Wavelength range optimisation using the Polar Qualification System

Karoly J. Kaffka and Zsolt Seregély
University of Horticulture and Food Industry, Ménesi ut 45, H-1118 Budapest, Hungary.

Introduction

The Polar Qualification System (PQS)—as a new data reduction and product qualification system—was introduced at the 3rd International Conference on Near Infrared Spectroscopy[1] 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 coordinate system. The investigation of a variety of about 40 different flavour and aroma samples was presented. As the first result achieved using the PQS, it was established that the different flavours and aromas could clearly be distinguished (identified). Even the sensory threshold (the minimal perceptible difference) could be well resolved.

At the 4th International Diffuse Reflectance Spectroscopy Conference,[2] four examples were introduced in order to demonstrate the versatility and usefulness of the PQS. In the first example, minced lean pork and minced fat bacon were mixed in different quantities; in the second example, ground roasted coffee and ground coffee substitute were blended in different ratios; in the third example, skimmed and fat milk powders were mixed giving different fat content, while in the fourth example, original goose liver samples with different fat content were used. It was established that PQS could be used for determining the quality of different food products and giving "polar distances" of the investigated samples from the standard sample it can be used to determine acceptability. It was also shown that the place of the quality point is practically independent from the noise of the spectrum; so no smoothing is needed for preparing spectra.

The PQS method was applied to qualify pharmaceutical substances. C. van der Vlies and his coworkers[3] reported results demonstrating that the method is able to detect small differences in chemical and physical properties, which makes it an ideal tool for qualifying pharmaceutical substances.

At the 7th International Conference on Near Infrared Spectroscopy, the reduction of spectral data for rapid quality evaluation was reported using the PQS method.[4] This method, carried out on several food and industrial products, demonstrated that PQS is a drastic, but meaningful, data-reduction method based on a geometrical view, therefore it is easy to imagine and at the same time it gives a good pictorial representation of the quality.

The PQS was further developed and three interpretations were given for the "centre" of the polar spectrum, resulting in three different formulae for determining the x and y coordinates 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 formulae. The results of this investigation and the character-

Near Infrared Spectroscopy: Proceedings of the 9th International Conference

istic features of the three formulae were introduced at the 8th International Conference on Near Infrared Spectroscopy.[5]

At the representation of a spectrum in the polar coordinate 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 wavelength as in the rectangular coordinate system. 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, a 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 compositions 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 ranges. It was found to be useful to span the wavelength difference between two absorption peaks of two components (for example, fat and water) to 90° or 270° in the polar diagram.

The task of the present study was to introduce new spectral manipulation methods which were developed to fulfil the requirements mentioned above. A computer program was completed for automatic determination of the optimal wavelength range for a given task and for the possibility of very flexible spectrum handling.

Materials and methods

The goal of the wavelength range optimisation is to determine a certain wavelength range (a certain 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 for "best distinction" must, of course, be defined. Three formulae were introduced with the help of which the best distinction could be determined. The maximum of the "absolute distance" or the "normalised distance" or the "sensitivity" are the possible criteria of the optimum. Interpretation of the formulae can be seen in Figure 1.

The formula of the "absolute distance" is as follows:

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

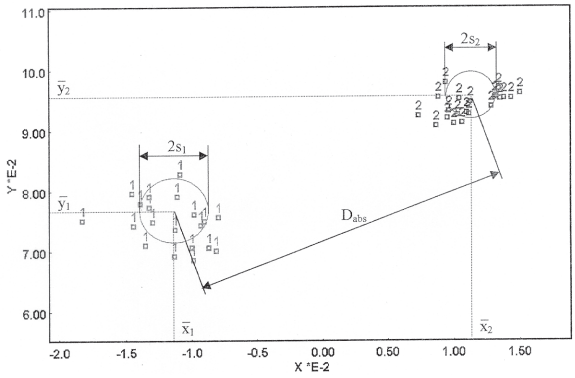

Figure 1. Graphical interpretation of the three formulae as possible criteria of the optimum wavelength range.

Where $\bar{x}_1$ and $\bar{x}_2$ are the average *x* coordinate values of the quality points of the two samples between which the distinction has to be made according to one chemical or physical property, $\bar{y}_1$ and $\bar{y}_2$ are the average *y* coordinate values of the quality points of the same two samples. The average values must be determined from at least five repeated measurements of the two samples.

The formula for the "normalised distance" is as follows:

$$D_{norm} = \frac{D_{abs}}{D_{abs} + s_1 + s_2}$$

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

$$s_1 = \sqrt{\sum_{i=1}^{n} \frac{(x_{1i} - \bar{x}_1)^2}{n \quad 1} + \sum_{i=1}^{n} \frac{(y_{1i} - \bar{y}_1)^2}{n \quad 1}}$$

$$s_2 = \sqrt{\sum_{i=1}^{n} \frac{(x_{2i} - \bar{x}_2)^2}{n-1} + \sum_{i=1}^{n} \frac{(y_{2i} - \bar{y}_2)^2}{n-1}}$$

where *n* is the number of the repeated spectrum measurements.

The formula for the "sensitivity" is as follows:

$$S = \frac{D_{abs}}{s_1 + s_2}$$

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

A computer program makes 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 which 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, ie repeated measurements of the two samples to be identified, 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 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 ranges 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. To be able to fulfil these requirements special spectrum handling possibilities must be given. Therefore, the program makes it possible for any wavelength ranges of the spectrum to be selected for use, to be set to zero, or to be omitted.

Results are introduced by using a milk-powder sample set, the fat content of which varies from 2.5 to 26 mass % in 1 mass % steps (except the first step, which varies from 2.6 to 3 mass %). Thus, the set contains 25 samples. The spectra of the outermost samples with 2.5 and 26 mass % fat contents were

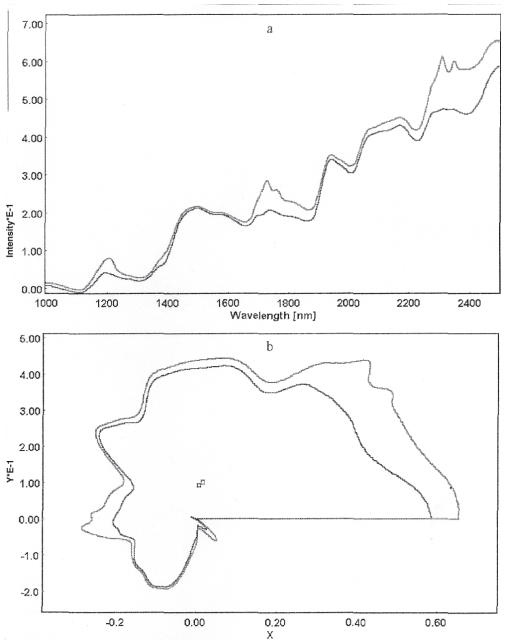

Figure 2. The log(1/*R*) spectra of two milk powder samples containing the lowest and the highest fat content, in rectangular (a) and polar (b) coordinate system.

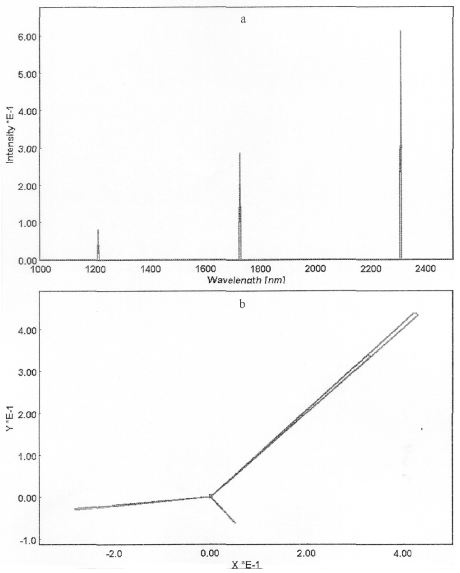

Figure 3. The spectra in the first, second and third optimal wavelength ranges giving the largest normalised distances and the largest sensitivities between the quality points of the two samples setting the spectrum values to zero outside the actually selected wavelength range, in rectangular (a) and polar (b) coordinate system.

measured eight times, while all the others will be measured three times. Altogether 85 log(1/*R*) spectra were recorded in the 1000–2500 nm wavelength region in 2 nm steps, using a Spectralyzer 1025 type spectrometer in reflectance mode.

The spectral values can be taken from the original [for example, log(1/*R*)] spectrum, or from any of its transformed (for example, smoothed, multiplicative scatter corrected or derivated) spectra.

Results

The log(1/*R*) spectra of two milk-powder samples containing the lowest and highest fat content can be seen in Figure 2 in rectangular (a) and in polar (b) coordinate systems. With automatic single wavelength range optimisation, setting the spectrum values to zero outside the actually selected wavelength range, the first optimal wavelength range was found between 2306 and 2310 nm.

Here, the value of the normalised distance was 0.979 and the sensitivity was 45.53. The second optimal wavelength range was found between 1724 and 1726 nm, the normalised distance was 0.971 and the sensitivity was 33.20. The third optimal wavelength range was found between 1208 and 1210 nm, where the normalised distance was 0.952 and the sensitivity was 19.83.

In Figure 3, the spectrum ranges of the two mentioned samples can be seen, which were found 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 in rectangular (a) and in polar (b) coordinate systems. There is a possibility of using all the three optimal wavelength ranges together by

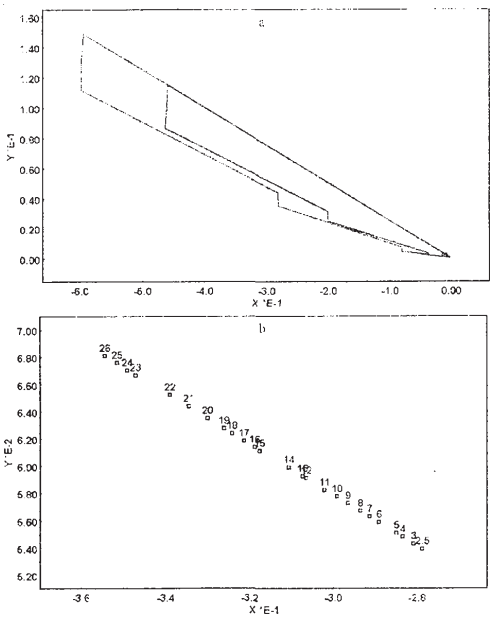

Figure 4. The polar spectra of the united three optimal spectrum ranges achieved by omitting the wavelength ranges between the selected optimal ones (a). The quality points of the 25 milk powder samples using this united spectrum.

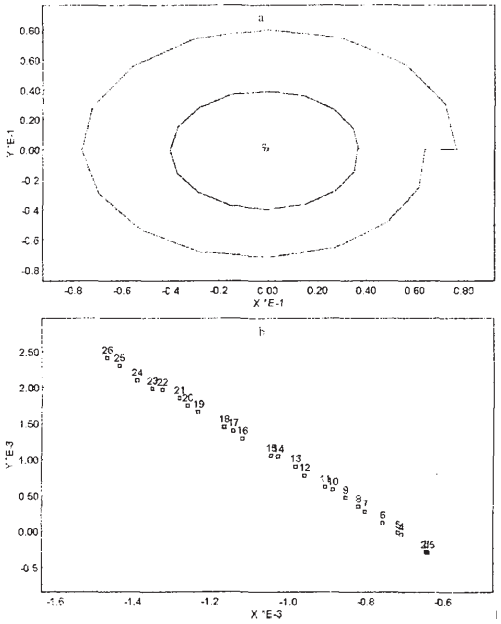

Figure 5. The spectra of the two samples in the first optimal wavelength range achieved by omitting the spectral values outside the actually selected wavelength ranges (a). The quality points of the 25 milk powder samples using this wavelength range.

omitting the wavelength ranges between the selected ones. In Figure 4, the polar diagram of this united spectrum ranges (a) and the quality points of the 25 milk powder samples (b) can be seen.

Using the three optimum wavelength ranges together, the result was better than only using one range. The normalised distance of 0.980 was achieved, while the sensitivity was 47.37.

Omitting the spectrum values outside the actually selected wavelength range, the first optimal wavelength range was found between 1182 and 1214 nm. The value of the normalised distance was 0.993 and the sensitivity was 136.00. Omitting the spectrum values outside the actually selected wavelength means that the actually selected wavelength range is spanned to 360°. The second optimal wavelength range was found between 1694 and 1734 nm, the value of the normalised distance was 0.992 and the sensitivity was 124.04. The third optimal wavelength range was found between 2274 and 2318 nm. The value of the normalised distance was 0.988 and the sensitivity was 81.36. Using this optimisation method, it is meaningless to put together the three optimal wavelength ranges. In Figure 5, the polar spectra of the selected optimum wavelength range of the two samples used for wavelength range optimisation (a) and the quality points of the 25 milk-powder samples (b) can be seen.

For comparison, in Figure 6, the LAB v. NIR plots are shown. The points belonging to the different fat content were determined by linear regression, describing the relationship between the spectral values measured at 2310 and 1726 nm and the fat content of the samples determined in the laboratory by chemical analysis. The three characteristic wavelengths found by using linear regression in the

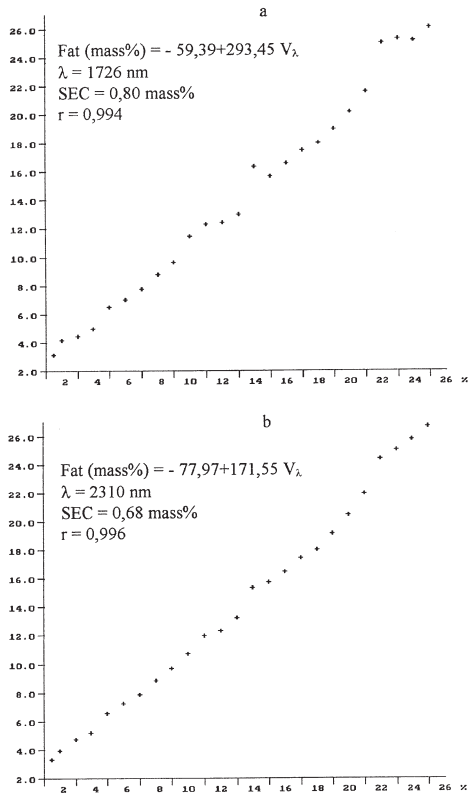

Figure 6. LAB v. NIR plots. The fat content was determined using the spectral values measured at 2310 and 1726 nm by linear regression.

1000–1500 nm, 1500–2000 nm and 2000–2500 nm regions were 1210, 1726 and 2310 nm, respectively.

The described investigations and results with milk powder samples was only to demonstrate the usefulness and versatility of the automatic wavelength range opimisation program. Better results could be achieved preparing the spectra (for example, smoothing, multiplicative scatter correction or using the second derivative). It must also be noted, 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 would be needed.

Discussion

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

Using the first method for 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 distinguising samples. These optimal wavelengths are the same as the characteristic wavelengths found by using linear regression in three wavelength regions (1000–1500 nm, 1500–2000 nm, 2000–2500 nm). Also the results are similar to PQS and MLR (multiple linear regression). Two essential difference between PQS and MLR must be mentioned: PQS, using automatic range optimisation, does not need an accurately analysed sample set, as MLR does; using PQS, the optimum wavelength ranges can be used together giving better results while it is not possible with MLR because of the collinearity problem.

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.

Conclusion

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

Acknowledgement

This work was supported by the grant from the Hungarian National Scientific Research Fund (OTKA) No.: T023020.

Thanks are due to K. Aschenbrenner for her conscientious technical assistance.

References

1. K.J. Kaffka and L.S. Gyarmati, "Qualitative (Comparative) analysis by near infrared spectroscopy", in *Proceedings of the Third International Conference on Near Infrared Spectroscopy*, Ed by R. Biston and N. Bartiaux-Thill. Agricultural Research Centre Publishing. Gembloux, Belgium, p. 135 (1991).
2. K.J. Kaffka, "Qualitative analysis applying NIR spectroscopy", in *Proceedings: International Diffuse Reflectance Spectroscopy Conferences*, Ed by R.A. Taylor. The Council for Near Infrared Spectroscopy, Gaithersburg, MD, USA, p. 63 (1992).
3. C. van der Vlies, W. Plugge and K.J. Kaffka, "Qualifying pharmaceutical substances by fingerprinting with NIR spectroscopy and the polar qualification system." *Spectroscopy,* **10(6),** p. 46 (1995).
4. K.J. Kaffka and L.S. Gyarmati, "Reduction of spectral data for rapid quality evaluation", in *Near Infrared Spectroscopy: The Future Waves,* Ed by A.M.C. Davies and P. Williams. NIR Publications, Chichester, UK, p. 209 (1996).
5. K.J. Kaffka and L.S. Gyarmati, *J. Near Infrared Spectrosc*. **6,** A191 (1998).