Trakia Journal of Sciences, Vol. 11, Suppl. 1, pp 462-472, 2013

Available online at:
http://www.uni-sz.bg

ISSN 1313-7069 (print)
ISSN 1313-3551 (online)

COMPUTER AIDED DIAGNOSTIC FOR BACTERIAL CONTAMINATION IN DAIRY PRODUCTS USING SPECTRAL DATA PROCESSING AND MULTIVARIATE ANALYSIS

P. Veleva-Doneva*

Department of Informatics and Mathematics, Trakia University, Stara Zagora, Bulgaria

ABSTRACT
The aim of the current research is to suggest clear, precise and quality methods regarding microbiological criteria for evaluating dairy products purity. The statistical models, which have been worked up for the classification of pathogenic microorganisms, are an istrument directed toward fulfilling the economic goals of analysis, which guarantees an evaluative mechanism permitting the maintenance of European and Bulgarian legislation related to European Union policies for the preservation of the public health.

Keywords food product hygiene, microbiological criteria for food purity, spectral and multivariate analysis

INTRODUCTION

European and Bulgarian legislation strives to provide regulations, connected to the general rules for the hygiene of food products in the respective member states. These rules are contained in the European Union's Common Agricultural Policy for the Preservation of Public Health, including Bulgaria as its member state[1]. The reason for this is that some food products can threaten people's health, if specific rules connected with their hygienic production, storage and transportation and health labeling are not maintained.

Thus, a number of countries' strategies connected with increasing food safety have determined raw and ready-for-consumption products, in which traces of a variety of bacteria and toxic elements have been registered due to their multiplication, as dangerous to consumers [1, 2]. At the same time, monitoring indicates that approximately 5% of food ready for consumption is contaminated with pathogenic microorganisms [3, 4]. A number of researchers describe cases of mass or sporadic poisoning with *Escherichia coli* and *Staphylococcus aureus* among others [5, 6, 7].

Striving towards a high level of public health protection is one of the main aims of the food law enacted via Regulation (EC) №1441/2007, related to microbiological criteria for food products. According to the law, food stuffs must not contain microorganisms or their metabolites in quantities, which present an unacceptable risk to human health.

The aforementioned regulation sets up mandatory testing of a series of types of pathogenic microorganisms and fixes criteria and norms for their presence in dairy products [8].
Main bacterial contamination in milk originates from the infected udder, the udder outside and from the equipment used in the milk storage and

***Correspondence to**: *Petya Veleva-Doneva*, *Department of Informatics and Mathematics, Trakia University, Stara Zagora, Bulgaria, tel: 042 699 439; 042 699 442, E-mail: pveleva@uni-sz.bg*

[1] Regulation (EC) №853/2004 of the European Parliament and the Council of the European Union from 29April 2004 and its follow-up regulations 1290/2005 from 21 June 2005 and 119/2010 from 9 February 2010.

processing. Temperature, time and presence of nutrition are the most important factors affecting bacterial growth. Food stuffs produced from raw milk, or milk processed at lower pasteurization temperatures, conceal a serious risk of developing a variety of microorganisms. Raw milk contaminated with pathogenic microorganisms has decreased quality indicators, which leads to a sharp decline in the quality of the dairy products produced, a decrease in their length of freshness and a serious risk to human health.

Raw milk, which does not contain pathogenic bacteria, is called "safe". According to Ordinance № 31 (BG) regarding the maximum acceptable quantity of contaminants in food, the presence of the pathogenic microorganisms *Staphylococcus aureus* and *Streptococcus agalactiae*, which are the subject of the current research, are "not permitted" in milk and dairy products.

Looked at as complex, polycomponent systems from the standpoint of their bacterial status, milk and dairy products may be the subject of modeling and research. The identification of the microorganisms, which cause bacterial infections, as well as the evaluation of milk and dairy product quality, based on microbiological, immunological and other such analyses are precise and objective, but also too laborious, lengthy, expensive and destructive. The necessity exists for faster, reliable methods and models to be developed, which would provide results equivalent to the classic methods, while allowing the testing of the raw materials and food stuffs not only in specialized laboratories, but also on location at the dairy farms, on the various food production lines, etc. That would then allow the respective corrective measures to be taken in a timely manner, directed toward protecting people's health, while attaining the desired socioeconomic effect on the community. These methods may be of varied character, such as complex chemometric methods, multivariate analysis, pattern recognition analysis, etc. They allow precise measurements and a multifaceted, complex influence upon the various factors connected with the purity of food products. The interconnection between quantifying and qualifying factors is established. They assist, at the same time, in the evaluation of the influence of the various microorganisms quickly and adequately.

One such method is spectral analysis in the near-infrared range (Near Infrared Spectroscopy – NIRS). Currently, a number of authors have reported positive results from using spectral analysis in the near-infrared range to ascertain the development of bacteria in various food stuffs [9, 10, 11, 12 and 13]. Its main advantages – speed (the analysis takes approximately one minute to complete), a lack of preliminary chemical preparation of the samples to be analyzed, the possibility to determine a number of components in the analyzed sample at once and the miniaturization of the measuring technology – have made it a good alternative to the classic methods of analysis. An immutable part of NIRS analysis is mathematical methods for quantitative and qualitative analysis. The development of mathematical methods for quality control, as well as computer technology allow the designing of new, effective instruments for quick processing and analysis of the data [14, 15 and 16]. The use of precise mathematical models creates the prerequisite indicators to perform analyses, without conditions and inaccuracies, which ensure speedy and objective diagnostics and guarantees certainty in relation to quality control of food stuffs without the subjective intervention of the researchers [17, 18].

The aim of this article is to suggest statistical regressive models for the ascertaining of the presence of a variety of types of pathogenic microorganisms in cow's milk via the use of spectral analysis in the near-infrared range and the instrument of multivariate analysis.

MATERIALS AND METHODS

Experimental research

Two type of raw milk samples were used in the investigation – quarter foremilk samples and experimental set of milk samples, contaminated with *Staphylococcus aureus* and *Streptococcus agalactiae.*

Udder quarter foremilk samples were collected before morning milking from 37 Holstein dairy cows and were sampled twice a month for a nine months period. Bacteriological analysis of milk was performed by classical microbiological methods for detection of *Staphylococcus aureus*,

Streptococcus agalactiae. Laboratory procedures include subculturing onto selective Baird Parker agar and nonselective blood agar media as well Camp-test and biochemical tests for bacterial identification and differentiation.

Experimental set of milk samples were prepared by contaminating of 9.9 ml milk with 0.1 ml culture media containing from 10^2CFU/ml to 10^{10} CFU/ml concentration of *Staphylococcus aureus* or *Streptococcus agalactiae* bacteria. A total of 35 milk samples were contaminated with *Staphylococus aureus* and 34 samples with *Streptococcus agalactiae*, respectively. All samples were incubated for 24 hours at temperature 5^{o}C before spectral measurement.

Classic microbiological analysis is used as a reference method to evaluate the accuracy of the developed models.

Spectral analysis

Near infrared spectra of all milk samples were obtained by NIRSystem 6500 spectrophotometer (FOSS NIRSystems, Silver Spring, MD, USA) in the spectral region from 600 to 1880 nm with 2 nm interval. The spectral data were collected as absorbance values [log (1/T)], where T = transmittance. Before the spectral analysis, each sample was warmed up to 40^{o}C in a water bath. The same temperature was maintenance during the spectra acquisition.

Spectra of the investigated samples are divided into classes as followed: samples contaminated with pathogenic bacteria (*Staphylococcus aureus* or *Streptococcus agalactiae*) are nominated as *Diseased* class, while the samples in which the presence of pathogenic bacteria is not established are declared as *Healthy* class. After that the milk samples contaminated with pathogenic bacteria (class *Diseased*) are divided additionally into two classes according to the type of bacteria: the samples which are contaminated with *Staphylococcus aureus* are nominated as *Staphylococcus* class, the samples contaminated with *Streptococcus agalactiae* are nominated as *Streptococcus* class.

The measured samples from each class are divided by a 3:1 ratio. Two thirds of them are used to form a calibration sample, while the others are used as test samples. Spectra from the calibration sample are used for the processing of all of the classes described above. The models received are tested against the data from the test sample.

Models for measuring, analysis and evaluation

Normally, when measuring with a spectrophotometer, a vast quantity of measurement data is accumulated (319 samples, for each of which 606 spectral characteristics are received). In order to decrease the data without losing significant information the so-called Principal Component Analysis (PCA) is used. In this manner the data is compressed and, normally, with the use of 10-15 factors (Principal Components), all possible spectral variations are described. The use of that many factors allows the evaluation of their influence on the accuracy of the classification when varying their number. For this report the number of factors varies as follows: $x_j = 7,9,15$.

The program Pirouette 4.5 (Infometrics, Inc., Woodinville, WA, USA) was used for the processing of the spectral characteristics and the calculation of the Principal Components (PC).

During the exponential growth phase, microorganisms utilize sugars, proteins or lipids dissolved in the media to reproduce themselves. Biochemical changes in the foods properties and increasing bacterial biomass are the major consequences during the food storage time. Acquisition of the real bacterial count and biochemical changes in the foods are features related to the food safety and determining food shelf-life [19, 20].

The relationship between chemical changes in dairy products as a result of bacterial growth and the measured spectral data could be presented in a general form by the following relation:

Bacterial status = f (spectral data) + E (error)

In some processes the combined influence of all the factors (Principal Components) observed is of substantial significance to the evaluation of their interdependence, as well as the result of their influence (increasing or decreasing) upon the sought after variables. The inclusion of all factors and the evaluation of their influence on the variables are made more difficult, if their number exceeds more than 5-6, due to their strong interdependence. So, in order to avoid

autocorrelation between them, specific models are sought out.

The interdependence between factors x_j, $(j=1,2,....i)$ where $(i=7,9,15)$ and corresponding to class (*Y*) according to the type of bacteria is represented by a linear regressive model like:

$$Y_j = \alpha + \beta_0.\chi_j + \sum_{i=1}^{j} \beta_i.f\left(\chi_{j-i}\right) + \varepsilon_j, \qquad (1)$$

where:

Y_j – are the values of the dependent variables (the relevant class);

α, β_0, β_i - are the regression coefficients $(i=1,....j)$;

$f\left(\chi_j\right)$ – the reduced spectra to a number of major components (the independent variables).

The application of some other type of functional dependence – for example, logarithmic function – is impossible because some of the factors have negative values.

In the case of more independent variables, which is our case, it is possible that degeneration of the matrix with eigenvectors may occur. If the dependence is not complete, but statistically significant, then, theoretically, it is possible to discover parameters of the model (1) according to the Method of Least Squares [18]. It is possible, however, for a distortion in evaluating the parameters to occur and, as a result, the regressive equation (1) loses its informative meaning. In order to be able to discover the values of the regressive coefficients $\beta_{j-1}(j=1,2,....j), i=7,9,15$ it is necessary to seek out other approaches, which rely upon assumptions [17].

In our case it is assumed that the factors are aligned according to the strength of influence from strongest to weakest, i.e. $x_j, x_{j-1},...,x_1$. With this assumption it is presumed that the coefficient β_0 is greater in value in comparison with the remaining coefficients. The anticipated decrease in the regressive coefficients may be subjected to various mathematical dependencies.

Two approaches are suggested, which form two mathematical models for identifying the presence of pathogenic bacteria in milk.

According to the first model it is assumed that the regressive coefficients $\beta_{j-1}(j=1,2,....j), i=7,9,15$ decrease by arithmetic progression. Assuming that the reduction starts after a certain event, for example $i=1$ and at a certain value of β_1, then the remaining values of the parameters β_i can be calculated at $i=2,3,...,14$ and $k \le 15$ using the formula:

$$\beta_i = \beta_1\left(1-\frac{j-1}{k}\right), \qquad (2)$$

If relation (2) is used, and the regression coefficients are substituted into equation (1), it takes the form:

$$Y = \alpha + \beta_0.x_j + \beta_1\sum_{i=1}^{j}\left[1-\frac{j-1}{k}\right].x_{j-i} + \varepsilon_j, \qquad (3)$$

After labeling with $Z = \sum_{i=1}^{j}\left[1-\frac{j-1}{k}\right].x_{j-i}$ equation (1) at $j=15$ for example becomes:

$$Y = \alpha + \beta_0.x_{15} + \beta_1 Z + \varepsilon_{15}, \qquad (4)$$

With this summary the degree of influence of one of the factors is evaluated. That factor is assumed to be the most influential and assessing the impact of other factors is done by the aggregate variable Zi.

According to the second model they decrease by geometric progression with the regression coefficient β_0 being the largest, followed by β_1. They are subject to the following relation: $\beta_i = \beta_0.q^i$, where the dividend q fulfills the inequality $0<q<1$. Under these conditions, equation (1) can be represented like this:

$$Y_j = \alpha + \beta_0.\chi_j + \sum_{i=0}^{j}\beta_0.q^{i+1}.\chi_{j-1},\ j=1\div k, \qquad (5)$$

To find the numerical values of the parameters α, β_0 and q the Method of least squares is applied again, but after a certain transformation known as "Koik's conversions" [21] to equation (5) it takes the form as a result:

$$Y_{15} = \gamma + q.Y_{14} + \beta_0.x_{15} + \eta_{15}, \text{ where } \gamma = \alpha.(1-q) \qquad (6)$$

Equation (6) is a mixed model of regression and autoregression. It contains 3 regression coefficients γ, β_0 and q, whose numerical values are found using the Method of least squares. Finding q allows to determine all other regression coefficients and to establish the relationship between the factors and the relevant class.

Accepting that each of the factors could be of the greatest level of influence leads to receiving various combinations of regressive coefficients and various values for the variable *Y* in both types of models (arithmetic and geometric). That, in turn, leads to referring the variable *Y* toward one or the other of the conditionally labeled classes (*Diseased* - *Staphylococcus* or *Streptococcus* and *Healthy)* within the framework of the given confidence interval. Of all the possible combinations of regressive coefficients, depending on the various numbers of factors of influence upon the sought after variable *Y*, those in which the percentage of correctly classified samples is highest – according to both models for each separate class – are chosen.

For classification of samples confidence intervals are used, which are obtained by Student distribution [17, 18]:

$$\hat{Y} - t_\alpha \frac{\sigma}{\sqrt{n}} \le Y \le \hat{Y} + t_\alpha \frac{\sigma}{\sqrt{n}}, \qquad (7)$$

where:

- t_α - is selected according to tabular degrees of freedom (*n*-1),

- $\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}{n}}$ - is dispersion;

- *n* – number of observation data (samples).

According to the obtained values of $\hat{Y}$, the samples belong to one or another of the specified classes of these confidence intervals [22]:

- for class *Health* - $\hat{Y} \in (0.5;1.5)$ and for class *Diseased* - $\hat{Y} \in (1.51;2.5)$;

- for class *Staphylococcus* - $\hat{Y} \in (4.00;6.00)$ and for class *Streptococcus* $\hat{Y} \in (6.01;8.00)$.

In order to calculate the parameters of the models for classification and to seek the variables the STATISTICA 8.0 (StatSoft. Inc.) is used.

To provide faster operation of the analysis based on the received models a program code is developed in the environment of MatLab 6.5, which could be used in systems for automatic quality control of dairy products.

RESULTS

This research uses a total of 319 samples of cow milk. Via the classic methods of analysis it was determined that 173 of them were without the presence of pathogenic bacteria. 57 of the milk samples were infected with the *Staphylococus aureus* and, respectively, another 89 samples were infected with *Streptococcus agalactiae*.

The selection of models we have suggested in various combinations (conditionally labeled as M_1, M_2, ... M_i, where *i*=7,9,15) of regressive coefficients is suitable according to the results displayed in the form of a table. In **Table 1** and **Table 2** are displayed the results for the models based on the assumption that the regressive coefficients decrease in arithmetic progression. The results for classes *Diseased* and *Healthy* are in **Table 1** and the results for classes *Staphylococus* and *Streptococcus* are in **Table 2**, respectively. In **Table 3** and Table 4 the results are displayed for the model based on the assumption that the regressive coefficients decrease in geometric progression – once again according to the two variants for division of the data.

It is obvious from the tables that the standard errors are small, i.e. the differences between the real and calculated values of the sought after variable are small.

Being that the experimental data were attained without iteration, in order to establish the adequacy of the model – that is, whether it can be used for a given type of research, the following null hypothesis is used (H_0): "the model is inadequate (does not explain anything)". In other words, the regression causes such dispersion of the calculated values $\bar{Y}_j, j = 1,2,...N$ approximating the average arithmetic $\bar{Y}$, as is the dispersion of the test data $Y_j, j = 1,2,...N$ near $\bar{Y}$ [18].

Table 1. *Statistical model assuming that the coefficients decrease in arithmetic progression for classes* Diseased and Healthy

model	α	β0	β1	R	R2	Sy	Dar-Wat	F
7 factors								
M_1	0,639	0,029	-0,008	0,335	0,112	0,474	0,465	13,734
M_2	0,377	0,057	0,259	0,315	0,099	0,477	0,489	11,995
M_3	1,257	0,518	0,051	0,483	0,233	0,439	0,645	33,204
M_4	0,131	-0,764	0,318	0,527	0,278	0,427	0,58	41,956
M_5	-0,07	-0,982	0,363	0,454	0,206	0,448	0,595	28,266
M_6	0,153	-0,784	0,311	0,384	0,147	0,464	0,526	18,838
M_7	0,36	-0,588	0,262	0,318	0,101	0,477	0,495	12,239
9 factors								
M_1	0,819	0,044	-0,809	0,461	0,213	0,446	0,605	29,425
M_2	1,016	-0,112	0,288	0,422	0,178	0,456	0,533	23,586
M_3	1,305	-0,393	0,127	0,432	0,187	0,453	0,57	25,041
M_4	1,654	1,225	-0,059	0,584	0,34	0,408	0,618	56,271
M_5	1,158	-0,824	0,201	0,396	0,157	0,462	0,502	20,225
M_6	1,292	0,43	0,126	0,252	0,063	0,486	0,437	7,388
M_7	1,302	1,418	0,121	0,273	0,074	0,486	0,439	8,756
M_8	1,273	0,586	0,137	0,247	0,061	0,487	0,43	7,086
M_9	1,279	-1,319	0,133	0,257	0,066	0,486	0,428	7,677
15 factors								
M_1	1,479	-38,68	0,36	0,338	0,114	0,473	0,44	14,083
M_2	1,472	-9,829	0,349	0,305	0,093	0,479	0,474	11,14
M_3	1,475	-18,15	0,367	0,333	0,111	0,474	0,493	13,552
M_4	1,469	-3,656	0,361	0,304	0,093	0,479	0,494	11,136
M_5	1,469	3,103	0,36	0,304	0,092	0,479	0,468	11,072
M_6	1,471	0,462	0,358	0,301	0,09	0,479	0,479	10,828
M_7	1,471	-1,902	0,371	0,308	0,095	0,478	0,465	11,4
M_8	1,468	-0,898	0,354	0,308	0,095	0,478	0,477	11,436
M_9	1,468	-2,441	0,378	0,363	0,132	0,468	0,483	16,569
M_{10}	1,471	-0,428	0,377	0,322	0,104	0,476	0,485	12,645
M_{11}	1,471	-0,488	0,386	0,335	0,112	0,476	0,488	13,75
M_{12}	1,481	-1,198	0,436	0,603	0,363	0,401	0,669	62,246
M_{13}	1,471	0,118	0,382	0,324	0,105	0,476	0,489	12,744
M_{14}	1,465	0,324	0,332	0,451	0,203	0,449	0,595	27,826
M_{15}	1,475	0,028	-0,898	0,364	0,133	0,468	0,45	16,686

Table 2. *Statistical model assuming that the coefficients decrease in arithmetic progression for classes* Streptococcus and Staphylococcus

model	α	β0	β1	R	R2	Sy	Dar-Wat	F
7 factors								
M_1	10,335	-0,091	-0,25	0,314	0,098	1,49	0,455	11,889
M_2	11,633	-0,277	-0,931	0,315	0,099	1,49	0,48	12,031
M_3	8,696	-1,478	-0,243	0,457	0,209	1,392	0,599	28,772
M_4	11,826	2,226	-0,984	0,51	0,26	1,347	0,549	38,323
M_5	12,374	2,718	-1,107	0,433	0,187	1,411	0,551	25,153
M_6	11,506	0,946	-0,902	0,333	0,111	1,476	0,492	13,608
M_7	11,205	3,031	-0,831	0,335	0,112	1,475	0,479	13,798
9 factors								
M_1	9,492	-0,124	2,08	0,413	0,171	1,425	0,544	22,459
M_2	9,109	0,381	-0,9	0,431	0,185	1,413	0,514	24,816
M_3	8,18	0,957	-0,378	0,367	0,134	1,456	0,515	16,918
M_4	7,193	-3,453	0,144	0,532	0,283	1,326	0,56	42,993
M_5	8,544	1,996	-0,562	0,339	0,115	1,473	0,479	14,169
M_6	8,177	-2,646	-0,357	0,274	0,075	1,505	0,43	8,874
M_7	8,156	-6,026	-0,348	0,289	0,084	1,499	0,449	9,943
M_8	8,225	0,44	-0,386	0,235	0,055	1,522	0,431	6,377
M_9	8,259	7,309	0,397	0,27	0,073	1,507	0,424	8,539
15 factors								
M_1	7,667	54,43	-1,076	0,297	0,088	1,495	0,443	10,568
M_2	7,671	77,579	-1,004	0,314	0,098	1,486	0,45	11,909
M_3	7,656	95,604	-1,121	0,375	0,141	1,451	0,492	17,832
M_4	7,687	27,35	-1,092	0,31	0,96	1,488	0,494	11,567
M_5	7,685	-13,12	-1,08	0,295	0,087	1,496	0,45	10,403
M_6	7,658	-26,48	-1,029	0,348	0,121	1,467	0,492	15,03
M_7	7,676	8,214	-1,125	0,302	0,091	1,492	0,445	10,932
M_8	7,69	5,62	-1,049	0,314	0,99	1,486	0,467	11,91
M_9	7,687	9,38	-1,148	0,381	0,145	1,448	0,494	18,469
M_{10}	7,687	1,183	-1,128	0,308	0,095	1,489	0,461	11,435
M_{11}	7,679	0,1	-1,106	0,294	0,087	1,496	0,463	10,348
M_{12}	7,651	3,216	-1,279	0,538	0,289	1,32	0,599	44,385
M_{13}	7,678	-1,122	-1,164	0,308	0,095	1,489	0,462	11,406
M_{14}	7,696	-1,015	-0,979	0,445	0,198	1,402	0,562	26,922
M_{15}	7,669	-0,081	2,003	0,326	0,106	1,48	0,439	12,942

Table 3. *Statistical model assuming that the coefficients decrease in geometric progression for classes* Diseased and Healthy

model	γ	q	βo	R	R2	Sy	Dar-Wat	F
7 factors								
M_1	0,114	0,750	0,009	0,787	0,619	0,310	1,770	176,360
M_2	0,354	0,765	-0,026	0,785	0,616	0,312	1,785	173,811
M_3	0,409	0,722	0,130	0,787	0,620	0,310	1,771	177,026
M_4	0,391	0,736	-0,298	0,791	0,625	0,306	1,774	181,068
M_5	0,338	0,770	-0,197	0,783	0,613	0,313	1,808	171,727
M_6	0,329	0,776	-0,163	0,782	0,611	0,313	1,809	170,579
M_7	0,324	0,779	-0,267	0,781	0,611	0,314	1,801	170,175
9 factors								
M_1	0,258	0,744	0,009	0,788	0,620	0,310	1,765	177,151
M_2	0,322	0,781	6,1E-05	0,781	0,610	0,314	1,808	169,601
M_3	0,370	0,751	-0,095	0,784	0,615	0,312	1,795	173,525
M_4	0,532	0,662	0,438	0,803	0,646	0,299	1,703	197,630
M_5	0,333	0,770	-0,175	0,783	0,614	0,312	1,808	172,373
M_6	0,326	0,778	0,155	0,781	0,611	0,314	1,803	170,210
M_7	0,335	0,773	0,606	0,783	0,613	0,313	1,786	171,833
M_8	0,319	0,782	0,233	0,781	0,610	0,314	1,805	169,853
M_9	0,323	0,779	-0,771	0,782	0,612	0,313	1,799	170,854
15 factors								
M_1	0,336	0,775	-22,59	0,786	0,618	0,311	1,739	175,353
M_2	0,327	0,779	-13,01	0,784	0,614	0,312	1,807	172,581
M_3	0,329	0,777	-5,307	0,782	0,612	0,313	1,796	170,791
M_4	0,310	0,790	5,736	0,784	0,615	0,312	1,799	173,157
M_5	0,318	0,783	3,360	0,783	0,612	0,313	1,778	171,424
M_6	0,322	0,781	0,149	0,781	0,610	0,314	1,809	169,609
M_7	0,320	0,782	-1,464	0,782	0,612	0,313	1,801	171,126
M_8	0,324	0,779	-0,622	0,782	0,611	0,313	1,801	170,668
M_9	0,336	0,771	-0,893	0,784	0,615	0,312	1,777	173,035
M_{10}	0,323	0,780	-0,083	0,781	0,610	0,314	1,809	169,785
M_{11}	0,325	0,779	-0,109	0,781	0,610	0,314	1,808	170,042
M_{12}	0,437	0,705	-0,425	0,795	0,633	0,305	1,770	186,881
M_{13}	0,321	0,781	-0,004	0,781	0,610	0,314	1,808	169,604
M_{14}	0,359	0,754	0,060	0,783	0,614	0,312	1,791	172,481
M_{15}	0,367	0,750	0,008	0,787	0,619	0,310	1,770	176,242

Table 4. *Statistical model assuming that the coefficients decrease in geometric progression for classes* Streptococcus and Staphylococcus

model	γ	q	βo	R	R2	Sy	Dar-Wat	F
7 factors								
M_1	2,654	0,755	-0,027	0,788	0,622	0,965	1,771	178,196
M_2	1,744	0,770	0,074	0,786	0,617	0,970	1,786	175,056
M_3	2,077	0,729	-0,407	0,789	0,623	0,963	1,778	179,533
M_4	1,969	0,743	0,915	0,792	0,628	0,957	1,791	183,106
M_5	1,735	0,774	0,581	0,784	0,615	0,973	1,811	173,490
M_6	1,671	0,783	0,051	0,783	0,613	0,976	1,810	171,552
M_7	1,696	0,779	0,985	0,783	0,614	0,975	1,798	172,350
9 factors								
M_1	2,288	0,749	-0,027	0,789	0,623	0,963	1,766	179,149
M_2	1,692	0,780	0,019	0,783	0,613	0,976	1,808	171,705
M_3	1,800	0,765	0,221	0,785	0,616	0,972	1,804	173,848
M_4	2,303	0,687	-1,230	0,802	0,644	0,936	1,726	196,158
M_5	1,727	0,777	0,411	0,784	0,615	0,973	1,819	173,158
M_6	1,737	0,773	-0,994	0,785	0,616	0,972	1,795	174,118
M_7	1,752	0,771	-2,033	0,785	0,616	0,972	1,793	174,112
M_8	1,669	0,783	-0,052	0,783	0,613	0,976	1,809	171,543
M_9	1,715	0,777	2,910	0,784	0,615	0,973	1,798	173,387
15 factors								
M_1	1,660	0,782	48,363	0,785	0,616	0,972	1,777	174,270
M_2	1,722	0,775	68,836	0,790	0,624	0,961	1,807	180,373
M_3	1,772	0,768	30,424	0,786	0,618	0,969	1,776	175,546
M_4	1,612	0,790	-10,319	0,784	0,614	0,974	1,813	172,712
M_5	1,662	0,784	-14,053	0,786	0,617	0,971	1,787	174,882
M_6	1,763	0,769	-9,657	0,786	0,618	0,970	1,810	175,298
M_7	1,654	0,784	6,312	0,785	0,617	0,972	1,803	174,510
M_8	1,715	0,777	2,277	0,784	0,615	0,974	1,802	173,014
M_9	1,776	0,769	2,643	0,785	0,617	0,971	1,787	174,601
M_{10}	1,677	0,782	0,282	0,783	0,613	0,976	1,808	171,765
M_{11}	1,670	0,783	-0,122	0,783	0,613	0,976	1,808	171,600
M_{12}	2,077	0,728	1,132	0,794	0,630	0,953	1,803	185,112
M_{13}	1,662	0,784	0,080	0,783	0,613	0,976	1,808	171,691
M_{14}	1,913	0,752	-0,228	0,787	0,619	0,969	1,788	175,979
M_{15}	1,874	0,756	-0,025	0,788	0,621	0,965	1,772	178,030

Being that the values of the F-criteria are large (Feq>Ft), the proposed H_0 is rejected and we can accept that the models we worked up in their great variety are adequate and can be used for analysis, evaluation and conclusions regarding the phenomena we studied.

The independent terms (α and γ) in the regressive coefficients of most models are larger than $\alpha = 0{,}05$. This means that their level of significance exceeds the standard error, i.e. they are statistically significant and, therefore the accepted H_0 is rejected. That allows us to conclude that we can use the Method of the Least Squares, the selection of models is suitable and they can be used for the needs of this research.

Of all the combinations of factors of influence upon the sought after variable in the models and for both types of assumptions regarding the regressive coefficients – regardless of whether they correspond to the requirements discussed above, those in which the percent of correctly classified samples were the highest for each separate class were chosen.

The classification results with the most suitable models (conditionally labeled M_1, M_2, ... M_i, where *i*=7,9,15) are presented in **Table 5** where the data are divided into classes *Diseased* and *Healthy* and **Table 6** where the data is divided into classes *Staphylococus* and *Streptococcus* for the calibration and test samples, respectively.

***Table 5**. Classification of calibration and test set by dividing the sample into classes Diseased and Healthy, for both assumptions (arithmetical and geometrical reduction of regression coefficients)*

Sample classes	Number of factors	Arithmetically decreasing coeff.			Geometrically decreasing coeff.		
		% Correct classifications			% Correct classifications		
		Model type	Calibration set	Test set	Model type	Calibration set	Test set
Healthy	7 factors	M_3	74,36 %	36,84 %	M_2	100 %	98,64 %
Diseased		M_3	70,19 %	100 %	M_3	84,26 %	93,18 %
Healthy	9 factors	M_4	72,65 %	68,42 %	M_2	87,15 %	91,04 %
Diseased		M_4	89,42 %	68,09 %	M_8	100 %	97,94%
Healthy	15 factors	M_{12}	69,23 %	57,89 %	M_{10}	72,56 %	70,16 %
Diseased		M_{12}	91,35 %	89,36 %	M_5	74,04 %	78,72 %

***Table 6.** Classification of calibration and test set by dividing the sample into classes Staphylococcus and Streptococcus, for both assumptions (arithmetical and geometrical reduction of regression coefficients)*

Sample classes	Number of factors	Arithmetically decreasing coeff.			Geometrically decreasing coeff.		
		% Correct classifications			% Correct classifications		
		Model type	Calibration set	Test set	Model type	Calibration set	Test set
Staphylococcus	7 factors	M_3	0 %	66,67 %	M_3	92,87 %	98,14 %
Streptococcus		M_1	95,16 %	100 %	M_2	100 %	0 %
Staphylococcus	9 factors	M_7	7,14 %	0 %	M_6	57,14 %	33,33 %
Streptococcus		M_3	100 %	96,25 %	M_2	73,67 %	86,14 %
Staphylococcus	15 factors	M_9	7,14 %	0 %	M_5	28,57 %	16,67 %
Streptococcus		M_{12}	100 %	93,12 %	M_{10}	69,73 %	62,29 %

The results received indicate that the application of some mathematical models ensure 100% accuracy of classification of types of bacteria researched for milk, in both the calibration and test samples. At the same time, they provide a basis to believe that the influence of various types of factors on the classes may be measured, analyzed and evaluated via these models. In terms of faster operation of the analysis by dividing the sample into classes Diseased and Healthy the most appropriate is the geometric model M_2 working with 7 factors and ensuring 100% accuracy of classification for the calibration set and 98.64% accuracy for the test set. When dividing the sample into classes Staphylococcus and Streptococcus the most appropriate is the arithmetic model M_3 working with 9 factors and ensuring 100% accuracy for the calibration set and 96.25% accuracy for the test set, respectively.

CONCLUSIONS

The mathematical models designed to analyze the quality of milk and dairy products could be used as a good alternative to the classic methods for analyzing the development of bacteria in food stuffs. The suggested mathematical models allow quick and accurate evaluation of the classification of various types of bacteria with extremely high precision with every alteration in the input data. The results received via this approach could be able to guarantee certainty regarding the quality control of food and the preservation of public health.

The approach is universal and can be applied to the evaluation of food stuffs from animals. The products, which answer to the criteria in the Regulations, will have no barrier to being sold. The food stuffs observed in accordance with Bulgarian standards will also correspond to international requirements. They will contribute to the creation of a domestic market, which will ensure a high level of preservation of public health in the country. At the same time, the European Union requirements for the preservation of the public health will be upheld and they will be able to freely enter the European market as desirable and quality food products.

REFERENCES

1. Hathaway, S., Harmonization of international requirements under HACCP based food control systems. *Food Control*, 6 (5):267-276, 1995.
2. Evers, E.G., Nauta, M.J., Havelaar, A.H. and Henken, A.M., 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 14 - 17 April 2002:35-45, 2002.
3. Hitchins, A.D., Assessment of alimentary exposure of Listeria monocytogenes. *Int. J. Food Microbiology*, 30:71-85, 1996.
4. Levine, P., Rose, B., Green, S., Ransom, G. and Hill, W., Pathogen testing of ready-to-eat meat and poultry products collected at federally inspected establishments in the United States, 1990-1999. *J. Food Prot.* 64:1188-1193, 2001.
5. Kitamoto, M., Kito, K., Niimi, Y., Shoda, S., Takamura, A., Hiramatsu, T., Akashi, T., Yokoi, Y., Hirano, H., Hosokawa, M., Yamamoto, A., Agata, N., Hamajima, N., Food poisoning by Staphylococcus aureus at a university festival. *Jpn J Infect Dis.* 62 (3):242-243, 2009.
6. Hart, J., Smith, G., Verocytotoxin-producing Escherichia coli O157 outbreak in Wrexham, North Wales, July 2009. Euro Surveill 14 (32):19300, 2009.
7. Schmid, D., Gschiel, E., Mann, M., Huhulescu, S., Ruppitsch, W., Bohm, G., Pichler, J., Lederer, I., Hoger, G., Heuberger, S., Allerberger, F., Outbreak of acute gastroenteritis in an Austrian boarding school, September 2006. Euro Surveill., 12 (3):224, 2007.
8. Annonymous, 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, 2007.
9. Naumann, D., Helm, D. and Labischinski, H., Microbial characterizations by FT-IR spectroscopy, *Nature*, 351:81-82, 1991.
10. Horvat, K., Seregely, Zs., Andrassy, E., Dalmadi, I. and Farkas, J., A preliminary study using near infrared spectroscopy to evaluate freshness and detect spoilage in sliced pork meat. *Acta Alimentaria*, 37:93-102, 2008.
11. Lin, M., Al-Holy, M., Mousavi-Hesary, M., Al-Qadiri, H., Cavinato, A.G. and Rasco, B.A., Rapid and quantitative detection of the microbial spoilage in chicken meat by diffuse reflectance spectroscopy (600–1100 nm). *Letters in Applied Microbiology*, 39:148–155, 2004.
12. Saranwong, S. and Kawano, S., Interpretition of near infrared calibration structure for determinating the total aerobic bacteria count of raw milk: interaction between bacterial metabolites and water absorptions, *J. Near Infrared Spectroscopic*, 16:497-504, 2008.
13. Tsenkova, R., Atanassova, S., Morita, H., Ikuta, K., Toyoda, K., Iordanova, I. and Hakogi, E., Near infrared spectra of cow's milk for milk quality evaluation: disease diagnosis and pathogen identification, *J. Near Infrared Spectroscopic,* 14:363-370, 2006.
14. Forina, M., Lanteri, S., Casale, M., Multivariate calibration. *Jurnal of Chromatography* A, 1158:61-93, 2007.
15. Fagan, C., O' Donell, C., Rudzik, L. and Wust, E., Milk and Dairy Products. In: *Infrared Spectroscopy for Food Quality and Control*. Sun, D.W. (ed.), Elsevier, Amsterdam: 241-273, 2009.
16. Martens, H., Naes, T., Miltivariate Calibration, Wiley, Chichester, 1991.
17. Manov, A., Statistics with SPSS. Sec. ed., Thrakia M, Sofia: 153-162, 2001.
18. Mitkov, A., Theory of the experiment. Library skills. Technical University of Rousse. ISBN 978-954-712-474-5: 39, 2010.
19. Gould, G.W., Methods for preservation and extension of shelf life. *International Journal of Food Microbiology*, 33 (1), 51–64, 1996.
20. Huis, in't Veld, J.H.J., Microbial and biochemical spoilage of foods: an overview. *International Journal of Food Microbiology*, Vol. 33, (1), 1-18, 1996.
21. Eliseeva, I., Kurisheva, S., Kosteeva, T., Econometrics. Finance and Statistics. Moscow, 2003.
22. Meilina, H., Kuroki, Sh., Jinendra, B., Ikuta, K., Tsenkova, R. Double threshold method for mastitis diagnosis based on NIR spectra of raw milk and chemometrics. *Biosystems engineering*, 04, 243-249, 2009.