18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Real-world recommendations: Do limits to validation constrain model usefulness?

Thornby, D. [1], **A. J. Diggle** [2] **and S.R. Walker** [1]

[1] *Queensland Department of Primary Industries and Fisheries, Toowoomba, Queensland*
[2] *Department of Agriculture and Food, Western Australia*
Email: david.thornby@dpi.qld.gov.au

Abstract: This paper describes efforts to validate a model of glyphosate resistance in awnless barnyard grass (*Echinochloa colona* (L.) Link.) in the sub-tropical northern grains region of Australia. Possible constraints to the perceived usefulness of the model due to hidden or sparsely documented system factors are discussed.

The evolution of herbicide resistance in real weed populations becomes apparent only after years or decades of selection. Modelling is therefore particularly useful in investigating and describing the change in the weed population and the relative importance of different factors pushing towards or mitigating against that change.

A key role of modelling in farming systems research is to improve understanding of issues such as the evolution of herbicide resistance. However, model results are increasingly important for formulating or adjusting recommendations, including for resistance prevention, to be used by land managers. Importantly, the research projects under which models of agricultural systems are produced are likely to include the development or refinement of recommendations for change as part of their expected outcomes.

The comprehensiveness of model validation can be thought to equate generally to the strength of an argument in favour of a model's hypotheses. By extension (and by convention), more comprehensive validation is often given to equate to a stronger argument in favour of practical recommendations that arise more or less directly from the model's outputs. That is, increasing confidence in the recommendations derived from a model is a result of the perceived validation of the model's predictions as much as of its hypotheses. Conversely, validation that is (by necessity or otherwise) partial or piecemeal, and which is structural rather than empirical, may not be seen as providing as secure an argument in favour of any recommendations for farmers that may be made. In the case of glyphosate resistance, hidden weed population variables and insufficiently detailed farming systems data make empirical, operational validation difficult. It is especially difficult to validate these models in time for them to be most useful in formulating practical recommendations for resistance prevention.

A model of glyphosate resistance in awnless barnyard grass, a key northern Australian weed has been constructed, and attempts made to validate it in order to encourage trust in the model's predictions and the recommendations to industry that might be made from them. Empirical validation was performed through comparison with a population of the weed that was confirmed to be glyphosate resistant in 2007. Structural and behavioural pattern validation were performed during model development.

In the historical dataset used for empirical validation, there are both hidden variables (in particular, the initial proportion of resistance-conferring alleles in the population before selection began) and sparsely documented variables (including the year in which glyphosate was first used, and the number and efficacy of glyphosate applications made to this weed population since then) that create difficulty in making direct comparisons between the real population and the model's predictions. The type of data that is available for use in empirical validation of the model is also a constraint. Herbicide resistance is identified as a point of failure in the agricultural system. Therefore, data from real fields is unlikely to contain information about whether or how the rate of evolution of resistance changes over time in response to the types of system actions that are included in the model. In this paper, potential validation methods and pitfalls are discussed, showing that recommendations can be made using the model's predictions with some confidence, but that herbicide model validation constraints may affect the credibility of the model's predictions particularly for non-scientist users.

Keywords: *Farming systems, glyphosate, herbicide resistance, validation, recommendations*

1. INTRODUCTION

Glyphosate is a broad-spectrum post-emergent herbicide that is recognized as being a critical tool for modern broadacre farming. It is widely available and has substantial advantages over many alternative herbicides in terms of price, operator safety, and environmental safety in a wide range of applications. In particular, reduced tillage farming for soil and moisture conservation relies heavily on the frequent use of glyphosate for broad spectrum weed control. A significant and growing proportion of farming worldwide uses glyphosate-resistant crop varieties. Therefore, conserving the usefulness of glyphosate on as broad as possible a spectrum of weed species is of critical importance.

Weed resistance to glyphosate was first identified in 1996, and as of 2009 there are confirmed populations of 15 species resistant to glyphosate in 13 countries (Heap, 2009). The development of strategies to conserve glyphosate efficacy on at-risk weed species is an ongoing aim of weed scientists.

Several models have been used to develop, refine, and support resistance prevention strategies and recommendations. Two Australian farming system models constructed by Diggle et al. (2003) and Werth et al. (2008) provided evidence that certain agronomic and ecological system factors were of particular importance in determining the risk of developing a glyphosate resistant weed population, and demonstrated that risk-minimising strategies using a greater variety of weed control and cropping options could be used to reduce the rate of evolution of glyphosate resistance. Richter et al. (2002) used modelling to investigate more general factors involved in the evolution of resistance. The model described herein (and in greater detail in Thornby et al., 2008) investigates these factors for awnless barnyard grass (*Echinochloa colona* (L.) Link) and demonstrates similarly that preventative strategies can be used to prevent or slow resistance evolution.

However, empirical validation of these models is problematic. Several key factors in the system are difficult to measure, and in validation data sets from real resistant populations, are only gathered after resistance has been identified as a significant problem, rather than over the 'life' of the system. Accordingly, empirical validation of these types of models using external data is uncommon. Further, resistance models have to date been formulated as 'white box' models – ones in which key factors and relationships between them are represented explicitly. This type of model resists empirical validation, which is better suited to 'black box' regression models in which processes may be bundled together and represented empirically or statistically (Barlas 1996). Because of the difficulties in validating glyphosate resistance models, there may be constraints to the adoption of recommendations generated or refined using the information gained from modelling the system. In this paper a new model of glyphosate resistance is briefly described and the validation of the model using historical data and with other methods is discussed. Implications for the uptake of practical recommendations are also discussed.

2. THE GLYPHOSATE RESISTANCE MODEL

As noted, several models have been developed to investigate the factors that affect the evolution of resistance to herbicides in various regions and farming systems. However, it was felt that glyphosate resistance evolution in grains farming in the subtropical northern Australian grains region was a sufficiently different problem to merit the construction of a regionally specific model. Of principal interest were the effects that the relatively high variation in climate and cropping regimes across the region and between years might have on resistance risk factors, and the effects on resistance risk of changes in farming and weed management practices made during the course of resistance evolution.

The crop modelling environment APSIM (Agricultural Production systems SIMulator, Keating et al., 2003) was used as the basis for an agroecosystem model of weeds evolving herbicide resistance. APSIM has been shown to produce realistic yield predictions for a wide variety of crops. Its scripting language allows the development of flexible simulations of management decisions (including cropping and weed control choices, frequencies and timings) that can change qualitatively over the life of a simulation.

Population dynamics factors such as entry of seeds into the seed bank, mortality in the seed bank and seed persistence are not modelled in APSIM. Nor are genetics factors such as gene flow from outside sources, gene frequencies, and results of mating between different biotypes. However, APSIM can be linked with mathematical process models implemented in Vensim (Ventana Systems, Inc.), which was used to model the 'missing' aspects of weed population dynamics (Smith et al., 2005). Sub-models in Vensim were devised for mating and seed bank processes, to fill the life cycle gaps in the crop model (Figure 1).

The model simulates a weed population with a single-gene glyphosate resistance mechanism that is assumed to provide partially dominant resistance strong enough to result in high levels of survivorship at field rates of glyphosate (Thornby et al., 2008). The software environment consists of a set of APSIM modules with control and parameter files for the APSIM simulation, and a Vensim mathematical model split into separate subsections for seed bank processes, mating processes, and communication between APSIM and Vensim. This approach is used to implement a combination of age-structured population dynamics and population genetics. The APSIM modules simulate management, weather, soil moisture and nutrient balance, output of variables, growth of crops and weed species, and competition between them. Variables are passed between the APSIM and Vensim sub-models using a daily time step. All user interaction is through APSIM, so variables for reporting are passed from Vensim where appropriate. The model can return a range of variables concerning crop and weed growth stage, biomass, stress level, seed production (or yield), seed bank density, soil moisture, fertility status, and proportion of resistance in the weed population. A diagrammatic version of the model is shown in Figure 1.

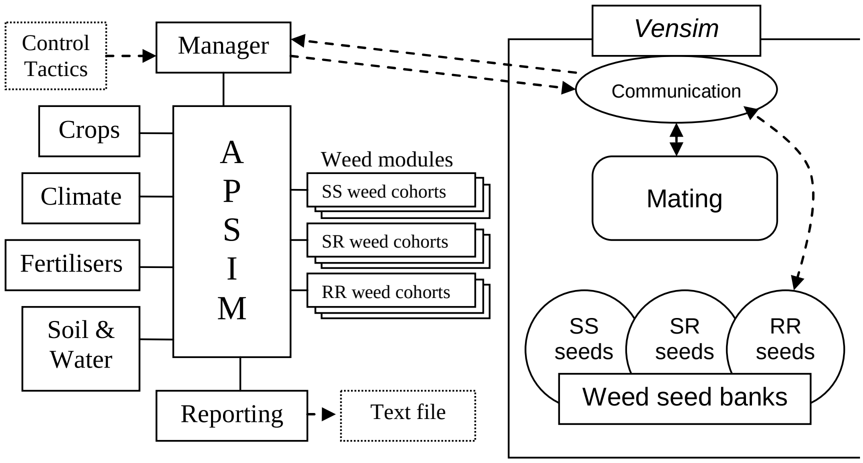

Figure 1. Diagram of model structure, showing those parts of the model (plant growth, agronomy, decision-making) controlled by APSIM and those parts (mating between resistant (RR), susceptible (SS) and heterozygous (SR) genotypes and seed bank dynamics) controlled by Vensim, and communication between them (represented by arrows). Note that outputs (reporting) and simulation-specific inputs (control tactics) are both controlled through APSIM.

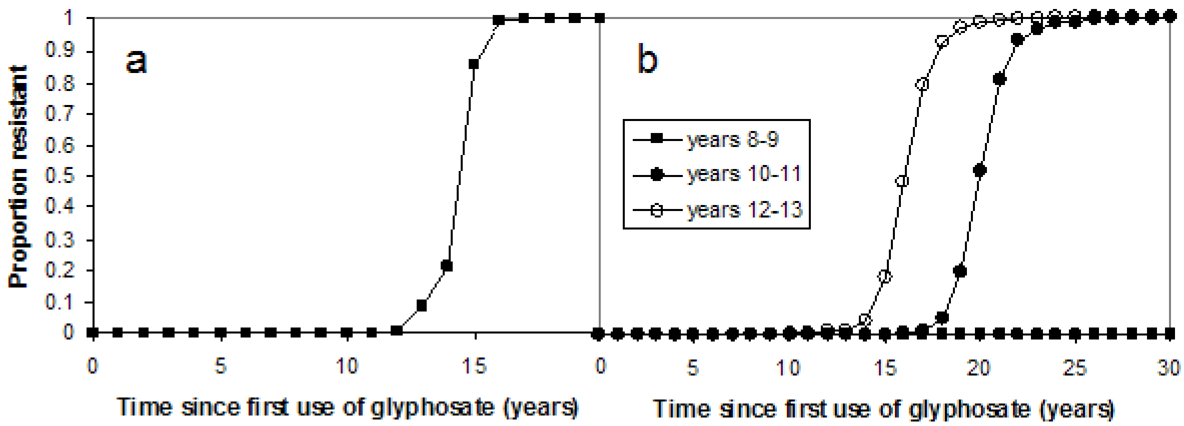

Figure 2. Rate of evolution of glyphosate resistance in awnless barnyard grass. Figure 2a shows the rate of evolution in continuous summer fallow with no tillage and only glyphosate used to control weed flushes. Figure 2b shows the rate of evolution where zero-till glyphosate only fallows for several years is followed by intensive use of double knock tactics for two years (either 8-9, 10-11, or 12-13 years after the first use of glyphosate) followed by the use of double knock on one weed cohort per year thereafter.

The model is able to simulate a range of weed species given adequate parameterisation. Due to its prevalence and relatively high resistance risk (Osten et al., 2007), awnless barnyard grass was chosen as a test case. The weed modules use APSIM's generic plant module, parameterised to simulate multiple barnyard grass cohorts or flushes. For further descriptions of the model's structure and performance, see Thornby et al. (2008).

The model demonstrates that, under continuous summer fallows with no tillage and glyphosate the only herbicide used to control flushes of barnyard grass, the population becomes dominated by glyphosate resistant individuals within 15 years (Figure 2a). This broadly agrees with the predictions made by other glyphosate resistance models (Richter et al., 2002; Diggle et al., 2003; Werth et al., 2008). As an example of the types of investigation of which the model is capable, Figure 2b shows the effect of switching from zero-tillage, glyphosate-only summer fallows to an intensive two year campaign of following glyphosate applications on every weed flush with a different herbicide a few days later (the double knock tactic), and then using the double knock on one flush in every year after that.

3. VALIDATION AND THEORIES OF VALIDATION

Ecological models, including agroecosystem models of herbicide resistance, are inherently challenging to validate (Rykiel, 1996). In previous resistance models, empirical validation (that is, testing of outputs against independent data sets) is either not reported (e.g. Richter et al., 2002 Werth et al., 2008) or qualitatively given as being broadly similar to coarse field data (e.g. Neve *et al*,. 2003). This is largely a limitation of herbicide resistance modelling rather than efforts made in constructing any particular models, as discussed below.

3.1. Empirical validation in system dynamics models

Model validation can take several forms. The simplest to present and interpret is empirical operational validation, wherein model output data on one or more dynamic characteristics of the model are compared to data gathered in the real system (Rykiel 1996). Empirical validation is commonly used to demonstrate that the model meets some purpose-related performance standards. However, statistical tests of empirical validity demand data conditions that system dynamics models rarely satisfy: system dynamics model outputs are not serially independent, and may be cross-correlated with other variables of interest (Barlas, 1996). Particularly where models attempt to examine causal relationships between real system factors (white-box models) they resist validation by statistical tests more appropriate for purely data-driven (black-box) models.

Beyond these statistical limitations, the data that are typically available (that is, gathered from real-world resistance cases) for comparison with models further limit empirical validation. In the field, herbicide resistance is detected as a single point of weed control failure. In reality, the evolution of resistance occurs as a continuum of changing gene frequencies. The ‘point of failure’ typically occurs, in the case of glyphosate resistance, after more than a decade of use. While resistance models give output in the form of a change in the proportion of resistance-conferring alleles in a simulated weed population, data in the field is never gathered on this key variable until the point of failure. Effectively, this represents an attempt to validate a stream of prediction data against a single point of real data which itself is usually an estimate. Clearly, this is a significant weakness in assuring that the model is valid.

3.2. Historical simulations for empirical validation

These limits notwithstanding, empirical validation of the current model was attempted, through simulating a known case of glyphosate resistance and comparing the predictions with what had been observed in the field. Towards the end of model development, the first glyphosate resistant population of barnyard grass was confirmed at Bellata, New South Wales (Cook et al., 2008), and information from this case was used to test the model. Information was gathered about the agronomic history of the paddock through discussions with consulting agronomists and weed specialists from the NSW Department of Primary Industries. The population is from a paddock that had historically been used for growing winter cereals, with some summer cropping only since 2001. First use of glyphosate was estimated to be in 1985, and the paddocks have been zero-tilled in most summers since 1990. Residual herbicides have been applied for grass control as part of recent summer cropping, but no significant rotation of knockdown herbicides away from glyphosate has been made, nor any attempts to control glyphosate survivors.

There were no records kept on glyphosate application number and efficacy throughout the evolution of resistance in the real population. However, expert opinion indicated that most flushes of barnyard grass would have been treated with glyphosate after its introduction. Given the further expert assumption that glyphosate efficacy was often not extremely high under local conditions, 100 simulations were run with a stochastic level of glyphosate efficacy between 80 and 100 per cent on susceptible plants for every application made.

The mean of 100 runs of the stochastic model predicts that 49 per cent of the population is resistant after 15 years, and that this proportion is over 97 per cent after 18 years (Figure 3). There are only partial quantitative estimates available as to the allele frequency in the population at the time resistance was confirmed. The results of herbicide efficacy studies on the population suggest that the proportion of resistant plants across the whole paddock was in the range of 40-60 per cent (Cook et al., 2008), but testing is necessarily done on a relatively small scale. The mean of the stochastic model runs for 40-60 per cent resistance proportion falls between 14 and 17 years after first use of glyphosate. If glyphosate was first used in 1985 (in line with the estimate), 21 years elapsed between first use of glyphosate and confirmation of resistance in the population. Under these assumptions, the model predicts somewhat more rapid evolution than actually occurred.

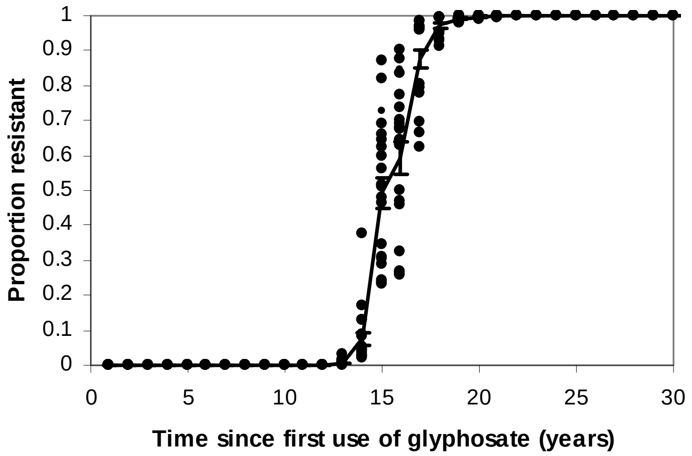

Figure 3. Annual proportion of barnyard grass population resistant to glyphosate for historical simulation of Bellata (NSW) glyphosate resistant barnyard grass. Closed circles represent predicted data from 20 random runs of the model. Line represents the mean of 100 runs; bars represent one standard error of the mean

The validation exercise undertaken here demonstrates two significant difficulties in validating herbicide resistance models empirically. The first is the issue of poorly documented historical variables. Over the 'life' of glyphosate use in the validation paddock, data were never kept on several factors that are likely to have had some bearing on the result; in particular, the exact number and efficacy of glyphosate and non-glyphosate weed control methods that were applied to barnyard grass cohorts. Similarly, the year in which glyphosate was first introduced was not recorded and could only be assumed based on local knowledge. The second problem is the issue of hidden system variables. Critically, the proportion of resistant to susceptible alleles in the population before selection started is not known, and would most likely not be known in any historical dataset. The frequency of resistance-conferring alleles in unselected populations of annual ryegrass (*Lolium rigidum* L.) is estimated at around $1x10^{-6}$ to $1x10^{-8}$ (Diggle et al., 2003, Werth et al., 2008), and similar frequencies have been assumed to apply in barnyard grass populations in the absence of any information suggesting otherwise. At these very low gene frequencies, measurement of actual resistance levels in any given weed population is a substantial undertaking and unlikely to be performed often, either for parameterising models or for comparing models with real cases. Evidence suggests that pre-selection gene frequency has a substantial influence on the time it takes for resistance to dominate the population (Neve et al., 2003; Werth et al., 2008). If a population varies substantially from the simulations' starting allele frequency, the predictions made by the model in terms of years-to-resistance may be substantially different from the real-world observations. This is not to say that there could not be parameters or mechanisms in the model that are incorrect or inappropriately applied; it only suggests that for herbicide resistance models, empirical operational validation is not straightforward. Given these uncertainties, it is difficult to say definitively whether this model, or any resistance model, 'passed' or 'failed' this empirical validation test.

3.3. Structural and pattern validation

The essence of the present model, and other herbicide resistance models, is the accumulation or accounting of birth, seed production, and death events in interbreeding sub-populations over a long time period. The processes for accruing values from one event to the next – addition and subtraction – are well understood. Where the values for individual events, such as germination rates or efficacy of field rates of herbicide on different biotypes, are each based on one or more real world experiments, accounting processes over time are not complex. Validation of the whole model arguably becomes a question of validating the results of each year's operations and relying on the predictable processes of accounting to deal with the 'dynamic' part of the system. Similarly, various sub-models are based on well understood processes, such as Mendelian genetics underlying the breeding sub-model. While the mathematical implementation of these sub-models requires scrutiny, the systems they describe are widely accepted. However, given that the model consists of a large number of factors and sub-models interacting, a computational approach is required, to investigate whether the interactions between elements in the system cause it to behave in a complex way.

Barlas (1996) suggests that tests of model structure and the patterns, rather than or as well as specifics, of a model's predictions should be used to validate it. Rykiel (1996) and Sargent (1999) also discuss similar structural tests of validity. These tests are particularly suited to 'white box' system dynamics models because they emphasize the importance of the way the system is represented, and which components of the real system are included, over quantifying the fit between predicted and observed data.

Qualitative 'expert knowledge' tests of model structure were conducted throughout the development of the model. However, the results or evidence from these tests are fairly subjective, and represented primarily in evolutionary structural change in the model. While redoing these tests for any given expert observer is trivial, the results for each new observer remain subjective. These tests, therefore, do not necessarily lead to substantially increased model credibility, particularly for non-expert stakeholders.

Extreme condition testing is a more directly quantitative method of testing model structural rigor. Zero values were assigned to a range of important parameters including initial weed seed bank populations of some or all genotypes, herbicide efficacy on some or all genotypes, and gamete and seed production in some or all genotypes. With these values held at zero, predictions about the system become much less complex – it can be readily assumed, for example, that when there are no resistant seeds in the seed bank and no novel mutation or gene import, the resistant portion of the population will never increase beyond zero. When efficacy of herbicides on all genotypes is zero, and assuming no fitness penalty accruing to the resistance allele, there would be no change in resistance frequency over time. The model was required to pass multiple extreme-condition tests during development, before any meaningful simulations were undertaken. However, as discussed below, it is difficult to promote these tests as rigorous tests of validity to non-modelling stakeholders, particularly as they effectively represent failure to invalidate the model structure and the relationships contained therein, and not 'positive' validation for a given set of inputs from which the user can (objectively or, more likely, subjectively) interpolate or extrapolate.

4. DO LIMITS TO VALIDATION CONSTRAIN THE MODEL'S USEFULNESS?

In agricultural systems, the uptake of models and recommendations derived from them is constrained by the level of acceptance of the model by individual users and by the industry sector as a whole (Greer et al., 1994; Kayande et al., 2006). User confidence in models has historically been low (Greer et al., 1994), and those who promote the usefulness of models in helping to influence decision-making are challenged with attempting to improve confidence in their models. In the case of this model, the use or uptake of the information generated through applying the model to specific circumstances is likely to be constrained to some extent by how readily accepted the model is by end-users and other non-scientist stakeholders.

Kayande et al. (2006) identify that a knowledge gap between the user's mental model and the computational model affects the perceived value of the computational model. Thus, the objective quality of a model may not be the major arbiter of industry acceptance and end-user uptake of recommendations. Greer et al. (1994) note that 'good explanations are the key to acceptance.' The validation process is a significant part of the explanation of a model. Validation demonstrates that the model 'works', which by extension improves user confidence in the model's predictions. Empirical validation – particularly direct comparison of model predictions and real world values – is arguably the simplest and most concise method of demonstrating model accuracy to non-scientist stakeholders. Empirical validation advertises itself as objective and knowledge-neutral, where structural and pattern validations require expert knowledge of the system to perform (for many structural analyses) or interpret. In these cases, the user is to a substantial extent left to trust the unidentified expert's validation, which may be no more persuasive than not presenting validation at all.

As discussed above, herbicide resistance modelling creates particular challenges for empirical validation, because important variables are hidden and others are unlikely to be present in historical datasets. An inability to validate herbicide resistance models 'positively' against obvious data therefore makes it more difficult to explain these models in the most straightforward way. Thus encouraging user confidence in the predictions models make is a difficult job.

5. CONCLUSIONS

Model validation that satisfies the philosophical and logical requirements of modellers and many other scientists may not necessarily encourage non-scientist stakeholders and model users to view the model as sufficiently credible. Where the aim of a modelling project is to provide or influence a set of recommendations for industry practice, the most easily presented and interpreted validation is likely to be the most persuasive. Non-empirical, non-operational validation techniques seem not likely to fit the bill in this case. The extent to which 'trust' in recommendations relies on the ability of a model to be shown alongside confirmatory real world data may depend on the industry for which the model is produced, but the uptake of recommendations, and hence model usefulness in narrowly defined terms, could be constrained by this issue of trust or acceptability. Herbicide resistance modelling, as discussed here, appears to have exactly the types of issues that act as pitfalls in the validation and trust-engendering process. Due to hidden and poorly documented system variables common in long term agricultural systems, direct empirical validation has been shown to be inadequate as a test of herbicide resistance model validity. Logically and philosophically appropriate validation of these models is possible, through structural analysis, pattern analysis and extreme condition testing, but the value of these tests lies in convincing ourselves of the accuracy of the model, and not in being useful for encouraging non-scientists to have confidence in the model's predictions.

ACKNOWLEDGMENTS

We acknowledge the Grains Research and Development Corporation as the source of funding for this work.

REFERENCES

Barlas, Y. (1996), Formal aspects of model validity and validation in system dynamics, *System Dynamics Review*, 12, 183-210.

Cook, T., A. Storrie, P. Moylan and B. Adams (2008), Field testing of glyphosate-resistant awnless barnyard grass (*Echinochloa colona*) in northern NSW, in: van Klinken, R.D., V.A. Osten, F.D. Panetta and J.C. Scanlan (eds), *Proceedings of the 16th Australian Weeds Conference*, Queensland Weeds Society, Brisbane.

Diggle, A.J., P.B. Neve and F.P. Smith (2003), Herbicides used in combination can reduce the probability of herbicide resistance in finite weed populations, *Weed Research*, 43, 371-382.

Greer, J.E., S. Falk, K.J. Greer, and M.J. Bentham (1994), Explaining and justifying recommendations in an agricultural decision support system. *Computers and Electronics in Agriculture*, 11, 195-214.

Heap, I. (2009) *The International Survey of Herbicide Resistant Weeds*, online at http://www.weedscience.org, accessed March 2009.

Kayande, U., A. de Bruyn, G. Lilien, A. Rangaswamy, and G. Van Bruggen (2006), The effect of feedback and learning on decision-support system adoption, in: Avlonitis G.J., N. Papavassiliou and P. Papastathopoulou (eds.), *Proceedings of the 35th EMAC Conference, Athens*, European Marketing Academy, Brussels.

Keating, B. A., P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D.P. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V.O. Snow, J.P. Dimes, D.M. Silburn, E. Wang, S.D. Brown, K.L. Bristow, S. Asseng, S.C. Chapman, R.L. McCown, D.M. Freebairn and C.J. Smith (2003), An overview of APSIM, a model designed for farming systems simulation, *European Journal of Agronomy*, 18, 267-288.

Neve, P., A.J. Diggle, F.P. Smith and S.P. Powles (2003), Simulating evolution of glyphosate resistance in *Lolium rigidum* I: population biology of a rare resistance trait, *Weed Research*, 43, 404-417.

Osten, V.A., S.R. Walker, A. Storrie, M. Widderick, P. Moylan, G.R. Robinson and K. Galea (2007), Survey of weed flora and management relative to cropping practices in the north-eastern grain region of Australia, *Australian Journal of Experimental Agriculture*, 47, 57-70.

Richter, O., P. Zwerger and U. Böttcher (2002), Modelling spatio-temporal dynamics of herbicide resistance. *Weed Research*, 42, 52-64.

Rykiel, E. (1996), Testing ecological models: the meaning of validation. *Ecological Modelling*, 90, 229-244.

Sargent, R.G. (1999), Verification and validation of simulation models. In: Medeiros, D.J., E.F. Watson, J.S. Carson, and M.S. Manivannan (eds) *Proceedings of the 1998 Winter Simulation Conference*, Institute of Electrical and Electronics Engineers, New York.

Smith, F.P., D.P. Holzworth and M.J. Robertson (2005), Linking icon-based models to code-based models: a case study with the agricultural production systems simulator, *Agricultural Systems*, 83, 135–151.

Thornby, D., S.R. Walker and J.P.M. Whish (2008), Modelling to estimate glyphosate resistance risk in barnyard grass in the northern Australian grain region, in: van Klinken, R.D., V.A. Osten, F.D. Panetta and J.C. Scanlan (eds), *Proceedings of the 16th Australian Weeds Conference*, Queensland Weeds Society, Brisbane.

Werth, J.A., C. Preston, I.N. Taylor, G.W. Charles, G.N. Roberts and J. Baker (2008), Managing the risk of glyphosate resistance in Australian glyphosate-resistant cotton production systems, *Pest Management Science*, 64, 417-421.