CIÊNCIA e NATURA

ISSN 2179-460X
OPEN ACCESS

Ci. e Nat., Santa Maria, v. 47, e85223, 2025 • https://doi.org/10.5902/2179460X85223
Submitted: 09/28/2023 • Approved: 04/14/2025 • Published: 12/19/2025

Statistics

Generalized additive models for location, scale and shape in the analysis of common bean productivity

Modelos aditivos generalizados para locação, escala e forma na análise da produtividade do feijão comum

Momate Emate Ossifo[I,II], Luiz Ricardo Nakamura[I], Cesar Pedro[III], Daniel Furtado Ferreira[I], Alex de Oliveira Ribeiro[I]

[I] Universidade Federal de Lavras, Lavras, MG, Brasil
[II] Universidade Eduardo Mondllane, Vilankulos, Moçambique
[III] Instituto de Investigação Agrária de Moçambique, Nampula, Moçambique

ABSTRACT

The common bean (*Phaseolus vulgaris L.*) is a leguminous plant and one of the world's most important crops, with substantial economic relevance. Hence, the main aim of this paper is to analyze the productivity of common beans, establishing relationships with specific variables. For this purpose, the following candidate explanatory variables were considered: plant height, number of branches per plant, days to flowering, days to maturity, number of seeds per pod, number of pods per plant, and seed mass. Because of its flexibility in explaining the behaviour of a response variable, the generalized additive models for location, scale, and shape (GAMLSS) were used for statistical modelling. Initially, three distinct distributions for the response variable (productivity) were considered: the inverse gamma (IGAMMA), the generalized gamma (GG), and the inverse Gaussian (IG). The covariates for the regression structures were selected using the so-called Strategy A, a stepwise-based method. Based on both Akaike and Schwarz criteria, the GAMLSS based on the IG distribution was chosen as the best fit. The variables number of pods per plant and days to maturity had a positive significant effect on average productivity, whereas the number of branches per plant presents a negative effect on its variability. Based on a residual analysis, we can conclude that the fitted GAMLSS based on the IG distribution is appropriate to explain the data.

Keywords: Distributional regression models; Model selection; *Phaseolus vulgaris L.*

RESUMO

O feijão comum (*Phaseolus vulgaris L.*) é uma leguminosa que constitui uma das culturas mais importantes do mundo, com grande relevância económica. Assim, o principal objetivo deste trabalho é analisar a produtividade do feijão comum, relacionando-a com variáveis específicas. Para o efeito, foram consideradas as seguintes variáveis explicativas candidatas: altura da planta, número de ramos por planta, dias para a floração, dias para a maturação, número de sementes por vagem, número de vagens por planta e massa de sementes. Devido à sua flexibilidade em explicar o comportamento de uma variável resposta, os modelos aditivos generalizados para localização, escala e forma (GAMLSS) foram utilizados para a modelagem estatística. Inicialmente, foram consideradas três distribuições distintas para a variável resposta (produtividade): a gama inversa (IGAMMA), a gama generalizada (GG) e a gaussiana inversa (IG). As covariáveis para as estruturas de regressão foram selecionadas utilizando a chamada Estratégia A, um método baseado em stepwise. Com base nos critérios de Akaike e Schwarz, o GAMLSS baseado na distribuição IG foi escolhido como o melhor ajuste. As variáveis número de vagens por planta e dias até à maturação tiveram um efeito significativo positivo na produtividade média, enquanto o número de ramos por planta apresenta um efeito negativo na sua variabilidade. Com base numa análise residual, podemos concluir que os GAMLSS ajustado com base na distribuição IG é adequado para explicar os dados.

Palavras-chave: Modelos de regressão distribucional; Seleção de modelos; *Phaseolus vulgaris L.*

1 INTRODUCTION

The common bean (*Phaseolus vulgaris L.*), a legume of the Fabaceae family, is one of the world's most important primary crops, with significant socioeconomic importance. It can fix atmospheric nitrogen in symbiosis with rhizobia bacteria, being a low-cost protein source for people in developing countries such as those in Latin America and Africa (Schmutz et al., 2014; Jannat et al., 2019), including Mozambique, where both fresh and dried beans are important sources of nutrition and income for small-scale farmers (Pedro et al., 2022a).

Despite its significant socioeconomic importance, Mozambique's yield hovers around 533.8 kg ha^{-1} (Mader, 2021), which is considered low when compared to the average yield of some of the major producing countries such as China: 1744.1 kg ha^{-1}, United Republic of Tanzania: 1343.7 kg ha^{-1}, Myanmar: 911.6 kg ha^{-1} (FAOSTAT, 2020), and Brazil: 1056 kg ha^{-1} (CONAB, 2023). Several causes might be attributed to the lower yield, including insufficient agricultural techniques, abiotic, and biotic stressors, as well as the use of low-genetic-potential varieties (Pedro et al., 2022a).

However, cultivar yield can be increased by evaluating and selecting progenitors with high genotypic value, or by crossing divergent parents drawn from autochthonous common bean varieties cultivated over many years by small-scale farmers and locally adapted, differing in a variety of agronomic traits, and many of them are on the verge of extinction. Autochthonous varieties, according to Stoilova et al. (2013) and Abdollahi et al. (2016), are plant strains that have been domesticated from nature through both natural and artificial selection, helping farmers and agricultural programmes adapt to new challenges like climate change.

Traditional statistical regression models have been widely used to assess bean productivity e.g., Kakhki et al. (2022), phenotypic stability of genotypes e.g., Rosse; Vencovsky (2000), cultivar growth and yield e.g., Pereira et al. (2013), and disease incidence risk e.g., Harikrishnan; Del Ro (2008). However, depending on the complexity of the data under study, more flexible models capable of explaining not only the mean of the response distribution, but also all of its properties (distribution parameters), may be required Kneib (2013).

In this sense, the generalized additive models for location, scale, and shape (GAMLSS) have gained substantial theoretical and practical importance across a wide range of disciplines, e.g., precision agriculture (Righetto et al., 2019; Ossifo et al., 2024) and animal production (Nakamura et al., 2022; Roquim et al., 2023). Within the GAMLSS framework, any distribution parameter can be modelled as a function of explanatory variables, allowing for different regression structures chosen for each of them, i.e., several covariates might be chosen to explain each of the response distribution’s parameters.

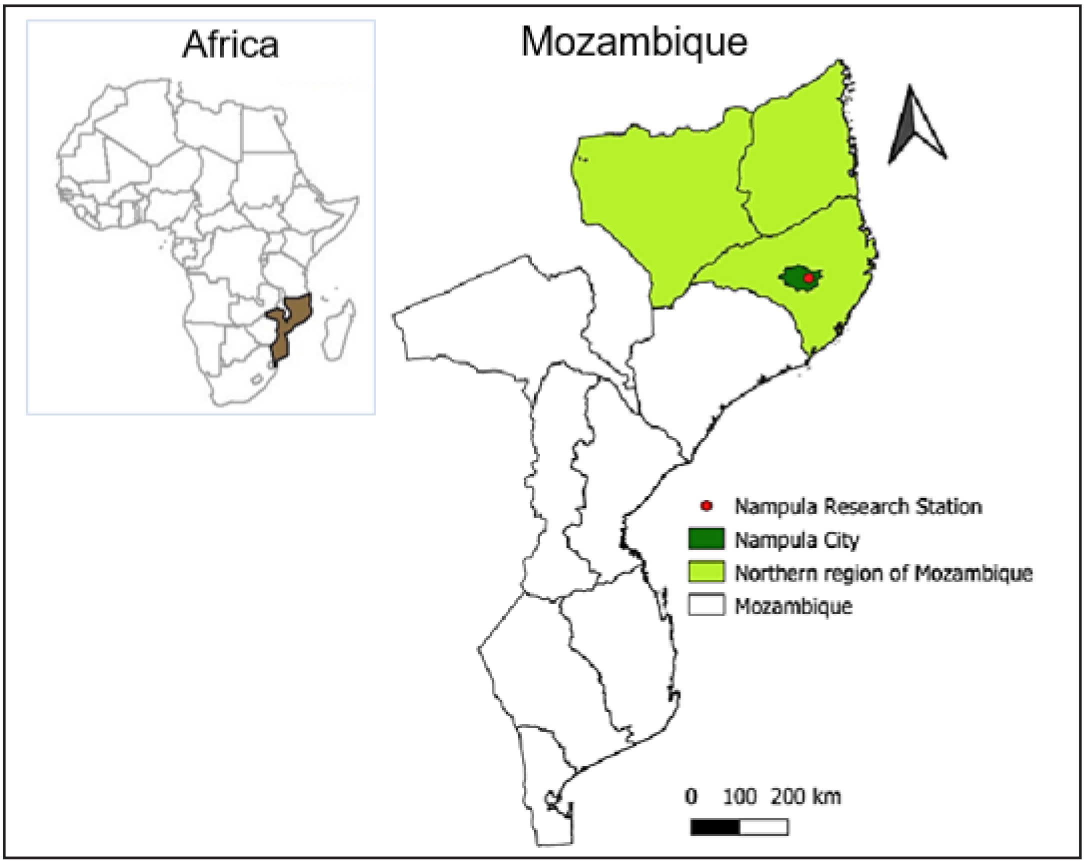
Given this highly flexible methodology, the main aim of this paper is to use the GAMLSS framework to model the productivity of common beans (*Phaseolus vulgaris L.*), relating it to potential candidate features: plant height, number of branches per plant, days to flowering, days to maturity, number of seeds per pod, number of pods per plant, and seed mass.

2 MATERIAL AND METHODS

2.1 The data

The experiment was carried out in the Agronomic Station of Nampula (15°09’01.8”S and 39°18’47.0”E) of the Mozambique Institute of Agricultural Research. Nampula is a district within the Province of Nampula, situated in the Northern Region of Mozambique, in the southeastern part of the African continent (Figure 1). The city of Nampula experiences a tropical climate classified as Aw according to the Köppen-Geiger climate classification. The annual precipitation is 959 mm, considering summers are much rainier than winters, and the average temperature is 23.9 °C.

Figure 1 – Location of the study area: Nampula district, Northern Region of Mozambique, in the Southeastern of the African Continent



Source: Source: Authorship (2023)

The characteristics evaluated were: plant height (PH; in cm), number of branches per plant (NBP), number of pods per plant (NPP), number of seeds per pod (NSP), days to flowering (DTF), days to maturity (DTM), seed mass (SM; in g), and the response, grain productivity (PROD, kg ha^{-1}) following the methodology outlined by IPGRI (2001). The experimental design employed was an augmented block design proposed by Federer (1956), consisting of four blocks.

Sowing was on 25 February 2021, with a seeding density of 10 seeds per linear metre in plots measuring 2 m in length and including two rows separated by 0.5 m. Crop management included two rounds of weeding, 15 and 30 days after the emergence of the plant. Two foliar fertilizations were performed using liquid Codafol NPK fertiliser (7-21-7) and two doses of the insecticide Cypermethrin (Hitcel 44EC) were provided for pest management during phenological stages V3 and V6. Harvesting, drying, threshing, and seed selection were all carried out within each plot. Following that, the weight of 100 seeds and seed mass per plot were calculated and converted into grain productivity. Further details are shown in Pedro et al. (2022).

2.2 The GAMLSS framework

GAMLSS is based on the assumption that the response variable (*Y*) may be represented by a distribution $D(\theta_k)$, where $\theta=(\theta_1,\ldots,\theta_p)^T$ represents the parameter vector of the distribution under consideration. The model is defined mathematically as follows:

$$g_k(\theta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^{J_k} s_{jk}(x_{jk}) \tag{1}$$

Where $g_k(\cdot)$, $k = 1, \ldots, p$, is a link function – e.g., if $\theta_k > 0$ then $g_k(\theta_k) = log(\theta_k)$ – X_k is a design matrix, β_k is a parameter vector, and $s_{jk}(\cdot)$ is a smoothing function (e.g., P-splines: Eilers; Marx, 1996) that explains the relationship between the covariate X_{jk} and the model's distribution parameter Rigby; Stasinopoulos (2005). If $\sum_{j=1}^{J_k} s_{jk}(x_{jk}) = 0$ then we have the full parametric GAMLSS Righetto et al. (2019)

The first stage in modelling common bean (*Phaseolus vulgaris L.*) productivity is to choose a distribution $D(\theta_k)$ that suits its behaviour well. In this context, marginal exploratory and residual analyses are often used, as Nakamura et al. (2017) emphasize. Given their superior marginal fit to bean productivity in a preliminary analysis, the following three candidate distributions capable of dealing with the asymmetric nature of the answer, as will be further discussed in Section 3, were considered in this study (Rigby et al., 2019): inverse gamma (IGAMMA), generalized gamma (GG), and inverse Gaussian (IG).

The IGAMMA probability density function (pdf) is written as follows

$$f(y|\mu,\sigma) = \frac{\mu^{\alpha}(\alpha+1)^{\alpha} y^{-(\alpha+1)}}{\Gamma(\alpha)} exp\left[\frac{-\mu(\alpha+1)}{y}\right] \quad (2)$$

Where $\mu > 0$ is the mode of the distribution and also a scaling parameter, $\sigma > 0$ is a dispersion parameter, $\alpha = 1/\sigma^2$ and $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$ is the gamma function

On the other hand, the GG pdf is given by

$$f(y \vee \mu,\sigma,\nu) = \frac{\nu \vee \theta^{\theta} y^{\nu\theta-1}}{\Gamma(\theta)\mu^{\nu}} exp\left(\frac{-\theta y^{\nu}}{\mu^{\nu}}\right) \quad (3)$$

Where $\mu > 0$ is a scaling parameter, $\sigma > 0$ is a dispersion parameter, $\nu > 0$ or $\nu < 0$ is a shape parameter and $\theta = 1/(\sigma^2\nu^2)$

Note that the IGAMMA distribution is a reparameterized special case of the GG distribution. Finally, the IG pdf can be written as

$$f(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2 y^3}} exp\left[\frac{-1}{2\mu^2\sigma^2 y}(y-\mu)^2\right] \quad (4)$$

Where $\mu > 0$ is the mean and $\sigma > 0$ is a dispersion parameter

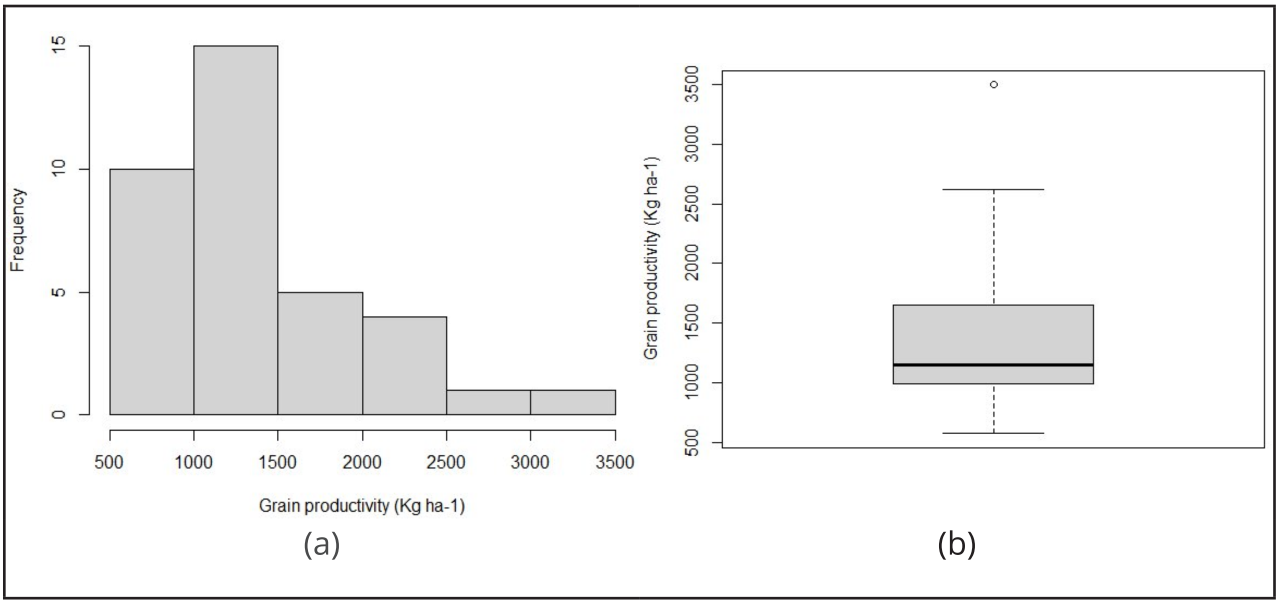
Since different explanatory variables can be included in any of the regression structures, several procedures for selecting variables for each of the parameters can be used Stasinopoulos et al. (2017). Among these, the most used is the so-called Strategy A Ramires et al. (2021), a stepwise approach-based method.

All fitted models are evaluated based on the generalized Akaike information criterion (GAIC) Voudouris et al. (2012), given by $GAIC(\kappa) = -2\hat{l}_p + \kappa \cdot df$, where $\hat{l}_p$ is the fitted total likelihood function, *df* is the total effective degrees of freedom and *k* is the penalty for each degree of freedom used. The lower the value of GAIC, the better the model fit. Note that if *k = 2* GAIC reduces to the Akaike information criterion (AIC) (Akaike, 1974) and *k = log(n)*, where *n* is the sample size, GAIC reduces to the Bayesian information criterion (BIC) Schwarz (1978).

After model fitting, we evaluate its assumptions using normalized quantile residuals Dunn; Smyth (1996), which are generally represented by worm plots (WP) Van Buuren; Fredriks (2001). Check Stasinopoulos et al. (2017) for further details.

3 RESULTS AND DISCUSSION

Figure 2 – Productivity of common bean (*Phaseolus vulgaris L.)*: (a) histogram and (b) box plot



Source: Authorship (2023)

The response variable marginal distribution (Figure 2) is positively skewed (skewness equals 1.47) and has heavy tails (leptokurtic), indicating a concentration of observations at its extremes (kurtosis equals 2.09). Based on these characteristics, the IGAMMA, GG, and IG distributions described in Section 2, are appropriate for modelling this response Rigby et al. (2019).

Figure 3 – Relationship between average productivity and explanatory variables: (a) plant height; (b) number of branches per plant; (c) number of pods per plant; (d) number of seeds per pod; (e) days to flowering; (f) days to maturity; and (g) seed mass (Continue)

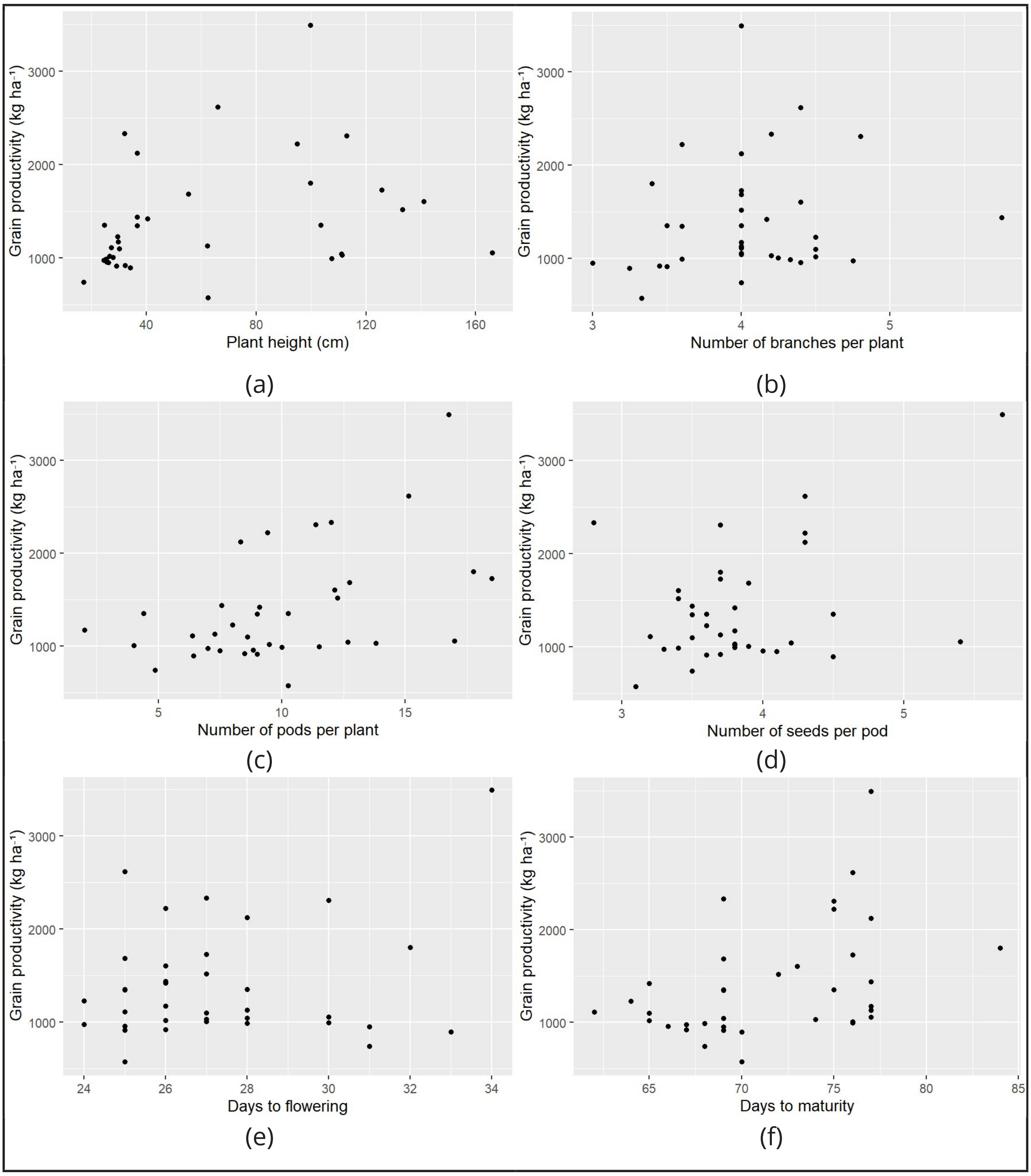
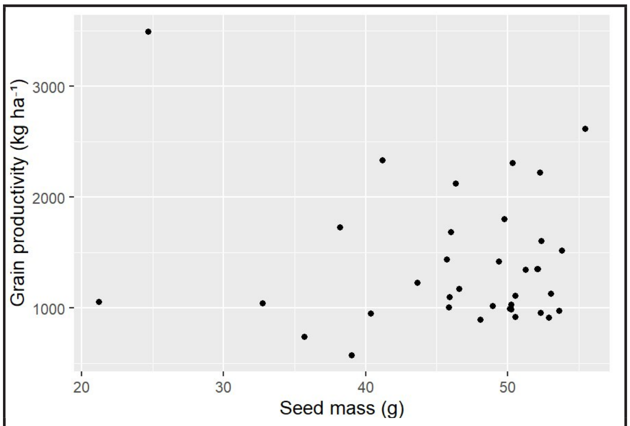


Figure 3 – Relationship between average productivity and explanatory variables: (a) plant height; (b) number of branches per plant; (c) number of pods per plant; (d) number of seeds per pod; (e) days to flowering; (f) days to maturity; and (g) seed mass

(Conclusion)



(g)

Source: Authorship (2023)

The relationship between productivity and each of the candidate explanatory variables is studied in Figure 3. Similar to Bisognin et al. (2019), positive correlations between the response variable and the covariates PH, NBP, NPP, and NSP are detected in Panels (a), (b), (c), and (d). This suggests that productivity is likely to be higher in taller plants with more branches, pods, and seeds. The apparent relationship between DTF and productivity (Panel e) is supported by Shahid and Kamaluddin (2013). They observed positive correlations between DTF and PROD in their assessment of the performance and relationship between variables contributing to autochthonous and exotic Rajmash bean materials under temperature conditions, indicating that this variable may have an impact on bean productivity. Similarly, Sofi et al. (2011) observed substantial differences in the days to flowering among common bean cultivars concerning their productivity.

Panel (f) displayed the relationship between DTM and PROD, with a higher variability within the 65-80 maturity range. Similar results were observed by Amanullah et al., (2006) with a range of 82-103 days for autochthonous varieties of common beans, and by Sofi et al., (2011) with a range of 77-124 days. Jannat et al., (2022) found a range of 68-115 days to maturity in their assessment of the biological diversity of autochthonous common bean types. The authors reported that early maturity of common beans is a requirement for cultivation preferences among agricultural communities, emphasizing the importance of days to maturity in crop appraisal.

The relationship between seed mass and PROD is displayed in Panel (g), where values typically exceed 45 g. This finding is consistent with the findings of Biscaro et al., (2009) who discovered greater correlations between SM and productivity following the administration of molybdenum by seed treatment in the cultivation of beans. This treatment increased leaf chlorophyll content, which resulted in increased seed mass and, as a result, increased productivity. Similarly, Fernandes et al., (2007) observed strong correlations between SM and PROD, claiming that foliar manganese delivery to the SM promoted bean productivity enhancement.

Following descriptive and exploratory analysis, three GAMLSS were fitted based on the IGAMMA, GG, and IG distributions. As described in Section 2.2, the Strategy was used to choose covariates for each regression structure. The goodness-of-fit measures obtained from these three regression models are shown in Table 1.

Table 1 – Goodness-of-fit measures for the three fitted GAMLSS

Model	AIC	BIC
Inverse Gaussian	537.10	545.01
Inverse gamma	541.40	549.37
Generalized gamma	545.22	553.14

Source: Authorship (2023)

According to Table 1, the model based on the IG distribution presented the lowest AIC (537.10) and BIC (545.01) values, hence it was chosen as the best fit for the data. Table 2 shows the variables selected for both parameters (μ and σ) of the inverse Gaussian (IG) distribution in the GAMLSS framework, as well as their estimates, standard errors, t-test values, and p-values.

Table 2 – Estimates for μ and σ, standard errors, t-test, and p-values, for the fitted GAMLSS based on the inverse Gaussian distribution

	μ (mean) – logarithm link function			
	Estimates	**Standard error**	**t**	**p-values**
Intercept	4.825	0.662	7.291	0.001*
Number of pods per plant	0.053	0.016	3.397	0.002*
Days to maturity	0.027	0.009	3.029	0.005*
	σ (shape parameter) – logarithm link function			
	Estimates	**Standard error**	**t**	**p-values**
Intercept	-1.863	1.464	-1.273	0.213
Number of branches per plant	-0.738	0.365	-2.025	0.052*

Source: Authorship (2023)
*Significant at the 10% level

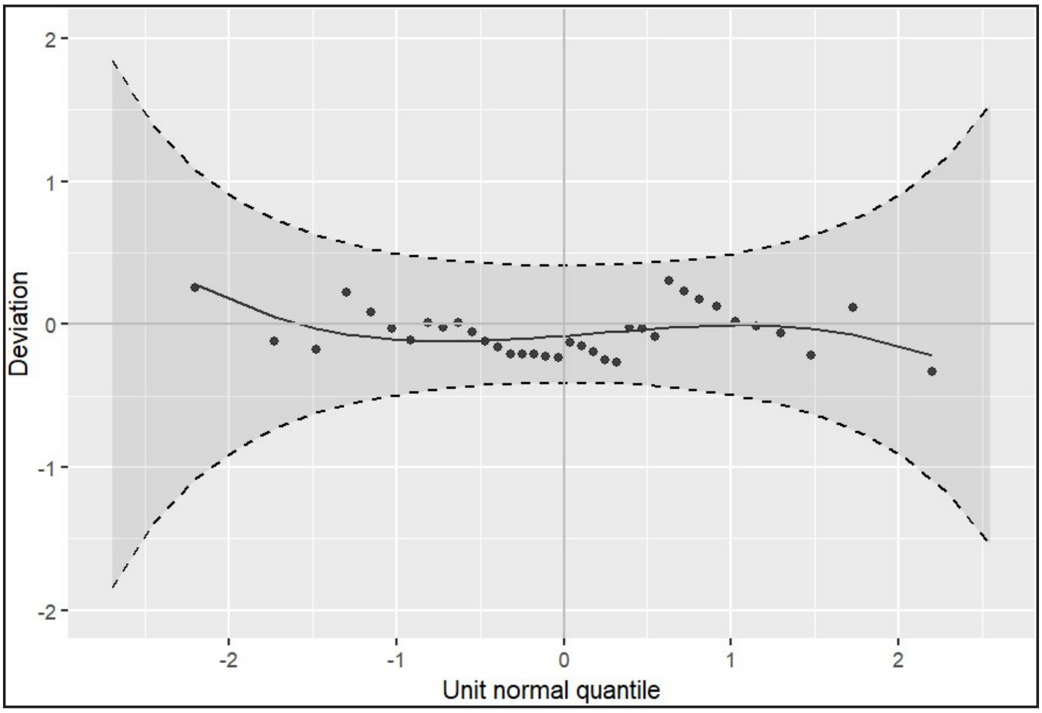
Table 2 reveals that only two covariates were considered in the μ structure and only one in the σ structure, where all of the terms are significant at the 10% level. The absence of the other four candidate covariates may be explained by the sophisticated statistical methodology we are using, which can model both of the response distribution parameters, i.e. beyond the mean regression.

Based on the fitted model, for each additional pod per plant, there is a multiplicative increase factor of *exp(0.053) = 1.0544* in the average of common bean productivity, i.e., an increase of 5.44% in the average of common bean productivity (kg ha^{-1}). Ribeiro et al., (2001) identified similar results when they observed positive correlations between the number of pods per plant and productivity in carioca beans, indicating that increasing NNP influences PROD enhancement. Zilio et al., (2011) also observed a positive correlation between productivity and NNP in three Santa Catarina municipalities during the 2008/2009 crop season.

Furthermore, for each extra day to maturity, there is an expected multiplicative increase factor of *exp(0.27) = 1,0274* in the common bean productivity (kg ha^{-1}), i.e., for each extra DTM there is an increase of 2.74% in productivity. This result corroborates the findings of Correa et al. (2012), who discovered a statistically significant positive correlation between DTM and PROD, although of minor significance, indicating that cultivars with later cycles are more productive. Machado et al., (2008) observed positive and statistically significant correlations, with higher magnitudes, between both variables.

Regarding the dispersion parameter σ, only the number of branches per plant was selected in its structure. For each additional branch per plant, there is a multiplicative decrease factor of *exp*(-0.738)=0.478 in the productivity's dispersion, i.e., a decrease of 52.2% in the response's dispersion. Arevalo et al., (2020) identified negative correlations between NBP and NSP (-0.63) and between NBP and NPP (-0.59), implying that plants with more branches produced fewer grains per pod as well as fewer pods per plant. This phenomenon may have an impact on the variability of productivity.

Figure 4 – Normalized quantile residuals obtained from the fitted GAMLSS based on the inverse Gaussian distribution



Source: Authorship (2023)

To assess the adequacy of the fitted model, Figure 4 displays the worm plot produced using normalized quantile residuals. It is clear that all residuals fall inside the 95% confidence bands, indicating that no inadequacies or significant deviations from the model's assumptions have been detected, i.e., the model was able to deal with the observed marginally high positive skewness and excess kurtosis of the response (Table 1). In other words, the fitted GAMLSS based on the IG distribution is appropriate to explain the common beans (*Phaseolus vulgaris* L.) productivity.

4 CONCLUSIONS

The generalized additive models for location, scale, and shape (GAMLSS) were appropriate for modelling the productivity of the common bean (*Phaseolus vulgaris* L.). Among the fitted models based on the three initially considered distributions, using the so-called Strategy A, the inverse Gaussian (IG) distribution was chosen to represent the response variable, demonstrating its adequacy for explaining the dataset. The variables number of pods per plant (NPP) and days to maturity (DTM) had significant effects on the mean parameter (μ), where for each NPP and DTM, an increase of 5.39% and 2.71%, respectively, in the average productivity of common beans (kg ha^{-1}) is expected. Furthermore, for the dispersion parameter, the variable number of branches per plant (NBP) was incorporated in its structure, revealing that for each NBP, an estimated decrease of 52.2% in response variability can be observed. The worm plot obtained using normalized quantile residuals revealed that the fitted GAMLSS is appropriate for explaining the dataset under consideration. Finally, the applied approach holds promise for future research into other varieties of crops.

ACKNOWLEDGEMENTS

To National Council for Scientific andTechnological Development (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for financing, in part, this study (Finance Code 001).

REFERENCES

Abdollahi, A.; Tahmasebpour, B.; Dehghanian, H. Factor analysis of Phonological and Morphological Traits in Common bean (*Phaseolus vulgaris* L.). *Biological Forum-An International Journal*, v. 8, n.1, p. 132–134, 2016.

Akaike, H. A new look at the statistical model identification.IEEE *Transactions on Automatic Control*. Notre Dame, v.19, n.6, p.716-723, 1974.

Amanullah, K. A.; Nawab, K.; Sohail, Q. Performance of promising common bean (*Phaseolus vulgaris* L.) germplasm at Kalam-Swat. Pakistan. *J. Biol. Sci*. v. 9, n. 14, p. 2642–2646, 2006.

Arevalo, A. C. M.; Cardoso, D. L.; Kraeski, M. J.; Santana D. C.; Arguelho S. G. Parâmetros genéticos, correlações e componentes principais para caracteres agronômicos em genótipos de feijoeiro comum do grupo carioca. *Research, Society and Development*, v. 9, n. 11, e3179119831, 2020.

Biscaro, G. A.; Junior, S. A. R. G.; Soratto, R. P.; Júnior, N. A. F.; Motomiya, A. V. A. M.; Filho, G. C. C. Molibdênio via semente e nitrogênio em cobertura no feijoeiro irrigado em solo de cerrado. *Ciênc. agrotec.*, v. 33, n. 5, p. 1280-1287, 2009.

Bisognin, M. B.; Pias, O. H. C.; Vian, A. L.; Basso, C. J.; Santi, A. L. Seed spacing variability reduces common bean yield. *Pesq. Agropec. Trop., Goiânia*, v. 49, e55134, 2019.

CONAB. 2023. *Acompanhamento da Safra Brasileira, Feijão,2022/23*, 4° levantamento. Available at: https://www.conab.gov.br. Accessed on June 2, 2023.

Correa, A. M.; Ceccon, G.; Correa, C. M. A.; Delben, D. S. Estimativas de parâmetros genéticos e correlações entre caracteres fenológicos e morfoagronômicos em feijão-caupi. *Rev. Ceres, Viçosa*, v. 59, n.1, p. 88 – 94, 2012.

Dunn, P.k.; Smyth, G.k. Randomized quantile residuals. *J. Comput. Graph Stat.* v. 5, p. 236–244, 1996.

Eilers, P.h.c.; Marx, B.d. Flexible smoothing with B-splines and penalties. *Statistical Science,* v. 11, p. 89–121 1996.

FAOSTAT-Food and Agriculture Organization Corporate Statistical Database. 2020. Available at: https://www.fao.org/faostat/en/. Accessed on July 28, 2023.

Federer, W.t. *Augmented (or Hoonuiaku) Designs*. Hawaiian Planters' Record, v. 55, p. 191-208, 1956.

Fernandes, D. S.; Soratto, R. P.; Kulczynski, S. M.; Biscaro, G. A.; Dos Reis, C. G. Produtividade e qualidade fisiológica de sementes de feijão em conseqüência da aplicação foliar de manganês. *Pesq. agropec. bras*., v.42, n.3, p.419-426, 2007.

Harikrishnan, R.; Del Río, L. E. A Logistic Regression Model for Predicting Risk of White Mold Incidence on Dry Bean in North Dakota. *Plant Disease*, v. 92, n. 1, 2008.

IPGRI. *Descritores para Phaseolus vulgaris.* International Plant Genetic Resources Institute, Rome. 2001.

Jannat, S.; Shah, A. H.; Hassan, M.; Sher, A.; Fiaz, S.; Elesawy, B. H.; Ismail, K. A.; El Askary, A.; Gharib, A. F.; Qayyum, A. Genetic diversity of common bean (*Phaseolus vulgaris* L.) ecotypes from Pakistan using Simple Sequence Repeats. *Saudi Journal of Biological Sciences*, v. 29, 2022.

Jannat, S.; Shah, A. H.; Sabir, S. M. Nutraceutical characterisation of common bean (*Phaseolus vulgaris* L.) germplasm from Pakistan. *International Food Research Journal*, v. 26, n. 2, p. 1835–1843, 2019.

Kakhki, S. H. N.; Taghaddosi, M. V.; Moini, M. R.; Naseri, B. Predict bean production according to bean growth, root rots, fly and weed development under different planting dates and weed control treatments. *heliyon journal,* v. 8, n. 11, 2022**.**

Kneib, T. Beyond mean regression. *Statistical Modelling*, v.13, p. 275–303, 2013.

Machado, C. F.; Teixeira, N. J.; Filho, F. R. F.; Rocha, M. R.; Gomes, R. L. F. Identificação de genótipos de feijão-caupi quanto à precocidade, arquitetura da planta e produtividade de grãos. *Revista Ciência Agronômica*, 39:114-123, 2008.

MADER-Ministério da Agricultura e Desenvolvimento Rural. 2021. *Inquérito Integrado Agrário 2020*. Maputo, Mocambique. Available at: https://www.agricultura.gov.mz/wp_content/uploads/2021/06/MADER_Inquerito_Agrario_2020.pdf. Accessed on June 20, 2023.

Nakamura, L. R.; Ramires, T. G.; Righetto, A. J.; Pescim, R. R.; Roquim, F. V.; Savian, T. V.; Stasinopoulos, D. M. Cattle reference growth curves based on centile estimation: A GAMLSS approach*. Computers and Electronics in Agriculture,* v. 192, p. 106572, 2022.

Nakamura, L. R.; Rigby, R. A.; Stasinopoulos, D. M.; Leandro, R. A.; Villegas, C.; Pescim, R. R. Modelling location, scale and shape parameters of the Birnbaum-Saunders generalized t distribution. *Journal of Data Science*, v. 15, p. 221–237, 2017.

Pedro, C.; Donça, M. C. B; Somueque, S. I.; Dique, J. E. L.; Bambo, E. C.; Colial; H. V.; Alexandre, D. C.; Gimo, S. T.; Divage, B. A. F.; Rico, D. M. R.; Barbosa, I. V.; Felizmino, V.; Serrote, R. I.; Moisés, Y. A. A.; Amós, N. R. Variability, Similarity Network and Genotypic Path Analysis of Common bean yield traits. *FPBJ - Scientific Journal*, v. 4, n. 2, p. 13–22, 2022a.

Pedro, C.; Donça, M. C. B.; Colial, H. V.; Dique, J. E. L.; Bambon, E. C.; Somueque, S. I.; Divage, B. A. F.; Rico, D. M. R.; Barbosa, I. P. Morpho-agronomic genetic diversity in common bean landraces based on blup values. *FPBJ - Scientific Journal,* v. 5, n. 3, 2022b.

Pereira, R. F.; Cavalcante, S. N.; Lima, A. S.; Filho, F. C. F. M; Santos, J. G. R. Crescimento e rendimento de feijão vigna submetido à adubação orgânica. *Revista Verde (Mossoró – RN - Brasil)*, v. 8, n. 3, p. 91 - 96, 2013.

Ramires, T. G.; Nakamura, L. R.; Righetto, A. J.; Pescim, R. R.; Mazucheli, J.; Rigby, R. A.; Stasinopoulos, D. M. Validation of Stepwise-Based Procedure in GAMLSS. *Journal of Data Science*, v. 19, n. 1, p. 96–110, 2021.

Ribeiro, N. D.; Mello, R. M.; Dalla, R. C.; Sluszz, T. Correlações genéticas de caracteres agromorfológicos e suas implicações na seleção de genótipos de feijão carioca. *Revista Brasileira de Agrociência*, v. 7, n. 2, p. 93-99, 2001.

Rigby, R. A.; Stasinopoulos, D. M. Generalized additive models for location, scale and shape (with discussion). *Appl. Statist.* v. 54, p. 507–554, 2005.

Rigby, R. A.; Stasinopoulos, D. M.; Heller, G. Z.; De Bastiani, F. *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R.* Chapman and Hall/CRC, 2019.

Righetto, A. J.; Ramires, T. G.; Nakamura, L. R.; Castanho, P. L. D. B.; Faes, C.; Savian, T. V. Predicting weed invasion in a sugarcane cultivar using multispectral image. *Journal of Applied Statistics,* v. 46, p. 1–12, 2019.

Rosse, L. N.; Vencovsky, R. Modelo de regressão não-linear aplicado ao estudo da estabilidade fenotípica de genótipo de feijão no estado do paraná. *Bragantia, Campinas*, v. 59, n. 1, p. 99 – 107, 2000.

Schmutz, J.; Mcclean, P. E.; Mamidi, S.; Wu, G. A.; Cannon, S. B.; Grimwood, J.; Jenkins, J.; Shu, S.; Song, Q.; Chavarro, C.; Torres-Torres, M.; Geffroy, V.; Moghaddam, S. M.; Gao, D.; Abernathy, B.; Barry, K.; Blair, M.; Brick, M. A.; Chovatia, M.; Gepts, P.; Goodstein, D. M.; Gonzales, M.; Hellsten, U.; Hyten, D. L.; Jia, G.; Kelly, J. D.; Kudrna, D.; Lee, R.; Richard, M. M. S.; Miklas, P. N.; Osorno, J. M.; Rodrigues, J.; Thareau, V.; Urrea, C. A.; Wang, M.; Yu, Y.; Zhang, M.; Wing, R. A.; Cregan, P. B.; Rokhsar, D. S.; Jackson, S. A. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, v. 46, n. 7, p. 707–713, 2014.

Schwarz, G. Estimating the dimension of a model. *Ann. Stat.,* v. 6, n. 2, p. 461–464,1978.

Shahid, A.; Kamaluddin, S. Correlation and path analysis for agro-morphological traits in rajmash beans under Baramulla- Kashmir region. *African Journal of Agricultural Research,* v. 8, n. 18, pp. 2027-2032, 2013.

Sofi, P. A.; Zargar, M. Y.; Debouk, D.; Graner, A. Evaluation of common bean (*Phaseolus vulgaris* L.) germplasm under temperate conditions of Kashmir Valley. *J. Phytol*., n. 3, v. 8, p. 47-52, 2011.

Stasinopoulos, D. M.; Rigby, R. A.; Heller, G. Z.; Voudouris, V.; De Bastiani, F. *Flexible Regression and Smoothing: Using GAMLSS in R*. Chapman and Hall/CRC, 2017.

Stoilova, T.; Pereira, G.; Tavares-De-Sousa, M. Morphological characterization of a small common bean (*Phaseolus vulgaris* L.) collection under different environments. *Journal of Central European Agriculture*. v. 14, n. 3, p. 1–11, 2013.

Van Buuren, S.; Fredriks, M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in medicine,* v. 20, p. 1259 – 1277, 2001.

Voudouris, V.; Gilchrist, R.; Rigby, R.; Sedgwick, J.; Stasinopoulos, D. Modelling skewness and kurtosis with the BCPE density in GAMLSS. *Journal of Applied Statistics*, v. 39, n. 6, p. 1279–1293, 2012.

Zilio, M.; Coelho, C. M. M.; Souza, C. A.; Santos, J. C. P.; Miquelluti, D. J. Contribuição dos componentes de rendimento na produtividade de genótipos crioulos de feijão (*Phaseolus vulgaris* L.). *Rev. Ciênc. Agron.,* v. 42, n. 2, p. 429 – 438, 2011.

Authorship contributions

1 – Momate Emate Ossifo (Corresponding Author)

Student, Undergraduate in Statistics and Agricultural Experimentation
http://orcid.org/0000-0002-5425-4087 • momateemate@gmail.com
Contribution: Conceptualization, Methodology, Validation, Data Curation, Formal Analysis, Investigation, Writing – Original Draft

2 – Luiz Ricardo Nakamura

Professor, Doctor in Statistics and Agricultural Experimentation
https://orcid.org/0000-0002-7312-2717 • luiz.nakamura@ufla.br
Contribution: Validation, Data Curation, Formal Analysis, Investigation, Supervision, Writing – Review & Editing

3 – Cesar Pedro

Professor, Doctor in Plant Breeding
https://orcid.org/0000-0002-2963-8652 • cpedroscientia@gmail.com
Contribution: Data Curation, Formal Analysis, Writing – Original Draft

4 – Daniel Furtado Ferreira

Professor, Doctor in Plant Breeding and Postdoctoral in Statistics
https://orcid.org/0000-0002-4371-5239 • danielff@ufla.br
Contribution: Supervision, Writing – Review & Editing

5 – Alex de Oliveira Ribeiro

Professor, Doctor in Statistics and Agricultural Experimentation
https://orcid.org/0000-0003-1499-5048 • alex.ribeiro@ufla.br
Contribution: Supervision, Writing – Review & Editing

How to quote this article

Ossifo, M. E., Nakamura, L. R., Pedro, C., Ferreira, D. F. & Ribeiro, A. O. (2025). Generalized additive models for location, scale and shape in the analysis of common bean productivity. *Ciência e Natura*, Santa Maria, 47, e85223. DOI: https://doi.org/10.5902/2179460x85223.