Predicting the day of storage of dairy products by data combination

M D Vasilev, G I Shivacheva and K I Krastev
Trakia University, Faculty of Technics and Technologies,
8600 Yambol, 38 Graf Ignatiev str., Bulgaria
e-mail: miroslav.vasilev@trakia-uni.bg

Abstract: Existing methods for tracking changes in the quality of dairy products are characterized by difficulty, time consuming, a large number of calculation procedures and resources, which makes them unsuitable for "on-line" monitoring. In the present work, feature vectors containing color components, spectral indices and physicochemical parameters of the products are used. Methods for selection of informative features based on consistently improving assessments have been applied. Models by principal components and latent variables are derived. It has been proven that the models are adequate and can be used to predict the day of storage of yellow cheese and white brined cheese. The advantage of the proposed data processing procedures, comparing them with those reported in the available literature, is that they require less computational resources, which makes them suitable for use in "on-line" monitoring of the condition of dairy products during storage.

Keywords: milk and dairy products, PCA model, feature vectors, residuals analysis, shelf-life, predictive model, model comparison.

1. Introduction

The market requirements for the quality of dairy products are clear that a certain product must have a constant quality, regardless of the manufacturer. This makes on-line monitoring and evaluation techniques necessary [1].

The variety of indicators that determine the suitability of dairy products for storage requires the use of different methods and tools for their evaluation. Both traditional methods and modern non-contact methods can be used for quantitative and qualitative evaluation of these indicators [2].

Traditional methods of analysis have been proven in practice. Monitoring the development of microorganisms harmful to human health during storage of dairy products is the main focus of modern research using laboratory methods, whose prediction accuracy is 67-71% [3].

The difficulty in preparing samples for analysis and the time-consuming nature of traditional methods can be partially offset by modern computer-based methods for obtaining, processing and analyzing data. Through such methods the accuracy of forecasting the condition of dairy products of 99% is achieved [4].

The main disadvantage of these methods is that they require the use of complex calculation procedures, are time consuming, which, like traditional laboratory methods, makes them inapplicable in "on-line" monitoring of the condition of dairy products during storage.

In his study, [5] offers a simplified method for the analysis of meat and dairy products through hyperspectral analysis. This method uses characteristic vectors with values obtained from different
sensors. By combining the data from them, the changes in the quality indicators of the studied dairy products are more effectively monitored, which contributes to a more accurate determination of the day of their storage.

The aim of the present work is to predict the day of storage of dairy products by vectors of features. The proposed methods and tools should be suitable for "on-line" monitoring of the condition of the products.

2. Material and methods
The object of the research in this article are samples of white brined cheese and yellow cheese, produced according to BDS. The dairy products were purchased from the trade network of Yambol, Bulgaria. The samples are stored at room temperature 20-22 °C and relative humidity 45% RH, on wooden racks with perforated polyethylene. The measurement period is 14 days under conditions that do not correspond to those specified by the manufacturer (0-4 °C).

Data based on spectral characteristics, color digital images and physico-chemical parameters of the examined white brined cheese (WBC) and yellow cheese (YC) samples were used in the formation of the feature vectors. Total of 25 features were used. The features are abbreviated as FYCx for yellow cheese and FWCx for white brined cheese (x=[1;25]). The feature vectors are abbreviated as FVYCx for yellow cheese and FVWCx for white brined cheese (x=[1;3]).

Color digital images were obtained using a DFK41AU02 industrial video camera (The Imaging Source Inc.) with a resolution of 1280x960 pixels.

Using a spectrophotometer model USB4000-UV-VIS (Ocean Optics Inc.) spectral characteristics were measured in the visible spectral range with a range of 360-1100nm.

Spectral indices were used according to [6] and [7]. The indices are calculated for reflection spectra.

The selected physico-chemical indicators are the more frequently used quality indicators, which change during storage of the studied dairy products.

Table 1 summarize the features used.

| White brined cheese | Yellow cheese | Feature | White brined cheese | Yellow cheese | Feature |
|---------------------|---------------|---------|---------------------|---------------|---------|
| FWC1                | FYC1          | R (RGB) | FWC14               | FYC14         | G       |
| FWC2                | FYC2          | G (RGB) | FWC15               | FYC15         | NExG    |
| FWC3                | FYC3          | B (RGB) | FWC16               | FYC16         | NGRDI   |
| FWC4                | FYC4          | L (Lab) | FWC17               | FYC17         | RGBVI   |
| FWC5                | FYC5          | a (Lab) | FWC18               | FYC18         | GLI     |
| FWC6                | FYC6          | b (Lab) | FWC19               | FYC19         | VARI    |
| FWC7                | FYC7          | C (CMYK)| FWC20               | FYC20         | ExG     |
| FWC8                | FYC8          | M (CMYK)| FWC21               | FYC21         | DM, %   |
| FWC9                | FYC9          | Y (CMYK)| FWC22               | FYC22         | WC, %   |
| FWC10               | FYC10         | REI     | FWC23               | FYC23         | TA, °T  |
| FWC11               | FYC11         | PTI     | FWC24               | FYC24         | pH      |
| FWC12               | FYC12         | CTI     | FWC25               | FYC25         | EC, mS/cm |
| FWC13               | FYC13         | TVI     | -                   | -             | -       |

The methods used for selection of traits based on consistently improving assessments are [8]:
- Method for ranking significant parameters for forecasting, RELIEFF;
- Method for selection of regression traits by analysis of adjacent components, FSRNCA;
- Selection method by analysis of adjacent components, FSNCA.

Methods used to reduce the amount of data are:
- Principal components obtained by the principal components analysis method, PCA;
Latent variables obtained by the method of partial regression of least squares, PLSR.

To predict the day of storage of the studied products on reduced data, the more common in practice polynomial model of the second order is applied [9].

\[ y = a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 x y + a_6 \]  \hspace{1cm} (1)

The following criteria were used to evaluate the model:

- Coefficient of determination \( R^2 \);
- Fisher's criterion - the obtained criterion needs to be higher than the critical one, reported from table \( F(\alpha, k_1, k_2) >> F_{cr}(\alpha, k_1, k_2) \), where \( \alpha \) is the significance level, \( k_1 \) and \( k_2 \) are degrees of freedom [10];
- P-level, must be less than the accepted significance level \( \alpha \);
- Standard error.

An analysis of the residues is made, which are determined by the difference between the values of the model and the actually measured ones. This assessed the adequacy of the model.

All data were processed at significance level \( \alpha = 0.05 \).

3. Results and discussion

An analysis of six models for each of the products – white brined cheese and yellow cheese. The obtained models are presented in their general form, and an analysis of the residues is also made.

3.1. Data analysis for yellow cheese

Figure 1 shows the results of a selection of features for yellow cheese. The first feature vector is selected by the RELIEFF method. It contains only the R (RGB) component, seven of the spectral indices and all physicochemical parameters without titratable acidity. The second vector, selected by the FSRNCA method, contains five of the color features represented mainly by the RGB and Lab color models. It also includes five of the spectral indices and three of the physicochemical parameters. The third vector, determined by the FSNCA method, contains eight of the color features, except that the b (Lab) color component that is not selected. Three of the physicochemical parameters are also included. It can be seen that no spectral indices were selected.

Figure 1. Results of selection of features for yellow cheese.

| Table 2. Selected feature vectors of yellow cheese. |
|--------------------------------------------------|
| Feature vector | Method | Features |
|----------------|--------|----------|
| FVYC1          | RELIEFF| FYC1, FYC10, FYC11, FYC14, FYC15, FYC16, FYC18, FYC19, FYC21, FYC22, FYC24, FYC25 |
| FVYC2          | FSRNCA | FYC1, FYC2, FYC3, FYC4, FYC8, FYC11, FYC15, FYC16, FYC18, FYC19, FYC21, FYC22, FYC24, FYC25 |
| FVYC3          | FSNCA  | FYC1, FYC2, FYC3, FYC4, FYC5, FYC7, FYC8, FYC9, FYC21, FYC22, FYC25 |
Table 2 shows the resulting vectors of yellow cheese traits, depending on the selection method used. After removing the insignificant coefficients from the basic model, which have p-Value >> α, it was found that the relationship between the day of storage and the first two principal components obtained from the selected feature vectors can be described by the following models:

\[
\begin{align*}
\text{FVYC1} & : D = 3.69 + 8.51.PC_1 - 11.33.PC_2 - 13.2.PC_2^2 \\
\text{FVYC2} & : D = 4.73 + 4.52.PC_1 - 16.36.PC_2 - 19.99.PC_2^2 \\
\text{FVYC3} & : D = 2.16 + 12.78.PC_1 - 15.13.PC_2 + 40.67.PC_1.PC_2
\end{align*}
\]

Table 3 shows the results of a comparative analysis of the obtained regression models by principal components. High values of the standard error are observed. The coefficients of all models are not significant because all p-Values are greater than the accepted significance level α = 0.05. It follows from this analysis that the models obtained by principal components do not describe the experimental data with sufficient accuracy and cannot be used to predict the day of storage of the product.

| Feature vector | Method   | R²  | F(3,17) | p-level | Standard Error |
|----------------|----------|-----|---------|---------|----------------|
| FVYC1          | RELIEFF  | 0.35| 3.05    | <0.06   | 3.78           |
| FVYC2          | FSRNCA   | 0.13| 0.85    | <0.48   | 4.37           |
| FVYC3          | FSNCA    | 0.28| 2.24    | <0.12   | 3.97           |

A comparative analysis of the obtained regression models by latent variables D=f(LV₁, LV₂), which have the form:

\[
\begin{align*}
\text{FVYC1} & : D = 8.21 + 7.76.LV_2 - 0.18.LV_1^2 - 6.83.LV_2^2 \\
\text{FVYC2} & : D = 10.52 - 0.4.LV_1^2 - 1.45.LV_2^2 + 0.9.LV_1.LV_2 \\
\text{FVYC3} & : D = 5.88 - 0.44.LV_1 + 6.78.LV_2 + 4.57.LV_2^2
\end{align*}
\]

Table 4 shows the results of a comparative analysis of the obtained regression models by latent variables. The standard error has the highest values in first and second models, obtained from feature vectors, in selection with the RELIEFF and FSRNCA methods. The coefficients of these two models are not significant because the p-Value is greater than the accepted significance level α = 0.05.

| Feature vector | Method   | R²  | F(3,17) | p-level | Standard Error |
|----------------|----------|-----|---------|---------|----------------|
| FVYC1          | RELIEFF  | 0.15| 0.93    | <0.45   | 4.35           |
| FVYC2          | FSRNCA   | 0.24| 1.79    | <0.19   | 4.09           |
| FVYC3          | FSNCA    | 0.62| 10.15   | <0.00   | 2.57           |

A comparative analysis of the results obtained shows that the model obtained with the FVYC3 feature vector selected by the FSNCA method containing latent variables describes the experimental data with sufficient accuracy. This is evidenced by its regression coefficient, which has the highest value compared to other models. The same trend is observed in the Fisher test, as well as a low value of the standard error.

These criteria are not sufficient to prove the adequacy of the model. An analysis of the residues was also made. Figure 2 shows the obtained model in general form. The first latent variable is plotted on the X axis. On the Y axis, the second latent variable. By Z are the predicted days of storage of the product.
Figure 2. Model of the type $D = F(LV_1, LV_2)$, obtained from FVYC3 data.

Figure 3 shows the results of residue analysis. A sign of the normal distribution of residues is their location in a straight line. From here it can be considered that the prerequisites of the regression analysis are fulfilled. As can be seen from the normal probability graph, the residues are close to the normal distribution. When analyzing the residues, it is established that there are no systematic deviations of the actual data from the theoretical ones, which is a sign of their normal distribution.

![Figure 3. Residue analysis of a yellow cheese model.](image)

The results show that the resulting model is adequate and can be used to predict the day of storage of the product with sufficient accuracy.

3.2. Analysis of white brined cheese data

Figure 4 shows the results of a selection of white brined cheese features. The first feature vector contains only the R (RGB) component. Of the spectral indices, only GLI and ExG were selected for selection. Three of the physicochemical parameters are included. The second vector of features contains six of the
color components, as well as five of the spectral indices, and of the physicochemical parameters, only the electrical conductivity is included. Unlike the previous two feature vectors, in the third only color features are selected, and no Y (CMYK) is selected.

Table 5 shows the resulting white brined cheese feature vectors, depending on the selection method used.

Table 5. Selected feature vectors of white brined cheese.

| Feature vector | Method | Features |
|----------------|--------|----------|
| FVWC1          | RELIEFF| FWC1, FWC10, FWC11, FWC12, FWC13, FWC14, FWC15, FWC16, FWC17, FWC19, FWC23, FWC24, FWC25 |
| FVWC2          | FSRNCA | FWC1, FWC2, FWC4, FWC5, FWC7, FWC8, FWC10, FWC11, FWC16, FWC17, FWC19, FWC25 |
| FVWC3          | FSNCA  | FWC1, FWC2, FWC3, FWC4, FWC5, FWC6, FWC7, FWC8 |

Based on the generalized quadratic model and after removing its insignificant coefficients, at which their p-level of significance is greater than α = 0.05, regression models were obtained for each of the used feature vectors. These models by principal components $D=f(PC_1, PC_2)$ have the form:

FVWC1

$$D = 8,3 + 9,3 \cdot PC_1, PC_2 - 4,45 \cdot PC_2^2$$ (8)

FVWC2

$$D = 18,36 - 49,88 \cdot PC_1 + 11,42 \cdot PC_2 + 9,98 \cdot PC_2^2 + 45,92 \cdot PC_1^2$$ (9)

FVWC3

$$D = 11,52 - 9,45 \cdot PC_1 + 12,74 \cdot PC_2 - 4,73 \cdot PC_1 \cdot PC_2$$ (10)

The obtained models were evaluated according to four criteria. Table 6 shows the results of a comparative analysis of the obtained regression models by principal components. It can be seen that the first and third models have lower values of the regression coefficient, higher standard errors, compared to the second model obtained after reducing the amount of data of the feature vector selected by the FSRNCA method.

Table 6. Comparative analysis of the obtained regression models by principal components.

| Feature vector | Method | $R^2$ | F(3,17) | p-level | Standard Error |
|----------------|--------|-------|---------|---------|----------------|
| FVWC1          | RELIEFF| 0,10  | 0,98    | <0,4    | 4,33           |
| FVWC2          | FSRNCA | 0,80  | 16,099  | <0,00   | 2,16           |
| FVWC3          | FSNCA  | 0,66  | 10,78   | <0,00   | 2,76           |

The obtained regression models by latent variables $D=f(LV_1, LV_2)$ have the form:
Table 7 shows the results of a comparative analysis of the obtained regression models by latent variables. It can be seen that the first and second models have lower values of the regression coefficient, larger standard errors, compared to the third model obtained after reducing the amount of data of the feature vector selected by the FSNCA method.

Table 7. Comparative analysis of the obtained regression models by latent variables.

| Feature vector | Method  | R²   | F(3,17) | p-level | Standard Error |
|----------------|---------|------|---------|---------|----------------|
| FVWC1          | RELIEFF | 0,58 | 7,91    | <0,002  | 3,03           |
| FVWC2          | FSRNCA  | 0,53 | 3,44    | <0,03   | 3,41           |
| FVWC3          | FSNCA   | 0,77 | 18,58   | <0,00   | 2,27           |

A comparative analysis of the results obtained shows that the model obtained with the FVWC2 trait vector selected by the FSRNCA method containing principal components describes the experimental data with sufficient accuracy. This is proved by its regression coefficient, which has the highest value compared to other models. The same trend is observed in the Fisher test, as well as the low value of the standard error.

Figure 5 shows the obtained model in General view. The first main component is plotted on the X axis. On the Y axis, the second main component. By Z are the predicted days of storage of the product.

![Figure 5. Model of type D = F (PC₁, PC₂), obtained from FVWC2 data.](image_url)
are fulfilled. As can be seen from the distribution of the residues and their location around the normal line, in the normal probability graph, they are close to the normal distribution.

![Normal Probability Plot of Residuals](image)

**Figure 6.** Residuals analysis of a model for white brined cheese.

The results show that the resulting model is adequate and can be used to predict the day of storage of the product with sufficient accuracy.

The results obtained in the present work confirm and supplement those reported in the available literature. With the use of the neural network method, an accuracy of 97-99% can be achieved [11] when predicting the shelf life of cheese, at low levels of errors for evaluation of the models developed by them. A disadvantage of the method is that it is difficult to apply in computing devices for "online" monitoring. The use of physicochemical and organoleptic characteristics of cheese as initial data achieves a prediction accuracy of 57-90% [12], but the methods of analysis they apply are time consuming, time consuming and require qualified personnel to use them [13]. These methods are not suitable for on-line monitoring of dairy products.

The models proposed in the present work for forecasting the change of quality indicators of white brined cheese and yellow cheese, depending on the stage of their storage, are entirely aimed at using methods that would be sufficiently effective in terms of quick and simple classification and at the same time give satisfactory accuracy according to the technological requirements for the quality indicators. This is evidenced by the published results in the available literature [5].

### 4. Conclusion

From the comparative analyzes it was found that for the prediction of the day of storage of yellow cheese a feature vector can be used, containing certain color components and physico-chemical indicators of the product. The amount of data is reduced by latent variables. Prediction of the day of white brined cheese storage is possible using a vector of features containing color components, spectral indices and electrical conductivity. The amount of data is reduced by principal components.

It was found that the use of predictive models for white brined cheese ($R^2 = 0.80$) gives better results in predicting the day of its storage than for yellow cheese ($R^2 = 0.62$).

The proposed methods and tools require fewer computational procedures than the use of neural networks and, which gives them the potential for application in on-line monitoring of the condition of dairy products during storage.

### Acknowledgments

This study was supported by Grant 1.FTT/2020, “International Conference on Technics, Technologies and Education ICTTE 2020” from the Faculty of Technics and Technologies – Yambol, Trakia University – Stara Zagora, Bulgaria.
5. References

[1] Zlatev Z, Georgieva-Nikolova M, Genchev A and Lukanov H 2020 A non-contact measuring device for determining poultry eggs weight J. Cent. Eur. Agric. JCEA 21 2 pp 222-230

[2] Van Boekel M, Steeg P and Dahoe A 2020 Co-optimization of safety, quality and legislation: opening Pandora’s box? Curr. Opin. Food Sci. 35 pp 65-71

[3] Martinez-Rios V, Gkogka E and Dalgaard P 2020 Predicting growth of Listeria monocytogenes at dynamic conditions during manufacturing, ripening and storage of cheeses – Evaluation and application of models Food Microbiol. 92 103578

[4] Goyal S and Goyal G 2013 Machine learning models for predicting shelf life of processed cheese Int. J. Open Inf. Technol. 1 7 pp 1-4

[5] Mladenov M, 2020 Model-based approach for assessment of freshness and safety of meat and dairy products using a simple method for hyperspectral analysis J. Food Nutr. Res. 59, 2, pp 108-119

[6] Cermakova I, Komarkova J, and Sedlak P, 2019 Calculation of Visible Spectral Indices from UAV-Based Data: Small Water Bodies Monitoring 14th Iberian Conference on Information Systems and Technologies CISTI pp 1-5

[7] Atanassova S, Nikolov P, Valchev N, Masheva S and Yorgov D 2019 Early detection of powdery mildew (Podosphaera xanthii) on cucumber leaves based on visible and near-infrared spectroscopy AIP Conf. Proc. 2075 pp 160014-1-160014-5

[8] Xu J-L, Esquerre C, Sun D-W 2018 Methods for performing dimensionality reduction in hyperspectral image classification J. Near Infrared Spec. 26 1 pp 61-75

[9] Stefanova R, Georgieva A, Krastev K 2017 Investigation of the biologically active substances contained in forest fruits for the purpose of using them as a raw material for the production of functional beverages J. Appl. Res. in Technics, Technologies and Education 5 4 pp 312-319

[10] Mitkov A, 2011 A theory of experiment ISBN 978-954-712-474-5

[11] Goyal S, Goyal G 2012 Evaluation of shelf life of processed cheese by implementing neural computing models Int. j. interact. multimedi. artif. intell. 1 5 pp 61-64

[12] Park J, Shin J, Bak D, Kim N, Lim K, Yang C and Kim J 2014 Determination of shelf life for butter and cheese products in actual and accelerated conditions, Korean J. Food Sci. An., 34 pp 245-251

[13] Baycheva, S 2016 Application of devices of measurement of colour in analysis of food products J. of Innov. and entrepreneurship 4 4 pp 43-59