research 45 (1), 2–17.

Cousens, R., Weaver, S.E., Porter, J.R., Rooney, J.M., Butler, D.R., Johnson, M.P., 1992. Growth and development of Avena fatua L. (Wild-oat) in the field. Annals of Applied Biology 120, 339–351.

Damiano, F., López, R.L., Vigna, M.R., Moschini, R., 2010. Evaluación del modelo microclimático del suelo STM^2 para estudios de emergencia de plántulas de Avena fatua. In: Proceedings of The I Congreso Internacional de Hidrología de Llanuras, Azul, Buenos Aires, Argentina, pp. 555–561.

De Gooijer, J.G., Kumar, K., 1992. Some recent developments in non-linear time series modelling, testing, and forecasting. International Journal of Forecasting 8 (2), 135–156.

Fennimore, S.A., Nyquist, W.E., Shaner, G.E., Doerge, R.W., Foley, M.E., 1999. A genetic model and molecular markers for wild oat (*Avena fatua* L.) seed dormancy. Theoretical and Applied Genetics 99, 711–718.

Foley, M.E., Fennimore, S.A., 1998. Genetic basis for seed dormancy. Seed Science Research 8 (2), 173–182.

Forcella, F., 1998. Real-time assessment of seed dormancy and seedling growth for weed management. Seed Science Research 8, 201–209.

Gardarin, A., Guillemin, J.P., Munier-Jolain, N.M., Colbach, N., 2012a. Estimation of key parameters for weed population dynamics models: base temperature and base water potential for germination. European Journal of Agronomy 32, 162–168.

Gardarin, A., Durr, C., Colbach, N., 2012b. Modeling the dynamics and emergence of multispecies weed seed bank with species traits. Ecological Modelling 240, 123–138.

Ghersa, C.M., Martínez-Ghersa, M.A., 2000. Ecological correlates of weed seed size and persistence in the soil under different tilling systems: implications for weed management. Field Crops Research 67 (2), 141–148.

Gianinetti, A., Cohn, M.A., 2007. Seed dormancy in red rice. XII. Population-based analysis of dry-after-ripening with a hydrotime model. Seed Science Research 17, 253–271.

Granger, C.W., 1993. Strategies for modelling nonlinear time-series relationships. Economic Record 69 (3), 233–238.

Grundy, A.C., 2003. Predicting weed emergence: a review of approaches and future challenges. Weed Research 43 (1), 1–11.

Haj Seyed-Hadi, M.R., González-Andújar, J.L., 2009. Comparison of fitting weed seedling emergence models with nonlinear regression and genetic algorithm. Computers and Electronics in Agriculture 65, 19–25.

Jana, S., Naylor, J.M., 1980. Dormancy studies in seed of *Avena fatua*. 11. Heritability for seed dormancy. Canadian Journal of Botany 58, 91–93.

Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modeling: an introduction. Ecological Modelling 120, 65–73.

Li, B., Foley, M.E., 1997. Genetic and molecular control of seed dormancy. Trends in Plant Science 2 (10), 384–389.

Lodovichi, M.V., Blanco, A.M., Chantre, G.R., Bandoni, J.A., Sabbatini, M.R., López, R., Vigna, M., Gigón, R., 2013. Operational planning of herbicide-based weed management. Agricultural Systems 121, 117–129.

Martinson, K., Durgan, D., Forcella, F., Wiersma, J., Spokas, K., Archer, D., 2007. An emergence model for wild oat (*Avena fatua*). Weed Science 55, 584–591.

Masin, R., Loddo, D., Bemvenuti, S., Otto, S., Zanin, G., 2012. Modeling weed emergence in Italian maize fields. Weed Science 60, 254–259.

Matlab, 2013. Global Optimization Toolbox User's Guide.

Mesgaran, M.B., Mashhadi, H.R., Alizadeh, H., Hunt, J., Young, K.R., Cousens, R.D., 2013. Importance of distribution function selection for hydrothermal time models of seed germination. Weed Research 53, 89–101.

Michalewicz, Z., 1996. Genetic Algorithms + Data Structures = Evolution Programs, 3rd ed. Springer, Berlin.

Moles, C.G., Mendes, P., Banga, J.R., 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Research 13, 2467–2474.

Moschini, R.C., Lopez, R.L., Vigna, M.R., Damiano, F., 2009. Modelos basados en tiempo térmico e hidrotérmico para predecir la emergencia de Avena fatua en lotes con y sin labranza estival, en Argentina. In: Sousa, E., Calha, I., Moreira, I., Monteiro, A., Rodrigues, L., Portugal, J. (Eds.), Herbologia e Biodiversidade numa agricultura sustentable. ISA Press, Lisboa, Portugal, pp. 239–242.

Naylor, J.M., Fedec, P., 1978. Dormancy studies in seed of *Avena fatua*. 8. Genetic diversity affecting response to temperature. Canadian Journal of Botany 56 (18), 2224–2229.

Naylor, J.M., Jana, S., 1976. Genetic adaptation for seed dormancy in *Avena fatua*. Canadian Journal of Botany 54 (3–4), 306–312.

Olden, J.D., Jackson, D.A., 2002. Illuminating the "black box"; a randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling 154, 135–150.

Page, E.R., 2004. Characterizing Spatially Variable Patterns of Wild Oat (*Avena fatua* L.) Emergence on the Palouse. Washington State University, Pullman, WA (Msc Thesis).

Page, E.R., Gallagher, R.S., Kemanian, A.R., Zhang, H., Fuerst, E.P., 2006. Modeling site-specific wild oat (*Avena fatua*) emergence across a variable landscape. Weed Science 54, 838–846.

Peters, N.C.B., 1982. The dormancy of wild oat seed (*Avena fatua* L.) from plants grown under various temperature and soil moisture conditions. Weed Research 22 (4), 205–212.

Rangaiah, G.P. (Ed.), 2010. Stochastic Global Optimization: Techniques and Applications in Chemical Engineering. Word Scientific Publishing, Singapore.

Sawhney, R., Naylor, J.M., 1980. Dormancy studies in seed of Avena fatua. 12. Influence of temperature on germination behavior of nondormant families. Canadian Journal of Botany 58 (5), 578–581.

Schutte, B.J., Regnier, E.E., Harrison, S.K., Schmoll, J.T., Spokas, K., Forcella, F., 2008. A hydrothermal emergence model for giant ragweed (*Ambrosia trifida*). Weed Science 56, 555–560.

Scursoni, J.A., Martín, A., Catanzaro, M.P., Quiroga, J., Goldar, F., 2011. Evaluation of post-emergence herbicides for the control of wild oat (*Avena fatua* L.) in wheat and barley in Argentina. Crop Protection 30, 18–23.

Spokas, K., Forcella, F., 2009. Software tools for weed seed germination modeling. Weed Science 57, 216–227.

Symons, S.J., Naylor, J.M., Simpson, G.M., Adkins, S.W., 1986. Secondary dormancy in *A. fatua*: induction on characteristics in genetically pure dormant lines. Physiologia Plantarum 68, 27–33.

Symons, S.J., Simpson, G.M., Adkins, S.W., 1987. Secondary dormancy in *Avena fatua*: effect of temperature and after-ripening. Physiologia Plantarum 70 (3), 419–426.

Vleeshouwers, L.M., Kropff, M.J., 2000. Modelling field emergence patterns in arable weeds. New Phytologist 148 (3), 445–457.

Zambrano-Navea, C., Bastida, F., González-Andújar, J.L., 2013. A hydrothermal seedling emergence model for *Conyza bonariensis*. Weed Research 53, 213–220.