Online Version ISSN: 1314-412X
Volume 4, Number 2
June 2012

AGRICULTURAL SCIENCE AND TECHNOLOGY

2012

An International Journal Published by Faculty of Agriculture,
Trakia University, Stara Zagora, Bulgaria

Scope and policy of the journal

Agricultural Science and Technology /AST/ – an International Scientific Journal of Agricultural and Technology Sciences is published in English in one volume of 4 issues per year, as a printed journal and in electronic form. The policy of the journal is to publish original papers, reviews and short communications covering the aspects of agriculture related with life sciences and modern technologies. It will offer opportunities to address the global needs relating to food and environment, health, exploit the technology to provide innovative products and sustainable development. Papers will be considered in aspects of both fundamental and applied science in the areas of Genetics and Breeding, Nutrition and Physiology, Production Systems, Agriculture and Environment and Product Quality and Safety. Other categories closely related to the above topics could be considered by the editors. The detailed information of the journal is available at the website. Proceedings of scientific meetings and conference reports will be considered for special issues.

Submission of Manuscripts

All manuscript written in English should be submitted as MS-Word file attachments via e-mail to ascitech@uni-sz.bg. Manuscripts must be prepared strictly in accordance with the detailed instructions for authors at the website
http://www.uni-sz.bg/ascitech/index.html
and the instructions on the last page of the journal. For each manuscript the signatures of all authors are needed confirming their consent to publish it and to nominate on author for correspondence.
They have to be presented by a submission letter signed by all authors. The form of the submission letter is available upon from request from the Technical Assistance or could be downloaded from the website of the journal. All manuscripts are subject to editorial review and the editors reserve the right to improve style and return the paper for rewriting to the authors, if necessary. The editorial board reserves rights to reject manuscripts based on priorities and space availability in the journal.

Internet Access

This journal is included in the Trakia University Journals online Service which can be found at www.uni-sz.bg.

Address of Editorial office:

Agricultural Science and Technology
Faculty of Agriculture, Trakia University
Student's campus, 6000 Stara Zagora
Bulgaria
Telephone.: +359 42 699330
+359 42 699446
http://www.uni-sz.bg/ascitech/index.html

Technical Assistance:

Nely Tsvetanova
Telephone.: +359 42 699446
E-mail: ascitech@uni-sz.bg

Online Version ISSN: 1314-412X

Volume 4, Number 2
June 2012

AGRICULTURAL SCIENCE AND TECHNOLOGY

2012

An International Journal Published by Faculty of Agriculture,
Trakia University, Stara Zagora, Bulgaria

AGRICULTURAL SCIENCE AND TECHNOLOGY, VOL. 4, No 2, pp 177 - 183, 2012

Product Quality and Safety

Modeling of spectral data characteristics of healthy and *Fusarium* diseased corn kernels

Ts. Draganova*

Department of Automatics, Information and Control Engineering, University of Ruse, 8 Studentska, 7017 Ruse, Bulgaria

Abstract. *Two approaches for identification of infected with Fusarium maize kernels using spectral characteristics in the near infrared region are described in the paper. They are used for spectral data reduction and for corn classification. First approach is based on analysis of discrete linear parametric models coefficients. Second approach is based on principal component analysis. Maximum percentage of correct rate achieve 74,67% for Fusarium infected corn kernels when linear discrete model coefficients are used as classification features. Principal components show overlapped classes but in combination with appropriate classifier as classification tree percentage of correct rate achieve 100% for Fusarium infected corn kernels. The main advantage for the classifier which is used (classification tree) is that it is created using simple procedure without any additional parameters accept training data set and target for each input variable.*

Keywords: corn kernels, Fusarium disease, near infrared spectral characteristics, autoregression linear discrete model, classification tree

Introduction

Fusarium disease, caused by several *Fusarium* species, is one of the major fungal diseased of cereals worldwide (Krska, 2008). *Fusarium* Graminearum and *Fusarium* moniliforme are the most significant, damaging and common pathogens in Bulgaria (Beev et al., 2011; Georgiev et al., 2011).

Corn kernels consists of four main parts: the germ, endosperm, pericarp and tip, through which the seeds are attached to the book. Endosperm is the - largest component by weight of maize seed and occupies about 83% by weight. With species of maize endosperm has a different texture - powdery, horny, waxy, and different colors - white, yellow with different shades.

Three traditional diagnostics methods for detection and identification of *Fusarium* species of corn kernels are used in Bulgaria. The first is based on visual assessment by expert. The second one is based on chemical analysis and the third is based on biology analysis of micro- and macro-morphological features of the *Fusarium* fungal. Some other methods like fluorometric analysis, gas and liquid chromatography are reliable and accurately. The basic disadvantages of these destructive and non-destructive methods of detection are time consuming, requires highly skilled experts and each method varies in the requirements for technical investments. The comparative analysis indicate that the computer systems using image processing and systems using spectral characteristics processing are the most common applied methods in precision farming and agriculture.

The grain industry requires high accuracy of *Fusarium* diagnostics, limited capital investments and short time of data processing, which could do with new information technologies. The implementation of new methods for reliable and fast identification and classification of seeds is of major technical and economical importance in the agricultural industry.

The problem of food quality control is determination of features for products classification in classes defined in the Standards. The development of an objective methodology for grading of food and agricultural products in the last three decades has significantly reduced the place of subjective methods. During the last years more popular method for different kind of food quality assessment is spectral data analysis (Damyanov, 2006; Damyanov et al., 2009; Mladenov, 2011).

The purpose of this paper is to illustrate corn kernels spectral data modeling using discrete linear models, to analyze the parameters of linear discrete models and significant to be used as criteria for separation of healthy and *Fusarium* infected corn kernels, to reduce the spectral information and to choose a classification procedure for corn kernels recognition.

Material and methods

1. Corn kernel samples

Grain samples were evaluated by expert at the Institute of corn - Bulgaria, where was established and cause infection - *Fusarium.* Reference samples included 150 healthy and 150 infected corn kernels. Colour images of healthy and *Fusarium* infected corn kernels are shown on Figure 1. Methodology to obtain the spectral characteristics of the kernels (Figure 2) includes the following main steps: forming two groups of maize kernels (healthy and infected), each group consists of 150 grains; each kernel characteristics are taken on both sides (of the germ side and the other side); forming the spectral characteristics of 300 kernels group and after three repeated measurements for a group of healthy and infected kernels group were collected in 900 spectral characteristics.

Two sets of data were formed. Training set includes 300 spectral characteristics for the germ side of corn kernels and 300 – for the other side for healthy and *Fusarium* infected respectively. Testing set includes 150 spectral characteristics for the germ side of corn kernels and 150 – for the other side for healthy and *Fusarium* infected respectively.

* e-mail: cgeorgieva@uni-ruse.bg

a) b) c) d)

Figure 1. Colour images of healthy and *Fusarium* infected corn kernels

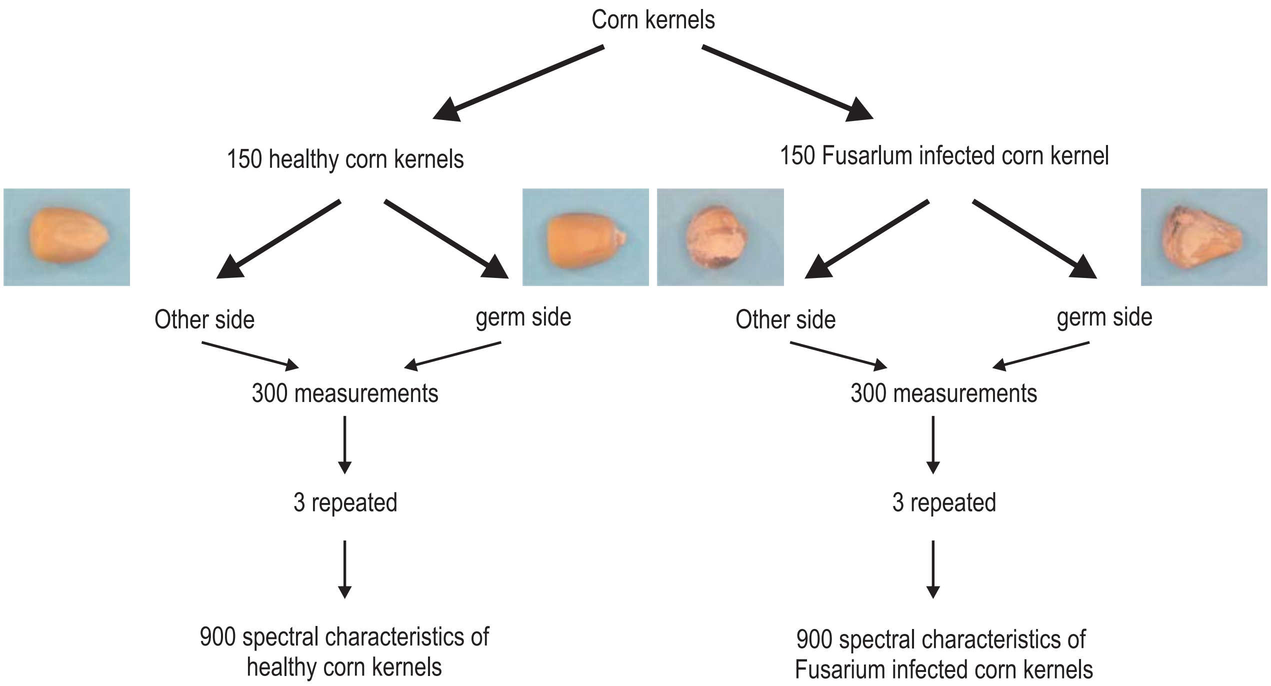


Figure 2. Methodology to obtain the spectral characteristics of the grains

2. Near Infrared Spectra Acquisition

Near infrared spectra of grain samples were obtained by NIRSystem 6500 spectrophotometer (FOSS NIRSystems, Silver Spring, MD, USA), in the spectral region from 400 to 2500 nm with 2 nm step. The spectrophotometer and computer system are presented on Figure 3. An additional tool which is plastic kernel holder was made. The kernel place is in the middle of the holder.

Spectral data are shown on Figure 4. The data which are in

Figure 3. Algorithm for identification of healthy and *Fusarium* infected corn kernels

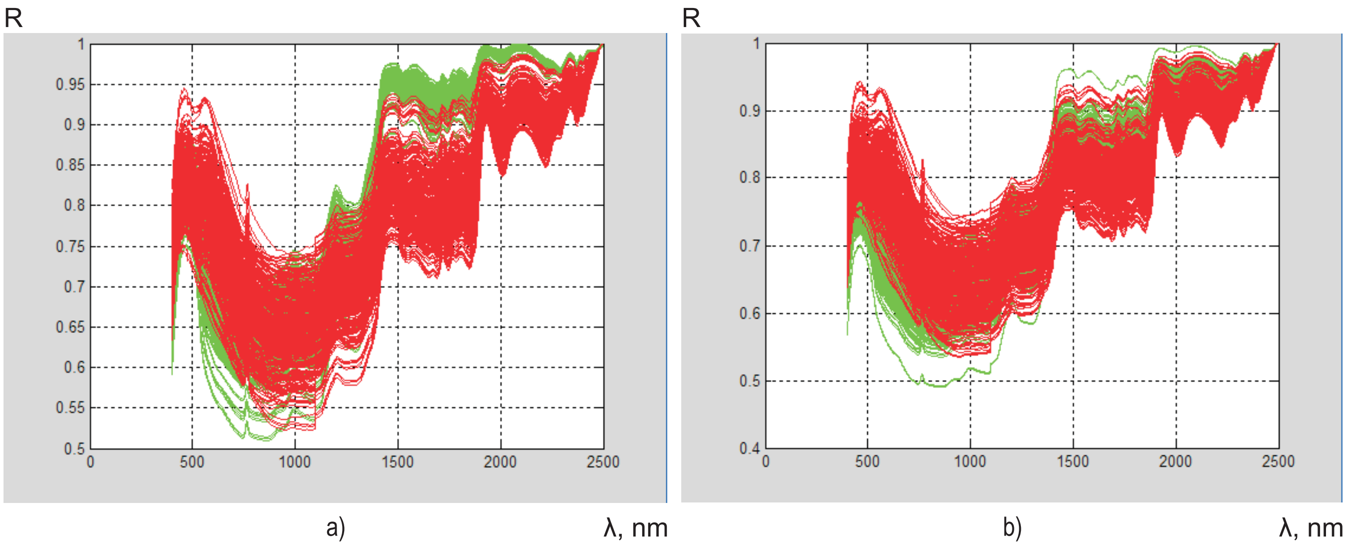


Figure 4. Normalized spectral characteristics of each kernel: a) – germ side of the kernel, b) – other side of the kernel

green are obtained from healthy kernels and these which are obtained from *Fusarium* infected kernels are shown in red colour.

Results show that there are not so much differences between characteristics and it is difficult to separate the kernels in two classes using raw data from spectrophotometer. Therefore methods for spectral characteristics modelling, data reduction and features extraction are applied to the obtained information.

3. Spectral data modeling, reduction and features extraction

Data modeling is realized using linear discrete model of the Autoregression (AR). Principal component analysis (Abdi, 2010) as a most popular method for data reduction is used for reduction of spectral information because obtained values for each characteristic are 1050. These values are too much for classifiers – creation, training and testing. Features extraction is based on the coefficients and principal components histograms – healthy and *Fusarium* infected kernels.

3.1. Data modeling using linear discrete model

The linear discrete model of the Autoregression (AR) type (Daskalov, 2010) is used for describing the spectral data of corn kernels with the following equation:

$$S_{\lambda}(k) + a_1 S_{\lambda}(k-1) + \dots + A_n S_{\lambda}(k-n) = e(k), \quad (1)$$

where k is the *k*-th value of the wavelength λ in nm;

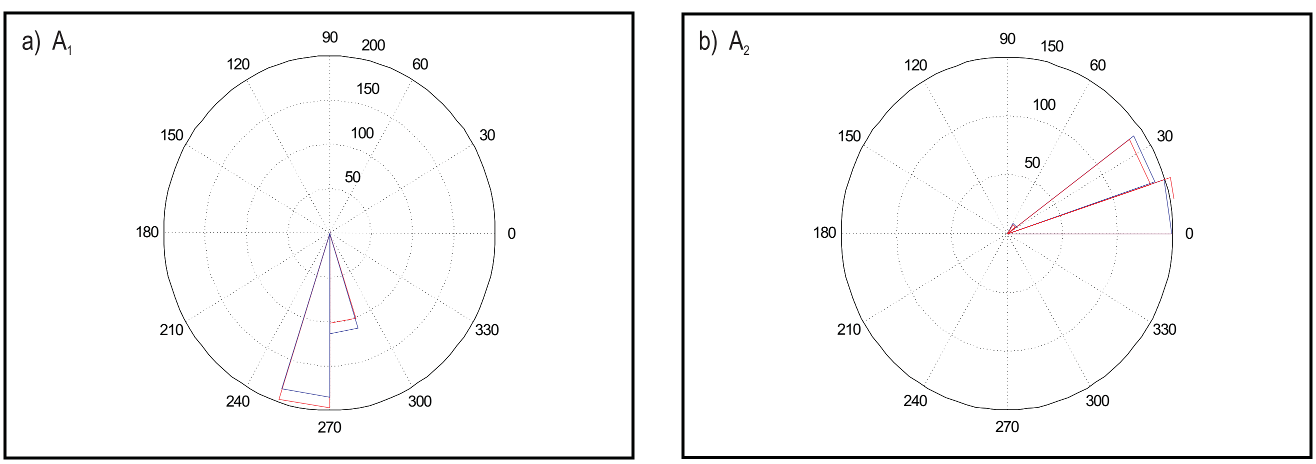
A_i, $i = (1 \div n)$ are the coefficients of the Autoregression model (AR);

S_{λ} (k) is the intensity of the radiation reflected from the maize kernels for the *k*-th value of the wavelength λ in nm;

n is the model series of the Autoregression type;

e(k) is the difference (error) between the modelled and the real characteristics.

Preliminary experiments (Daskalov, 2010) show that first ten coefficients are informative for classification of the kernels. Therefore the histograms of the first ten coefficients from linear models of healthy and *Fusarium* infected corn kernels are shown on Figure 5. The histograms are in polar plot and for angle histogram creation is used MATLAB. The plot showing the distribution of values grouped according to their numeric range, showing the distribution of theta in 20 angle bins. The vector of linear model coefficients, expressed in radians, determines the angle of each bin from the origin. The length of each bin reflects the number of elements in the vector that fall within a group, which ranges from 0 to the greatest number of elements deposited in any one bin.



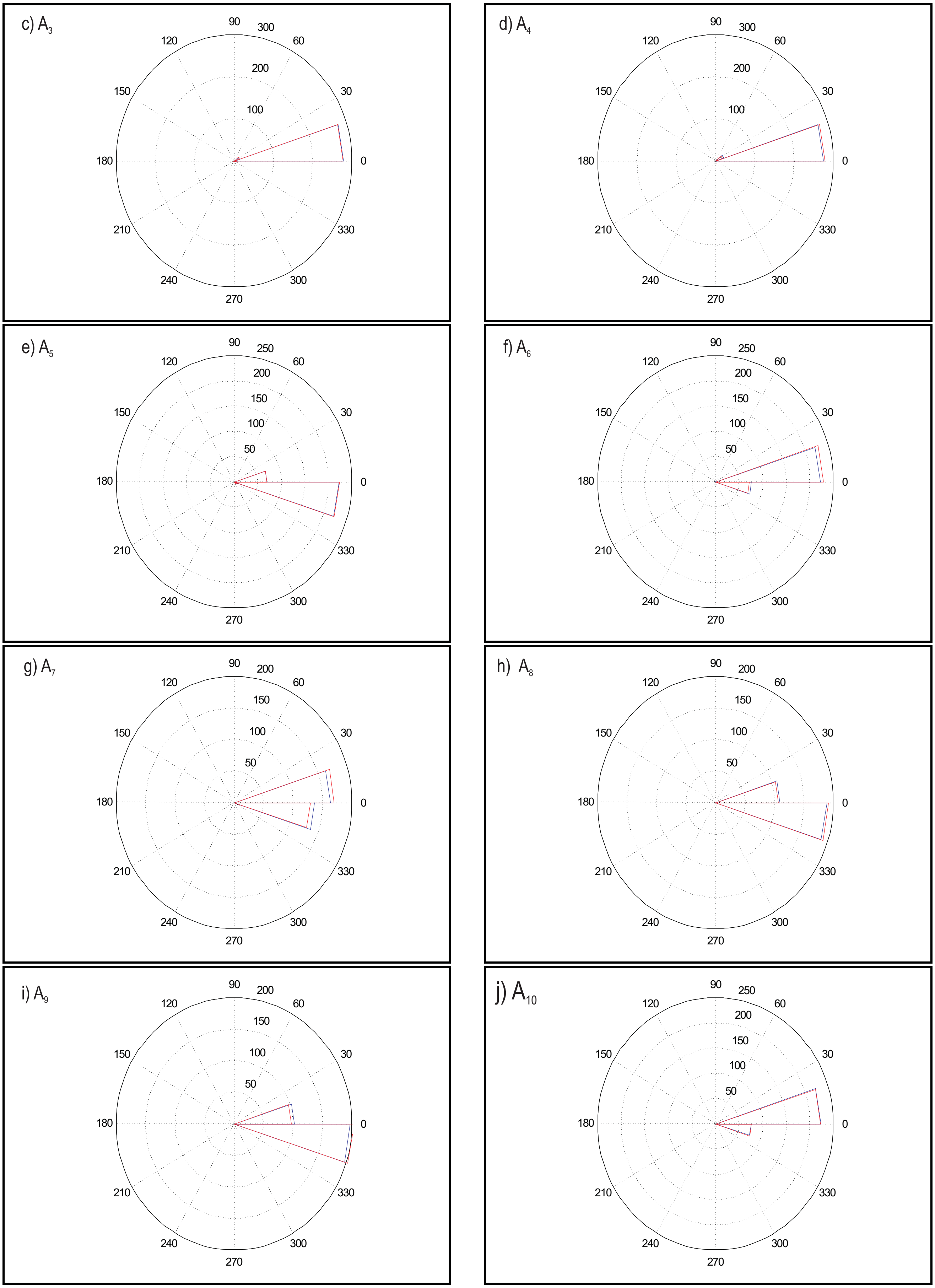
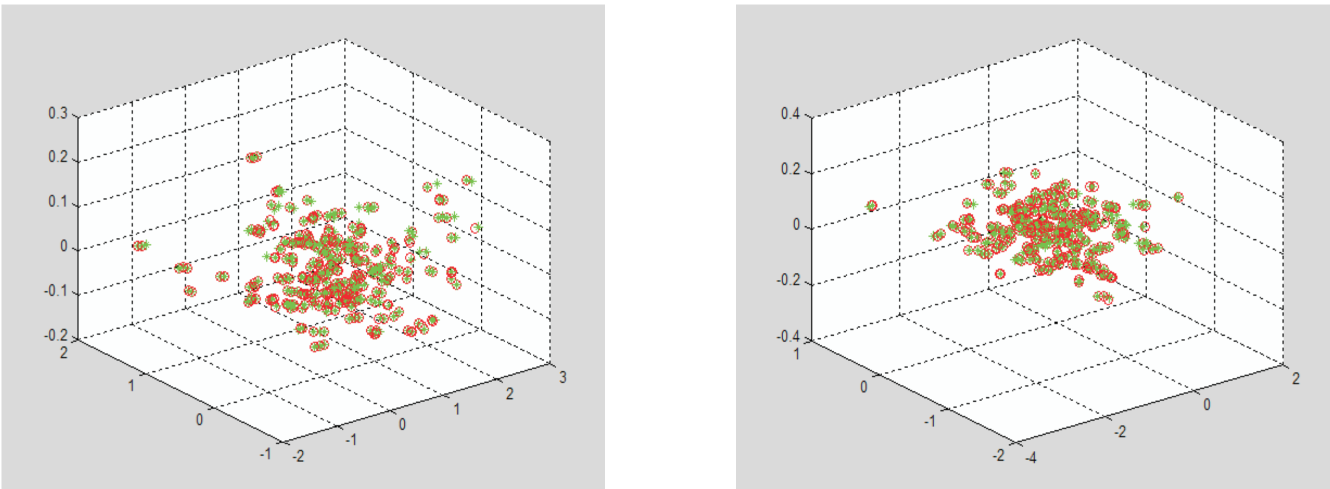


Figure 5. Histograms of coefficients from A_1 to A_{10} for two classes of corn kernels – healthy (blue) and *Fusarium* infected (red)

The results are obtained for the germ side of the corn kernels. The results from the other side of the kernels have similarity shapes. The differences between classes are not so clearly define and probably only three coefficients could be use as a classification features – A_1, A_2 and A_7.

3.2. Data reduction using Principal component analysis

Principal component analysis which is most popular method for data reduction (Abdi, 2010) is used for spectral data transformation in the space of the principal components and for reduction the vector of the obtained reflection coefficients from 1050 to maximum ten features. Obtained first three principal components for two classes of corn kernels (healthy and *Fusarium* infected) are shown on Figure 6. Class healthy kernels are shown in green colour and class *Fusarium* infected – in red colour. The results show that the classes are overlap. It is necessary to use appropriate classification procedure for these overlapped classes of corn kernels in which linear model coefficients and principal components will use as an input variables.



a) b)

Figure 6. First three principal components for two classes of corn kernels (healthy and *Fusarium* infected): a) germ side of the kernels, b) other side of the kernels

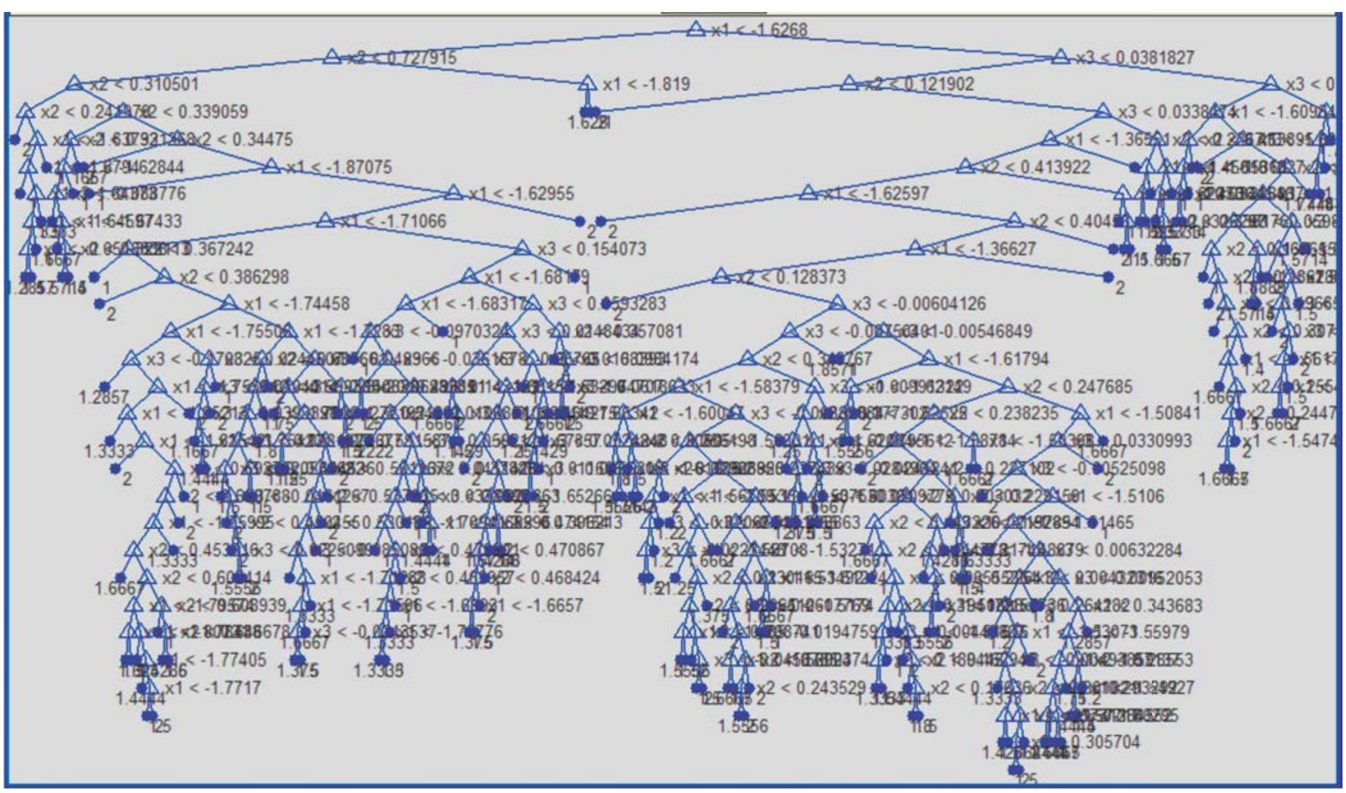


Figure 7. Classification tree with linear models coefficients

4. Classification procedure

Classification trees are widely used as nonparametric classifiers (Venables and Ripley, 2002). Therefore this method does not require specification of any functional form. Classification trees predict responses to data. To predict a response, follow the decisions in the tree from the root (beginning) node down to a leaf node. The leaf node contains the response. Classification trees give responses that are nominal, such as 'true' or 'false'. This classiciation procedure can easily handle outliers. Outliers can negatively affect the results of some statistical models, like Principal Component Analysis (PCA) and linear regression. But the splitting algorithm of the tree will easily handle noisy data. Classification trees isolate the outliers in a separate node. This property is very important, because spectral data of real objects as corn kernels very often have outliers due to variety characteristics or climate conditions. Tree – based methods work in practice and have some advantages as follow algorithmic basis instead of mathematical; give good results in some cases when classical methods are less satisfactory (Venables and Ripley, 2002).

The classification tree created from the training set (which combines the data for two side of the corn kernels – germ and other side) is shown on Figure 7. In this case linear model coefficients are used as input values for the tree.

MATLAB software is used for programming, creation and testing the classification tree. Part of classification rules which are created automatically are shown in Table 1.The classification trees aims are first to find a rule for classifying cases using a step-by-step approach, one variable at a time and second to produce a rule for classifying objects into categories. High dimensions and complicated rules give over-optimistic performance.

Table 1. Classification tree rules

1	if x1<-1.6268 then node 2 elseif x1>=-1.6268 then node 3 else 1
2	if x2<0.720207 then node 4 elseif x2>=0.720207 then node 5 else 1
3	if x3<0.0381093 then node 6 elseif x3>=0.0381093 then node 7 else 2
4	if x1<-1.85213 then node 8 elseif x1>=-1.85213 then node 9 else 1
5	class = 2
6	if x2<0.121902 then node 10 elseif x2>=0.121902 then node 11 else 2
7	if x3<0.107102 then node 12 elseif x3>=0.107102 then node 13 else 2
8	class = 1
9	if x2<0.310501 then node 14 elseif x2>=0.310501 then node 15 else 1
10	class = 1
11	if x3<0.0338474 then node 16 elseif x3>=0.0338474 then node 17 else 2
12	if x1<-1.60984 then node 18 elseif x1>=-1.60984 then node 19 else 2
13	if x3<0.146618 then node 20 elseif x3>=0.146618 then node 21 else 1
14	if x2<0.241979 then node 22 elseif x2>=0.241979 then node 23 else 1
15	if x2<0.339059 then node 24 elseif x2>=0.339059 then node 25 else 1
16	if x1<-1.36551 then node 26 elseif x1>=-1.36551 then node 27 else 2
17	if x2<0.236713 then node 28 elseif x2>=0.236713 then node 29 else 1
18	if x2<0.453895 then node 30 elseif x2>=0.453895 then node 31 else 1
19	if x1<-1.60522 then node 32 elseif x1>=-1.60522 then node 33 else 2
20	class = 1
21	class = 2
22	class = 2

Results and discussion

The corn kernels from the test set are analyzed using their spectral characteristics. The linear model coefficients A_1, A_2 and A_7 and first three principal components are calculated. The data from the test set were divided in for subsets – Test group 1 (healthy corn kernels – germ side), Test group 2 (healthy corn kernels – other side), Test group 3 (*Fusarium* infected corn kernels – germ side), Test group 4 (*Fusarium* infected corn kernels – other side). The results from classification the data using created classification tree are shown in Table 2.

Percentage of correct rate was calculated using following equation (Venables and Ripley, 2002):

$$\%correct \quad rate = \frac{Nc}{N}.100\% \tag{2}$$

where N_c is the number of corn kernels which are correct classified,

N – the total number of the corn kernels in the test subset.

Percentage of misclassification rate was calculated using following equation :

$$\%misclassification \quad rate = \frac{Nm}{N}.100\% \tag{3}$$

where N_m is the number of corn kernels which are non - correct classified,

N – the total number of the corn kernels in the test subset.

The classification results show that the percentage of correct rate when linear model coefficients are used as input features for classification tree are in the range from 60,67% for class healthy to 74,67% for class *Fusarium* infected. The classification results for the percentage of correct rate when principal components are used as input features for classification tree are better from 81,33% to 86% for class healthy and 100% for class *Fusarium* infected.

Table 2. Classification results

Groups	Linear model coefficients		Principal components	
	Class healthy	Class *Fusarium* infected	Class healthy	Class *Fusarium* infected
Test group 1				
Corn kernels	91	59	129	21
% correct rate	60.67		86	
%misclassification rate		39.33		14
Test group 2				
Corn kernels	91	59	122	28
% correct rate	60.67		81.33	
%misclassification rate		39.33		18.67
Test group 3				
Corn kernels	44	106	0	150
% correct rate		70.67		100
%misclassification rate	26.67		0	
Test group 4				
Corn kernels	38	112	0	150
% correct rate		74.67		100
%misclassification rate	25.33		0	

Conclusion

Linear discrete model is not so appropriate method for spectral data modeling – the coefficients of the model are overlapped for the classes and the accuracy of identification is lower. Maximum percentage of correct rate achieve 74,67% for *Fusarium* infected corn kernels.

Principal components also show overlapped classes but in combination with appropriate classifier as classification tree percentage of correct rate achieve 100% for *Fusarium* infected corn kernels.

The main advantage for the classifier which is used (classification tree) is that it is created using simple procedure without any additional parameters accept training data set and target for each input variable.

Acknowledgement

The study was supported by contract № BG051PO001-3.3.04/28, "Support for the scientific staff development in the field of engineering research and innovation". The project is funded with support from the Operational Program "Human Resources Development" 2007-2013, financed by the European Social Fund of the European Union.

References

Abdi H and Williams L, 2010. Principal component analysis. Journal of Computational Statistics, 2, 4, 433-459.

Beev G, Denev S and Pavlov D, 2011. Occurrence and distribution of *Fusarium* species in wheat grain. Agricultural Science and Technology, 3, 2, 165-168.

Damyanov Ch, 2006. Non-destructive identification of the quality in food products automatic sorting systems. Academic publisher of the University of Food Technologies-Plovdiv.

Damyanov Ch, Nachev V and Titova T, 2009. Modified VIS / NIR - photometric system for grading of food agricultural products. Journal of Automation and Information, 3, 41-48.

Daskalov P, Mancheva V, Draganova Ts and Tsonev R, 2010. An approach for *Fusarium* diseased corn kernels recognition using linear discrete models. Agricultural Science and Technology, 2, 2, 90 – 95.

Georgiev M, Pavlov D, Beev G, Gergzikova M and Bazitov R, 2011. Species composition of weeds in wheat and barley. Agricultural Science and Technology, 3, 2, 143-149.

Krska R, Schubert-Ullrich P, Molinelli A, Sulyok M, Macdonald S and Crews C, 2008. Mycotoxin analysis: An update. Food additives and contaminants, 25, 2, 152-163.

Mladenov M, 2011. Analysis and evaluation of the quality of grain. Academic publisher of the University of Ruse.

Venables WN and Ripley BD, 2002. Modern Applied Statistics with S-PLUS, (4th Edition).

Instruction for authors

Preparation of papers

Papers shall be submitted at the editorial office typed on standard typing pages (A4, 30 lines per page, 62 characters per line). The editors recommend up to 15 pages for full research paper (including abstract references, tables, figures and other appendices)

The manuscript should be structured as follows: Title, Names of authors and affiliation address, Abstract, List of keywords, Introduction, Material and methods,Results, Discussion, Conclusion, Acknowledgements (if any), References, Tables, Figures.

The title needs to be as concise and informative about the nature of research. It should be written with small letter /bold, 14/ without any abbreviations.

Names and affiliation of authors The names of the authors should be presented from the initials of first names followed by the family names. The complete address and name of the institution should be stated next. The affiliation of authors are designated by different signs. For the author who is going to be corresponding by the editorial board and readers, an E-mail address and telephone number should be presented as footnote on the first page. Corresponding author is indicated with *.

Abstract should be not more than 350 words. It should be clearly stated what new findings have been made in the course of research. Abbreviations and references to authors are inadmissible in the summary. It should be understandable without having read the paper and should be in one paragraph.

Keywords: Up to maximum of 5 keywords should be selected not repeating the title but giving the essence of study.

The introduction must answer the following questions: What is known and what is new on the studied issue? What necessitated the research problem, described in the paper? What is your hypothesis and goal ?

Material and methods: The objects of research, organization of experiments, chemical analyses, statistical and other methods and conditions applied for the experiments should be described in detail. A criterion of sufficient information is to be possible for others to repeat the experiment in order to verify results.

Results are presented in understandable tables and figures, accompanied by the statistical parameters needed for the evaluation. Data from tables and figures should not be repeated in the text.

Tables should be as simple and as few as possible. Each table should have its own explanatory title and to be typed on a separate page. They should be outside the main body of the text and an indication should be given where it should be inserted.

Figures should be sharp with good contrast and rendition. Graphic materials should be preferred. Photographs to be appropriate for printing. Illustrations are supplied in colour as an exception after special agreement with the editorial board and possible payment of extra costs. The figures are to be each in a single file and their location should be given within the text.

Discussion: The objective of this section is to indicate the scientific significance of the study. By comparing the results and conclusions of other scientists the contribution of the study for expanding or modifying existing knowledge is pointed out clearly and convincingly to the reader.

Conclusion: The most important consequences for the science and practice resulting from the conducted research should be summarized in a few sentences. The conclusions shouldn't be numbered and no new paragraphs be used. Contributions are the core of conclusions.

References:

In the text, references should be cited as follows: single author: Sandberg (2002); two authors: Andersson and Georges (2004); more than two authors: Andersson et al.(2003). When several references are cited simultaneously, they should be ranked by chronological order e.g.: (Sandberg, 2002; Andersson et al., 2003; Andersson and Georges, 2004).

References are arranged alphabetically by the name of the first author. If an author is cited more than once, first his individual publications are given ranked by year, then come publications with one co-author, two co-authors, etc. The names of authors, article and journal titles in the Cyrillic or alphabet different from Latin, should be transliterated into Latin and article titles should be translated into English. The original language of articles and books translated into English is indicated in parenthesis after the bibliographic reference (Bulgarian = Bg, Russian = Ru, Serbian = Sr, if in the Cyrillic, Mongolian = Mo, Greek = Gr, Georgian = Geor., Japanese = Ja, Chinese = Ch, Arabic = Ar, etc.)

The following order in the reference list is recommended:

Journal articles: Author(s) surname and initials, year. Title. Full title of the journal, volume, pages. Example:

Simm G, Lewis RM, Grundy B and Dingwall WS, 2002. Responses to selection for lean growth in sheep. Animal Science, 74, 39-50

Books: Author(s) surname and initials, year. Title. Edition, name of publisher, place of publication. Example:

Oldenbroek JK, 1999. Genebanks and the conservation of farm animal genetic resources, Second edition. DLO Institute for Animal Science and Health, Netherlands.

Book chapter or conference proceedings: Author(s) surname and initials, year. Title. In: Title of the book or of the proceedings followed by the editor(s), volume, pages. Name of publisher, place of publication. Example:

Mauff G, Pulverer G, Operkuch W, Hummel K and Hidden C, 1995. C3-variants and diverse phenotypes of unconverted and converted C3. In: Provides of the Biological Fluids (ed. H. Peters), vol. 22, 143-165, Pergamon Press. Oxford, UK.

Todorov N and Mitev J, 1995. Effect of level of feeding during dry period, and body condition score on reproductive performance in dairy cows,IXth International Conference on Production Diseases in Farm Animals, Sept.11 – 14, Berlin, Germany, p. 302 (Abstr.).

Thesis:

Penkov D, 2008. Estimation of metabolic energy and true digestibility of amino acids of some feeds in experiments with muscus duck (Carina moshata, L). Thesis for DSc. Agrarian University, Plovdiv, 314 pp.

The Editorial Board of the Journal is not responsible for incorrect quotes of reference sources and the relevant violations of copyrights.

AGRICULTURAL
SCIENCE AND TECHNOLOGY

Volume 4, Number 2
June 2012





CONTENTS

Journal web site:
www.uni-sz.bg/ascitech/index.html

Publisher:
www.alfamarket.biz