STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, XLV, 2, TOM II, 2010

Dedicated to the memory of Prof. dr. Ioan Silaghi-Dumitrescu marking 60 years from his birth

ARTIFICIAL NEURAL NETWORKS USED FOR INVESTIGATION OF FATTY ACID CONTENT OF ROMANIAN SUNFLOWER OILSEEDS GENOTYPES

DORINA BRĂTFĂLEAN, VASILE MIRCEA CRISTEA*, PAUL ŞERBAN AGACHI

ABSTRACT. The objective of this study was to investigate the efficiency of Artificial Neural Networks (ANNs) in classifying and predicting the fatty acid content from Romanian sunflower oilseeds genotypes, as solutions of computational engineering problems. The two-layer probabilistic ANN, using a radial basis layer and a competitive layer, has been used for classification. There were two criteria of classification, the degree of polyunsaturation and the linoleic/oleic acid ratio, which allowed the defining of two categories. The first ANN has been designed for classifying the first category into three groups defined by the polyunsaturated fatty acid content: group 1 of less than 40% polyunsaturated fatty acid, group 2 of 40%-50% polyunsaturated fatty acid, and group 3 of higher than 50% polyunsaturated fatty acid. The classification was based on the following acids in the samples: C14:00, C15:00, C16:00, C16:01, C17:00, C18:00, C18:01, C18:02, C18:03, C20:00, C20:01 and C22:00. The second designed ANN has been used for classifying the category of linoleic/oleic acid ratio into three groups: group 1 of linoleic/oleic acid ratio higher than 2, group 2 of linoleic/oleic acid ratio between 1 and 2 and group 3 of linoleic/oleic acid ratio less than 1. The results of both classifications revealed a good accuracy of the trained ANNs for classifying the sunflower oilseeds. The numerical tests demonstrated the computational advantages of the prediction methodology.

Keywords: *ANN, fatty acids, oilseeds, polyunsaturated fatty acid, sunflower genotypes.*

INTRODUCTION

Fats are a subclass of lipids, but the term “fat” is often used instead of the term denoting “lipid”. For nutrition labeling purposes, fat has been defined as triglycerides (substances extracted with organic solvent) or total lipids. Triglycerides are the building blocks of fats and oils, such as sunflower oil, which is extracted from sunflower seeds and is the most common vegetable fat.

* *Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11. Arany Janos, RO-400028 Cluj-Napoca, Romania, mcristea@chem.ubbcluj.ro*

From the economical point of view sunflower oil is a very good source for valorisation due to the variety of fatty esters contained. Fatty acids are differentiated by their molecular composition [1]. One differentiating characteristic is the degree of saturation, according to which they are: saturated, monounsaturated or polyunsaturated. Other differentiating characteristics are the chain length and the number of carbon atoms in the fatty acid molecule. Fatty acids represent 95 of every 100 grams of fat or oil. Fatty acid variability and profile in the sunflower oil depend on the sunflower genotype. They found numerous applications as food (human and animal feeds), and non-food products (pharmaceutical field, cosmetics, bio-detergents, biodiesel, etc).

Due to their many applications, it is important to be able to assess the quality of different oilseeds. This may be performed on the basis of the fatty acids content. This is an important reason for searching and developing new methodologies that allow more flexible and better-controlled means for the classification of fatty acid content of fats.

The quality of food or other products is monitored based on their composition through laboratory analyses, while computational screening may be used for prediction and simulation, thus offering economic benefits [2]. Methodologies based on Artificial Neural Networks (ANNs) have become an efficient tool in many studies involving property classification and prediction.

Modelling and data mining methodologies based on ANNs are able to represent information on complex systems [3]. ANNs are adaptive systems that change their structure based on external or internal information that flows through the network. Neural networks are non-linear statistical data modelling tools [4]. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

The aim of the present work has been the comparative investigation of fatty acid content for different Romanian sunflower genotypes, based on ANNs classification aptitude. Several genotypes of Romanian sunflower oilseeds have been considered in the investigation of the fatty acid content, using laboratory analytical techniques. The fatty acid methyl esters (FAMEs) were analyzed by gas-chromatography, and the results were used to test the ability of the specially trained ANNs for predicting the classis of genotypes, according to the type of their fatty acid contained.

Samples

The Romanian National Agricultural Research and Development Institute Fundulea provided the sunflower seeds used in this study. Before they were analyzed, the sunflower seeds were stored at a temperature of 5°C and low humidity [5]. First, the dry weight determination of each sample was performed in the laboratory. Around 10g of sunflower seeds at harvesting

maturity were sampled for each sunflower genotype [6]. Ten inbred lines and five Romanian hybrids have been considered in this study. The genotypes are presented in Table 1.

Table1. Investigated inbred lines and Romanian sunflower hybrids

Genotype	LC-L1	LC-L2	LC-L3	LC-L4	LC-L5	LC-L6	LC-L7
Type of variety	Semi-precocious	Semi-precocious	Preco-cious	Tardif	Semi-precocious	Semi-Tardif	Semi-Tardif

enotype	LC-L8	LC-L9	LC-L10	R-H1	R-H2	R-H3	R-H4	R-H5
Type of variety	Semi-precocious	Precocious	Tardif	Alex	Favorit	Rapid	Top75	Turbo

The samples were collected for two types of laboratory analysis, qualitative and quantitative, and one computational method. The first method (quantitative) was the detection of the percent of fat from oilseeds by Accelerated Solvent Extractor (ASE) method, and the second method (qualitative) was gas the chromatograph analysis. The computational method was based on the ANNs classification capacity, using the Matlab software environment and its accompanying Neural Network Toolbox.

Procedure of total lipids extraction

Extraction of the total lipids from the sunflower seeds has been the initial step of this study. The ASE method was used according to Official Method of Analysis, 1999 [6]. This method has a similar operating principle as the Soxhlet Extraction method used for lipids extraction. ASE method involves the gravimetric determination of the oil using the light petroleum extract from sunflower seeds. The equipment consists in a metallic cell having six commune broiler based extractors. The petroleum extract is called “oil content”. Following the weighting step, the seed material was placed in a cellulose extraction cartridge and then introduced into the extractor. The parameters of the extraction process and the number of cycles were programmed by means of an electronic interface [7]. The advantage of the ASE method consists in the relatively high number of samples which can be investigated and the reduced time of extraction (6 samples and 30 min of extraction for one sample). The fat content has been computed by the direct expression:

$$Fat\ content[\%] = \frac{Mass\ of\ the\ extract}{Mass\ of\ the\ sample} \cdot 100 \quad (1)$$

The ASE method was used to determine quantitatively the fat from the sunflower oilseed.

Fatty acid methyl esters (FAMEs) preparation

FAMEs were produced in order to determine the content of total fatty acids in the extracted oils [6]. The first step for FAMEs preparation consists in the saponification procedure [2]. Two samples of 0.1g lipid extracted from each sunflower genotype were treated with 1 mL NaOH solution (1M) in methanol by refluxing for 1h at 100°C. After free fatty acid removal, the samples were esterified in the presence of H_2SO_4 as catalyst. The FAMEs were extracted with petroleum ether from the salt saturated mixture. After FAMEs drying, the Na_2SO_4 anhydride was added and the supernatant was poured in the specific cell.

Gas chromatography system

The fatty acids have been analyzed by gas chromatography (GC) according to the published methods [6]. The esters were separated by gas chromatograph (Carlo Erba model equipped with FID type FRATOVAP 300, and split/splitless injector and associated column (L=2m, d=3mm), containing 10% diethylene glycol succinat on Chromosorb W, AW). The temperature of the injection port was maintained at constant value of 190°C and the detector temperature at 225°C. The oven temperature was programmed to increase from 80°C to 200°C at a rate of 5°C /min using Argon as carrier gas, fed with a reference flow rate of 24mL/min. The peaks were identified based on their retention times using standard fatty acids methyl esters. All samples were run in duplicate.

Identification of FAMEs

The total amount of Miristic acid-C14:00, Pentadecanoic acid-C15:00, Palmitic acid-C16:00, Palmtoleic acid-C16:01, Heptadecanoic-C17:00, Stearic acid-C18:00, Oleic acid-C18:01, Linoleic acid-C18:02, Linolenic C18:03, Arachidic acid-C20.00, Behenic acid-C22:00, Gadoleic acid-C20:01 was computed from the GC peak area of each FAME. No correction factors or internal standards were used. The average values for the two replicates were accepted as the concentrations of the sunflower oil samples. The precision of the results is given in the description of ISO 5508/9 standard method.

ANN Classification

Probabilistic ANNs have been used for classification [8]. The designed ANN is a two-layer probabilistic neural network consisting in a radial basis layer and a competitive layer. The radial basis layer computes distances from the input vector to the training input vectors and produces a vector whose elements measure how close the particular input is to a training input. The competitive layer cumulates these contributions for each class of inputs to produce, as its net output, a vector of probabilities. The layer uses a compete

transfer function for selecting the maximum of these probabilities and generates a vector having as elements a 1 (one) for the particular class and 0 (zero) for the other classes. The probabilistic ANN is guaranteed to converge to a Bayesian classifier. Artificial Neural Network Toolbox of Matlab software was used to build the present application [8, 9]. The input and hidden layer consist of fifteen neurons and the output layer includes three neurons. Output values of the network are rounded to the nearest integer to obtain the classes.

RESULTS AND DISCUSSION

Degree of fat extraction

First, the set of 15 sunflower seeds genotypes considered in the study (ten inbred lines and five hybrids) have been submitted to the investigations of fats and fatty acid content. The laboratory investigation determined the quantity of fat extracted by ASE method, as presented in Fig. 1.

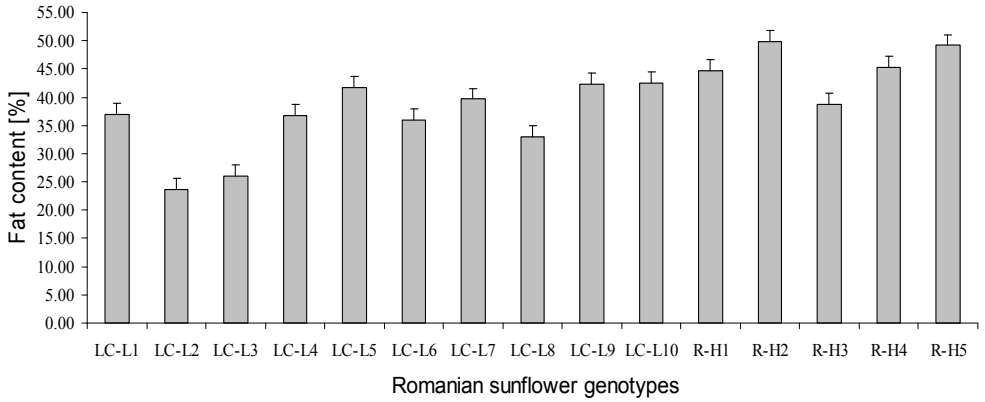

Figure 1. Fat content in the sunflower seeds of the investigated Romanian genotypes

For the case of the inbred lines the fat content ranges from 24.01% for the LC-L2 line to 42.49% for the LC-L10 line, while for the hybrids the fat content varies between 38.75% for the R-H3 line and 49.15% for the R-H2 line. Results confirm the large variability of fat content characterizing the Romanian sunflower seeds genotypes. The experimental results are show in Supporting Information.

Second, the profile of fatty acid analyzed thought GC-FID methods show that twelve types of fatty acids were determined [10]. The degree of saturation for the investigated inbred lines and Romanian sunflower hybrids are presented in Table 2.

Table 2. The degree of saturation results for the investigated inbred lines and Romanian sunflower hybrids.

Inbred lines and hybrids	Total Saturated fatty acids [%]	Total Monounsaturated fatty acids [%]	Total Polyunsaturated fatty acids [%]	Total Unsaturated fatty acids [%]	Saturated/ unsaturated acid ratio	Linoleic /oleic acid ratio
LC-L1	11.441	31.719	56.840	88.559	0.129	1.787
LC-L2	13.291	48.724	37.985	86.709	0.153	0.772
LC-L3	10.652	42.144	47.204	89.348	0.119	1.116
LC-L4	12.765	40.528	46.707	87.235	0.146	1.149
LC-L5	10.704	39.015	50.281	89.296	0.120	1.291
LC-L6	10.007	41.716	48.277	89.993	0.111	1.153
LC-L7	13.994	39.865	46.141	86.006	0.163	1.153
LC-L8	12.432	47.099	40.469	87.568	0.142	0.852
LC-L9	13.957	43.345	42.698	86.043	0.162	0.973
LC-L10	10.545	35.731	53.724	89.455	0.118	1.502
R-H1	12.081	29.138	58.781	87.919	0.137	2.014
R-H2	13.275	25.369	61.356	86.725	0.153	2.429
R-H3	14.060	27.330	58.610	85.940	0.164	2.149
R-H4	13.470	26.010	60.520	86.530	0.156	2.344
R-H5	13.836	26.548	59.616	86.164	0.161	2.256

ANN screening data

The classification of the samples according to the fatty acid content into three groups has been performed using ANNs. There were two criteria of classification, the degree of polyunsaturation and the linoleic/oleic acid ratio, which allowed the defining of two *categories*. As a result, the two considered categories are: the category of polyunsaturated fatty acids and the category of linoleic/oleic acid ratio. For each of the two considered categories, ANNs were used for classifying the samples into three *groups*.

The first classification, of polyunsaturated fatty acid category, was performed by the ANNs for dividing the samples into three groups according to the percentage of polyunsaturated fatty acid content. The ANN was trained using a set of input–output literature data [11-16]. Inputs of the ANN have been considered the contents of: Miristic acid-C14:00, Pentadecanoic acid-C15:00, Palmitic acid-C16:00, Palmitoleic acid-C16:01, Heptadecanoic-C17:00, Stearic acid-C18:00, Oleic acid-C18:01, Linoleic acid-C18:02, Linolenic C18:03, Arachidic acid-C20.00, Behenic acid-C22:00, Gadoleic acid-C20:01. The outputs of the ANN are: *group 1* of samples having the polyunsaturated fatty acid content less than 40%, *group 2* of samples having the polyunsaturated fatty acid content between 40%-50% and *group 3* of samples with the polyunsaturated fatty acid content higher than 50%. The ANN testing step followed the training procedure. A randomly selected testing set of 15 experimental samples obtained from the Romanian inbred line genotypes (not yet seen by the ANN) was given to the previously trained ANN in order to test its ability to properly perform the classification. The investigated sunflower genotypes have the polyunsaturated fatty acids content higher than 40%,

except one inbred line having less than 40% polyunsaturated fatty acids (LC-L2 with 37.985%). The classification results for this testing set of samples are presented in Fig. 2, where the ANN predicted data are compared to the experimental results.

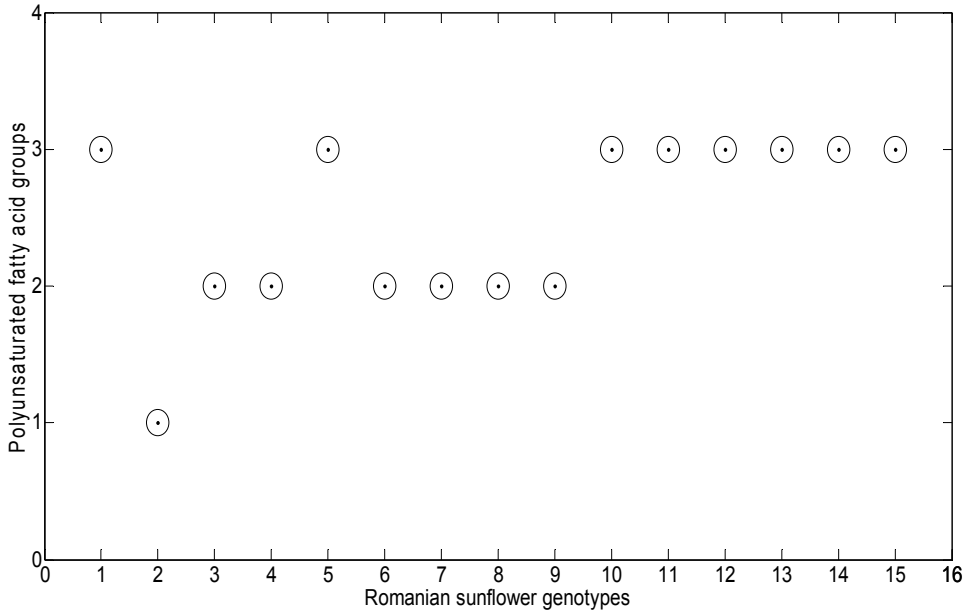

Figure 2. Classification results of polyunsaturated fatty acid groups, for the testing set of data; Group 1 of polyunsaturated fatty acid ≤40%, Group 2 of polyunsaturated fatty acid 40% - 50%, Group 3 of polyunsaturated fatty acid ≥ 50%; the ANN groups predictions, marked by "⊙", and the desired polyunsaturated fatty acid groups (experimental results), marked by "•".

The three groups predicted by ANN simulation for the classification of the polyunsaturated fatty acid reveal a very good fit between the classification groups performed by the ANN and the groups obtained by the ASE method.

The second ANN classification of the experimental data, according to the linolenic/oleic acid ratio criterion, was also based on a previous step of ANN training. Again, literature data was used to train the ANN [11-16]. Values of the linoleic/oleic acid ratio reveal which fatty acid is predominant in the oilseed, i.e. either the monounsaturated fatty acid (oleic acid) or the polyunsaturated fatty acid (linoleic acid). In order to test the ability of the trained ANN for classifying the data, the testing was performed on the set of 15 experimental samples obtained from the Romanian inbred lines genotypes. Classification results are presented in Fig. 3.

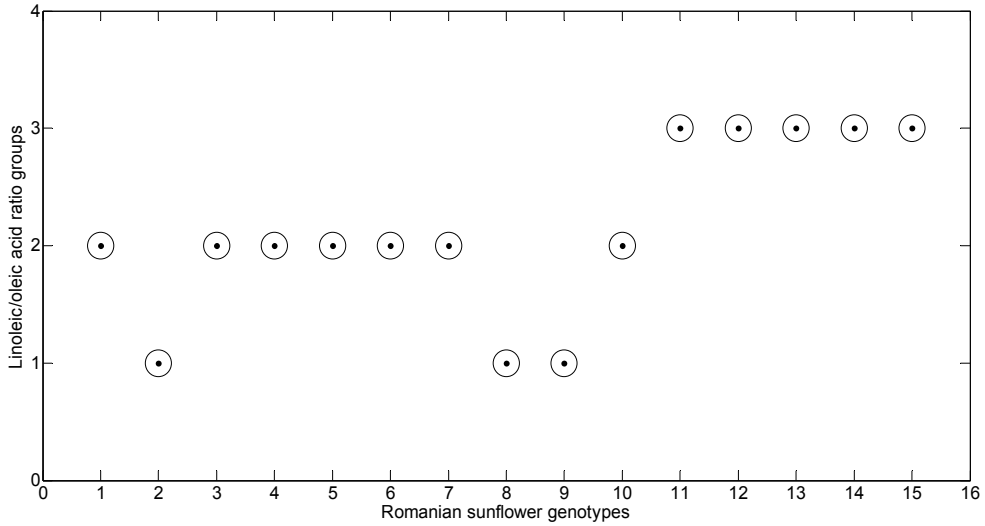

Figure 3. Classification results of linoleic/oleic acid ratio groups, for the testing set of data; Group 1 of linoleic/oleic acid ratio ≥ 2, Group 2 of linoleic/oleic acid ratio between 1-2, Group 3 of linoleic/oleic acid ratio ≤ 1;the ANN groups predictions, marked by "◌", and the desired linoleic/oliec acid ratio groups (experimental results), marked by "•".

Again, the classification reveals a perfect fit between the groups obtained by the experimental investigations and the results predicted by the ANN.

The obtained results prove the capacity of the trained ANNs for making exact sunflower oil seed group classifications in the cases of samples of unknown group membership.

Prediction of determined classes

Furthermore, by merging the two categories presented in Fig. 2 and Fig. 3 the combined results presented in Fig. 4 have been obtained.

Results presented in Fig. 4 reveal the prediction of three *classes* of the Romanian sunflower seeds genotypes. The first class consists of ten investigated sunflower oilseed genotypes with a high content of polyunsaturated fatty acids and a linoleic/oleic acid ratio less than 1. The second class consists of two sunflower inbred lines (LC-L8 and LC-L9), where the linoleic acid is dominant and the polyunsaturated fatty acid content places it into group 2 of the first classification category. The third class is composed of three sunflower lines (LC-L1, LC-L5 and LC-L10) where the oleic acid is dominant and the polyunsaturated fatty acid content is higher than 50%, corresponding to group 3 of the first classification category. It may be also noticed that all five considered hybrid lines are oleic hybrids, as they predominantly contain the oleic acid.

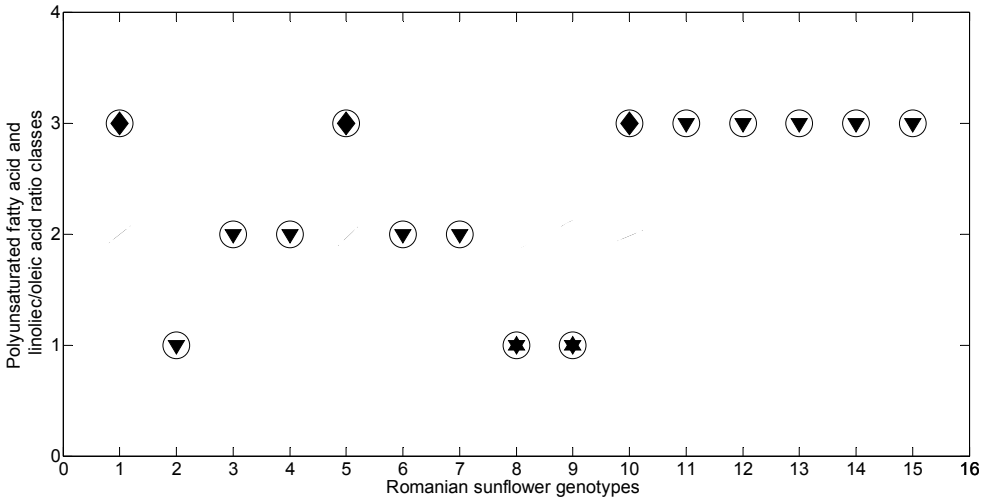

Figure 4. Classes of investigated Romanian sunflower genotypes; the ANN predicted classes are marked by "◌" and the classes obtained from experimental data are: class 1 marked by "▼", class 2 marked by "✸" and class 3 marked by " ♦ ".

It may be concluded that the investigated lines may be used in crossed hybridization in order to obtain new hybrids with high content of either oleic or linoleic acid. The investigated hybrid lines may be appreciated for their nutritional potential, due to their high polyunsaturated fatty acid content.

CONCLUSIONS

Several genotypes of Romanian sunflower oilseeds have been considered in the investigation of the fatty acid content, using laboratory analytical techniques. Methodologies based on the two layer probabilistic ANNs, consisting in a radial basis layer and a competitive layer, have been used for classification of sunflower oilseed genotypes according to fatty acid content. The ANNs were trained with experimental data from the literature. Two classifications (categories) were obtained, function of (1) the degree of polyunsaturation and (2) the linoleic/oleic acid ratio. Three groups have been considered for each investigated category. Three classes have been also predicted by the investigation of the Romanian sunflower oilseeds genotypes.

The classification aptitude of the designed ANNs proved to be very good, as the group classification has been performed with no error, for both categories. Moreover, the advantages of the classification based on ANNs are significant when taking into consideration the limited number of parameters required as ANN inputs and the number of samples needed for the training set. Further improvement may be obtained by extending the number of input parameters and by carefully filtering the training set of data.

Starting from the methodology of designing ANNs, aimed to predict and classify the sunflower seeds, further classification procedures may be developed for similar applications. ANNs incentives consist in the help they bring for characterizing the sunflower oils with economical implication as oil composition assessment is an essential feature to provide its potential use in different fields, such as: nutrition, pharmacy, health care or other nonfood applications.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support for providing the sunflower seeds genotypes to the Agricultural Research and Development Institute Fundulea, and to the Romania-Slovenia bilateral cooperation project.

REFERENCES

1. L. D. Villamide, *British Poultry Science*, **2000**, *41*, 182.
2. D. Brodnjak-Voncina, Z. C. Kodbab, M. Novic, *Chemometrics and Intelligent Laboratory Systems*, **2005**, *75*, 31.
3. Y. Tsompanakis, N. D. Lagaros, G. E.Stavroulakis, *Advances in Engineering Software,* **2008**, *39*, 612.
4. L. Fengjun, X. Zongben, *Chaos, Solitons and Fractals*, **2008**, *36*, 572.
5. D. Bratfalean, V. M. Cristea, P. S. Agachi, D. F. Irimie, *Revue Roumanie de Chimie*, **2008**, *53*, 881.
6. AOAC, "Official Method of Analysis" Association of Official Analytical Cereal Chemists, Washington, DC, USA, **1999**.
7. B. Matthaüs, L. Brühl, *Journal of the American Oil Chemist's Society*, **2001**, *78*, 95.
8. D. Bratfalean, M. V. Cristea, P. S. Agachi, D. F. Irimie, A. Sarrafi, M. Petitprez, Proceedings of the *17th European Symposium on Computer Aided Process Engineering – ESCAPE17*, May, Bucharest, **2007**, 1.
9. M. V. Cristea, S. Varvara, L. Muresan, I. C. Popescu, *Indian Journal of Chemistry*, **2003**, *42A*, 764.
10. D. Bratfalean, D. F. Irimie, P. S. Agachi, *Studia Universitatis Babes-Bolyai, Chemia*, **2007**, *LII 2*, 47.
11. B. Perez-Vich, L. Velasco, J. M. Fernandez-Martinez, *Journal of the American Oil Chemist's Society*, **1998**, *75*, 547.
12. B. Perez-Vich, R. Garces, J. M. Fernandez-Martinez, *Heli,* **2000**, *23*, 77.
13. A. V. Vrânceanu, "*Floarea-soarelui Hibridă"*, Ceres Press, Bucharest, **2000**, p.15-25, p. 650-720.
14. G. J. Seiler, "Potential Source of Reduced Palmitic and Stearic Fatty Acids in Sunflower Oil From a Population of Wild Helianthus Annuus*", ASHS Press, Alexandria* VA, **2002**, p.150-154.
15. N. Izquierdo, L. A. F. Andrade, V. Pereyra, *Field Crops Research*, **2002**, *77*, 115.
16. K. Warner, B. Vick, L.K. Leingrtner, R. Isaak, K. Doroff, Proceedings from the 18th session of the CCFO – Report of the eighteenth session of the codex committee on fats and oil, U.K., **2003**.