Péllico Netto & Behling *New Zealand Journal of Forestry Science* (2026) 56:1
https://doi.org/10.33494/nzjfs562026x413x
E-ISSN: 1179-5395
published on-line: 04/01/2026

RESEARCH ARTICLE Open Access

New Zealand Journal of Forestry Science

New probability density function for biopopulations modelling

Sylvio Péllico Netto and Alexandre Behling*

Federal University of Paraná, Department of Forest Sciences, Graduate Program in Forestry. Av. Pref. Lothário Meissner, 900, Jardim Botânico, Campus III, 80210-170 Curitiba, PR, Brazil

*Corresponding author: alexandre.behling@yahoo.com.br

(Received for publication 11 August 2024; accepted in revised form 4 December 2025)
Editor: Euan G. Mason

Abstract

Background: Biological populations were studied to understand their ecology and to evaluate the relationships between living beings that comprise them. Mathematical functions used in probabilistic models should present multifunctionality, sensitivity, and flexibility to appropriately describe a natural phenomenon. The objective of this study was to develop a new probabilistic distribution with five parameters to maximize its flexibility and ensure a better goodness of fit when compared to other important distributions, such as Beta, Burr, Silva and Pareto.

Methods: New distribution estimators were derived using the mathematical expectation of central and dispersion moments. Estimated values of the parameters were obtained using an optimization process developed by Abel Soares Siqueira, research software engineer at the Netherlands eScience Center in Amsterdam. Data for the application of the developed distribution method were collected at different sites in Brazil, where asymmetry and kurtosis were detected.

Results: The Pellico-Behling Probability Distribution (5P) was applied to fit the datasets for *Cariniana legalis*, *Acacia mearnsii*, and *Eucalyptus saligna*. For the average mortality of 124 species, it was used with (4P). The distribution fitted to sampled datasets was compared with the fitted Beta and Burr (4P) distributions, except for Silva's polynomial distribution that was fitted to the heights of the species *Eucalyptus saligna* and the Pareto distribution to mortality of 124 tropical species from a fragment of a semideciduous seasonal forest, to evaluate and verify its potential and robustness.

Conclusions: The new distribution with five parameters is flexible and produced better goodness of fit than those obtained from the other distributions used for comparative purposes.

Keywords: Transformed Beta; asymmetry; flexibility; aggregative method; ecological experiments.

Introduction

Biological populations are studied to understand their ecology and to evaluate the relationships between living beings that comprise them. Although numerous mathematical models have been used to characterise the phenomena and behaviours of living beings throughout their life process, it is difficult to describe them accurately because life processes are, in essence, complex, and many of them are strongly influenced by fluctuations arising from the actions of biotic and abiotic environmental variables.

Mathematical functions used as probabilistic models require certain features, such as multifunctionality, sensitivity, and flexibility to appropriately describe a natural phenomenon.

In various circumstances working with random variables in the forest environment, we encounter unusual occurrences such as positive and negative asymmetries, discontinuity points within the sample datasets, and the occurrence of accentuated kurtosis. In these samples, few probabilistic distributions are able to assimilate these characteristics.

As a result of these experiences, we decided to deepen our knowledge of how such occurrences could be described by a single probability function, and the only plausible way to ensure this was to increase the flexibility of a mathematical model whose structural composition could be integrated by more than three parameters, although this path is generally discouraged by mathematical statisticians.

The studies and research reported in this study were the result of over 20 years of working with certain distributions, especially Gamma and Beta, because of their characteristics and special ability to describe the behaviours of many biological variables.

Genesis and historical remarks

The first function of interest for describing biological behaviour is known as the gamma function, introduced by Leonhard Euler (born April 15, 1707, in Basel, Switzerland, who died September 18, 1783, in Saint Petersburg, Russia), that is, the factorial of a series of integers, as presented in (1):

$$\int_0^\infty u^{\alpha-1} e^{-u} d_u = \Gamma(\alpha) = (\alpha-1)! = 1.2.3.4.5....(\alpha-1) \qquad (1)$$

As can be seen, the gamma function is not convergent at the point where $u = 0$. To convert it as a probability density function (*pdf*), a constant β was added to it, and with transformations, it became convergent from zero to infinity and was called the Gamma distribution.

$$f(x) = \frac{u^{\alpha-1} e^{\frac{u}{\beta}}}{\beta^\alpha \Gamma(\alpha)} \qquad \text{for } x \geq 0 \qquad (2)$$

Leonhard Euler and Adrien-Marie Legendre (born on September 18, 1752, in Paris, France and who died on January 10, 1833, in the same place) simultaneously developed the integral resulting from the product of two factorials, which became known as the Euler integral of the first type in the early nineteenth century. The derivation of the product of the two gamma functions is presented in (3).

$$\Gamma(x)\,\Gamma(y) = \int_0^\infty e^{-u} u^{x-1} d_u \int_0^\infty e^{-v} v^{y-1} d_v = (x-1)!(y-1)! \qquad (3)$$

After applying these two integrals and transforming them into polar coordinates to operate them more rationally, they finally obtained:

$$\Gamma(x)\Gamma(y) = \int_0^\infty e^{-u} u^{(x+y-1)} d_u \; 2\left(\frac{1}{2}\int_0^1 u^{x-1}(1-u)^{y-1} d_u\right)$$

This last integral was later named by Jacques Philippe Marie Binet (born on February 2, 1786, in Rennes, France and who died on May 12, 1856, in Paris, France) as a beta function and is identified as *B(x,y)* in (4).

$$\Gamma(x)\Gamma(y) = \int_0^\infty e^{-u} u^{(x+y-1)} d_u \; 2\left(\frac{1}{2}\int_0^1 u^{x-1}(1-u)^{y-1} d_u\right) = \Gamma(x+y)\cdot B(x,y) \qquad (4)$$

Therefore, $B(x,y) = \dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ (5)

Next, we consider the derivation of Legendre's double formula, which results in a more attractive and felixible form of the beta function, that is, taking it in the condition obtained in (6).

$$B(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1} d_u \qquad (6)$$

Where Re (x) > 0 and Re (y) > 0

Making up $m = \sqrt{u}$ or $u = m^2$ and $d_u = 2m\, d_m$

$$B(x,y) = \int_0^1 m^{2(x-1)}(1-m^2)^{y-1}\, 2m\, d_m$$

$$B(x,y) = 2\int_0^1 m^{2x-1}(1-m^2)^{y-1} d_m$$

If the beta function is placed in the condition in which the representations of the Bessel and hypergeometric functions can be developed, then taking the form *B (x+1, y+1)* to facilitate the derivation, we have (7).

$$B(x+1, y+1) = 2\int_0^1 m^{2x}(1-m^2)^y d_m \qquad (7)$$

Making up:

$$t = \frac{m^2}{1-m^2},\; t\left(1-m^2\right) = m^2,\; t - t\,m^2 = m^2,$$

$$t = m^2 + t\,m^2,\; t = m^2(1+t) \;\; e \;\; m^2 = \frac{t}{1+t}$$

$$\frac{2d_m}{d_t}=\frac{(1+t)-t}{(1+t)^2}=\frac{1}{(1+t)^2} \quad \text{and} \quad d_m=\frac{1}{2(1+t)^2}d_t$$

Substituting in (7) and rationalising we have:

$$B(x+1,y+1)=\int_0^1 \frac{t^x}{(1+t)^{x+y+2}}d_t$$

and, consequently, in the most usual way we have:

$$B(x,y)=\int_0^1 \frac{t^{x-1}}{(1+t)^{x+y}}d_t \qquad (8)$$

The beta function, as defined in (6), assumes that the variable *u* only exists in the interval $0 \le u \le 1$. Extending this interval to $a \le u \le b$ such that $b > a$, then placing it with a variation in this range results in:

$$B(x,y)=\int_a^b (t-a)^{x-1}(b-t)^{y-1}d_t \qquad a\le t\le b$$

and, consequently, the beta function, in these circumstances, results in (9):

$$B(x,y)=\frac{\Gamma(x)\,\Gamma(y)(b-a)^{x+y-1}}{\Gamma(x+y)} \qquad a\le t\le b \qquad (9)$$

Its application to the theory of probabilities, using the property that the integral of the *pdf*, in the interval of variable *x* between zero and infinity, must be equal to one, Cramér (1951), results in:

$$f(t)=\frac{\Gamma(x+y)(t-a)^{x-1}(b-t)^{y-1}}{\Gamma(x)\Gamma(y)(b-a)^{x+y-1}} \qquad a\le t\le b$$

In modern statistical notation, variables of a probability function are denoted by the letters *x* and *y* to avoid misinterpretation. The parameters of the beta function are now named α and β, and the variable is considered as *x*; thus, the general form of the beta function is presented in (10):

$$f(x)=\frac{\Gamma(\alpha+\beta)(x-a)^{\alpha-1}(b-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)(b-a)^{\alpha+\beta-1}} \qquad a\le t\le b \qquad (10)$$

Loetsch, Zöhrer, & Haller (1973) applied it to fit diameter distributions using datasets from forests of Germany, where *a* and *b* are the minimum and maximum diameters, respectively.

Scolforo (2006) presented it with substitutions to facilitate the practical comprehension by ecologists and forest managers as follows:

$$f(x)=\frac{\Gamma(\alpha+\beta)(d_i-d_{\min})^{\alpha-1}(d_{\max}-d_i)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)(d_{\max}-d_{\min})^{\alpha+\beta-1}}$$

$$d_{\min}\le t\le d_{\max} \qquad (11)$$

Here, we introduce the work of Burr (1942), who suggested 12 different forms of cumulative distribution functions which might be useful for fitting to variables from a variety of datasets, including biological and ecological variables from the forest environment. The relevance of choosing one of these forms of distribution is to facilitate the mathematical analysis to which it is subjected, while attaining the condition of goodness of fit as much as possible (Tadikamalla 1980).

Many attempts have been made to clarify and apply the 12 probability distributions derived from cumulative functions that integrate Burr's system (Burr 1968) and others, to different datasets, such as by Burr & Cislak (1968), who showed that Burr's system covered almost all domains of the main Pearson Type IV and VI distributions, and an important part of that is the Type I or the beta distribution. Hatke (1949), after evaluating the cumulative probability function:

$$F(x)=1-\left(1+x^{\frac{1}{c}}\right)^{-\frac{1}{k}}$$

as proposed by Burr (1942), stated that this approach was a practical tool for fitting a smooth curve to observed data. The fitting method was comparable to that reported by Pearson (1900) and others and was accomplished with simple calculations. These distributions are fitted by the method of moments, and their theoretical frequencies are obtained by the evaluation of consecutive values of *F(x)* using calculating machines and logarithms, and by taking the derivative of *NF(x)*. No integration or heavy interpolation is involved, such as may be required in fitting a classical frequency function; Rodriguez (1977) focused specifically on Burr's family of distributions of type XII, with the generic cumulative:

$$F(x)=1-\left(1+x^{c}\right)^{-k}$$

function, which yields a wide range of values of skewness and kurtosis; and Ferose & Aslam (2013), who derived maximum likelihood estimators (MLE) to obtain the parameters of a Burr type V distribution based on left-censored samples, including confidence intervals for the parameters. A simulation study was also conducted to investigate the performance of point and interval estimates.

Type XII is the most well-known applied Burr distribution to many datasets of different scientific origins because it yields a wide range of values of skewness and kurtosis and can be used to fit almost any given set of unimodal data (Nadarajah et al. 2012); its *pdf* is presented in (12).

$$f(x,c,k)=ck\,x^{c-1}\left(1+x^{c}\right)^{-(k+1)} \qquad (12)$$

Burr's distributions have appeared in the literature under different names because of their relationship with various other distributions, namely, the Pareto Type II (Lomax) when *c=1*, Srivastava (1965); when *k=1* it becomes the Fisk distribution (Fisk 1961), which is a special case of the Champernowne distribution (Champernowne 1952); in its inverse case; for *1/X* it becomes Dagum's distribution (Dagum 1977), and other special cases such as the Compound Weibull, Weibull-Exponential, logistic, log-logistic, Weibull, and Kappa family of distributions (Tadikamalla, 1980). These distributions can be used to model a wide variety of phenomena, including forest variables, to describe ecological and production information throughout a forest's lifetime.

The objective of this study was to develop a new distribution with five parameters; very flexible and with better goodness of fit when compared to Beta, Burr (3P), Silva, and Pareto distributions.

Methods

Development of the new probability density function

Before presenting the derivation of the probabilistic distribution proposed by Péllico-Behling, some considerations should be made:

(1) The statistical procedures used by Burr (1942) presuppose designing a cumulative probabilistic function by order statistics and deriving it to obtain the probability density function. This is called the *derivative method*; however, as can be seen in the several functions mentioned above, this procedure generates parental parameters in the numerator of the ratio that makes up the resulting function; and

(2) Since we wished to obtain a probability density function with five parameters capable of attaining maximum flexibility of the resulting function, we decided to generate the probability density function using another statistical procedure called the *aggregative method*.

Considering the beta function as presented in Equation 8 and with the transformations already incorporated into it, we have (13).

$$B(\alpha,\beta)=\int_0^1 \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}\, d_t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad (13)$$

Several applications of the beta distribution to diameter and height data of trees from native and planted forests in Brazil showed that it was not sufficiently flexible for good fitness in cases of severe asymmetry and kurtosis, duly evaluated by applying the Kolmogorov-Smirnov test (Kolmogorov 1933; Smirnov 1948).

The flexibility of the beta function was initially achieved by expanding it from two to three parameters as follows:

Include transformations in the Bessel function without mischaracterising it as the beta function, i.e.,

$$\int_0^1 B_{(\alpha,\beta)} d_t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\left(\frac{1}{b}\right)^{\alpha} b^{\alpha+\beta}\,\Gamma(\alpha+\beta)} \qquad (14)$$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)\, b^{\alpha}}{b^{\alpha+\beta}\,\Gamma(\alpha+\beta)} = \frac{\Gamma(\alpha)\Gamma(\beta)\, b^{\alpha} b^{-(\alpha+\beta)}}{\Gamma(\alpha+\beta)} = \frac{\Gamma(\alpha)\Gamma(\beta)\, b^{\alpha-(\alpha+\beta)}}{\Gamma(\alpha+\beta)}$$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)\, b^{-\beta}}{\Gamma(\alpha+\beta)}$$

Taking only the partial components presented below, it can be shown that they are the result of a transformed integral:

$$\frac{\Gamma(\alpha)\Gamma(\beta)\, b^{-\beta}}{\Gamma(\alpha+\beta)} = b^{-\beta}\int_0^1 \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}\, d_t = b^{\alpha-1+1-\alpha-\beta}\int_0^1 \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}\, d_t$$

$$= b^{\alpha-1} b b^{-(\alpha+\beta)}\int_0^1 \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}\, d_t = \frac{b^{\alpha-1}\, b}{b^{\alpha+\beta}}\int_0^1 \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}\, d_t = \int_0^1 \frac{b\, b^{\alpha-1}\, t^{\alpha-1}}{b^{\alpha+\beta}(1+t)^{\alpha+\beta}}\, d_t$$

$$= \int_0^1 \frac{b\, b^{\alpha-1}\, t^{\alpha-1}}{(b+bt)^{\alpha+\beta}}\, d_t = \int_0^1 \frac{b(bt)^{\alpha-1}}{(b+bt)^{\alpha+\beta}}\, d_t \qquad (15)$$

By setting $bt = u$, deriving $bd_t = d_u$, $d_t = d_u/b$, and replacing them, we have:

$$= \int_0^1 \frac{b\, u^{\alpha-1}}{(b+u)^{\alpha+\beta}}\,\frac{d_u}{b} = \int_0^1 \frac{u^{\alpha-1}}{(b+u)^{\alpha+\beta}}\, d_u \qquad (16)$$

Consequently, by naming the resulting integral with the three parameters of *PB (b, α, β)* we have:

$$\frac{\Gamma(\alpha)\Gamma(\beta)\, b^{-\beta}}{\Gamma(\alpha+\beta)} = \int_0^1 \frac{u^{\alpha-1}}{(b+u)^{\alpha+\beta}}\, d_u = \int_0^1 PB(b,\alpha,\beta)\, d_u \qquad (17)$$

Naming $\alpha - 1 = a$, $\alpha+\beta = c$ and $u = x$, the new function with three parameters is set to:

$$f(x) = x^a/(b+x)^c \qquad (18)$$

Some families of distributions have been derived to approximate some already known distributions as much as possible. These families are commonly referred to as distribution or frequency curve systems. Although theoretical explanations may highlight the relevance of a system, such arguments must first be evaluated in terms of their practicality. Some additional requirements are the ease of computing and algebraic manipulation; however, it is desirable to include as few parameters as possible in defining a system member. In most circumstances, it is sufficient to have up to four parameters. Normally, at least three parameters are needed, and the inclusion of a fourth parameter can bring about a notable improvement, but it should be critically evaluated whether this decision is worthwhile (Johnson & Kotz 1970). Even considering this important warning, we decided to incorporate two more parameters c and d into the beta distribution to make it as flexible as possible.

$$\frac{1}{d\,c^{\alpha}}\int_0^1 PB(b,\alpha,\beta)\,d_u = \frac{1}{d\,c\,c^{\alpha-1}}\int_0^1 \frac{u^{\alpha-1}}{(b+u)^{\alpha+\beta}}\,d_u \quad (19)$$

By making $\alpha = \tau/d$ and substituting it into the resulting new function, we have:

$$\frac{1}{d\,c^{\alpha}}\int_0^1 PB(b,\alpha,\beta)\,d_u = \frac{1}{d\,c\,c^{\frac{\tau}{d}-1}}\int_0^1 \frac{u^{\frac{\tau}{d}-1}}{(b+u)^{\frac{\tau}{d}+\beta}}\,d_u$$

$$= \frac{1}{d}\int_0^1 \frac{u^{\frac{\tau}{d}-1}c^{-\left(\frac{\tau}{d}-1\right)}}{(b+u)^{\frac{\tau}{d}+\beta}}\,d_u = \frac{1}{d}\int_0^1 \left(\frac{u}{c}\right)^{\frac{\tau}{d}-1}(b+u)^{-\left(\frac{\tau}{d}+\beta\right)}d_u = \frac{1}{d}\int_0^1 \frac{\left(\frac{u}{c}\right)^{\frac{\tau-1}{d}}}{\left(\frac{u}{c}\right)^{\frac{d-1}{d}}}(b+u)^{-\left(\frac{\tau}{d}+\beta\right)}d_u \quad (20)$$

Note that,

$$\frac{\tau-1}{d}-\frac{d-1}{d}=\frac{\tau-d}{d}=\frac{\tau}{d}-1$$

and

$$= \frac{1}{d}\int_0^1 \left(\frac{u}{c}\right)^{\frac{\tau}{d}-1}(b+u)^{-\left(\frac{\tau}{d}+\beta\right)}\frac{d_u}{\left(\frac{u}{c}\right)^{\frac{d-1}{d}}} = \int_0^1 \left(\frac{u}{c}\right)^{\frac{\tau}{d}-1}(b+u)^{-\left(\frac{\tau}{d}+\beta\right)}\frac{d_u}{d\,\left(\frac{u}{c}\right)^{\frac{d-1}{d}}}$$

Making:

$$x = \left(\frac{u}{c}\right)^{\frac{1}{d}};\; u = c\,x^{d};\; d_u = c\,d\,x^{d-1}d_x \text{ and } d_x = \frac{d_u}{c\,d\,x^{d-1}}$$

Replacing them we have:

$$\int_0^1 x^{\tau-1}\left(b+cx^{d}\right)^{-\left(\frac{\tau}{d}+\beta\right)}d_x = \int_0^1 \frac{x^{\tau-1}}{\left(b+cx^{d}\right)^{\left(\frac{\tau}{d}+\beta\right)}}\,d_x \quad (21)$$

Finally, making τ – 1 = a, (τ/d) + β = e and replacing them in the integral, we have:

$$\int_0^1 \frac{x^{a}}{\left(b+cx^{d}\right)^{e}}\,d_x = PB\left(a,b,c,d,e\right) \quad (22)$$

The new function with five parameters is therefore defined as:

$$f(x) = \frac{x^{a}}{(b+cx^{d})^{e}} \quad (23)$$

where a, c, d, and e are the parameters responsible for the shape of the curve, which makes it flexible in its fitting to occurrences of asymmetry and kurtosis, which forces it to reach the modal point, and b is the parameter responsible for the change in position along the x-axis.

For the transformation of the proposed model into a probability density function, considering the property specified by Cramér (1951), we have:

$$\int_0^1 f(x)\,d_x = \int_0^1 \frac{x^{a}}{\left(b+c\,x^{d}\right)^{e}}\,d_x = \frac{\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{a+1}{d}} b^{e}\,d\,\Gamma(e)} \quad (24)$$

Consequently, the $pdf_{(x)}$ is defined as Case 1:

$$pdf_{(x)} = \frac{\left(\frac{c}{b}\right)^{\frac{a+1}{d}} b^{e}\,d\,\Gamma(e)\,x^{a}}{\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)(b+c\,x^{d})^{e}} = k_1\,\frac{x^{a}}{(b+c\,x^{d})^{e}} \quad (25)$$

where k_1 is the inverse of the result of the integration of $f(x)$

$$k_1 = \frac{\left(\frac{c}{b}\right)^{\frac{a+1}{d}} b^{e}\,d\,\Gamma(e)}{\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)} \quad (26)$$

The cumulative distribution function can be obtained by integrating it into cumulative intervals for variable x for $x \geq 0$, as shown in (20):

$$\int_0^{x_i} f\,dp_{(x)}dx = \frac{\left(\frac{c}{b}\right)^{\frac{a+1}{d}} b^{e}\,d\,\Gamma(e)}{\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)}\int_0^{x_i}\frac{x^{a}}{(b+c\,x^{d})^{e}}\,d_x \quad x \geq 0 \quad (27)$$

The solution for this integral, instead of applying complicated hypergeometric functions to obtain the cumulative function of the Péllico-Behling distribution, was to calculate the additive rectangular areas of small classes (1/1,000 in size), as an approximation to the areas obtained by the successive integration procedure, as presented in (21):

$$F(x) = \frac{\sum_{i=1}^{x} f(x)}{\sum_{i=1}^{k} f(x)} \quad (28)$$

where k is the number of class intervals that are as small as possible, which in the present case was 1,000.

Derivation of Distribution Parameters

The distribution estimators were derived using the mathematical expectation of the central and dispersion moments.

Arithmetic Mean:

$$m_1 = E(X) = \int_0^1 x\,f(x)\,d_x = k\int_0^1 \frac{x\,x^{a}}{\left(b+c\,x^{d}\right)^{e}}\,d_x = k\int_0^1 \frac{x^{a+1}}{\left(b+c\,x^{d}\right)^{e}}\,d_x \quad (29)$$

$$E(X)=\frac{\left(\frac{c}{b}\right)^{\frac{a+1}{d}} b^{e}\, d\,\Gamma(e)}{\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)}\cdot\frac{\Gamma\left(e-\frac{a+2}{d}\right)\Gamma\left(\frac{a+2}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{a+2}{d}} b^{e}\, d\,\Gamma(e)}=\frac{\Gamma\left(e-\frac{a+2}{d}\right)\Gamma\left(\frac{a+2}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{1}{d}}\Gamma\left(e-\frac{a+1}{d}\right)\left(\Gamma\left(\frac{a+1}{d}\right)\right)} \quad (30)$$

$$\mu_x=\frac{\Gamma\left(e-\frac{a+2}{d}\right)\Gamma\left(\frac{a+2}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{1}{d}}\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)} \quad \text{Arithmetic mean (31)}$$

Variance:

$$m_2=E\left(X-\mu_x\right)^2=E\left(X^2\right)-\mu^2 \quad (32)$$

$$E\left(X^2\right)=k\int_0^1 x^2\, f(x)\,d_x=k\int_0^1\frac{x^2\,x^a}{\left(b+c\,x^d\right)^e}d_x=$$

$$k\int_0^1\frac{x^{a+2}}{\left(b+c\,x^d\right)^e}d_x=k\,\frac{\Gamma\left(e-\frac{a+3}{d}\right)\Gamma\left(\frac{a+3}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{a+3}{d}} b^e\,d\,\Gamma(e)} \quad (33)$$

$$\sigma_x^2=\frac{\Gamma\left(e-\frac{a+3}{d}\right)\Gamma\left(\frac{a+3}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{2}{d}}\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)}-\mu_x^2 \quad \text{Variance (34)}$$

and the standard deviation is equal to:

$$\sigma_x=\sqrt{\frac{\Gamma\left(e-\frac{a+3}{d}\right)\Gamma\left(\frac{a+3}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{2}{d}}\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)}-\mu_x^2} \quad (35)$$

Mode:
Taking the function as formalised in (23), we have:

$$f(x)=\frac{x^a}{\left(b+c\,x^d\right)^e} \quad (36)$$

Deriving it with respect to *x*, we have

$$\frac{d\,f(x)}{d_x}=\frac{\left(b+c\,x^d\right)^e a\,x^{a-1}-e\left(b+c\,x^d\right)^{e-1}\left(c\,d\,x^{a+d-1}\right)}{\left(b+c\,x^d\right)^{2e}} \quad (37)$$

Making $d\,f(x)/d_x = 0$ and equalising, we have:

$$x^a\,e\left(b+c\,x^d\right)^{e-1}c\,d\,x^{d-1}=\left(b+c\,x^d\right)^e a\,x^{a-1}$$

$$\frac{x^a e\left(b+c\,x^d\right)^{e-1}c\,d\,x^{d-1}}{\left(b+c\,x^d\right)^e}=a\,x^{a-1}$$

$$a\left(b+c\,x^d\right)=\frac{x^a\,x^{d-1}\,e\,c\,d}{x^{a-1}}$$

$$a\left(b+c\,x^d\right)=x^d\,e\,c\,d$$

$$a\,b=x^d e\,c\,d-a\,c\,x^d$$

$$a\,b=x^d\left(e\,c\,d-a\,c\right) \quad (38)$$

$$x=\left(\frac{a\,b}{c\,(e\,d-a)}\right)^{\frac{1}{d}} \quad \text{Mode} \quad (39)$$

Inflection Points:
Consider the result of the first derivative of *f(x)*, as stated in (40), and we have:

$$\frac{d\,f(x)}{d_x}=\frac{\left(b+c\,x^d\right)^e a\,x^{a-1}-e\left(b+c\,x^d\right)^{e-1}\left(c\,d\,x^{a+d-1}\right)}{\left(b+c\,x^d\right)^{2e}} \quad (40)$$

Simplifying, we have:

$$\frac{d\,f(x)}{d_x}=\frac{\left(b+c\,x^d\right)^{e-1}\left[\left(b+c\,x^d\right)a\,x^{a-1}-c\,d\,e\,x^{a+d-1}\right]}{\left(b+c\,x^d\right)^{2e}}$$

$$\frac{d\,f(x)}{d_x}=\frac{\left(b+c\,x^d\right)a\,x^{a-1}-c\,d\,e\,x^{a+d-1}}{\left(b+c\,x^d\right)^{e+1}}=$$

$$\frac{a\,b\,x^{a-1}+a\,c\,x^{a+d-1}-c\,d\,e\,x^{a+d-1}}{\left(b+c\,x^d\right)^{e+1}}$$

$$\frac{d\,f(x)}{d_x}=\frac{a\,b\,x^{a-1}+x^{a+d-1}\left(a\,c-c\,d\,e\right)}{\left(b+c\,x^d\right)^{e+1}} \quad (41)$$

Taking now the second derivative of (41), we have:

$$\frac{d^2 f(x)}{d_x}=\frac{\left(b+c\,x^d\right)^{e+1}\left[a\,b(a-1)x^{a-2}+c(a-d\,e)(a+d-1)x^{a+d-2}\right]}{\left(b+c\,x^d\right)^{2(e+1)}}.$$

$$-\frac{\left[a\,b\,x^{a-1}+c\left(a-d\,e\right)x^{a+d-1}\right]\left(e+1\right)\left(b+c\,x^d\right)^e c\,d\,x^{d-1}}{\left(b+c\,x^d\right)^{2(e+1)}} \quad (42)$$

Making $d^2 f(x)/d_x = 0$ and equalising, we have:

$$\frac{\left(b+cx^d\right)^{e+1}\left[\left(a^2b-ab\right)x^{a-2}+c\left(a-de\right)\left(a+d-1\right)x^{a+d-2}\right]}{\left(b+cx^d\right)^e}= \\ =\left[abx^{a-1}+c\left(a-de\right)x^{a+d-1}\right]\left(e+1\right)cdx^{d-1} \quad (43)$$

Simplifying we have:

$$\left(b+cx^d\right)\left[\left(a^2b-ab\right)x^{a-2}+c\left(a-de\right)\left(a+d-1\right)x^{a+d-2}\right]= \\ =\left[abx^{a-1}+c\left(a-de\right)x^{a+d-1}\right]\left(e+1\right)cdx^{d-1} \quad (44)$$

Expanding the terms we have:

$$\left(a^2b^2-ab^2\right)x^{a-2}+bc\left(a-de\right)\left(a+d-1\right)x^{a+d-2}+\left(a^2b-ab\right)cx^{a+d-2} \\ +c^2\left(a-de\right)\left(a+d-1\right)x^{a+2d-2}= \\ =ab\left(e+1\right)cdx^{a+d-2}+c^2d\left(a-de\right)\left(e+1\right)x^{a+2d-2} \quad (45)$$

Note that with all equality terms are multiplied by x^{a-2} and simplifying we have:

$$\left(a^2b^2-ab^2\right)+bc\left(a-de\right)\left(a+d-1\right)x^d+\left(a^2b-ab\right)cx^d+c^2\left(a-de\right)\left(a+d-1\right)x^{2d}- \\ -ab\left(e+1\right)cdx^d-c^2d\left(a-de\right)\left(e+1\right)x^{2d}=0 \quad (46)$$

As can be seen, the algebraic result is an equation of the second degree, which is presented in a more appropriate manner, resulting in:

$$x^{2d}\left[c^2\left(a-de\right)\left(a+d-1\right)-c^2d\left(a-de\right)\left(e+1\right)\right]+ \\ x^d\left[bc\left(a-de\right)\left(a+d-1\right)+\left(a^2b-ab\right)c-abcd\left(e+1\right)\right]+ \\ ab^2\left(a-1\right)=0 \quad (47)$$

Rationalising, we have:

$$x^{2d}\left\{c^2\left(a-de\right)\left[\left(a+d-1\right)-d\left(e+1\right)\right]\right\}+x^d\left\{c\left[b\left(a-de\right)\left(a+d-1\right)\right.\right. \\ \left.\left.+ab\left(a-1\right)-abd\left(e+1\right)\right]\right\}+ab^2\left(a-1\right)=0 \quad (48)$$

Making up $x^{2d} = z^2$ and solving it, we have:

$$z_i=\frac{-\{c[b(a-de)(a+d-1)+ab(a-1)-abd(e+1)]\}\pm}{2\{c^2(a-de)[(a+d-1)-d(e+1)]\}}$$

$$\frac{\sqrt{\{c[b(a-de)(a+d-1)+ab(a-1)-abd(e+1)]\}^2-}}{2\{c^2(a-de)[(a+d-1)-d(e+1)]\}}$$

$$\frac{\sqrt{-4\left\{c^2\left(a-de\right)\left[\left(a+d-1\right)-d\left(e+1\right)\right]\right\}ab^2\left(a-1\right)}}{2\left\{c^2\left(a-de\right)\left[\left(a+d-1\right)-d\left(e+1\right)\right]\right\}} \quad (49)$$

Rationalising, we have:

$$z_i=\frac{-bc[2a(a-1)+de(1-d-2a)]\pm}{2c^2[a(a-1)+de(de+1-2a)]}$$

$$\frac{\sqrt{b^2c^2[2a(a-de-1)+de(1-d)]^2-4b^2c^2[a(a-1)+de(de+1-2a)]a(a-1)}}{2c^2[a(a-1)+de(de+1-2a)]}$$

and $x_i = z_i^{\frac{1}{d}}$ gives us the inflection points (50)

Asymmetry (A):

Considering the result of the third moment of *pdf(x)* in (23), we have:

$$m_3=E\left(X-\mu_x\right)^3=E\left(X^3\right)-3\left[E\left(X\right)E\left(X^2\right)\right]+2\left[E\left(X\right)\right]^3 \quad (51)$$

or: $$m_3=E\left(X^3\right)-3\left[\mu_x\left(\sigma_x^2+\mu_x^2\right)\right]+2\left(\mu_x\right)^3 \quad (52)$$

and: $$E\left(X^3\right)=\frac{\Gamma\left(e-\frac{a+4}{d}\right)\Gamma\left(\frac{a+4}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{3}{d}}\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)} \quad (53)$$

The asymmetry A is obtained by: $A=\frac{m_3}{\sigma^3}$ (54)

Where σ is the standard deviation derived in (35).

Kurtosis (K):

Considering the result of the fourth moment of $pdf_{(x)}$ in (23), we have:

$$m_4=E\left[X-E\left(X\right)\right]^4=E\left(X^4\right)-4\left[E\left(X\right)E\left(X^3\right)\right]+ \\ 6\left\{\left[E\left(X\right)\right]^2E\left(X^2\right)\right\}-3\left[E\left(X\right)\right]^4 \quad (55)$$

or: $$m_4=E\left(X^4\right)-4\left\langle\mu_x\left\{m_3+3\left[\mu_x\left(\sigma_x^2+\mu_x^2\right)\right]-2\left(\mu_x\right)^3\right\}\right\rangle+ \\ 6\left[\mu_x^2\left(\sigma_x^2-\mu_x^2\right)\right]-3\left(\mu_x\right)^4 \quad (56)$$

and: $$E\left(X^4\right)=\frac{\Gamma\left(e-\frac{a+5}{d}\right)\Gamma\left(\frac{a+5}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{4}{d}}\Gamma\left(e-\frac{a+1}{d}\right)\Gamma\left(\frac{a+1}{d}\right)} \quad (57)$$

The Kurtosis K, given that its value in the normal distribution is equal to three, is obtained by:

$$K=3-\frac{m_4}{\sigma^4} \quad (58)$$

Where σ is the standard deviation derived in (Equation 35).

Conditions for the existence of the probability density function

The existence of the $pdf_{(x)}$ requires that the following conditions must be met:

(1) The coefficients must assume values greater than zero, that is, $a > 0$, $b > 0$, $c > 0$, $d > 0$, $e > 1$, and $e \geq 2a$, which are imposed by $pdf_{(x)}$ itself (Equation 26) to ensure the characteristic of unimodality and avoid the occurrence of indeterminacy; and

(2) Considering the parameters, arithmetic mean (Equation 31) and variance (Equation 34), the restriction $d \neq (a + 2)$ is added.

Simplifications of the derived function

As can be seen, the derived function has five parameters, namely, *a, b, c, d*, and *e*, as presented in (Equation 25), which is the most generic unimodal distribution and named Case 1.

Case 2. By setting $a = 0$, the distribution becomes hyperbolic and can be summarised as follows:

The Péllico-Belling Hyperbolic function:

$$f(x_2) = k_2 \frac{1}{\left(b + c\,x^{d}\right)^{e}} \tag{59}$$

Arithmetic mean $$\mu_{x2} = \frac{\Gamma\left(e - \frac{2}{d}\right)\Gamma\left(\frac{2}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{1}{d}}\Gamma\left(e - \frac{1}{d}\right)\Gamma\left(\frac{1}{d}\right)} \tag{60}$$

Variance $$\sigma_{x2}^{2} = \frac{\Gamma\left(e - \frac{3}{d}\right)\Gamma\left(\frac{3}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{2}{d}}\Gamma\left(e - \frac{1}{d}\right)\Gamma\left(\frac{1}{d}\right)} - \mu_{x2}^{2} \tag{61}$$

Mode: $x_2 = 0$ There is no mode (62)

Inflexion Points: In this case, there are no inflexion points.

Asymptry: $$E\left(X_2^3\right) = \frac{\Gamma\left(e - \frac{4}{d}\right)\Gamma\left(\frac{4}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{3}{d}}\Gamma\left(e - \frac{1}{d}\right)\Gamma\left(\frac{1}{d}\right)} \tag{63}$$

and: $$m_{32} = E\left(X_2^3\right) - 3\left[\mu_{x2}\left(\sigma_{x2}^2 + \mu_{x2}^2\right)\right] + 2\left(\mu_{x2}\right)^3 \tag{64}$$

The asymmetry is obtained in (Equation 65):

$$A_2 = \frac{m_{32}}{\sigma_2^3} \tag{65}$$

Kurtosis: $$E\left(X_2^4\right) = \frac{\Gamma\left(e - \frac{5}{d}\right)\Gamma\left(\frac{5}{d}\right)}{\left(\frac{c}{b}\right)^{\frac{4}{d}}\Gamma\left(e - \frac{1}{d}\right)\Gamma\left(\frac{1}{d}\right)} \tag{66}$$

and:

$$m_{42} = E\left(X_2^4\right) - 4\left\langle \mu_{x2}\left\{m_{32} + 3\left[\mu_{x2}\left(\sigma_{x2}^2 + \mu_{x2}^2\right)\right] - 2\left(\mu_{x2}\right)^3\right\}\right\rangle + 6\left[\mu_{x2}^2\left(\sigma_{x2}^2 - \mu_{x2}^2\right)\right] - 3\left(\mu_{x2}\right)^4 \tag{67}$$

The kurtosis is obtained in (68) is:

$$K_2 = 3 - \frac{m_{42}}{\sigma_2^4} \tag{68}$$

Data

Data for the application of the developed distribution were collected at different sites in Brazil, where appropriate circumstances of asymmetry and kurtosis were detected.

Mortality and evaluation of species in tropical forests (Positive Asymmetry, Kurtosis, and application of case 2 of the distribution)

Sample data were collected in a fragment of the semi-deciduous seasonal forest located at Reata Farm, in Cassia, MG, Brazil. Nine sampling units of 1 ha each were measured and subdivided into 100 subunits of 100 m^2, totalling 9 ha as primary units and 900 subunits, in which the census was conducted.

The area is in the municipality of Cassia, southern region of the state of Minas Gerais, Brazil, with approximately 200 ha and 90 ha of seasonal semideciduous forest in a climax state that is untouched and located between: latitude 20°20' and 20°40' S and longitude 46°40' and 47°00' W.

According to RADAMBRASIL (1978), the studied region is characterised by the remaining morpho-structural domain of folded chains, showing traces of these structures, with occasional exposures of their basements. The area in question is in the Alto do Rio Grande Plateau Region, with average altitudes around 680 m above sea level.

According to UFV (2010), in this region, soil variations classified as Dystrophic Red Latosol occur, featuring mineral, non-hydromorphic soils, and more specifically, the typical moderate A, medium texture, sub-deciduous forest phase, flat relief, and smooth undulated type.

The climate in the region of Cassia, MG, Brazil, according to Köppen's classification, is of the Cwa (altitudinal tropical) type, with rigorous and rainy summers, an annual precipitation of 1200–1400 mm, and average annual temperatures of 26.5 °C (maximum) and 19.5 °C (minimum).

Data on the height distribution of the species *Cariniana legalis* (Mart.) Kuntze and an average mortality of 124 species, evaluated in the period 1996–2001 from this area, were used to illustrate the fit of the new probability distribution to height data with strong positive asymmetry and mortality by diameter classes under conditions common in tropical forests, such as those with negative exponential shape.

Plantation 1 (Symmetry)

Data from eight temporary plots sampled in commercial plantations of black wattle in the municipalities of Cristal and Piratini in the state of Rio Grande do Sul, Brazil within seven years were used for the present study. The areas are located at 30°59'59" S and 52°02'54" W, and 31°26'52" S and 53°06'14" W, respectively. The local altitudes vary between 320 m and 370 m above sea level.

According to the Köppen climate classification, the climate of the region for Cristal is Cfa or Piratini Cfb. The average annual temperature is 18.0 °C for Cristal and 16.5 °C for Piratini and the average annual precipitation is 1309 mm for Cristal and 1507 mm for Piratini.

The study sites were gently undulating to undulating, with the soil type Neossols Regolithic for Cristal and soil type Litholic Neossols for Piratini.

Data from the diameter distribution of *Acacia mearnsii* De Wild from these areas were used to illustrate the fit of the new probability distribution for the situation in which the data are close to normality.

Plantation 2 (Negative Asymmetry)

Data illustrating negative asymmetry were sampled in an experiment implemented in the Ibiti Forest Park owned by the company Ripasa S.A. Cellulose and Paper, located in the municipality of Itararé, SP (data kindly made available by Professor Dr. Carlos Roberto Sanquetta – Federal University of Paraná). The local geographic coordinates are: 24°09’ S and 49°19’ W at an altitude of 900 m. According to the Köppen climate classification, the climate of the region is Cfa, sub-humid sub-temperate, with an annual average temperature of 20.3 °C and an annual average precipitation of 1371 mm.

The study site is slightly undulating, with soil of a deep typical dark red dystrophic Latosol type.

Data of the species *Eucalyptus saligna* Sm. from this area was used to illustrate the fit of the new probability distribution to a height dataset with strong negative asymmetry.

The data of the three species mentioned and mortality of 124 tropical species are summarised in Table 1, where the frequencies for graphical presentations were transformed into relative dispersions to configure a harmonious view of different datasets.

Fitting the Péllico-Behling distribution

The Péllico-Behling distribution was fitted to different datasets, first using different softwares, such as SAS, MATLAB, Statistics, SPSS, R, and Table Curve, but the results for the coefficients were not the same in each one of them as expected. We noted that this divergence is due to a laerge number of possible combinations of the five parameters that can reach goodness of fit for the proposed distribution. We searched for an alternative solution for this problem and discovered a function proposed by mathematician Dr. Abel Soares Siqueira, who used an optimisation procedure, and the resulting function was denominated OPTMBEL and is elucidated in (40), using the Julia language (Julia language 2025, Bezanson et al. 2025). The following packages were used: CSV, DataFrames, Plots, ADNLPModels, ForwardDiff, NLPModels, JSOSolvers, LinearAlgebra, Logging, Printf, Percival,

Table 1. Datasets used to fit the Pellico-Behling Probability Distribution.

Cariniana legalis			***Acacia mearnsii***			***Eucalyptus saligna***			**Mortality**		
Heights (1) (m)			**DBH (2) (cm)**			**Heights (3) (m)**			**Tropical species (124) (n)**		
Classes		**Freq.**	**Classes**		**Freq.***	**Classes**		**Freq.***	**Classes**		**Freq.***
Abs.	**Rel.**		**Abs.**	**Rel.**		**Abs.**	**Rel.**		**Abs.**	**Rel.**	
2.5	0.045	4	4	0.050	29	3	0.045	1	14	0.062	323
7.5	0.136	78	6	0.150	55	5	0.136	5	22	0.188	112
12.5	0.227	141	8	0.250	178	7	0.227	7	30	0.312	55
17.5	0.318	98	10	0.350	146	9	0.318	13	38	0.438	28
22.5	0.409	70	12	0.450	292	11	0.409	22	46	0.562	6
27.5	0.500	48	14	0.550	335	13	0.500	62	54	0.688	1
32.5	0.591	29	16	0.650	230	15	0.591	122	62	0.812	6
37.5	0.682	12	18	0.750	116	17	0.682	168	70	0.938	1
42.5	0.773	6	20	0.850	51	19	0.773	37			
47.5	0.863	4	22	0.950	14	21	0.863	3			
52.5	0.954	3				23	0.954	0			
Total		493			1446			440			532

Abs: absolute values. Rel: relative values. Freq: Frequency.

and NLPModelsIpopt to develop the function. In the OPTMBEL function, *xdata* represents the frequency classes and *ydata* represents the frequency of the classes. The outputs of the function are the values of the five coefficients of the Péllico-Behling distribution. The primary objective of the function is to minimise the sum of squared residuals between the observed frequency values and those estimated by the probability density function *(pdf*). This process involves optimising the *pdf* to ensure that the estimated frequencies closely match the observed data, thereby enhancing the accuracy and reliability of the model.

The details of the OPTMBEL function are provided in Appendix 1. The OPTMBEL function fits a nonlinear mathematical model defined by the following expression:

$$f_{(x)} = \frac{x^{p1}}{(p2+p3\ x^{p4})^{p5}}$$

to the observed data, *xdata* and *ydata*. Using constrained optimisation with the Ipopt solver via the ADNLPModels package, the function minimises the sum of squared residuals between the model and the data, along with a regularisation term to prevent overfitting. The model is subject to three nonlinear parameter constraints to ensure the desired mathematical properties, as described above. Upon completion of the fitting process, the function prints the estimated parameters, the optimisation status, and the residual of the objective function. It also generates output files in both text and LaTeX formats, containing the initial parameters, optimal parameters, evaluated constraints, and process status, thus supporting documentation and reproducibility of the results

Considering the complexity of the statistical estimators of the distribution, a function called ESPN was developed to obtain them using the Julia language. In this function, *a, b, c, d,* and *e* are the distribution parameters. The output results were the mean, mode, variance, standard deviation, skewness, kurtosis, and inflection points.

The details of the ESPN function are provided in Appendix 2. This function computes and displays various statistical measures and characteristics of the distribution parameterised by *a*, *b*, *c*, *d*, and *e* (corresponding to *p1*, *p2*, *p3*, *p4*, and *p5* obtained from the OPTMBEL function) of the Péllico-Behling distribution. Specifically, the function calculates the mean, variance, standard deviation, and mode of the distribution using expressions involving the gamma function (Γ), and determines the inflection points of the density curve. Additionally, it computes measures of skewness and kurtosis, providing a detailed description of the distribution's shape. These calculations are carried out analytically based on nonlinear relationships among the parameters, enabling a comprehensive analysis of the statistical properties of the defined distribution. The results are printed to the console to facilitate user interpretation.

Goodness of fit

The goodness of fit of the fitted distributions was evaluated by the application of the Kolmogorov-Smirnov test (Kolmogorov 1933; Smirnov 1948) at a 95% probability.

Distributions additionally selected to fit the datasets to be compared to the Péllico-Behling distribution.

Beta distribution

From Euler's work, his mathematical functions became statistical matrices for the development of probabilistic distributions, from which a family of continuous ones defined in the interval (0-1) emerged, composed of two positive parameters, denoted by α and β, which appear as exponents of the random variable *X* and control the shape of the distribution.

The Beta distribution has been applied to model the behaviour of variables, limited to finite-size intervals, in a great diversity of populations.

However, in biological and forestry applications, the interval for variable *X* lies between two finite values *a* and *b*, which implies that this distribution, as the mother of the Pellico-Behling distribution, will be fitted to all datasets in the form already mentioned in (10), to be compared with the proposed distribution by the authors and verify their real potentiality and robustness. The fitting of the beta distribution was performed using SAS (*on demand for academics*) and the Julia language.

Burr's distribution

The Burr distribution is a continuous probability distribution for a non-negative random variable, and is one of several different distributions, sometimes called the "generalized log-logistic distribution," which is most used to model household income but can also be applied to fit ecological and production variables in native and planted forests.

This distribution as presented in (Equation 12) will be fitted to all databases to establish a comparison with the Péllico-Behling distribution. Burr's distribution was fitted using SAS (*on demand for academics*) and the Julia language.

As previously mentioned, we applied Burr's distributions to compare their fitting with the Péllico-Behling distribution. In the case of the exponential type of distribution, we evaluated the application of the Pareto Type II (Lomax) distribution when the parameter *c=1*, but the results were not satisfactory as the estimated frequencies in the upper diameter classes diverged significantly from the observed frequencies, even though the KS test attests no significance. Therefore, we decided to fit the original Pareto distribution (Pareto 1897) to this dataset, and the results were deemed appropriate.

Silva's distribution

The function proposed by Eduardo Quadros da Silva, Silva et al. (2003) was chosen to be fitted to the height database of the species *Eucalyptus saligna* because of its special condition of strong asymmetry and kurtosis, which made it attractive for the flexibility resulting from the polynomial of n^{th} degree in the second segment inside the limits $l_1 \leq x \leq l_2$, as presented in (Equation 69).

$$\frac{1}{k}\begin{Bmatrix} c_1 x^d & if\ 0 < x < l_1 \\ a_1 x^n + a_2 x^{n-1} + \ldots + a_m & if\ l_1 \leq x \leq l_2 \\ \frac{c_2}{x^h} & if\ x > l_2 \end{Bmatrix} \quad (69)$$

$$0 \qquad otherwise$$

where n, d, and h are positive integers; a_1, a_2, a_3, ... a_m are the coefficients of the polynomial; c_1 and c_2 are coefficients of the complementary functions of Silva's truncated distribution; x is the selected variable; k is the value of the integral:

$$\int_0^\infty \left[x^d c_1 + \left(a_1 x^n + a_2 x^{n-1} + a_3 x^{n-2} + \ldots + a_m\right) + \frac{c_2}{x^h} \right] d_x = k$$

l_1 is the upper limit of the class in which the function $c_1 x^d$ is fitted; and l_2 is the upper bound of the data class for the polynomial fit.

This polynomial distribution is composed of three mathematical functions, as presented in (Equation 69), in which the first part consists of a positive increasing potential function, the second is a polynomial adjusted by the least squares method, and the third is a hyperbolic descending function, which has a straight-line $y = 0$ as an asymptote. The three segments must meet the requirements of a probability density function; that is, they must be continuous, with non-negative functional values and convergent in $(0 + \infty)$. Silva et al. (2003) emphasised that five steps are required to fit these functions.

Recently, an adaptation of the Silva's function was conducted, with refitting of the polynomial with the help of the Julia language. This adaptation was conducted aiming at a better refitting of the complementary functions to achieve smoothness on the resultant truncated function composed with the polynomial curve. The coefficient value in the function $g_1(x) = c_1 x^d$ was optimised to approximate the estimated values to the observed ones. The values of h in the function $g_3(x) = c_2/x^h$ were also optimised.

The cut points on the polynomial curve:

$$g_2(x) = x^n + a_2 x^{n-1} + a_3 x^{n-2} + \ldots + a_m$$

were chosen to smooth the junction of these with $g_1(x)$, and to approximate the estimated values to the real ones, with less squared error; therefore, the degrees of the polynomials ranged from the 2nd to 5th degree. The polynomial coefficients were obtained through simple linear regression, which made the estimates more accurate. After completing the fittings, with adaptation of Silva's probability function, this distribution improved in describing the height behaviour of the species *Eucalyptus saligna*. The fitting of de Silva's distribution was performed using the Julia language.

Fitting the Pellico-Behling distribution by cross validation

To evaluate the generalisation capacity of the Pellico-Behling distribution, fittings were performed using cross-validation. The DBH database for the black wattle species (Plantation 1) was used for this purpose. In this case, a sample dataset was considered, consisting of 674 values corresponding to data from a single stand. Two hundred fittings were conducted, with each fitting using 80% of the values to fit the distribution and the remaining 20% to evaluate the goodness of fit of the fitted distribution using the KS test at a 95% probability. The selection of values for fittings and goodness of fit evaluation was done randomly for each replication. The relative classes presented in Table 1 were created to obtain the frequency for fitting the data and testing their goodness of fit. The 200 replications allowed us to evaluate the behaviour of the Pellico-Behling distribution coefficients through their frequency, as well as the probability and cumulative density functions, in addition to the performance of the goodness of fit test.

The fittings via cross-validation were performed using the OPTMBELR function, developed in the Julia language. Details of the OPTMBELR function are provided in Appendix 3. This function was designed to validate the robustness and stability of the proposed model fitting through a repeated cross-validation (resampling) procedure applied to diameter data of *Acacia mearnsii*. Based on the original dataset, the function performs R repetitions of a random partition into two subsets: one for model fitting (80%) and another for testing (20%). In each repetition, the data are normalised and organised into 12 frequency classes using fixed interval limits. Model fitting is carried out using constrained nonlinear optimisation via the Ipopt solver, following the mathematical formulation implemented in the OPTMBEL function, which minimises the sum of squared residuals with an added regularisation term.

In each iteration, the estimated parameters were stored, fitted curves were generated, and the observed and estimated cumulative frequencies were compared using the Kolmogorov-Smirnov (KS) test. The test statistic was calculated and compared to the critical value to evaluate the goodness-of-fit. At the end of the process, the function returned the proportion of iterations in which the calculated KS statistic was lower than the critical value, indicating the percentage of statistically acceptable fits. Additionally, various graphical outputs were saved, including histograms of the estimated parameters, fitted frequency and cumulative curves, and the distribution of KS values, providing a comprehensive overview of model quality and parameter stability.

The OPTMBELR function required four input arguments:

(i) a DataFrame X containing at least one column named DAP, with the diameter values to be analysed;

(ii) an integer R, indicating the number of cross-validation repetitions;

(iii) a vector p0 containing five initial values for the model parameters; and

(iv) a string out, specifying the path to the directory where the graphical output files will be saved. The function is included as an annex to this document to ensure reproducibility and to support future applications.

The OPTMBEL and OPTMBELR functions required the following packages for their implementation: CSV is the package for importing, creating, and manipulating files in CSV format, DataFrames is the package for creating and manipulating data tables, Plots is the package for generating graphics, ADNLPModels is the package that provides implementation of automatic differentiation-based models and ForwardDiff is the package that implements methods to obtain derivatives, gradients, Jacobians, Hessians, and higher-order derivatives of native Julia functions using automatic differentiation in direct mode, NLPModels is the package that provides general guidelines for representing nonlinear programming problems in Julia and a standardised API for evaluating functions and their derivatives, JSOSolvers is the package that provides optimisation solvers for unconstrained optimisation, LinearAlgebra is the package for computing matrices, Logging is the package that provides basic features for logging output in Julia, Printf is the Package that provides basic features for formatted printing in Julia, Percival is the package that provides implementation of the augmented Lagrangian solver and NLPModelsIpopt is the package that provides a thin IPOPT wrapper for NLPModels.

To compare the Péllico-Behling, Burr, and Beta distributions, one iteration was randomly selected from the 200 data partitions to evaluate the Kolmogorov-Smirnov (KS) test on both the fitting (80%) and testing (20%) datasets. The fitting dataset was used to fit the three distributions, and the resulting models were then evaluated on the testing dataset using the KS goodness-of-fit test. During the fitting process, the diameter variable was standardised to the [0, 1] range, as previously described.

Evaluation of the Péllico-Behling distribution on independent datasets

A new sample consisting of 271 diameter at breast height (DBH) measurements of *Acacia mearnsii* (black wattle) was collected in May 2025 from three circular plots of 400 m^2 each, randomly installed in a 7.3-year-old stand located in the municipality of Piratini, in the state of Rio Grande do Sul, Brazil. It is important to note that this dataset was not used in the initial fitting of the distributions described in Table 1 and was reserved exclusively for the external validation of the fitted probability functions. The aim of this approach was to evaluate the predictive performance of the Péllico-Behling distribution when applied to independent data obtained under conditions similar to those used for model fitting. The fitted distributions were assessed on the testing dataset using the Kolmogorov-Smirnov (KS) goodness-of-fit test. During the fitting process, the diameter variable was standardised to the [0, 1] range, as previously described.

Results

The Pellico-Behling Probability Distribution (5P) was applied to fit the datasets for *Cariniana legalis* (Figure 1), *Acacia mearnsii* (Figure 2), and *Eucalyptus saligna* (Figure 3), as presented in Equation 25. The average mortality of 124 species was obtained using case 2, that is, with (4P). The fitting of the distribution to the data of average mortality of 124 species using case 2 with four parameters (Equation 59), is presented in (Figure 4). In all these cases the cumulative function was calculated using the additive rectangular areas of small classes of 1/1,000 size, approximative to the area obtained by the integration of function (Equation 28). The results of the estimates for the populations are presented in Table 2. The results of the fitted distributions, including their statistics and the goodness-of-fit evaluation are presented in Table 3. The results of the estimates for the populations obtained by Burr distribution are summarised in Table 4 and Figure 5. The results of the estimates for the populations using the Beta distribution are summarised in Table 5 and Figure 6. The results of

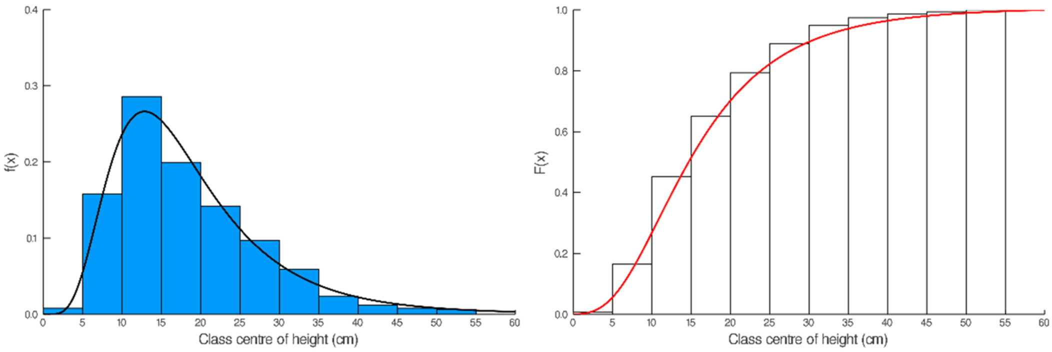

Figure 1. Péllico-Behling Probability Distribution fitted to heights of the species *Cariniana legalis.*

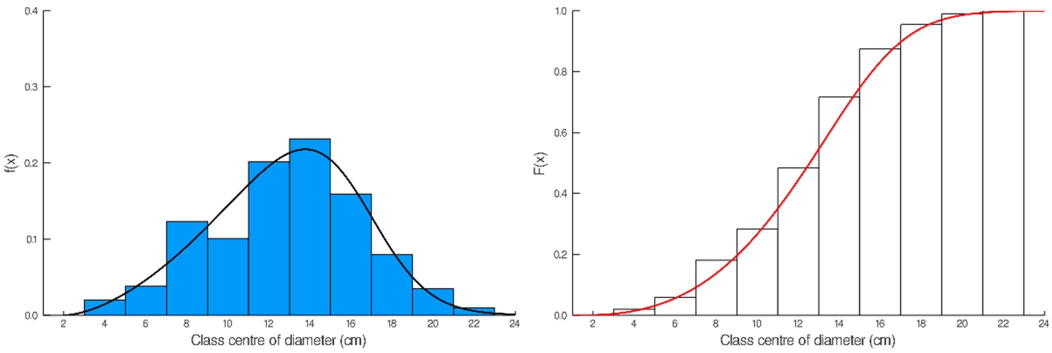

Figure 2. Péllico-Behling Probability Distribution fitted to *DBH* of *Acacia mearnsii*.

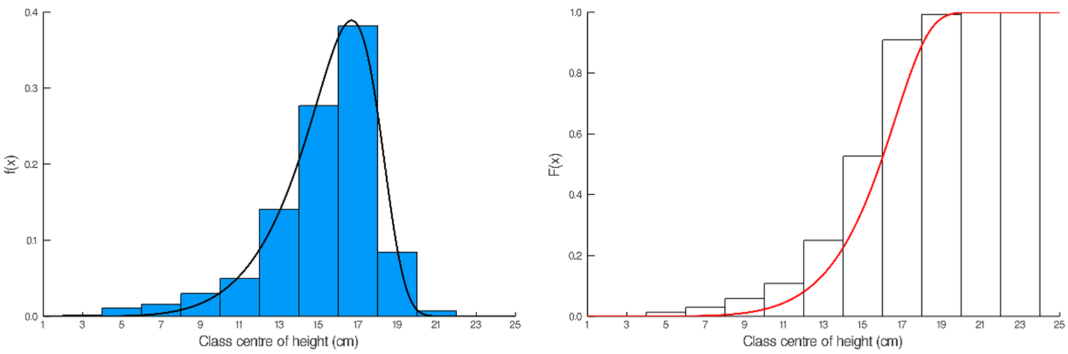

Figure 3. Péllico-Behling Probability Distribution fitted to heights of the species *Eucalyptus saligna.*

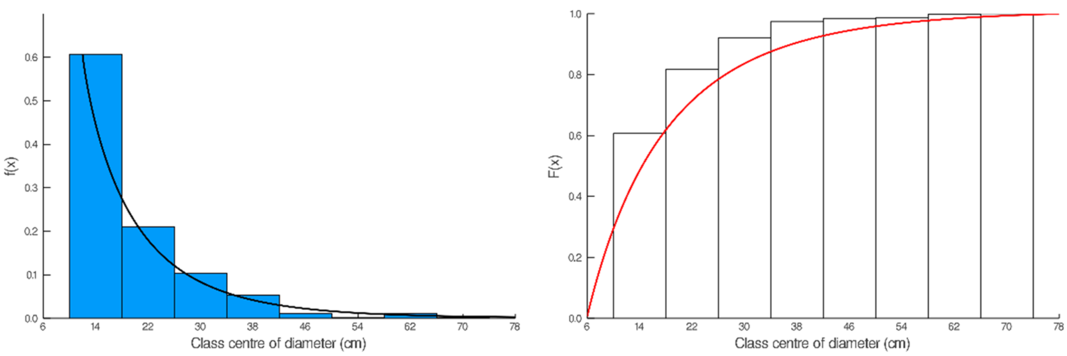

Figure 4. Péllico-Behling Probability Distribution fitted to the average mortality of 124 tropical species, evaluated from 1996 to 2001, in a fragment of the semideciduous seasonal forest in Cassia, MG, Brazil.

Table 2. Summary of the results for fitting the Péllico-Behling distribution to different sets of data.

Application cases	Relativised (*x*)	*f(x)*	*F(x)* k = 1000 classes
Cariniana legalis	$x = \frac{x_i}{60}$	$f(x) = \frac{x^{5.85445}}{\left(0.29542 + 1.33537\,x^{1.16224}\right)^{11.70896}}$	$F(x) = \frac{\sum_{i=1}^{x} f(x)}{\sum_{i=1}^{k} f(x)}$
Acacia mearnsii	$x = \frac{x_i - 2}{24 - 2}$	$f(x) = \frac{x^{1.65945}}{\left(1.07521 + 6.75713\,x^{7.04916}\right)^{3.31892}}$	$F(x) = \frac{\sum_{i=1}^{x} f(x)}{\sum_{i=1}^{k} f(x)}$
Eucalyptus saligna	$x = \frac{x_i - 1}{25 - 1}$	$f(x) = \frac{x^{5.19183}}{\left(0.89138 + 12.58504\,x^{14.81904}\right)^{14.13236}}$	$F(x) = \frac{\sum_{i=1}^{x} f(x)}{\sum_{i=1}^{k} f(x)}$
Mortality	$x = \frac{x_i - 6}{78 - 6}$	$f(x) = \frac{1}{\left(1.0308 + 0.3653\,x^{0.8710}\right)^{16.4772}}$	$F(x) = \frac{\sum_{i=1}^{x} f(x)}{\sum_{i=1}^{k} f(x)}$

the estimates for the populations using the Silva's distribution are summarised in Table 6 and Figure 7.

Results of fitting the Pellico-Behling distribution by cross validation

The fittings, considering cross-validation for 200 replications are presented in Figure 8, applied to the distribution of diameters of *Acacia mearnsii* De Wild (n = 674). The fitted functions proved suitable to highlight the distinct variabilities of the observed distributions, showing robustness in representing the data, as well as stability across the 200 simulations performed.

The proportion of times where the calculated Kolmogorov-Smirnov (KS) test value was equal to or less than the tabulated KS value reached 97.50% across 200 replications of the fittings, as illustrated in Figure 9, showing the goodness of fit of the tested

Table 3. The Péllico-Behling distribution statistics.

Statistics	*Acacia mearnsii*	*Eucalyptus saligna*	*Cariniana legalis*	Mortality
	DBH	**Heights**	**Heights**	**124 trop. spp.**
Mean	0.4939	0.6047	0.318	0.1942
Variance	0.0269	0.0089	0.033	0.0517
Standard deviation	0.1642	0.0946	0.1827	0.2275
Mode	0.5349	0.6528	0.2144	0
Inflexion points	[0.3644, 0.6844]	[0.5808, 0.7230]	[0.1111, 0.3174]	---
Asymmetry	-0.1395	-0.8585	2.2373	2.9471
Kurtosis	-0.2521	0.803	12.0516	12.44
KS	0.0317^{ns}	0.0387^{ns}	0.0161^{ns}	0.0120^{ns}
Tab. (5%)	0.0366	0.0648	0.0613	0.0590

ns: not significant.

Table 4. Summary of the results for fitting the Burr distribution to *Cariniana legalis, Acacia mearnsii, Eucalyptus saligna,* and Pareto distribution for mortality of 124 tropical species.

Application cases	***f(x)***	**KS test**
Cariniana legalis	$f_{(x)} = \dfrac{1.75895\ \ 2.91640\left(\dfrac{x}{21.24594}\right)^{2.91640}}{x\left[1+\left(\dfrac{x}{21.24594}\right)^{2.91640}\right]^{(1.75895+1)}}$	0.0318^{ns}
Acacia mearnsii	$f_{(x)} = \dfrac{311.64591\ \ 3.91762\left(\dfrac{x}{61.59325}\right)^{3.91762}}{x\left[1+\left(\dfrac{x}{61.59325}\right)^{3.91762}\right]^{(311.64591+1)}}$	0.0323^{ns}
Eucalyptus saligna	$f_{(x)} = \dfrac{9.10632\ \ 7.12003\left(\dfrac{x}{22.50700}\right)^{7.12003}}{x\left[1+\left(\dfrac{x}{22.50700}\right)^{7.12003}\right]^{(9.10632+1)}}$	0.0387^{ns}
Mortality	$f_{(x)} = \dfrac{1.4651\ \ 46.6328^{1.4651}}{x^{(1.4651+1)}}$	0.0369^{ns}

KS – Kolmogorov-Smirnov test

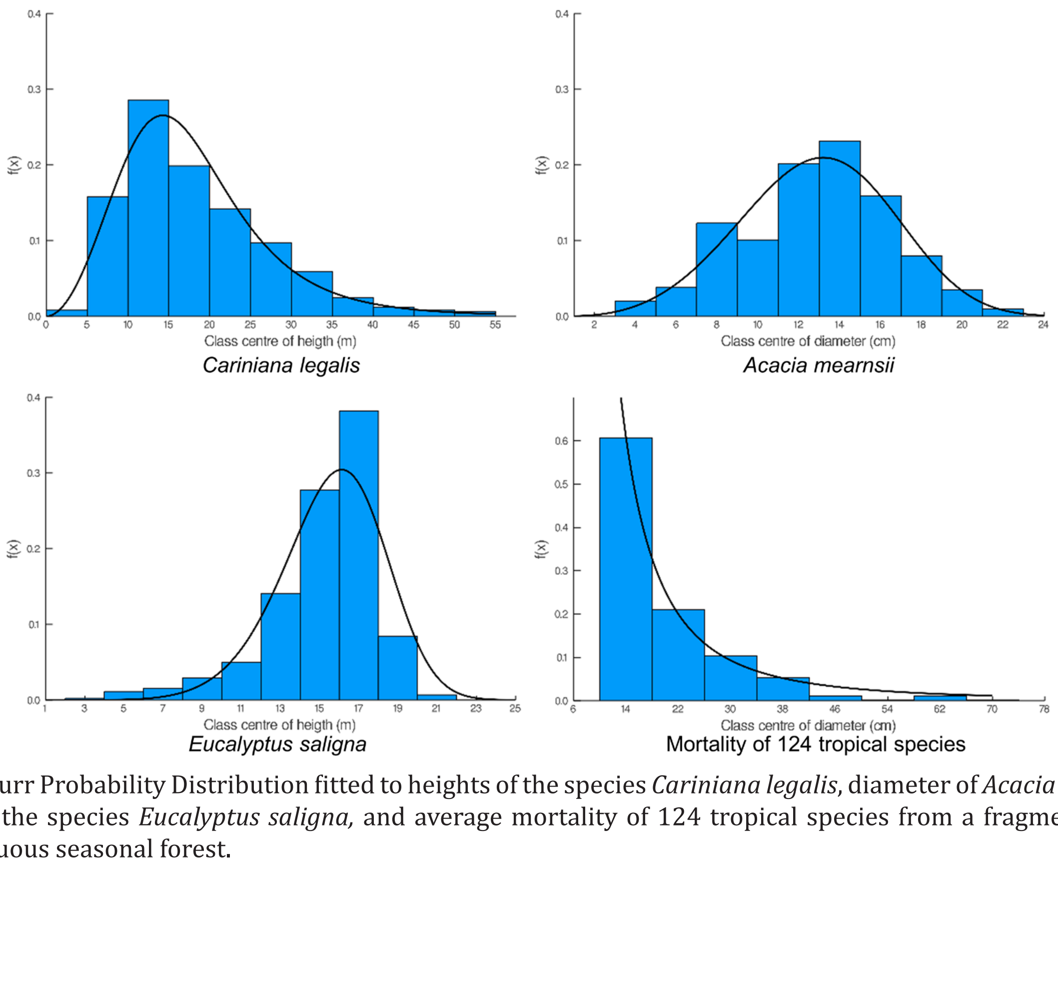

Figure 5. Burr Probability Distribution fitted to heights of the species *Cariniana legalis*, diameter of *Acacia mearnsii*, heights of the species *Eucalyptus saligna,* and average mortality of 124 tropical species from a fragment of the semideciduous seasonal forest.

Table 5. Summary of the results for fitting the beta distribution to different sets of data.

Application cases	f(x)	KS
Cariniana legalis	$f_{(x)} = \frac{1}{B(2.0266\,,4.0551)} \frac{(x-2.5)^{2.0266-1}\,(52.5-x)^{4.0551-1}}{(52.5-2.5)^{2.0266+4.0551-1}}$	0.1021^{ns}
Acacia mearnsii	$f_{(x)} = \frac{1}{B(6.8948\,,5.7513)} \frac{(x-2)^{6.8948-1}\,(24-x)^{5.7513-1}}{(24-2)^{6.8948+5.7513-1}}$	0.1348*
Eucalyptus saligna	$f_{(x)} = \frac{1}{B(8.9922\,,5.8327)} \frac{(x-3)^{8.9622-1}\,(23-x)^{5.8387-1}}{(23-3)^{8.9622+5.8387-1}}$	0.0727*
Mortality	$f_{(x)} = \frac{1}{B(0.0713\,,2.3436)} \frac{(x-14)^{0.0713-1}\,(78-x)^{2.3436-1}}{(78-14)^{0.0713+2.3436-1}}$	0.4572*

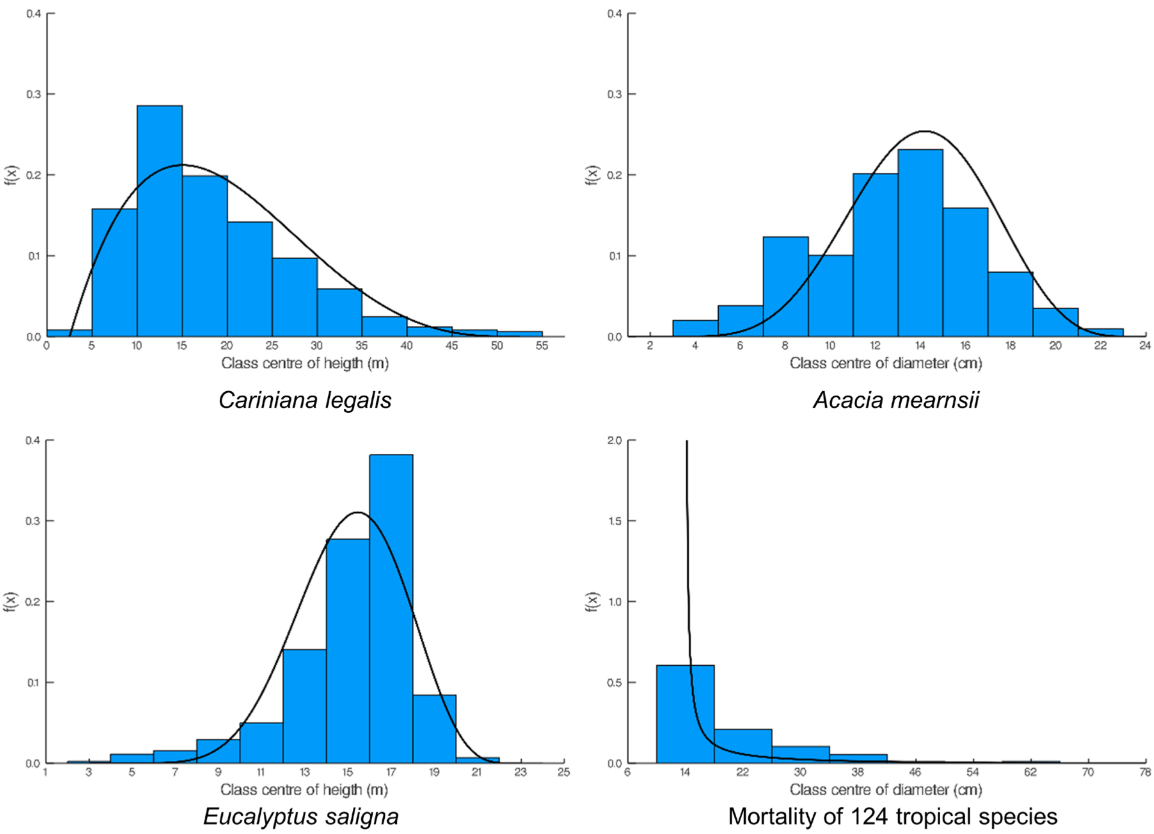

Figure 6. Beta Probability Distribution fitted to heights of the species *Cariniana legalis*, diameter of *Acacia mearnsii*, heights of the species *Eucalyptus saligna,* and average mortality of 124 tropical species in a fragment of the semideciduous seasonal forest.

Table 6. Summary of the results for fitting the Silva's distribution to heights of the species *Eucalyptus saligna.*

Component	f(x)	KS
1	$f_{(x)} = 0.07447 \ x^{2.37660}$	
2	$f_{(x)} = -0.3359\, x^4 + 18.104 \ x^3 - 361.58 \ x^2 + 3196.9\, x - 10571$	0.012^{ns}
3	$f_{(x)} = \frac{96.81705e19}{x^{14.30643015}}$	

distribution. This high percentage of fit highlights the consistency and reliability of the distribution to assess the characteristics of the variable used.

Approximately, the values of coefficient *a* ranged from 0.7 to 1.3, while coefficient *b* ranged from slightly above 1 to 1.3. Coefficient *c* ranged between 2 and 7, coefficient *d* between 3 and 5, and coefficient *e* between 2 and 6, Figure 10. No outlier values were observed for the coefficients across the 200 simulations effectuated, indicating that they are stable and reliable to be used.

The fitting through the optimisation of the proposed *pdf*, which minimises the sum of squared residuals between the observed and estimated frequencies, proved to be an excellent fitting method. The optimisation procedure showed its effectiveness in assessing the characteristics of the data distribution, ensuring that the adjusted coefficients fell within an acceptable range of values. By minimising residuals, the method provided a close approximation between observed and estimated frequencies, highlighting its reliability and applicability for fitting the proposed *pdf*, with consistent and accurate results, reinforcing the overall robustness of the statistical analysis. Consequently, the performance of the procedures using the OPTMBEL and OPTMBELR functions enables to assert that the obtained results were reliable and statistically robust.

The fitting results of the Péllico-Behling distribution, applied to 80% of the *Acacia mearnsii* De Wild DBH data and tested on the remaining 20%, representing one instance among the 200 cross-

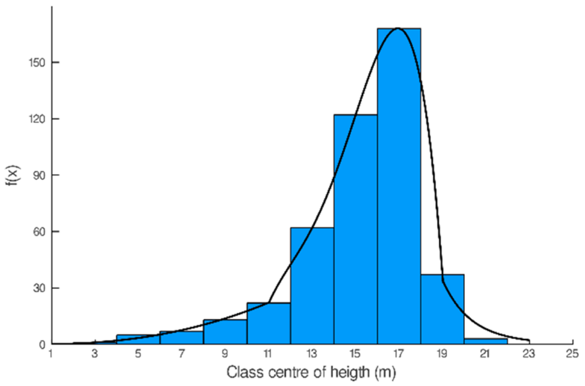

Figure 7. Silva's Probability Distribution fitted to heights of the species *Eucalyptus saligna.*

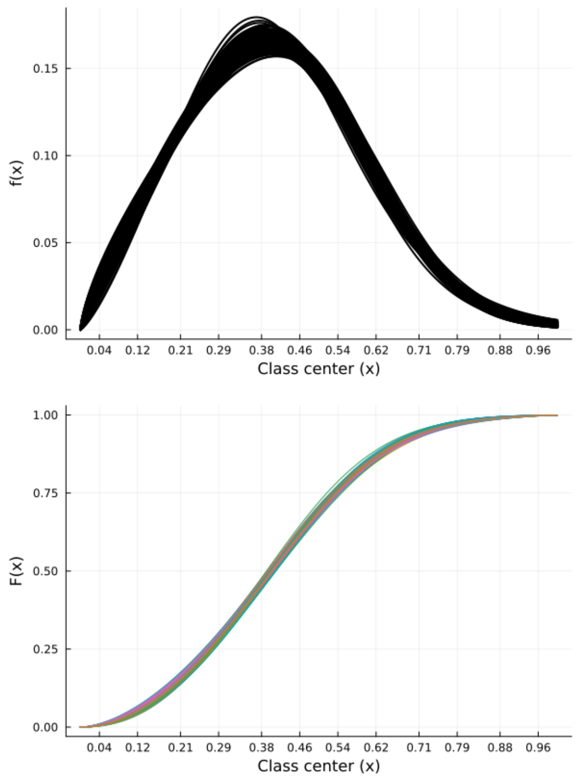

Figure 8. Fitting of the Péllico-Behling Probability Distributions [f(x), F(x)] through cross-validation, repeated 200 times, for the database *DBH* of *Acacia mearnsii*.

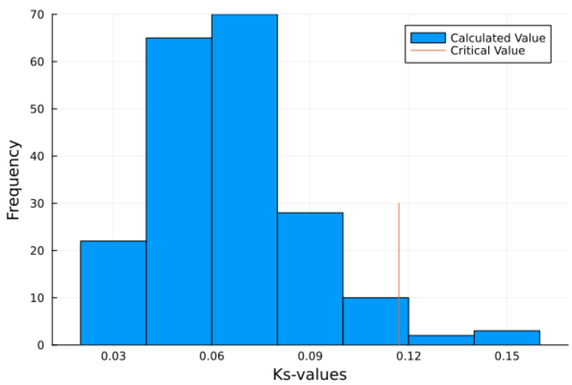

Figure 9. KS tests applied to fittings with the Péllico-Behling Probability Distribution using the database for goodness of fit assessment, repeated 200 times, to *DBH* of *Acacia mearnsii*.

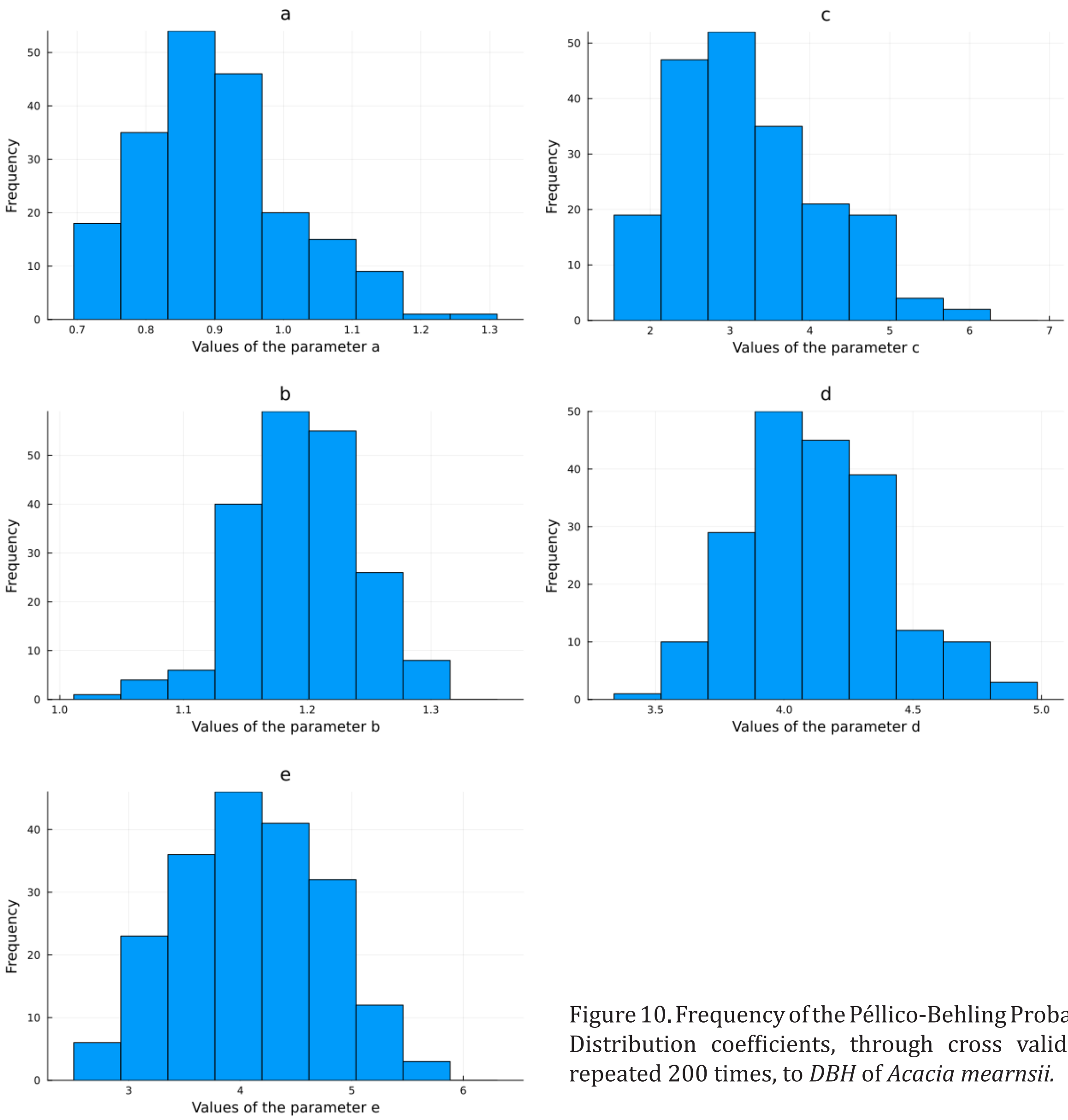

Figure 10. Frequency of the Péllico-Behling Probability Distribution coefficients, through cross validation, repeated 200 times, to *DBH* of *Acacia mearnsii.*

validation repetitions, are presented in Figure 11 and Table 7. The corresponding results for the Beta and Burr distributions are shown in Figure 12 and also summarised in Table 7.

Evaluation of the Péllico-Behling Distribution Using Independent Datasets

The fitted Péllico-Behling (Table 2), Burr (Table 4), and Beta (Table 5) distributions were evaluated for their ability to predict the diameter at breast height (DBH) distribution in a new *Acacia mearnsii* stand dataset that was not used during the model fitting stage. The results showed that the Péllico-Behling distribution demonstrated greater flexibility in capturing the observed diameter structure, being the only one among those tested that did not show statistical significance in the Kolmogorov-Smirnov (KS) test—indicating a satisfactory fit to the observed data (Figure 13). This underscores both the potential of the Péllico-Behling distribution for modeling DBH distributions in *Acacia mearnsii* stands and the flexibility of the function.

Discussion

Probability distributions are very important models to describe the behaviour of biotic variables in forest ecosystems, as well as their evolution in a time horizon that represents a climax life cycle. The modeling of biological and ecological phenomena through probabilistic distributions makes it possible to treat them quantitatively and evaluate them statistically. The use of distributions for such circumstances has been widely used by renowned researchers dealing with biological, forestry and ecological experiments, such as, Bailey (1980), Bailey & Dell (1973), Bowling

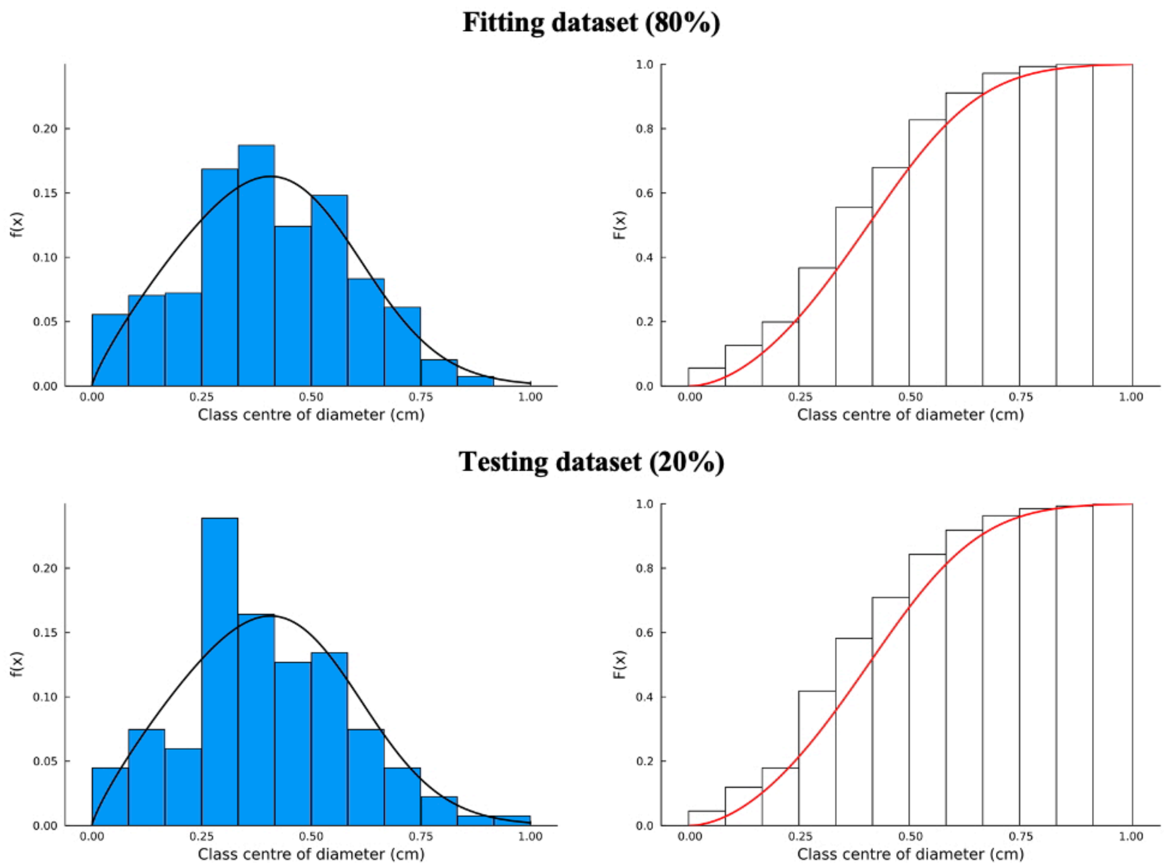

Figure 11. Péllico-Behling distribution fitted to 80% of the DBH data from *Acacia mearnsii* and tested on the remaining 20%, representing one instance from the 200 repetitions used in the cross-validation.

et al. (1989), Borders et al. (1990), Xu et al. (1992), Cao (1997), Fonseca et al. (2009), Nanosa et al. (2000), Zhang et al. (2001), Liu et al. (2002), Zhang et al. (2003), Ivkovi & Rozenberg (2004), Liu et al. (2004), Qin et al. (2007), Breidenbach et al. (2008), Machado et al. (2008), Binotti et al. (2012), Sandoval et al. (2012), Rupšys & Petrauskas (2017), Miranda et al. (2018), Chen et al. (20019), Schmidt et al. (2019), Duchateau et al. (2020), Piva et al. (2020), Schmidt et al. (2020), Ciceu et al. (2021), Cao (2022), Guo et al. (2022), Goodwin (2022), Waldy et al. (2022).

In the last 30 years, we have encountered numerous datasets whose fitted distributions presented asymmetry, accentuated kurtosis, and other peculiarities that have led to non-goodness-of-fit when using the most applied probability distributions to forest data, such as Gamma, Beta, Weibull, and Burr's system. Our group, formed by forest researchers and two mathematicians interested in collaborating in the search for alternatives to improve such fittings, initially proposed the application of a polynomial of degree n as a mathematical model for improving the goodness-of-fit condition of a probability distribution.

A polynomial distribution of degree n was developed, which satisfied the previously appointed constraints very well. This polynomial distribution consists of three mathematical functions and is a "truncated pdf," which shows great flexibility when fit to forest data (Silva et al. 2003). In this model, the first part consists of a positive potential increasing function, the second part is a polynomial fit using the least-squares method, and the third is a hyperbolic function with $y = 0$ as an asymptote. The three functions must meet the assumptions of a *pdf*, that is, they must be continuous with non-negative and convergent values in the interval $(0 + \infty)$.

As mentioned in Section 4 of the manuscript, when a probability distribution has more than three parameters, the possibility of combinatorial solutions grows exponentially; that is, their estimates obtained with different software programs did not generate appropriate values for variance, skewness, kurtosis, and inflection points in many fitted cases. Only after the development of an optimisation function by the mathematician Dr. Abel Soares Siqueira, called OPTIMBEL, this was possible to obtain a single realistic solution for the five parameters of the new proposed distribution.

Additionally, the scale transformation of the variable X in the interval between zero and one allowed us to obtain consistent parameters for all sampled datasets to which the proposed distribution

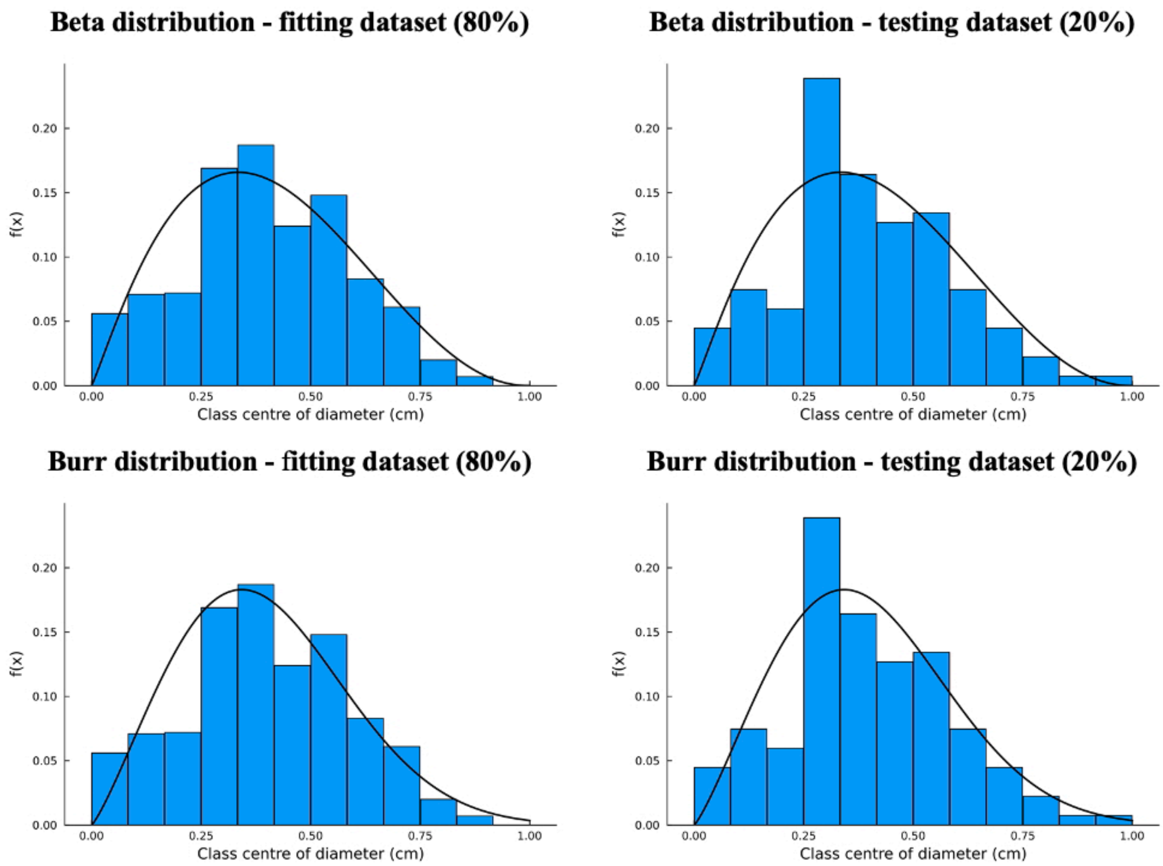

Figure 12. Beta and Burr distributions fitted to 80% of the DBH data from *Acacia mearnsii* and tested on the remaining 20%, representing one instance from the 200 repetitions used in the cross-validation.

Table 7. Summary of the results from fitting the Péllico-Behling, Beta and Burr distributions to 80% of the DBH data (n = 674) from *Acacia mearnsii*, and testing on the remaining 20%. This represents a single instance from the 200 repetitions used in the cross-validation.

Distribution	***f(x)***	**KS - fitting dataset (80%)**	**KS - testing dataset (20%)**
Péllico-Behling	$f_{(x)} = \frac{x^{0.81713}}{(1.22506 + 2.81626\, x^{4.40239})^{4.39206}}$	0.0322^{ns}	0.0595^{ns}
Beta	$f_{(x)} = \frac{1}{B(2.1157\,, 3.2194)} \frac{(x-0)^{2.1157-1}\,(1-x)^{3.2194-1}}{(24-2)^{2.1157+3.2194-1}}$	0.0832*	0.1027^{ns}
Burr	$f_{(x)} = \frac{1825\;\; 2.2033 \left(\frac{x}{13.6236}\right)^{2.2033}}{x\left[1+\left(\frac{x}{13.6236}\right)^{2.2033}\right]^{(1825+1)}}$	0.0965*	0.0804^{ns}

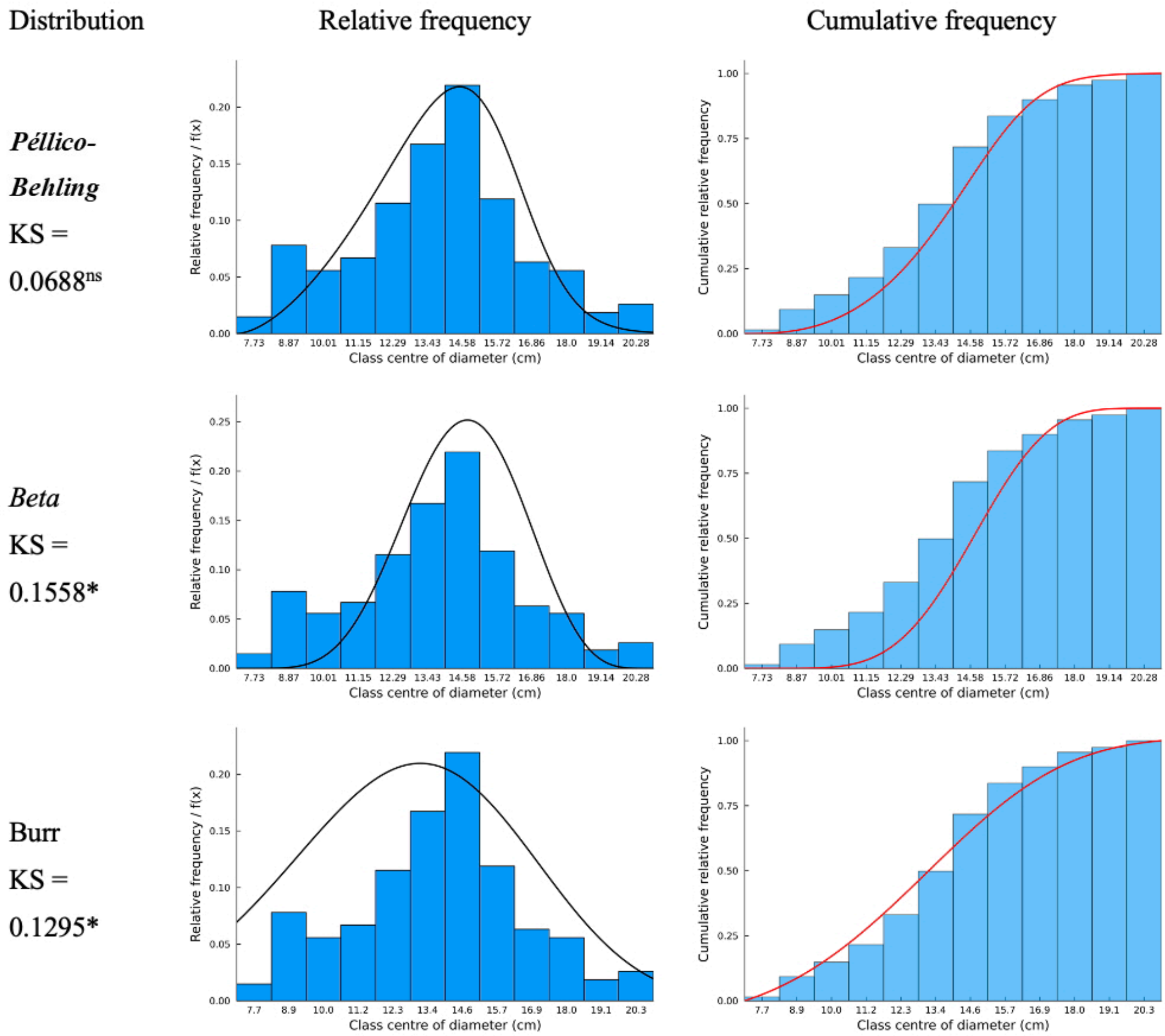

Figure 13. Fitted Péllico-Behling, Beta and Burr distributions tested on a new dataset from a 7.3-year-old *Acacia mearnsii* stand.

was fitted. Evaluating the calculations of the statistical estimates obtained from these fittings allowed us to suggest new inclusion of restrictions for the adjustments of the distribution, as presented in the methodology. It is important to highlight that the transformed scales open the possibility of using the estimated parameters as seeds for fitting new datasets.

The Péllico-Behling distribution after fitting to the sampled datasets was compared with the fitted Beta and Burr (4P) distributions, except for Silva's polynomial distribution that was fitted to the heights of the species *Eucalyptus saligna* and the Pareto distribution to mortality of 124 tropical species from a fragment of a semideciduous seasonal forest, to evaluate and verify its potential and robustness. The first dataset is the height of *Cariniana legalis* collected in a mixed tropical forest fragment in Cassia, MG, Brazil; the second dataset is the *DBH* of *Acacia mearnsii* collected in the plantation areas of Cristal and Piratini in the state of Rio Grande do Sul, Brazil; the third dataset is the heights of *Eucalyptus saligna* from an experiment implemented in the Ibiti Forest Park owned by the company Ripasa S.A. Celulose and Paper, located in the municipality of Itarare, SP; and the fourth dataset is the mortality of tropical species (124) collected in a mixed tropical forest fragment in Cassia, MG, Brazil. The proposed distribution revealed reliable goodness-of-fit in the fitted cases. The goodness of fit evaluated with the KS test in the other distributions revealed in almost all cases to be non-significant, but not better than the results obtained for the proposed distribution, except for the fitting of Silva's distribution applied to *the Eucalyptus saligna* dataset, because its *pdf*, as mentioned before, is a truncated distribution with exceptional flexibility due to the inclusion of a fourth-degree polynomial segment in that model, capable of assimilating the extreme cases of skewness.

Through the cross-validation process applied to the DBH data of *Acacia mearnsii*, it was observed that the optimisation of the proposed PDF proved to be an effective and highly accurate fitting method. By ensuring that the adjusted coefficients remained within an acceptable range, the procedure effectively captured the characteristics of the data distribution. This approach resulted in a close alignment between observed and estimated frequencies, highlighting the capability and robustness of the statistical analysis. Consequently, the use of the OPTMBEL and OPTMBELR functions confirmed the reliability and statistical robustness of the obtained results.

The results from one of the 200 cross-validation iterations highlighted the stability of the fitting performance of the Péllico-Behling distribution. The KS test results were not significant for either the fitting or testing datasets, indicating a good fit to the data. In contrast, for the Burr and Beta distributions, the KS test was non-significant only for the testing dataset. This indicates that while the Péllico-Behling distribution provided a consistent fit across both datasets, it also demonstrated superior performance in the goodness-of-fit assessment. These findings were further supported when the previously fitted distributions (shown in Tables 2, 3, and 4) were applied to an independent dataset collected from a 7.3-year-old *Acacia mearnsii* stand not included in the original model fitting. Once again, the Péllico-Behling distribution was the only one that the Kolmogorov-Smirnov test was not significant, reinforcing its robustness, flexibility, and predictive capability for modeling diameter distributions in independent stands of this species.

The proven flexibility of the proposed distribution in different cases of asymmetries and kurtosis in forest datasets opens the possibility of using it for monitoring one variable or a set of variables in long-term experiments to assimilate their variations in shape over time in the same forest population.

Conclusions

In this research, a new model called the Péllico-Behling distribution was developed with five parameters, capable of attaining maximum flexibility of the resulting function.

The probability density function was generated without using the *derivative method*, but instead a new procedure called the *aggregative method* to avoid the occurrence of parental parameters in the numerator of the ratio that makes up the resulting function.

After fitting this new distribution with various software to different datasets, the results were not equal, and this divergence is due to an enormous number of possible combinations of the five parameters. Even though goodness of fit was obtained in all adjustments, the statistical estimates were not correct in most cases, requiring an alternative solution for this problem.

An optimisation function, OPTMBEL, was developed to provide an appropriate solution for the parameters and correct the statistical estimates of the distribution.

The scale transformation of variable X in the interval between zero and one was favored to obtain consistent parameters for all sampled datasets, to which the proposed distribution was fitted.

The new distribution is quite flexible and presents better reliability when compared with beta, Burr (3P), Silva, and Pareto distributions. The plots of the *pdf* and *cdf* clearly show the flexibility of the proposed distribution.

The proposed distribution fitted well to symmetric and asymmetric data in unimodal cases and to exponential occurrences of the variable X, which qualifies it to model and monitor its behaviour in different realities occurring in biopopulations.

List of abbreviations

cdf: continuous distribution function
pdf: probability density function
MLE: maximum likelihood estimators

Authors' contributions

SPN: developed the derivation of the new distribution, with its parameters and statistical estimators. He also collected part of the data to illustrate its application and evaluated the quality of the results obtained in all tested populations; AB: dedicated his time adjusting the distributions to the datasets, as well as to illustrate the Pellico-Behling distribution and compare it with other selected ones to evaluate its performance, especially in skewed data behaviour.

Competing interests

The author(s) declare that they have no competing interests.

Acknowledgements

We acknowledge the assistance of the mathematician Dr Abel Soares Siqueira, who searched for an alternative solution to obtain reliable values for the parameters of the Péllico-Behling distribution using an optimisation procedure. The resulting function was denominated OPTMBEL and was developed using the Julia language. We also thank PhD student Claiton Nardini for his collaboration in the final formatting of the manuscript for submission to the *New Zealand Journal of Forestry Science.* We acknowledge CNPq for support, Project 309824/2023-0.

References

Bailey TL, Dell TD 1973. Quantifying diameter distributions with the Weibull function. Forest Science 19(2), 97-104. https://doi.org/10.1093/forestscience/19.2.97

Bailey RL 1980. Individual tree growth derived from diameter distribution models. Forest Science 26(4), 626-632. https://doi.org/10.1093/forestscience/26.4.626

Bezanson J, Karpinski S, Shah VB, Edelman A 2025. Why we created Julia. https://julialang.org/blog/2012/02/why-we-created-julia/

Binet JFM (1839. Memoire sur les intégrales définies eulériennes et sur leur application à la théorie des suites. France: Hachette Livre Bnf.

Binoti DHB, Binoti MLMS, Leite HG, Oliveira J 2012. Probability density functions for description of diameter distribution in thinned stands of *Tectona grandis*. Cerne 18(2), 185-196. https://doi.org/10.1590/S0104-77602012000200002

Borders BE, Patterson WD 1990. Projecting stand tables: a comparison of the Weibull diameter distribution method a percentile-based projection method and a basal area growth projection method. Forest Science 36(2), 413-424. https://doi.org/10.1093/forestscience/36.2.413

Bowling EH, Burkhart HE, Burk TE, Beck DE 1989. A stand-level multispecies growth model for Appalachian hardwoods. Canadian Journal of Forest Research 19(4), 405-412. https://doi.org/10.1139/x89-064

Breidenbach J, Gläser C, Schmidt C 2008. Estimation of diameter distributions by means of airborne laser scanner data. Canadian Journal of Forest Research 38(6), 1611-1620. https://doi.org/10.1139/x07-237

Burr IW 1942. Cumulative frequency functions. The Annals of Mathematical Statistics 13(2), 215-232. https://doi.org/10.1214/aoms/1177731607

Burr IW 1968. On a general system of distributions III. The simple range. Journal of the American Statistical Association 63(322), 636-643. https://doi.org/10.1080/01621459.1968.11009282

Burr IW, Cislak PJ 1968. On a general system of distributions I. Its curve shaped characteristics; II. The sample median. Journal of the American Statistical Association 63(322), 627-635. https://doi.org/10.1080/01621459.1968.11009281

Cao QV 1997. A method to distribute mortality in diameter distribution models. Forest Science 43(3), 435-442. https://doi.org/10.1093/forestscience/43.3.435

Cao QV 2022. Predicting future diameter distributions given current stand attributes. Canadian Journal of Forest Research 52(4), 561-567. https://doi.org/10.1139/cjfr-2021-0216

Champernowne DG 1952. The graduation of income distributions. Econometrica 20(4), 591-615. https://doi.org/10.2307/1907644

Chen Y, Wu B, Min Z 2019. Stand diameter distribution modeling and prediction based on maximum entropy principle. Forests 10(10), 859. https://doi.org/10.3390/f10100859

Ciceu A, Pitar D, Badea O 2021. Modeling the diameter distribution of mixed uneven-aged stands in the Southwestern Carpathians in Romania. Forests 12(7), 958. https://doi.org/10.3390/f12070958

Cramér H 1951. Mathematical methods of statistics. Princeton: Princeton University Press.Dagum CA 1977. New model of personal income distribution: Specification and estimation. Économie Appliquée 30 413-437. https://doi.org/10.3406/ecoap.1977.4213

Duchateau E, Schneider R, Tremblay S, Dupont Leduc L 2020. Density and diameter distributions of saplings in naturally regenerated and planted coniferous stands in Québec after various approaches of commercial thinning. Annals of Forest Science 77(2), 1-18. https://doi.org/10.1007/s13595-020-0929-5

Ferose N, Aslam M 2013. Maximum likelihood estimation-MLE of Burr Type V distribution under left censored samples. WSEAS Transactions on Mathematics 12(6), 657-669.

Fonseca TF, Marques CP, Parresol BR 2009. Describing maritime pine diameter distributions with Johnson's SB distribution using a new all parameter recovery approach. Forest Science 55(4), 367-373. https://doi.org/10.1093/forestscience/55.4.367

Fisk PR 1961. The graduation of income distributions. Econometrica 29(2), 171-185. https://doi.org/10.2307/1909287

Goodwin AN 2022. Constraint shock in diameter distribution models: the cost of ignoring negative bounds. Forest Science 68(3), 246-258. https://doi.org/10.1093/forsci/fxac014

Guo H, Lei X, You L, Zeng W, Lang P, Lei Y 2022. Climate sensitive diameter distribution models of larch plantations in north and northeast China. Forest Ecology and Management 506: 119947. https://doi.org/10.1016/j.foreco.2021.119947

Hatke MA 1949. A certain cumulative probability function. The Annals of Mathematical Statistics 20(3), 461-463. https://doi.org/10.1214/aoms/1177730002

Ivkovi M, Rozenberg P 2004. A method for describing and modelling of within ring wood density distribution in clones of three coniferous species. Annals of Forest Science 61(8), 759-769. https://doi.org/10.1051/forest:2004072

Johnson NL, Kotz S 1970. Distributions in Statistics: Continuous Univariate Distributions - 1. New York: John Wiley & Sons.

Julia Language 2025. The Julia Programming Language. https://julialang.org/

Kolmogorov A 1933. Sulla determinazione empirica di una legge di distribuzione. Giornale dell'Istituto Italiano degli Attuari 4, 83-91.

Loetsch F, Zöhrer F, Haller KE 1973. Forest Inventory. Munich: BLV Verlagsgesellschaft.

Liu C, Zhang L, Davis CJ, Solomon DS, Gove JH 2002. A finite mixture model for characterizing the diameter distributions of mixed species forest stands. Forest Science 48(4), 653-661. https://doi.org/10.1093/forestscience/48.4.653

Liu C, Zhang SY, Lei Y, Newton PF, Zhang L 2004. Evaluation of three methods for predicting diameter distributions of black spruce (*Picea mariana*), plantations in central Canada. Canadian Journal of Forest Research 34(12), 2424-2432. https://doi.org/10.1139/x04-117

Machado SA, Bartoszeck ACPS, ueiredo Filho A, Oliveira EB 2008. Efeito da densidade e do sítio sobre as curvas de distribuição diamétrica em bracatingais nativos (*Mimosa scabrella* Benth.), da região metropolitana de Curitiba. Ambiência 4(1), 37-50.

Miranda R, Fiorentin L, Péllico Netto S, Juvanho R, Dalla Corte A 2018. Prediction system for diameter distribution and wood production of *Eucalyptus*. Floresta e Ambiente 25(3): e20160548. https://doi.org/10.1590/2179-8087.054816

Nadarajah S, Pogány TK, Saxena RK 2012. On the characteristic function for Burr distributions. Statistics 46(3), 419-428. https://doi.org/10.1080/02331888.2010.513442

Nanos N, Wubalem T, Montero G, Gil L, Alia R 2000. Modelling resin production distributions for *Pinus pinaster* Ait using two probability functions. Annals of Forest Science 57(4), 369-377. https://doi.org/10.1051/forest:2000128

Pareto V (1897. Cours d'Économie Politique. Lausanne and Paris: Rouge and Cie Vol. I.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, 5(50), 157-172.

Piva LRO, Ruza MS, Heck TO, Corte APD, Péllico Netto S, Behling A 2020. Effect of high forest and coppice regimes on the diameter distribution of *Eucalyptus urophylla* S.T. Blakely stands. Scientia Forestalis 48(127), 1-11. https://doi.org/10.18671/scifor.v48n127.15

Qin J, Cao QV, Blouin DC 2007. Projection of a diameter distribution through time. Canadian Journal of Forest Research 37(1), 188-194. https://doi.org/10.1139/x06-216

RADAMBRASIL 1978. Levantamento de Recursos Naturais. Brasília: Departamento Nacional da Produção Mineral Ministério das Minas e Energia 561 p.

Rodriguez RN 1977. A guide to the Burr Type XII distributions. Biometrika 64(1), 129-134. https://doi.org/10.1093/biomet/64.1.129

Rupšys P, Petrauskas E 2017. A new paradigm in modelling the evolution of a stand via the distribution of tree sizes. Scientific Reports 7(1), 1-13. https://doi.org/10.1038/s41598-017-16100-2

Sandoval S, Cancino J, Rubilar R, Esquivel E, Acuna E, Muñoz F, Espinosa M 2012. Probability distributions in high-density dendroenergy plantations. Forest Science 58(6), 663-672. https://doi.org/10.5849/forsci.11-028

Sanquetta CR, Oliveira Filho PC, Bartoszeck ACPS, Durigan ME, Klechowicz N, Nisgoski S 1998. Efeito do espaçamento de plantio em reflorestamentos. I. *Eucalyptus grandis* Hill ex Maiden e *Eucalyptus saligna* Sm em Itararé SP. Revista Acadêmica 9(2), 41-52.

Schmidt LN, Machado SA, Pelissari AL, Silva GF 2019. Dynamics of eucalyptus diameter distribution in the state of Minas Gerais. Floresta e Ambiente 26(2), 1-9. https://doi.org/10.1590/2179-8087.015617

Schmidt LN, Sanquetta MNI, McTague JP, Silva GF, Fraga Filho CV, Sanquetta CR, Scolforo JRS 2020. On the use of the Weibull distribution in modeling and describing diameter distributions of clonal eucalypt stands. Canadian Journal of Forest Research 50(10), 1050-1063. https://doi.org/10.1139/cjfr-2020-0051

Scolforo JRS 2006. Biometria Florestal: Modelos de Crescimento e Produção. Lavras: UFLA/FAEPE.

Silva EQ, Péllico Netto S, Machado SA, Sanquetta CR 2003. Função densidade de probabilidade aplicável à ciência florestal. Floresta 33(3), 285-294. https://doi.org/10.5380/rf.v33i3.2262

Smirnov N 1948. Table for estimating the goodness of fit of empirical distributions. The Annals of Mathematical Statistics 19(2), 279-281. https://doi.org/10.1214/aoms/1177730256

Srivastava MS 1965. Characterizations of Pareto's distributions. The Annals of Mathematical Statistics 36 361-362.

Tadikamalla PR 1980. A look at the Burr and related distributions. International Statistical Review 48(3), 337-345. https://doi.org/10.2307/1402945

UFV- Universidade Federal de Viçosa 2010. Mapa de Solos do Estado de Minas Gerais - Legenda Expandida. Belo Horizonte.

Xu L, Wood GR, Woollons RC, Whyte AGD 1992. Stand table prediction with Reverse Weibull and Extreme Value density functions: some theoretical considerations. Forest Ecology and Management 48(1-2), 175-178. https://doi.org/10.1016/0378-1127(92)90129-W

Waldy J, Kershaw Jr JA, Weiskittel A, Ducey MJ 2022. Diameter distribution model development of tropical hybrid *Eucalyptus* clonal plantations in Sumatera Indonesia: A comparison of estimation methods. New Zealand Journal of Forestry Science 52: 1-14. https://doi.org/10.33494/nzjfs522022x151x

Zhang L, Gove JH, Liu C, Leak WB 2001. A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid uneven-aged stands. Canadian Journal of Forest Research 31(9), 1654-1659. https://doi.org/10.1139/x01-086

Zhang L, Packard KC, Liu C 2003. A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce-fir stands in northeastern North America. Canadian Journal of Forest Research 33(7), 1340-1347. https://doi.org/10.1139/x03-054

Appendicies

Appendix 1: OPTMBEL function

```
using CSV, DataFrames, Plots, ADNLPModels,
ForwardDiff, NLPModels, JSOSolvers, LinearAlgebra,
Logging, Printf,Percival, NLPModelsIpopt
gr()

function OPTMBEL(xdata,ydata)
  lvar = [1e-16; zeros(4)]
  model = (x,p) -> x^p[1] / ((p[2] + p[3] *
x^p[4])^p[5])
  erro(x, y, p) = model(x, p) - y

  λ = 1e-6
  f(p) = sum(erro(xi, yi, p)^2 for (xi, yi) in zip(xdata,
ydata)) + λ * norm(p)^2
  c(p) = [
   p[5] - 2p[1];
   p[5] * p[4] - p[1] - p[4] - 2;
   p[5] * p[4] - 2p[4] + 1;
  ]
  lcon = zeros(3)
  ucon = fill(Inf, 3)
  nlp = ADNLPModel(f, p0, lvar, fill(Inf, 5), c, lcon, ucon)
  # nlp = ADNLPModel(f, p0, zeros(5), fill(Inf, 5))
  local output
  try
   output = ipopt(nlp)
   # output = percival(nlp, max_eval=10000, subsolver_
logger=ConsoleLogger())
   # output = with_logger(NullLogger()) do
    # tron(nlp, variant=:Newton, max_eval=10000)
    # percival(nlp, max_eval=10000)
   # end
  catch ex
   println("Falhou para $filename: $ex")
   return
  end
  p = output.solution
  residuo = round(norm(model.(xdata, Ref(p)) - ydata),
digits=4)

  println("p = $p")
  println("status = ", output.status)
  println("residuo = ", residuo)

  open("sylvio/saida.txt", "w") do io
   println(io, "original = [")
   for i = 1:length(output.solution)
    println(io, p0[i])
   end
   println(io, "]")
   println(io, "solution = [")
   for i = 1:length(output.solution)
    println(io, output.solution[i])
   end
   println(io, "]")
   println(io, "restrição: [")
   cp = c(output.solution)
   for i = 1:3
    println(io, cp[i])
   end
   println(io, "]")
   println(io, "status = ", output.status)
  end

  open("sylvio/saida.tex", "w") do io
   println(io, raw"\begin{center}")
   println(io, raw"\begin{tabular}{r|r}")
   println(io, "inicial & otima \\\\ \\hline")
   r = round.(p, sigdigits=6)
   for i = 1:5
    println(io, "$(p0[i]) & $(r[i]) \\\\")
   end
   println(io, raw"\end{tabular}")
   println(io, raw"\end{center}")
  end


end

#Function OPTMBEL(xdata,ydata)
OPTMBEL(X, Y)
```

Appendix 2: ESPN function

```
using SpecialFunctions

function ESPN(a,b,c,d,e)
  Média=((gamma(e-((a+2)/d)))*(gamma((a+2)/d)))/
(((c/b)^(1/d))*(gamma(e-((a+1)/
d)))*gamma((a+1)/d))
  Var=(((gamma(e-((a+3)/d)))*(gamma((a+3)/d)))/
(((c/b)^(2/d))*(gamma(e-((a+1)/d)))*(gamma((a+1)/
d))))-(Média^2)
  dp=sqrt(Var)
  Moda=((a*b)/(c*(e*d-a)))^(1/d)
  #Pontos de inflexão
  ZiU=(-b*c*((2*a)*(a-1)+d*e*(1-d-
2*a))+(((b^2)*(c^2)*((2*a*(a-d*e-
1)+d*e*(1-d))^2)-4*(b^2)*(c^2)*((a*(a-
1))+d*e*(d*e+1-2*a))*a*(a-1))^0.5))/
(2*(c^2)*(a*(a-1)+d*e*(d*e+1-2*a)))
  ZiL=(-b*c*((2*a)*(a-1)+d*e*(1-d-
2*a))-(((b^2)*(c^2)*((2*a*(a-d*e-
1)+d*e*(1-d))^2)-4*(b^2)*(c^2)*((a*(a-
1))+d*e*(d*e+1-2*a))*a*(a-1))^0.5))/
(2*(c^2)*(a*(a-1)+d*e*(d*e+1-2*a)))
  xiU=ZiU^(1/d)
  xiL=ZiL^(1/d)
  #Assimetria (S)
  EX3=((gamma(e-((a+4)/d)))*(gamma((a+4)/d)))/
(((c/b)^(3/d))*(gamma(e-((a+1)/
d)))*gamma((a+1)/d))
  m3=EX3-3*(Média*(Var+Média^2))+2*(Média^3)
  S=m3/(dp^3)
  #Curtose
  EX4=((gamma(e-((a+5)/d)))*(gamma((a+5)/d)))/
(((c/b)^(4/d))*(gamma(e-((a+1)/
d)))*gamma((a+1)/d))
  m4=EX4-4*(Média*(m3+3*(Média*(Var+Média^2))-
2*Média^3))+6*(Média^2*(Var+Média^2))-3*Média^4
  Curtose=(m4/(dp^4))-3
  println("Média: ", Média)
  println("Variância: ", Var)
  println("Despio-padrão: ", dp)
  println("Moda: ", Moda)
  println("Pontos de infelxão: ", [xiL,xiU])
  println("Assimetria: ", S)
  println("Curtose: ", Curtose)
end

ESPN(a,b,c,d,e)
```

Appendix 3: OPTMBELR function

```
using CSV, DataFrames, Plots, ADNLPModels,
ForwardDiff, NLPModels, JSOSolvers,
LinearAlgebra,Logging, Printf,Percival,
NLPModelsIpopt, StatsBase

 #Dados de biomassa total e dap de acácia negra
 X=CSV.read("acacian.csv", DataFrame)

 function OPTMBEL(xdata,ydata,p0)
  lvar = [1e-16; zeros(4)]
  model = (x,p) -> x^p[1] / ((p[2] + p[3] *
x^p[4])^p[5])
  erro(x, y, p) = model(x, p) - y

  λ = 1e-6
  f(p) = sum(erro(xi, yi, p)^2 for (xi, yi) in zip(xdata,
ydata)) + λ * norm(p)^2
  c(p) = [
   p[5] - 2p[1];
   p[5] * p[4] - p[1] - p[4] - 2;
   p[5] * p[4] - 2p[4] + 1;
  ]
  lcon = zeros(3)
  ucon = fill(Inf, 3)
  nlp = ADNLPModel(f, p0, lvar, fill(Inf, 5), c, lcon, ucon)

  #nlp = ADNLPModel(f, p0, zeros(5), fill(Inf, 5))
  local output
  try
   output = ipopt(nlp)
   # output = percival(nlp, max_eval=10000, subsolver_
logger=ConsoleLogger())
   # output = with_logger(NullLogger()) do
    # tron(nlp, variant=:Newton, max_eval=10000)
    # percival(nlp, max_eval=10000)
   # end
  catch ex
   println("Falhou para $filename: $ex")
   return
  end
  return output.solution
 end

 # Função de validação
 function OPTMBELR(X::DataFrame, R::Int64,
p0::Vector{Float64}, out::String)
  if length(p0) !== 5
   error("Erro: O tamanho do vetor não pode ser
diferent de 5.")
  end
  plot()
  Tabela1 = DataFrame()
  Tabela2 = DataFrame()
  Tabela3 = DataFrame()
  Tabela4 = DataFrame()
  resultado_grap = Matrix{Float64}(undef, 1000,R)
  coef = Matrix{Float64}(undef, R, 5)
  KScalc = Vector{Float64}(undef,R)
```

```
  for i in 1:R

    Mínimo=minimum(X.DAP)
    Máximo=maximum(X.DAP)
    Z=(X.DAP.-Mínimo)./(Máximo-Mínimo)

    #Limites de classes (   |-  Fechado para a esquerda e
aberto para a direita)
    x=DataFrame(ZDAP=Z)
    x=x.ZDAP
    n_individuos = size(x,1)
    df = DataFrame(ZDAP=Z)

    global sample_1 = sample(1:nrow(df),
Int64(round((n_individuos*0.80), digits = 0)),
replace=false)
    df1 = df[sample_1, :]
    test_rows = setdiff(1:nrow(df), sample_1)
    global df_teste = df[test_rows, :]
    println(df_teste)
    x1=df1.ZDAP
    #Limites de classes (   |-  Fechado para a esquerda e
aberto para a direita)
    x=x1

    Classe_1=x[(0 .<= x .< 0.0833)]
    Classe_2=x[(0.0833 .<= x .< 0.1667)]
    Classe_3=x[(0.1667 .<= x .< 0.2500)]
    Classe_4=x[(0.2500 .<= x .< 0.3333)]
    Classe_5=x[(0.3333 .<= x .< 0.4167)]
    Classe_6=x[(0.4167 .<= x .< 0.5000)]
    Classe_7=x[(0.5000 .<= x .< 0.5833)]
    Classe_8=x[(0.5833 .<= x .< 0.6667)]
    Classe_9=x[(0.6667 .<= x .< 0.7500)]
    Classe_10=x[(0.7500 .<= x .< 0.8333)]
    Classe_11=x[(0.8333 .<= x .< 0.9167)]
    Classe_12=x[(0.9167 .<= x .< 1)]

    #Frequência absoluta 80% dos dados
    f1_Classe1=length(Classe_1)/size(sample_1, 1)
    f1_Classe2=length(Classe_2)/size(sample_1, 1)
    f1_Classe3=length(Classe_3)/size(sample_1, 1)
    f1_Classe4=length(Classe_4)/size(sample_1, 1)
    f1_Classe5=length(Classe_5)/size(sample_1, 1)
    f1_Classe6=length(Classe_6)/size(sample_1, 1)
    f1_Classe7=length(Classe_7)/size(sample_1, 1)
    f1_Classe8=length(Classe_8)/size(sample_1, 1)
    f1_Classe9=length(Classe_9)/size(sample_1, 1)
    f1_Classe10=length(Classe_10)/size(sample_1, 1)
    f1_Classe11=length(Classe_11)/size(sample_1, 1)
    f1_Classe12=length(Classe_12)/size(sample_1, 1)

    x2 = df_teste.ZDAP

    Classe_1=x2[(0 .<= x2 .< 0.0833)]
    Classe_2=x2[(0.0833 .<= x2 .< 0.1667)]
    Classe_3=x2[(0.1667 .<= x2 .< 0.2500)]
    Classe_4=x2[(0.2500 .<= x2 .< 0.3333)]
    Classe_5=x2[(0.3333 .<= x2 .< 0.4167)]
    Classe_6=x2[(0.4167 .<= x2 .< 0.5000)]
    Classe_7=x2[(0.5000 .<= x2 .< 0.5833)]
    Classe_8=x2[(0.5833 .<= x2 .< 0.6667)]
    Classe_9=x2[(0.6667 .<= x2 .< 0.7500)]
    Classe_10=x2[(0.7500 .<= x2 .< 0.8333)]
    Classe_11=x2[(0.8333 .<= x2 .< 0.9167)]
    Classe_12=x2[(0.9167 .<= x2 .< 1)]

    #Frequência absoluta 20% dos dados
    f1_Classe1_30=length(Classe_1)/(nrow(X) -
size(sample_1, 1))
    f1_Classe2_30=length(Classe_2)/(nrow(X) -
size(sample_1, 1))
    f1_Classe3_30=length(Classe_3)/(nrow(X) -
size(sample_1, 1))
    f1_Classe4_30=length(Classe_4)/(nrow(X) -
size(sample_1, 1))
    f1_Classe5_30=length(Classe_5)/(nrow(X) -
size(sample_1, 1))
    f1_Classe6_30=length(Classe_6)/(nrow(X) -
size(sample_1, 1))
    f1_Classe7_30=length(Classe_7)/(nrow(X) -
size(sample_1, 1))
    f1_Classe8_30=length(Classe_8)/(nrow(X) -
size(sample_1, 1))
    f1_Classe9_30=length(Classe_9)/(nrow(X) -
size(sample_1, 1))
    f1_Classe10_30=length(Classe_10)/(nrow(X) -
size(sample_1, 1))
    f1_Classe11_30=length(Classe_11)/(nrow(X) -
size(sample_1, 1))
    f1_Classe12_30=length(Classe_12)/(nrow(X) -
size(sample_1, 1))

    Classe=[1,2,3,4,5,6,7,8,9,10,11,12]
    global Centro_de_Classe=[0.0417, 0.1250, 0.2083,
0.2917, 0.3750, 0.4583, 0.5417, 0.6250, 0.7083, 0.7917,
0.8750, 0.9583]

    f1=[f1_Classe1,f1_Classe2,f1_Classe3, f1_Classe4,f1_
Classe5, f1_Classe6,f1_Classe7, f1_Classe8,f1_Classe9,f1_
Classe10,f1_Classe11,f1_Classe12]
    Tabela1=DataFrame(i=Classe, CC=Centro_de_
Classe,fi=f1)

    #Tabela de frequência
    f1=[f1_Classe1_30,f1_Classe2_30,f1_Classe3_30, f1_
Classe4_30,f1_Classe5_30, f1_Classe6_30,f1_Classe7_30,
    f1_Classe8_30,f1_Classe9_30,f1_Classe10_30,f1_
Classe11_30,f1_Classe12_30]
    Tabela2=DataFrame(i=Classe, CC=Centro_de_
Classe,fi=f1)

    #Frequencia Observada
    Tabela3 = cumsum(Tabela2.fi)./sum(Tabela2.fi)

    #Inserir o resultado do ajuste
    p1 = DIST.OPTMBEL(Tabela1.CC, Tabela1.fi, p0)

    coef[i,:] = p1

    f(x) =x^coef[i,1] / ((coef[i,2] + coef[i,3] *
x^coef[i,4])^coef[i,5])
```

```
    xGrid = range(0, 1, length = 1000)
    XG = range(0, 1, length = 1000)
    YG1=f.(xGrid)

    resultado_grap[:,i] = YG1

    if i === 1
     plot()
     global plt_normal = plot(XG,[resultado_grap[:,1]],
linewidth = 2,linecolor = :black, legend = false, xlabel =
"Class center (x)", ylabel = "f(x)", xticks = round.(Centro_
de_Classe, digits = 2))
    else
     global plt_normal = plot!(XG,[resultado_grap[:,i]],
linewidth = 2,linecolor = :black, legend = false)

    end

    if i === R
     for j in 1:R
      if j === 1
       plot()
       global plt_acumulado = plot(XG,
cumsum((resultado_grap[:,1]./sum(resultado_
grap[:,1])), dims = 1), xlabel = "Class center (x)", ylabel
= "F(x)", xticks = round.(Centro_de_Classe, digits = 2))
      else
       global plt_acumulado = plot!(XG,
cumsum((resultado_grap[:,j]./sum(resultado_grap[:,j])),
dims = 1), legend = false)
      end
     end
    end

    # Frequencia estimada
    Tabela4 = f.(Centro_de_Classe) #normal

    AcumuladoFest = cumsum(Tabela4)/sum(Tabela4)
#Acumulada

    #Tabela para calculo KS
    global Tabela5 = DataFrame(CC = Centro_de_Classe,
FO = Tabela3, FE = AcumuladoFest, Diff = abs.(Tabela3
.- AcumuladoFest))

    #Kscalculado
    global KScalc[i] = maximum(Tabela5.Diff)

    #Ks tabelado
    global Kstab = 1.36/(sqrt((nrow(X) - size(sample_1,
1))))
    global teste = coef
   end
   menor_kscalc = (length(KScalc[(KScalc .<=
Kstab)])/R) * 100

   minimum_values = minimum(coef, dims = 1)
   maximum_values = maximum(coef, dims = 1)

   #Eixo X colocar centro de classes
   histogram(KScalc, bins = 10, label = "Calculated
Value", xlabel = "Ks-values", ylabel = "Frequency")
   savefig(plot!([Kstab, Kstab], [0,30], label = "Critical
Value"), "$out\\Ks_values")
   savefig(histogram(coef[:,1], bins = range(minimum_
values[1], maximum_values[1], 10), title = "a", legend
= false, xlabel = "Values of the parameter a", ylabel =
"Frequency"), "$out\\Results of paramter a")
   savefig(histogram(coef[:,2], bins = range(minimum_
values[2], maximum_values[2], 10), title = "b", legend
= false, xlabel = "Values of the parameter b", ylabel =
"Frequency"), "$out\\Results of paramter b")
   savefig(histogram(coef[:,3], bins = range(minimum_
values[3], maximum_values[3], 10), title = "c", legend
= false, xlabel = "Values of the parameter c", ylabel =
"Frequency"), "$out\\Results of paramter c")
   savefig(histogram(coef[:,4], bins = range(minimum_
values[4], maximum_values[4], 10), title = "d", legend
= false, xlabel = "Values of the parameter d", ylabel =
"Frequency"), "$out\\Results of paramter d")
   savefig(histogram(coef[:,5], bins = range(minimum_
values[5], maximum_values[5], 10), title = "e", legend
= false, xlabel = "Values of the parameter e", ylabel =
"Frequency"), "$out\\Results of paramter e")
   savefig(plt_normal, "$out\\Plot Normal")
   savefig(plt_acumulado, "$out\\Plot Acumulado")
   println("$menor_kscalc%, proporção de vezes que
o valor de KS calculado foi menor que o valor de KS
tabelado")
 end

 OPTMBELR(X, 100, [1.0,2.0,3.0,4.0,5.0], "G:\\
Resultados2\\")
```