A regression model to predict seed filling in ruzi grass (Brachiaria ruziziensis)

Edna Antony[1*], Dibeyendu Deb[2], K. Sridhar[1], Vinod Kumar[1] and Dharm Singh Meena[3]

[1]Southern Regional Research Station, ICAR-IGFRI, Dharwad-580005, India
[2]ICAR- Indian Grassland and Fodder Research Institute, Jhansi-284003, India
[3]ICAR-National Bureau of Plant Genetic Resources, New Delhi-110012, India
*Corresponding author e-mail: edna.antony@googlemail.com

Abstract

Brachiaria ruziziensis (congo signal or ruzi grass) is an important fodder crop in India. Cultivation of ruzi grass through seed is restraint due to a higher component of chaffy seeds in a seed lot. A number of filled seeds in a lot are an important seed quality parameter in ruzi grass. Currently, X-ray radiography and manual estimation of filled seeds are means to identify filled seeds in a seed lot. X-ray radiography method is an expensive and manual estimation of a number of filled seeds is tedious and time-consuming. Hence, a regression model was developed to estimate the number of filled seeds in a seed lot based on the weight of 100 seeds. Various regression diagnostics like standard residual plot, Normal Q-Q plot, Scale–location plot and Leverage plots were used to validate the model. A third-degree polynomial with 100 seed weight as a predictor was found to be the best fit to predict the number of filled seeds in a seed lot.

Keywords: *Brachiaria ruziziensis*, Filled seeds, Regression model

Brachiaria ruziziensis is an important fodder crop in India. The area under cultivation of this grass is not well documented in India even though it is widely cultivated in Kerala, parts of Karnataka and Goa. This fodder grass is well suited for coastal lands with high rainfall and humidity. However, its widespread adoption is affected due to poor seed quality and germination. Hence in India, root slips are most commonly used as planting material in *Brachiaria ruziziensis*. Nevertheless, root slips being bulky, its transportation to longer distance is cumbersome and expensive. During long-distance transport, drying of root slips could also cause problems in the establishment of the grass. Propagation through seeds is a better option. Nonetheless, availability of good quality filled seeds poses a real problem in India (Antony *et al.,* 2013a). It is known that the proportion of filled to unfilled

seeds depends on the method of harvest and seed collection (Antony *et al.*, 2017; Hare *et al.*, 2007a-c). The appropriate method of harvest depends on growth habit, synchrony of crop development, standing seeds, fallen seeds, availability of labour and on expertise. Several seed collection methods including single destructive harvest and sweating, multiple non-destructive harvests like tying in living sheaves and hand harvesting are also reported in *Brachiaria* (Kowithayakorn and Phaikaew, 1993; Phaikaew *et al.,* 1993; Loch and de Souza, 1999).

Even though grasses by nature produce a huge amount of seeds, many are not filled as there is no synchrony in flowering leading to staggered maturity (Loch and de Souza, 1999). The problem of predicting the stage for harvest is compounded by the weather factors during ripeness. Therefore, the chance of having varying proportions of filled and unfilled seeds in a seed lot is high. A number of filled seeds in a seed lot determine its quality and only filled seeds have the capacity to germinate. In *Brachiaria* there are no existing seed standards for defining the quality of seeds. Separation of filled seeds and unfilled seeds from the seed lot is very difficult due to light weight and small size of seeds. Presently, manual examination of seeds and soft X-ray radiography are the only options to estimate the number of filled seeds (Bahukandi *et al.,* 2013; Antony *et al.,* 2013b). The manual examination is very tedious and time-consuming due to the small size of seeds and X-ray radiography facility is expensive and rarely available. Another alternative is to estimate germination percentage. However, it was observed that *Brachiaria ruziziensis* seeds have only 10-20% germination and many un-germinated were hard and viable (Antony *et al.,* 2013a; 2017).

When estimating biological parameters like leaf area, leaf area index, root length, aphid count, tree height becomes time - consuming and expensive, biologists

always have depended and developed models to save time and money (Narayan *et al.*, 2014; Oner *et al.,* 2011; Ghoreishi *et al.,* 2012; Verghese *et al.,* 2002). Mathematical models are easy to use, save time and may not involve any expensive instruments (Welham *et al.*, 2014). Currently available techniques to estimate the number of filled seeds in ruzi seed lot is time-consuming and involves expensive equipment. In *Bromus tomentellus* (Jafri *et al.,* 2007) and in tall fescue (Akfar *et al.,* 2007), test weight was correlated to a number of seeds in a panicle. In this study, a regression model was tried to evolve for predicting the number of filled seeds in *Brachiaria ruziziensis* with 100 seed weight as a predictor.

The crop was grown following the recommended management practices at Southern Regional Research Station, ICAR-IGFRI, Dharwad, located at 15°12' N latitude and 76°34'E longitude with an altitude of 678 m above the mean sea level. The samples were collected from plants grown with and without fertiliser application, with and without hormonal treatment as well as from plants grown under irrigated and rainfed conditions for two years. This ensured that seeds samples represented a wide range of conditions. In this study, a single destructive harvest (SDH) method followed by shade drying and periodic collection of seeds (Antony *et al.,* 2017; Antony *et al.,* 2013b) was adopted. At maturity, single destructive harvest was carried out when seeds of initial flush (20% flowering) dropped and the inflorescence harvested were shade dried by tying upside down (Antony *et al.,* 2013b). Seeds were collected at periodic intervals by shaking the inflorescence at 2, 5, 8, 10, 12, 20 and 30 days after harvest (DAH). This method ensured not the only collection of uniformly matured seeds but also the separation of filled seeds from unfilled seeds. The seeds were dried to 10% moisture content. One hundred seeds were counted from each lot and weighed. The seed filling was estimated by manual threshing of seeds and verified by soft X-ray radiography (Bahukhandi *et al.,* 2017; Bahukhandi *et al.,* 2013).

All statistical data analysis was done on free open R-software (Venables and Smith 2004). For model fitting the total available dataset was divided into two parts; Two third of the total data was used for model fitting and rest one third was kept separated for model testing (Snee, 1977; Rizvi *et al.*, 2007). As here the count of filled seeds was the response variable and it was to be regressed on the variable 100 seed weight, so the correlation between 100 seed weight and a number of seeds filled

were worked out and graphically the relation between them was explored. To find a best-fit equation in order to explain the relationship between these two variables, different regression models were tried.

Various residual diagnostics like standard residual plot, Normal Q-Q plot, Scale –location plot and Leverage plots (Fox, 2008) were used to check if the selected model satisfies the assumptions of the regression equation. The model output was validated by taking seeds at random and the seed filling was estimated using the model output equation. The output was verified biologically using X-ray radiography. The dataset kept for model testing was used to test the best-selected model.

The method of seed collection using single destructive harvest followed by a periodic collection of seeds ensured better-filled seeds and well separation of good seeds from bulky unfilled seeds (Antony *et al.*, 2013a). Using this method of seed collection, seeds filling were in the range of 10-90% and their spread was normally distributed. The 100 seed weight data points had a mean of 0.45 with a standard error of 0.015 and standard deviation 0.22. Skewness and kurtosis were 0.65 and -1.2, respectively. Filled seed number had a mean value of 35.63 and with standard error 2.14 and standard deviation 30.24. Skewness and kurtosis were 0.42 and -1.5, respectively. The correlation between 100 seed weight and number of seeds was high and significant ($P<.0.01$). In many kinds of grasses, the test weight was correlated to the number of seeds (Akfar *et al.,* 2007; Jafri *et al.,* 2007).
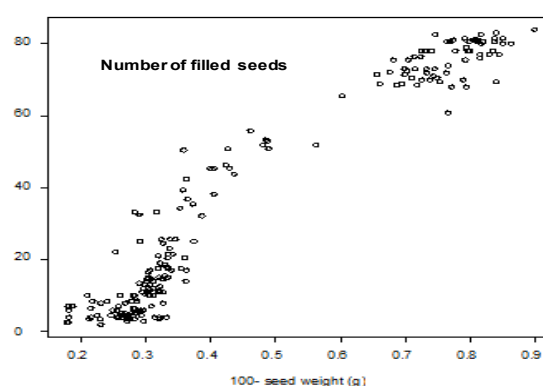
Fig 1. Scatter plot showing relationship between 100-seed weight and number of filled seeds

The relation between these two variables was graphically explored through a scatter plot (Fig 1). Applying simple linear regression (using R 3.1.2) between these two variables where number of seeds filled was the response

and 100 seed weight was the effect variable, the equation obtained (Fig 2) was as follows:

number of seeds filled = -24.07 + 130.48 * (100 seed weight) with R^2 = 0.93.

As the scatterplot revealed a curvilinear shape, polynomial equations were tried to see if it provided an improvement over the linear model. 2nd, 3rd and 4th order polynomials gave a better fit over the linear model with higher R^2 values as well as reducing residual standard errors (Table 1), respectively. The 4th polynomial provided a little improvement over 3rd, a polynomial in respect of both R^2 and residual standard error. It was also more unstable due to its complex nature. Therefore, the 3rd order polynomial was considered as the best fit to explain the present relation (Fig 3). The 3rd order polynomial model applied had the following equation:

$$Y = -13.98 - 3.08 * (X) + 431.9 * (X^2) - 355.49 * (X^3)$$

Where Y= number of seeds filled and X= 100 seed weight

All the coefficients of the models with their significances were recorded (Table 2). Here the null hypothesis was that 'the coefficients are equal to zero, if it appears significant then the null hypothesis is rejected which implies that the model is significant.
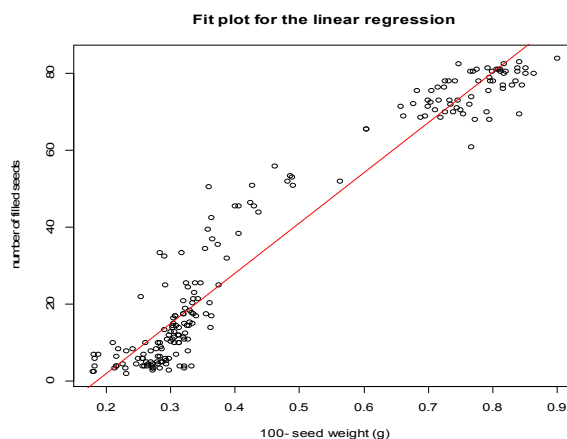
Table 1. Fit statistics of the models

Models	R^2	Residual standard error
Linear model	0.930	7.813
2nd order polynomial	0.942	7.183
3rd order polynomial	0.945	6.869
4th order polynomial	0.945	6.867

Fig 2. Fit plot for the linear regression
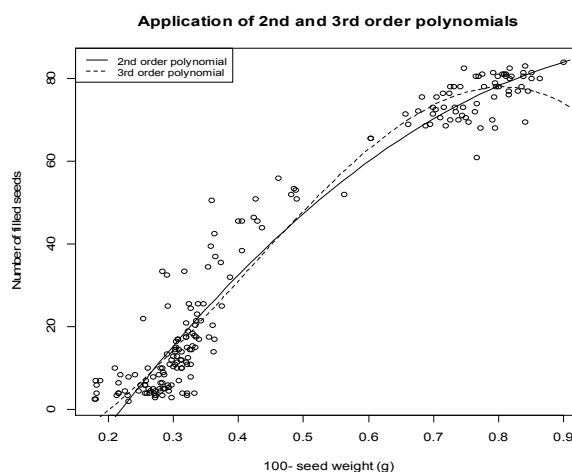
Fig 3. Polynomial models showing the relationship between 100 seed weight and number of filled seeds

Various regression diagnostics were applied on the best model selected (3rd order polynomial model, Fig 4) to test how good the model fit was achieved. The plots Fig
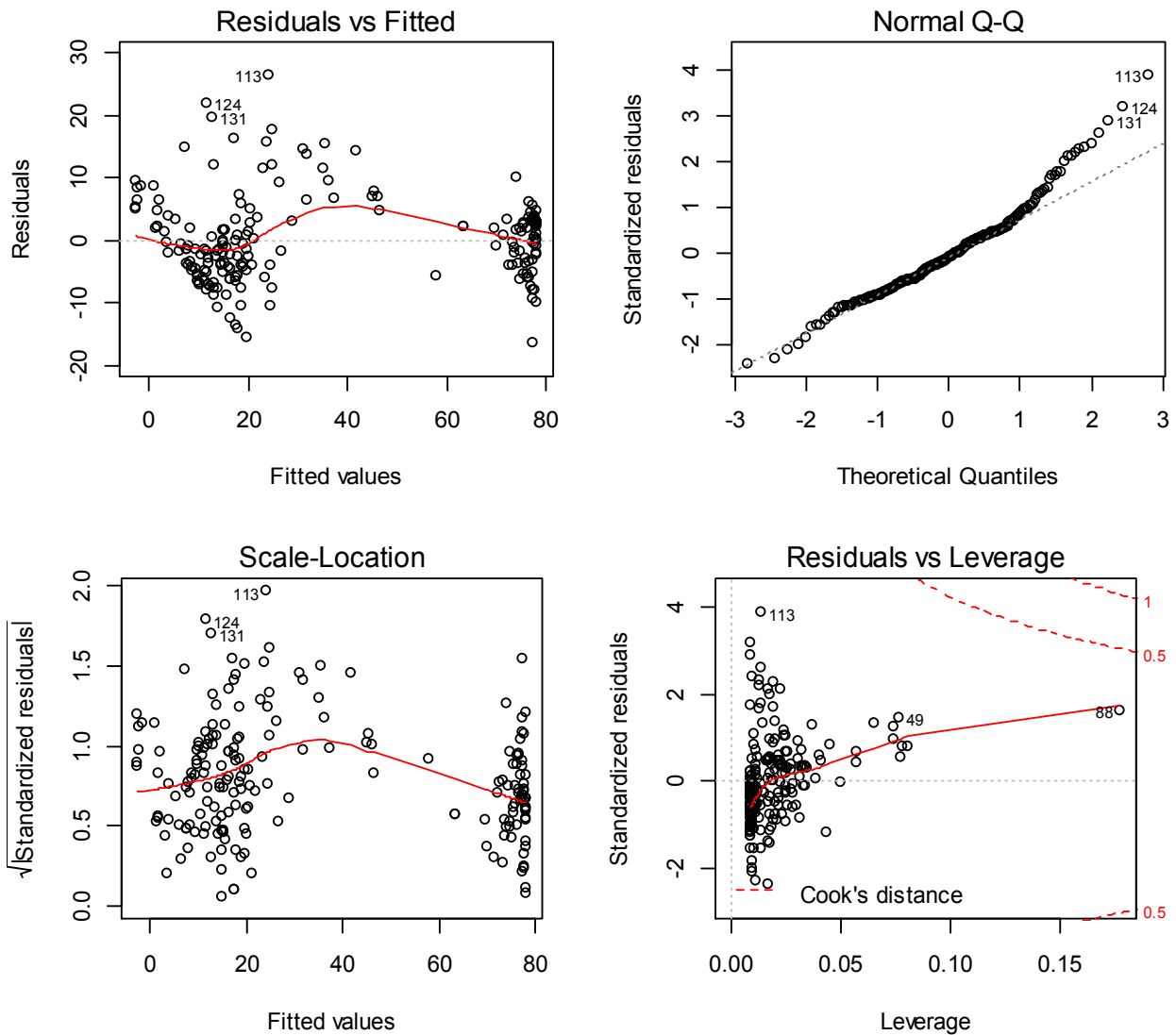
Table 2. Coefficients of the models with their significance

Equation	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
$y = a + b.x$	Intercept	-24.0660	1.2637	-19.04	0.0000**
	Seed weight	130.4849	2.4828	52.56	0.0000**
$y = a+b.x+c.x^2$	Intercept	-49.8131	4.3866	-11.36	0.0000**
	Seed weight	251.4339	20.0011	12.57	0.0000**
	I(seed_weight^2)	-114.2061	18.7626	-6.09	0.0000**
$y = a + bx + cx^2 + dx^3$	Intercept	-13.9808	9.1668	-1.53	0.1288
	Seed weight	-3.0831	60.9722	-0.05	0.9597
	I(seed_weight^2)	431.9001	125.5108	3.44	0.0007**
	I(seed_weight^3)	-355.4974	80.8642	-4.40	0.0000**

*(P< 0.05); **(P< 0.01)

Fig 4. Regression diagnostics

4a and 4b depicted standard residual plot showing residuals against fitted values. Points that tend towards being outliers like 113, 124, and 131 were also marked. As the residuals did not show any particular pattern, the error variance was constant and the assumption of homoscedasticity was satisfied which was also confirmed by studentized Breusch-Pagan test (Breusch and Pagan, 1979). The second plot (4b) illustrated that the residuals followed the line of theoretical quantiles and thus established the assumption of their normality. The labelled points on plot 4d represented cases with undue influence on the regression relationship. One-third data of the total available dataset kept separated for testing the model. So after validating the model, the test dataset was used for testing the model. It was obse-

-rved that the estimated values closely matched the observed ones (Fig 5).
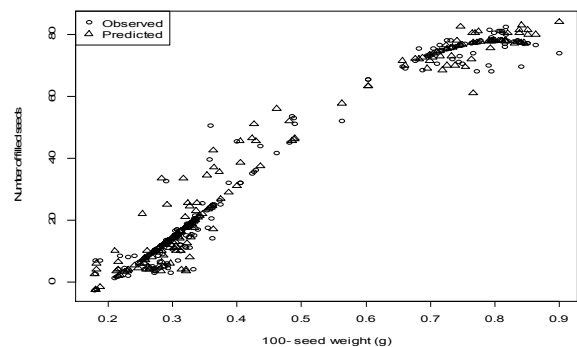
Fig 5. Estimated values and observed values of the number of filled seeds in *Brachiaria ruziziensis*

In this study, a regression model for predicting the seed quality of *Brachiaria ruziziensis* has been discussed. Using regression model saves time without using any expensive instruments and reduces the hard labour involved in predicting the seed quality. By virtue of its small size, grass seeds are difficult to handle. Hence mathematical equations can be used for estimating seed quality or filling percentage in grasses including other Bracharia species like *Bracharia decumbence, Brachiaria brizantha*. This study has practical utility for researchers involved in fodder research and extension.

Acknowledgement

References

Afkar, S., A. Jafari and G. Karimzadeh. 2007. Evaluation of seed yield and seed components in tall fescue (*Festuca arundinacea* Schreb.) through correlation, regression and path analysis. *Bioforsek Focus* 2: 39-43.

Antony, E., K. Sridhar and V. Kumar. 2017. Effect of chemical sprays and management practices on *Brachiaria ruziziensis* seed production. *Field Crops Research* 211: 19-26.

Antony, E., K. Sridhar and V. Kumar. 2013a. Flowering and seed filling pattern in *Brachiaria* species. In: *Proc National conference on Plant Physiology on Current Trends in Plant Biology Research (*13-16 December). Directorate of Groundnut Research, Junagadh. pp. 910-911.

Antony, E., K. Sridhar, and V. Kumar. 2013b. Shade drying of *Brachiaria ruziziensis* inflorescence minimises seed loss due to shedding. *IGFRI Newsletter* 19: 4-5.

Bahukhandi, D., N. Manjunatha, D. Vijay, and H.V. Singh. 2017. Effect of seed treatments and storage containers on seed quality of berseem (*Trifolium alexandrinum* L). *Range Management and Agroforestry* 38: 227-233.

Bahukhandi, D., D.R. Malaviya, S. Pradeep and H.C. Pandey. 2013. Detection of pest-pathogen in forage legume seeds with X-ray radiography. *Range Management and Agroforestry* 34: 119-121.

Breusch, T. S. and A.R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47: 1287-1294.

Fox, J. 2008. *Applied Regression Analysis and Generalized Linear Models*. 2nd edn. SAGE Publications Inc., Thousand Oaks, CA, US.

Ghoreishi, M., Y. Hossini and M. Maftoon. 2012. Simple models for predicting leaf area for Mango (*Mangifera indica* L*.). Journal of Biology and Earth Sciences* 2: B45-B53.

Hare, M.D., P.Tatsapong and K. Saipraset. 2007a. Seed production of two brachiaria hybrid cultivars in north-east Thailand. 1. Method and time of planting. *Tropical Grasslands* 41: 26-34.

Hare, M.D., P.Tatsapong and K. Saipraset. 2007b. Seed production of two brachiaria hybrid cultivars in north-east Thailand. 2. Closing date defoliation. *Tropical Grasslands* 41: 35-42.

Hare, M.D., P.Tatsapong and K. Saipraset. 2007c. Seed production of two brachiaria hybrid cultivars in north-east Thailand. 3. Harvesting method. *Tropical Grasslands* 41: 43-49.

Jafari, A.. B. Rouzbahani and E. Rahmani. 2007. Seed and herbage yield in Bromus tomentellous Boiss grown under optimal and drought stress conditions. *Bioforsek Focus* 2:33-38.

Kowithayakorn, K and C. Phaikaew. 1993. Harvesting and processing techniques of tropical grass and legume seeds for small farmers. In: *Proceedings of the XVII International Grassland Congress, Palmerston North and Rockhampton.* pp. 1809-1813.

Loch, D.S. and F.H.D. De Souza.1999. Seed harvesting and drying: grasses. In: D.S. Loch and J.E. Ferguson (eds). *Forage Seed Production, Volume 2: Tropical and Subtropical Species.* CAB International, Oxon, UK. pp. 191-212.

Narayanan, S., R. M. Aiken, P. V. VaraPrasad, Z .Xin, Ge Paul and J. Yu. 2014. A simple quantitative model to predict leaf area index in sorghum. *Agronomy Journal* 106: 219-226.

Oner, F., M. S. Odabas, I. Sezer and F. Odabas. 2011. Leaf area prediction for corn (*Zea mays* L.) cultivars with multi-regression analysis. *Photosynthetica* 49: 637-640.

Phaikaew, C., C. Manidool and P. Devahuti. 1993. Ruzi grass (*Brachiaria ruziziensis*) seed production in northeast Thailand. In: *Proceedings of the XVII International Grassland Congress. Palmerston North and Rockhampton.* pp. 1766–1767.

Rizvi, R. H., V.K Gupta, Ajith, S. P. Ahalwat and A. Datta. 2007. Preliminary growth models for *Dalbergia sissoo* ROB plantation in semiarid regions of central India. *Indian Journal of Forestry* 30: 421-425.

Snee, D. R. 1977. Validation of regression models: methods and examples. *Technometrics* 19: 415-428.

Venables, W.N. and D.M. Smith. 2004. *An Introduction to R: Notes on R, A Programming Environment for Data Analysis and Graphics, v. 2.0. 1.* Network Theory, Bristol, UK.

Verghese, A. and P. D. Kamala Jayanthi. 2002. A technique for quick estimation of aphid numbers in the field. *Current Science* 82: 1165-1168.

Welham, S.J., S.A. Gezan, S.J. Clark and Mead. 2014. *Statistical Methods in Biology: Design and Analysis of Experiments and Regression*. CRC Press. ISBN No.10: 1439808783.