**Crop-Weed Relationship Studies in Additive Design: Selecting the Top Model**

Maxwel Coura Oliveira1, Gustavo Antonio Mendes Pereira2, Rodrigo Werle3, and Stevan Knezevic1

1Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Concord, NE, USA.

2Department of Phytotechnic, Federal University of Viçosa, Viçosa, MG, Brazil

3Department of Agronomy and Horticulture, University of Nebraska-Lincoln, North Platte, NE, USA.

**Corresponding author:** Maxwel Coura Oliveira, E-mail: maxwelco@gmail.com

**Category:** Biometry, Modeling, and Statistics.

**ABSTRACT**: Crop-weed competition is extensively studied in weed science. The additive design, where weed density varies, and the crop density is kept constant, is the most used design for competition studies. However, most crop-weed competition research is conducted by non-statisticians, which sometimes report misleading results because of lack of knowledge with this type of data analysis. The objective of this study is to demonstrate selection of top model for describing crop-weed relationship in additive design to non-statisticians. We evaluated three models routinely used to interpret competition studies, including polynomial quadratic, logistic, and rectangular hyperbola. Based on statistical criteria and meaningfulness of parameters, we demonstrated the rectangular hyperbola to be the top model to describe crop-weed competition studies in additive design. Moreover, we showed that at low densities *C. benghalensis* is more competitive than *Richardia brasiliensis* in corn, but both weed species compete similarly at higher densities. We propose the use of the rectangular hyperbola as a standardized model for crop-weed competition in additive design.

**Keywords**: AIC criterion, model selection, crop-weed competition, rectangular hyperbola

**Nomenclature:** *Commelina benghalensis*, *Richardia brasiliensis*

**Introduction**

Studies have described the relationship function of crop yield loss in response to weed density using additive design. Despite several review papers recommending the use of rectangular hyperbola in the weed science literature (Ritz et al. 2015; Swanton et al. 2015), there is no common sense to which model is the most appropriate (Silva et al. 2015; Strieder et al. 2007; Trezzi et al. 2015; Voll et al. 2002). Four major regressions curves are commonly used: linear (Figure 1A), polynomial quadratic (Figure 1B), sigmoid (Figure 1C), and rectangular hyperbola (Figure 1D).

The commonly used criteria for selection of linear and nonlinear regression models is the equation with higher R-squared (R2). The R2 tests the goodness of fit and is statistically inadequate for nonlinear model selection (Archontoulis and Miguez 2015; Zuur et al. 2007). There are several appropriate statistical criteria for selecting the best nonlinear model for datasets: Alkaike’s information criterion (AIC), Bayesian information criterion (BIC), F-test, and likelihood ratio (Anderson 2008; Lewis et al. 2011; Zucchini 2000). Non-nested models are models with different structure and parameters, in this case, the AIC and BIC are indicated for top model selection. However, F-test or any of the model selection techniques described above are applicable for nested models. Nested are models that are a special case of each other and have identical terms whereas one must have one or more extra terms. From a practical standpoint, the top model should be selected upon a balance between statistics and biological relevance, which will help scientists answer their research questions (Archontoulis and Miguez 2015). Therefore, in crop-weed relationship studies in additive design, the model that provides a good fit and meaningful biological parameters are considered a strong candidate model.

The advances in statistical software should facilitate the use of standardized nonlinear regression analysis that could performed by non-statisticians (Knezevic et al. 2007). Here, we compare three non-nested candidate models (polynomial quadratic, logistic, and a rectangular hyperbola) for describing the crop-weed relationship. We used data from an experiment of corn (*Zea mays* L.) in competition with two weed species, *Richardia brasiliensis* and *Commelina benghalensis*. First, we demonstrate the selection process for the top model to describe the crop-weed relationship for each weed species. Second, we test the hypothesis that *C. benghalensis* and *R. brasiliensis* compete similarly with corn.

**Material and Methods**

**Plant Material**. On March 2011, inflorescences of *R. brasiliensis* were harvest on roadsides near Diamantina, Minas Gerais (MG), Brazil. Inflorescences of *R. brasiliensis* were dried at room temperature (25 °C), cleaned, and stored at 5 °C until the onset of the experiment. Ten days before the experiment began (September 2011), stolon (vegetative propagules) of *C. benghalensis* were collected in wetlands, near Diamantina, MG. Seeds of *R. brasiliensis* and stolon of *C. benghalensis* were seeded and transplanted to separate trays (1210 cm3) filled with red latosol (pH 6.1 and 1% organic matter). A single seed of glyphosate-resistant (GR) corn (AG8088) was sown in 8 dm3 plastic pots filled with the aforementioned soil source. The soil was fertilized following the local recommendations, and N was applied at 15 and 30 DAE (days after corn emergence) at a rate of 55 mg dm-3 of ammonium sulfate. Greenhouse conditions were 26/19 °C day/night, and pots were watered daily.

**Experimental procedures**. The experiment was conducted under greenhouse conditions at the Federal University of Jequitinhonha and Mucuri, MG. In this study, the additive design for competition studies was used, whereas weed densities varied, and corn density was kept constant (Swanton et al. 2015). The treatment design was a factorial with two weed species, *R. brasiliensis* and *C. benghalensis*, and five weed densities (0, 1, 2, 3, and 4 plants pot-1), in a completely randomized design with four replications.

Corn biomass was harvested at 60 DAE from each experimental unit. Shoot biomass were oven-dried at 65 °C until reaching constant weight, and dry weight recorded. The corn biomass (g) data (shoot) were converted into yield loss (%) compared with the corn control treatment (no weeds):

eq. (1)

where *Μ* is the mean biomass (g) of the untreated replicates, and is the biomass (g) of an individual treated experimental unit.

**Statistical Analysis**

Three models were fitted to yield loss data (%) in response to weed density (plants pot-1):

*Rectangular hyperbola model* proposed by (Cousens 1985):

eq. (2)

where *I* represent YL (yield loss) per unit weed density as *D* (density) approaches 0, and *A* represents YL as D approaches ∞ (or maximum expected yield loss). The rectangular hyperbola model was fitted using the *nls* function of R version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria).

*Logistic model* (four parameter log-logistic curve):

eq. (3)

where *c* is the (lower limit or YL at low weed density), *d* is the asymptote (upper limit or YL at high weed density), and *e* represents the weed density (weeds pot-2) that cause 50% yield loss (inflection point). The parameter *b* is the relative slope around the parameter *e*, and *D* is the number of weeds pot-1. Parameters for the sigmoidal model (four-parameter logistic) were estimated using the *drm* function of drc package (Ritz and Streibig 2005) in R software.

*Polynomial quadratic model* (second order):

eq. (4)

where *α* is the intercept in the y-axis (maximum YL in the absence of weed), *a* represents the slope of the model. The parameter *b* is the quadratic term of the model, and *D* is the number of weeds plot-1. The parameters for the polynomial quadratic equation were estimated using the *lm* function of R software.

**Top model selection to describe crop-weed competition.** The AIC criterion, which is indicated for non-nested model selection (Anderson 2008), was calculated as:

eq. (5)

where is the likelihood function and is the number of estimated parameters in the model. According to the AIC criterion, the top model has the lowest AIC value. AIC values for each model was estimated using the *AIC* command in R software.

**Model selection to evaluate weed competitiveness with the crop.** Assuming that rectangular hyperbola is the top model, the impact of *R. brasiliensis* and *C. benghalensis* on corn YL is accessed through the variance-ratio or F-test performed using equation [2] (Lindquist et al. 1996). This statistical procedure evaluates the difference of residual sum squares (RSS) of nested models (Werle et al. 2014). F-test is calculated as (CITATION):

eq. (6)

Where RSSFULL and RSSRED represent the minimized residual sum squares of the parameters estimated for the full (step 1) and reduced model (step 2, 3, or 4), respectively; dfFULL and dfRED represent the degrees of freedom of the full and reduced models, respectively. F-value greater than the F-critical value (P-value <0.05) indicates that two models are different, thus, the full model should be used. F-value smaller than the F-critical value indicates that two models are not different (P-value >0.05); therefore, a model with fewer parameters (reduced model) can be used to describe the data. When P-value >0.05 we fail to reject the null hypothesis and a reduced model should be used (no difference of *I* and *A* parameter values between weed species). However, if P-value <0.05, the null hypothesis is rejected and the full model should be used (different *I* and *A* parameter values for each weed species). The F-test principle for nonlinear regression analysis was calculated for each model using *nls* *ANOVA* command in R software (Ritz and Streibig 2008).

Four major steps need to be completed to compare the parameters using this method (see supplemental file):

First, fit Equation [2] to the data of each species individually (*R. brasiliensis* and *C. benghalensis*); this represents the full model, where four parameter values (*I* and *A* for each weed species) will be estimated.

Second, pool the data for both species (*R. brasiliensis* and *C. benghalensis*) and fit Equation [2]. This represents the reduced model, where two parameter values (*I* and *A* for both weed species combined) are estimated for the polled data. This step will allow testing the hypothesis that *I* and *A* do not vary between species, which means that both species compete similarly with corn. If the hypothesis is accepted (P-value>0.05), stop here. Otherwise, there are two more hypothesis to be tested.

Third, fit equation [2] setting a single parameter *I*, but different *A* parameter for each species. This is a reduced model and three parameters will be estimated. This step tests the second hypothesis, that weed species compete similarly at low densities (*I*), but different at higher densities (*A*).

Fourth, fit equation [2] setting a single parameter *A*, but different *I* parameters for each species. This is a reduced model, and three parameters will be estimated. This step tests the third hypothesis, that weed species compete similarly at higher densities (*A*), but different at low densities (*I*). Additional AIC was also performed for the nested model selection for confirming the F-test model selection.

**Goodness-of-fit of the models.** Root mean squared error (RMSE), model efficiency (ME), and R2 were calculated and used to test the goodness-of-fit of non-nested and nested models (Archontoulis and Miguez 2015; Mayer and Butler 1993; Roman et al. 2000):

eq. [7]

eq. [8]

*R2* eq. [9]

where RSS and RSTare the sums of squares for the residual and total, respectively; is the number of data points; is the number of model parameters; is the observed, is the predicted, and is the mean observed value. The ME values range from -∞ and 1, with values closer to 1 indicating better predictions (Werle et al. 2014). R2 values range from 0 to 1, and it was used only for the polynomial quadratic model, which is a form of linear regression.

**Results**

**Top model selection to describe crop-weed competition.** The retangular hyperbola model proposed by Cousens (1985) resulted in the lowest AIC (326.4), followed by a sigmoid model (333.3) and a polynomial quadratic model (333.9) (Table 1). The model woth the smallest value of AICc, and highest probability (AICw) was considered the “top model”or the best descriptor of the full reality given the set of candidate models and the data (Anderson, 2010). Therefore, the Cousens model was the top model to describe the corn yield loss (%) in response to both *R. brasiliensis* and *C. benghalensis* competition (Figure 2 and Table 2). In addition to AIC, the Cousens model had a low RSME, and high ME for both species, which indicate god fit of the model (Table 1). Four parameters were estimated using the Cousens model, *I* and *A* for *R. brasiliensis* and *C. benghalensis* (Table 2). However, there are biological issues with the parameter estimates, *I* and *A* of *C. benghalensis* were estimated over 100% (Table 2). The parameter *I* of *C. benghalensis* curve had a steep inclination, which is likely that at this pot size used, *C. benghalensis* is very competitive in corn. Therefore, bigger pots and lower *C. benghalensis* densities would have been necessary for such a study. The parameter *A* for *C. benghalensis* was over 100%, which is likely that the pots were too small that final constant yield was reached too fast with C. benghalensis. (Weiner and Freckleton 2010). The asymptote (upper limit) was not completely reached using four plants pot-1 in our experiment; therefore, parameter *A* was overestimated (Table 2). Despite this common issue, there is no lack of fit of parameters *I* and *A* estimated for *R. brasiliensis* and *C. benghalensis* (P<0.05).

According to AIC, the logistic model was the second best model to describe the data (Table 1). This model had the lowest RSME (12.5), but the lowest ME, 0.83 and 0.89 for *R. brasiliensis* and *C. benghalensis*, respectively. The maximum corn yield loss caused by the competition of *R. brasiliensis* and *C. benghalensis* (*d*) was 61.5% and 94.3%, respectively. The parameter *e* represents the weed density (plants pot-1) that caused 50% in yield loss (%) and was 1.3 and 0.76 plants pot-1 of *R. brasiliensis* and *C. benghalensis*, respectively. The common issue of fitting a sigmoid curve for crop-weed competition study is the systematic lower limit (*c*) lack of fit (Table 3). The P-value for parameter *c* is not significant (P>0.05); therefore, this parameter is similar to zero. Also, the standard error is bigger than *c* estimated values, which means that the sample mean is not close to the actual population mean.

Similar to the logistic model, the polynomial quadratic model provided the highest AIC (333.9), which means that this model was statistically the least appropriate for describing the data (Table 1). Nonetheless, by ME and R2, this model resulted in a good fit to the data. However, ME and R2 are not adequate for model selection. Even so, a quadratic model has a biologically implausible turning point for crop-weed studies (Figure 1B). The slope and the intercept parameters estimated from a polynomial quadratic model could potentially have biological meaning. However, the quadratic parameter have not. Also, in the present study, there is a lack of fit for the intercept of *R. brasiliensis* and *C. benghalensis* (P>0.05) (Table 4).

**Model selection to evaluate weed competitiveness with the crop.** According to AIC, the Cousens model was the top model to describe the data (Table 1). Thus, we conducted the F-test to evaluate whether *R. brasiliensis* and *C. benghalensis* competed similarly with corn. The F-test of the Cousens model for corn yield loss (%) indicated that a reduced model (P=0.28) with different parameter *I* (competition at low weed densities) and similar parameter *A* (competition at higher densities) was the best model to describe corn competition to *R. brasiliensis* and *C. benghalensis* (Table 5). According to the parameter estimates in the rectangular hyperbola, at low densities (*I*), corn yield loss was 39.1 and 235.4 % in competition to *R. brasiliensis* and *C. benghalensis*, respectively. However, at higher densities, *R. brasiliensis* and *C. benghalensis* compete similarly, and corn yield loss was 105.4% (Figure 5 and Table 6). AIC corroborates to F-test (Table 5). The model selected by the F-test (different I, but similar A) resulted in the lowest AIC of 325.7. The best fit was also demonstrated as this model provides the lowest RSME and the highest ME for *R. brasiliensis* and *C. benghalensis*.

**Discussion**

**Selecting the top (non-nested) model.** According to AIC criteria (Table 1), when compared to polynomial quadratic and logistic models, the rectangular hyperbola model was the best model to describe corn yield loss (%) in response to *R. brasiliensis* and *C. benghalensis* (Figure 2) competition. This model was also the best for describing corn leaf area, height, and stem diameter reduction in response to *R. brasiliensis* and *C. benghalensis* (data not shown). The rectangular hyperbola model provides two biological parameters (*I* and *A*) that are meaningful in agronomic terms and can help answer research questions in crop-weed competition studies (Table 2).

To understand the nature of crop-weed competition modeling, one needs to comprehend the concept of constant final yield (CFY). The CFY, well revised by (Weiner and Freckleton 2010), is described from low to high densities, whereas the relationship between total biomass per unit area and density is initially linear, but eventually reaches a constant biomass that remains constant. To use the rectangular hyperbola, CFY needs to be reached; otherwise, parameter estimates will not be statistically and biologically meaningful. For example, in our study, the CFY was not completely reached, and the parameter *A* for *R. brasiliensis* and *C. benghalensis* was over 100% (Table 2). A competition study that shows a linear relationship trend has not reached CFY (Figure 1A); therefore, it is likely that either the appropriated weed density for the study was not selected or there was no time enough for competition occur.

The problem with using the logistic model to describe crop-weed competition is that this model does not have a flexible inflection point (which is related to rate of change). One of the assumptions when using the logistic model is that the inflection point is always at 50% of the asymptotic size, which is clearly not the case with your data… Do some further review and find some citations to support this.

The logistic model does not seem to be appropriate to describe the data from crop-weed competition studies (Figure 3). The main reason is that this sigmoid curve of the logistic model is symmetric (S-shape) around the parameter *e* (Figure 1C), which the yield loss (%) or any response (*Y*-axis value) is forced to zero at lower densities. The logistic function assumes that yield loss (%) is near zero at lower weed densities. Though the logistic model is not recommended for competition studies, it is one of the most commonly used and appropriate models in other weed research topics. Logistic curves are extensively used for herbicide dose-response and critical time for weed removal (Knezevic and Datta 2015; Ritz et al. 2015). For example, in dose-response studies, the parameter *e* is meaningful and important for comparison of herbicide doses that control 50% of weed population (Knezevic et al. 2007; Oliveira et al. 2017).

The polynomial quadratic model had the highest AIC; therefore, it was the least relevant model (Table 1). This model does not provide any meaningful biological parameters that would improve the discussion, test hypothesis, and help researchers understand the results from crop-weed competition studies (Figure 4). The polynomial quadratic curve is symmetric around its maximum response value, which makes such response biologically unlikely in a crop-weed relationship (Figure 1B). For example, the maximum corn yield loss (%) is lower at four plants pot-1 than in three plants pot-1 (Figure 2). Therefore, a polynomial quadratic curve should not be encouraged to fit regression in competition studies.

In crop-weed competition studies, because of misleading R2 model selection, it is common to find multiple equations fitting response variables (Ferreira et al. 2015; Silva et al. 2015). For example, in a study of two weed species ( ) in competition with neotropical trees, two different models (one for each weed species) were used to describe the biomass (g) of *Senegalia polyphylla* (Monquero et al. 2015). Moreover, other models were used to describe additional response variables reported in the study. It becomes difficult to evaluate and compare the competitiveness of weeds when different equations with different parameters are used.

**Weed competition with corn model (nested) selection.** We rejected the null hypothesis that *R. brasiliensis* and *C. benghalensis* compete similarly in corn (Table 5). The F-test indicated that competition of *R. brasiliensis* was lower than *C. benghalensis* at low densities, but similar at higher densities (Figure 5).

A complete review of model parameter *I* and *A* of the rectangular hyperbola is provided by a (Cousens 1985). Also, this model is recommended for crop-weed studies in weed research (Ritz et al. 2015; Swanton et al. 2015). Many authors have used this model to answer their research questions and improve weed control decision-making (Fischer et al. 2004a; Lindquist et al. 1996; Lindquist et al. 1999; Werle et al. 2014). For example, using the rectangular hyperbola model, it was concluded the higher competitiveness of *Amaranthus palmeri* in corn and *Kochia scoparia* in sunflower (Lewis and Gulden 2014; Massinga et al. 2001). Parameters *I* and *A* are also useful for estimating weed competition across different locations and for calculating weed thresholds ((Fischer et al. 2004b; Lindquist et al. 1996). Additionally, using parameters *I* and *A,* it was demonstrated that organic cropping systems have the potential to tolerate great abundance of weeds compared to conventional system (Ryan et al. 2009). Thus, the rectangular hyperbola proposed by Cousens (1985) and the F-test nested model selection are important and useful tools in crop-weed competition research.

**Conclusions**

Here we demonstrated that the rectangular hyperbola proposed by Cousens (1985) was the best model to describe crop-weed competition data in additive design. We also showed potential parameter issues when using the Cousens model. Nonetheless, the Cousens (1985) model has a shape that fits well with the expect results of crop-weed competition studies. The parameters *I* and *A* are easily interpreted and biologically meaningful in agronomic terms. We propose Cousens (1985) as a standardized model for crop-weed competition studies in additive design. If the data trend seems linear, the experimental design needs to be adjusted by increasing the weed density to achieve the CFY. If CFY is immediately reached at low density treatments, density treatments need to be adjusted or experimental units (e.g., pots) increased in size. Sigmoid curves are recommended to other set of studies in weed research (e.g., herbicide dose-response), and polynomial quadratic curves are not recommended. We also showed that *C. benghalensis* and *R. brasiliensis* are very competitive with corn. In areas infested with *C. benghalensis*, weed management will increase, as *C. benghalensis* showed are competitive even at lower densities.

**Acknowledgments**

The authors thank CAPES (Brazilian Government Foundation) - Proc. no 9112-13-8, for financial support to the graduate student involved in this study.