Trends in Agricultural and Environmental Sciences
Pantanal Editora
ISSN: 2966-2621
10.46420/TAES.e250006

A baru seed quality index based on a fuzzy inference system

Bruno Rodrigues de **Oliveira**[1,*], Eder Pereira **Neves**[2], Marco Aparecido Queiroz **Duarte**[2], Alan Mario **Zuffo**[3], Francisco Charles dos Santos **Silva**[3], Leandris **Argentel-Martínez**[4]

[1] Pantanal Editora;
[2] UEMS - State University of Mato Grosso do Sul, Cassilândia, MS, Brazil;
[3] UEMA - State University of Martanhão, Balsas, MA, Brazil;
[4] National Technological Institute of Mexico/IT Yaqui Valley, Bácum, México.

* Correspondence: bruno@editorapantanal.com.br

Received: 2025-11-24
Accepted: 2025-12-11
Published: 2025-12-23

Main Editors
Jorge González Aguilera

For citation: Oliveira, B. R.; Neves, E. P.; Duarte, M. A.; Zuffo, A. M.; Silva, F. C.; Argentel-Martínez, L (2025). A baru seed quality index based on a fuzzy inference system. Trends in Agricultural and Environmental Sciences, (e250006), DOI: https://doi.org/TAES.e250006

Abstract: This study presents a fuzzy inference system (FIS) developed to evaluate the morphological quality of *Dipteryx alata* (baru) almonds based on four biometric variables: diameter, width, length, and weight. The membership functions were automatically calibrated using descriptive statistics of the dataset, whereas the inference rules were defined by experts. The resulting fuzzy quality index (FQI) effectively captured interannual biometric variability and allowed the identification of almonds with higher and lower morphological qualities at different harvests (2012 and 2023). The centroid-based defuzzification method provided stable and interpretable results, confirming the model's robustness and suitability for physical quality assessment. The proposed system can serve as an initial morphological filter in sustainable production chains, enabling rapid, transparent, and nondestructive classification of baru almonds and supporting precision agriculture practices.

Keywords: Fuzzy Logic; Morphological Quality; Baru Almond; Seed Classification; Precision Agriculture.

1. Introduction

Dipteryx alata Vogel, known as baru, is a species native to the Brazilian Cerrado that has gained increasing prominence in the global market because of its high nutritional value and economic potential (Arruda et al., 2021; Valadão & Souza, 2024). Nutritionally, baru almonds (nuts) are widely recognized as "superfoods" because of their high protein content, surpassing the protein content of traditional nuts such as almonds and pistachios (Egea et al., 2023; Fernandes et al., 2010; Miyashita et al., 2025). They are rich in healthy lipids, dietary fiber, and essential minerals such as iron, zinc, and calcium, and the presence of bioactive compounds endows them with significant antioxidant capacity (Miyashita et al., 2025; Sousa et al., 2024). Economically, the sustainable trade of baru generates income, strengthens family farming, and contributes to the conservation of the Cerrado's biome (Santos et al., 2024). The growing demand for healthy foods has increased market value, with projections that the global baru market will reach US$47 million by 2032 (Santos et al., 2024; Sousa et al., 2024).

Quality control in agriculture is often limited by the uncertainty and imprecision of environmental and biological data, which can make traditional methods subjective or inefficient at a large scale (Erdoğdu et al., 2025). In this context, artificial intelligence-based systems, such as fuzzy logic, are emerging as robust and effective tools for precision agriculture (Özoğul, 2025; Widayat et al., 2025). Fuzzy logic is particularly suited for addressing vague and ambiguous information, enabling adaptive and flexible decision-making in complex agricultural environments by replicating expert human reasoning (Morchid et al., 2025; Prasad et al., 2024). Its applications in the agricultural sector include monitoring and optimization of irrigation systems, identification and classification of crops and diseases, decision support system (DSS) planning, environmental and greenhouse monitoring, traceability and quality monitoring, optimization of irrigation systems and decision support for yield forecasting and crop monitoring (Widayat et al., 2025).

Studies (Botezelli et al., 2000; Mota et al., 2020; Zuffo et al., 2014) have reported significant morphological variability among baru populations and harvests, which directly affects uniformity. In addition, another study revealed that the evaluation of physical characteristics, such as the biometrics of fruits and seeds, is a fundamental and primary step in optimizing and ensuring the sustainability of the value chain of native species such as baru (Sousa et al., 2024). The reviewed literature establishes that baru is a rich source of protein and fatty acids. These substances are related to biometric attributes, and a review article supports the principle that seed weight and size are therefore partial predictors of oil extraction yield and quality (Santos et al., 2024). These findings highlight the need for models to assess seed (almond) quality exclusively through biometric parameters and evaluate their physical and morphological characteristics.

This paper presents a fuzzy inference system (FIS) for classifying the morphological quality of baru almonds based on biometric attributes (diameter, width, length, and weight). The methodology adopts a data-driven approach, in which membership functions are objectively calibrated using descriptive statistics of the dataset, such as quartiles, minima, and maxima. This statistical calibration eliminates the subjectivity inherent in manual adjustments, ensuring that the model automatically adjusts to the specific and reduced scale of the almonds. On the other hand, the definition of inference rules is performed by experts. The proposed system provides a continuous fuzzy quality index (FQI) capable of capturing the biometric variability of baru almonds.

2. Materials and methods

2.1 Description of the experiment

The study of (Zuffo et al., 2014) was conducted in the nature reserve of Fazenda União, a remnant of the Cerrado with little human disturbance, located in the Serra Azul Valley, 28 km from the municipality of Nova Xavantina, eastern Mato Grosso state, Brazil (14° 50' 41" S; 52° 22' 49" W), at an altitude of 290 m. The approximately 100-hectare area contained several native fruit species, including *baruzeiro* (baru tree), which is not exploited for extractive purposes and serves as a food source for wildlife.

The region's climate, according to the Köppen climate classification, is Aw, characterized by two well-defined seasons: a dry season (generally from May to September) and a rainy season (October to April). The average annual temperature is 24°C, with an average rainfall of 1,500 mm. Climate data were obtained from the meteorological station of the National Institute of Meteorology (INMET) (Silva et al., 2008).

Ripe fruits were collected in August, during the first week of the month, in 2012 and 2013, coinciding with the region's dry season. Fruits were collected from the ground in the canopy projection of 10 *baruzeiro* trees, which averaged 15.0 ± 5.00 m in height. The trees were randomly selected from the study area and spaced at least 200 m apart, and the same trees were harvested in both years.

Following the suggestion of (Gonçalves et al., 2013) for biometric studies of fruits from tropical tree species, 10 fruits were collected per plant, totaling 100 fruits, all of which were visually healthy, intact, and undistorted. After harvest, the fruits were packed in boxes and transported to the Plant Biology Laboratory of Mato Grosso State University, Nova Xavantina campus, Mato Grosso.

The fruits were assessed for longitudinal length (L), width (W), and thickness (TH) using a digital caliper (Clarke-150 mm). Fresh weight (FW) was determined on a precision analytical balance (0.001 g). Fruit volume (FV) was obtained using the displaced water volume method by immersing the fruit in a measuring cylinder, as described by (Basso, 1999).

Seed (almond) removal was performed after breaking the endocarp of the fruits with a hammer. Longitudinal length (L), width (W), thickness (W), and fresh weight (FW) were determined for 100 seeds. The seed volume index (SVI) was calculated by multiplying the length, width, and thickness of each seed (CLS × ES × LS). The fresh weight of the seeds was also obtained using a precision analytical balance (0.001 g).

2.2 Fuzzy Rule-Based System

A fuzzy rule-based system (FRBS) is a framework designed to model complex, uncertain, and imprecise processes using linguistic rules rather than precise mathematical equations. These systems are inspired by human reasoning and the concept of approximate inference introduced by Zadeh (1965), allowing them to operate effectively in environments where classical logic cannot adequately represent uncertainty (Morchid et al., 2025). According to (Castillo & Melin, 2023), in recent years, FRBSs have gained prominence in control, decision-making, pattern recognition, and data analysis applications.

An FRBS generally consists of three main steps: fuzzification, inference (rule base), and defuzzification. Fuzzification is the initial and fundamental step in a fuzzy logic system, in which crisp input values originating from the external environment are transformed into the uncertainty domain. This process is performed by applying membership functions that associate each input with a degree of membership (variable between 0 and 1) to one or more predefined linguistic terms. The format of a membership function can be triangular, trapezoidal, Gaussian, sigmoid, Z-shaped, S-shaped, or generalized Bell. The kind of membership function to be used in the FRBS depends on the system requirements and the desired smoothness in the transition between fuzzy sets; it is crucial because it defines how ambiguity and imprecision will be mathematically modeled, that is, how the system will perform approximate reasoning (Castillo & Melin, 2023; Erdoğdu et al., 2025; Magdalena, 2015; Mittal et al., 2020; Morchid et al., 2025; Prasad et al., 2024; Shihabudheen & Pillai, 2018; Widayat et al., 2025).

Triangular membership functions are defined by three parameters (a, b, c), which determine the shape and position of the triangle on the universe of the discourse axis. A triangular membership function $\mu_A(x)$ is defined by equation (1):

$$\mu_A(x) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right), \quad (1)$$

where a is the lower limit (where membership begins to increase); b is the point of maximum membership: $\mu_A(b) = 1$; and c is the upper limit (where membership returns to zero). An illustration of this membership function can be found in Figure 1.

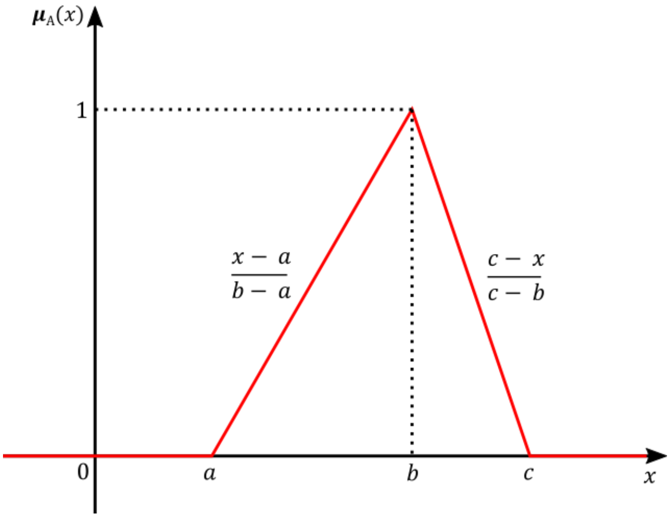


Figure 1. Representation of a triangular membership function of a fuzzy set A.

The trapezoidal membership function is defined by the parameters (a, b, c, d), whose function $\mu_A(x)$ is defined by equation (2) and illustrated in Figure 2.

$$\mu_A(x) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right) \tag{2}$$

Now, a and d represent the left and right trapezoid feet, respectively, and parameters b and c are the maximum points that control the left and right shoulders, respectively.

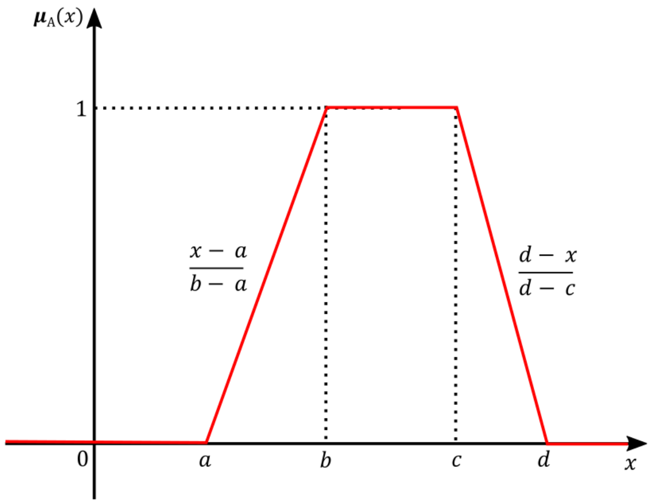


Figure 2. Representation of a trapezoidal membership function of a fuzzy set A.

The next step is inference and rule-based development. The rule base constitutes the logical core of the FRBS and represents the expert's or the system's knowledge about the relationship between inputs and outputs. A fuzzy system with two noninteractive inputs, x_1 and x_2 (antecedents), and a single output y (consequent), is expressed by a series of r IF-THEN linguistic propositions in the form of rules (Ross, 2010).

$$R_j: \text{IF } x_1 \text{ is } A_j \text{ AND/OR } x_2 \text{ is } B_j \text{ THEN } y \text{ is } C_j, j = 1, 2, 3, \ldots, r,$$

where A_j and B_j are the fuzzy sets representing the j-th antecedent or premise pairs and C_j is the fuzzy set representing the j-th consequent. For the operators "AND" and "OR", we have what is called a conjunctive or disjunctive antecedent system, respectively.

The inference engine combines all applicable rules using fuzzy operators such as AND, OR, and NOT, generating a fuzzy output set. (Lu et al., 2024) highlighted that the most commonly used methods are Mamdani, in which the output is fuzzy and highly interpretable, and Takagi-Sugeno, where the output is a mathematical function suitable for optimization and efficient calculations. The inference process allows the system to evaluate multiple conditions simultaneously, weighting the activation levels of each rule and ensuring that the final decision adequately reflects all available information.

In the final step, known as defuzzification, the resulting fuzzy set from inference is converted into a numeric value (crisp) that can be used for action or decision. One of the most commonly used methods is the centroid method, which, in discrete form, is calculated by equation (3):

$$Y^* = \frac{\sum_{i=1}^{n} \mu(y_i).y_i}{\sum_{i=1}^{n} \mu(y_i)}, \tag{3}$$

where Y^* is the crisp output value, $\mu(y_i)$ is the membership function of the output fuzzy set for point i, and n is the total number of discrete points considered in the output variable's discourse universe. According to (Mendel, 2024), other methods, including the mean of the maxima and the max of the maxima, can be used depending on the application.

2.3 Proposed modeling

The fuzzy inference system (FIS) proposed for baru almond morphological quality scoring was developed using a data-driven approach, ensuring that the model's parameters definitions were justified by statistical calculations extracted from a biometric dataset of 100 almonds (seeds) collected in 2012 and 2013, according to the experiment described in Section 2.1. The system's framework follows the classic fuzzy logic processing cycle, starting with fuzzification, in which the input biometric (numeric) values are converted into membership degrees to fuzzy sets. Next, inference combines these membership degrees using a set of "IF-THEN" logic rules. Finally, defuzzification translates the aggregated fuzzy output into a single numerical value, the fuzzy quality index, which represents the almond's final score.

The FIS was built with four input variables, namely, diameter, width, length, and weight, and a single output variable, namely, quality (which refers to morphological quality). For the input variables, the linguistic sets "Small/Short/Slim," "Middle," and "Large/Long/Heavy" were defined. The output variable, Quality, was subdivided into the sets "Low," "Middle," and "High". Table 1 details the assignments of the language sets to the FIS model variables.

Table 1. Discrimination of biometric and output variables and their respective linguistic sets.

Type	Category	Variable	Linguistic sets
Input	Biometric	Diameter	Small, Middle, Large
Input	Biometric	Width	Short, Middle, Long
Input	Biometric	Length	Short, Middle, Long
Input	Biometric	Weight	Slim, Middle, Heavy
Output	Of the model	Quality	Low, Middle, High

Model calibration was conducted strictly objectively, using descriptive statistics (minimum, maximum, and quartiles) from the biometric dataset to define the points of the membership functions. This statistical approach to anchoring the functions aims to eliminate the subjectivity inherent in manual adjustments, allowing the system to automatically adapt to the actual data distribution.

The universe of discourse for each input variable was delimited by its minimum (min) and maximum (max) values. The quartiles (Q_1, Q_2, and Q_3) were used to position the membership functions according to the following: the "Small/Short/Slim" membership function was defined as trapezoidal, with its starting and peak points anchored at the minimum value and the first quartile (Q_1), respectively, and the base endpoint at the second quartile (Q_2, the median), ensuring overlap with the "Middle" that was defined as triangular, with its center point at the median (Q_2) and the base points at the first (Q_1) and third quartiles (Q_3). Finally, the "Large/Long/Heavy" function was defined as trapezoidal, with its base starting at the second quartile (Q_2) and the peak at the third quartile (Q_3), extending to the maximum value of the universe of discourse. This calculation methodology avoids the creation of "dead zones" between fuzzy sets, ensuring that each seed presents a degree of relevance greater than zero in at least one category, which increases the robustness of the system and prevents classification errors.

Figure 3 shows the methodological diagram of the fuzzy inference system (FIS) proposed for scoring the morphological quality of baru almonds, which follows a data-driven approach. The knowledge base is formed by the experiment involving the collection of 100 almonds in 2012 and 2013 in the Cerrado region of Mato Grosso and by the knowledge of the experts who conducted the experiment. The framework uses biometric variables (diameter, width, length, and weight) and descriptive statistics (minimum, maximum, and quartiles) to objectively calibrate the membership functions. The FIS consists of the following modules: a fuzzifier (which converts crisp inputs into fuzzy sets), an inference engine (which uses a base of rules), and a defuzzifier (which translates the fuzzy result into a crisp output value, the fuzzy quality index).

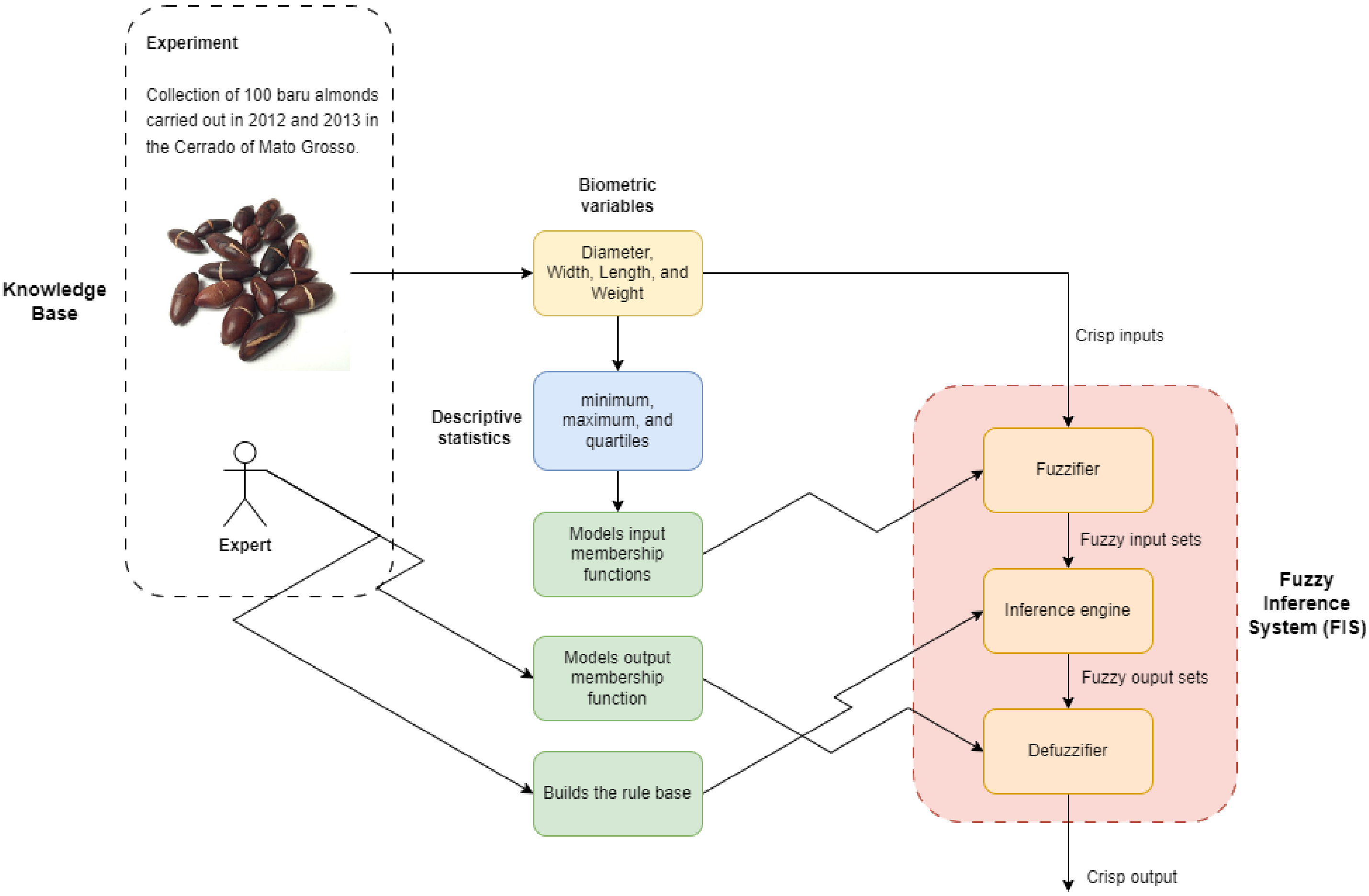


Figure 3. Schematic diagram of the fuzzy inference system (FIS) for scoring the morphological quality of baru almonds. The data-driven model uses four biometric input variables and has membership functions objectively calibrated using descriptive statistics. Processing follows the steps of fuzzification, rule-based inference, and defuzzification, resulting in a fuzzy quality index.

3. Results and discussion

The descriptive statistics of the biometric data for baru almonds collected in 2012 and 2013, including diameter, are presented in Table 2. These variables, width, length, and weight, were used as parameters to construct the input membership functions of the proposed fuzzy inference system (FIS). For each variable, the minimum, quartile (Q_1, Q_2, and Q_3), and maximum values define the linguistic limits, ensuring that the fuzzy partitions reflect the real distribution of the data rather than arbitrary boundaries.

The results show a consistent biometric pattern between harvests, but with noticeable interannual variation in all dimensions, particularly in length and weight, as previously noted. In 2012, almonds presented greater average dimensions (length = 25.99 mm; weight = 1.36 g) than those from 2013 (length = 21.94 mm; weight = 1.14 g), suggesting the influence of environmental factors such as rainfall and temperature on almond development (Zuffo et al., 2014). Similar fluctuations in almond morphology between years have been reported in other studies on *Dipteryx alata*, where climatic differences between seasons affected the deposition of storage compounds and overall fruit size (Mittal et al., 2020; Miyashita et al., 2025; Rinaldi et al., 2021).

These variations are particularly relevant in the context of fuzzy modeling because they expand the range of the universe of discourse for each variable, allowing the membership functions to capture natural biological variability. By defining trapezoidal and triangular membership functions on the basis of quartiles, the model adapts to the specific distribution of each harvest and avoids the subjectivity associated with manual parameter tuning. This approach is consistent with best practices in data-driven fuzzy modeling described by (Prasad et al., 2024; Widayat et al., 2025), who emphasize that automatic calibration using descriptive statistics improves robustness and interpretability, especially in agricultural systems with heterogeneous inputs.

Table 2. Descriptive statistics of biometric data from baru almonds collected in 2012 and 2013.

Statistic	Year 2012				Year 2013			
	Diameter	Width	Length	Weight	Diameter	Width	Length	Weight
min	8.230000	7.180000	22.530000	0.900000	7.620000	7.870000	16.650000	0.24000
Q_1	9.852500	8.400000	24.945000	1.245000	8.500000	8.787500	20.837500	1.01750
Q_2	10.25000	8.950000	25.990000	1.355000	8.890000	9.165000	21.940000	1.14000
Q_3	11.20000	9.415000	26.940000	1.432500	9.185000	9.612500	23.955000	1.30000
max	12.63000	10.57000	28.980000	1.700000	9.950000	10.73000	29.560000	1.91000

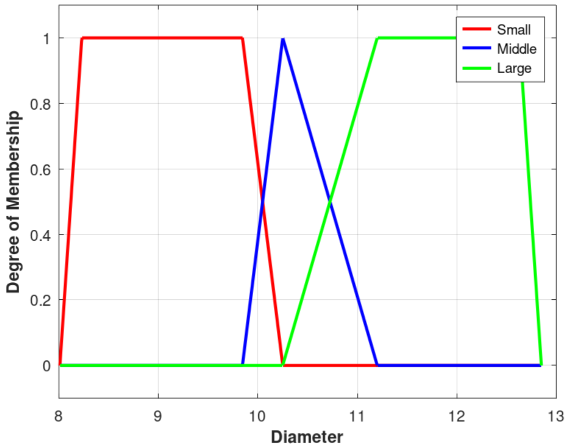
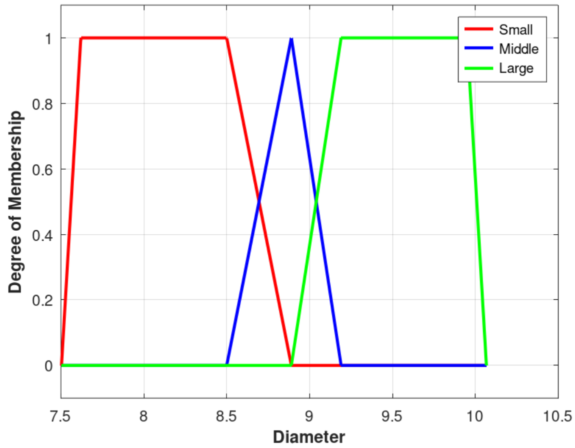
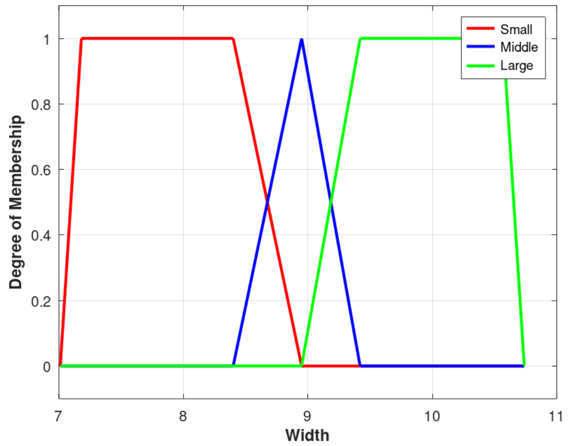
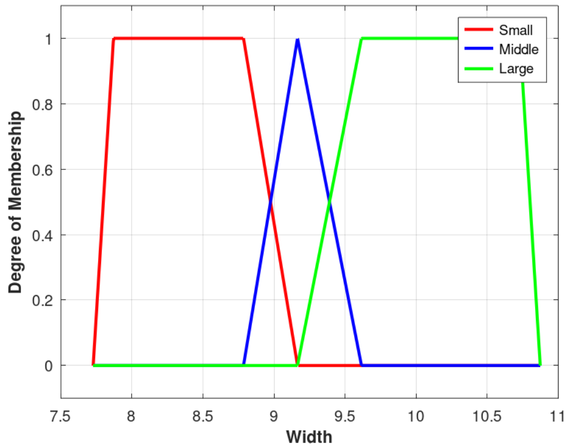
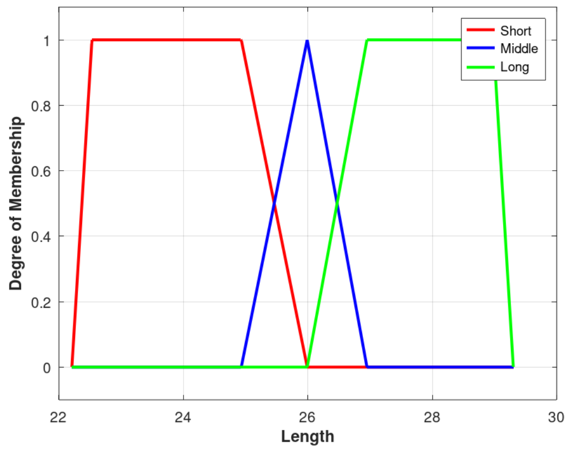
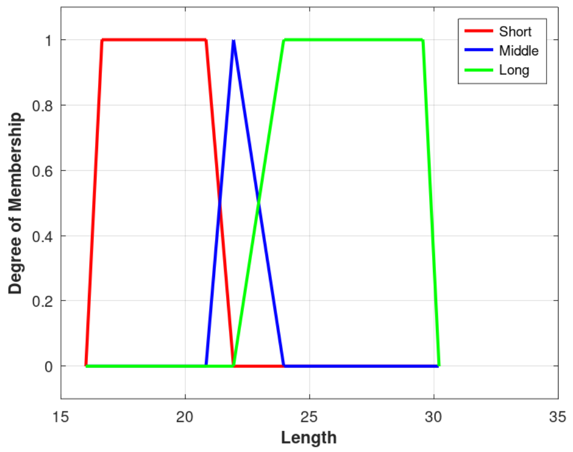
min and max are the minimum and maximum values, respectively; Q_1, Q_2, and Q_3 are the quartiles that divide the data into 25%, 50%, and 75%, respectively.

Table 3 lists the resulting membership functions for each biometric attribute and the output variable, Quality, generated separately for the years 2012 and 2013. The graphical models clearly show consistency in the structure of the sets, although the domain intervals shift slightly according to the biometric differences observed between years. For example, in 2012, the "Large" class for Diameter spans approximately 10.5--12.5 mm, whereas in 2013, this range shifted downward, from approximately 9.0--9.5 mm. This reduction reflects the smaller average almond size observed at the second harvest. Similar behavior is observed for the length and weight variables. The fuzzy sets for 2013 exhibit narrower support and overlap, which indicates a more compact distribution of values, typical of harvests with less morphological variation. Conversely, the 2012 fuzzy sets display broader transitions between the small–middle and middle–large categories, indicating greater heterogeneity among samples. This pattern supports field observations that environmental variability in Cerrado ecosystems contributes to morphological diversity within *Dipteryx alata* populations (Mittal et al., 2020).

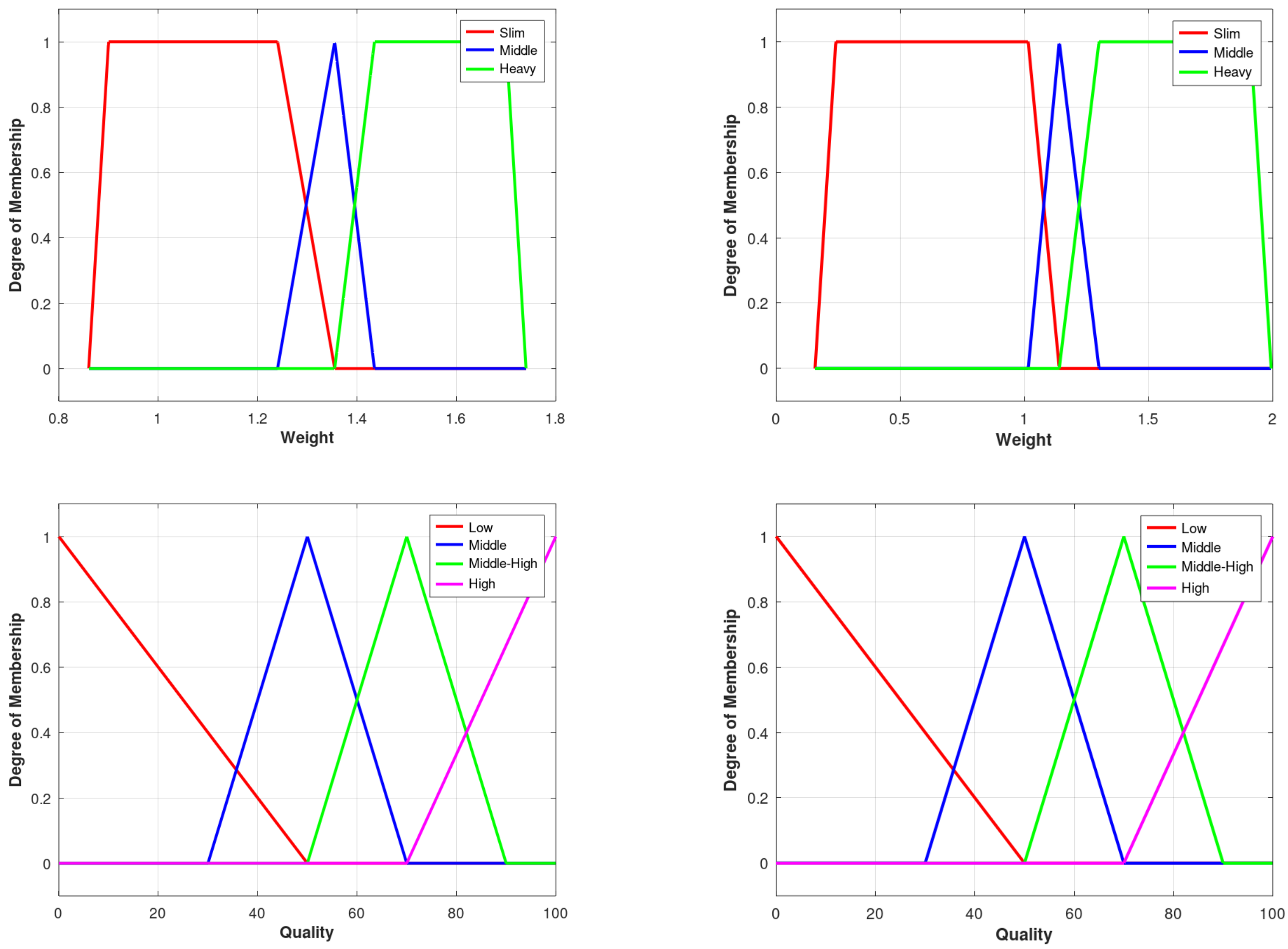
The shape and symmetry of the membership functions also provide insights into the biological relationships among the parameters. For example, width and diameter present nearly symmetric triangular sets, suggesting uniform variability, whereas length and weight exhibit slightly skewed functions, reflecting a tendency for these variables to accumulate higher frequencies around intermediate values. This behavior is consistent with allometric patterns in seed morphology, where longitudinal growth and mass accumulation respond nonlinearly to environmental stimuli (Sousa et al., 2024).

The fuzzy partitions obtained for the output variable Quality (bottom plots in Table 3) illustrate a progressive transition among four linguistic categories: low, middle, middle-high, and high. The inclusion of the intermediate Middle-High proved particularly relevant, as it captures cases in which biometric attributes partially compensate for each other. For example, almonds with high weights but moderate lengths or medium weights with long lengths. This intermediate zone avoids abrupt classification jumps and provides a smoother interpretation of quality.

Table 3. Graphs of the models of the set of membership functions obtained for each variable separated by the years 2012 and 2013, built from descriptive statistics (input variables) and expert knowledge (output variables).

Year 2012	Year 2013
	
	
	

Year 2012	Year 2013



Morphologically, the results reveal that width and diameter are less variable than length and weight are, as indicated by the smaller differences between quartiles and narrower fuzzy intervals. It suggests that while baru almonds maintain a relatively stable cross-sectional profile, their longitudinal development and mass accumulation are more influenced by environmental or physiological factors. This finding aligns with that of (Rinaldi et al., 2021; Santos et al., 2024; Zuffo et al., 2014), who emphasized that nutrient availability and water stress modulate fruit elongation and seed filling in baru.

Finally, the separation of membership functions by year reinforces the adaptability of the proposed method. The proposed fuzzy system is capable of reconfiguring its linguistic boundaries automatically from updated descriptive statistics, which provides consistency across different harvests or environmental contexts. This adaptive feature is essential for modeling dynamic biological systems and aligns with current trends in intelligent agricultural modeling (Erdoğdu et al., 2025; Morchid et al., 2025; Özoğul, 2025).

The fuzzy rules base, constructed after consulting with experts (Zuffo et al., 2014), is presented in Table 4. The rules translate the four biometric input variables (weight, length, diameter, and width) into an output variable (quality). The rule base consists of 11 logical propositions in the "IF-THEN" format that seek to capture the complexity and morphological quality criteria for *Dipteryx alata* almonds.

The high-quality criteria are established by Rules 1, 4, and 7. Rule 1 is fundamental since it identifies seeds that combine the density attribute (weight is heavy) with the longitudinal size attribute (length is long), essential characteristics for superior quality. Rule 4 imposes a more stringent criterion of excellence, requiring the almond to be heavy and large in all transverse and longitudinal dimensions (the weight is heavy AND the diameter is large AND the length is long), and it is activated only for samples that exceed standards. In turn, Rule 7 emphasizes the importance of filling, activating the high linguistic category for the variable quality, for heavy and wide almonds.

Table 4. Fuzzy rules base used by the fuzzy inference system (FIS) to classify quality. The table lists the linguistic propositions in the form "IF-THEN" that connect the input biometric attributes (weight, length, diameter, and width) to determine the output quality.

Rule	Linguistic proposition	Description
Rule 1	IF Weight is Heavy AND Length is Long THEN Quality is High	Identifies seeds that combine density and size for superior quality
Rule 2	IF Weight is Middle AND (Diameter is Middle OR Width is Middle) THEN Quality is Middle	Covers most almonds, classifying those with typical proportions
Rule 3	IF Weight is Slim OR Length is Short THEN Quality is Low	Disqualification criteria: deficiency in weight or length is sufficient to classify the quality as low
Rule 4	IF Weight is Heavy AND Diameter is Large AND Length is Long THEN Quality is High	Strict rule that requires excellence in all dimensions (large and heavy)
Rule 5	IF Weight is Slim AND the Width is Small THEN Quality is Middle	Complements Rule 3 by identifying underdeveloped almonds (slim and narrow)
Rule 6	IF Diameter is Large AND Length is Short THEN Quality is Middle	Includes almonds with average weight or that demonstrate a trade-off between a desirable and an undesirable attribute
Rule 7	IF Weight is Heavy AND Width is Large THEN Quality is High	High quality grade for almonds that are heavy and wide, indicating good fill
Rule 8	IF Weight is Slim AND Diameter is Small THEN Quality is Low	Classifies low-quality almonds that are light and small in diameter, a strong indication of underdevelopment
Rule 9	IF Weight is Middle AND Length is Long THEN Quality is Middle-High	Classifies almonds that exceed the average by having a medium weight (neutral) and a superior length (positive attribute)
Rule 10	IF Weight is Heavy AND Length is Middle THEN Quality is Middle-High	Classifies almonds that exceed the average by having a high weight (positive attribute) with a medium length (neutral)
Rule 11	IF Length is Long AND (Width is Small OR Diameter is Small) THEN Quality is Middle	Rates quality as average in a trade-off scenario: the desirable length (Long) is offset by an undesirable dimension (Width or Small Diameter)

In contrast, Rules 3, 5, and 8 establish the criteria for low quality and disqualification. Rule 3 acts as a primary criterion, where a deficiency in just one of the critical dimensions (weight is Slim or length is short) is sufficient to negatively impact the activation of the membership function. In addition, Rule 5 identifies underdeveloped almonds that are both light and narrow. Rule 8 reinforces this indicator by classifying almonds that are light and small in diameter as strong indicators of underdevelopment.

The Rules for Middle and Middle-High Quality Cover the majority of the population and manage compensation scenarios. Rule 2 classifies almonds with middle weights and typical proportions in the cross-sectional dimensions, encompassing the majority of almonds in the normal range. The more complex rules, such as Rules 6 and 11, demonstrate the model's ability to handle attribute trade-offs. For example, Rule 11 assigns Middle Quality when a desirable attribute, such as Long Length, is offset by a deficient cross-sectional dimension (Small Width OR Small Diameter). Rules 9 and 10 define the middle-high threshold for those almonds that exceed the average standard by combining a neutral attribute (middle) with a superior attribute (long or heavy). The structure of these rules ensures that the classification is robust and capable of reflecting the heterogeneity and uncertainty intrinsic to the biometric attributes of baru almonds.

Figures 5 and 6 show the distributions of fuzzy quality index (FQI) values for samples from the 2012 and 2013 harvests. The values are calculated by combining the outputs of all the rules activated in the respective membership functions of each input variable. The centroid method then calculates the center of mass of this aggregated area. Figure 5 (2012 harvest) reveals a relatively broad distribution, with values ranging from ~20--90, with the mean index marked at ~43.5 (red dashed line). Figure 6 (2013 harvest) shows an equally broad distribution but with a slightly lower mean (~41.0) and a more concentrated frequency at low values (~20--40), although there are still high values (~70--90).

Results in figures 5 and 6 suggest several relevant characteristics. First, the fact that the average FQI for 2012 (~43.5) is slightly greater than that for 2013 (~41.0) indicates that, according to the biometric parameters collected for 2012 (see Table 2), the seeds from that harvest had, on average, a greater "morphological physical quality"— that is, larger dimensions and greater mass — than those from the 2013 harvest. Consistent with the statistics presented, the 2012 harvest produced seeds of greater average length and weight. Second, the similar range in both years (with values varying up to ~90) demonstrates that, although the average decreased, there were still seeds with high morphological performance in 2013 — which highlights the intraharvest variability and justifies the use of a fuzzy model with a continuous quality gradient.

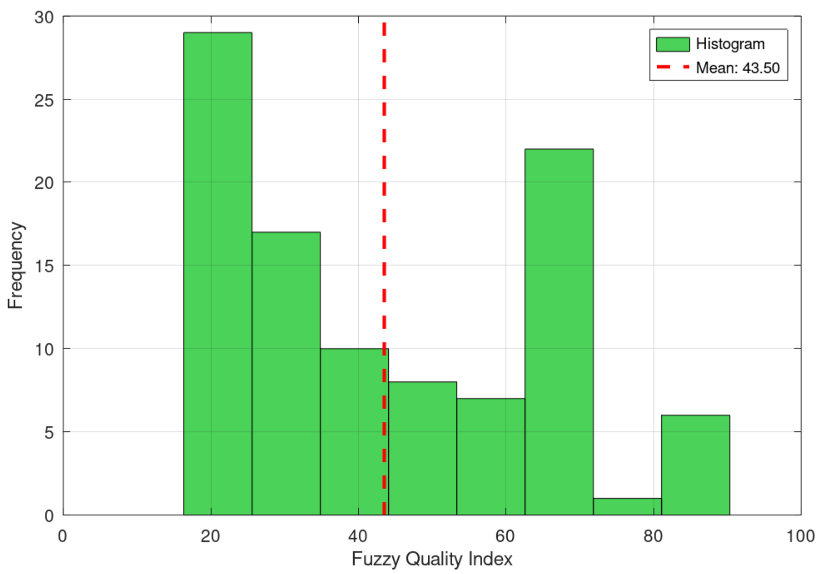


Figure 4. Histogram of fuzzy quality indices obtained by applying the proposed FIS to all samples of baru almonds from 2012.

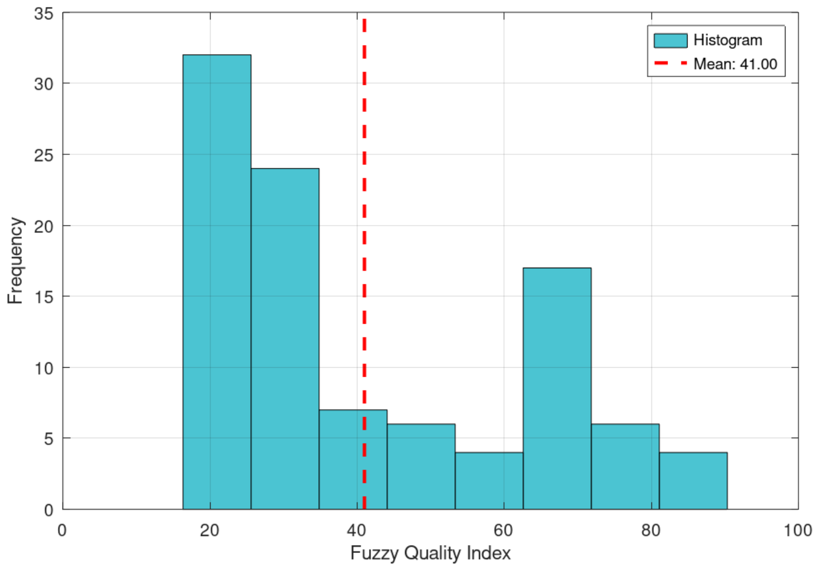


Figure 5. Histogram of fuzzy quality indices obtained by applying the proposed FIS to all samples of baru almonds from 2013.

In the literature, this morphological variability between harvests is well documented for baru and native Cerrado species. (Sousa et al., 2024) noted that the biometric properties of baru vary according to genetics, environmental conditions, and degree of maturation. Such information supports the fact that seeds from 2013 presented, on average, lower values, which was reflected in the shift of the distribution curve to median-low values in Figure 6. Furthermore, a review by (Egea et al., 2023) highlighted that seed size, mass, and integrity are critical factors for the potential quality of baru for industrial and commercial use.

The shape of the histograms is also relevant for discussing the performance of the proposed fuzzy system. The histogram for the 2012 harvest shows a bimodal distribution or at least a clear concentration in the lower range (~20--35) and a tail extending to higher values (~60--90). This tail suggests a set of almonds that present all the favorable biometric attributes (high weight + long length + larger width/diameter), conditions that the model maps to "High" quality. In the 2013 harvest, although there was a similar tail, the frequency of samples in the higher range was lower, and a more significant population was observed in the lower values (~20--30). It may indicate a greater proportion of almond with reduced biometrics, possibly due to climatic effects and unfavorable maturation conditions (Zuffo et al., 2014). This type of distribution supports the justification for including an intermediate category (Middle-High) in the system's consequent to capture transitions between "Middle" and "High" more gradually, as suggested in the literature on fuzzy systems adapted to seed and agricultural product classification, which recommend avoiding rigid thresholds to reflect the continuous nature of biological variability (Erdoğdu et al., 2025; Magdalena, 2015; Mittal et al., 2020; Morchid et al., 2025; Ross, 2010; Widayat et al., 2025).

Furthermore, the relatively low average (~40--45) in both years shows that most almonds did not reach the maximum possible morphological quality levels, reinforcing the idea that using a continuous fuzzy index is more appropriate than a dichotomous classification, such as good quality or poor quality. In other words, the developed model allows us to differentiate gradients of physical performance, not just extreme classes. This approach is consistent with studies on biometrics and seed quality that point to the importance of measuring continuous variability (e.g., in almonds or grains) and not relying solely on fixed categories (Mahmood et al., 2023).

In practical terms, the visualization provided in Figures 5 and 6 confirms that the FIS captured both the internal almond variability and the differences between harvests. The lower average in 2013 may indicate the need for adjustments in the harvesting, maturation, or sorting process for the following harvest, and it also reinforces the index's role as a benchmark tool for future monitoring. Furthermore, the overlapping quality ranges, with some 2013 almonds achieving very high levels, underscore that even though the year was "less favorable", there are samples that met the full biometric criteria, and the system detected them appropriately. The consistency between the histograms and the modeled membership functions (see Table 3) reinforces the methodological consistency of the study: the functions were calibrated using descriptive statistics, and the resulting distributions show that the FQI is distributed according to the defined universe of discourse. It strengthens the reliability of the system as an exploratory instrument for determining the physical quality of baru seeds.

To detail the step-by-step process of the proposed FIS, tables 5 and 6 show how the system generates the lowest and highest FQIs, respectively, using the centroid method and the values of the variables responsible for generating these indices. Table 5 shows the results for the input: Diameter=10.00, Width=9.69, Length=25.72, and Weight=1.13 for the year 2012, and Diameter=9.35, Width=9.39, Length=25.70, and Weight=0.76 for the year 2013. Table 6 presents the results for the input: Diameter=10.25, Width=9.60, Length=27.13, and Weight=1.55 for the year 2012, and Diameter=8.92, Width=9.61, Length=25.09, and Weight=1.31 for the year 2013. These examples demonstrate the interpretability and internal coherence of the model, allowing a transparent analysis of how combinations of biometric variables influence the final quality classification.

Table 5. Example of the proposed FIS, considering the values of the variables that generate the lowest fuzzy quality index, for 2012 and 2013 harvests.

Variable and membership function activation				**Rule and activation**
Diameter	**Width**	**Length**	**Weight**	
		Year 2012		
Small: 0.6173	Small: 0.0000	Short: 0.2547	Slim: 1.0000	Rule 4 - Strength: 1.0000
Middle: 0.3827	Middle: 0.0000	Middle: 0.7453	Middle: 0.0000	Rule 5 - Strength: 0.2547
Large: 0.0000	Large: 1.0000	Long: 0.0000	Heavy: 0.0000	Rule 10 - Strength: 0.6173
		Year 2013		
Small: 0.0000	Small: 0.0000	Short: 0.0000	Slim: 1.0000	
Middle: 0.0000	Middle: 0.5000	Middle: 0.0000	Middle: 0.0000	Rule 4 - Strength: 1.0000
Large: 1.0000	Large: 0.5000	Long: 1.0000	Heavy: 0.0000	

In Table 5, the most strongly activated rules are Rules 4, 5, and 10. Despite the partial activation of the "High" quality rules, the strong presence of the thin weight category and the short length category ultimately reduced the final FQI. In 2013, the lowest FQI resulted from a sample with diameter = 9.35 mm and weight = 0.76 g, corresponding to strong activation of Rule 4 and confirming the system's sensitivity to weight reduction.

The behavior of the FIS is consistent with empirical findings reported in the literature. Several studies emphasize that the mass and size of baru almonds are key indicators of physical maturity and reserve content (Fernandes et al., 2010; Sousa et al., 2024). Seeds with lower weights generally have less embryo filling and lower oil contents (Mittal et al., 2020). Similarly, other studies reported that dimensional reduction is correlated with lower extraction efficiency and, consequently, with lower physical quality (Egea et al., 2023; Valadão & Souza, 2024). Therefore, the low FQI obtained for the lighter and smaller seeds confirms the biological validity of the FIS, which is associated with biometric reduction and inferior morphological quality.

On the other hand, the results in Table 6 for the 2013 harvest show that the "best" almond activated Rules 1 and 9 more strongly, both of which are associated with high morphological uniformity. These patterns reflect a stable correlation between length and mass, two primary factors that define the physical integrity and filling degree of baru almonds.

This behavior is consistent with recent literature that emphasized the role of morphometric uniformity as a predictor of industrial performance and mechanical integrity (Erdoğdu et al., 2025; Özoğul, 2025). According to (Mahmood et al., 2023), centroid-based fuzzy classification produces stable results when applied to biological materials with nonlinear variability, such as nuts and grains, ensuring robustness to small-dimensional variations. Furthermore, the combination of high length and mass is a morphological indicator of maturation uniformity and high storage potential, reinforcing the validity of fuzzy system outputs (Arruda et al., 2021; Rinaldi et al., 2021).

Table 6. Example of the proposed FIS, considering the values of the variables that generate the highest fuzzy quality index, for 2012 and 2013 harvests.

Variable and membership function activation				Rule and activation
Diameter	**Width**	**Length**	**Weight**	
		Year 2012		
Small: 0.0000	Small: 0.0000	Short: 0.0000	Slim: 0.0000	Rule 1 - Strength: 1.0000
Middle: 1.0000	Middle: 0.0000	Middle: 0.0000	Middle: 0.0000	Rule 9 - Strength: 1.0000
Large: 0.0000	Large: 1.0000	Long: 1.0000	Heavy: 1.0000	
		Year 2013		
Small: 0.0000	Small: 0.0000	Short: 0.0000	Slim: 0.0000	Rule 1 - Strength: 1.0000
Middle: 0.9000	Middle: 0.0111	Middle: 0.0000	Middle: 0.0000	Rule 6 - Strength: 0.1000
Large: 0.1000	Large: 0.9889	Long: 1.0000	Heavy: 1.0000	Rule 9 - Strength: 0.9889

In both years, the consistent activation of the same high-quality rules (Rule 1 and Rule 9, Table 6) or low-quality rules (Rule 4, Table 5) reveals the stability and robustness of the fuzzy model, even when dealing with biometric fluctuations between harvests. This consistency is essential for the reproducibility of the FIS, demonstrating its potential application as a morphological quality index for classifying baru seeds.

4. Conclusion, limitations, and future work

This study presented and evaluated a fuzzy inference system (FIS) designed to classify the morphological quality of *Dipteryx alata* (baru) almonds based on four biometric attributes: diameter, width, length, and individual mass. Using a strictly data-driven strategy for calibrating the membership functions (anchors at minimum, quartiles, and maximum), the model was capable of interpretably representing the inter- and intracrop morphological variability observed in the 2012 and 2013 samples. The distributions of the fuzzy quality index and the step-by-step examples (Tables 5 and 6) demonstrated consistency between the rules defined by the experts and the defuzzified outputs using the centroid method, indicating that the FIS captures attribute compositions that are typically associated with greater or lesser physical integrity of the seeds.

Despite its robust interpretation and automatic adaptation of functions to each harvest, the study also highlights the natural limitations of the morphometric approach: by considering only dimensional variables and mass, the index reflects only physical/morphological quality, being unable to replace physiological or biochemical evaluations (germination, vigor, and oil/protein content). Furthermore, the model's sensitivity to specific configurations of the membership functions and the rule base (constructed by experts) needs additional tests to validate reproducibility and generalizability in samples from other regions and harvests.

Additionally, a promising expansion of the methodology would involve incorporating digital image analysis. This approach would enable the automation of biometric variable extraction (diameter, width, and length), replace manual measurements and increase system scalability. Furthermore, it allows the inclusion of new attributes such as color, brightness, and texture, which could serve as indicators of the maturity stage or the presence of damage. In this way, the assessment would no longer rely solely on physical measurements, overcoming one of the limitations of the current model. The integration of these new image-based predictors into the fuzzy inference system has the potential to refine the index, making it more robust and accurate.

Considering the results and identified limitations, the proposed model fulfills its role as an exploratory tool for morphological standardization and as an initial filter in the baru seed sorting chain — that is, as a practical and computationally inexpensive step to separate batches with greater potential for storage and industrial processing. Its linguistic and explainable nature is particularly valuable for incorporation into participatory workflows with technicians and producers, where decisions need to be transparent.

As a perspective for continuity, the fuzzy system should be improved by incorporating new physical, physiological, and biochemical variables, such as moisture content, density, vigor, germination, and chemical composition, to expand its predictive power and more fully represent the quality of baru almonds. Furthermore, the use of optimization and rule learning techniques (genetic algorithms and neuro-fuzzy models) is recommended to automatically adjust the membership functions and weights of the rules, as is the integration of spectral and image data to capture attributes that are not measurable morphometrically. Future studies should also include cross-validation between harvests and regions, multicenter sample expansion, and the development of prototypes applicable in real sorting systems, aiming to consolidate the FIS as a robust, interpretable, and operational tool for quality control of barcodes.

5. References

Arruda, L. C. de, Faria, W. C. S., Siqueira, N. F. M. P., Campos, M. F. de S., & Barros, W. M. de. (2021). Development of cereal bar with baru almonds and pineapple peels. *Research, Society and Development*, *10*(5), e21610514684. https://doi.org/10.33448/rsd-v10i5.14684

Basso, S. M. S. (1999). *Caracterização morfológica e fixação biológica de nitrogênio de espécies de Adesmia DC. e Lotus L.* [Doutorado]. Universidade Federal do Rio Grande do Sul.

Botezelli, L., Davide, A. C., & Malavasi, M. M. (2000). Characteristics of fruits and seeds of four provenances of dipteryx alata vogel. *CERNE*, 9–18.

Castillo, O., & Melin, P. (Eds.). (2023). *Hybrid Intelligent Systems Based on Extensions of Fuzzy Logic, Neural Networks and Metaheuristics* (Vol. 1096). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28999-6

Egea, M. B., de Oliveira Filho, J. G., Campos, S. B., & Lemes, A. C. (2023). The potential of baru (Dipteryx alata Vog.) and its fractions for the alternative protein market. *Frontiers in Sustainable Food Systems*, *7*. https://doi.org/10.3389/fsufs.2023.1148291

Erdoğdu, A., Dayi, F., Yildiz, F., Yanik, A., & Ganji, F. (2025). Combining Fuzzy Logic and Genetic Algorithms to Optimize Cost, Time and Quality in Modern Agriculture. *Sustainability*, *17*(7), 2829. https://doi.org/10.3390/su17072829

Fernandes, D. C., Freitas, J. B., Czeder, L. P., & Naves, M. M. V. (2010). Nutritional composition and protein value of the baru (*Dipteryx alata* Vog.) almond from the Brazilian Savanna. *Journal of the Science of Food and Agriculture*, *90*(10), 1650–1655. https://doi.org/10.1002/jsfa.3997

Gonçalves, L. G. V., Andrade, F. R., Marimon Junior, B. H., Schossler, T. R., Lenza, E., & Marimon, B. S. (2013). Biometria de frutos e sementes de mangaba (Hancornia speciosa Gomes) em vegetação natural na região leste de Mato Grosso. *Revista de Ciências Agrárias*, *36*(1), 31–40.

Lu, J., Ma, G., & Zhang, G. (2024). Fuzzy Machine Learning: A Comprehensive Framework and Systematic Review. *IEEE Transactions on Fuzzy Systems*, *32*(7), 3861–3878. https://doi.org/10.1109/TFUZZ.2024.3387429

Magdalena, L. (2015). Fuzzy Rule-Based Systems. In *Springer Handbook of Computational Intelligence* (pp. 203–218). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43505-2_13

Mahmood, A. M. R., Faraj, J. M., Sharef, A. A., Majeed, H. O., Hamasalih, F. M., & Tahir, N. A. (2023). Evaluation of Wild Almond Genotypes Grown as a Rain-fed Crop in Sulaimani Governorate using RAPD and ISSR Markers. *Kufa Journal for Agricultural Sciences*, *15*(1), 19–33. https://doi.org/10.36077/kjas/2023/v15i1.10214

Mendel, J. M. (2024). *Explainable Uncertain Rule-Based Fuzzy Systems*. Springer International Publishing. https://doi.org/10.1007/978-3-031-35378-9

Mittal, K., Jain, A., Vaisla, K. S., Castillo, O., & Kacprzyk, J. (2020). A comprehensive review on type 2 fuzzy logic applications: Past, present and future. *Engineering Applications of Artificial Intelligence*, *95*, 103916. https://doi.org/10.1016/j.engappai.2020.103916

Miyashita, N. M. R., Hudson, E. A., Rezende, J. de P., Vidigal, M. C. T. R., & Pires, A. C. dos S. (2025). Baru Proteins: Extraction Methods and Techno-Functional Properties for Sustainable Nutrition and Food Innovation. *Foods*, *14*(8), 1286. https://doi.org/10.3390/foods14081286

Morchid, A., Said, Z., Abdelaziz, A. Y., Siano, P., & Qjidaa, H. (2025). Fuzzy logic-based IoT system for optimizing irrigation with cloud computing: Enhancing water sustainability in smart agriculture. *Smart Agricultural Technology*, *11*, 100979. https://doi.org/10.1016/j.atech.2025.100979

Mota, E. E. S., Novaes, C. R. D. B., Silva, L. B. e, & Chaves, L. J. (2020). Structure of the phenotypic variability of fruit and seeds of Dipteryx alata vogel (Fabaceae). *Revista Brasileira de Fruticultura*, *42*(5). https://doi.org/10.1590/0100-29452020003

Özoğul, G. (2025). Applications of artificial intelligence technologies in agriculture: advantages, challenges, risks, prospects, and recommendations. *Cogent Food & Agriculture*, *11*(1). https://doi.org/10.1080/23311932.2025.2568199

Prasad, R., Srivastava, A. K., & Tiwari, R. (2024). Fuzzy Logic-Based Sprinkler Controller for a Precision Irrigation System: A Case Study of Semi-Arid Regions in India. *ECSA-11*, 103. https://doi.org/10.3390/ecsa-11-20504

Rinaldi, M. M., Rocha, F. S., Santos, R. M. dos, Pereira, M. dos S., Queiroz, D. B. V. de, & Morais, F. M. de. (2021). *Production, physical, chemical and functional characterization of baru fruits and seeds (Dipteryx alata Vog., Fabaceae) from Embrapa Cerrados e Arinos, MG harvest 2019*.

Ross, T. J. (2010). *Fuzzy Logic with Engineering Applications*. Wiley. https://doi.org/10.1002/9781119994374

Santos, J. M. dos, Borges, J. A. T., Santos, S. M. dos, Silva, R. M. M. F., Trichez, V. D. K., & Formagio, A. S. N. (2024). Baru (Dipteryx alata): a comprehensive review of its nutritional value, functional foods, chemical composition, ethnopharmacology, pharmacological activities and benefits for human health. *Brazilian Journal of Biology*, *84*. https://doi.org/10.1590/1519-6984.278932

Shihabudheen, K. V., & Pillai, G. N. (2018). Recent advances in neuro-fuzzy system: A survey. *Knowledge-Based Systems*, *152*, 136–162. https://doi.org/10.1016/j.knosys.2018.04.014

Silva, F. A. M., Assad, E. D., & Evangelista, B. A. (2008). Caracterização climática do bioma Cerrado. In S. M. Sano, S. P. Almeida, & J. F. Ribeiro (Eds.), *Cerrado: ecologia e flora* (pp. 69–88).

Sousa, T. L. de, Filho, J. G. de O., Cabassa, I. de C. C., Lemes, A. C., & Egea, M. B. (2024). Unraveling the Potential of Baru (Dipteryx alata Vog.) Fruit Fractions as a Sustainable Food Ingredient: Chemical and Technological Characteristics and Prebiotic Potential. *Sustainability*, *16*(24), 10976. https://doi.org/10.3390/su162410976

Valadão, G., & Souza, Á. N. de S. (2024). Financial and economic viability analysis of baru almond (Dipteryx alata Vogel) agroextractivism in the Urucuia River Valley, Arinos/MG. *Sustainability in Debate*, *15*(2), 103–136. https://doi.org/10.18472/SustDeb.v15n2.2024.54132

Widayat, I. W., Arsyad, A. A., Mantau, A. J., Adhitya, Y., & Köppen, M. (2025). Fuzzy Methods in Smart Farming: A Systematic Review. *Informatica*, 453–489. https://doi.org/10.15388/24-INFOR579

Zuffo, A. M., Andrade, F. R., & Zuffo Júnior, J. M. (2014). Biometric characterization of fruits and seeds baru (Dipteryx alata Vog.) in eastern Mato Grosso, Brazil. *Revista de Ciências Agrárias*, *37*(4), 463–471.

6. Additional Information

6.1 Acknowledgments

We thank the State University of Mato Grosso do Sul, the State University of Mato Grosso, and the State University of Maranhão.

6.2 Funding

There are no funds that sponsored this research.

6.3 conflicts of interest

Describe if there is any conflict of interest between the research presented and the professional life of the authors, or state that there is not.

6.4 Data availability

The data and scripts for implementing the presented methodology are available here: https://github.com/brunobro/a-baru-seed-quality-index-based-on-a-fuzzy-inference-system