



Predicting field weed emergence with empirical models and soft computing techniques

J L GONZALEZ-ANDUJAR*, G R CHANTRE†, C MORVILLO*‡, A M BLANCO§ & F FORCELLA¶

*Instituto de Agricultura Sostenible (CSIC), Córdoba, Spain, †Departamento de Agronomía/CERZOS, Universidad Nacional del Sur/CONICET, Bahía Blanca, Buenos Aires, Argentina, ‡Departamento de Producción Vegetal, Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina, §Planta Piloto de Ingeniería Química, Universidad Nacional del Sur/CONICET, Bahía Blanca, Buenos Aires, Argentina, and ¶USDA-ARS North Central Soil Conservation Research Laboratory, Morris, MN, USA

Received 13 October 2015

Revised version accepted 13 June 2016

Subject Editor: Rob Freckleton, Sheffield, UK

Summary

Seedling emergence is one of the most important phenological processes that influence the success of weed species. Therefore, predicting weed emergence timing plays a critical role in scheduling weed management measures. Important efforts have been made in the attempt to develop models to predict seedling emergence patterns for weed species under field conditions. Empirical emergence models have been the most common tools used for this purpose. They are based mainly on the use of temperature, soil moisture and light. In this review, we present the more popular empirical

models, highlight some statistical and biological limitations that could affect their predictive accuracy and, finally, we present a new generation of modelling approaches to tackle the problems of conventional empirical models, focusing mainly on soft computing techniques. We hope that this review will inspire weed modellers and that it will serve as a basis for discussion and as a frame of reference when we proceed to advance the modelling of field weed emergence.

Keywords: artificial neural networks, genetic algorithms, predictive modelling, nonlinear regression, weed control, day degrees, d °C.

GONZALEZ-ANDUJAR JL, CHANTRE GR, MORVILLO C, BLANCO AM & FORCELLA F (2016). Predicting field weed emergence with empirical models and soft computing techniques. *Weed Research*. doi: 10.1111/wre.12223.

Introduction

Seedling emergence is one of the most important phenological processes that influences the success of weed species (Forcella *et al.*, 2000) and, therefore, predicting weed emergence timing plays a critical role in scheduling weed management measures (Ghersa, 2000). For example, from the chemical control perspective, many post-emergence herbicides will not control weeds that have not yet emerged at the time of application. Conversely, if control is applied too late, early emerged weeds may be too large for adequate control (Dalley

et al., 2004). Moreover, if weeds are controlled too late, potential crop yield can be reduced. Organic (ecological) farming practices using mechanical weed control techniques also would benefit from weed emergence predictive tools to better define intervention times (Oriade & Forcella, 1999; Forcella, 2012).

Since the early 1990s, important efforts have been made in the attempt to develop models to predict seedling emergence patterns for several weed species (e.g. Dorado *et al.*, 2009; Izquierdo *et al.*, 2013; Zambrano-Navea *et al.*, 2013; Werle *et al.*, 2014) as a fundamental step in the development of IWM strategies. These

efforts vary from empirical weed emergence models to more complex mechanistic models.

Although mechanistic or reductionist models can help to provide a close description of the basic eco-physiological processes underlying weed emergence (i.e. seed dormancy, germination and pre-emergence shoot growth) (Vleeshouwers & Kropff, 2000; Batlla *et al.*, 2004; Gardarin *et al.*, 2012), they usually require a large amount of often unavailable or difficult to gather experimental information to be developed, calibrated and finally validated (Grundy, 2003). Conversely, empirical weather-based models aim to identify the relation between soil environmental variables and emergence data collected under field conditions with the final objective to provide real-time information on weed management programmes.

Empirical models simply describe the general shape of the data set to which the researcher is trying to fit a curve. They provide a simpler framework for practical support decisions regarding optimal time for intervention. The most common means of analysing such relationships is some form of regression, which typically is nonlinear.

In this review, we have restricted our attention to empirical models, as they are the most common models used for modelling weed emergence under field conditions. We identify some statistical and biological limitations that could affect their predictive accuracy and, finally, we present briefly a new generation of algorithm modelling approaches to overcome their limitations, focusing mainly on soft computing techniques.

Predicting weed emergence

Empirical emergence models are tools for predicting the percentage of total weed emergence occurrence based mainly on the use of temperature and soil moisture (Forcella *et al.*, 2000; Grundy, 2003) and, more recently, light (Royo-Esnal *et al.*, 2015). They are based on the thermal- or hydrothermal-time concepts (Gummerson, 1986; Roman *et al.*, 2000; Martinson *et al.*, 2007), which assume that seeds need to accumulate a certain amount of growing degree days (d °C; GDD) before completing germination and emergence. They link emergence with GDD and can be classified, in a general way, as thermal-time, hydrothermal-time and photohydrothermal-time models. In thermal-time models (TT), daily mean soil temperature (eventually air temperature) is accumulated above a specific threshold during the cropping season until weed emergence is completed. Thermal-time accumulation can start at various times depending upon specific situations. These dates could be the last soil tillage event or burndown herbicide application, crop sowing or an arbitrary date such as 1 January for summer-

growing weeds in cold climates of the Northern Hemisphere, or winter-growing weeds in warm climates of the Southern Hemisphere.

A more complex approach includes the integration of soil water potential with soil temperature into hydrothermal time (HTT). These models can be better at predicting emergence than TT models (McGiffen *et al.*, 2008), as they include soil water availability, which is an important factor for seed germination and shoot growth. In these models, GDD are accumulated when the daily average of soil water potential and soil temperature are above specified threshold values below which seedlings cannot emerge (Gummerson, 1986). If experimental soil temperature and water potential data are missing, the weed modeller can easily and accurately estimate such variables using free access software (e.g. soil temperature and moisture model (STM²); Spokas & Forcella, 2009). In photohydrothermal-time models (PhHTT), photoperiod is used to modify TT based upon day length (Royo-Esnal *et al.*, 2015). The rationale is that with longer day lengths, soil is irradiated longer and accumulates more heat than with shorter day lengths, even when maximum and minimum air temperatures are identical between days with long and short day lengths. The output for these models is percentage of the total annual emergence of a given species.

A typical nonlinear regression model (NLR) for weed emergence is like the following

$$Y = f(x, \varphi) + \varepsilon \quad (1)$$

where $\varepsilon \approx N(0, \sigma^2)$, Y is cumulative emergence, x is cumulative GDD (TT, HTT or PhHTT), and φ is specific function parameter. The term $f(x, \varphi)$ is a nonlinear S-shaped function. S-shaped curves start at some fixed point, then increase up to an inflection point, after which the slope of the curve decreases and the curve approaches the upper asymptote (Ratkowsky, 1983) (Fig. 1). Different parametric S-type models have been commonly fitted to emergence data (e.g. Schutte *et al.*, 2008; Dorado *et al.*, 2009) and, among these, the Logistic, Weibull and Gompertz have been used most widely in predicting weed emergence (Table S1).

Logistic:

$$Y = k / (1 + \exp(-a(x - p))) \quad (2)$$

Gompertz:

$$Y = k \exp(-\exp(-a(x - p))) \quad (3)$$

Weibull

$$Y = k(1 - \exp(-(a(x - p))^c)) \quad (4)$$

In these expressions, Y and x as in eqn 1, a is the slope parameter (emergence rate), p represents the inflection point on the x axis, c is a shape factor that

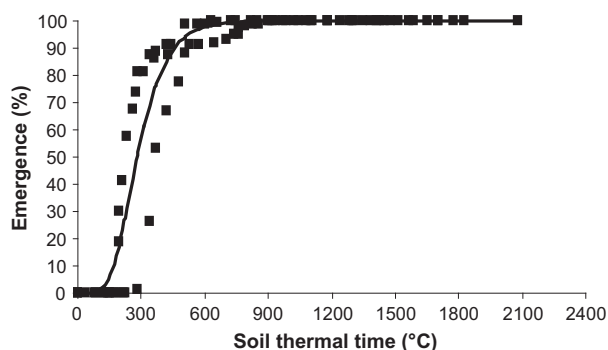


Fig. 1 Typical S-type model fitted to emergence data.

determines the skewness and kurtosis of the distribution, and k is the maximum emergence fraction of the model.

Gompertz and Logistic are very similar functions. Both are special cases of a generalised Logistic model (Calvo Haro *et al.*, 1994). The main differences between the Gompertz and Logistic functions is that the Gompertz function approaches the asymptote much more gradually, which often matches observations of late-emerging seedlings. The Weibull model is a more flexible function than the Logistic and Gompertz. It can acquire the characteristics of other types of distributions based on the value of the shape parameter c (eqn 4). It typically has been used for modelling seedling emergence patterns (Table S1).

Comparison among models has been performed (e.g. Dorado *et al.*, 2009). However, a clearly general favoured model has not emerged. Even the same species (e.g. *Chenopodium album* L. and *Avena fatua* L.) fitted different models best under diverse crops (see Table S1).

Model limitations

Statistical limitations

The aim of NLR models is to predict weed emergence easily and provide a practical tool to make informed management decisions. In this sense, NLR fitting presents several statistical limitations that could affect their predictive accuracy (Onofri *et al.*, 2010; Cao *et al.*, 2011). Some of these issues are briefly presented below:

- Fitting NLR models requires good initial parameter estimates. Poor initial estimates could lead to wrong solutions or may yield no solutions at all (Holmström & Petersson, 2002). The algorithm used to find the solutions of nonlinear equations may incorrectly pinpoint local minima and thus only find local optima and, therefore, resulting parameter estimations are biased.

- The parameters are estimated by NLR fitting routine using algorithms such Marquardt–Levenberg (ML) or Gauss–Newton (GN) (Ratkowsky, 1983). There is a wide range of statistical software to fit NLR models, which likely is the cause of NLR’s popularity. Choice of the appropriate fitting algorithm is very important. For instance, ML is more robust than GN, which means that in many cases, it finds a solution even if it starts very far from the global minimum. Unfortunately, the algorithms used in the fitting procedures may not be selected adequately and users often rely on their statistical software’s default options.

- Observations are censored. When we make periodic seedling counts, the ‘new emergence’ events could have occurred at any time in the interval between the last and current counts. Neglecting the existence of censored data may lead to biased results.

- Observed cumulative emergence values obtained from consecutive monitoring are not statistically independent, resulting in positive autocorrelation of the residuals and, therefore, which leads to erroneous predictions.

Biological/Ecological limitations

Because TT, HTT and PhHTT models are developed based on environmental conditions, they might be used to predict weed emergence across different years and geographical regions. Nevertheless, validation results show that empirical models may not be accurate if environmental conditions vary significantly from the original conditions in which the experiment was conducted (e.g. Izquierdo *et al.*, 2013). Many reasons can explain these differences. For instance, different soil management (e.g. cultivation operation) may reduce the accuracy of the model by varying the vertical distribution of the seeds within the soil profile (Grundy *et al.*, 1996). As a result, seeds will have different temperature, soil moisture and light conditions, and the rate of dormancy release and germination will be different (Cao *et al.*, 2011). In addition, deeply buried seeds need more time to emergence than seeds near the soil surface, thereby delaying and/or extending the emergence flush. Also, model accuracy may be limited by possible population differentiation at diverse geographical locations, as a result of different selection pressures in different environmental conditions (Dorado *et al.*, 2009).

In addition, the use of microclimate indices (TT, HTT and PhHTT) as single explanatory variables is based on the following assumptions: (i) the seedling emergence process is considered as ‘a whole’ without proper discrimination between seed dormancy, germination and post-germination growth subprocesses and (ii) population-based thermal (T_b = base temperature)

and hydric (Ψ_b = base water potential) parameters are assumed independent of soil microclimatic variables (soil water potential and soil temperature, respectively). As suggested by Dahal and Bradford (1994) and Keereab and Murdoch (1999), interactions between T_b and Ψ , as well as between Ψ_b and T , are to be expected, mainly when estimated parameters are not constant within the population (Bradford, 2002).

Further developments

Recently, a new generation of modelling approaches is beginning to tackle the statistical problems of conventional NLR models. For instance, Davis *et al.* (2013) used nonlinear mixed-effect models, which can reduce the number of restrictive assumptions presented by NLR models. Onofri *et al.* (2010, 2011, 2014) presented new developments that maintain the positive features of parametric approaches, and Cao *et al.* (2013) described a nonparametric approach as a very flexible tool to model weed emergence.

Other new conceptual approaches come from the algorithmic modelling culture (*sensu* Breiman, 2001). Algorithmic modelling departs from the concept of an ‘ideal system’ described by *complete-and-precise information* and heads towards a real, uncertain and complex system where precise and deterministic models hardly apply. As highlighted by Bonissone (1997) ‘as we attempt to solve real-world problems we realize that they are typically ill-defined systems, difficult to model and with large-scale solution spaces’. In this context, soft computing techniques (also called computational intelligence) are capable of dealing with complex systems because it does not require strict mathematical definitions. Unlike conventional computing (i.e. hard computing), soft computing is tolerant of uncertainty, imprecision and partial truth, providing a better rapport with reality (Das *et al.*, 2013). Among soft computing techniques (SCTs), artificial neural networks (ANNs) (Chantre *et al.*, 2012, 2014) and genetic algorithms (GAs) (Haj Seyed Hadi & Gonzalez-Andujar, 2009; Blanco *et al.*, 2014) have been proposed as promising tools for weed emergence prediction.

Artificial neural networks (ANNs)

Artificial neural networks are machine learning computer techniques that provide a practical and flexible modelling framework for multivariate nonlinear mapping (Lek & Guégan, 1999). Such models are inspired by the operation of biological neural networks, specifically the animal brain. ANNs are generally represented as a system of interconnected processing units (also called nodes or neurons), which exchange signals (i.e.

information) between each other. The connections have numeric weights that are adjusted during the training process using a given learning algorithm. Therefore, an ANN model is characterised by (i) its architecture (i.e. the pattern of connections among nodes), (ii) its method of determining the weights on the connections (i.e., learning or training process) and (iii) its activation functions (i.e. mathematical functions that process input data). Briefly, a feed-forward neural network (also called multilayer perceptron) consists of input variables (x_n), output variables (Y_n) and a given number of hidden layers containing n nodes (for further details on ANN architectures the reader is referred to Fausett (1994)).

A basic three-layer feed-forward ANN was implemented successfully by Chantre *et al.* (2014) for *Avena fatua* field emergence prediction in different temperate regions of the United States, Canada and South Australia (Fig. 2).

As observed in Fig. 2, each of the input nodes of the model receives a given input variable (note that cumulative thermal time and hydrotime are considered independent variables) and broadcasts it to each one of the hidden neurons. Hidden neurons compute their activation functions and broadcast their results (z_1, z_2, z_3) to the single output neuron that finally produces the response of the network (Y). The output signal of each hidden neuron (z_j) is calculated as follows:

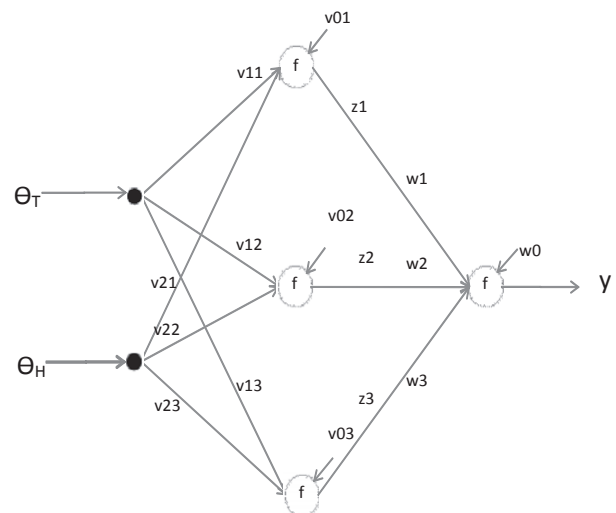


Fig. 2 A three-layer feed-forward ANN. The network has two inputs thermal time ($\Theta_T = (T - T_b)t_g$) and hydrotime ($\Theta_H = (\psi - \psi_b)t_g$), three neurons in the hidden layer and a single output variable (y = cumulative emergence). The term $f(\cdot)$ = the activation function of the network; v_{ij} = connection weights between input- and hidden-layer neurons and w_j = connection weights between hidden- and output-layer neurons; v_{0j} = bias on hidden neuron j ; w_0 = output neuron bias (adapted from Chantre *et al.*, 2014).

$$z_j = f\left(\sum_{i=1,2} v_{ij}x_i + v_{0j}\right) \quad j = 1, \dots, 3 \quad (5)$$

while the output of the network is given by:

$$z_j = f\left(\sum_{i=1,3} w_j z_i + w_0\right) \quad (6)$$

In Eqns (5) and (6), $f(\cdot)$ is the activation function of the network, x_i is the corresponding input variable, v_{ij} is the weight of the connection between the input and hidden neurons, and v_{0j} is the bias of hidden neuron j . Similarly, w_j represents the weight of the connection between hidden and output neurons, and w_0 is the bias of the output neuron.

From a statistical point of view, ANNs have many positive features that should be highlighted. They (i) effectively implement a wide range of nonlinear mapping problems due to their excellent potential for data interpolation, (ii) admit multivariate input/output mapping, (iii) have no requirements for a given *a priori* shape fitting function (i.e. ANNs can easily generate an outcome in the form of a polynomial function of degree n) and (iv) for certain cases involving complex emergence patterns, ANNs require fewer adjustable parameters than conventional multivariate techniques to obtain a similar parsimony.

Among the negative aspects of ANNs implementation, we might cite the following: First, they have very low extrapolation capacity. Thus to adequately optimise a given objective function (e.g. error minimisation), a wide range of observed scenarios is needed to exploit ANN interpolation capability. Second, ANNs employ a 'black-box' model approach with limited explanatory power from a mechanistic biological perspective. Third, trial and error is usually required to identify the network that renders the largest parsimony.

Although ANNs have been used intensively to solve highly complex nonlinear mapping problems in agronomic and biological systems (Saberli *et al.*, 2007; Fortin *et al.*, 2010; Dai *et al.*, 2011), their application for modelling weed emergence remains largely unexplored. Recent studies performed by Chantre *et al.* (2012, 2014) have shown that ANNs are able to outperform conventional NLR models. As discussed by Chantre *et al.* (2014), the adoption of two independent input variables (i.e. thermal time and hydrotime) instead of a single explanatory variable (i.e. hydrothermal time) weighted differently within the neural network which produced better results. This possibly was due to a more intimate discrimination of the germination and post-germination growth processes. In other words, the use of two (or more) independent variables instead of a single aggregated variable enhances the

statistical power of ANNs compared with NLR. Thus, additional variables such as (i) seed burial depth (tillage system), (ii) dispersal or seeding date, (iii) photoperiod or (iv) seedbank dormancy level might be included. However, overparameterisation may occur by increasing the number of input variables, thus generating less parsimonious models.

Genetic Algorithms (GAs)

Genetic algorithms are stochastic optimisation techniques, based on the evolution of sets of potential solutions (also called individuals or phenotypes) following natural selection rules (i.e. selection, crossover, mutation). Each individual has a set of specific properties (chromosomes in a biological sense), which can suffer crossover and mutation. Basically, from an initial population of randomly generated individuals, the GA algorithm iteratively performs a stochastic search in which successive generations (i.e. populations) are obtained. The fitness (i.e. the value of the objective function in the optimisation problem) of every individual in successive populations is evaluated, and the evolutionary process proceeds until the best fitness value (i.e. solution) is obtained. These techniques have demonstrated good performance in nonlinear unconstrained models (Michalewicz, 1996; Rangaiah, 2010; Blanco *et al.*, 2014). In fact, GAs show a good balance between exploration and exploitation of the search space, increasing the probability of convergence to global optima (i.e. expected solutions) instead of reaching local minima (Blanco *et al.*, 2014). In addition, stochastic techniques only use objective function values (e.g. by minimising an error function) in the calculation compared with deterministic algorithms, which also require calculation of derivatives). This feature facilitates their computational implementation and robustness of the search, due to a higher convergence speed. Either GA or ANN approaches can be implemented with average computing time of 15–20 min (personal observation).

For more information on GAs, the interested reader is referred to Michalewicz (1996). Based on the mentioned strengths, Haj Seyed Hadi and Gonzalez-Andujar (2009) suggested that GAs are more appropriate to deal with ill-conditioned optimisation problems than the deterministic optimisation algorithms usually used in the NLR approach. The authors compared GAs with NLR for fitting emergence data of six weed species and found that GAs often provided a better fit than NLR.

Recently, Blanco *et al.* (2014) developed an emergence model for *Avena fatua* based on the disaggregation of seed dormancy release (i.e. after-ripening process) and germination/pre-emergence growth

processes. Logistic functions (eqn 2) were adopted to model separately (i) seedbank dormancy release as a function of after-ripening thermal-time accumulation and (ii) germination/pre-emergence growth as a function of hydrothermal-time accumulation (Fig. 3A). In such work, a GA optimisation approach was implemented for parameter estimation. The logic of the model illustrated in Fig. 3A is that on each calendar day, a specific seedbank fraction loses its dormant condition due to after-ripening thermal-time accumulation (θ_{AT}). The accumulated dormancy release function ($AcDr$) is obtained by integrating such fractions over time following a Logistic distribution (Fig. 3A). Thereafter, each specific non-dormant fraction undergoes the germination/post-germination process by accumulating hydrothermal time (θ_{HTG}). Finally, accumulated

emergence (AcG) 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.

The scheme of the model (i.e. Genetic Algorithm Block) is shown in Fig. 3B. Model input consists of the soil microclimatic data (daily mean soil temperature and soil water potential). Thereafter, θ_{AT} and θ_{HTG} (cumulative time variables) constitute the input of the dormancy release ($AcDr$) and accumulated emergence (AcG) Logistic functions, respectively. In the present example, the solutions are represented by individuals, each one constituted by the parameters of the Logistic functions = a_{dr}, a_g, p_{dr}, p_g (see Fig. 3). The GA algorithm performs selection, crossover and mutation operations (Forrest, 1993) on such

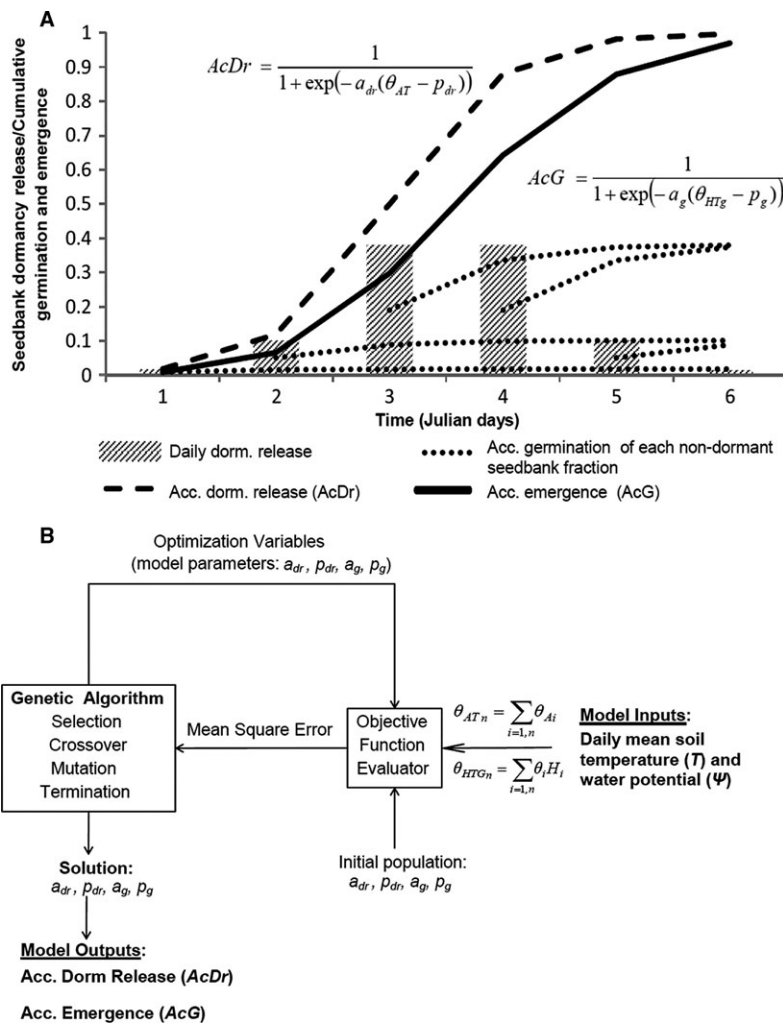


Fig. 3 Genetic algorithm based model for *Avena fatua* field emergence prediction (adapted from Blanco *et al.*, 2014). (A) Schematic representation of the model showing that accumulated dormancy release (dashed line) is obtained by integrating daily dormancy release (bars). Dotted lines represent accumulated 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 during the following days according to 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. (B) Genetic algorithm based parameter estimation scheme.

individuals (i.e. optimisation variables) from a randomly selected initial population along a pre-specified number of generations or until some convergence criterion is met. The main output of the model is the accumulated emergence function, which is obtained by integrating the daily germinated fraction of *A. fatua* corresponding to each non-dormant seedbank fraction which, in turn, is finally represented by the cumulative dormancy release curve (Fig 3A).

As discussed by Blanco *et al.* (2014), the GA optimisation approach provided enough flexibility to closely represent complex *A. fatua* emergence patterns in the semi-arid region of Argentina. In this case, the possibility to quantify separately dormancy release requirements (after-ripening thermal-time accumulation) from hydrothermal-time needs for germination and emergence clearly outperformed the prediction accuracy of previously developed NLR and ANN models. Although the model proposed by Blanco *et al.* (2014) is rather simple in terms of the equations used (Logistics) and parameters interpretation, a negative aspect of such a modelling approach is the fact that it requires moderate programming skills and sufficient level of expertise to develop its heuristics.

Conclusion

Models based on SCTs can provide enough flexibility to represent weed seedling emergence patterns better than NLR models and, therefore, offer more reliable results. An important feature of these new modelling alternatives is the lack of substantial conceptual drawbacks observed in NLR models (Cao *et al.*, 2011), such as the requirement for initial parameter estimates to start the optimisation process that conditions the final solution. Failure to provide an accurate prediction has practical consequences for weed emergence models. Economic losses may occur if control measures do not coincide with the seasonal dynamics of the weed populations, not only in terms of crop/weed competition, but harvest efficiencies and production of seeds by escaped weeds as well.

From a practical point of view, the implementation of SCT models for use by farmers and technicians can be performed effectively. A good example of a successful implementation is the AVEFA-Bordenave model developed by Blanco *et al.* (2014), which was described previously. Output from this model can be accessed freely on line at <http://www.meteobahia.com.ar/productoavefa.php>. Extension of this system to multiple sites within a region as well as multiple regions can be envisioned easily. The implementation of these techniques can also be seen as submodels integrated within operational planning of herbicide-based (Lodovichi

et al., 2013) or strategic IWM programmes (Beltran *et al.*, 2012).

Despite their advantages, some caveats of ANNs and GAs should be mentioned (i) to obtain reliable model predictions, a good pool of input data (i.e. series of cumulative field emergence data with their corresponding soil microclimatic data) are required. These should be representative of a wide range of observed field scenarios to take advantage of the maximum interpolation capacity of both ANNs and GAs during the training process.

(ii) Although SCTs are relatively simple to develop and implement, they require a minimum degree of familiarity with specific programming platforms. For example, the MatLab environment (MathWorks, Inc., Natick, Massachusetts, United States) provides well-developed toolboxes for modelling ANNs and implementing GA optimisation.

Nevertheless, SCT models are a valuable alternative to the classical parametric nonlinear regression models to describe and predict weed emergence. However, further research is required to establish more generally the usefulness of these approaches.

Acknowledgements

J. L. Gonzalez-Andujar was supported by FEDER (European Regional Development Funds) and the Spanish Ministry of Economy and Competitiveness funds (Projects AGL2012-33736). G. R. Chantre was supported by grants from Consejo Nacional de Investigaciones Cientificas y Tecnicas (CONICET) and Universidad Nacional del Sur (PGI-UNS).

References

- BATLLA D, KRUK BC & BENECH-ARNOLD RL (2004) Modelling changes in dormancy in weed soil seed banks: implications for the prediction of weed emergence. In: *Handbook of Seed Physiology. Applications to Agriculture* (eds RL BENECH-ARNOLD & RA SANCHEZ), 245–264. Haworth Press Inc, New York, USA.
- BELTRAN JC, PANNELL DJ, DOOLE GJ & WHITE B (2012) A bioeconomic model for analysis of integrated weed management strategies for annual barnyardgrass (*Echinochloa crus-galli* complex) in Philippine rice farming systems. *Agricultural Systems* **112**, 1–10.
- BLANCO AM, CHANTRE GR, LODOVICHI MV *et al.* (2014) Modeling seed dormancy release and germination for predicting *Avena fatua* L. field emergence: a genetic algorithm approach. *Ecological Modelling* **272**, 293–300.
- BONISSONE PP (1997) Soft computing: the convergence of emerging reasoning technologies. *Soft Computing* **1**, 6–18.
- BRADFORD KJ (2002) Applications of hydrothermal time to quantifying and modeling seed germination and dormancy. *Weed Science* **50**, 248–260.

- BREIMAN L (2001) Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Statistical Science* **16**, 199–231.
- CALVO HARO RM, GONZALEZ-ANDUJAR JL & PEREZ S (1994) *Manual de modelos no lineales en los ámbitos agronómico, ganadero y forestal*. INIA, Madrid, Spain.
- CAO R, FRANCISCO-FERNANDEZ M, ANAND A, BASTIDA F & GONZALEZ-ANDUJAR JL (2011) Computing statistical indices for hydrothermal times using weed emergence data. *The Journal of Agricultural Science* **149**, 701–712.
- CAO R, FRANCISCO-FERNÁNDEZ M, ANAND A, BASTIDA F & GONZALEZ-ANDUJAR JL (2013) Modeling *Bromus diandrus* seedling emergence using nonparametric estimation. *Journal of Agricultural, Biological, and Environmental Statistics* **18**, 64–86.
- CHANTRE GR, BLANCO AM, LODOVICHI MV *et al.* (2012) Modeling *Avena fatua* seedling emergence dynamics: an artificial neural network approach. *Computers and Electronics in Agriculture* **88**, 95–102.
- CHANTRE G, BLANCO A, FORCELLA F, VAN ACKER R, SABBATINI M & GONZALEZ-ANDUJAR JL (2014) A comparative study between non-linear regression and artificial neural network approaches for modelling wild oat (*Avena fatua*) field emergence. *The Journal of Agricultural Science* **152**, 254–262.
- DAHAL P & BRADFORD KJ (1994) Hydrothermal time analysis of tomato seed germination at suboptimal temperature and reduced water potential. *Seed Science Research* **4**, 71–80.
- DAI X, HUO Z & WANG H (2011) Simulation for response of crop yield to soil moisture and salinity with artificial neural network. *Field Crops Research* **121**, 441–449.
- DALLEY CD, KELLS JJ & RENNER KA (2004) Effect of glyphosate application timing and row spacing on corn (*Zea mays*) and soybean (*Glycine max*) yields. *Weed Technology* **18**, 165–176.
- DAS SK, KUMAR A, DAS B & BURNWAL A (2013) On soft computing techniques in various areas. *International Journal of Information Technology and Computer Science* **3**, 59–68.
- DAVIS AS, CLAY S, CARDINA J *et al.* (2013) Seed burial physical environment explains departures from regional hydrothermal model of giant ragweed (*Ambrosia trifida*) seedling emergence in US Midwest. *Weed Science* **61**, 415–421.
- DORADO J, SOUSA E, CALHA I, GONZALEZ-ANDUJAR JL & FERNANDEZ-QUINTANILLA C (2009) Predicting weed emergence in maize crops under two contrasting climatic conditions. *Weed Research* **49**, 251–260.
- FAUSETT LV (1994) *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications* (ed. LV FAUSETT). Prentice-Hall, Englewood Cliffs, New Jersey, USA.
- FORCELLA F (2012) Air-propelled abrasive grit for postemergence in-row weed control in field corn. *Weed Technology* **26**, 161–164.
- FORCELLA F, BENECH-ARNOLD R, SANCHEZ R & GHERSA CM (2000) Modeling seedling emergence. *Field Crops Research* **67**, 123–139.
- FORREST S (1993) Genetic algorithms: principles of natural selection applied to computation. *Science* **261**, 872–878.
- FORTIN JG, ANCTIL F, PARENT L-É & BOLINDER MA (2010) A neural network experiment on the site-specific simulation of potato tuber growth in Eastern Canada. *Computers and Electronics in Agriculture* **73**, 126–132.
- GARDARIN A, DÜRR C & COLBACH N (2012) Modeling the dynamics and emergence of a multispecies weed seed bank with species traits. *Ecological Modelling* **240**, 123–138.
- GHERSA CM (2000) Plant phenology and the management of crop-weed interactions. *Field Crops Research* **67**, 91–93.
- GRUNDY A (2003) Predicting weed emergence: a review of approaches and future challenges. *Weed Research* **43**, 1–11.
- GRUNDY A, MEAD A & BOND W (1996) Modelling the effect of weed seed distribution in the soil profile on seedling emergence. *Weed Research* **36**, 375–384.
- GUMMERSON R (1986) The effect of constant temperatures and osmotic potentials on the germination of sugar beet. *Journal of Experimental Botany* **37**, 729–741.
- HAI SEYED HADI MR & GONZALEZ-ANDUJAR JL (2009) Comparison of fitting weed seedling emergence models with nonlinear regression and genetic algorithm. *Computers and Electronics in Agriculture* **65**, 19–25.
- HOLMSTRÖM K & PETERSSON J (2002) A review of the parameter estimation problem of fitting positive exponential sums to empirical data. *Applied Mathematics and Computation* **126**, 31–61.
- IZQUIERDO J, BASTIDA F, LEZAUN J, SÁNCHEZ DEL ARCO M & GONZALEZ-ANDUJAR JL (2013) Development and evaluation of a model for predicting *Lolium rigidum* emergence in winter cereal crops in the Mediterranean area. *Weed Research* **53**, 269–278.
- KEBREAB E & MURDOCH A (1999) Modelling the effects of water stress and temperature on germination rate of *Orobanche aegyptiaca* seeds. *Journal of Experimental Botany* **50**, 655–664.
- LEK S & GUÉGAN J-F (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* **120**, 65–73.
- LODOVICHI MV, BLANCO AM, CHANTRE GR *et al.* (2013) Operational planning of herbicide-based weed management. *Agricultural Systems* **121**, 117–129.
- MARTINSON K, DURGAN B, FORCELLA F, WIERSMA J, SPOKAS K & ARCHER D (2007) An emergence model for wild oat (*Avena fatua*). *Weed Science* **55**, 584–591.
- MCGIFFEN M, SPOKAS K, FORCELLA F, ARCHER D, POPPE S & FIGUEROA R (2008) Emergence prediction of common groundsel (*Senecio vulgaris*). *Weed Science* **56**, 58–65.
- MICHALEWICZ Z (1996) *Genetic Algorithms + Data Structures = Evolution Programs* (ed. Z MICHALEWICZ). Springer Science & Business Media, Berlin, Germany.
- ONOFRI A, GRESTA F & TEI F (2010) A new method for the analysis of germination and emergence data of weed species. *Weed Research* **50**, 187–198.
- ONOFRI A, MESGARAN M, TEI F & COUSENS R (2011) The cure model: an improved way to describe seed germination? *Weed Research* **51**, 516–524.
- ONOFRI A, MESGARAN M, NEVE P & COUSENS R (2014) Experimental design and parameter estimation for threshold models in seed germination. *Weed Research* **54**, 425–435.

- ORIADE C & FORCELLA F (1999) Maximizing efficacy and economics of mechanical weed control in row crops through forecasts of weed emergence. *Journal of Crop Production* **2**, 189–205.
- RANGAIAH GP (2010) *Stochastic Global Optimization: Techniques and Applications in Chemical Engineering* (ed. GP RANGAIAH). World Scientific, Singapore.
- RATKOWSKY D (1983) *Nonlinear Regression Analysis*. (ed. D RATKOWSKY). Marcel Decker Inc., New York, USA.
- ROMAN ES, MURPHY SD & SWANTON CJ (2000) Simulation of *Chenopodium album* seedling emergence. *Weed Science* **48**, 217–224.
- ROYO-ESNAL A, NECAJEVA J, TORRA J, RECASENS J & GESCH RW (2015) Emergence of field pennycress (*Thlaspi arvense* L.): comparison of two accessions and modelling. *Industrial Crops and Products* **66**, 161–169.
- SABERALI S, NOORI S, KHAZAEI J & HEJAZI A (2007) Artificial neural network modelling of common lambsquarters biomass production response to corn population and planting pattern. *Pakistan Journal of Biological Sciences* **10**, 326–334.
- SCHUTTE BJ, REGNIER EE, HARRISON SK, SCHMOLL JT, SPOKAS K & FORCELLA F (2008) A hydrothermal seedling emergence model for giant ragweed (*Ambrosia trifida*). *Weed Science* **56**, 555–560.
- SPOKAS K & FORCELLA F (2009) Software tools for weed seed germination modeling. *Weed Science* **57**, 216–227.
- VLEESHOUWERS L & KROPFF M (2000) Modelling field emergence patterns in arable weeds. *New Phytologist* **148**, 445–457.
- WERLE R, SANDELL LD, BUHLER DD, HARTZLER RG & LINDQUIST JL (2014) Predicting emergence of 23 summer annual weed species. *Weed Science* **62**, 267–279.
- ZAMBRANO-NAVEA C, BASTIDA F & GONZALEZ-ANDUJAR JL (2013) A hydrothermal seedling emergence model for *Conyza bonariensis*. *Weed Research* **53**, 213–220.

Supporting Information

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

Table S1 Empirical and soft computing based models for weed emergence published from 1996 to 2015. This list does not pretend to be exhaustive, but only gives a sample of the multitude of empirical weed emergence models developed in the last 20 years.