Notulae Botanicae Horti Agrobotanici Cluj-Napoca

An Analysis of the Distribution of Seed Size: A Case Study of the Gymnosperms

Lorentz JÄNTSCHI[1,2], Rodica C. SOBOLU[1], Sorana D. BOLBOACĂ[3*]

[1]*Horticulture and Landscaping Chair, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Mănăștur, 400372 Cluj-Napoca, Romania; lorentz.jantschi@gmail.com*

[2]*Department of Chemistry, Technical University of Cluj-Napoca, 103-105 Muncii Bdv., 400641 Cluj-Napoca, Romania*

[3]*Department of Medical Informatics and Biostatistics, "Iuliu Hațieganu" University of Medicine and Pharmacy Cluj-Napoca, 6 Louis Pasteur, 400349 Cluj-Napoca, Romania; sbolboaca@umfcluj.ro (*corresponding author)*

Abstract

Seed morphology is one of the most addressed issues in seeding plants studies due to its importance in the propagation of seeding plants, which can be related to the influence of the environment of the genetic structure in plant populations. A distribution analysis was conducted on extreme values (minimum and maximum) of seed width and length for gymnosperms spread within the Carpathian Mountains region. Combining the probabilities from independent tests successfully limited the best-fit distribution to a small number of distribution laws. Analyses revealed that the extreme values of investigated seed width and length best fit a log-logistic distribution or one of its generalised forms. The left-weighting of the distribution (to small sizes) revealed a better adaptation of small-sized seeding species. The extreme values of seed dimensions could be used to predict the dimension of a random observation, while the composition of the seeds, which is related to dimension, could provide phylogenetic information.

Keywords: adaptation, Burr distribution, gymnosperms, log-logistic distribution, seed size

Introduction

Several studies have addressed seed morphology as a relevant trait for the prediction of successional trajectories (Grime *et al.*, 1997), the role of herbivores in vegetation succession (Aarrestad *et al.*, 2011), the effect of eutrophication on ecosystems (Kadoya *et al.*, 2011) and for the effect of climate change on species diversity (Floran *et al.*, 2011; Sieck *et al.*, 2011; Walck *et al.*, 2011). Grime and co-authors suggested that investigation of seed morphology might provide a Darwinian underpinning (Grime *et al.*, 1997) for Odum's theory of ecosystem maturity (Odum, 1969). Seed morphology serves as a characteristic for making classifications within sub-generic groups and deducing phylogenetic relationships (Coulter and Chamberlain, 1910).

The documented molecular evidence reliably shows that similarities in plants and their seed morphology can be derived independently without strong phylogeny support (Bowe *et al.*, 2000; Chaw *et al.*, 2011). For example, studies conducted to solve Darwin's mystery regarding the origin of angiosperms led to the conclusion that Gnetales and various fossil groups are related to angiosperms, forming the anthophytes and sustaining the idea that angiosperm origins and homologies should be sought among extinct seed plant groups (Bowe *et al.*, 2000; Chaw *et al.*, 2011).

One important issue often addressed in relation to plant morphology is how the environment can drive and explain the genetic structure in plant populations (Givnish, 2010; Loveless and Hamrick, 1984). For instance, the distribution of seeds can be considered from various points of view: phenology and geographical distribution in relation to seed morphology (Norman, 1994), dispersal of seeds by animals (Nathan *et al.*, 2008; Schupp, 1993), and plant-animal interactions within morphological parameters (Szentesi and Jermy, 1995). The overall conclusion of these studies is that bruchids or other seed predators do not likely drive the evolution of the seed size of the plant species (Nathan *et al.*, 2008; Norman, 1994; Schupp, 1993; Szentesi and Jermy, 1995).

Seed size is a central trait of plant ecology and evolution (Moles *et al.*, 2005a; 2005b), conditioning the probability of seed abundance and dispersal (Guo *et al.*, 2000), predation, germination (Pearson *et al.*, 2002), and seedling survival, even within a single species (Obeso *et al.*, 2011). Additionally, evidence of early plant performance can be found by examining the distribution of seed size (Rodríguez-Calcerrada *et al.*, 2011).

Studies of the link between seed size and growth have been reported since 1908 (Zavits, 1908). For cereals, origination from large seeds led to higher productivity, competitive abilities against weeds, and pests compared to those grown from small seeds (Baalbaki *et al.*, 1997).

The question of whether the living gymnosperms form a clade still remains (Burleigh *et al.*, 2004; Jiao *et al.*, 2011; Rydin *et al.*, 2002), especially as to whether they represent a "natural" group in early classification systems (Schmidt and Schneider-Poetsch, 2002). This question inspired the present research. In this study, extreme values of seed size (minimum and maximum values of both width and length) and the overall distribution from 79 species assigned to the gymnosperms group were analysed. The distribution of seed size among species of gymnosperms is expected to provide further information regarding the origins of the group.

Materials and methods

The measurements of seeds from the Carpathian Mountains region (79 species) were included in the present study. The observed data were obtained from the study by Bojňanský and Fargašová (2007) and are presented in Tab. 1.

All continuous probability density functions available in EasyFit Professional (v 5.2, MathWave Technologies, USA) were used in the analysis. Kolmogorov-Smirnov (Kolmogoroff, 1941; Smirnov, 1948), Anderson-Darling (Anderson and Darling, 1952), and Chi-Squared (Fisher, 1922a, 1924; Pearson, 1900) statistics were applied to measure the agreement between the observations and the model. A global measure of agreement between the observation and the model was calculated for each given probability density function (PDF) using the Fisher's Chi-Squared (abbreviated as F-C-S) formula (Fisher 1948). The global probability of observation of a specific value based on Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Squared (C-S) used is presented in Eq (1):

$$X^2 = \sum_{i=1}^{n} \ln(1/p_i) \sim \chi^2(n, \overline{p}) \qquad (1)$$

where X^2 = value of the Chi-Square statistic; p_i = probability of the ith test; χ^2 = value of the Chi-Squared parameter from the Chi-Square distribution; n = number of tests; and $\overline{p}$ = global probability.

A step-based procedure was applied to reduce the number of likely PDF alternatives:

Step 1: For every list of observations for the 79 species, including minimum and maximum length and width, the series of alternatives were constructed using a maximum likelihood estimate (Fisher, 1922b).

Step 2: The alternative distribution laws that failed to meet the criterion to not be rejected at 20% risk of error by either the Kolmogorov-Smirnov or Anderson-Darling statistics were removed. It is well known that both the Kolmogorov-Smirnov and Anderson-Darling statistics are more frequently susceptible to type I than to type II errors (Jäntschi and Bolboacă, 2009).

Step 3: All alternative distributions that failed to meet the criterion to not be rejected at 5% risk of error by the Chi-Squared statistic were removed from further analysis. This criterion was imposed knowing that the Chi-Squared statistic is more susceptible to type II than to type I errors (Bolboacă *et al.*, 2011; Neyman and Pearson, 1967).

Step 4: The global probabilities of observations were computed using the probabilities given by the Kolmogorov-Smirnov and Chi-Square statistics. As a combined statistic, we assumed that the associated Chi-Squared statistic would not be further exposed to type II errors, and consequently, a filter with 20% risk of error was applied to further reduce the list of alternatives.

Step 5: The four obtained lists were cross-referenced to obtain the distribution law that fit best for all the lists.

Step 6: The procedure described by Fisher (1948) was applied to the revised list, but for eight probabilities in this instance instead of two (to verify the assumption that a given distribution law fits for every one of the four independent sets of observations - maximum and minimum length and width).

Step 7: We next removed from the intersected list all alternatives with negative values (negative-domain). This step was implemented because the investigated character (the seed sizes) could not take on any negative values. It would be inappropriate to apply this step at the beginning of the analysis because in the general case, it is possible to accept a distribution with a negative domain if the probability of its negative values falls within the range of sampling or in the range of computational error.

Results

The summary of the results obtained after applying the first four steps of the analysis approach is presented in Fig. 1.
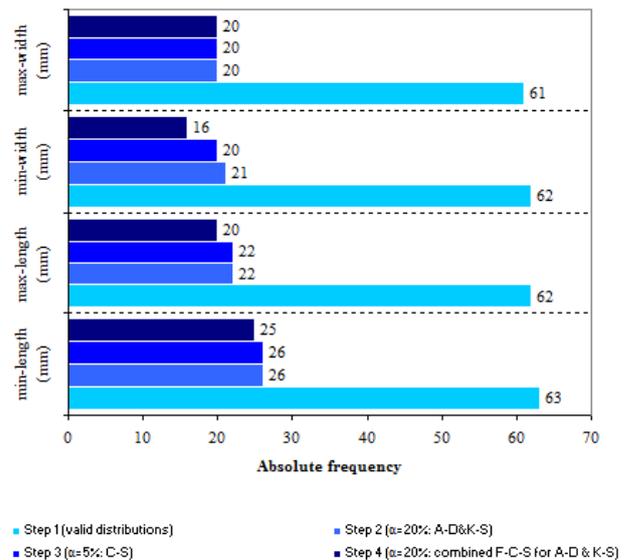

Fig. 1. Reduction of PDF alternatives from steps 1-4 (A-D = Anderson-Darling statistic; K-S = Kolmogorov-Smirnov statistics)

48

Tab. 1. Seed width and length for gymnosperms observed in the Carpathian Mountains region - from Bojňanský and Fargašová (2007)

Morphology Species	Length (mm)		Width (mm)		Morphology Species	Length (mm)		Width (mm)	
	min.	max.	min.	max.		min.	max.	min.	max.
Cycas revoluta	30.0	45.0	25.0	35.0	*Pinus rigida*	3.8	4.6	2.4	3.0
Ginkgo biloba	15.0	20.0	12.0	15.0	*Pinus sabiniana*	17.0	27.0	9.0	12.0
Taxus baccata	6.0	8.0	4.0	5.0	*Pinus sibirica*	5.0	6.0	2.0	3.0
Abies alba	10.0	12.0	5.5	6.5	*Pinus strobus*	6.5	7.5	3.2	3.8
Abies balsamea	5.0	6.0	3.7	4.2	*Pinus sylvestris*	3.0	5.0	2.2	3.5
Abies cephalonica	10.0	13.0	5.0	6.0	*Pinus uncinata*	4.0	6.0	2.0	3.0
Abies concolor	8.0	10.0	5.0	7.0	*Pinus wallichiana*	8.0	10.0	5.0	6.0
Abies grandis	8.0	10.0	5.0	6.0	*Larix decidua*	3.7	4.3	2.2	2.5
Abies nobilis	9.0	12.0	4.0	6.0	*Larix polonica*	3.0	4.0	2.0	3.0
Abies nordmanniana	7.5	8.5	3.8	4.2	*Larix sibirica*	3.0	4.0	2.2	2.4
Abies pinsapo	9.0	12.0	5.5	6.5	*Pseudolarix kaempferi*	4.2	4.8	2.4	2.6
Abies sibirica	8.0	10.0	4.0	6.0	*Chrysolarix amabilis*	7.0	8.5	5.0	5.5
Abies veitchii	5.0	6.0	2.5	3.5	*Cedrus atlantica*	10.0	12.0	5.5	6.5
Pseudotsuga menziesii	5.0	7.0	3.0	4.0	*Cedrus deodara*	12.0	16.0	6.0	7.0
Tsuga canadensis	2.5	4.0	2.5	3.0	*Cedrus libani*	10.0	13.0	4.0	5.0
Picea abies	4.0	5.0	2.0	3.0	*Libocedrus decurrens*	14.0	18.0	6.0	8.0
Picea engelmanii	2.0	3.0	1.5	2.0	*Taxodium distichum*	10.0	12.0	5.5	7.0
Picea glauca	2.6	3.1	1.4	1.7	*Cryptomeria japonica*	5.0	7.0	2.0	3.0
Picea mariana	1.5	2.3	0.8	1.2	*Sequoia sempervirens*	4.4	5.0	2.8	3.2
Picea obovata	3.5	4.0	1.8	2.1	*Sequoiadendron giganteum*	5.0	6.2	4.0	4.6
Picea omorica	3.0	3.6	1.8	2.2	*Metasequoia glyptostroboides*	4.0	5.0	4.0	5.0
Picea orientalis	2.5	3.5	1.8	2.2	*Cupressus sempervirens*	4.0	6.0	3.0	4.0
Picea pungens	3.8	4.2	2.2	2.4	*Chamaecyparis lawsoniana*	4.3	4.7	3.4	4.0
Picea sitchensis	3.3	3.8	1.5	2.0	*Chamaecyparis nootkatensis*	5.0	6.0	4.8	5.2
Pinus banksiana	3.0	5.0	1.5	2.5	*Chamaecyparis obtusa*	3.5	4.0	3.3	3.7
Pinus bungeana	8.0	10.0	5.0	6.5	*Chamaecyparis pisifera*	2.6	3.0	3.8	4.2
Pinus cembra	8.0	12.0	6.0	8.0	*Chamaecyparis thyoides*	2.0	2.5	2.0	2.5
Pinus contorta	4.0	5.0	2.5	3.2	*Thuja occidentalis*	5.8	6.2	2.9	3.1
Pinus flexilis	10.0	12.0	7.5	8.5	*Thuja plicata*	5.3	6.0	2.8	3.2
Pinus halepensis	5.5	6.5	3.0	3.5	*Thujopsis dolabrata*	4.2	4.8	3.0	3.5
Pinus jeffreyi	10.0	12.0	6.0	7.0	*Platycladus orientalis*	5.2	6.2	3.0	3.6
Pinus montezumae	5.0	5.4	3.8	4.2	*Juniperus chinensis*	5.0	5.5	4.0	4.8
Pinus mugo	4.0	4.5	2.5	3.0	*Juniperus communis*	4.0	5.0	2.0	3.0
Pinus nigra	6.8	7.5	3.8	4.3	*Juniperus excelsa*	4.4	5.0	3.0	3.5
Pinus pallasiana	7.0	9.0	4.2	4.6	*Juniperus oxycedrus*	6.0	7.0	4.0	4.5
Pinus peuce	7.5	8.5	5.0	6.0	*Juniperus sabina*	3.7	4.5	2.5	3.0
Pinus pinaster	7.6	8.0	4.2	4.6	*Juniperus sibirica*	4.5	5.0	3.0	3.5
Pinus pinea	15.0	20.0	7.0	11.0	*Juniperus virginiana*	3.5	4.0	2.4	2.8
Pinus ponderosa	7.0	9.0	5.0	6.0	*Ephedrea distachya*	4.8	5.4	2.5	3.0
Pinus radiata	6.0	8.0	3.5	4.5					

The identified alternatives for min-length, max-length, min-width and max-width were intersected (Step 5), and the alternative list was reduced to the following PDFs: Burr (3P), Burr (4P), Dagum (3P), Dagum (4P), Generalised Extreme Value (3P), log-Logistic (2P), log-Logistic (3P), lognormal (3P), Pearson 5 (2P), Pearson 5 (3P), Pearson 6 (3P), and Pearson 6 (4P). No alternatives were removed from the list in the sixth step.

Seven PDFs were removed after encountering negative domains for the sizes of Gymnosperm seeds. The excluded probability distribution functions from Step 7 of the analysis are presented in Tab. 2.

The final list of distribution laws obtained by applying the proposed approach is presented in Tab. 3.

As the results presented in Tab. 3 show, the investigated seed sizes of the gymnosperm group best fit a generalised log-Logistic (3P) distribution. This distribution was plotted for both maximum and minimum values in Fig. 2.

The small dispersal of seed dimensions for both width and length is observed on the plotted distribution (Fig. 2).

Tab. 2. Rejected distribution laws based on negative domains

Distribution	EstMin	Observable - Discussion
Dagum(13.77; 3.62; 3.5592; -2.8574)	-2.8574	min-length - Negative part in the domain of Dagum (4P);
Dagum(2.0446; 2.7581; 3.7359)	0.0000	Dagum (4P) is rejected, and this attracted the rejection of the Dagum (3P) as well;
GenExtreme(0.25224; 2.1148; 4.4014)	-3.9826	min-length - Negative part in the domain of GenExtreme (3P);
Pearson5(3.8847; 18.389)	0.0000	min-length - Negative part in the domain of Pearson5 (3P);
Pearson5(4.5623; 23.984; -0.42052)	-0.4205	Pearson5 (3P) is rejected and also attracted the rejection of the Pearson5 (2P);
Pearson6(38.353; 5.25; 0.45002; -0.1967)	-0.1967	min-width - Negative part in the domain of Pearson6 (4P);
Pearson6(19.021; 5.1948; 0.85351)	0.0000	Pearson6 (4P) is rejected and also attracted the rejection of the Pearson6 (3P)

EstMin = estimated minimum value; Dagum (4P: k, α, β, γ): three continuous shape parameters (k >0, α > 0, and β > 0) and one continuous location parameter (γ); Dagum (3P: k, α, β); GenExtreme (3P: k, σ, μ): Generalised Extreme Value Distribution with shape parameter (k), scale parameter (σ > 0), and location parameter (μ); Pearson5 (3P: α, β, γ): shape parameter (α > 0), scale parameter (β > 0), location parameter (γ, for $\gamma \equiv 0 \rightarrow$ Pearson5 (2P)); Pearson6 (4P: $\alpha1$, $\alpha2$, β, γ): shape parameters ($\alpha1$ > 0, $\alpha2$ > 0), scale parameter (β > 0), location parameter (γ, for $\gamma \equiv 0 \rightarrow$ Pearson6 (3P))

Discussion

The extreme values of seed size (width and length) from 79 species assigned to the gymnosperm group were successfully analysed with reference to probability distribution laws. A seven-step approach was developed and applied to identify the distribution law that best fit the investigated characteristics. In the first four steps, the reduction of alternatives varied from 0% (min-length, 3rd step; max-length, 3rd step; max-width, 3rd step; and max-width, 4th step) to 67% (max-width, 2nd step). The most reductions were observed in the second step relative to the first step, followed by the 4th step. A list of twelve alternatives was ultimately obtained after the four lists were intersected. The negative domains for size of gymnosperm seeds further led to a narrower list of probable distributions (see Tab. 3).

The analysis identified five probable distributions, and four of them were generalizations of the log-logistic distribution. According to the ΣMLE scores (Dey and Kundu, 2010; Holcomb *et al.*, 1999; Nixon and Thompson, 2004), the descending classification of distributions is as follows: log-logistic (3P) - log-normal - log-logistic (2P) - Burn (4P) - Burn (3P). The log-normal distribution was ranked somewhere between two log-logistic distributions, in agreement with the literature Dey and Kundu (2010). The log-logistic (3P) distribution proved to be able to characterise the extreme values of length and width of investigated seeds according to the ΣMLE score criterion. The three-parameter log-logistic distribution is frequently used in models of flood frequency (Ahmad *et al.*, 1988; Hosking and Wallis, 1997; Robson and Reed, 1999) and is related to the modelling of environmental conditions. The relationship between seed size and environmental conditions may facilitate adaptation. Foster and Janson (1985), for example, demonstrated a relationship between large seed size and establishment in shady, stable plant associations. Moreover, Eriksson and Kainulainen showed that the selection for increasing seed size associated with the expansion of modern type tropical forests spurred a competition/colonization trade-off, initiating a reversed evolutionary trajectory towards smaller seeds (Eriksson and Kainulainen, 2011).

The distribution of extreme values may serve for breeding, increasing probabilities of finding seeds of specific size across the species of the genus (Bettge *et al.*, 2002; Cheng *et al.*, 2010; Cilas *et al.*, 2010). Some studies have shown that adaptation is directly linked to seed size, that is, the smaller the better (Howe and Smallwood, 1982). However, in
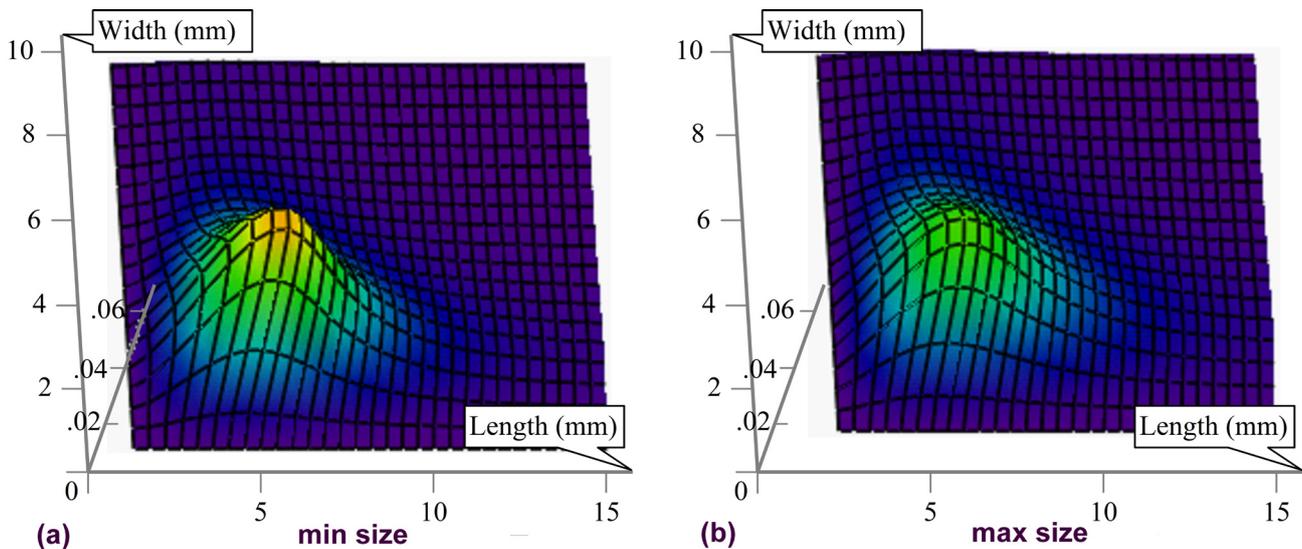

Fig. 2. Seed size distribution (a) minimum values and (b) maximum values

50

Tab. 3. Seed size: probable distribution laws

Observable	Distribution	№P	Probability	$\overline{p}$	MLE	ΣMLE
max-length	Burr(1.0288; 2.1756; 4.4271; 1.9759)		0.59		-207.8	
max-width	Burr(0.6364; 3.7892; 2.8714; 0.5206)	4	0.80	0.948	-265.8	-793.0
min-length	Burr(1.2657; 2.3018; 4.6780; 1.1740)		0.75		-162.5	
min-width	Burr(0.9867; 3.0063; 2.8401; 0.4266)		0.71		-156.9	
max-length	Log-Logistic(2.2083; 4.3587; 1.9602)		0.59		-211.7	
max-width	Log-Logistic(2.7717; 3.1166; 0.8826)	3	0.82	0.929	-170.9	-731.9
min-length	Log-Logistic(2.5976; 4.1923; 1.0288)		0.62		-194.6	
min-width	Log-Logistic(2.9832; 2.8511; 0.4343)		0.71		-154.7	
max-length	Burr(0.3609; 5.5830; 4.4803)		0.70		-510.2	
max-width	Burr(0.4882; 5.0397; 3.1351)	3	0.79	0.906	-351.8	-1387.6
min-length	Burr(0.6078; 4.1014; 4.2994)		0.48		-308.5	
min-width	Burr(0.7443; 3.9066; 2.9328)		0.69		-217.1	
max-length	Log-Normal(0.7504; 1.5518; 1.6985)		0.54		-212.0	
max-width	Log-Normal(0.6329; 1.1999; 0.7416)	3	0.40	0.764	-170.7	-734.2
min-length	Log-Normal(0.6291; 1.5175; 0.7256)		0.68		-195.4	
min-width	Log-Normal(0.5800; 1.1015; 0.3049)		0.58		-156.1	
max-length	Log-Logistic(3.2880; 6.6024)		0.21		-223.5	
max-width	Log-Logistic(3.6494; 4.0681)	2	0.77	0.384	-181.2	-774.1
min-length	Log-Logistic(3.3311; 5.2960)		0.36		-204.3	
min-width	Log-Logistic(3.5263; 3.2758)		0.24		-165.1	

№P = number of parameter associated with the probability distribution function; $\overline{p}$ = global probability;
Burr (4P: k, α, β, γ): three shape parameters (k >0, α > 0, and β > 0) and one location parameter (γ, for γ ≡ 0 → Burr(3P)); LogLogistic (3P: α, β, γ): shape parameter (α > 0), scale parameter (β > 0), location parameter (γ, for γ ≡ 0 → LogLogistic(2P)); Lognormal (3P: σ, μ, γ): two continuous parameters (σ > 0, μ), location parameter (γ)
MLE = Maximum-Likelihood Estimation (MLE = max(Σln(PDFDist(Xi)))); Σ = sum

other cases, seed size maximization is a breeding objective (Damayanti et al., 2010; Dansi et al., 2010; Saxena, 2008). The present study identified the most likely probability distribution function for the extreme values of seed size across Gymnosperms. The probability of finding seeds of a given size could thus be obtained from the probability distribution function of extreme seed size values. Our study showed that the investigated seed sizes of the gymnosperm group most likely fit a generalised log-logistic distribution. It is well known that maximum and minimum values are order statistics and depend on the sample size and study design. The main limit of our study is that the analysis was performed based on the assumption that the sample size was sufficient and that the design utilised to measure seed width and length was reliable and valid (Bojňanský and Fargašová, 2007). The sample's extreme values could be viewed as measures of dispersion when the range is of interest as well as a measure of location when the mid-range is of interest. However, the sample's minimum and maximum values could be used to obtain the prediction interval as estimators for values outside the sample (e.g., for n = 15, a prediction interval of 93% is obtained for the next random observation; the 16th observation falls between the smallest and the largest observation with a chance of 93%) (Whitmore, 1986). However, why are the dimensions of seeds of interest? Seed gymnosperms produce proteins (Jensen and Berthold, 1989; Konarev et al., 2008) that are important characters used in phylogenetic

analysis (Braun et al., 1996; Jensen, 1994). Moreover, the proteins could later be used as a source of food plant proteins, and given a relationship between seed dimensions and composition, it is therefore plausible the results of our study could be relevant.

Conclusions

The minimum and maximum values for seed width and length of investigated gymnosperms best fit a generalised log-logistic distribution. This information was obtained by combining the probabilities from independent tests. The extreme values of seed dimensions could be used to predict the dimension of a random observation.

References

Aarrestad PA, Masunga GS, Hytteborn H, Pitlagano ML, Marokane W, Skarpe C (2011). Influence of soil, tree cover and large herbivores on field layer vegetation along a savanna

landscape gradient in northern Botswana. J Arid Environ 75(3):290-297.

Ahmad MI, Sinclair CD, Werritty A (1988). Log-logistic flood frequency analysis. J Hydrol 98(3-4):205-224.

Anderson TW, Darling DA (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Ann Math Stat 23(2):193-212.

Baalbaki RZ, Copeland LO (1997). Seed size, density and protein content effect on field performance of wheat. Seed Sci Technol 25:511-521.

Bettge AD, Morris CF, DeMacon VL, Kidwell KK (2002). Adaptation of AACC Method 56-11, solvent retention capacity, for use as an early generation selection tool for cultivar development. Cereal Chem 79(5):670-674.

Bojňanský V, Fargašová A (2007). Atlas of seeds and fruits of central and east-European flora. The Carpathian Mountains Region. Springer, Dordrecht, Holland, 33 p. and 1046 p.

Bolboacă SD, Jäntschi L, Sestraş AF, Sestraş RE, Pamfil DC (2011). Pearson-Fisher Chi-Square Statistic Revisited. Information 2(3):528-545.

Bowe LM, Coat G, DePamphilis CW (2000). Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc Natl Acad Sci USA 97(8):4092-4097.

Braun H, Czihal A, Shutov AD, Bäumlein H (1996). A vicilin-like seed protein of cycads: similarity to sucrose-binding proteins. Plant Mol Biol 31(1):35-44.

Burleigh JG, Mathews S (2004). Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. Am J Bot 91(10):1599-1613.

Chaw S-M, Parkinson CL, Cheng Y, Vincent TM, Palmer JD (2000). Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc Natl Acad Sci USA 97(8):4086-4091.

Cheng J, Yan J, Sela H, Manisterski J, Lewinsohn D, Nevo E, Fahima T (2010). Pathogen race determines the type of resistance response in the stripe rust - Triticum dicoccoides pathosystem. Physiol Plantarum 139(3):269-279.

Cilas C, Machado R, Motamayor J-C (2010). Relations between several traits linked to sexual plant reproduction in *Theobroma cacao* L.: Number of ovules per ovary, number of seeds per pod, and seed weight. Tree Genet Genomes 6(2):219-226.

Coulter JM, Chamberlain CJ (1910). Morphology of Gymnosperms. University of Chicago Press: Chicago, US.

Damayanti F, Lawn RJ, Bielig LM (2010). Genotypic variation in domesticated and wild accessions of the tropical tuberous legume *Vigna vexillata* (L.) A. Rich. Crop and Pasture Science 61(10):771-784.

Dansi A, Adoukonou-Sagbadja H, Vodouhè R (2010). Diversity, conservation and related wild species of Fonio millet (Digitaria spp.) in the northwest of Benin. Genet Resour Crop Ev 57(6):827-839.

Dey AK, Kundu D (2010). Discriminating Between the Log-Normal and Log-Logistic Distributions. Commun Stat A-Theor 39(2):280-292.

Eriksson O, Kainulainen K (2011). The evolutionary ecology of dust seeds. Perspect Plant Ecol Evol Syst 13(2):73-87.

Fisher RA (1922a). On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. J Roy Stat Soc 85:87-94.

Fisher RA (1922b). On the Mathematical Foundations of Theoretical Statistics. Philos T Roy Soc A 222:309-368.

Fisher RA (1924). The Conditions under Which χ2 Measures the Discrepancy between Observation and Hypothesis. J Roy Stat Soc 87:442-450.

Fisher RA (1948). Combining independent tests of significance. Am Stat 2(5):30.

Floran V, Sestraş RE, Gil MRG (2011). Organelle genetic diversity and phylogeography of scots pine (*Pinus sylvestris* L.). Not Bot Horti Agrobo 39(1):317-322.

Foster SA, Janson CH (1985). The relationship between seed size and establishment conditions in tropical woody plants. Ecology 66(3):773-780.

Givnish TJ (2010). Ecology of plant speciation. Taxon 59(5):1329-1366.

Grime JP, Thompson K, Hunt R, Hodgson JG, Cornelissen JHC, Rorison IH, Hendry GAF, Ashenden TW, Askew AP, Band SR, Booth RE, Bossard CC, Campbell BD, Cooper JEL, Davison AW, Gupta PL, Hall W, Hand DW, Hannah MA, Hillier SH, Hodkinson DJ, Jalili A, Liu Z, Mackey JML, Matthews N, Mowforth MA, Neal AM, Reader RJ, Reiling K, Ross-Fraser W, Spencer RE, Sutton F, Tasker DE, Thorpe PC, Whitehouse J (1997). Integrated screening validates primary axes of specialisation in plants. Oikos 79(2):259-281.

Guo Q, Brown JH, Valone TJ, Kachman SD (2000). Constraints of seed size on plant distribution and abundance. Ecology 81(8):2149-2155.

Holcomb DL, Smith MA, Ware GO, Hung Y-C, Brackett RE, Doyle MP (1999). Comparison of six dose-response models for use with food-borne pathogens. Risk Anal 19(6):1091-1100.

Hosking JRM, Wallis JR (1997). Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press: Cambridge, Great Britain.

Howe F, Smallwood J (1982). Ecology of seed dispersal. Annu Rev Ecol Syst 13:201-228.

Jäntschi L, Bolboacă SD (2009). Distribution Fitting 2. Pearson-Fisher, Kolmogorov-Smirnov, Anderson-Darling, Wilks-Shapiro, Cramer-von-Misses and Jarque-Bera statistics. Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Horticulture 66(2): 691-697.

Jensen U, Berthold H (1989). Legumin-like proteins in

52

gymnosperms. Phytochemistry 28(5):1389-1394.

Jensen U, Vogel-Bauer I, Nitschke M (1994). Leguminlike Proteins and the Systematics of the Euphorbiaceae. Ann Mo Bot Gard 81(2):160-179.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Hong M, Leebens-Mack J, DePamphilis CW (2011). Ancestral polyploidy in seed plants and angiosperms. Nature 473:97-100.

Kadoya T, Akasaka M, Aoki T, Takamura N (2011). A proposal of framework to obtain an integrated biodiversity indicator for agricultural ponds incorporating the simultaneous effects of multiple pressures. Ecol Indic 11(5):1396-1402.

Kolmogoroff A (1941). Confidence Limits for an Unknown Distribution Function. Ann Math Stat 12(4):461-463.

Konarev AV, Lovegrove A, Shewry PR (2008). Serine proteinase inhibitors in seeds of Cycas siamensis and other gymnosperms. Phytochemistry 69(13):2482-2489.

Loveless MD, Hamrick JL (1984). Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst 15:65-95.

Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Pitman AJ, Westoby M (2005a). Factors that shape seed mass evolution. Proc Nat Acad Sci 102:10540-10544.

Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Westoby M (2005b). A brief history of seed size. Science 307:576-580.

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008). A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci USA 105(49):19052-19059.

Neyman J, Pearson ES (1967). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference, Part I. reprinted at pp. 1-66 in Joint Statistical Papers; Neyman J, Pearson ES (Eds.). Cambridge University Press: Cambridge, UK. (originally published in 1928).

Nixon RM, Thompson SG (2004). Parametric modelling of cost data in medical studies. Stat Med 23(8):1311-1331.

Norman EM (1994). A Re-Examination of Sanango racemosum. 1. Morphology and Distribution. Taxon 43(4):591-600.

Obeso JR, Martínez I, García D (2011). Seed size is heterogeneously distributed among destination habitats in animal dispersed plants. Basic Appl Ecol 12(2):134-140.

Odum EP (1969). The strategy of ecosystem development. Science 164(3877):262-270.

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 5th Ser 50:157-175.

Pearson TRH, Burslem DFRP, Mullins CE, Dalling JW (2002). Germination ecology of neotropical pioneers: Interacting effects of environmental conditions and seed size. Ecology 83(10):2798-2807.

Robson A, Reed D (1999). Statistical Procedures for Flood Frequency Estimation. In: Flood Estimation Handbook, Wallingford, Institute of Hydrology.

Rodríguez-Calcerrada J, Nanos N, Aranda I (2011). The relevance of seed size in modulating leaf physiology and early plant performance in two tree species. Trees-Struct Funct 25(5):873-884.

Rydin C, Källersjö M, Friis EM (2002). Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems and the monophyly of conifers. Int J Plant Sci 163(2):197-214.

Saxena KB (2008). Genetic Improvement of Pigeon Pea - A Review. Trop Plant Biol 1(2):159-178.

Schmidt M, Schneider-Poetsch HAW (2002). The evolution of gymnosperms redrawn by phytochrome genes: The Gnetatae appear at the base of the gymnosperms. J Mol Evol 54(6):715-724.

Schupp EW (1993). Quantity, quality and the effectiveness of seed dispersal by animals. Vegetatio 107-108(1):15-29.

Sieck M, Ibisch PL, Moloney KA, Jeltsch F (2011). Current models broadly neglect specific needs of biodiversity conservation in protected areas under climate change. BMC Ecology 11, art. no. 12, DOI: 10.1186/1472-6785-11-12. Available online: http://www.biomedcentral.com/1472-6785/11/12 (accessed on 17 April 2012).

Smirnov NV (1948). Table for estimating the goodness of fit of empirical distributions. Ann Math Stat 19(2):279-281.

Szentesi Á, Jermy T (1995). Predispersal Seed Predation in Leguminous Species: Seed Morphology and Bruchid Distribution. Oikos 73(1):23-32.

Walck JL, Hidayati SN, Dixon KW, Thompson K, Poschlod P (2011). Climate change and plant regeneration from seed. Global Change Biol 17(6):2145-2161.

Whitmore GA (1986). Prediction Limits for a Univariate Normal Observation. Am Stat 40(2):141-143.

Zavits GD (1908). Relation between the size of seed and the yield of plants in farm crops. Agron J 1:98-104.