THESES OF PHD DISSERTATION

István Kertész

Budapest

2021

Hungarian University of Agriculture and Life Sciences

Mathematical Modelling of Food Quality Parameters Based on Multivariate and Spectral Data

István Kertész

Budapest

2021

A doktori iskola megnevezése: Élelmiszertudományi Doktori Iskola

Tudományága: Élelmiszertudományok

vezetője: Simonné Prof. Dr. Sarkadi Livia
Egyetemi tanár, DSc
Magyar Agrár- és Élettudományi Egyetem
Élelmiszertudományi Intézet
Élelmiszerkémiai és Táplálkozástudományi Tanszék

Témavezetők: Dr. Felföldi József
Egyetemi tanár, PhD
Magyar Agrár- és Élettudományi Egyetem
Élelmiszertudományi Intézet
Élelmiszerpari Méréstechnika és Automatizálás Tanszék

Dr. Baranyai László
Egyetemi tanár, PhD
Magyar Agrár- és Élettudományi Egyetem
Élelmiszertudományi Intézet
Élelmiszerpari Méréstechnika és Automatizálás Tanszék

....

Az iskolavezető jóváhagyása A témavezetők jóváhagyása

1. Introduction and objectives

The quality of edible products has been an important factor since the beginning of human history, but the means of quality inspection have since evolved tremendously. One of the key aspects of methods for quality measurement is objectivity, making sure that standardized comparisons are possible irrespective of time and location, preferably with the minimization of the effect of measurement on the inspected product itself. This led to the rise of non-destructive testing techniques, or NDT for short. Some of these techniques capitalize on mechanical oscillatory properties, such as acoustic and ultrasonic response of the material.

Objectives

The primary goal was to find a mathematical procedure for estimation of different quality attributes of complex materials, such as food products, based on non-destructive testing data. In my thesis, I focus on the application of various multivariate analysis methods to extract information on the attributes of different food products corresponding to quality properties.

The experiments presented in the following thesis were conducted with the following objectives:

1. Developing a signal analysis process for extraction of the most amount of information from spectral data, primarily ultrasound signals, which is fast and applicable in an industrial setting.
2. Finding an apt experimental setup as well as calibration and estimation protocols for indirect viscosity measurement of enzymatic curdling of milk, measured with ultrasound.
3. Developing an experimental setup for eggshell crack detection based on acoustic response signals, and development of a procedure for accurate classification from an industrial standpoint.
4. Evaluation of the developed signal analysis protocol with a multiclass differentiation problem with a food product.

2. Materials and Methods

2.1. Signal form and analysis

For spectral analysis in the experiments presented Fast Fourier Transformation (FFT), and Continuous Wavelet Transformations (CWT) were used. In the ultrasonic measurements, a chirp signal was applied, which is a linearly sweeping wave signal (in this case between 50-450 kHz) enveloped by a Hanning Filter. An example of the input and response signal is shown in figure 1. (unnormalized and denoised).

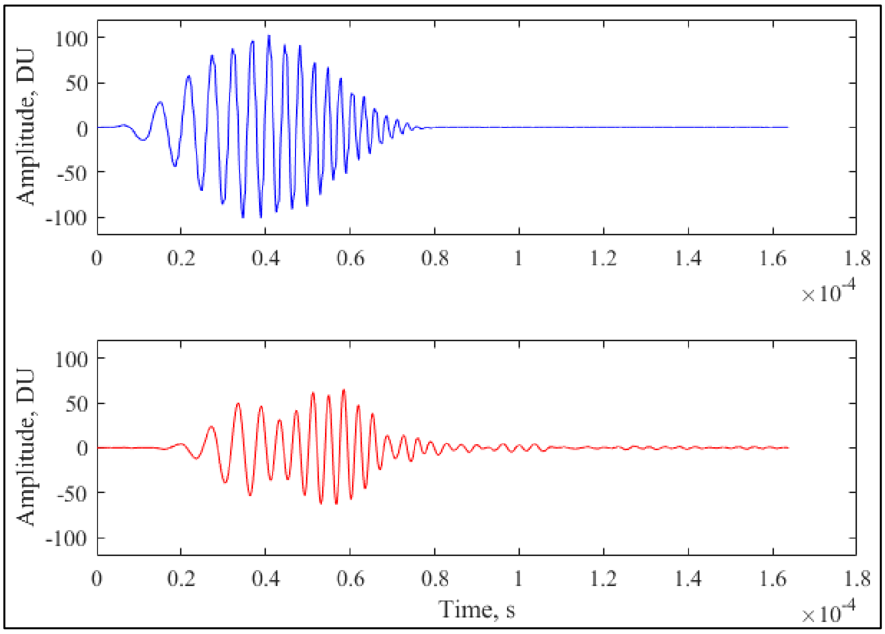

Figure 1. Example input (a) and response (b) signals

2.2. Milk clotting measurements

Altogether, 13 measurements were conducted, at three different milk fat contents, 1.5%, 2.8%, and 3.5%. A transducer-receiver couple was submerged in a beaker containing the milk sample, measurements were conducted for 120 minutes at 38°C, after addition of rennet enzyme.

Measurements for the viscosity change during the milk clotting experiments were carried out with a HAAKE RotoVisco 1 (Haake Technik GmbH, Germany) rotational viscometer, at 38°C.

The trend in the viscosity curves followed an inverse negative exponential equation, which was fitted for all measurements.

Goodness of fit was characterized with the adjusted coefficient of determination (R^2_{adj}), and their values for the measurements are shown in table 1.

Table 1. R^2_{adj} values for viscosity estimations

fat content (m/m%)	$\mathbf{R^2_{adj}}$
1.5	0.9985
	0.9948
	0.9980
	0.9792
	0.9984
2.8	0.9862
	0.9510
	0.9973
	0.9858
3.5	0.9869
	0.7251
	0.9101
	0.9917

The resulting curves were further used for estimation by the ultrasonic response signals.

Experiments were carried out with a piezoelectric ultrasonic transducer-receiver couple with a nominal frequency of 250 kHz (The Ultran Group, USA). To capture the input and response signals, a Velleman PCSGU250 Oscilloscope (Velleman NV, Belgium), datapoints were recorded every minute for 120 minutes. Further analysis was carried out in Matlab 2017a (The Mathworks Inc., USA). Signals were denoised, adjusted in time to cancel the effects of shifting, and normalized.

2.3. Estimation of viscosity from ultrasound signals

For estimation of the fitted viscosity curves by the ultrasonic signals, CWT was carried out. The variance of the individual coefficients were calculated in time, the vector field of the variances were calculated and the peaks of the field were identified. It was assumed that these wavelet coefficients are descriptive of the neighboring coefficients, making covariance calculation obsolete. The number of the extracted coefficients was between 7 and 32. The extracted coefficients were used as predictor variables for Partial Least Squares (PLS) regression. The optimum number of latent variables in the PLS model was selected according to the lowest validated root mean squared error (RMSE) value.

2.4. Cheese classification

In a preliminary experiment, 60 samples of sliced mozzarella cheese (from Szarvasi Mozzarella Kft., Hungary), smoked and untreated were compared in order to test new methods for calculation of the Time-of-Flight (TOF) parameter. TOF was calculated with three different approaches: cross-correlation (XC) method, the Short-Time Average/Long-Time Average (STA/LTA) and the Autoregressive Akaike Information Criterion Picker (AR-AIC) methods, the latter two are used in earthquake detection. At the same time, FFT and CWT coefficients were used for classification, with 66 and 65 coefficients, respectively. The classifiers tested included an array of linear and non-linear methods: Decision trees, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), K-Nearest Neighbors Classifier (KNN), and different ensemble methods.

In the next experiment three classes were tested normal, smoked and lactose-free variants, as the previous experiment did not yield satisfying results. On this occasion, AR-AIC and XC method were tested for TOF calculation, FFT and CWT coefficients were used for classification as well, calculated from the response signal and the XC curves. This time, a variable extraction protocol described in Section 2.3 was utilized to gather the most relevant coefficients for prediction.

2.5. Eggshell crack detection

Acoustic measurements were conducted through a course of six weeks on medium-sized eggs. 20 eggs were weekly, a very small, most of the time invisibly thin microcrack was inflicted on the pointy tip. Acoustic testing was done before and after this destructive action, totaling in 705 measurements. The test itself was done by excitation with a hollow metal rod, meaning a single light knock on the shell in an upright (North-South, NS) and a laid down (East-West, EW). For stabilization and noise reduction, in both positions, the samples were placed on a foam pad, hollowed out for insertion of a microphone. The recorded sound files were processed with a program compiled in Matlab 2017a environment. FFT of the signals were calculated with different window sizes, and padding zeros to assess the effect of increasing resolution. During the assay, I utilized LDA, QDA, SVM, KNN and Ensemble classifiers for crack detection.

3. Results and Discussion

3.1. Milk clotting measurements

PLS regression estimations were carried out in three arrangements of data:

1. Estimation of viscosity for individual experimental runs
2. Estimation of viscosity for different fat contents with joint measurement datasets
3. Estimation of viscosity for all measurements at once

The results of the evaluation of the first set (all samples modeled individually) are shown in table 2. with cross-validated R^2_{adj} and Residual Prediction Deviation values included.

Table 2. Coefficients of determination and Residual Prediction Deviation values of the models

Fat content (m/m%)	R^2_{adj}	RPD
1.5	0.9944	5.07
	0.9697	4.49
	0.9876	6.96
	0.9966	10.88
2.8	0.9688	5.20
	0.9758	4.53
	0.9983	14.22
	0.9877	6.23
3.5	0.9632	4.38
	0.9943	10.09
	0.9973	13.99
	0.9684	4.71
	0.9910	8.27

GOF and accuracy statistics of the second set of models is shown in table 3.

Table 3. Coefficients of determination and Residual Prediction Deviation values at the lowest mean squared error values for prediction

fat content (m/m%)	R^2_{adj}	RPD
1.5	0.8929	2.72
2.8	0.9464	2.73
3.5	0.8606	2.51

For the last, third set, 135 latent variables returned an R^2_{adj} of 0.9708 and an RPD of 5.85, which means a very good, highly explanatory model with a high prediction accuracy.

Estimation for the third set is shown in figure 4.

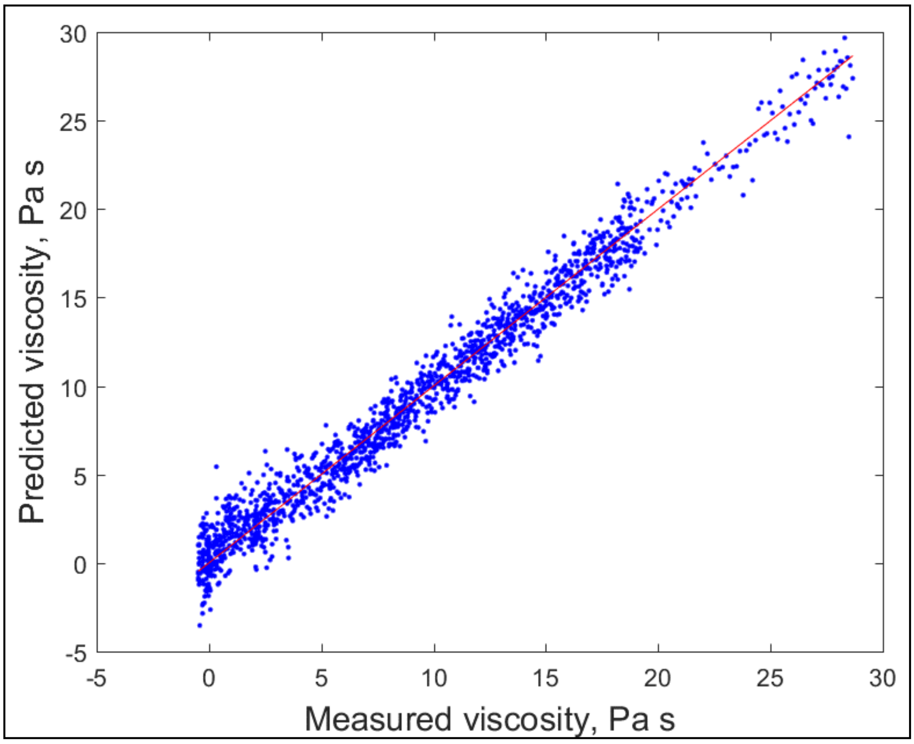

Figure 4. Model prediction with 135 latent variables

3.2. Cheese classification

The total misclassification errors for the preliminary experiments were high, but showed that FFT and CWT should be used. Table 4. shows the maximum correct classification rates for the different post-processing algorithms (all validated results).

Table 4. Correct classification rates for different post-processing methods

	response-FFT	**XC-FFT**	**response-CWT**	**XC-CWT**
Tree	97.8%	55.6%	97.8%	95.6%
LDA	96.7%	54.4%	98.9%	96.7%
QDA	84.4%	52.2%	98.9%	97.8%
SVM	96.7%	61.1%	98.9%	97.8%
KNN	97.8%	56.7%	100.0%	97.8%
Ensemble	97.8%	55.6%	100.0%	97.8%

3.3. Eggshell crack detection

First, a general prediction was done with all the settings to get an overall picture of which methods should be taken into consideration for later calculations. Results for the three performance indicators are shown in tables 5., 6. and 7.

Table 5. Group prediction accuracy (%) of different classification methods. Intensity of the color corresponds to better prediction accuracy

	Multiplier	**2^1**	**2^2**	**2^3**	**2^1**	**2^2**	**2^3**	**2^1**	**2^2**	**2^3**
	Window size		**2^{12}**			**2^{13}**			**2^{14}**	
Classifier	LDA	94.8	94.2	94.5	94.0	94.6	94.6	94.6	95.0	94.7
	QDA	95.5	95.0	94.3	96.3	95.6	95.6	97.2	96.2	97.0
	SVM	95.2	95.9	95.2	95.9	95.3	95.6	96.3	96.3	96.0
	KNN	93.2	92.8	93.3	94.7	94.2	94.9	94.9	94.9	94.7
	Ensemble	94.5	92.5	94.5	94.6	94.6	94.5	94.5	95.0	94.7

Table 6. False positives for crack presence (%) of different classification methods. Intensity of the color corresponds to better prediction accuracy

	Multiplier	**2^1**	**2^2**	**2^3**	**2^1**	**2^2**	**2^3**	**2^1**	**2^2**	**2^3**
	Window size		**2^{12}**			**2^{13}**			**2^{14}**	
Classifier	LDA	6	6	6	6	6	6	6	5	6
	QDA	3	3	3	3	3	3	2	3	2
	SVM	5	4	5	3	5	4	3	3	3
	KNN	6	6	6	5	5	4	4	5	5
	Ensemble	4	5	6	5	6	6	6	5	6

Table 7. Speed of classification (s^{-1}) of different classification methods. Intensity of the color corresponds to higher classification speed

	Multiplier	**2^1**	**2^2**	**2^3**	**2^1**	**2^2**	**2^3**	**2^1**	**2^2**	**2^3**
	Window size	**2^{12}**			**2^{13}**			**2^{14}**		
Classifier	LDA	610	450	130	430	250	220	170	160	91
	QDA	610	500	270	410	210	230	210	100	92
	SVM	780	270	140	480	220	270	260	150	95
	KNN	750	380	140	470	270	260	260	100	91
	Ensemble	460	490	180	300	200	220	120	88	63

Window size contributed to the classification performance, but the multiplier value did not. Speed of classification shows very high values, but they decline rapidly with multiplier and window size as well. QDA and SVM classifiers showed the best performance, the performance difference between them becomes evident when the false positive rates are interpreted correctly: for two faulty eggs missed by QDA there is a third missed by SVM. This makes it clear that further only the QDA classifier should be used. After verification of the applicability of QDA, the method should be rigorously tested.

This in the end resulted in 2.1% prediction error for the total number of misclassifications and 0.87% for false positives with 40 LV, which accounted for 99.6% explained variance, as shown in table 8.

Table 8. Classification matrix of the Quadratic Discriminant Analysis algorithm with 40 latent variables.

		true	
		intact	cracked
predicted	intact	566	6
	cracked	9	112

4. Conclusions, recommendations

The developed wavelet-based variable reduction algorithm has proven to be effective, for different materials, classification and estimation of physical change over time as well. The least amount of corrections was applied in the presented experiments in order to prove the robustness of the algorithm itself, but in real applications, accuracy of a highly adapted method is more important than general applicability, and hence corrections are desired.

The setup used for the clotting experiment includes a submerged through-transmission mode pulser-receiver couple, but it might be interesting to test whether a pitch-catch mode setup might be suitable, as the change in viscosity probably affects the waves traveling on the surface of the curdling liquid, and changes in the matrix might affect the signals significantly. Also, pitch-catch mode can be used in a non-contact setup, which is preferred over immersion setups. The proposed ultrasonic method is suitable, applicable, and recommended for industrial use in different fields of the food industry with adaptation to the technology at hand.. In theory the use of the method can be extended for different products such as yoghurt, sour cream, puddings, *etc.*, to monitor texture properties in phase changing processes.

Egg crack detection resulted in a very high validated accuracy based only on the FFT spectrum magnitude values used as predictor variables. In an industrial setting, the most problematic aspects are external sources of noise and automation of a steady excitation process. The problem of noises can be addressed by using a tunnel with quasi-soundproof insulation, also, the use of a high-sensitivity lobar polar pattern condenser microphone is suggested. The main purpose of the foam pad in the setup is to minimize the conduction of ambient vibrations, but other, nonrigid materials may also be suitable for supporting the samples.

5. New scientific results

1. I have developed a novel post-processing method for variable reduction of multidomain signals, and the algorithm has been tested on two distinct, independent materials for estimation and classification purposes.

   The process is based on continuous wavelet transformation of sound signals from time domain to a set of joint time and spectrum domain matrices, using Morse wavelet. These matrices consisting of wavelet coefficients are analyzed across observations for the highest peaks in standard deviation, which thereby represent the coefficients carrying the highest amount of information on the differences between samples. The identified coefficients are extracted from each observation, significantly reducing the number of variables used for further analysis, which would be infeasible with the original matrices due to the vast number of coefficients.

   The proposed post-processing method, in theory, can be applied to other multidomain data, where domains contain continuously changing variables, and therefore peaks can effectively represent nearby data points, such as in spatial or spectral data.

2. An ultrasonic measurement setup was developed for the determination of viscosity of milk during enzymatic curdling, and therefore the state of the curdling process. The setup consists of an immersed through-transmission mode piezoelectric transducer-receiver couple with 250 kHz nominal frequency and an oscilloscope with $4 \cdot 10^{-8}$ second resolution. The signal form is a linear chirp with a frequency sweep across 50-450 kHz and enveloped with a Hanning filter.

   A new calibration process was developed for ultrasonic viscosity measurement during clotting of milk. The procedure includes estimation of the viscosity values during enzymatic curdling of cheese by an inverse negative exponential curve (R^2: 0.7251 – 0.9985), and estimation of the modeled values by the proposed algorithm. The ultrasonic signals were normalized, de-noised and put through wavelet-decomposition, relevant peaks in the standard deviation matrix of the wavelet coefficients representing high amount of the information in the signal are extracted from the wavelet transforms of each measurement and used for prediction of viscosity by partial least squares regression. The calibration process yielded high adjusted coefficient of determination values (0.9632-0.9983) and residual prediction deviation (4.38-14.22), signifying that the process is capable of high accuracy estimation.

A new process was developed for calibration of indirect ultrasonic measurement of viscosity of milk during enzymatic curdling with fat contents between 1.5-3.5%. The calibration was carried out for responses of measurements with a 50-450 kHz linear ultrasonic chirp signal, using a 250kHz immersion through-transmission mode piezoelectric transducer-receiver couple. Analysis of the normalized and denoised signals was done according to the post-processing algorithm proposed and resulted in high prediction ability for viscosity in terms of precision (R^2 = 0.9708) and accuracy (RPD = 5.85).

3. A new setup and acoustic method for eggshell crack detection were developed. The setup consists of a soft foam pad with a built-in high-sensitivity microphone, and a hollow metal rod for a single-hit excitation on a horizontally aligned egg in the equatorial line, in 0° direction from the microphone. The sound of the acoustic response is recorded at 96000 Hz sampling frequency.

   Faulty eggs with hairline cracks were identified by quadratic discriminant analysis at 2.1% total error rate and a 0.87% false negative rate at 40-fold cross-validation. The predictor variables for the discriminant analysis were Fourier coefficients of the acoustic response of eggs to a single hit excitation in a laid down position. The coefficients were obtained using increased precision by padding the sound signals with zeros for double-length and using a window size of 2^{14} points.

4. The proposed new variable reduction method was shown to be applicable with great success for classification in a multiclass classification problem. The applicability was demonstrated on three different types of sliced cheese samples (normal, smoked, lactose free), using only the sound signals of ultrasonic measurements with a chirp signal in through-transmission configuration. Two of the three groups were identical in composition, all three tested different for thickness and were different in treatment. With 5-fold cross-correlation, the proposed post-processing method applied to the ultrasonic response signals yielded a 100% correct classification using a K-nearest neighbor classifier, compared to 97.8% based on the Fourier coefficients of the signals, providing evidence that the method increases prediction accuracy. Also, the algorithm dramatically increased classification accuracy compared to the predictions based on Fourier coefficients, when the cross-correlate of the input and response signals were processed instead of only the response signals.

6. List of publications in the field of studies

Kertész, I., Nagy, D., Baranyai, L., Pásztor-Huszár, K., Varsányi, K., Nguyen, L. P. L., Felföldi, J. (2021a): Development of a novel ultrasonic spectroscopy method for estimation of viscosity change during milk clotting. *Molecules.* (accepted for publication)

Kertész, I., Zsom-Muha, V., András, R., Horváth, F., Németh, C., Felföldi, J. (2021b): Development of a Novel Acoustic Spectroscopy Method for Detection of Eggshell Cracks. *Molecules,* 26(15), 4693;

Kertész István, Zsomné Muha Viktória, Zsom Tamás, András Rebeka, Nagy Dávid, Felföldi József (2019): Tojás minőségének és a héj repedésének roncsolásmentes vizsgálata, *Animal welfare, etológia és tartástechnológia*, 15. (1), 19-35.

Kertész, I., Felföldi, J. (2016): Comparison of sound velocity estimation and classification methods for ultrasonic testing of cheese, *Progress in Agricultural Engineering Sciences,* 12 (1), 51-62.

Kertesz, I., Simon, N., Kaszab, T., Zsom Muha, V., Felfoldi, J. (2015): Ultrasonic method for cheese quality evaluation, In: Walker, J. (Ed.) *American Society of Agricultural and Biological Engineers Annual International Meeting 2015*