African Journal of Agricultural Research Vol. 6(2), pp. 618-623, 4 February, 2011
Available online at http://www.academicjournals.org/AJAR
DOI: 10.5897/AJAR10.777
ISSN 1991-637X ©2011 Academic Journals

Full Length Research Paper

Sunflower *(Helianthus annuus L.)* hybrids seeds distribution modelling: Normal, lognormal and weibull models

Hussein Oraki*, Iraj Alahdadi and Fataneh Parhizkar khajani

Department of Plant and Animal Science, College of Abouraihan, The University of Tehran, Tehran, Iran.

Accepted 9 December, 2010

Sunflower (*Helianthus annus L.*) is one of the most cultivated oil crops in the world. Seed oil of cultivated sunflower has good quality that they are important source of unsaturated oil. Various types of cleaning, grading and separation equipment are designed on the basis of the physical properties of grains or seeds. However knowing the average of seed dimensions are important in designing some apparatuses, but knowing the dimensional distribution is necessary for designing apparatuses such as sorters, separators, and planters. The aim of this study was to model the size distribution of two hybrid of sunflower (hysun33 and hysun25) dimension by three statistic models namely, Normal, Lognormal and Weibull distribution functions. For comparison the goodness of fit of the models, three statistic tests namely, Kolmogrov-Smirnov, Anderson-Darling and Chi-squared tests were used. The results showed that for hysun33 Hybrid, the length, width and thickness of seeds varied from 6.36 to 12.87, 4.02 to 7.57 and 2.65 to 5.52 mm, respectively. Corresponding values for Hysun25 variety were 7.57 to 10.99, 2.54 to 6.82 and 2.03 to 4.93 mm, respectively. According to the results, the normal function was the best model to predict the length, width and thickness distributions of the two hybrids.

Key word: Sunflower seed, physical properties, size distribution, statistic distribution.

INTRODUCTION

Sunflower (*Helianthus annuus* L.) seed is considered to be an important oilseed crop because it contains highly nutritious oil in large quantity (Shukla et al., 1992). Sunflower kernel has 50% oil content, of which 30% is the essential fatty acid, linoleic acid. The Monoun saturated fatty acid, oleic acid that accounts about 10% of the total fat content, make it nutritionally superior than other oilseeds (Earle et al., 1986). The hull comprises 20 to 30% of the seed, depends on the variety, and contains mostly crude fiber and an insignificant quantity of fat (Tranchino et al., 1984). In 1998, the seed world production was about 28.5 million tonnes (FAO, 1999). Physical and engineering properties are important in many problems associated with the design of machines and the analysis of the behavior of the product during agricultural process operations such as handling, planting, harvesting, threshing, cleaning, sorting and drying.

Solutions to problems in these processes involve knowledge of their physical and engineering properties (Irtawange, 2000). Principal axial dimensions of seeds are useful in selecting sieve separators, in calculating grinding power during size reduction. They can also be used to calculate surface area and volume of kernels which are important during modeling of grain drying, aeration, heating, and cooling (Majdi et al., 2006). Maybe knowing the average of seed dimensions is important in designing the apparatuses, but knowing the dimensional seed size distribution is necessary for designing apparatuses such as sorters, separators, and planters. So we were used the statistic distributions for predict the seed size distribution. The normal distribution is applicable to a very wide range of phenomena and is the

*Corresponding author. E-mail: oraki.hussein@gmail.com. Tel: +989163046779.

most widely used distribution in statistics.

The Weibull distribution is one of the important distributions in reliability theory; it is the distribution that received maximum attention in the past few decades. Weibull models have been used in many different applications and for solving a variety of problems from many different disciplines such as yield strength of steel, size distribution of fly ash, fiber strength of Indian cotton,(Weibull, 1951), for predicting the size distribution of Douglas fir (Knowe and Stein, 1995), eastern cottonwood (Knowe et al., 1994), Scots pine (von Gadow, 1984; Sarkkola et al., 2005), loblolly pine (Borders and Patterson, 1990; Cao, 2004; Matney and Sullivan, 1982), jack pine (Bailey and Dell, 1973) and different species mixtures (Siipilehto, 1999; Chen, 2004).

The lognormal distribution is applicable to random variables that are constrained by zero but have a few very large values. The application of a logarithmic transformation to the data can allow the data to be approximated by the symmetrical normal distribution, although the absence of negative values may limit the validity of this procedure. But a review of literature found no studies on application of statistical models in Sunflower seed size distribution. Kolmogrov- Smirnov test is used to decide if a sample comes from a hypothesized continuous distribution. It is based on the empirical cumulative distribution function. The Anderson-Darling procedure is a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. The Chi-Squared test is used to determine if a sample comes from a population with a specific distribution. This test is applied to binned data, so the value of the test statistic depends on how the data is binned.

The objectives of this research were thus to: (i) modeling Sunflowers Hybrid seed dimensions and mass with three statistical distributions namely, Normal, Lognormal and Weibull distributions, (ii) comparison the models and (iii) find the best model.

MATERIALS AND METHODS

Seed samples were gathered from plant improvement institute in Karaj, Iran. In this study, various physical properties of Sunflower seed cultivars such as length, width, thickness, equivalent diameter, geometric diameter, volume, surface area, sphericity, aspect ratio, mass of 1000 seeds and mass of one seed were determined at a moisture content of 10% (wet basis). For measurement; the length, width and thickness, we selected 120 seeds from seed heap randomly and used dial caliper with 0.01 mm accuracy. Mass of 1000 seeds were measured by means of digital balance with 0.01g accuracy and mass of one seed were measured by using the digital balance with 0.001g accuracy. According to the Mohsenin (1970) the geometric mean diameter (D_g) and degree of sphericity (Φ) can be expressed as follows:

$$D_g=(LWT)\frac{1}{3} \quad (1)$$

$$\Phi=\frac{(LWT)^{\frac{1}{3}}}{L} \quad (2)$$

Grain volume (V) and surface area (S) were calculated using (Jain and Bal, 1997).

$$V = 0.25[(\frac{\pi}{6})L(W+T)^2] \quad (3)$$

$$S = \frac{\pi B L^2}{(2L-B)} \quad (4)$$

Where:

$$B = (WT)^{\frac{1}{2}} \quad (5)$$

The aspect ratio (R_a) was calculated by (Maduako and Faborode, 1990).

$$R_a = \frac{W}{L} \quad (6)$$

Where: L, W and T are length, width and thickness of seeds in mm, respectively.

For each of the three seed dimensions for two varieties, location and scale parameters for the normal and lognormal distribution, and shape and scale parameters for the Weibull distribution were estimated by using the Statistic toolbox procedure in Matlab 7.2 (Copyright 1984 to 2005 by The MathWorks, Inc.). The normal and lognormal probability density functions (PDF), used in this study to characterize seed size distribution has the following form of Equations 8 and 9, respectively:

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \quad (7)$$

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left[\frac{-(Lnx-\mu)^2}{2\sigma^2}\right] \quad (8)$$

Where: x is seed dimension, μ is the location parameter ($-\infty<\mu<\infty$) and σ>0 is the scale parameter.

The probability density function (PDF) of the two-parameter Weibull distribution is:

$$f(x|\alpha,\beta) = \beta\alpha^{-\beta}x^{\beta-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] \quad (9)$$

Where: x is seeds dimension (mm), $\beta>0$ is the scale parameter, $\alpha>0$ is the shape parameter.

For lognormal distribution there is no closed form expression for

Table 1. Some physical properties of Sunflower hybrids seed at 10% moisture content (wet basis).

Property	Number of observation	Hysun33					Hysun25		
		Mean	Mean	Max	Min	SD	Max	Min	SD
Length, mm	150	10.41	9.52	10.99	7.57	0.69	8712.	6.36	0.80
Width, mm	150	5.55	4.89	6.82	2.54	0.73	7.57	4.02	0.74
Thickness, mm	150	3.95	3.24	4.93	2.03	0.68	5.52	2.65	0.66
Geometric diameter, mm	150	6.08	5.29	6.84	4.01	0.61	7.52	4.84	0.55
Sphericity	150	0.58	0.55	0.70	0.41	0.05	0.97	0.45	0.06
Volume, mm^3	150	124.92	84.91	173.43	35.80	28.07	230.89	61.17	32.76
Surface area, mm^2	150	99.08	75.71	123.78	45.13	16.80	149.72	62.11	17.56
Aspect ratio	150	0.53	0.51	0.77	0.27	0.07	1.19	0.37	0.09

the cdf:

$F(x|\mu,\sigma)=\int_{-\infty}^{x} f(x|\mu,\sigma)dx$ of x.

The cumulative distribution function (cdf) of the Weibull distribution is:

$$F(x)=1-\exp[(-\frac{x}{\beta})^{\alpha}] \quad (10)$$

Depending on the chosen goodness of fit test, we calculated goodness of fit statistics for each of the fitted distributions. We used three methods for testing the goodness of fit, using the EasyFit5.1 Softwar (EasyFit package, MathWave, Inc.).

The Kolmogorov-Smirnov statistic (KS) is based on the largest vertical difference between the theoretical and the empirical cumulative distribution function:

$$D = \max\left(F(x_i)-\frac{i-1}{n}, \frac{i}{n}-F(xi)\right) \qquad (1\leq i \leq n) \quad (11)$$

Where: x_i is the dimension of *i*th seed and n is the number of test.

The Anderson-Darling statistic (AD) is defined as (Anderson and Darling, 1954):

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1).\,[\mathrm{Ln}\,F(x_i) + \mathrm{Ln}(1 - F(x_{n-i+1}))] \quad (12)$$

Chi-Squared statistic is most probably the oldest method which is being used for goodness of fit or for model discrimination. The Chi-Squared statistic is defined as:

$$\chi^2 = \sum_{i=1}^{k}\frac{(Oi-Ei)^2}{Ei} \quad (13)$$

Where O_i is the observed frequency for bin i, and E_i is the expected frequency for bin i calculated by:

$$E_i=F(x_2)-F(x_1) \quad (14)$$

Where F is the cumulative density function of the probability distribution being tested, and x_1, x_2 are the limits for bin i. The data were analyzed statistically using SAS software preliminary, and goodness of fit test based on the simultaneous 95% confidence band.

RESULTS AND DISCUSSION

The physical properties for Hysun33 and 25 cultivars were studied in this study shown in Table 1. There were clearly differences between the cultivars Hysun23 and Hysun33 for all of their properties (Table 1). The sphericity values of sunflower seeds fall within the range of 0.32 to 1 reported by Mohsenin (1986) for most agricultural materials. The results for sphericity agreed with Mohsenin suggestion. From Table 1, it is obvious that the length, thickness, geometric diameter, volume, surface area and 1000 seeds mass for Hysun33 hybrid greater than Hysun25 hybrid, while there is a little difference between their width, sphericity and aspect ratio. Table 2 indicates estimated parameters for Normal, Lognormal and Weibull distributions, it is obvious that in case of Normal and Lognormal distributions, the greater shape and scale parameters are indicator for greater seed dimensions (Table 2). While in case of Weibull distribution, for Hysun25 length shape parameter is greater than shape parameter for Hysun33 hybrid, while the Hysun25 length smaller than the Hysun33 length (Table 1 and 2).

The summaries of the Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared goodness of fit statistics are displayed in Table 3. For three tests, whatever the goodness of fit statistic is smallest, the models have better fit. Table 3 and Figure 1(a) shows that the best model to predict the Hysun33 length is normal while the results of the two tests (Kolmogrov and Anderson) indicate that the Normal model has best fit to predict the Hysun25 length (Figure 1 b). Furthermore according to the results Kolmogorov-Smirnov and Anderson criteria the best model to predict the Hysun33 width is Normal (Table 3 and Figure 2), but according to the Chi-Squared criterion, Weibull has the best fit to predict the Hysun33 width (Table 3). Thus from table 3 it is obvious that the best model to predict the Hysun25 width is Normal and according to the Kolmogorov-Smirnov, Anderson- Darlingand Chi-Squared criteria the worst model is weibull and according to the Kolmogorov-Smirnov criterion Lognormal model has the best fit to predict the Hysun25

Table 2. Estimated parameters for different distribution functions for two varieties.

Property	Distribution	Hysun33		Hysun25	
		Shape	Scale	Shape	Scale
Length	Normal	10.42	0.8	9.53	0.69
	Lognormal	2.34	0.08	2.25	0.07
	Weibull	14.96	10.77	16.75	9.82
Width	Normal	5.55	0.74	4.89	0.73
	Lognormal	1.7	0.13	1.58	0.15
	Weibull	9.02	5.83	7.92	5.18
Thickness	Normal	3.95	0.67	3.24	0.69
	Lognormal	1.36	0.17	1.15	0.21
	Weibull	7.00	4.20	5.65	3.50

Table 3. Kolmogorov-smirnov, Anderson-Darling and Chi-Squared criterion for comparison the models.

Property	Distribution	Hysun33			Hysun25		
		Anderson	Kolmogrov	Chi-squared	Anderson	Kolmogrov	Chi-squared
Length	Normal	0.089	1.60	17.10	0.039	0.23	3.75
	Lognormal	0.106	2.29	26.20	0.045	0.29	3.25
	Weibull	0.109	2.34	30.20	0.079	1.60	4.69
Width	Normal	0.056	0.75	15.90	0.066	0.52	7.33
	Lognormal	0.081	1.09	10.59	0.062	0.71	11.00
	Weibull	0.073	1.69	9.64	0.095	1.19	20.22
Thickness	Normal	0.062	0.90	6.33	0.058	0.83	18.67
	Normal	0.089	1.60	17.10	0.039	0.23	3.75
	Lognormal	0.106	2.29	26.20	0.045	0.29	3.25

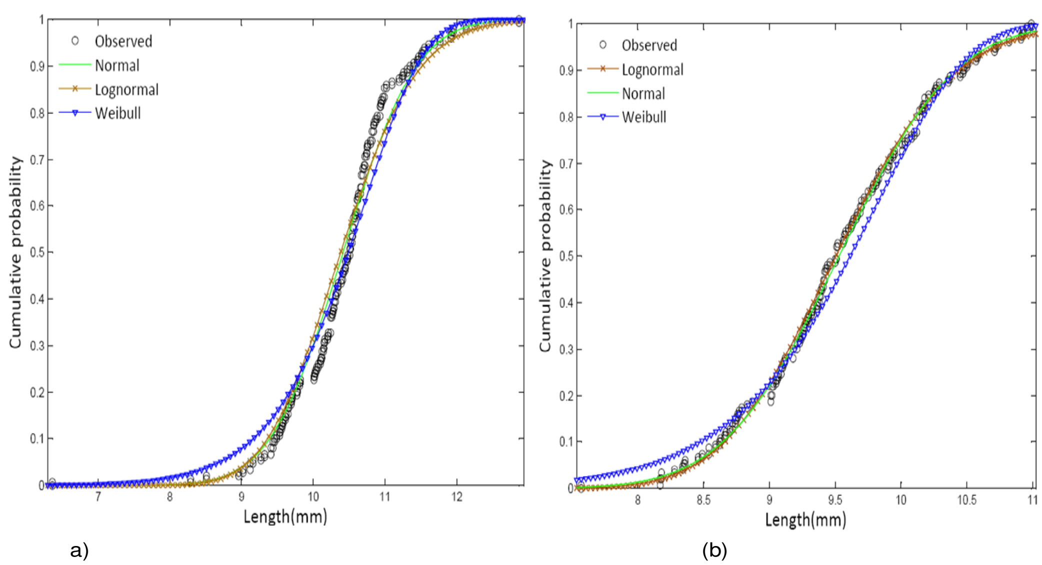
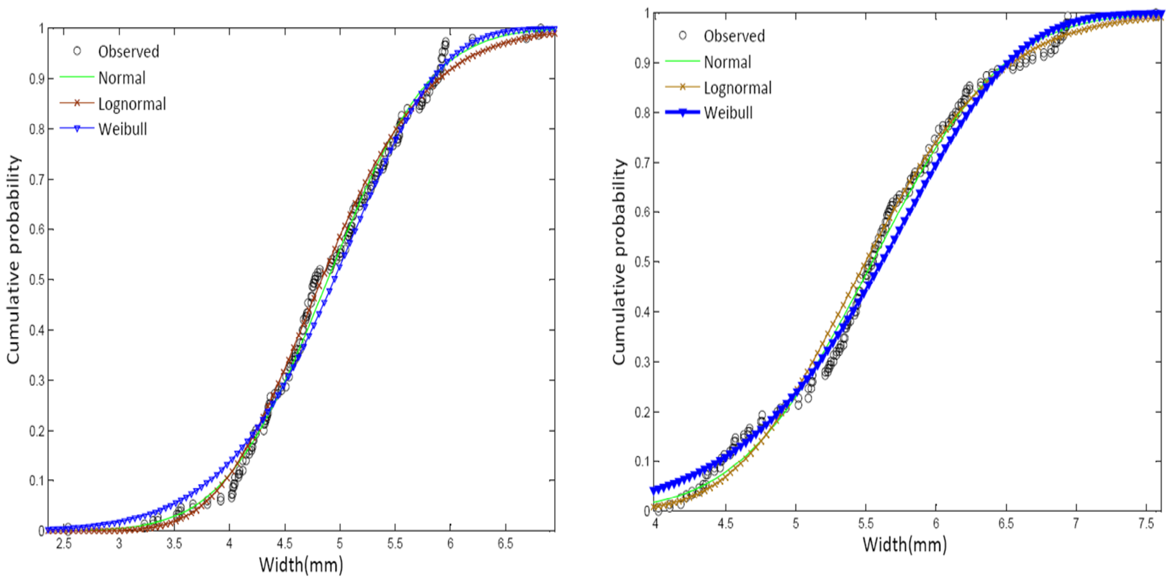
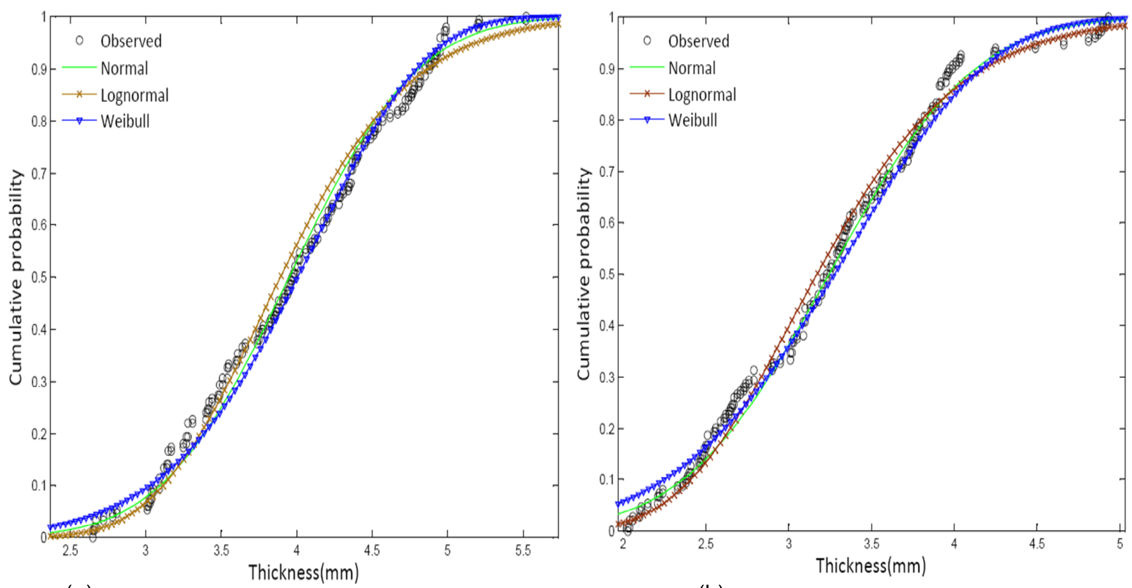


Figure 1. Cumulative probability functions for Normal, Lognormal and Weibull distributions: a) Hysun33 length and b) Hysun25 length.



a) b)

Figure 2. Cumulative probability functions for Normal, Lognormal and Weibull distributions: a) Hysun33 width and b) Hysun25 width.



(a) (b)

Figure 3. Cumulative probability functions for Normal, Lognormal and Weibull distributions: a) Hysun33 thickness and b) Hysun25.

width (Table 3). In case of Hysun33 thickness, on the base of results of three tests, the Lognormal have the best and result of two tests (Kolmogorov and Chi-Squared) Weibull models have the worst fit (Table 3 and Figure 3). To predict the Hysun25 thickness, according to the results of two tests, the Normal models have the best and according to Chi-Squared test Weibull criterion have the worst fit (Table 3 and Figure 3).

Conclusions

The ability to predict grain and seed size distribution by suitable statistical distributions is important for designing the various agricultural equipments and machines. In this study Normal, Lognormal and Weibull distributions were fitted to observed seed dimension of two Sunflower hybrids seeds. Base on the three statistical tests namely, Kolmogorov-Smirnov, Anderson-Darling and Chi-Squired, to predict the length, width and thickness of Hysun33 hybrid, Normal was the best model while in case of Hysun25 hybrid, the fitted Normal provided a closer fit to the data of length, and to predict the Hysun25 width, base on the Anderson-Darling and Chi-Squired test Normal is the best model and according to the Kolmogorov-Smirnov test Lognormal is the best model, and according to the two statistical tests (Kolmogorov-Smirnov and Anderson-Darling), Normal is the best model to predict the thickness of Hysun25 hybrid and result of the chi-squired test Weibull is the best model to predict the Hysun25 thickness.

Cumulative probability density function plots are applicable to find the best fitted model to observed data.

REFERENCES

Anderson TW, Darling DA (1954). A test of goodness of fit. J. Am. Stat. Assoc., 49: 765-769.

Bailey RL, Dell TR (1973). Quantifying diameter distributions with the Weibull function. For. Sci., 19, 97-04.

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. For. Sci., 36: 413-424.

Cao QV (2004). Predicting parameters of a Weibull function for modeling diameter distribution. For. Sci., 50: 682-685.

Earle FR, Vanetten CH, Clark TF, Wolff IA (1986). Compositional data on sunflower seed. J. Am. Oil Chemist Soc., 45: 876-879.

FAO (1999). Sunflower. http//www.fao.org.

Irtawange SV (2000). The effect of accession on some physical and engineering properties of African yam bean. Unpublished Ph.D. Thesis, Department of Agricultural Engineering, University of Ibadan, Nigeria.

Jain RK, Bal S (1997). Properties of pearl millrt. J. Agric. Engine. Res., 66: 85-91.

Knowe SA, Foster GS, Roussau RJ, Nance WL (1994). Eastern cottonwood clonal mixing study: predicted diameter distributions. Can. J. For. Res., 24: 405-414.

Knowe SA, Stein WI (1995). Predicting the effects of site preparation and protection on development of young Douglas-fir plantations. Can. J. For. Res., 25: 1538–1547.

Maduako JN, Faborode MO (1990). Some physical properties of cocoa pods in relation to primary processing. Ife J, Technol: 2: 1–7.

Majdi A, Al-Mahasneh T, Rababah M (2006). Effect of moisture content on some physical properties of green wheat. J. food Engine., 79: 1467-1473.

Mohsenin NN (1986). Physical Properties of Plant and Animal Materials, second ed. Gordon and Breach Science Publishers, New York.

Siipilehto J (1999). Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fennica, 33: 281-301.

Shukla BD, Srivastava PK, Gupta RK (1992). Oilseed Processing Technology. Central Institute of Agricultural Engineering Publications, Bhopal, India.

Tranchino L, Melle F, Sodini G (1984). Almost complete dehulling of high oil sunflower seed. J. Am. Oil Chemists Soc., 61: 1261-1265.

von Gadow, K (1984). Die Erfassung von Durchmesserverteilungen in gleichaltrigen Kiefernbesta¨nden. Forstwissenschaftliches Centralblatt, 103: 360-374.

Weibull W (1951). A statistical distribution function of wide applicability. J. Appl. Mech., 18: 293-297.