Europ. J. Agronomy 68 (2015) 89–96

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Seed yield prediction of sesame using artificial neural network

Samad Emamgholizadeh[a,*], M. Parsaeian[b], Mehdi Baradaran[b]

[a] Associate professor, Department of Water and Soil Engineering, Shahrood University, Shahrood, Iran
[b] Assistant Professor, Department of Agronomy and Plant Breeding, Shahrood University, Shahrood, Iran

ARTICLE INFO

Article history:
Received 29 September 2014
Received in revised form 23 April 2015
Accepted 30 April 2015
Available online 15 May 2015

Keywords:
Seed yield estimation
Sesame
Artificial neural networks
Multiple regression model

ABSTRACT

The prediction of seed yield is one of the most important breeding objectives in agricultural research. So, in this study, two methods namely artificial neural network (ANN) and multiple regression model (MLR) were employed to estimate the seed yield of sesame (SYS) from readily measurable plant characters (e.g., flowering time of 100% (days), the plant height (cm), the capsule number per plant, the 1000-seed weight (g) and the seed number per capsule). The ANN and MLR were tested using field data. Results showed that the ANN predicts the SYS accurately with a root-mean-square-error (RMSE) of 0.339 t/ha and a determination coefficient (R^2) of 0.901. Also, it was found that the ANN model performed better than the MLR model with a RMSE of 0.346 t/ha, and R^2 of 0.779. Finally, sensitivity analysis was conducted to determine the most and the least influential characters affecting SYS. It was found that the capsule number per plant and the flowering time of 100% had the most and least significant effects on SYS, respectively.

1. Introduction

Sesame (*Sesamum indicum* L.) is one of the oldest oil seed crops and is widely grown in tropical and subtropical areas of Asia (Ashri, 1998). There is no consensus among researchers regarding the center of origin of the cultivated sesame and it seems that it is a crop with more than one origin. The primary center of sesame origin has been predicted to be Fertile Crescent, the Indian subcontinent or the Iran–Afghanistan region (Ashri, 1989). Also, the wild species of this plant can be found in Africa. Nowadays, our knowledge about the nutritional value and health benefits of sesame, has increased the universal demand for its seed and oil. The superior oil quality of sesame, which is ascribed by the low level of saturated fatty acids as well as the activity of unique natural antioxidants together with tocopherols make it distinctive among other oil seed crops. However, sesame has a low ranking in the world production of edible oil seeds which is mainly attributed to low yielding varieties with an indeterminate growth habit, uneven ripening of capsules, seed shattering, susceptibility to environmental stresses and the lack of adequate researches in this crop (Bhat et al., 1999). The increase of seed yield is one of the most important breeding objectives in all crop plants. Seed yield is a quantitative polygenic and complex trait and also, it is a resultant of different factors. Its phenotypic expression is predominantly affected by the environmental conditions; hence it has a low heritability. Therefore, the response to the direct selection for this character could lead to a low profit. In contrast, the selection of high-heritable characters associated with seed yield is good prospect for increasing its performance. Seed yield components not only directly affect the seed yield but also indirectly affect the performance through negative or positive ways (Solanki and Gupta, 2001; Khan et al., 2007; Ibrahim and Khidir, 2012).

A number of studies have tried to relate the seed yield of sesame to its components and morpho-phenological properties of plant using multivariate analysis such as multiple linear regressions (MLR), path analysis (PA), factor analysis (FA) and other techniques. For example Ganesh and Sakila (1999) and Ibrahim and Khidir (2012) used simple phenotypic or genotype correlations to find the association of seed yield and its contributing characters in sesame. They found that seed yield was positively correlated with the capsule number per plant, plant height, 1000- seed weight and days to maturity. Parimala and Mathur (2006) were employed multiple linear regression (MLR) to establish a model between seed yield of sesame and its component such as capsules per plant, the capsule length, number of branches, the plant height, the number of seeds per capsule, the 1000-seed weight and days to first flower. Their findings indicated that the number of capsules per plant was the most important character and could explain the high amount of the total variation of seed yield in sesame. Based on the predicted equation for grain yield Shim et al. (2006) found that the plant height and number of capsules per plant had significant contribution to sesame grain yield under the early sowing condition, while there

* Corresponding author. Tel.: +98 9111194389; fax: +98 234 5224620.
E-mail address: s.gholizadeh517@yahoo.com (S. Emamgholizadeh).

http://dx.doi.org/10.1016/j.eja.2015.04.010
1161-0301/

was no significant effect for 1000-seed weight. Furthermore, path coefficient analysis (PCA) has been widely used in sesame breeding programs to interpret the nature of relationships between seed yield and yield-determining traits (Yingzhong and Yishou, 2002; Shim et al., 2001; Mothilal, 2005; Kurdistani et al., 2011; Azeez and Morakinyo, 2011). Ganesh and Sakila (1999) also studied path analysis for some parental genotypes and their progeny in sesame and reported that the traits of number of capsules per plant, plant height and number of branches per plant had a high positive direct effect on the plant yield in sesame. Also the literature review shows that many researchers such as Shim et al. (2006) and Ashraf (2013) have used MLR to predict the seed yield of sesame.

The main shortcoming of regression-based models is that they cannot capture the highly nonlinear and complex relationship between seed yield and the relevant crop plant properties. In contrast to the traditional methods such MLR and PA models, the application of artificial intelligence (AI) models such as artificial neural networks (ANN), genetic expression programming (GEP) and adaptive neuro-fuzzy inference system (ANFIS) were recently attracted the attention of researchers in agriculture science (Azamathulla and Ghani, 2011; Shahinfar et al., 2012; Emamgholizadeh et al., 2013a,b; Samadianfard et al., 2014; Silva et al., 2014; Iquebal et al., 2014). Alvarez (2007) used the ANN approach to predict average regional yield and production of wheat in the Argentine Pampas. Other researchers applied the ANN as a feasible alternative method for the discrimination and identification of *Camellia japonicaCamellia japonica* L. (Mugnai et al., 2008) and tea genotypes (Pandolfi et al., 2009). Yong-Jun et al. (2011) also reiterated on the application of artificial neural network in genomic selection for crop improvement.

Barbosa et al. (2011) used the ANN to study the genetic diversity analysis of *Carica papaya* based on the model proposed by Kohonen. Zaefizadeh et al. (2011) also compared two methods of multiple linear regressions (MLR) and the artificial neural network (ANN) in predicting the yield using its components in the hulless barley. Other researchers applied the ANN as a feasible alternative method for the discrimination and identification of *C. japonica* L. (Mugnai et al., 2008) and tea genotypes (Pandolfi et al., 2009). Yong-Jun et al. (2011) utilized ANN in genomic selection for crop improvement.

The objective of this study is to develop and test the ANN model to predict the seed yield of sesame (SYS). It is worth noting that this is the first study which explores the potential ability of the ANN to model the seed yield of sesame.

2. Materials and methods

2.1. Field experiments

Nine diverse genotypes of sesame including five genotypes representing different sesame growing regions of Iran (Varamin 2822, Yekta, Darab1, TN234 and TN240) and four Asian genotypes, belonging to Pakistan, Indian, Turkey and Iraq (Pakistani, Indian, EM and Iraq 22) were crossed in a full-diallel mating design. Iranian genotypes were produced by selection within different Iranian landraces. The field experiment was carried out at the research farm of Isfahan University of Technology (32°2′N and 51°32′E, 1630 m as 1) during the two years of 2007–2008 with a cropping season on a Typic Haplargid soil with clay loam texture, pH 7.5 and organic matter content of 1%. All parents, 81 F_1s (progenies of first filial), and 45 F_2s (progenies of second filial, without reciprocals) were agro- morphologically evaluated using a randomized complete block design with three replications. Each plot consisted of four rows 1.5 m in length and 50 cm row to row spacing with inter-row plant distance of 7 cm. Fertilizer were applied at 80 kg N/ha and 100 kg P/ha prior to sowing and 40 kg N/ha top dressed four weeks after planting. Days to flowering and days to maturity were visually recorded on plot basis. The plant height, the height to the first fruiting node, the number of fruiting branches per plant, the capsule number per plant, the seed number per capsule, the 1000-seedweight and seed yield per plant were recorded using 10 (F_1 experiment) and 30 (F_2 experiment) randomly selected plants from each plot and their average were used. Two middle rows of each plot were harvested to determine the seed yield.

2.2. Statistical analysis

Data were subjected to analysis of variance (ANOVA) using general linear model of SAS statistical program (2010). The association of agro-morphological traits was evaluated by computing the Pearson correlation coefficients among all combinations of characters using SAS statistical package. The ANN was carried out by Qnet 2000 model using seed yield as the dependent variable and the remaining traits as independent variables. For training and testing the ANN model, 378 samples were collected from the study area. To have an overview of the measured variables (i.e., flowering time (10% and 100%), seed maturity, plant height, capsule number per plant, 1000-seed weight, seed number per capsule, plant height to the first fruiting branch, plant height to the first fruiting node, capsule number per branch and branch number), the statistical indices are shown in Table 1.

2.3. Artificial neural network (ANN)

Artificial neural network (ANN) is an intelligence model and it imitates the procedure of the human brain works (Tufail et al., 2008). In other words, the ANN has certain performance characteristics like biological neural networks of the human brain (Haykin, 1994). This computational technique was first introduced in 1943 (McCulloch and Pitts, 1943). A typical ANN consists of a number of simple processing elements called neurons or nodes. These neurons are organized into groups termed layers. Each neuron is connected to other neurons by means of direct communication links, each with

Table 1
Statistical indices of measured data.

Variables	Synonym	Min	Max	Mean	Std. deviation
Flowering time of 10% (days)	FT10	40.00	55.00	47.13	2.55
Flowering time of 100% (days)	FT100	40.00	60.00	52.27	2.69
Seed maturity (days)	SM	105.00	151.00	129.82	13.11
Plant height (cm)	PH	100.70	200.40	144.95	16.96
Capsule number per plant	CNPP	24.80	100.50	51.92	12.44
1000-seed weight (g)	TSW	1.99	4.20	3.40	0.37
Seed number per capsule	SNPC	27.58	100.48	55.56	14.17
Plant height to first fruiting branch (cm)	PHFFB	26.10	88.90	60.68	10.69
Plant height to first fruiting node (cm)	PHFFN	0.00	5.20	1.81	0.92
Capsule number per branch	CNPB	0.00	62.80	30.78	14.87
Branch number	BN	0.00	18.20	8.33	4.22
Seed yield of sesame (t/ha)	SYS	1.24	5.04	2.76	0.71