

## ИНФОРМАЦИОННИ ХАРАКТЕРИСТИКИ НА СПЕКТРАЛНИТЕ ДАННИ НА ЦАРЕВИЧНИ СЕМЕНА ПРИ ДИАГНОСТИЦИРАНЕ НА ЗАБОЛЯВАНЕТО ФУЗАРИУМ

### INFORMATIVE CHARACTERISTICS OF CORN KERNELS SPECTRAL DATA USED FOR FUSARIUM DISEASE DIAGNOSTIC

**Цветелина Драганова**

Русенски Университет “Ангел Кънчев”  
cgeorgieva@uni-ruse.bg

**Пламен Даскалов**

Русенски Университет “Ангел Кънчев”  
daskalov@uni-ruse.bg

**Петя Велева - Донева**

Технически Университет Габрово  
pveleva@server.uni-sz.bg

#### Abstract

*Informative characteristics of corn kernels spectral data used for Fusarium disease diagnostic are obtained and described in the paper. Two approaches – first one based on wavelet transformation and second one based on method Soft Independent Modeling of Class Analogy (SIMCA) are used for spectral data processing. First approach is useful for obtaining the informative area of spectral region. Second approach is used for obtaining the informative wavelengths which corresponds to Fusarium cell wall enzymes – chitin and glucan. The results show that these approaches reduce the spectral data from 1050 points to 8 which could be used as a informative features for Fusarium diagnostic.*

**Keywords:** Corn Kernels, Fusarium, NIRS, Wavelet Transformation, Soft Independent Modeling of Class Analogy

#### ВЪВЕДЕНИЕ

Заболяването Фузариум по царевични зърна променя както външната така и вътрешната им структура [5]. По зърната се образува бял до бледорозов налеп, като гладката им повърхност се променя. Намалява влагата в зърната вследствие на което се променя и тяхната форма. В стените на клетките на гъбите от рода Фузариум присъстват ензимите  $\alpha$  – glucan и  $\beta$  – chitin.

Методите, които се използват за оценяване на качествени показатели на селскостопански продукти са базирани основно на анализ на изображения и на спектрален анализ. Анализът на изображения се използва когато има налични видими признаци, характерни за съответния клас продукти [3,7]. Спектралният анализ е приложим в случаите когато се търси връзка между

показатели за качество и химични структури, характерни за определени заболявания както във видимия, така и в близката инфрачервена област на електромагнитния спектър [1,6].

Спектралният анализ е използван за определяне на промените, които настъпват в зърната по време на тяхното съхранение [2]. Спектрални данни се използват и като инструмент за качествена оценка на жито и фураж [8].

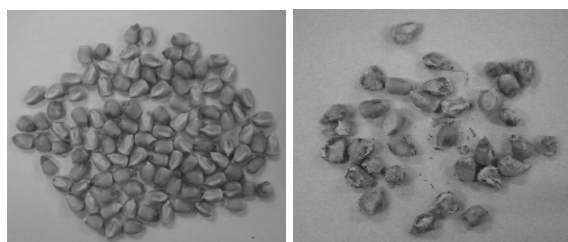
Проблемът при използване на спектрални данни е това, че трябва да се определят първо подходящ диапазон, в който да се снимат тези данни и второ подходящи признаци за всеки от изследваните класове, които да се използват за диагностициране на заболяването по царевичните зърна.

Целта на статията е да се илюстрират два подхода за обработка на спектрални данни

на царевични зърна, които взаимно се допълват и позволяват да се определят информационни характеристики, полезни за разпознаването на заболяването Фузариоза.

## ИЗЛОЖЕНИЕ

Използват се царевични зърна здрави и заразени от Фузариум, изображенията на част от които са представени на фиг. 1.



а) б)  
**Фиг. 1.** Царевични зърна:  
а) здрави, б) заразени от Фузариум

Използвани са 150 здрави и 150 заразени от Фузариум царевични зърна. Всяко зърно се измерва по три пъти. В резултат на това са получени 450 спектрални характеристики за клас здрави и 450 за клас заразени. Данните са разделени в две извадки обучаваща и тестова. Обучаващата се състои от по 300 спектъра за двата класа зърна или общо 600. Тестовата се състои от по 150 спектъра за двата класа или общо 300 спектъра.

За диагностицирането на заболяването се използва информацията от спектралните данни на царевичните зърна. Спектралните им характеристики се снемат като се използва спектрофотометър на NIRSystem 6500. Измерва се спектралното разпределение на химичните компоненти, намиращи се в зърната, като се отчита коефициента на дифузно отражение  $r(\lambda_i)$  в диапазона от 400 до 2500 nm със стъпка 2 nm. Информация за двата класа зърна се съдържа в стойностите на тези коефициенти, които формират спектралната отражателна характеристика  $R\{r(\lambda_i)\}$  на всеки от изследваните обекти.

Спектралните характеристики на царевичните зърна се нормират с цел минимизиране на влиянието на външни “смушчаващи” фактори като естествени биологични различия в обектите, съответно се подобрява точността и надеждността на резултатите:

$$R_n^j(\lambda_i) = \frac{r(\lambda_i)}{r_{max}^j}, \quad (1)$$

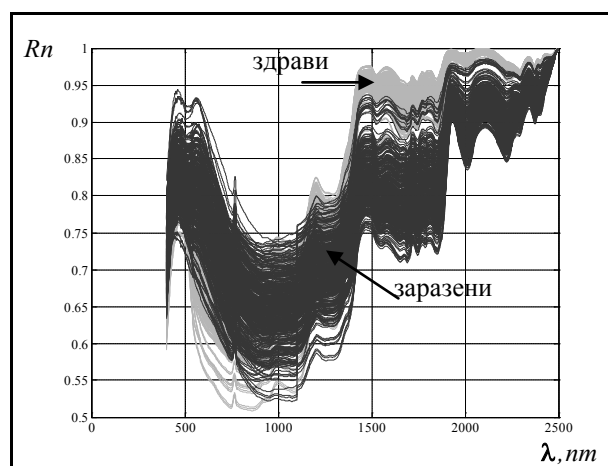
където:

$R_n^j(\lambda_i)$  е спектрална характеристика на царевично зърно с номер  $j$ ;

$r(\lambda_i)$  - коефициента на отражение за дължина на вълната  $\lambda_i$ ;

$r_{max}^j$  - максимална стойност на коефициента на отражение за спектрална характеристика на царевично зърно с номер  $j$ .

Нормираните спектрални характеристики на здрави и заразени царевични зърна са представени на фиг. 2.



**Фиг. 2.** Спектрални характеристики на здрави и заразени царевични зърна

Получените данни показват, че те не могат да бъдат използвани като признак за диагностициране на заболяването без допълнителна обработка. От тях трябва да се отделят информативните участъци от спектъра, както и информативни признаци, по които да се формулират критерии за разпознаването на вида на зърната. Това се налага и от факта, че трябва да се намали големия обем от данни (за всяка характеристика са снети по 1050 точки).

За целта спектралните характеристики се обработват по два подхода.

Първият подход се основава на уейвлет трансформации [4,9]. Преобразуване от семействата на непрекъснатите уейвлет преобразувания се прилага върху спектралните характеристики. Те се описват с уравнение:

$$c_{a,b} = \int Rn(\lambda) \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) dt, \quad (2)$$

където:

$c_{a,b}$  са уейвлет коефициентите;

$a, b$  - променливите за времевата и честотната област;

$\Psi$  - уейвлет функцията.

Използва се функция от семейството на т. нар. “Мексиканска шапка”, описващо се с уравнение:

$$\Psi(x) = \left(\frac{2}{\sqrt{3}} \pi^{-1/4}\right) (1-x^2) e^{-x^2/2}. \quad (3)$$

При това се получава декомпозиране на характеристиката и едновременно ѝ разглеждане във временната и в честотната област. По този начин се получава детайлно представяне на спектралните данни.

Вторият подход за обработка на спектралните данни се основава на метода SIMCA (Soft Independent Modeling of Class Analogy). Този метод се основава на подобие и дава информация за дължините на вълните, на които се базира разграничаването между отделните класове т.нар. способност за разделяне (*discriminating power*).

Чрез използване на втора производна за предварителна обработка на данните могат да се получат характерни дължини на вълните от спектъра, които да покажат има ли наличие на ензимите, присъстващи в стени на клетките за гъбите и да се оцени вида на зърното – заразено от Фузариум или не.

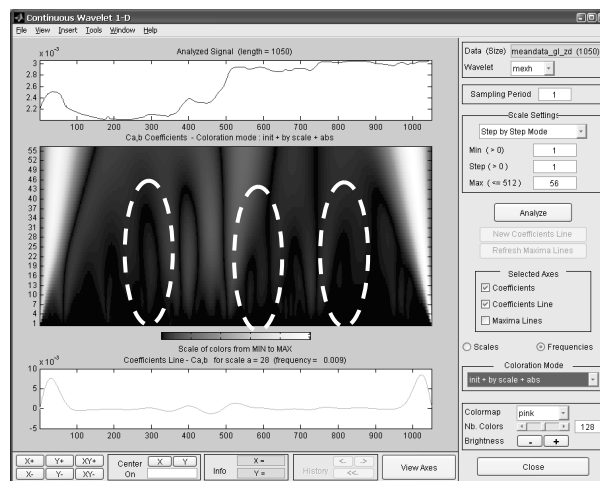
Въз основа на тези дължини на вълните се създават водни индекси, индекси за наличие на Фузариум и др., използващи коефициента на дифузно отражение, които служат като критерии за класификация на зърната.

За прилагане на метода се използва програмата Pirouette 3.1 (Infometrics, Inc., Woodinville, WA, USA), която позволява и създаване на модели за класификация на зърната.

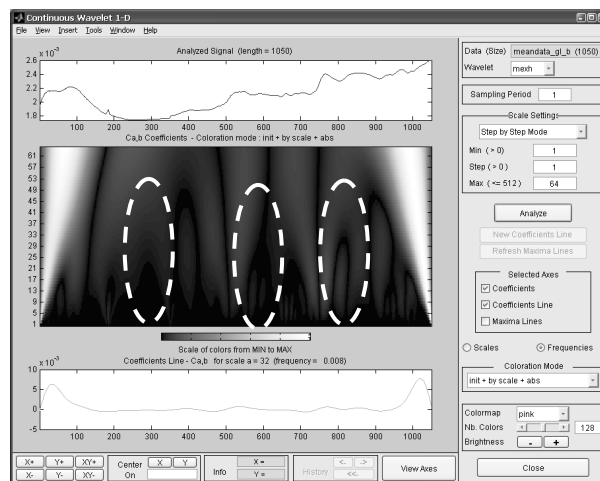
## РЕЗУЛТАТИ

Програмната реализация на уейвлет трансформациите се извършва в средата MATLAB

[10]. Резултатът от преобразуванията на осреднените характеристики за двата класа царевични зърна – здрави и заразени са представени на фиг. 3 и 4.



Фиг. 3. Уейвлет трансформации на спектрални данни на здрави царевични зърна



Фиг. 4. Уейвлет трансформации на спектрални данни на заразени царевични зърна

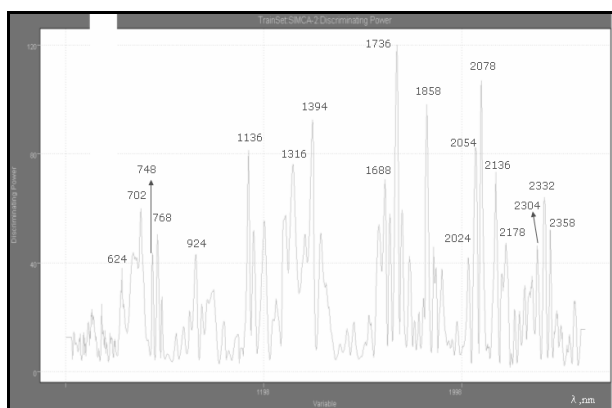
Времева съставка  $b$  (от 1 до 1050) съответства на диапазона на дължината на вълната (от 400 до 2500 nm).

Получените коефициенти за този диапазон показват области от диапазона на спектъра, които се различават за двата класа –  $b=1 \div 114$ ,  $b=250 \div 350$ ,  $b=550 \div 650$  и  $b=750 \div 900$ . На тези области съответстват следните дължини на вълните:  $\lambda = 400 \div 626 \text{ nm}$ ,  $\lambda = 898 \div 1098 \text{ nm}$ ,  $\lambda = 1498 \div 1698 \text{ nm}$  и  $\lambda = 1898 \div 2198 \text{ nm}$ .

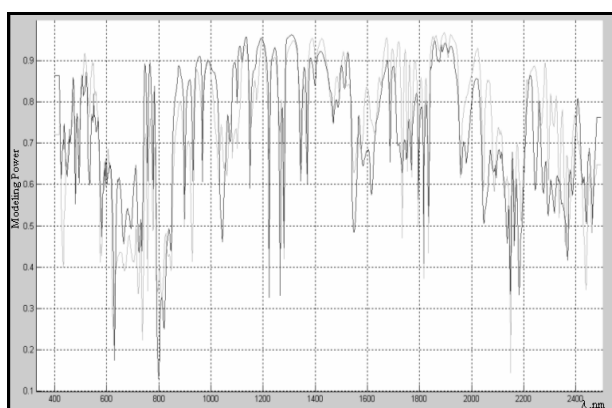
Но конкретна дължина на вълната не може да се посочи.

Резултатите от обработката на спектралните характеристики по метода SIMCA са представени на фиг. 5. Показана е способността за разделяне на двата класа (discriminating power), при която максимумите съответстват на информативни дължини на вълната. Информативни дължини на вълните са: 624, 702, 748, 768, 924, 1136, 1316, 1394, 1688, 1736, 1858, 2024, 2054, 2078, 2136, 2178, 2304, 2332, 2358 nm.

На фиг. 6 е представена способността за моделиране на двата класа здрави и заразени царевични зърна. Този показател дава информация относно областите от спектъра, в които могат да се търсят разлики в двата класа.



Фиг. 5. Способност за разделяне (discriminating power)



Фиг. 6. Способност за моделиране на класовете (modeling power)

Част от дължините на вълните характеризират водната абсорбция на семената (624, 924, 1136 nm). При заболяването розова фузариоза заразените семена са с по – ниско съдържание на вода от здравите, поради което се проявяват именно тези дължини на вълните. Но влагосъдържанието в семената може да се промени под влиянието и на

други фактори, например климатичните условия.

Поради това тези дължини на вълната не са достатъчни за разпознаване на заразени царевични семена. От друга страна останалите дължини на вълните са свързани с химичния състав на въглехидрата *chitin* и полизахарида  $\lambda$  – *glucan*, които присъстват в стените на клетките на гъбите от рода Фузариум. Наличието на такива химични групи в спектъра на семето са показател за инфицирано с фузариоза царевично семе.

При използването на двата подхода последователно може да се определи следното:

- от първия подход, обработката чрез уейвлет трансформациите – областите от спектъра, които са информативни и в които се наблюдават различия в получените коефициенти за двата класа;

- от втория подход, обработката със SIMCA метода – конкретните информативни дължини на вълните от спектъра, които съответстват на водна абсорбция или на ензимите, характерни за наличието на Фузариум в зърната.

От първата област определена от уейвлет преобразуването  $\lambda = 400 \div 626 \text{ nm}$  целесъобразно е да се използва дължината на вълната – 624 nm, определена по втория метод. За областта  $\lambda = 898 \div 1098 \text{ nm}$  - 924 nm;  $\lambda = 1498 \div 1698 \text{ nm}$  - 1688 nm, за  $\lambda = 1898 \div 2198 \text{ nm}$  - 2024, 2054, 2078, 2136, 2178 nm. По този начин данните се редуцират от 1050 стойности на коефициента на отражение до 8, които са информативни и могат да се използват като признаци за диагностициране и като входни данни за класификатори.

## ЗАКЛЮЧЕНИЕ

Получените резултати показват, че при съпоставянето на получените дължини на вълните с областите от спектъра могат да се изберат подходящи, които да бъдат използвани като признаци при диагностицирането на заболяването. По тези дължини на вълните могат да се съставят индекси (водни, за наличие на ензими, вегетативни и др.), а могат да се използват непосредствено и коефициентите на отражение съответстващи на тези дължини на вълните.

## БЛАГОДАРНОСТ

Изследванията са подкрепени по договор № BG051PO001-3.3.04/28, “Подкрепа за развитие на научните кадри в областта на инженерните научни изследвания и иновациите”. Проектът се осъществява с финансовата подкрепа на Оперативна програма “Развитие на човешките ресурси” 2007-2013, съфинансирана от Европейския социален фонд на Европейския съюз“.

## ЛИТЕРАТУРА

- [1] Atzberger C., M. Guerif, F. Baret, W. Werner, (2010), Comparative analysis of three chemometrics techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat, Computers and electronics in agriculture, 73.
- [2] Cassells J., R. Reuss, B. Osborne, I. Wesley, (2007), Near infrared spectroscopic studies of changes in stored grain, Journal of near infrared spectroscopy, 15, pp. 161 – 167.
- [3] Chen X., Yi Xun, W. Li, J. Zhang, (2010), Combining discriminant analysis and neural networks for corn variety identification, Computers and electronics in agriculture, 71.
- [4] Graps A., (1995), An introduction to wavelets, IEEE Computational Science and Engineering, Vol. 2 (2).
- [5] Krska R., P. Schubert-Ullrich, A. Molinelli, M. Sulyok, S. Macdonald, C. Crews, (2008), Mycotoxin analysis: an update, Food additives and contaminants, vol. 25(2), pp. 152 – 163.
- [6] Pimstein A., A. Karnieli, D. Bonfil, (2007), Wheat and maize monitoring based on ground spectral measurements and multivariate data analysis, Journal of Applied Remote Sensing, Vol. 1, pp. 1 – 16.
- [7] Singh C., D. Jayas, J. Paliwal, N. White, (2010), Identification of insect – damaged wheat kernels using short – wave near – infrared hyperspectral and digital colour imaging, Computers and electronics in agriculture, 73.
- [8] Sinnaeve G., J. Herman, V. Baeten, Y. Sadaoui, M. Frankinet, P. Dardenne, (2009), Quality assessment of wheat and forage using diode array NIR instrument on the harvester. Proceeding of NIRS.
- [9] Wu L., Y. Wen, X. Deng, H. Peng, (2009), Identification of weed/corn using BP network based on wavelet features and fractal dimension, Scientific research and essay, Vol. 4(11), pp. 1194 – 1200.
- [10] [www.mathworks.com](http://www.mathworks.com)