

PQS (POLAR QUALIFICATION SYSTEM) THE NEW DATA REDUCTION AND PRODUCT QUALIFICATION METHOD

K. J. KAFFKA^{a*} and ZS. SEREGÉLY^b

^aFaculty of Food Science, Szent István University, H-1118 Budapest, Ménesi út 45. Hungary

^bMetrika R&D Co. H-1119 Budapest, Petzvál J. u. 25. Hungary

(Received: 13 July 2000; revision received: 18 June 2001; accepted: 7 August 2001)

The near infrared spectra are useful information sources relating to quality (e.g. composition) of a material examined. To obtain and interpret useful information requires in most cases the application of sophisticated methods of mathematical statistics. A method different from those mentioned above, implementing a large scale data reduction based on geometrical consideration is the PQS. According to this method, the quality of a material can be characterised by the centre of its spectrum represented in a polar co-ordinate system. In many cases it is enough to know whether the investigated product deviates in a certain degree from a given “standard product” or not. This can be decided by determining special “distances” between the two (investigated and standard) products using their near infrared spectra. Besides the successfully used Euclidean and Mahalanobis distances a new one, the “polar distance” was introduced giving the distance between the two centres (quality points) of the spectra of the two products examined. A method was elaborated to select the optimal wavelength range giving the maximum normalised distance between the two quality points of the investigated products. The so called “wavelength range optimisation” can not be used to work with non spectral data sets. While in case of NIR spectra the sequence of the data are determined by nature, in several cases the order of the data can be freely varied and the goal is the determination of the optimal data sequence. By introducing the “sequence optimisation” PQS could be generalised and used from the field of near infrared spectroscopy to solve any kind of multivariate tasks. The advantage of the PQS optimisation method is its simplicity. Since PQS was developed to extract the needed information from NIR spectra, the basic principles of the technique are introduced with the help of near infrared spectra of some milk powder samples of different fat content. The sequence optimisation is demonstrated with the sensor signal responses of an electronic nose (chemosensor array) instrument measuring different steam distilled volatile oil samples.

Keywords: near infrared spectroscopy, electronic nose, polar qualification system, PQS, wavelength range optimisation, sequence optimisation, polar distance, data reduction, qualification

The PQS – as a new data reduction and product qualification system – was introduced at the 3rd International Conference on Near Infrared Spectroscopy (KAFFKA & GYARMATI, 1991) in Brussels. According to PQS a “quality point” was defined on a two dimensional “quality plane” and a “polar distance” was used for describing quality differences. The quality point of the investigated material was given on the quality plane by the centre of its spectrum represented in a polar co-ordinate system.

* To whom correspondence should be addressed.

Tel: 372-62-00/6060; Fax: 372-63-21; E-mail: kkaiffka@omega.kee.hu

It was demonstrated using several food-industrial products that PQS is a drastic but meaningful data-reduction method based on geometrical view, therefore it is easy to imagine and at the same time it gives a good pictorial representation of the quality (KAFFKA, 1992).

When representing a spectrum in a polar co-ordinate system the radius is a function of the spectral value, while the angle is a function of the wavelength. A wavelength scale can be placed on the circle surrounding the polar diagram of the spectrum centrally. The absorption peaks of this polar spectrum are, of course, at the same wavelengths as in the rectangular co-ordinate system. By changing the composition of a sample the quality point shifts parallel with the direction of the absorption peak of the changing component. From the direction of the quality point's shift conclusion can be drawn with regards to which component has changed. In the quality plane one quality point belongs to one sample but theoretically samples of different composition may have the same quality point. To resolve this very rare coincidence, the wavelength range must be selected very carefully, or two polar diagrams should be studied with different wavelength range. It was found useful to span the wavelength difference between two absorption peaks of two components (e.g. fat and water) to 90° or 270° in the polar diagram (VAN DER VLIES, et al., 1995; KAFFKA & GYARMATI, 1996).

The PQS was further developed, three interpretations were given for the "centre" of the polar spectrum, resulting in three different formulas for determining the x and y co-ordinates of the quality point. The effect of the change in the amplitude of the absorption peaks, the effect of the noise of the spectrum, the effect of the shifting and tilting the base-line of the spectrum on the location of the quality point were investigated using the three formulas. The results of this investigation and the characteristic features of the three formulas were introduced at the 9th International Diffuse Reflectance Conference (KAFFKA & GYARMATI, 1998).

The task of the present study was to summarise the most important steps of development of the polar qualification system, to introduce the latest manipulation possibilities of the spectra, to investigate and analyse the properties of the PQS method and by introducing the "sequence optimisation" to generalise PQS from the field of near infrared spectroscopy to solve any kind of multivariate tasks.

1. Materials and methods

As it can be seen in Fig. 1 and as it is suggested by the name, the basic idea behind this method is to represent the spectra in the polar co-ordinate system and determine the so-called "quality points", the centres of the polar spectra. The polar spectrum might be unusual but the two spectra are equivalent.

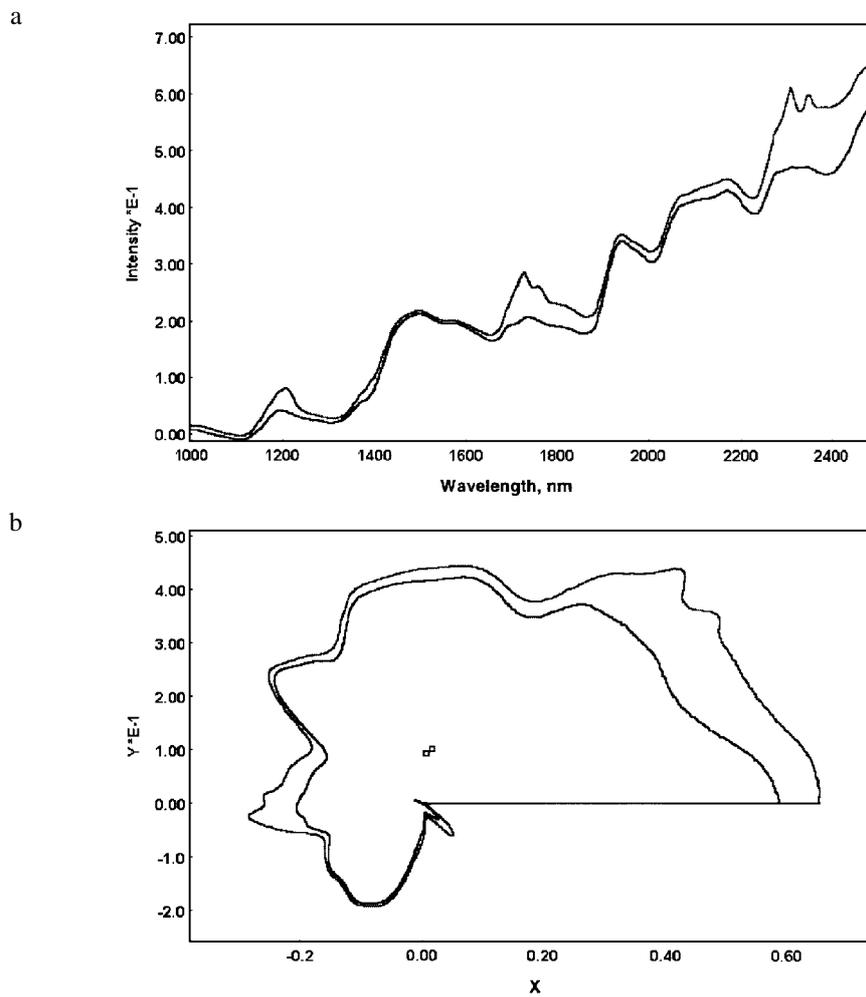


Fig. 1. The $\log(1/R)$ spectra of two milk powder samples containing the lowest and the highest amount of fat, in rectangular (a) and polar (b) co-ordinate system

There are three possibilities for the calculations of the quality point co-ordinates (Fig. 2). According to the “point method” the centre is defined as the centre of gravity of the unit masses placed in each spectral points. According to the “line method” the centre is defined as the centre of gravity of a wire shaped as the polar spectrum. At the “surface method” the centre is defined as the centre of gravity of a surface surrounded by the spectrum.

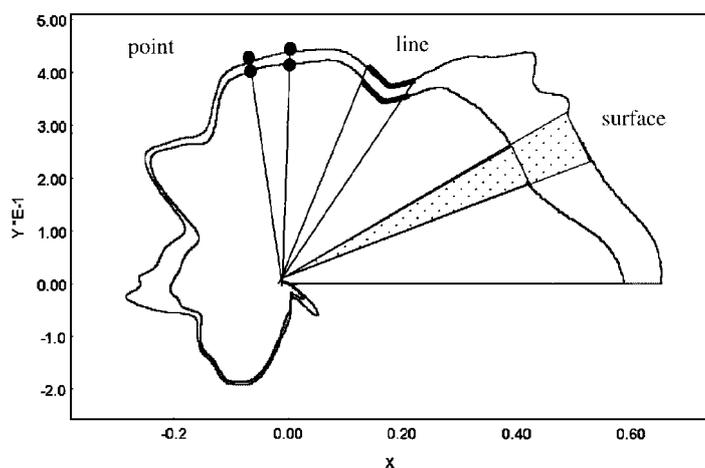


Fig. 2. Explanation of the three ideas how the centre of the spectrum can be defined

As it can be seen in the figures the location of the centres of the polar spectra can be defined by their x and y co-ordinates in the Descartes co-ordinate system placed in the centre of the polar diagrams.

The detailed investigation of the properties of the three methods with the help of artificial $\log(1/R)$ spectra was already published in an earlier paper (KAFFKA & GYARMATI, 1998). The idealised $\log(1/R)$ spectra used in this study contained two absorption peaks caused by components A and B. The peaks at 1175 nm and 2300 nm were described with Gaussian curves the bandwidth of which were 50 nm for the A and 150 nm for the B component. In the paper mentioned four experiments were performed for the detailed investigation.

In the first experiment a constant 0.5 $\log(1/R)$ unit peak-amplitude was chosen for the B component, while the peak-amplitude of the A component was changed from 0.0 to 1.0 unit in 0.1 unit steps. The baseline of the spectra was constantly 0.0. In the second experiment the spectra remained the same as during the first investigation but the baseline was shifted vertically with 0.2 unit. In the third experiment the baseline shifted spectra were loaded with 0.1 unit (peak to peak) noise. In the fourth experiment the baseline was rotated around the origin to a slope of 0.5 $\log(1/R)/1500$ nm. Figure 3 shows the spectra in the polar co-ordinate system of components A and B, where the masses of A and B were changed from 0.0 unit to 1.0 unit in 0.1 unit steps (top) and the corresponding quality points using the point method (below).

From the four experiments the following conclusions were drawn. Using “point method” of the PQS, the relations are linear. It means on the one hand that changing the concentration of one component the “quality points” lie on a straight line on the x, y “quality plane”, and on the other hand the distance (polar distance) between two “quality points” is proportional to the concentration difference of the relevant samples.

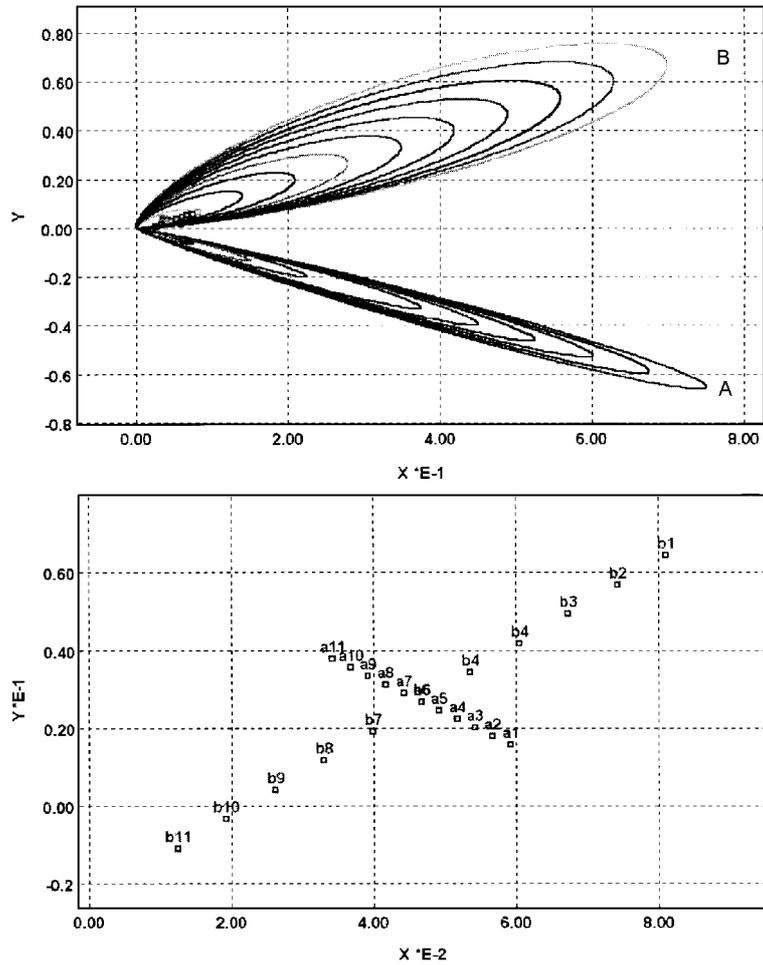


Fig. 3. Idealised absorbance spectra in the polar co-ordinate systems of A and B components described with Gaussian curves(top). The masses of both components are changed. The location of the quality points is changing with the ratio of the components using the point method (bottom)

Using “point method” of the PQS, the location (position) of the “quality points” changes if the baseline is tilted, but the polar distance between two “quality points” belonging to two samples with given different composition remains the same.

Using “line” or “surface method” of the PQS the relations are non-linear. Changing the concentration of one component the “quality points” lie on a curved line on the x, y “quality plane”, and the distances (polar distances) between two “quality points” are not

proportional to the concentration difference of the relevant samples. However the polar distances are much higher than using the “point method”. It means that the sensitivity of the “line” and “surface method” is higher compared to the point method.

Using the “point” and “surface” method of the PQS, the location of the “quality points” do not change significantly if the spectrum becomes noisy. These methods are insensitive to noise, while the “line” method is very sensitive to it.

Besides the successfully used Euclidean and Mahalanobis distances the new one, the “polar distance” was introduced giving the distance between the two centres (quality points) of the spectra of the two products examined. Figure 3 demonstrates how the “quality point” shifts in the direction of the changing component’s peaks in the polar spectrum. This makes it possible to draw conclusion with regards to which component has changed if the “quality point” shifts to somewhere. If there is a need to separate the effect of the changes in the concentration of two components on the location of the “quality point” it was found useful to span the wavelength difference between the two peaks of the main components to 90° or 270° in the polar diagram.

Hereby we would like to remark that the absorption peaks of the same components can be situated 180° to each other balancing their shifting effect to the location of the quality point. The so-called “wavelength range optimisation” solves this problem. The goal of the wavelength range optimisation is to determine that wavelength range (that part) of the spectrum which gives the best distinction of two samples according to one selected chemical or physical property using their quality points. The criterion of the “best distinction” must be of course defined. Three formulas were introduced with the help of which the best distinction could be determined (KAFFKA & SEREGÉLY, 2000). The maximum of the “absolute distance” or the “normalised distance” or the “sensitivity” are the possible criteria of the optimum. Interpretation of the formulas can be seen in Fig. 4.

The formula of the “absolute distance” is as follows:

$$D_{\text{abs}} = \sqrt{(\bar{x}_2 - \bar{x}_1)^2 + (\bar{y}_2 - \bar{y}_1)^2} \quad (1)$$

where \bar{x}_1 and \bar{x}_2 are the average x co-ordinate values of the quality points of the two samples for which the distinction have to be made according to one chemical or physical property, while \bar{y}_1 and \bar{y}_2 are the average y co-ordinate values of the same two samples. The average values must be determined from at least 5 repeated measurements of the two samples.

The formula for the “normalised distance” is as follows:

$$D_{\text{norm}} = \frac{D_{\text{abs}}}{D_{\text{abs}} + s_1 + s_2} \quad (2)$$

where s_1 and s_2 are the standard deviations of the quality points of the two samples.

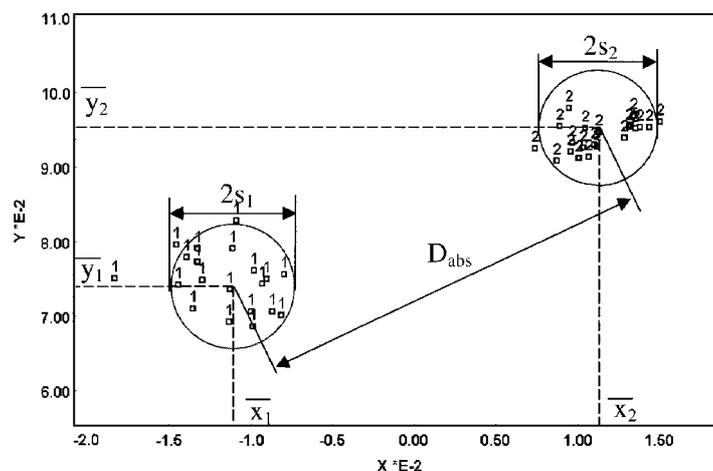


Fig. 4. Graphical interpretation of the three formulas as possible criteria of the optimum wavelength range

The formula for the “sensitivity” is as follows:

$$S = \frac{D_{\text{abs}}}{s_1 + s_2} \quad (3)$$

By calculating these terms the effectiveness of the classification can be expressed numerically making it possible to compare the results obtained by using different classification models.

The values of the selected chemical or physical property of the two samples should not be exactly known, but it must be certain that a – possibly large – difference exists in this property between the two samples.

A self made computer program and algorithms make it possible to perform single wavelength range optimisation automatically. In this case parameters such as the first and the last wavelength (within the optimal wavelength range has to be searched for), the gap (the initial wavelength range), the gap shift and the gap broadening must be specified. Two groups of spectra, repeated measurements of the two samples to be distinguished have to exist. The “gap” is the length of the initial wavelength range in nm, which is then shifted with the “gap shift” in nm during the optimisation process from the first wavelength until the gap reaches the last one. Then the gap broadens with the „gap broadening” in nm until the gap reaches the length of the whole wavelength region i.e. the difference between the last and the first wavelength.

After the determination and discussion of the first optimal wavelength range it is possible to restart the automatic wavelength range optimisation program, so the best second, the best third etc. wavelength range can also be calculated. In this case the

previously determined optimal wavelength range or ranges must be omitted and the optimisation has to be repeated only in the ranges located below, above or between them (Fig. 5).

As it can be seen in Fig. 5 by determining the best second and third wavelength ranges and by using the whole near infrared region, the optimal ranges can be situated opposite to each other resulting in a worse classification.

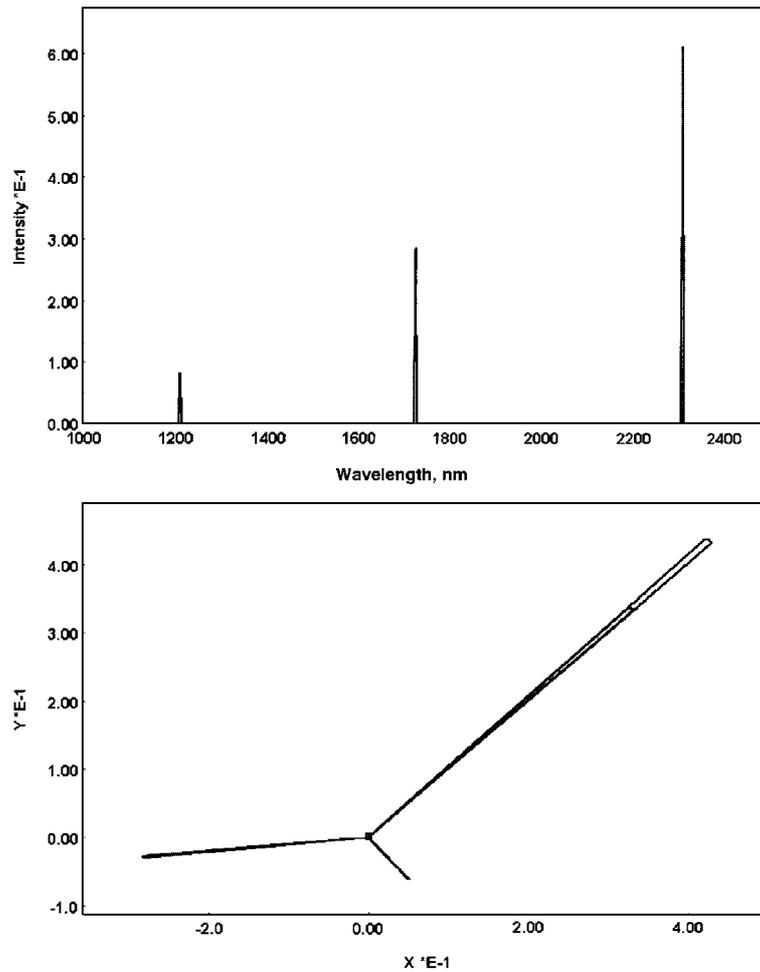


Fig. 5. The spectra in the first, second and third optimal wavelength ranges giving the largest normalised distances and the largest sensitivities between the quality points of the two samples setting the spectrum values to zero outside the actually selected wavelength range, in rectangular (top) and polar (bottom) coordinate system

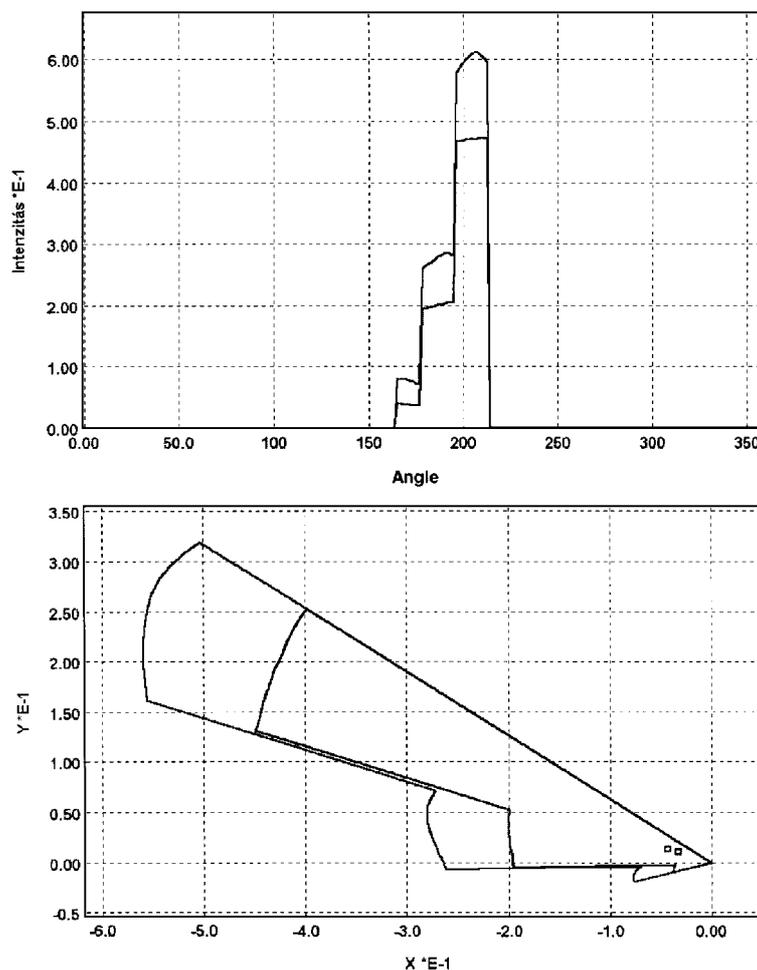


Fig. 6. The Descartes (top) and polar (bottom) spectra of the three united optimal spectrum ranges achieved by omitting the wavelength ranges between the selected optimal ones

By omitting the ranges among the selected ones the optimal ranges can be collected together, the direction of their shifting effect to the location of the quality point can be turned to the same direction, summarising their effects (Fig. 6).

The optimal wavelengths are the same as the characteristic wavelengths found by using linear regression in three wavelengths regions (1000–1500 nm, 1500–2000 nm, 2000–2500 nm). Also the results are similar with PQS and MLR (multiple linear regression). Two essential differences between PQS and MLR must be mentioned: PQS using automatic range optimisation does not need an accurately analysed sample set, as

does at MLR (where the calibration models are created to determine the relationship between the NIR spectra of the sample and their quality parameters determined by reference methods), at PQS the optimum wavelength ranges can be used together getting better results, while it is forbidden at MLR because of the collinearity problem.

The described investigations and results with milk powder samples were only to demonstrate the usefulness and versatility of the automatic wavelength range optimisation program. Better results could be achieved by preparing the spectra (e.g. smoothing, multiplicative scatter correction or using the second derivative). It must also be noticed, that the demonstrated optimisation refers only to the fat content. If the task is to separate two samples according to their water content, then two samples with different water content is needed.

The “wavelength range optimisation“ can not be used to work with non spectral data sets. While the sequence of the data are determined by nature in NIR spectroscopy, there are several multivariate tasks where the order of the data can be changed freely and the goal is the determination of the optimal data sequence in the respect of the given classification (which gives the best distinction of two samples using the quality points of these two samples). For this purpose a new evaluation method, the “sequence optimisation” was developed. By elaborating the “sequence optimisation”, PQS could be generalised and extended from the field of near infrared spectroscopy to solve any kind of multivariate tasks.

Figure 7 shows the repeatedly measured normalised data of the sensor signals of a chemosensor-array, the so called “electronic nose” instrument, of two selected volatile oils represented in Descartes co-ordinate system as a spectrum (top). In the lower part of the Fig. 7 the same data set is shown in polar co-ordinate system.

Figure 8 shows the locations of the quality points of the seven investigated volatile oils calculated from the normalised data of the sensor signal response of the electronic nose evaluated also by principal component analysis in Fig. 9. By comparing the results it can be established that the investigated samples can be well identified in one step on the quality plane calculated from the optimal data sequence determined by sequence optimisation (SEREGÉLY & KAFFKA, 1999).

2. Results

In this chapter of our paper few fields of application are presented, without much comments, where we achieved results using PQS in the food industrial quality control.

Figures 10–12 represent the application possibilities of near infrared spectroscopy and the polar qualification system in the brewing industry (SEREGÉLY, 2000). Figure 10 shows the discrimination of spring and winter barley varieties. The aim of this work was to present a feasible procedure for the prediction of the quality parameters of barley samples from their NIR spectra in the respect of their brewing consumption.

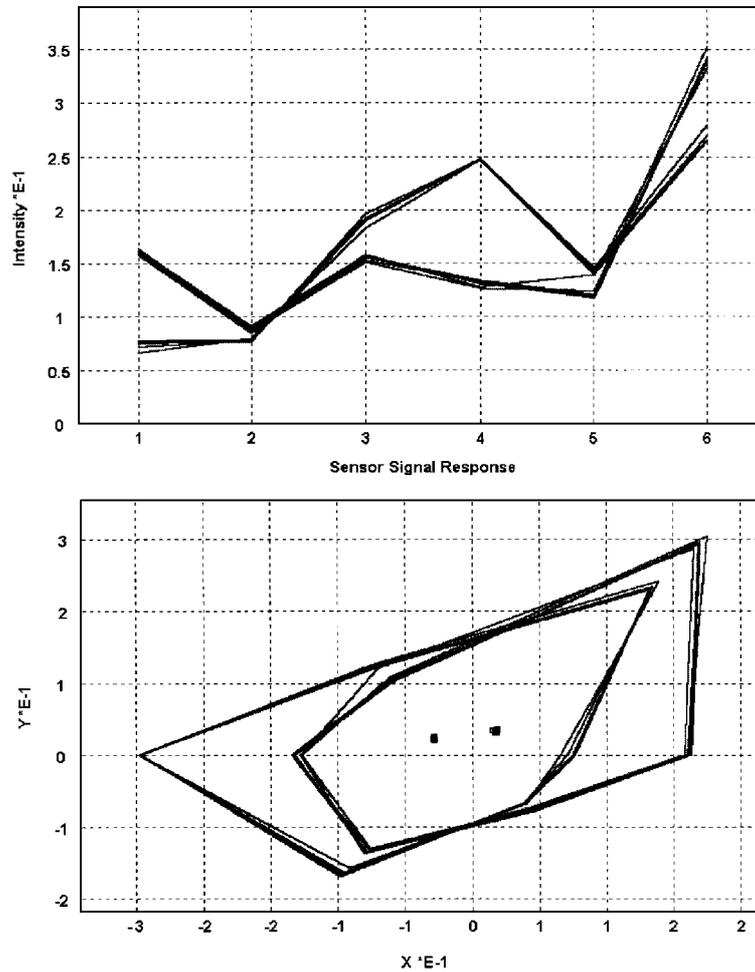


Fig. 7. The normalised data of the sensor signals of a chemosensor-array the so called “electronic nose” instrument of the two volatile oils represented in Descartes co-ordinate system as a spectrum (top). The same data sets and the relevant quality points are shown in polar co-ordinate system (bottom)

The task of the second step of this work presented in Fig. 11 was – in spite of the large variability caused by variety and harvest area – to determine the joint features of brewing barley varieties to be able to separate them from spring varieties having lower quality used for animal feeding. Finally samples were taken during malt production, the different enzymatic processes were stopped by drying and the movement of the quality point of germinating barley on the quality plane can be observed from barley to ready made malt in Fig. 12.

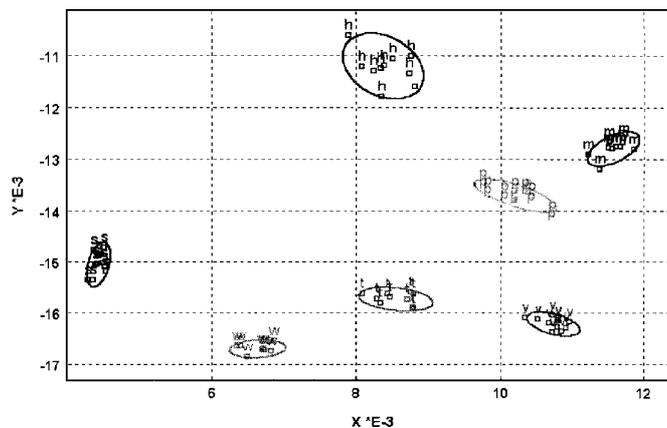


Fig. 8. Discrimination of seven volatile oil samples using electronic nose and sequence optimisation (t: thyme; p: peppermint; h: oreganum hirt.; v: oreganum vulg.; w: dill weed; s: dill seed; m: mentax)

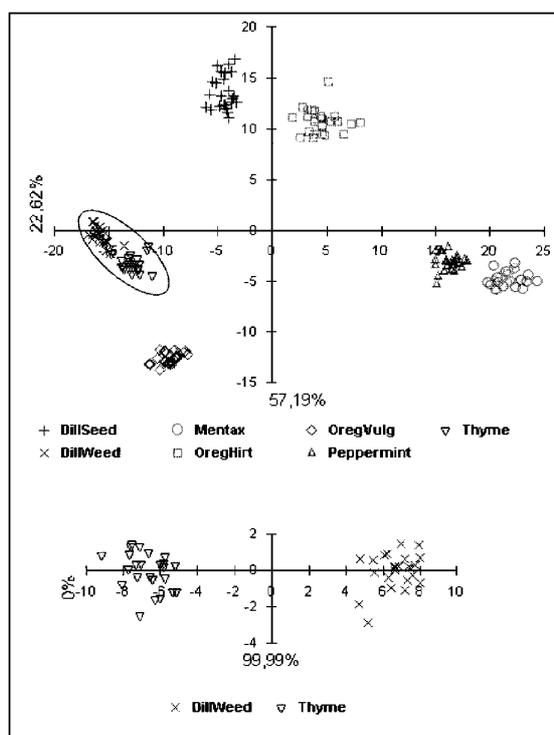


Fig. 9. Discrimination of seven volatile oil samples in two steps using electronic nose and principal component analysis

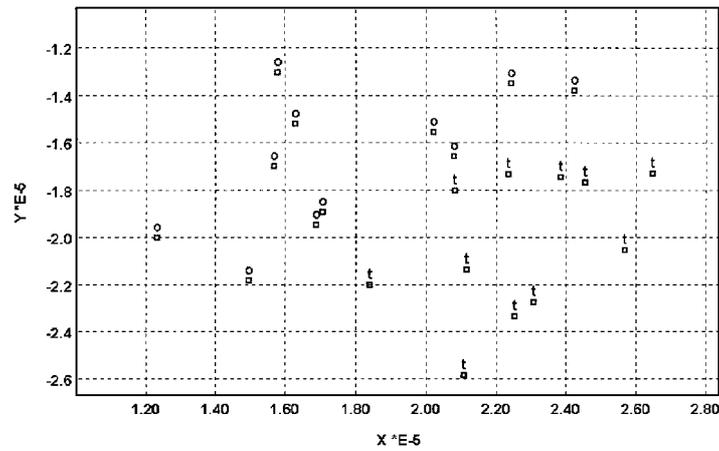


Fig. 10. The discrimination of winter and spring barley varieties (o: winter; t: spring varieties)

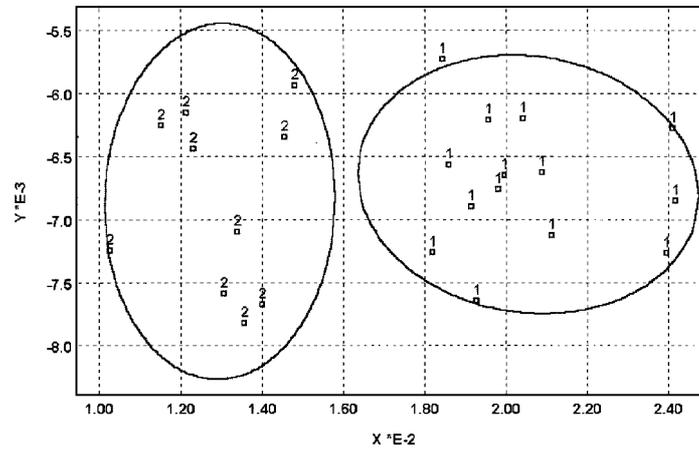


Fig. 11. The discrimination of spring barley varieties having “brewing quality”, from those used only for animal feeding (1: brewing; 2: crop)

The demonstration of the results of the PQS applications aimed at the processing of the information content of near infrared spectra is illustrated with the help of some research works focused on the determination of quantitative composition differences.

Figure 13 shows the quality points of milk powder samples having different fat content, while Fig. 14 demonstrates the location of the quality points of coffee samples having different proportion of coffee substituent content, and finally the quality plane of the red paprika samples of different mould content (KISKÓ, et al., 1998) are presented in Fig. 15.

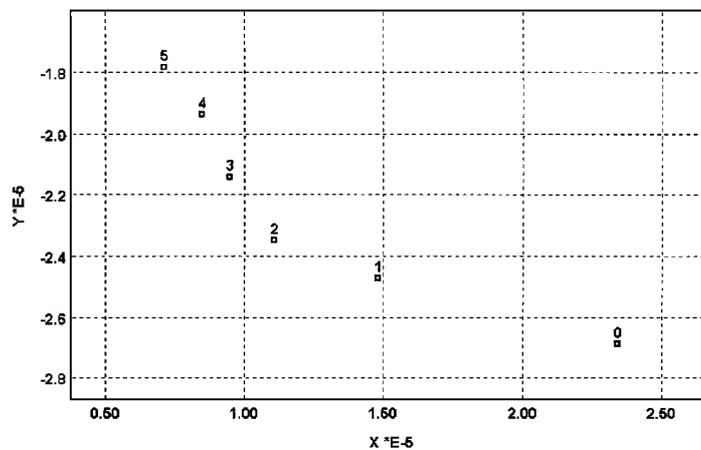


Fig. 12. Tracking the changes during malt production from barley to ready made malt (0: 0th day, barley; 1: 1st day;... 5: 5th day, ready made malt)

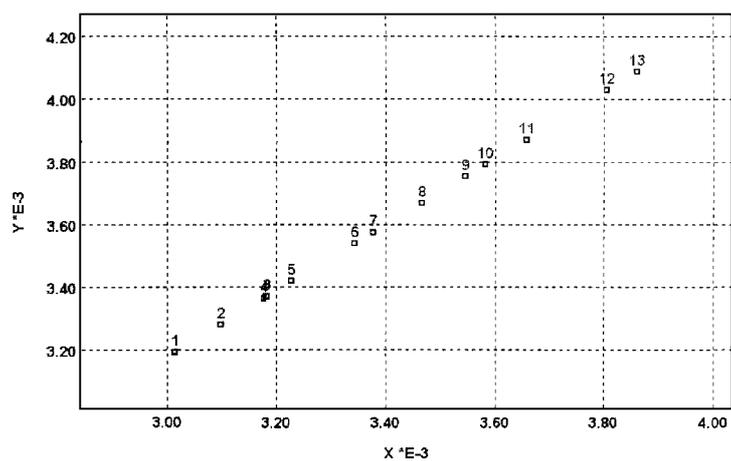


Fig. 13. Milk powder samples having different fat content (1: 2.5%; 2: 4%; 3: 6%; 4: 8%; 5: 10%; 6: 12%; 7: 14%; 8: 16%; 9: 18%; 10: 20%; 11: 22%; 12: 24%; 13: 26%)

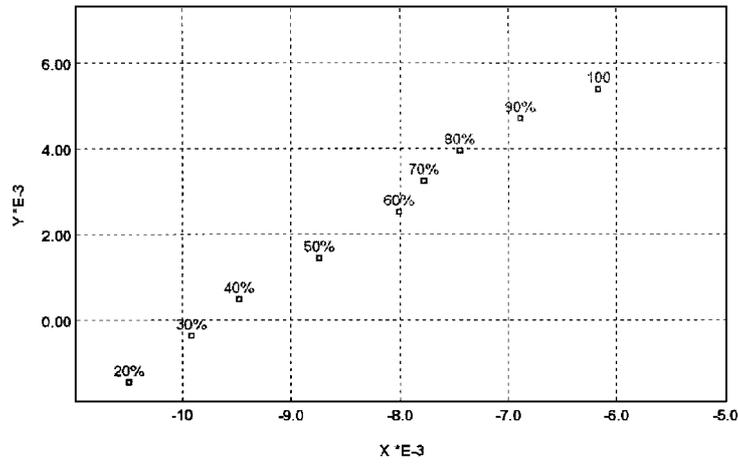


Fig. 14. Coffee samples having different coffee substituent content (%: mass percentage of the coffee substituent)

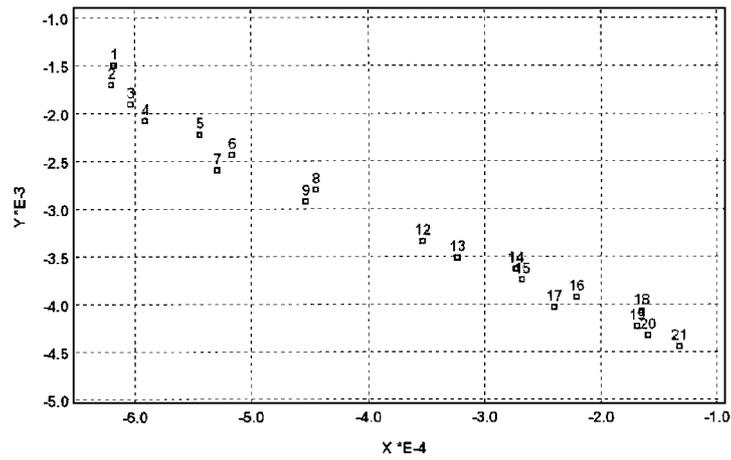


Fig. 15. Red paprika samples having different mould content (1: low mould content;21: high mould content)

Figures 16–17 show the locations of the quality points of the germinating barley – investigated also by near infrared spectroscopy – calculated from the normalised data of the sensor signal response of the electronic nose in original (Fig. 16) and in optimal (Fig. 17) data sequence. As it can be seen in the Fig. 17, although the normalised data of the sensor signal response provide the information suitable for tracking the changes

during malt production, however using the original data sequence the samples can not be separated. Contrary to this by using the data sequence, determined by the optimisation, the samples can be identified.

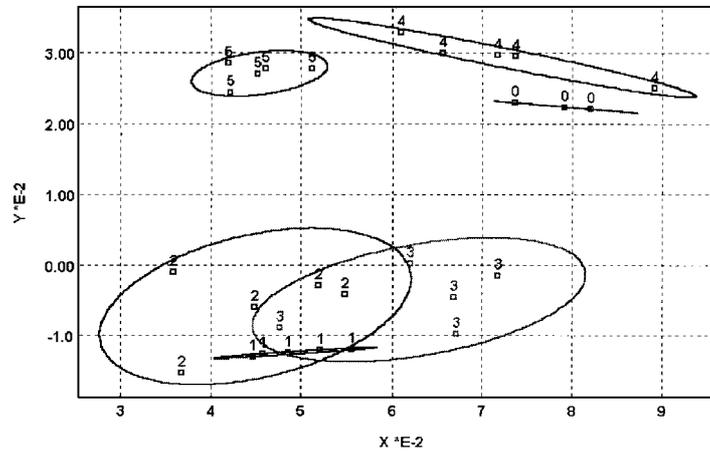


Fig. 16. Tracking the changes during malt production from barley to ready made malt using electronic nose (0: 0th day, barley; 1: 1st day;... 5: 5th day, ready made malt) Original sequence

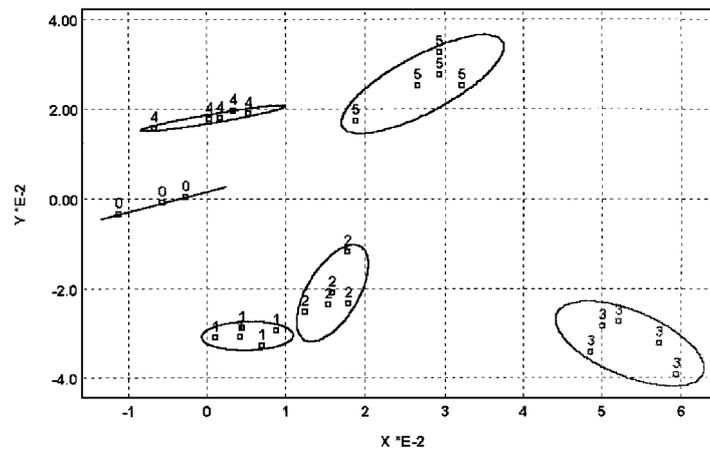


Fig. 17. Tracking the changes during malt production from barley to ready made malt using electronic nose and sequence optimisation (0: 0th day, barley; 1: 1st day;... 5: 5th day, ready made malt) Optimal sequence

3. Conclusion

A data reduction method based on geometrical (graphical) view, polar qualification system was introduced in detail and its properties were investigated. The “geometrical” data reduction method is easy to visualise in contrast to the existing, different “mathematical” data reduction methods. As the near infrared spectrum is a fingerprint of the investigated material, PQS compresses (reduces) the spectral data in a “quality point” on a two dimensional “quality plane” for qualification. The location (position) of the quality point is influenced (defined) by all, normally several hundreds, spectral data.

The investigation of the properties of the PQS was performed using idealised $\log(1/R)$ spectra of the A and B components. The artificial absorption peaks were described with Gaussian curves; with peaks at 1175 nm for component A and at 2300 nm for component B with 50 nm and 150 nm bandwidth, respectively.

As it can be seen in the paper, there are two basically different methods at performing the automatic wavelength range optimisation. At the first method the spectral values of the spectral ranges outside the actually selected wavelength range are set to zero, while at the second method the spectral ranges outside the actually selected wavelength range are omitted.

Using the first method of the automatic wavelength range optimisation, the optimal wavelength ranges are narrow and at or close to the maximum of the absorption bands of the component the variation of which was used for distinction of samples. Using the second method of the automatic wavelength range optimisation significantly broader optimal wavelength ranges were found and the results were considerably, astonishingly better. The best sensitivity achieved by using the first method was 45.53, while using the second method the sensitivity was 136.00 in case of milk powder samples.

The handling of spectra, namely the possibility to select, or set to zero or to omit optional wavelength ranges of the spectrum combined with the automatic single wavelength range optimisation, which can be repeated as well as the sequence optimisation in case of non spectral data sets, opens new perspectives in application of PQS offering a rapid, accurate, cheap and simple method for qualifying or identifying products, using near infrared spectra or any other data sets.

*

Thanks are due to K. ASCHENBRENNER and G. BAYER for their conscientious technical assistance.

Sponsored by grants from the National Scientific Research Found (OTKA) No. 023020 and No. 032814.

References

- KAFFKA, K.J. (1992): Qualitative analysis applying NIR spectroscopy. TAYLOR, R.A. (Ed.) *Proceedings: International Diffuse Reflectance Spectroscopy Conferences*, The Council for Near Infrared Spectroscopy, Gaithersburg, MD. (USA) pp. 63–70.
- KAFFKA, K.J. & GYARMATI, L.S. (1991): Qualitative (Comparative) analysis by near infrared spectroscopy. BISTON, R. & BARTIAUX-THILL, N. (Eds.) *Proceedings of the Third International Conference on Near Infrared Spectroscopy*, Agricultural Research Centre Publishing, Gembloux (Belgium) pp. 135–139.
- KAFFKA, K.J. & GYARMATI, L.S. (1996): Reduction of spectral data for rapid quality evaluation. DAVIES, A.M.C. & WILLIAMS, P. (Eds.) *Near Infrared Spectroscopy: The Future Waves*, NIR Publications, Chichester (U.K.) pp. 209–213.
- KAFFKA, K.J. & GYARMATI, L.S. (1998): Investigating the polar qualification system. *J. Near Infrared Spectroscopy*, 6, A191–A200.
- KAFFKA, K.J. & SEREGÉLY, ZS. (2000): Wavelength range optimisation using the PQS (Polar Qualification System). DAVIES, A.M.C. & GIANGIACOMO, R. (Eds.) *NIR Spectroscopy: Proceedings of the 9th International Conference*, NIR Publications, Chichester (U.K.) pp. 261–267.
- KISKÓ, G., STEGEMAN, H. & FARKAS, J. (1998): Detection of moulds in paprika powder by enzym-linked immunosorbent assay. *Acta Alimentaria*, 27, pp. 97–103.
- SEREGÉLY, ZS. (2000): Investigating the NIR characteristics of winter and spring barley and malt. DAVIES, A.M.C. & GIANGIACOMO, R. (Eds.) *NIR Spectroscopy: Proceedings of the 9th International Conference*, NIR Publications, Chichester (U.K.) pp. 543–546.
- SEREGÉLY, ZS. & KAFFKA, K.J. (1999): Result achieved at qualification of volatile oils using chemosensor-array (electronic nose). *Trends in AgriFood Sector*, Nitra.
- VAN DER VLIES, C., PUGGE, W. & KAFFKA, K.J. (1995): Qualifying pharmaceutical substances by fingerprinting with NIR spectroscopy and the polar qualification system. *Spectroscopy*, 10, 46–49.