

СТАТИСТИЧЕСКИ МОДЕЛИ ЗА ОКАЧЕСТВЯВАНЕ НА КАШКАВАЛ

Петя Велева-Донева¹, Стефка Атанасова¹, Цветелина Георгиева²

¹Тракийски Университет, Студентски град, Стара Загора, 6000, България,

pveleva@uni-sz.bg

²Русенски Университет „Ангел Кънчев”, ул. Сudentcka 8, п.к.7017, Русе, България,

cgeorgieva@uni-ruse.bg

STATISTICAL MODELS FOR EVALUATION OF YELLOW CHEESE QUALITY

¹Trakia University, Student's campus, Stara Zagora, 6000, BULGARIA, e-mail: pveleva@uni-sz.bg,

²University of Ruse, 8 Studentska Street, 7017 Ruse, BULGARIA; e-mail: cgeorgieva@uni-ruse.bg

ABSTRACT

An approach for quality control of cheese based on spectral analysis in the Near Infrared region and the tools of mathematical modeling is presented in this paper. Two types of models (Arithmetic and Geometric) are developed for evaluation of the bacterial status of the product. The obtained results show that the models could be used as a good alternative to traditional methods of analysis, which is an argument in the direction of the economic objectives of the analysis that guarantee an evaluating mechanism aimed at compliance with the European and Bulgarian legislation with EU policy for people's health protection.

Keywords: yellow cheese, NIR spectroscopy, microbiological criteria for food purity, multivariate analysis

ВЪВЕДЕНИЕ

Европейското и българското законодателство се стремят да осигурят регламенти, свързани с общите правила за окачествяване на хранителните продукти в съответните страни-членки. Причината за разработването им е, че някои хранителни продукти, не отговарят на хигиенните правила, свързани с техния произход, съхраняване, транспортиране, здравна маркировка и др., поради което могат да станат причина за застрашаване здравето на хората. Стратегиите на редица държави, свързани с повишаване безопасността на храните, определят суровините и готовите за консумация храни, в които е регистрирано наличие на различни видове бактерии или токсични продукти вследствие на тяхното размножаване като опасни за потребителя [4,7]. В същото време провежданият мониторинг показва, че около 5% от готовите за консумация храни са контаминирани с патогенни микроорганизми. Редица изследователи описват случаи на масови или спорадични отравяния с *Escherichia coli*, *Listeria monocytogenes* и др. [8,10,13].

Стремежът към високо ниво на защита на общественото здраве е една от основните цели на закона за храните, така както е постановен в Регламент (ОЕ) No: 1441/2007, касаещ микробиологичните критерии за хранителните продукти. Според него, храните не трябва да съдържат микроорганизми или техни метаболити в количества, които представляват недопустим риск за човешкото здраве. Регламентът определя задължителен контрол над редица видове патогенни микроорганизми и фиксира критерии и норми за тяхното присъствие в месните и млечните хранителни продукти [3]. Нормите определят правилото „за отсъствие“ на бактерията в 25 g продукт или до 100 CFU/g в края на срока на годност на продукта, в това число микроорганизми от вида *Listeria monocytogenes* и *Escherichia coli*, които са обект на настоящото изследване.

Разглеждани като сложни многокомпонентни системи от гледна точка на техния бактериален статус, млечните продукти могат да бъдат обекти на моделиране и изследване.

Традиционните микробиологични методи за оценка на качествени показатели на млечните продукти се провеждат в лабораторни условия. Те са свързани с употребата на

химически реактиви и субстрати и отнемат от 24 до 72 часа. От една страна идентификацията на микроорганизмите, основаващи се на имунологичен, генетичен, физико-химичен и т.н. анализи, са точни и обективни, но са твърде скъпи и трудоемки [5]. От друга страна покриването на критериите за микробиологична безопасност на храните е обвързано с постоянен контрол върху готовата продукция, което налага разработване на бързи методи, даващи еквивалентни на класическите методи резултати. С тях ще може да се установи степента на настъпилите промени в храните, а от там и тяхната безопасност и срок на годност. Това от своя страна ще позволи навреме да се предприемат съответните корективни действия, насочени към опазване здравето на хората и постигане на съответния социален и икономически ефект за обществото.

Един такъв метод е спектралният анализ в близката инфрачервена област (Near Infrared Spectroscopy – NIRS). Понастоящем редица автори докладват положителни резултати от използването на спектралния анализ за идентификация, класификация и диференциране на различни бактериални видове, както и бактериално замърсяване на храни [2,9,11]. Основните му предимства като бързина, липса на предварителна химична обработка на пробите, възможност за едновременно определяне на няколко компонента в анализираната проба и минитюаризация на измерителната техника го превръщат в добра алтернатива на класическите методи за анализ. Неотменима част от NIRS анализа са статистическите методи за количествен и качествен анализ. Развитието на статистическите методи за анализ позволява разработване на нови ефективни способи за бърза обработка и анализ на данните [6]. Това създава предпоставка показателите да се анализират без условности и неточности, което осигурява експресно и обективно диагностициране и гарантира сигурност по отношение на контрола на качеството на храните без субективната намеса на изследователите [1].

Целта на статията е да се предложат статистически регресионни модели за установяване наличието на различни видове патогенни микроорганизми в кашкавал чрез използване на спектрален анализ в близката инфрачервена област и инструментариума на математическото моделиране.

МАТЕРИАЛИ И МЕТОДИ

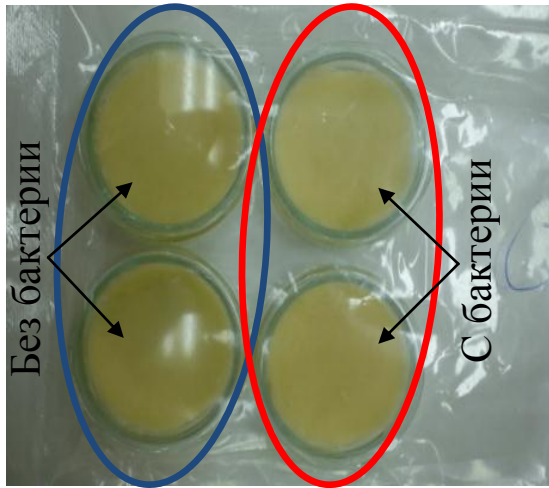
1. Експериментално изследване

За провеждане на експеримента е използван готов за консумация млечен продукт (кашкавал), закупен от търговската мрежа. Пробите са подготвени и контаминирани с бактерии от вида *Listeria monocytogenes* и *Escherichia coli* при спазване на общите правила и изисквания за лабораторни изследвания съгласно БДС 1670-82 (Фиг.1).

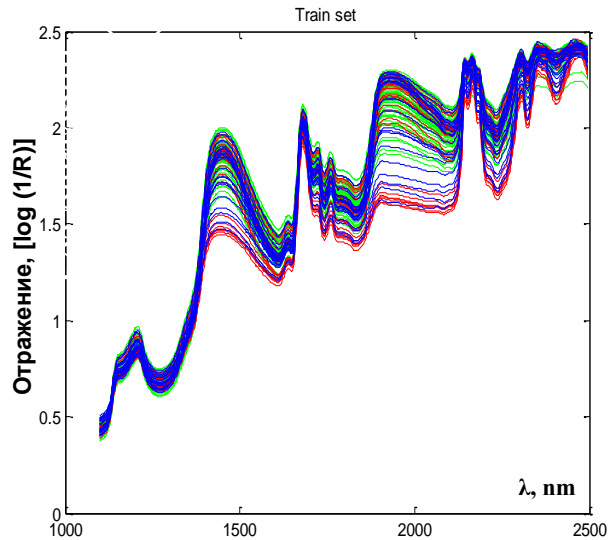
От така подготвения комплект от проби (контрола и контаминирани) през равен интервал от 48 часа за 14 дневен период са вземани проби за спектрално и микробиологично изследване. Всяко измерване включва 2 проби от контролната група и по 2 проби от всяка от групите, контаминирани с даден вид бактерия. От повърхността на всяка проба са измерени спектри на дифузно отражение чрез сканиращ спектрометър NIRSystem 5000 (FOSS NIRSystems, Silver Spring, MD, USA) в диапазон 1100 – 2500 nm със стъпка 2 nm. Непосредствено след спектралното измерване на всяка от пробите е извършен микробиологичен анализ с цел удостоверяване вида на бактериите.

На база на анализа те условно са разделени на класове в два варианта. Първоначално пробите са разделени на два класа: на тези, в които не е установено наличие на патогенни бактерии условно е присвоен клас „Без бактерии“, а на тези, които са контаминирани с патогенни бактерии (*Listeria monocytogenes* или *Escherichia coli*) е присвоен клас „С бактерии“. След това пробите, в които е установено наличието на бактерии т.е. клас „С бактерии“, допълнително са разделени на два класа според вида на бактериите: на тези, които са контаминирани с *L. monocytogenes* е зададен клас „L.Mono“, а на контаминираните

с *Escherichia coli* съответно е зададен клас „*E.Coli*“.



Фиг. 1 Вакуумни опаковки от кашкавал



Фиг. 2 Спектри на проби от кашкавал

Измерените спектри на пробите се представят като $[\log (1/R)]$, където R представлява коефициент на отражение от пробата (Фиг.2). Спектралните характеристики на незаразените проби от клас „Без бактерии“ са построени със зелен цвят, а пробите от класове „*L.Mono*“ съответно с червен цвят и „*E.Coli*“ – със син цвят.

Спектрите са разделени в съотношение 2:1. С две трети от тях са формирани калибровъчни извадки за двата варианта на изследване, а с останалите - тестови извадки. За разработване на моделите са използвани спектрите от калибровъчните извадки. Получените модели са тествани с данните от тестовите извадки.

2. Статистически модели

Тъй като статистическите методи за анализ са теоретично добре изяснени през годините те намират широко приложение при изследване на показатели на хранителни продукти. Моделирането и изследването на качествените показатели на млечни продукти на базата на техните спектри спада към т.нар. класификационни задачи, които от своя страна са частен случай на задачите за разпознаване на образи. В същността си те представляват представяне на изследваните показатели с подходящи характеристики и отнасяне на характеристиките, описващи обектите за класифициране към един или друг предварително зададен клас [14].

От анализа на представените на Фиг.2 спектри се вижда, че спектрите за различните класове са много близки по форма, на места се препокриват и почти е невъзможно да се определят области от спектъра, които да са характерни само за единия или другия клас. Получените спектрални данни се характеризират с много голям обем, а това значително затруднява класификацията им. В такива случаи се търсят вторични признаци за окачествяване, които без загуба или с минимална такава, да описват достатъчно добре обекта на изследване и да позволят обективното му класифициране. Изборът на информативни признаци често е свързан с редуциране на N -мерното пространство на измерените характеристики до ново признаково пространство при запазване на ценността на информацията и разделителните свойства между обектите. При решаване на конкретната задача за редуциране на признаковите пространства и избор на информативни признаци без загуба на значима информация, е използван т. нар. Анализ на Главните компоненти (PCA). Чрез него се получава компресиране на данните и обикновено с около 9-10 фактора (главни

компонента) се описват всички възможни вариации в спектрите.

За обработка на спектралните характеристики и изчисляване на главните компоненти (PC) се използва програмата Pirouette 4.5 (Infometrics, Inc., Woodinville, WA, USA).

В различните фази на растеж, микроорганизмите използват захари, протеини или липиди от дадения продукт, за да се размножават. Биохимичните промени в състава на млечните продукти, както и увеличаването на бактериалната биомаса са в пряка зависимост от времето за съхранението им. Това дава основание връзката между химичните промени в продукта, в резултат на развитието на бактериите и измерените спектрални данни да бъде представена в общ вид със следното равенство:

$$\text{Бактериален статус} = f(\text{спектрални данни}) + E (\text{грешка})$$

Взаимовръзката между факторите x_j , ($j=1,2,\dots,i$) при ($i=5,7,9$) и класа, съответстващ на бактериалния статус (Y) може да се изрази с линеен регресионен модел от вида:

$$Y_j = \alpha + \beta_0 \cdot x_j + \sum_{i=1}^j \beta_i \cdot f(x_{j-i}) + \varepsilon_j, \quad (1)$$

където: Y_j – са стойностите на зависимите променливи (съответния клас);

α, β_0, β_i - са регресионните коефициенти ($i=1,\dots,j$);

$f(x_j)$ – редуцираните спектри до определен брой (j) главни компоненти (независимите променливи).

Прилагането на друг вид функционална зависимост, като напр. логаритмична функция е невъзможно, тъй като някои от факторите имат отрицателни стойности. В настоящото изследване броят на факторите варира съответно: $x_j=5,7,9$. При използване на по-голям брой фактори от посочените, е възможно да настъпи израждане на матрицата на собствените им вектори. Ако зависимостта е непълна, но статистически значима, тогава теоретически е възможно намирането на параметрите на модел (1) по Метода на най-малките квадрати (МНК). Възможно е обаче да се получи изкривяване в оценките на параметрите, а като следствие от това регресионното уравнение (1) да загуби своя познавателен смисъл. За да могат да се намерят стойностите на регресионните коефициенти β_{j-1} ($j=1,2,\dots,i$), $i=5,7,9$ се налага да се търсят други подходи, почиващи на допускания.

В нашия случай се предполага, че факторите са подредени по силата на влияние от по-силния към по-слабия, т.е. x_j, x_{j-1}, \dots, x_1 . При това допускане се предполага, че коефициентът β_0 е най-големият по стойност в сравнение с останалите коефициенти. Очакваното намаление на регресионните коефициенти може да бъде подчинено на различни математически зависимости. С цел оценка влиянието на броя фактори върху точността на модела са направени два вида допускания:

- регресионните коефициенти намаляват в аритметична прогресия (полученият модел за удобство е наричан „Аритметичен“);

- регресионните коефициенти намаляват в геометрична прогресия (моделът е наричан „Геометричен“).

Класификационните модели са получени и при двата варианта разделяне на пробите по класове, а именно: при разделяне на класове „С бактерии“ и „Без бактерии“ и разделяне на класове „L. Mono“ и „E.coli“.

След като се получат статистическите оценки на регресионните коефициенти α, β_0 и β_i , търсеният параметър в неизвестните проби се определя чрез уравнението:

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_0 \cdot x_j + \sum_{i=1}^j \hat{\beta}_i \cdot f(x_{j-i}), \quad (2)$$

където $f(x_{j-i})$ са спектралните данни за неизвестната проба, трансформирани по същия

начин както спектрите на пробите, използвани за получаване на калибровъчното уравнение.

За пресмятане параметрите на моделите за класификация и търсените променливи е използван статистическият пакет STATISTICA 8.0 (StatSoft. Inc.).

Приемането, че всеки един от факторите би могъл да е със най-силна степен на влияние, води до получаване на различни комбинации от регресионни коефициенти, което от своя страна води до получаване на различни стойности на променливите \hat{Y} и при двата вида модели (Аритметичен и Геометричен). От всички възможни комбинации на моделите (условно обозначени с M_1, M_2, \dots, M_i , при $i=5,7,9$) се избират тези, при които общата класификационна грешка (e_o) е най-малка. За класификация на пробите се използват доверителни интервали, получени чрез разпределението на Стюдънт [1]:

$$\hat{Y} - t_\alpha \frac{\sigma}{\sqrt{n}} \leq Y \leq \hat{Y} + t_\alpha \frac{\sigma}{\sqrt{n}}, \text{ като } t_\alpha \text{ - се избира таблично според степените на свобода } (n-1).$$

Според получените стойности на \hat{Y} пробите се отнасят към един или друг от зададените класове при следните доверителни интервали [12]:

- за клас „Без бактерии“ - $\hat{Y} \in (0.5; 1.5)$ и за клас „С бактерии“ - $\hat{Y} \in (1.51; 2.5)$;
- за клас „L. Mono“ - $\hat{Y} \in (1.5; 2.5)$ и за клас „E.coli“ - $\hat{Y} \in (2.51; 3.5)$.

АНАЛИЗ НА РЕЗУЛТАТИТЕ

В изследването са използвани общо 174 проби от кашкавал. Чрез класическите методи за анализ е установено, че в 97 от тях няма наличие на патогенни бактерии. 40 проби са инфектирани с бактерии от вида *L. monocytogenes* и съответно 37 проби са инфектирани с *Escherichia coli*. Резултатите от класификационните процедури със селектираните статистически модели (Аритметични и Геометрични) за калибровъчните извадки от кашкавал са представени в Таблица 1.

Таблица 1. Класификация на калибровъчните извадки проби от кашкавал с избраните статистически модели за анализ при различен брой на главните компоненти

Брой РС	Аритметичен модел				Геометричен модел			
	Без/ С бактерии		L.Mono/E.Coli		Без/ С бактерии		L.Mono/E.Coli	
	Модел	$e_o, \%$	Модел	$e_o, \%$	Модел	$e_o, \%$	Модел	$e_o, \%$
5 РС	M_3	30,36	M_1	30,61	M_2	39,93	-	-
	M_5	29,46						
7 РС	M_1	32,14	M_2	34,69	M_1	32,14	M_2	32,65
	M_2	34,82						
9 РС	M_1	39,28	M_9	36,73	M_2	45,53	-	-

От Таблица 1 се вижда, че най-ниска класификационна грешка при разделяне на пробите на класове „С бактерии“ и „Без бактерии“ за аритметичния модел е получена при модел M_5 с използване на 5 признака ($e_o=29,46\%$), а за геометричния модел е съответно при модел M_1 ($e_o=32,14\%$) с 7РС. При разделяне на пробите на класове „L.Mono“ и „E.Coli“ минималната обща грешка за аритметичния модел е на стойност $e_o=30,61\%$ при използване съответно на 5 признака за класификация. За геометричния модел M_2 грешката съответно е $e_o=32,65\%$ при 7РС. От гледна точка на бързодействието на процедурите е целесъобразно да се изберат моделите с по-малък брой признаци, но окончателният избор е направен след проверка на точността им с контролните извадки.

Резултатите от класификацията на контролните извадки при избраните модели и брой признаци са представени в Таблица 2. При класифициране на пробите от кашкавал в класове „С бактерии“ и „Без бактерии“, въпреки че за геометричния модел M_4 са необходими 7 признака, той работи със значително по-малка грешка в режим на тестване ($e_o=29,03\%$) в

сравнение с аритметичния модел M_5 ($e_o=33,87\%$), за който са необходими само 5 главни компонента. Ако условието за бързодействие на анализа не е от съществена важност, би могъл да се предпочете модел M_1 . При тестване на моделите за видово определяне на бактериите най-малка класификационна грешка е получена за аритметичен модел M_1 с 5PC: $e_o=21,42\%$.

Таблица 2 Класификация на контролните извадки с избраните статистически модели за анализ на проби от кашкавал при различен брой на главните компоненти

Вид модел, брой PC					
клас	Аритм. M_5 (5 PC)	Геом. M_4 (7 PC)	клас	Аритм. M_1 (5 PC)	Геом. M_2 (7 PC)
<i>C/ Без бактерии</i>	$e_o=33,87\%$	$e_o=29,03\%$	<i>L.Mono/ E.Coli</i>	$e_o=21,42\%$	$e_o=32,14\%$

ЗАКЛЮЧЕНИЕ

Разработените математически модели за качествен анализ на млечни продукти биха могли да се използват като добра алтернатива на класическите методи за анализ на бактериален статус. Предложеният подход има универсален характер и би могъл да се използва при оценка на друг вид качествени показатели на млечните продукти, както и за други хранителни продукти.

Получените по този подход резултати биха могли да гарантират сигурност по отношение на контрола на качеството на храните и опазване здравето на хората. Наблюдаваните хранителни продукти, отговаряйки на българските стандарти, ще съответстват и на международните изисквания, което ще доприне за създаването на такъв вътрешен пазар, който ще осигурява висока степен на защита на здравето на хората в страната, а това от своя страна ще позволи тези продукти да навлязат свободно на европейския пазар като желани и качествени хранителни продукти.

ЛИТЕРАТУРА

1. Митков, А., 2001. Теория на експеримента. Библ. за докторанта. ТУ-Пусе, ISBN 978-954-712-474-5
2. Alexandrakis, D., G. Downey, A. Scannell, 2008. Detection and Identification of Bacteria in an Isolated System with Near-Infrared Spectroscopy and Multivariate Analysis., J. Agric. Food Chem., 56(10), 3431–3437
3. Anonymous, 2007. Commission Regulation (EC) 1441/2007 amending Regulation (EC) No 2073/2005 on microbiological criteria for foodstuffs. *Official Journal of the European Union* L 322, 12-29
4. Evers, E., M. Nauta, A. Havelaar and A. Henken, 2002. Quantitative risk assessment of food borne pathogens – a modeling approach. Chapter 4 in “New approaches to food-safety economics”. *Proceedings of the Frontis workshop on new approaches to food-safety economics, Wageningen, the Netherlands*, 35-45
5. Ewerett, D., M. Auty, 2008. Cheese structure and current methods of analysis. *Int. Dairy J.*, 18, 759-773
6. Forina, M., S. Lanteri, M. Casale, 2007. Multivariate calibration. *J. of Chromatography A*, 1158, 61-93
7. Hathaway, S., 1995. Harmonization of international requirements under HACCP based food control systems. *Food Control*, 6 (5), 267-276
8. Hart, J., G. Smith, 2009. Verocytotoxin-producing *Escherichia coli* O157 outbreak in Wrexham, North Wales, *Euro Surveill.* 14 (32), 19300

9. Horvat, K., Zs. Seregely, E. Andrassy, I. Dalmadi, and J. Farkas, 2008. A preliminary study using near infrared spectroscopy to evaluate freshness and detect spoilage in sliced pork meat. *Acta Alimentaria*, 37, 93-102
10. Levine, P., B. Rose, S. Green, G. Ransom, & W. Hill, 2001. Pathogen testing of ready-to-eat meat and poultry products collected at federally inspected establishments in the United States, *J. Food Prot.* 64:1188 -1193
11. Lin, M., M. Mousavi, M. Al-Holy, A. Cavinato, B. Rasco, 2006. Rapid near infrared spectroscopic method for the detection of spoilage in rainbow trout (*Oncorhynchus mykiss*) fillet. *J. Food Sci.*, 71, S18-S23
12. Meilina, H., Sh. Kuroki, B. Jinendra, K. Ikuta, R. Tsenkova, 2009. Double threshold method for mastitis diagnosis based on NIR spectra of raw milk and chemometrics. *Biosystems engineering*, 04, 243-249
13. Schmid, D., E. Gschiel, M. Mann, S. Huhulescu, W. Ruppitsch, G. Bohm, J. Pichler, I. Lederer, G. Hoger S. Heuberger, F. Allerberger, 2007. Outbreak of acute gastroenteritis in an Austrian boarding school, September 2006. *Euro Surveill.*, 12 (3), 224
14. Sun, D. (ed.), 2009. Computer Vision Technology for Food Quality Evaluation. Food Science and Technology, Int. Series