Non-linear regression models in the management of accumulated production of parchment coffee in Peru

Modelos de regressão não linear na gestão da produção acumulada de café pergaminho no Peru

Modelos de regresión no lineal en el manejo de la producción acumulada de café pergamino en Perú

Diana Del Rocío Rebaza Fernández
Natiele de Almeida Gonzaga
Marcelo Ângelo Cirillo
Joel Augusto Muniz

Abstract

Parchment coffee results from washing the coffee cherry, and its production has achieved a significant increase in the coffee-growing regions of Peru. Knowing the production pattern of this grain is essential to help coffee producers make decisions in the economic and social sector. As growth curves generally have sigmoidal behavior, which is well fit by non-linear models, this study aimed to model the cumulative production pattern of parchment coffee as a function of time (in months) in the year 2022, comparing the fit of the non-linear Logistic, Gompertz and von Bertalanffy models. The cumulative national production, and production of the departments of Huánuco and San Martín, in Peru, were analyzed. Data used to fit the models were obtained from the Ministry of Development and Irrigation (MIDAGRI) of Peru. To check the assumptions of normality, homoscedasticity, and independence of residuals, the

1 Magister Scientiae em Applied Statistics, Universidade Federal de Lavras (UFLA), Universidade Nacional Agraria La Molina (UNALM), Instituto de Ciências Exatas e Tecnológicas, Lavras, Minas Gerais, Brazil. E-mail: dianarebaza@lamolina.edu.pe Orcid: https://orcid.org/0000-0002-6105-5588
2 Master in Statistics and Agricultural Experimentation, Universidade Federal de Lavras (UFLA), Instituto de Ciências Exatas e Tecnológicas, Lavras, Minas Gerais, Brazil. E-mail: natiele.gonzaga1@estudante.ufla.br Orcid: https://orcid.org/0000-0002-4916-9056
3 Doctor in Statistics and Agricultural Experimentation, Universidade Federal de Lavras (UFLA), Instituto de Ciências Exatas e Tecnológicas, Lavras, Minas Gerais, Brazil. E-mail: macufla@ufla.br Orcid: https://orcid.org/0000-0003-2026-6802
4 Doctor in Statistics and Agricultural Experimentation, Universidade Federal de Lavras (UFLA), Instituto de Ciências Exatas e Tecnológicas, Lavras, Minas Gerais, Brazil. E-mail: joamuniz@ufla.br Orcid: https://orcid.org/0000-0002-1069-4136
Non-linear regression models in the management of accumulated production of parchment coffee in Peru

Shapiro-Wilk, Breusch-Pagan, and Durbin-Watson tests were used, respectively. The model parameters were estimated using the least squares method using the Gauss-Newton algorithm in the R software. The goodness-of-fit of the models was tested using goodness-of-fit measures such as Coefficient of Determination ($R^2$), Residual Standard Deviation (RSD), Akaike Information Criterion (AIC), and nonlinearity measures. Based on the models’ goodness-of-fit measures, the Gompertz model with a first-order autoregressive error term (AR1) fit best to national production data, and the Logistic model was the most suitable for describing the production of the departments of Huánuco, and San Martín.

Keywords: Production. Coffee. Logistic. Gompertz. Von Bertalanffy.

Resumo

O café pergaminho é um produto como resultado de um processo da lavagem do fruto do cafeeiro e sua produção tem alcançado um aumento significativo nas regiões cafeeiras no Peru. Conhecer o padrão na produção deste grão é fundamental para auxiliar os produtores de café na tomada de decisões, no setor econômico e social. Como geralmente as curvas de crescimento apresentam comportamento sigmoidal, no qual é bem ajustado através dos modelos não lineares, o objetivo deste trabalho foi modelar o padrão da produção acumulada do café pergaminho em função do tempo (em meses) no ano de 2022, comparando os ajustes dos modelos não lineares Logístico, Gompertz e von Bertalanffy. Foi analisada a produção acumulada nacional e dos departamentos de Huánuco e San Martín, no Peru. Os dados utilizados para o ajuste dos modelos foram obtidos através do Ministério do Desenvolvimento e Irrigação (MIDAGRI) do Peru. Para verificar os pressupostos de normalidade, homocedasticidade e independência residual foram utilizados os testes de Shapiro-Wilk, Breusch-Pagan e Durbin-Watson, respectivamente. A estimação dos parâmetros dos modelos foi feita pelo método dos mínimos quadrados utilizando o algoritmo de Gauss-Newton, no software R. A qualidade do ajuste dos modelos foi verificada por meio dos avaliadores de qualidade de ajuste, tais como Coeficiente de Determinação ($R^2$), Desvio Padrão Residual (DPR), Critério de Informação de Akaike (AIC) e medidas de não linealidade. Com base nos avaliadores de qualidade de ajuste dos modelos, o modelo Gompertz com estrutura de erros autoregressivos de primeira ordem (AR1) se ajustou melhor aos dados de produção nacional e o modelo Logístico foi o mais adequado para descrever a produção dos departamentos de Huánuco e San Martín.

Resumen

El café pergamino es un producto de un proceso de lavado del café y su producción ha logrado un aumento significativo en las regiones cafeteras del Perú. Conocer el patrón en la producción de este grano es fundamental para asistir a los productores de café en la toma de decisiones, en el sector económico y social. Dado que las curvas de crecimiento suelen mostrar un comportamiento sigmoide, en el que se ajusta bien a través de los modelos no lineales, el objetivo de este trabajo fue modelar el patrón de la producción acumulada de café pergamino en función del tiempo (en meses) en el año 2022, comparando los ajustes de los modelos no lineales Logistics, Gompertz y von Bertalanffy. Se analizó la producción acumulada en el país y en los departamentos de Huánuco y San Martín, en Perú. Los datos utilizados para el ajuste de los modelos se obtuvieron a través del Ministerio de Desarrollo y Riego (MIDAGRI) del Perú. Se utilizaron las pruebas de Shapiro-Wilk, Breusch-Pagan y Durbin-Watson para verificar los supuestos de normalidad, homocedasticidad e independencia residual, respectivamente. La estimación de los parámetros del modelo se realizó por el método de mínimos cuadrados utilizando el algoritmo de Gauss-Newton, en el software R. La calidad del ajuste de los modelos se verificó por medio de la calidad de los evaluadores de ajuste, tales como Coeficiente de Determinación (R2), Desviación Estándar Residual (DPR), Criterio de Información de Akaike (AIC) y medidas de no linealidad. Con base en los evaluadores de calidad del ajuste de los modelos, el modelo de Gompertz con estructura de error autorregresiva (AR1) de primer orden se ajustó mejor a los datos de producción nacionales y el modelo de Logística fue el más adecuado para describir la producción de los departamentos de Huánuco y San Martín.


Introduction

Coffee is grown in tropical and subtropical countries and is, according to the Food and Agriculture Organization (FAO, 2023), the most traded basic product worldwide, being one of the most consumed drinks worldwide.

Despite the impact of climate change on coffee production and the farmers’ vulnerability (MORALES et al., 2022), the volume of Peruvian coffee production ranks nine in world production and three in South America (USDA, 2023). According to the National
Coffee Board in Peru, Peruvian coffee is currently the main export product due to its higher production and availability in ecological floors under climates suitable for growing coffee in some regions of the country (JNC, 2023).

Parchment coffee goes through a special process to obtain the beans. The steps considered in the process are washing and drying until the parchment is obtained, maintaining good agricultural practices in production such as benefit, transportation, and the entire coffee production chain. Furthermore, it is important to prevent the development of mold, which leads to diseases in cultivation (Peruvian Technical Standard Implementation Guide, 2021).

Analyzing coffee exports has been the research focus of several scholars. However, according to Diel et al. (2019), an option for evaluating the productive behavior of crops is growth models, which are measurements observed over time, which can be, for example, weight, height, production, etc., and can describe the production behavior of coffee crops. An advantage of this technique is its ease of interpreting the parameters involved in modeling using non-linear models.

In this context, there are several studies on non-linear models associated with growth curves, such as the study by Fernandes et al. (2014), who compared the Gompertz and Logistic models for describing coffee fruit growth curves considering the heterogeneity of sampling variances and concluded that both models were suitable for describing the fruit dry mass considering the heterogeneity, which provided more accurate parameter estimates. In the same year, Sousa et al. (2014) adjusted the previously mentioned models with independent and autoregressive first-order error structures, AR(1), and the Gompertz model proved to be better for describing the germination of coffee seeds.

Other cultivars studied using non-linear models were the growth of lettuce Gloriosa, Pira Verde, and Stella (CARINI et al., 2020), for which the Logistic model best described the growth of these cultivars. Recent studies were presented by Machado et al. (2023), where the germination of Brachiaria brizantha seeds was analyzed by fitting non-linear models, and the model that best fit the data was Gompertz. Other research to mention a few, Puiatti et al. (2020), Paine et al. (2018).

Fitting non-linear models is useful for describing plant characteristics due to their sigmoidal behavior. However, it is critical to choose an appropriate parameterization as there may be problems in inference related to the estimated parameters of the model (FERNANDES et al., 2015). The least squares method can be applied to obtain parameter estimates for non-linear models using the Gauss-Newton iterative algorithm.
Although there is sufficient research literature on coffee at an international level, studies have not yet been found describing the growth of coffee production, specifically parchment, in Peru, using non-linear models, knowing that it is the third country in South America that produces more coffee. In this context, this study aimed to fit the non-linear Logistic, Gompertz, and von Bertalanffy models to data on the cumulative production of parchment coffee for 2022 at the national level and regions with the highest sales.

Having a detailed knowledge about agricultural production is critical, as it allows managers and administrators to act in a more informed and strategic way, adapting its methods and decisions according to market demands, environmental challenges and opportunities for innovation. This understanding is essential for effective and successful management in the agricultural sector, which can contribute to both the economic and social sector of the country.

**Methodology**

To fit the non-linear models, data on parchment coffee production were obtained directly from the Ministry of Development and Irrigation (MIDAGRI) of Peru, corresponding to the monthly production (in tons) from January to December 2022. Furthermore, we considered the cumulative national production and the production of two departments with the highest production, Huánuco and San Martín.

The following parameterizations of non-linear models were fit to the production data (Table 1):

<table>
<thead>
<tr>
<th>Model</th>
<th>Form</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>[ Y_i = \frac{\alpha}{1 + \exp(\beta - kx_i)} + \epsilon_i ]</td>
<td>( \alpha ) is the asymptotic value, ( \beta ) is the location parameter, ( k ) controls the slope of the curve, and ( \epsilon_i ) is the random error.</td>
</tr>
<tr>
<td>Gompertz</td>
<td>[ Y_i = a \exp\left(-\exp(\beta - kx_i)\right) + \epsilon_i ]</td>
<td></td>
</tr>
<tr>
<td>Von Bertalanffy</td>
<td>[ Y_i = \alpha \cdot \left(1 - \exp\left(-k(x_i - \beta)\right)\right)^3 + \epsilon_i ]</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 lists the equations used to fit the models to the data, where \( Y_i \) represents the ith observation of the variable cumulative production of parchment coffee, with \( i=1,2,\ldots,n; \ x_i \) is the ith observation of the independent variable, which, in this study, refers to the months of the year; \( \alpha \) is the maximum cumulative value of coffee production, \( \beta \) is the abscissa of the
inflection point, at which production starts to slow down, $k$ is an index associated with the speed of growth, the higher its value, the faster the curve reaches its inflection point, $\varepsilon_i$ are the random errors assumed to be independent and identically distributed to a standard normal, $\varepsilon_i \sim N(0,1\sigma^2)$.

In the logistic model, at the inflection point, $x$, the curve changes in concavity. At this point, the growth rate reaches its maximum, the asymptotic value; therefore, this function is symmetric to the inflection point (FERNANDES et al., 2014). Andrade, Cirillo and Beijo (2014) also analyzed the logistic model in the proposal of a bootstrap procedure using measures of influence in the presence of outliers, obtaining satisfactory results.

The Gompertz model has similar properties to the logistic model; however, it is not symmetrical to the inflection point, but the sigmoidal pattern is the same (WINDSTOR, 1932). A similar characteristic occurs with the von Bertalanffy model (1957) compared to the Gompertz model, where the first applies to growth studies in animals such as mammals, crustaceans, and fish.

For analysis of model residuals, tests were applied for normality (Shapiro-Wilk (SW)), homoscedasticity (Breusch-Pagan (BP)), and independence (Durbin-Watson (DW)). Furthermore, the least squares method was used to estimate the parameters using the Gauss-Newton iterative algorithm, and the initial values were obtained by observing the function’s behavior using the graphical method.

When the assumptions were not met, in the case of independence of errors, the autoregressive component AR(1) was estimated using the argument correlation=AR(1) of the gnlss function of the RStudio 4.2.3 software (R core Team, 2023).

To compare the models, the goodness-of-fit was assessed using the coefficient of determination ($R^2$), Akaike information criterion (AIC), residual standard deviation (RSD), and the parametric non-linearity was evaluated by Bates and Watts curvature (1988), in which the lowest value of intrinsic ($c^j$) and parametric ($c^0$) non-linearity was selected as the best model. A significance level of 5% was set when inferring the parameters.

The critical points of the model selected as the best in each studied region were calculated, namely the point of maximum acceleration (PMA), inflection point (IP), point of maximum deceleration (PMD), and asymptotic deceleration point (ADP).

All analyses were run in statistical software R (R Core Team, 2023), using the nlme, car, lmtest, and qpcR packages.
Results and Discussion

In the analysis of the cumulative national coffee production, a lack of convergence of the Logistic model was found, and for the departments of Huánuco and San Martín, the von Bertalanffy model also had a convergence problem. Therefore, such models were disregarded from the analysis.

After fitting the models to the data, residual analysis was carried out to check the assumptions of normality, homoscedasticity, and independence of residuals using the Shapiro-Wilk, Breusch-Pagan, and Durbin-Watson tests, respectively. These same tests were used by Fernandes et al. (2014) in the selection of non-linear models to describe the coffee bean growth curves. For the cumulative national production of coffee, according to the Durbin-Watson test, the assumption of independence of residuals was violated in the Gompertz and von Bertalanffy models (p-value < 0.05). In this way, a first-order autoregressive structure (AR1) was incorporated to model the dependence of residuals. A similar problem was observed for coffee production in the departments of Huánuco and San Martín using the Gompertz model, and a first-order autoregressive structure was also used as a measure to correct the dependence of residuals. These alternatives were also adopted by Mazzini et al. (2005), Sousa et al. (2014), and Prado et al. (2013), who improved their model estimates using this autoregressive term. According to the results obtained in the Shapiro-Wilk and Breusch-Pagan tests, the residuals followed a normal distribution and showed homogeneity of variances.

Table 2 presents the results obtained from the goodness-of-fit measures $R^2$, SRD, AIC, and the non-linearity measure (Bates and Watts curvatures); the best model is the one with the highest $R^2$ value and the lowest values of the other criteria.

### Table 2

*Measures to assess the goodness-of-fit of non-linear models to the parchment coffee cumulative national production, and production of the departments of Huánuco and San Martín.*

<table>
<thead>
<tr>
<th>Production</th>
<th>Model</th>
<th>$R^2$</th>
<th>DPR</th>
<th>AIC</th>
<th>$C^1$</th>
<th>$C^9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>Gompertz</td>
<td>0.997</td>
<td>8626.100</td>
<td>249.769</td>
<td>0.109</td>
<td>0.270</td>
</tr>
<tr>
<td></td>
<td>Von Bertalanffy</td>
<td>0.990</td>
<td>16414.600</td>
<td>266.763</td>
<td>0.556</td>
<td>1.141</td>
</tr>
<tr>
<td>Huánuco</td>
<td>Logistic</td>
<td>0.996</td>
<td>345.1440</td>
<td>173.408</td>
<td>0.125</td>
<td>0.498</td>
</tr>
<tr>
<td></td>
<td>Gompertz</td>
<td>0.999</td>
<td>70.740</td>
<td>140.819</td>
<td>0.021</td>
<td>0.068</td>
</tr>
</tbody>
</table>
In the selection of the models in Table 2, for the cumulative national production of parchment coffee, the Gompertz model with a first-order autoregressive (AR) structure presented a better performance due to a high value of the coefficient of determination (R²) and low values in SRD, AIC and measures of intrinsic Cᵢ and parametric C₀ non-linearity, showing the superiority of the Gompertz model to describe the cumulative national production of parchment coffee. Similar results were reported by Machado et al. (2023) and Oliveira et al. (2008), who described the germination of seeds of *Brachiaria brizantha* cv. Marandu and Fernandes et al. (2014), in the analysis of coffee fruit.

As for parchment coffee production in the departments of Huánuco and San Martín, the Logistic model was selected in both regions as it presented a better goodness-of-fit concerning R², SRD, and AIC. Furthermore, the Bates and Watts curvature measures point to the superiority of the Logistic model due to the low non-linearity values, thus corroborating Frühauf et al. (2022) in their study of predicting the height of bean plants.

**Table 3**

95% confidence intervals for the parameters of the models that best fit the data for parchment coffee cumulative national production, and production of the departments of Huánuco and San Martín.

<table>
<thead>
<tr>
<th>Production</th>
<th>Model</th>
<th>Parameters</th>
<th>Estimate</th>
<th>LI</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>Gompertz</td>
<td>α</td>
<td>360882.000</td>
<td>342195.500</td>
<td>379568.400</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β</td>
<td>3.100</td>
<td>2.675</td>
<td>3.526</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k</td>
<td>0.600</td>
<td>0.560</td>
<td>0.740</td>
</tr>
<tr>
<td></td>
<td></td>
<td>φ</td>
<td>0.728</td>
<td>0.213</td>
<td>0.926</td>
</tr>
<tr>
<td>Huánuco</td>
<td>Logístico</td>
<td>α</td>
<td>12590.000</td>
<td>12485.720</td>
<td>12699.700</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β</td>
<td>4.167</td>
<td>4.054</td>
<td>4.286</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k</td>
<td>0.740</td>
<td>0.720</td>
<td>0.770</td>
</tr>
<tr>
<td>San Martín</td>
<td>Logístico</td>
<td>α</td>
<td>69070.000</td>
<td>68223.718</td>
<td>69919.542</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β</td>
<td>4.768</td>
<td>4.449</td>
<td>5.119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k</td>
<td>1.144</td>
<td>1.066</td>
<td>1.231</td>
</tr>
</tbody>
</table>

Based on fitting the Gompertz and Logistic models for the cumulative production of parchment coffee, Table 3 lists the estimates and 95% confidence intervals, where all parameters were significant (p < 0.001). In addition, their confidence intervals do not contain zero. According to the table, the maximum cumulative national production of parchment coffee in 2022 is 360,882.000 tons. In the department of Huánuco, the maximum cumulative
production is 12,590,000 tons, and in the department of San Martín, 69,070,000 tons. Therefore, higher production of parchment coffee is found in the department of San Martín than in the department of Huánuco.

Along with the parameters’ interpretations, the first four partial derivatives for the time variable of the Logistic and Gompertz models were used; the first derivative was the production rate. The point of inflection (PI) of the curve is obtained by equating the second derivative to zero. Equating the third derivative to zero, the point of maximum acceleration (PMA) and the point of maximum deceleration (PMD) are obtained, and the fourth derivative provides information about the asymptotic deceleration point (ADP) (MISCHAN; PINHO, 2014; SILVA et al., 2021).

Table 4 presents the critical points of the curves using the parameter values estimated for the models that best fit the data on the annual cumulative national production of parchment coffee, and production of the departments of Huánuco and San Martín. Such critical points were also analyzed in the germination of Brachiaria brizantha cv. Marandu (MACHADO et al., 2023), lettuce (CARINI et al., 2020), pout pepper (DIEL et al., 2020), and coconut fruit (SILVA et al., 2021).

**Table 4**

*Critical points: point of maximum acceleration (PMA), inflection point (IP), point of maximum deceleration (PMD), and asymptotic deceleration point (ADP) of the selected models fitting the cumulative national production of parchment coffee, and production of the departments of Huánuco and San Martín.*

<table>
<thead>
<tr>
<th>Production</th>
<th>Model</th>
<th>Point</th>
<th>PAM</th>
<th>PI</th>
<th>PDM</th>
<th>PDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>Gompertz</td>
<td>Abscissa</td>
<td>1.496</td>
<td>3.100</td>
<td>4.703</td>
<td>6.095</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ordered</td>
<td>43403.980</td>
<td>132761.100</td>
<td>320714.400</td>
<td>305775.500</td>
</tr>
<tr>
<td>Huánuco</td>
<td>Logistic</td>
<td>Abscissa</td>
<td>2.396</td>
<td>4.167</td>
<td>5.937</td>
<td>7.244</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ordered</td>
<td>2660.580</td>
<td>6295.000</td>
<td>9929.420</td>
<td>11435.060</td>
</tr>
<tr>
<td>San Martín</td>
<td>Logistic</td>
<td>Abscissa</td>
<td>3.615</td>
<td>4.768</td>
<td>5.920</td>
<td>6.771</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ordered</td>
<td>14596.210</td>
<td>34535.000</td>
<td>54473.790</td>
<td>62733.880</td>
</tr>
</tbody>
</table>

It is important to find the critical points, as the abscissa determines productive performance (SARI et al., 2018). In Table 4, for national parchment coffee production, PMA and PMD abscissas were approximately 2 and 5 months, respectively, with the inflection point occurring approximately at three months with a cumulative production of 132,761,100 tons.
ADP occurred approximately in the 7th month, with a cumulative production of 305,775,500 tons.

For production in the Huánuco department, PMA and PMD abscissas were approximately 3 and 6 months, with the inflection point at approximately 4 months with cumulative production of 6,295,000 tons. APD occurred approximately in the 7th month, with a cumulative production of 11,435,060 tons. For production in the department of San Martín, PMA and PMD occurred approximately at 4 and 6 months, IP approximately 5 months with a production of 34,535,000 tons, and ADP also occurred approximately in the 7th month, with cumulative production of 62,733,880 tons.

Figure 1 illustrates the good fit of the Gompertz model to the cumulative national production of parchment coffee over the months. The von Bertalanffy model is initially underestimating and, in the end, overestimating. Figures 2 and 3 show that the Logistic model had a better fit to the parchment coffee production data in the departments of Huánuco and San Martín; the Gompertz model was initially underestimating and, in the end, overestimating it. Similar results were observed by Oliveira, Lôbo and Pereira (2000) when describing the growth of females of the Guzerat breed. In general, the cumulative coffee production in 2022 had a sigmoidal behavior, which was well fit by non-linear regression models.

**Figure 1**
*Fit of the Gompertz and Von Bertalanffy models for the parchment coffee cumulative national production.*
Figure 2

*Fit of the Gompertz and Von Bertalanffy models for the parchment coffee cumulative production in the department of Huánuco.*

![Figure 2](image1)

Figure 3

*Fit of the Gompertz and Von Bertalanffy models for the parchment coffee cumulative production in the department of San Martín.*

![Figure 3](image2)
Conclusions

The non-linear regression models were suitable for describing the cumulative production of Peruvian parchment coffee in 2022. The Gompertz model presented the best performance for describing national production, and the Logistic model was the most suitable for the departments of Huánuco and San Martín. These models provided practical interpretations of the fundamental parameters for decision-making in the management of accumulated production of Peruvian parchment coffee.

Acknowledgments

The authors would like to thank the National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES) for the financial support in carrying out this work.

References


Non-linear regression models in the management of accumulated production of parchment coffee in Peru


Received: 02.01.2024
Accepted: 02.23.2024