

Article

Rotated Lorenz Curves of Biological Size Distributions Follow Two Performance Equations

Peijian Shi ^{1,*} , Linli Deng ¹ and Karl J. Niklas ^{2,*} 

¹ Bamboo Research Institute, College of Science, Nanjing Forestry University, 159 Longpan Road, Nanjing 210037, China; denglinli@njfu.edu.cn

² School of Integrative Plant Science, Cornell University, 236 Tower Road, Ithaca, NY 14853, USA

* Correspondence: pjshi@njfu.edu.cn (P.S.); kjn2@cornell.edu (K.J.N.); Tel.: +86-25-8542-7231 (P.S.)

Abstract: The Lorenz curve is used to describe the relationship between the cumulative proportion of household income and the number of households of an economy. The extent to which the Lorenz curve deviates from the line of equality (i.e., $y = x$) is quantified by the Gini coefficient. Prior models are based on the simulated and empirical data of income distributions. In biology, the Lorenz curves of cell or organ size distributions tend to have similar shapes. When the Lorenz curve is rotated by 135 degrees counterclockwise and shifted to the right by a distance of $\sqrt{2}$, a three-parameter performance equation (PE), and its generalized version with five parameters (GPE), accurately describe this rotated and right-shifted curve. However, in prior studies, PE and GPE were not compared with the other Lorenz equations, and little is known about whether the skewness of the distribution could influence the validity of these equations. To address these two issues, simulation data from the beta distributions with different skewness values and six empirical datasets of plant (organ) size distributions were used to compare PE and GPE with three other Lorenz equations in describing the rotated and right-shifted plant (organ) size distributions. The root-mean-square error and Akaike information criterion were used to assess the validity of the two performance equations and the three other Lorenz equations. PE and GPE were both validated in describing the rotated and right-shifted simulation and empirical data of plant (organ) distributions. Nevertheless, GPE worked better than PE and the three other Lorenz equations from the perspectives of the goodness of fit, and the trade-off between the goodness of fit and the model structural complexity. Analyses indicate that GPE provides a powerful tool for quantifying size distributions across a broad spectrum of organic entities and can be used in a variety of ecological and evolutionary applications. Even for the simulation data from hypothetical extreme skewed distribution curves, GPE still worked well.

Keywords: line of absolute equality; inverted U-shaped curves; nonlinear regression; parameter estimation; size distribution



Citation: Shi, P.; Deng, L.; Niklas, K.J. Rotated Lorenz Curves of Biological Size Distributions Follow Two Performance Equations. *Symmetry* **2024**, *16*, 565. <https://doi.org/10.3390/sym16050565>

Academic Editor: Calogero Vetro

Received: 5 March 2024

Revised: 15 April 2024

Accepted: 22 April 2024

Published: 5 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

One of the hallmarks of biology is variations in the size, shape, and mass of individuals belonging to the same hierarchical level of organization (e.g., cells and organs), which affect a broad range of physiological, ecological, and evolutionary phenomena. Consequently, the accurate quantification of size, shape, or mass frequency distributions is critical to understanding such phenomena. Here, we draw from economic theory and show that a particular approach has broad applicability to understanding and quantifying organic size distributions.

Specifically, the Gini coefficient is widely used in economics to measure the inequality of household incomes. It is based on the Lorenz curve (LC), which is a graphical representation of the cumulative proportion (or percentage) of household income plotted against the cumulative proportion (or percentage) of the number of households in the economy [1]. The Gini coefficient is equal to the area formed by the LC and the line of absolute equality

(i.e., $y = x$) divided by one half of the area of the unit square. Botanists have used the Gini coefficient to quantify the variation in tree and seed size [2,3], but few studies have checked the validity of the LC in describing the size distributions of other organic structures. Likewise, economists have tested the LC's validity, however, mainly focusing on income distributions [4–6]. Therefore, there is a pressing need to explore and refine the applicability of the LC and Gini index in both biology and economics.

In an effort to improve the accuracy of the LC–Gini index approach using plants as test organisms, we note that the rotated LC of leaf area distribution on a plant resembles a thermal performance curve (Figure 1a,b), which can be generated by a nonlinear equation proposed by [7]:

$$y = c \left(1 - e^{-K_1(x-x_1)} \right) \left(1 - e^{K_2(x-x_2)} \right), \quad (1)$$

where y represents the jumping distance of the green frog (*Rana clamitans* Latreille) at body temperature x , and c , K_1 , K_2 , x_1 and x_2 are parameters to be estimated. K_1 and K_2 influence the instantaneous rates of change on the left and right parts of the performance curve, respectively; x_1 and x_2 represent the minimum and maximum threshold temperatures, i.e., the lower and upper intersections between the performance curve and the x -axis. In this equation, the thermal performance is hypothesized to terminate below x_1 and above x_2 , i.e., $y = 0$ when $x \leq x_1$ and $x \geq x_2$. Equation (1), henceforth denoted as PE, generates symmetrical and asymmetrical inverted U-shaped curves.

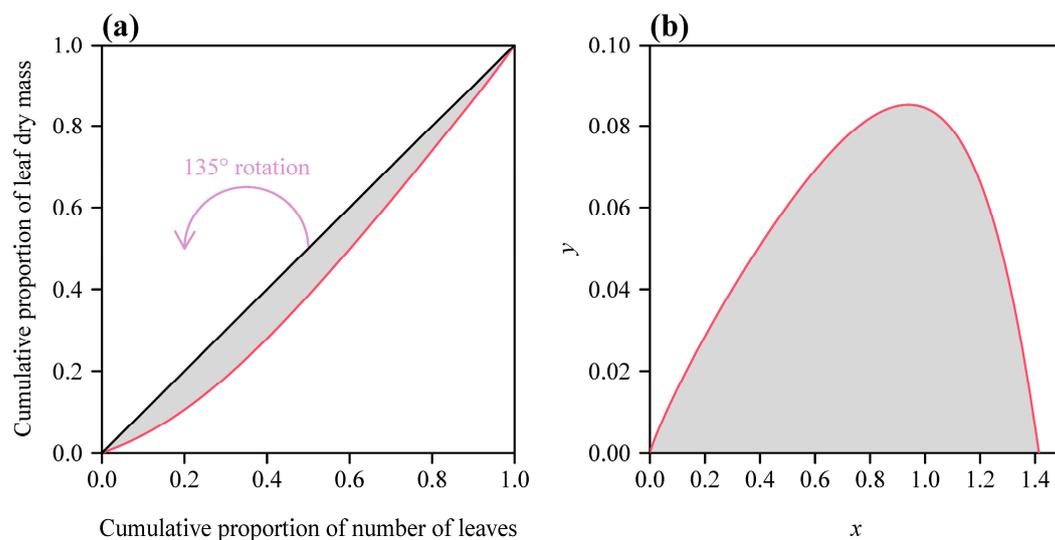


Figure 1. An illustration of the original Lorenz curve (a) and its rotated and right-shifted version (b) using data for the individual leaf dry mass distribution of a bamboo (*Shibataea chinensis* Nakai) culm. The Gini coefficient equals two times the shaded area. Panel (a) shows the cumulative proportion of the individual leaf dry mass on the culm plotted against the cumulative proportion of the number of leaves on the culm. The red curve in panel (b) was obtained by rotating the Lorenz curve (red curve) in panel (a) by 135° counterclockwise and shifting it to the right by a distance of $\sqrt{2}$.

To increase the flexibility of data fitting, two additional parameters, α and β , can be introduced to Equation (1) [8]:

$$y = c \left(1 - e^{-K_1(x-x_1)} \right)^\alpha \left(1 - e^{K_2(x-x_2)} \right)^\beta. \quad (2)$$

Equation (2) is referred to as the generalized performance equation (GPE) hereinafter. Note that the two pairs of parentheses on the right-side of Equation (1) are transformed by using two power functions with powers α and β in Equation (2). According to the definition of the LC (Figure 1a), the numerical range of the cumulative proportion of the number (i.e., the x -axis) and that of the cumulative proportion of the size (i.e., the y -axis)

are both between 0 and 1. Thus, the length of the diagonal of the unit square is $\sqrt{2}$, and the rotation and right-shift do not influence its numerical value. In this way, x_1 and x_2 in Equations (1) and (2) are actually known constants (i.e., 0 and $\sqrt{2}$, respectively). GPE was found to be better than PE in describing the leaf area distribution at an individual plant level [8].

However, there are still two critical questions that are not answered by prior studies related to the use of PE and GPE: (i) Does the skewness of biological size distributions influence the validity of the two performance equations and other Lorenz equations? (ii) Are the two performance equations better than other Lorenz equations in nonlinear regression? To address these two important questions, we used simulation data from beta distributions with different skewness values, and six empirical datasets of plant organ size distributions, including stomatal area (i.e., the area of the profile formed by two guard cells), leaf dry mass, tepal projection area, fruit volume, seedhead length, and the diameter (at breast height) distribution of a temperate forest (provided in the online Supplementary Tables S1–S6), to test the validity of PE and GPE and three other Lorenz equations, and to compare the two performance equations with the three other Lorenz equations. The data of the cumulative proportion of plant organ size vs. the cumulative proportion of the number of organs per plant (or per quadrat) were rotated by 135 degrees counterclockwise and shifted to the right by a distance of $\sqrt{2}$ to meet the requirements for fitting these equations.

2. Materials and Methods

2.1. The Three Other Lorenz Equations

Three other Lorenz equations were used to make comparisons with PE and GPE, which are found to be able to describe many size distributions of abiotic and biotic areas [9]:

(i) The Sarabia equation (henceforth denoted as SarabiaE; see [10]):

$$y_L = (1 - \lambda + \eta)x_L + \lambda x_L^{a_1+1} - \eta \left[1 - (1 - x_L)^{a_2+1} \right], \quad (3)$$

where x_L and y_L represent the cumulative proportion of the number, and the cumulative proportion of income or size, respectively; and λ , η , a_1 , and a_2 are constants to be estimated, where $a_1 \geq 0$, $a_2 + 1 \geq 0$, $\eta a_2 + \lambda \leq 1$, $\lambda \geq 0$, and $\eta a_2 \geq 0$.

(ii) The Sarabia-Castillo-Slottje equation (henceforth denoted as SCSE; see [11]):

$$y_L = x_L^\gamma [1 - (1 - x_L)^{\alpha_1}]^{\beta_1}, \quad (4)$$

where x_L and y_L represent the cumulative proportion of the number, and the cumulative proportion of income or size, respectively; and α_1 , β_1 , and γ are constants to be estimated, where $0 < \alpha_1 \leq 1$, $\beta_1 \geq 1$, and $\gamma \geq 0$.

(iii) The Sitthiyot-Holasut Equation (henceforth denoted as SHE; see [9]):

$$y_L = (1 - \rho) \left[\left(\frac{2}{P+1} \right) \left(\frac{x_L - \delta}{1 - \delta} \right) \right] + \rho \left[(1 - \omega) \left(\frac{x_L - \delta}{1 - \delta} \right)^P + \omega \left\{ 1 - \left[1 - \left(\frac{x_L - \delta}{1 - \delta} \right) \right]^{1/P} \right\} \right], \quad (5)$$

when $x_L > \delta$; and $y_L = 0$, when $x_L \leq \delta$. Here, x_L and y_L represent the cumulative proportion of the number, and the cumulative proportion of income or size, respectively; and δ , ρ , ω , and P are constants to be estimated, where $0 \leq \delta < 1$, $0 \leq \rho \leq 1$, $0 \leq \omega \leq 1$, and $P \geq 1$.

We rotated and right-shifted these three Lorenz equations to make the abscissa values range between 0 and $\sqrt{2}$.

2.2. The Simulation Data from Beta Distributions

To examine the generality of our approach, we tested the validity of the two performance equations and the three other Lorenz equations for 120 random numbers, i.e., realizations, simulated from beta distributions with different skewness values [12]:

$$f(z) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} z^{a-1}(1-z)^{b-1}, \quad (6)$$

where f is the density function of a random variable z . The skewness (S_k) of the beta distribution is given by the formula

$$S_k = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}. \quad (7)$$

We set different combinations of the two parameters, a and b , both ranging between 0.5 and 10, which generated a range of skewness from -2.3 to 2.3 (Figure 2). However, fixing one parameter's value and varying the other parameter's value from 0.5 to 10 did not render skewness values to span the range between -2.3 and 2.3 . Thus, we used the circular formula $a^2 + b^2 = 10^2$ to get b when a varies from 0.5 to 10 in 300 equidistant values. We then obtained 300 different skewness values, and simulated 120 random numbers as the hypothetical size values from each of the beta distributions corresponding to the 300 skewness values (shown by the blue arc in Figure 2). Based on prior experience, it is likely that a range of skewness from -2.3 to 2.3 spans the skewness of many actual biological size distributions. Here, we present four examples from the 300 skewness values: (i) extremely left-skewed, (ii) moderately left-skewed, (iii) moderately right-skewed, and (iv) extremely right-skewed, by setting (i) $a = 10$ and $b = 0.5$, (ii) $a = 10$ and $b = 2$, (iii) $a = 2$ and $b = 10$, and (iv) $a = 0.5$ and $b = 10$ (Figure 3).

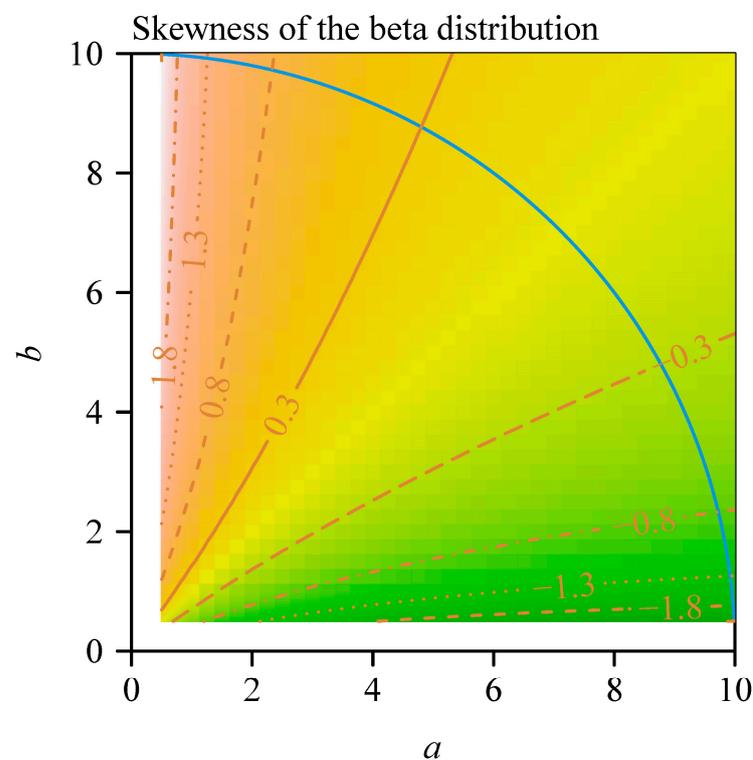


Figure 2. Isolines of the skewness of beta distributions generated from different combinations of parameters a and b . The numbers associated with different types of lines correspond to the skewness values, and the blue curve was generated using the circular equation $a^2 + b^2 = 10^2$.

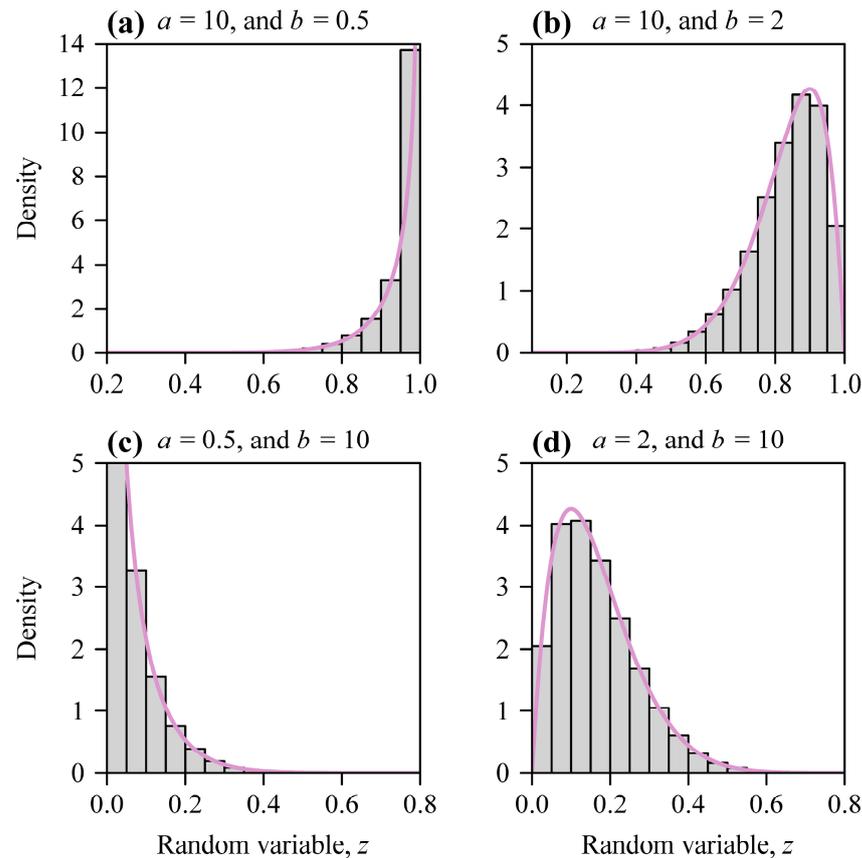


Figure 3. Density curves of four beta distributions. The parameters a and b are the model parameters of the beta density function. The purple curves are the density curves with different combinations of parameters a and b . (a) $a = 10$ and $b = 0.5$. (b) $a = 10$ and $b = 2$. (c) $a = 0.5$ and $b = 10$. (d) $a = 2$ and $b = 10$.

In practice, the Weibull function is frequently used to describe biological size distributions [13]. However, it is somewhat difficult to generate the same or approximate numerical range of random variables from different degrees of skewed curves using the Weibull function. Additionally, the effect of the skewness on the validity of the two performance equations in data fitting manifests a negligible difference from beta functions. Thus, we did not use the Weibull function and other distribution functions, such as the gamma function and the log-normal function [14], to generate the simulation data of biological size.

2.3. Data of Plant (Organ) Size Distributions

Six empirical datasets of plant (organ) size distributions were used to test the generality of the two performance equations, and the goodness of fit was used to examine the validity of the equations.

(i) Dataset 1: Data of 73 stomata in a $662 \mu\text{m} \times 444 \mu\text{m}$ micrograph of *Magnolia denudata* Desr. The stomatal length (L) and width (W) were directly measured using ImageJ software (version 1.54g; <https://imagej.nih.gov/ij/index.html> (accessed on 1 March 2024)), and stomatal area (i.e., the area profile formed by two guard cells) was estimated as $0.811 \times LW$. The site and sampling information are provided in [15], and the validity of the estimation method on the stomatal area is provided in [16].

(ii) Dataset 2: Data of 23 leaves of a bamboo (*Shibataea chinensis* Nakai) culm. Envelopes containing laminae were placed into a ventilated oven (XMTD-8222; Jinghong Experimental Equipment Co., Ltd., Shanghai, China) at $80 \text{ }^\circ\text{C}$ for at least 72 h to determine the dry mass of each leaf lamina using an electronic balance (ME204/02, Mettler Toledo Company, Greifensee, Switzerland, with a measurement accuracy of 0.0001 g). The site and sampling information can be found in [8].

(iii) Dataset 3: Data of nine tepals of a *Magnolia × soulangeana* Soul.-Bod. flower. Each tepal was scanned at 600-dpi resolution with a photo scanner (V550, Epson Indonesia, Batam, Indonesia). Adobe Photoshop 2021 (version 22.4.2; Adobe Systems Incorporated, San Jose, CA, USA) was used to obtain black and white images of tepal profiles that were saved as .bmp image at a 600-dpi resolution. The protocols proposed by prior studies [17,18] based on Matlab (version $\geq 2009a$; MathWorks, Natick, MA, USA) were used to obtain the planar coordinates of the tepal boundary by calculating the pixel values of each image. The projection area of each tepal was calculated using the “bilat” function of the “biogeom” package (version 1.4.3) [19] based on R (version 4.3.3) [20]. Details of the data acquisition methods can be found in [21].

(iv) Dataset 4: Volume data of 35 fruits of an individual *Cucumis melo* L. var. *agrestis* Naud. plant. Volume is the volume of each fruit using a graduated cylinder with a 3 cm diameter. The site and sampling information can be found in [22].

(v) Dataset 5: Length data of 144 seedheads of *Setaria viridis* (L.) P. Beauv. in a 1 m \times 1 m quadrat. The study area is located in a field of Baima Experiment Station of Nanjing Forestry University (31°37'55" N, 119°07'42" E) in September 2015. The seedheads of *S. viridis* in the 15 m \times 15 m study area were measured, and the study area was divided into 125 quadrats of 1 m \times 1 m. One quadrat was randomly selected in the present study.

(vi) Dataset 6: Breast height diameter data of 81 trees with DBH ≥ 1 cm in a 50 m \times 50 m quadrat of a temperate forest in Beijing Songshan National Nature Reserve, China (40°30' 50" N, 115°49' 12" E), censused in August 2014. The site and forest census information can be found in [23].

2.4. Data Fitting and Model Assessment

Hypothetical simulations and empirically determined data of plant (organ) size distributions were first rotated by 135 degrees counterclockwise and shifted to the right by a distance of $\sqrt{2}$ to meet the requirements for fitting the two performance equations, i.e., PE and GPE. The three other Lorenz equations were rotated and right-shifted to fit the rotated and right-shifted data. The parameters of each equation were then estimated using the Nelder-Mead optimization algorithm [24] to minimize the residual sum of squares (RSS) between the observed and predicted y values.

The “fitLorenz” function in the “biogeom” package (version 1.4.3) [19] based on the statistical software R (version 4.3.3) [20] was used to fit the size distributions, and the root-mean-square error (RMSE) was used to measure the goodness of fit for each of the five equations. The Akaike information criterion (AIC) was also used to compare the five equations, which is usually recommended in nonlinear regression, because it provides a rigorous reflection of the tradeoff between the goodness of fit and model structural complexity compared to the adjusted coefficient of determination [25].

3. Results

The results from simulations of the hypothetical size using beta distributions with skewness values ranging between -2.3 and 2.3 validated the two performance equations. GPE was found to be the best among the five equations, and PE was better than SCSE and SHE apart from being less effective than GPE and SarabiaE, from both the perspective of the goodness of fit and the perspective of the tradeoff between the goodness of fit and model structural complexity (Figure 4). Figure 5 provides an example of fitting simulations from the beta distribution with $S_k = -0.81$ using the five equations. Figure 6 shows the results of fitting the data from the four types of representative skewness cases corresponding to Figure 3.

Analyses indicated that the rotated and right-shifted observations of the six empirical datasets were well fitted by GPE (Figure 7). For each dataset, GPE had the lowest RMSE and AIC values compared to the other four equations, i.e., GPE was superior to the others from the perspective of the goodness of fit, and the tradeoff between the goodness of fit and model structural complexity. PE ranked the third best model following the second

best, SarabiaE (Table 1). Therefore, the introduction of parameters α and β in Equation (2) increased the goodness of fit but not at the cost of increasing the model's complexity, although for a small sample size, GPE tended to overfit the rotated and right-shifted observations, producing a slightly concave part approaching the (0, 0) point (Figure 7c). Relative to the estimated GPE, the estimated PE always exhibited inverted U-shaped curves. Overall, for the six investigated datasets, the two performance equations and SarabiaE are more valid than SCSE and SHE.

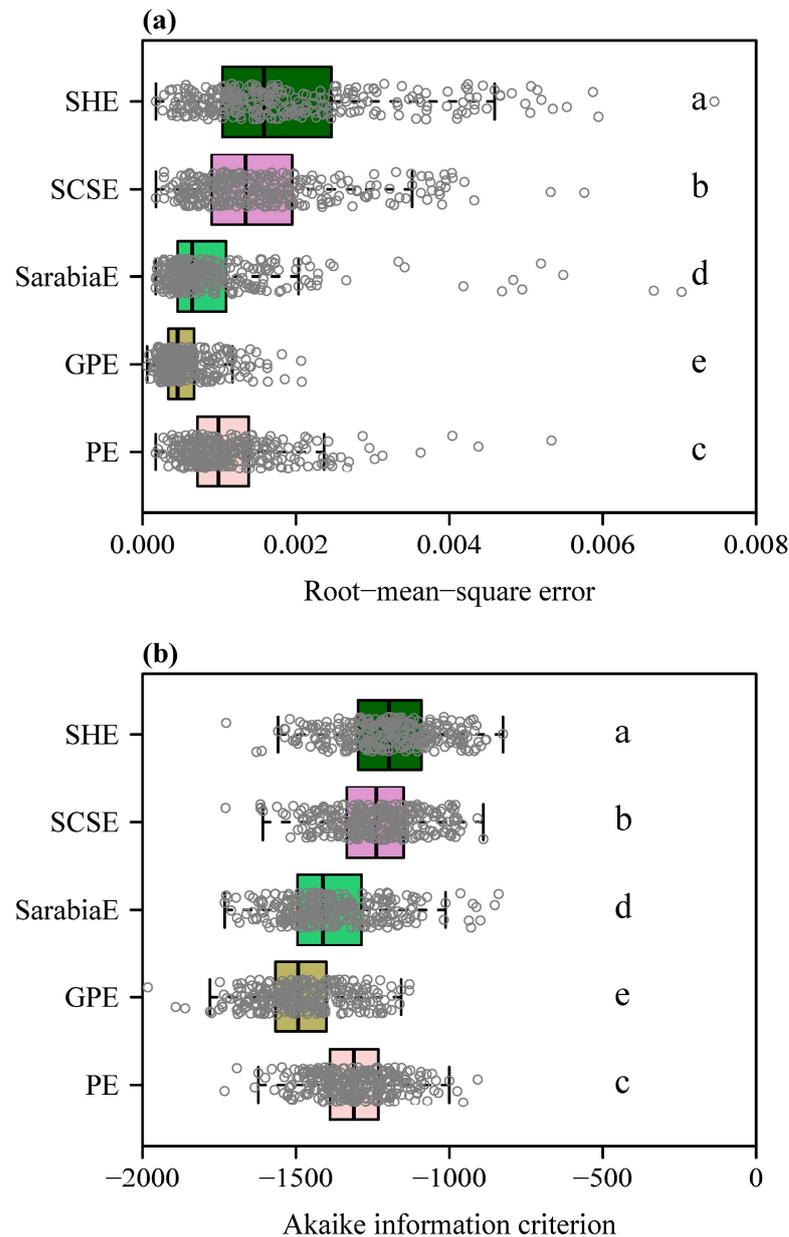


Figure 4. Boxplots of the root-mean-square errors (a), and Akaike information criteria (b), compared between any two of the five equations (i.e., PE, GPE, SarabiaE, SCSE, and SHE) for the 300 simulated datasets from the beta distributions with skewness values ranging from -2.3 to 2.3 . Significant differences between any two equations using the paired t -test at the 0.05 significance level are marked in different letters. The vertical solid line in each box represents the median; the whiskers extend to the most extreme data point, which is no more than 1.5 times the interquartile range from the box. The gray open circles show one dimensional scatter plots of the given data.

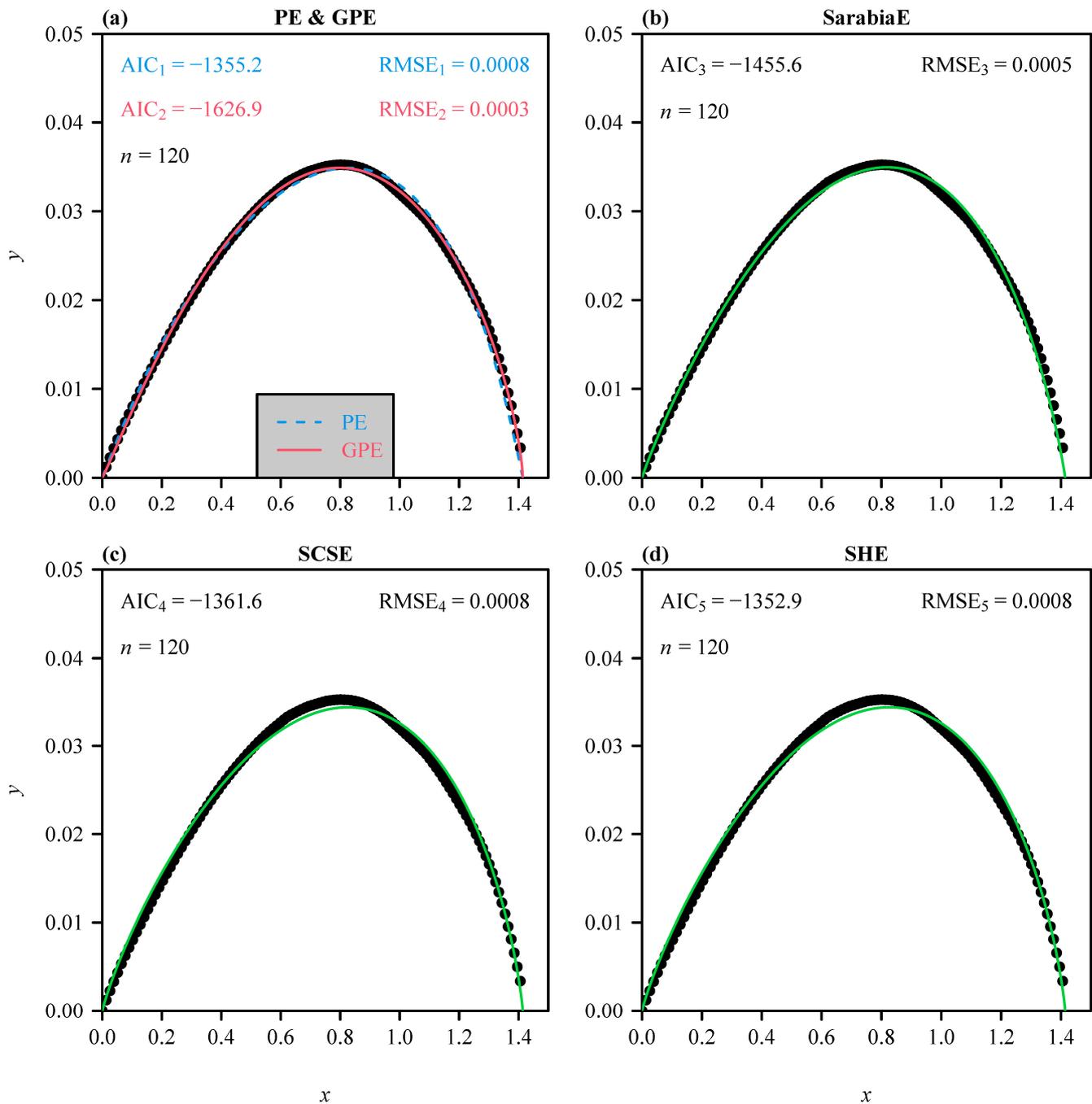


Figure 5. The rotated and right-shifted Lorenz curves (blue, red, and green lines) and simulations (closed circles) from the beta distribution with $S_k = -0.81$ (presented as an example of the simulated 300 skewness values) using the five equations. Panel (a) shows the results of fitting the two performance equations, i.e., PE and GPE; panels (b–d) show the results of fitting the three rotated and right-shifted Lorenz equations, i.e., SarabiaE, SCSE, and SHE. AIC_1 to AIC_5 are the Akaike information criteria of the estimated PE, GPE, SarabiaE, SCSE, and SHE, respectively; $RMSE_1$ to $RMSE_5$ are the root-mean-square errors of the estimated PE, GPE, SarabiaE, SCSE, and SHE, respectively; and n is the sample size of each dataset.

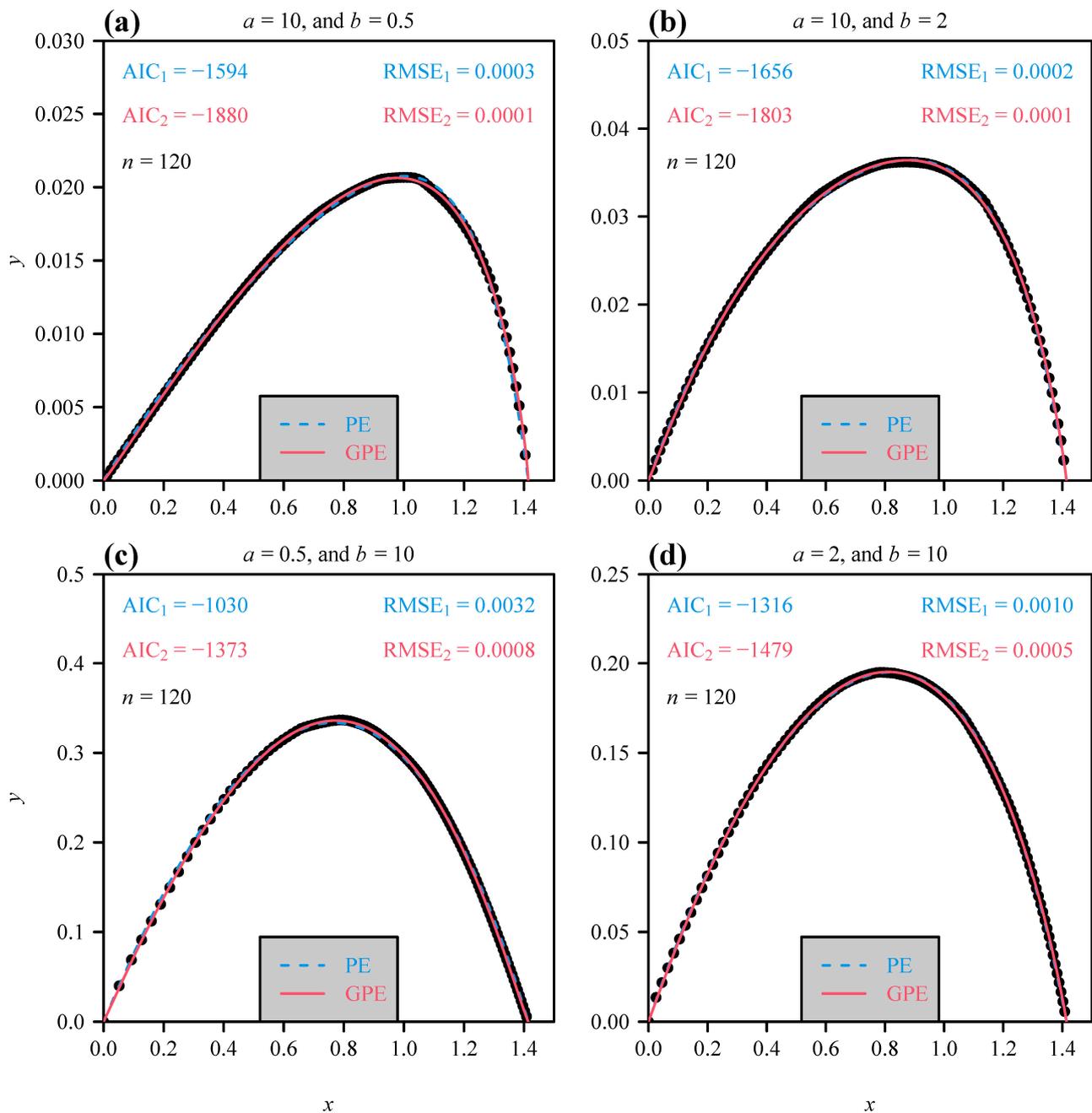


Figure 6. The rotated and right-shifted Lorenz curves (blue and red lines) and the data (closed circles) for the four simulated datasets fitted by the two performance equations. PE is the three-parameter performance equation (i.e., Equation (1)) obtained by fixing the two intersections with the x -axis to 0 and $\sqrt{2}$; GPE is the five-parameter performance equation (i.e., Equation (2)). AIC_1 and AIC_2 are the Akaike information criteria of the estimated PE and GPE, respectively; $RMSE_1$ and $RMSE_2$ are the root-mean-square errors of the estimated PE and GPE, respectively; and n is the sample size of each dataset, which are realizations from the beta distributions with different combinations of parameters a and b . (a) $a = 10$ and $b = 0.5$. (b) $a = 10$ and $b = 2$. (c) $a = 0.5$ and $b = 10$. (d) $a = 2$ and $b = 10$.

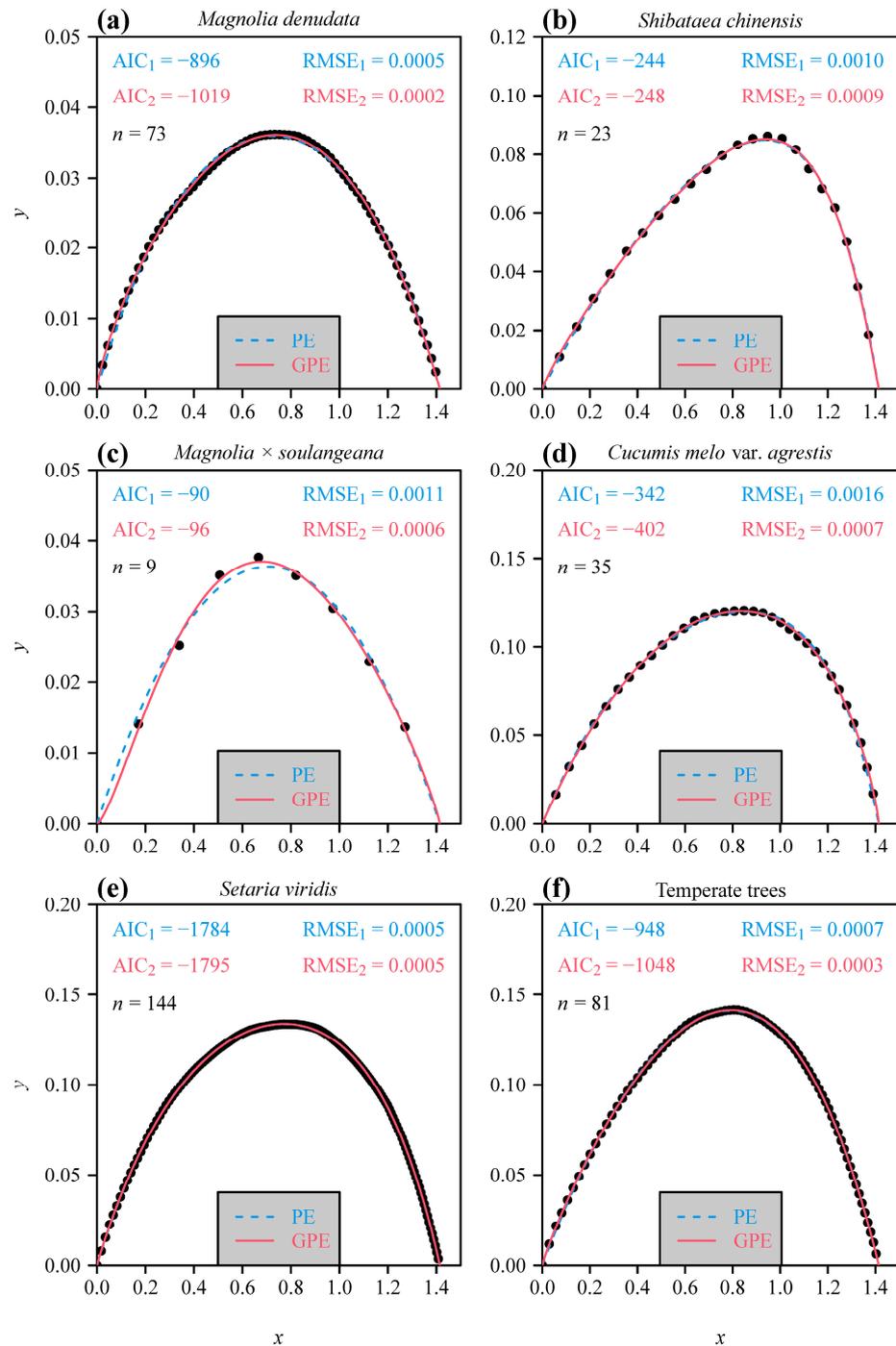


Figure 7. The rotated and right-shifted Lorenz curves (blue and red lines) and data (closed circles) for the six empirical datasets fitted by the two performance equations. PE is the three-parameter performance equation (i.e., Equation (1)) obtained by fixing the two intersections with the *x*-axis to 0 and $\sqrt{2}$; GPE is the five-parameter performance equation (i.e., Equation (2)). AIC₁ and AIC₂ are the Akaike information criteria of the estimated PE and GPE, respectively; RMSE₁ and RMSE₂ are the root-mean-square errors of the estimated PE and GPE, respectively; and *n* is the sample size of each dataset. (a) Stomatal area (i.e., the area of the profile formed by two guard cells) distribution in a 662 $\mu\text{m} \times 444 \mu\text{m}$ micrograph of *Magnolia denudata*. (b) Leaf dry mass distribution of a bamboo (*Shibataea chinensis*) culm. (c) Tepal area distribution of a *Magnolia × soulangeana* flower. (d) Fruit volume distribution of an individual *Cucumis melo var. agrestis* vine. (e) Seedhead length distribution in a 1 m \times 1 m quadrat of *Setaria viridis*. (f) Breast height diameter distribution in a 50 m \times 50 m quadrat of a temperate forest.

Table 1. Results of fitting the five equations to the six empirical datasets.

Species	Indicator	PE	GPE	SarabiaE	SCSE	SHE
<i>Magnolia denudata</i> (<i>n</i> = 73)	RMSE	0.000495	0.000208	0.000339	0.001084	0.001065
	AIC	−896.07	−1018.87	−949.32	−781.64	−782.21
<i>Shibataea chinensis</i> (<i>n</i> = 23)	RMSE	0.001009	0.000854	0.002536	0.002536	0.002410
	AIC	−244.09	−247.73	−199.67	−201.67	−202.02
<i>Magnolia × soulangeana</i> (<i>n</i> = 9)	RMSE	0.001066	0.000614	0.001031	0.002510	0.002529
	AIC	−89.64	−95.58	−88.26	−74.24	−72.10
<i>Cucumis melo</i> var. <i>agrestis</i> (<i>n</i> = 35)	RMSE	0.001629	0.000654	0.000769	0.000778	0.001190
	AIC	−342.05	−401.99	−392.59	−393.76	−362.02
<i>Setaria viridis</i> (<i>n</i> = 144)	RMSE	0.000480	0.000456	0.000972	0.001157	0.001694
	AIC	−1784.25	−1795.15	−1579.08	−1530.83	−1418.96
Temperate trees (<i>n</i> = 81)	RMSE	0.000664	0.000347	0.000446	0.003821	0.003667
	AIC	−947.58	−1048.44	−1010.12	−664.02	−668.68

4. Discussion

The results based on 300 simulated datasets with significantly different degrees of skewness and six empirical datasets of plant (organ) size distributions validate the general approach used in this study and the predictions of the performance equation (PE) and its generalized version (GPE). GPE performed the best, SarabiaE worked the second best, and PE worked the third best; all three were better than the remaining two Lorenz equations, i.e., SCSE and SHE. PE and GPE were tested using 120 random numbers manifesting extremely left-skewed, moderately left-skewed, moderately right-skewed, and extremely right-skewed beta distributions. Both equations fitted the rotated and right-shifted LCs for each simulation. We also tested two super-extreme skewed cases based on a beta distribution function [12] by setting (i) $a = 10$ and $b = 0.1$ (representing a super-extreme left-skewed distribution curve with $S_k = -5.45$), and (ii) $a = 0.1$ and $b = 10$ (representing a super-extreme right-skewed distribution curve with $S_k = 5.45$), which significantly exceeded empirically observed biological size distributions. PE and GPE effectively fitted the rotated and right-shifted LCs of the simulated super-extreme left-skewed distribution curve, which were significantly better than the three other Lorenz equations (not shown due to space limitations). However, the two equations did not fit the simulations of the super-extreme right-skewed distribution curve well; they were significantly worse than SCSE (Figure 8). This latter case simulates a distribution in which most organs are small and approximately equal in size but with several very large organs, which is not generally seen in most biological systems. For this case, the rotated data exhibited an approximate isosceles triangle, which PE and GPE did fit, whereas SCSE and SHE exhibited a better goodness of fit than the two performance equations. Fortunately, a super-extreme right-skewed distribution is rarely observed in real biological size distributions. In contrast, an extreme left-skewed size distribution is not uncommon, e.g., most fruits are large and equal in size but several small stunted fruits exist. Such cases can be effectively described by PE and GPE. Thus, the two performance equations and SarabiaE are applicable for the majority of abiotic and biotic samples, e.g., the distributions of bird egg size in a nest, of fish size in a school, and the energy aftershock releases within a seismic belt in a period. In 300 simulated datasets, 120 random numbers were used as hypothetical biological size values for each dataset. We also checked whether the number of random numbers could influence the results of model comparisons by setting 60, 120, 180, 240, 300, and 360 random numbers, respectively, for each of 300 beta distributions with significantly different skewness values. We found that the numbers of random variables did not influence the results, i.e., GPE ranked the best, SarabiaE ranked the second best, PE ranked the third best, and all three were better than the remaining two Lorenz equations (i.e., SCSE and SHE) in the range of skewness between -2.3 and 2.3 based on the comparisons of RMSEs and AICs (the same as those shown in Figure 4).

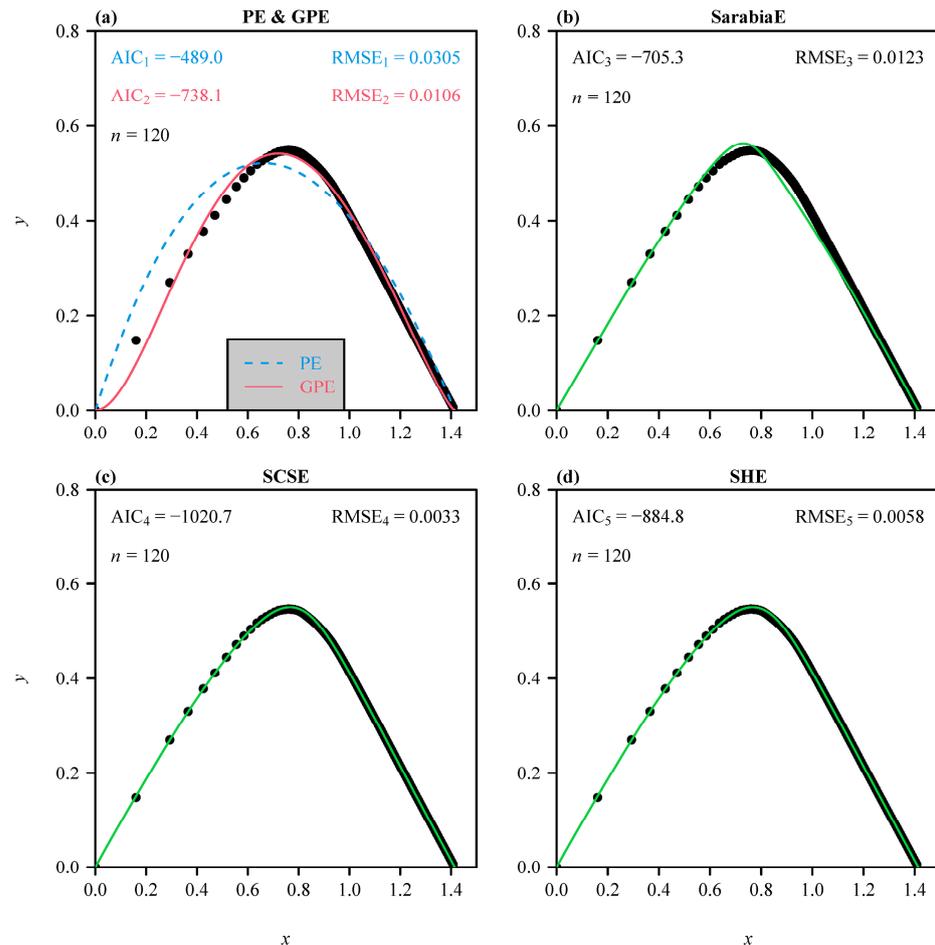


Figure 8. The rotated and right-shifted Lorenz curves (blue, red, and green lines) and simulations (closed circles) from the beta distribution with $a = 0.1$ and $b = 10$ (representing a super-extreme right-skewed distribution curve) using each of the five equations. Panel (a) shows the results of fitting the two performance equations, i.e., PE and GPE; panels (b–d) show the results of fitting the three rotated and right-shifted Lorenz equations, i.e., SarabiaE, SCSE, and SHE. AIC_1 to AIC_5 are the Akaike information criteria of the estimated PE, GPE, SarabiaE, SCSE, and SHE, respectively; $RMSE_1$ to $RMSE_5$ are the root-mean-square errors of the estimated PE, GPE, SarabiaE, SCSE, and SHE, respectively; and n is the sample size of each dataset.

5. Conclusions

Rotated Lorenz curves for biological size distributions are well described by PE and GPE, although GPE provides a better fit and tradeoff between the goodness of fit and the model structural complexity of the six empirical datasets. Both equations also fit simulations with significantly different degrees of skewness (based on beta distribution functions). Therefore, simulated and empirically determined size frequency distributions provide robust evidence that the two performance equations and the Sarabia equation apply broadly to many biotic and abiotic datasets. In conclusion, this approach provides a general protocol for describing size frequency distributions, which is essential to understanding many important ecological and evolutionary phenomena.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/sym16050565/s1>, Table S1: Data of 73 stomata in a $662 \mu\text{m} \times 444 \mu\text{m}$ micrograph of *Magnolia denudata*; Table S2: Data of 23 leaves of a bamboo (*Shibataea chinensis*) culm; Table S3: Data of nine tepals of a *Magnolia \times soulangeana* flower; Table S4: Volume data of 35 fruits of an individual *Cucumis melo* var. *agrestis* plant; Table S5: Length data of 144 seed-

heads of *Setaria viridis* in a 1 m × 1 m quadrat; Table S6: Breast height diameter data of 81 trees with DBH ≥ 1 cm in a 50 m × 50 m quadrat of a temperate forest.

Author Contributions: Formal analysis, P.S., L.D. and K.J.N.; investigation, P.S., L.D. and K.J.N.; writing—original draft preparation, P.S.; writing—review and editing, K.J.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data of plant size and plant organ size distributions are tabulated in the online Supplementary Tables S1–S6.

Acknowledgments: We thank Jie Gao, Johan Gielis, Meng Lian, Jinfeng Wang, Lin Wang, Yujun Wang, Weihao Yao, Kexin Yu, Liuyue Zhang, and Xiao Zheng during the preparation of this work. We also thank three reviewers for their constructive comments on the earlier version of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Lorenz, M.O. Methods of measuring the concentration of wealth. *Am. Stat. Assoc.* **1905**, *9*, 209–219. [CrossRef]
- Metsaranta, J.M.; Loeffers, V.J. Inequality of size and size increment in *Pinus banksiana* in relation to stand dynamics and annual growth rate. *Ann. Bot.* **2008**, *101*, 561–571. [CrossRef] [PubMed]
- Chen, B.J.W.; During, H.J.; Vermeulen, P.J.; Anten, N.P.R. The presence of a below-ground neighbour alters within-plant seed size distribution in *Phaseolus vulgaris*. *Ann. Bot.* **2014**, *114*, 937–943. [CrossRef] [PubMed]
- Gastwirth, J.L. A general definition of the Lorenz curve. *Econometrica* **1971**, *39*, 1037–1039. [CrossRef]
- Gastwirth, J.L. The estimation of the Lorenz curve and Gini index. *Rev. Econ. Stat.* **1972**, *54*, 306–316. [CrossRef]
- McDonald, J.B. Some generalized functions for the size distribution of income. *Econometrica* **1984**, *52*, 647–663. [CrossRef]
- Huey, R.B.; Stevenson, R.D. Integrating thermal physiology and ecology of ectotherms: A discussion of approaches. *Am. Zool.* **1979**, *19*, 357–366. [CrossRef]
- Lian, M.; Shi, P.; Zhang, L.; Yao, W.; Gielis, J.; Niklas, K.J. A generalized performance equation and its application in measuring the Gini index of leaf size inequality. *Trees Struct. Funct.* **2023**, *37*, 1555–1565. [CrossRef]
- Sitthiyot, T.; Holasut, K. A universal model for the Lorenz curve with novel applications for datasets containing zeros and/or exhibiting extreme inequality. *Sci. Rep.* **2023**, *13*, 4729. [CrossRef]
- Sarabia, J.-M. A hierarchy of Lorenz curves based on the generalized Tukey’s lambda distribution. *Econom. Rev.* **1997**, *16*, 305–320. [CrossRef]
- Sarabia, J.-M.; Castillo, E.; Slottje, D.J. An ordered family of Lorenz curves. *J. Econom.* **1999**, *91*, 43–60. [CrossRef]
- Lenth, R.V. Algorithm AS 226: Computing noncentral beta probabilities. *J. R. Stat. Soc. Ser. C. Appl. Statist.* **1987**, *36*, 241–244. [CrossRef]
- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1995; Volume 1.
- Cohen, J.E.; Xu, M. Random sampling of skewed distributions implies Taylor’s power law of fluctuation scaling. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 7749–7754. [CrossRef] [PubMed]
- Yu, K. Quantification of Stomatal Morphology and the Relationship between Stomatal Size and Stomatal Density in 12 *Magnoliaceae* Species. Master’s Thesis, Nanjing Forestry University, Nanjing, China, 2023.
- Zhang, L.; Niklas, K.J.; Niinennets, Ü.; Li, Q.; Yu, K.; Li, J.; Chen, L.; Shi, P. Stomatal area estimation based on stomatal length and width of four *Magnoliaceae* species: Even “kidney”-shaped stomata are not elliptical. *Trees Struct. Funct.* **2023**, *37*, 1333–1342. [CrossRef]
- Shi, P.; Ratkowsky, D.A.; Li, Y.; Zhang, L.; Lin, S.; Gielis, J. A general leaf area geometric formula exists for plants—Evidence from the simplified Gielis equation. *Forests* **2018**, *9*, 714. [CrossRef]
- Su, J.; Niklas, K.J.; Huang, W.; Yu, X.; Yang, Y.; Shi, P. Lamina shape does not correlate with lamina surface area: An analysis based on the simplified Gielis equation. *Glob. Ecol. Conserv.* **2019**, *19*, e00666. [CrossRef]
- Shi, P.; Gielis, J.; Quinn, B.K.; Niklas, K.J.; Ratkowsky, D.A.; Schrader, J.; Ruan, H.; Wang, L.; Niinemets, Ü. ‘biogeo’: An R package for simulating and fitting natural shapes. *Ann. N. Y. Acad. Sci.* **2022**, *1516*, 123–134. [CrossRef]
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024. Available online: <https://www.rproject.org/> (accessed on 1 April 2024).
- Wang, J.; Shi, P.; Yao, W.; Wang, L.; Li, Q.; Tan, R.; Niklas, K.J. The scaling relationship between perianth fresh mass and area: Proof of concept using *Magnolia soulangeana* Soul.-Bod. *Trees Struct. Funct.* **2024**, *38*, 241–249. [CrossRef]
- He, K.; Hui, C.; Yao, W.; Wang, J.; Wang, L.; Li, Q.; Shi, P. Evidence that field muskmelon (*Cucumis melo* L. var. *agrestis* Naud.) fruits are solids of revolution. *Plants* **2023**, *12*, 4186. [CrossRef]
- Shi, P.; Quinn, B.K.; Chen, L.; Gao, J.; Schrader, J. Quantifying α -diversity as a continuous function of location – A case study of a temperate forest. *J. For. Res.* **2023**, *34*, 1683–1691. [CrossRef]

24. Nelder, J.A.; Mead, R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308–313. [[CrossRef](#)]
25. Spiess, A.-N.; Neumeyer, N. An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach. *BMC Pharmacol.* **2010**, *10*, 6. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.