Application of the simulated annealing method for the assessment of the time evolution of samples by near infrared spectroscopy

Dominique Bertrand,[1] Nathalie Allosio-Ouarnier,[2] Patrick Boivin[2] and El Mostafa Qannari[3]

[1]*INRA, PHYTEC Laboratory, rue de la Géraudière, 44316 Nantes Cedex 3, France.*

[2]*IFBM, Pôle Technologique de Brabois, 7 rue du Bois-de-la-Champelle, 54512 Vandoeuvre Cedex, France.*

[3]*ENITIAA/INRA, Laboratoire de Statistique Appliquée à la Caractérisation des Aliments, rue de la Géraudière, 44316 Nantes Cedex 3, France.*

Introduction

In the food industries, most of the processes involve at least one continuous time-dependent transformation of raw materials into the final products. A certain degree of transformation has to be reached at a given time. The quality control of the transformation requires having an analytical method, making it possible to assess the degree of progression of the industrial process. In some situations, it is possible to find a single component, easy to analyse, which could be used as a marker for the progression. The concentration of such a relevant marker must increase or decrease continuously as a function of time. Rapid techniques such as near infrared (NIR) spectroscopy can then be used as an analytical method for the analysis of the identified marker.

However, in some industrial situations, it is difficult to identify a single component which would be representative of the progress in the transformation. When the process involves natural transformations such as fermentation, ageing or ripening, a single marker can rarely be found. In the present work, the production of malt from barley in the malting industry will be taken as an illustrative example of such a process. The malting process consists of three main steps. First, barley grains are steeped to increase their water content from about 12% to 44%. Thereafter, malting involves germination of the grains in order to degrade or modify the endosperm. The process is designed to produce grains from which the maximum amount of carbohydrates can be extracted during mashing. The yeast will use these carbohydrates during beer fermentation. Finally, the grain has to be kilned in order to stop sprouting, to develop the future flavours and colour of the final beer and to allow for storage. The germination stage, which lasts about five days, has to be carefully controlled. In order to obtain final products of constant quality, the malster must regulate the progression of the process. The barley sprouting is, indeed, a time dependent process, in which it is difficult to find a single relevant marker for the modification. This situation makes NIR difficult to use as an analytical technique for the quality control of barley sprouting, because usual statistical techniques, such as multilinear regression or partial least squares are inappropriate for the calibration of the spectrometer. One possibility would be to calibrate

Near Infrared Spectroscopy: Proceedings of the 9th International Conference
© IM Publications Open LLP 2000

the instrument taking the absorbances as predictive variables and the time itself as the dependent variable. It would be possible to establish a calibration using the linear model:

$$t = \boldsymbol{x\beta}' + \beta_0 + \varepsilon \quad (1)$$

with t the time, $\boldsymbol{x}$ the row vector of the P absorbances of the spectrum, $\boldsymbol{\beta}'$ the column vector of the P coefficients, β_0 the intercept value and ε the error term. For control purposes, the discrepancy between the observed time and the predicted time will be an indication of the degree of progression of the process, in comparison with reference (normal) conditions. However, this approach is not satisfying for two reasons. First, the relevant information of the predictive model lies in the error term (ε = *predicted time – observed time*), which is not statistically relevant. Second, the existence of a linear model, as defined in Equation (1), is doubtful. In a given time-dependent process, the rate of modification of a sample is rarely a constant value. For example, in a process involving the growth of organisms, it is generally observed that the evolution is slow at the beginning of the process, then it gets faster and slows down at the end. The growth often presents a sigmoidal shape. In the present study, a method for the assessment of the evolution over time of samples by NIR spectroscopy is presented. This method, which is based on a simulated annealing (SA)[1,2] procedure, has been tested for the control of barley sprouting during malting by NIR.

Theory

Description of the problem

Let $\boldsymbol{X}$ be a three-way array dimensioned $N \times P \times T$ of a reference spectral collection of samples collected at T different periods of the processes, during N independent processing (batches). The sampling periods are the same from one batch to another and the spectra are ordered according to time (dimension T) in $\boldsymbol{X}$. Each spectrum comprises P spectral measurements such as absorbance values. An element of this table, x_{ijk} represents the jth spectral value of the kth spectrum of the ith batch. It is not assumed that, from one batch to another, the degree of progression of the sample of the same index k would be the same. Such an array can be processed by a direct three-way table approach, such as PARAFAC.[3] We want here to assess, from each spectrum, a numerical value y which will represent the degree of transformation of the corresponding sample. As in the usual regression methods, the developed predictive model can eventually be applied on "unknown" spectra. The assessed values of y will be an indicator of the progression of the transformation and will be used as indicators for managing the process. Contrary to usual regression methods, it is supposed that a quantitative reference value, which could be exploited for establishing the predictive model, is not available.

Principle and utilisation of the simulated annealing method

Basic assumptions

The method of SA[1,2] is a technique that has attracted significant attention as being suitable for optimisation problems in which a desired global extremum is hidden among many, poorer, local extrema. This method is particularly adaptable to complex combinatorial problems and to non-linear optimisation. Before describing the SA variant used in the present study, it is necessary to have precise basic assumptions which are necessary for establishing a predictive model.

1) It is assumed that the spectra contain information about the degree of transformation of the corresponding sample. The degree of transformation of the sample collected at time t in the batch n is characterised by a numerical value y_{nt}, which can be assessed as a linear combination of the corresponding spectral data:

© IM Publications Open LLP 2000

$$y_{nt} = \sum_{j=1}^{P} x_{njt}\beta j + \varepsilon \tag{2}$$

The linear model is chosen owing to its simplicity. However, Equation (2) does not imply that y is linearly dependent on time. The problem is to estimate the coefficients of vector β given the three-way array $\boldsymbol{X}$. The vector β is of course assessed once for all the batches in the calibration set.
2) It is also supposed that y is *monotonous* as a function of time. The degree of transformation of a sample cannot decrease when the time increases:

$$\forall \{t_2, t_1; t_2 > t_1\} \quad y_{nt2} \geq y_{nt1} \tag{3}$$

This assumption is reasonable for most of the natural transformations, in which at least one biochemical component appears or disappears in the course of the process. This assumption leads us to define a quality criterion in the estimation of $\boldsymbol{\beta}$. Let us suppose that at a moment of the iterative procedure of the SA method, we have two concurrent estimations of $\boldsymbol{\beta}$, namely $\hat{\beta}_1$ and $\hat{\beta}_2$. We have

$$y_{1nt} = \sum_{j=1}^{P} x_{njt}\hat{\boldsymbol{\beta}}_{1j} \text{ and } y_{2nt} = \sum_{j=1}^{P} x_{njt}\hat{\boldsymbol{\beta}}_{2j} \tag{4}$$

Where $\hat{y}_{1nt}$ and $\hat{y}_{2nt}$ are the estimation of the degree of transformation of the sample of batch n at time t obtained by using the vectors $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$, respectively. For a given batch n, the estimatd values of $\boldsymbol{y}$ obtained from the two concurrent models can be gathered in two vectors of T elements $\hat{\boldsymbol{y}}_1$ and $\hat{\boldsymbol{y}}_2$, respectively. The assumption given in Equation (3) imposes that the rank in increasing order of the estimations of $\boldsymbol{y}$ within a given batch should be equal to the rank of the sampling time. Among the two vectors $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ it is, therefore, reasonable to prefer the one which leads to a more correct ranking of the values within a batch. Thus, a statistical criterion of the accuracy of the model is easily obtained from the correlation coefficient of ranks (Spearman's coefficient) between the expected ranks defined from the sampling time and the observed ones defined by ordering the values in increasing order. When several batches are available ($N > 1$) it is easy to use the same criterion. We do not try to establish any direct comparison of the degree of progression between samples belonging to different batches. The $N \times T$ estimated values of y obtained from Equation (2) are ranked within the corresponding batch (i.e. from 1 to T). For a given estimation of $\boldsymbol{\beta}$, a single Spearman's coefficient between the expected and observed ranks can then be assessed for the whole set of data. This Spearman's coefficient is the desired criterion of accuracy of the model. It must be noticed that this criterion does not allow a completely unambiguous estimation of $\boldsymbol{\beta}$. The multiplication of $\boldsymbol{\beta}$ by any arbitrary positive constant would indeed lead to the same value of the Spearman's coefficient, because the observed ranks would not change. In order to eliminate this ambiguity, the estimated vector $\hat{\boldsymbol{\beta}}$ is normalised.

Application of simulated annealing

A SA iterative procedure is used for estimating β from the experimental three way array $\boldsymbol{X}$. The heart of the method of SA is an analogy with thermodynamics, specifically with the way that liquids freeze and crystallise or metals cool and anneals. The basis of the analogy and the rationale of the fundamental algorithm will not be detailed here, but can be found in many articles[1,2]. In order to apply SA, one must define the following elements:
1. A description of the possible system configurations. In the present case, a configuration is equivalent to a given estimation $\hat{\boldsymbol{\beta}}$.
2. A control parameter T (analogous of temperature) and *an annealing schedule*. The meaning of T will be explained later. T is set to an initial value T_0 at the beginning of the SA procedure and regularly decreases in the course of the iteration process.

© IM Publications Open LLP 2000

3. An objective function E (analogous to energy) whose *maximisation* (in the present case) is the goal of the procedure. E is here the Spearman 's coefficient between observed and expected ranks of $\boldsymbol{y}$, as previously described.
4. A generator of random changes in the configuration. These changes are the "options" presented to the system. The role of this generator is to propose to the system some modification of the current estimation . The choice of such a generator is not easy. We have chosen to generate changes able to take into account the continuous nature of the spectra. The procedure consists in adding a random Gaussian curve to the current vector . Two numbers μ and α are randomly generated. μ represents the centre of the Gaussian curve and can range from *1* to *P*. α regulates the intensity of the perturbation. The perturbation is given by:

$$\hat{\beta}_{m_i} = \hat{\beta}_{k_i} + \alpha T \left(\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{i-\mu}{\sigma})^2} \right) \quad \{i = 1,...,P\} \tag{5}$$

Where $\hat{\beta}_{m_i}$ and $\hat{\beta}_{k_i}$ represented the ith element of the modified vector $\hat{\boldsymbol{\beta}}_m$ and the ith element of the non-perturbed vector $\hat{\boldsymbol{\beta}}_k$, respectively. As the temperature T decreases with the number of iterations, the perturbation is less and less important when the number of iterations increases. σ is a constant regulating the "narrowness" of the perturbation.
Thus, the SA algorithm is as follows:
#1 *Initialisation*: The temperature T is set to T_0. The elements of $\hat{\boldsymbol{\beta}}_k$ (current estimation of $\boldsymbol{\beta}$) are also set to random small positive and negative values. The value of the current energy E_k (Spearman's coefficient) is assessed using $\hat{\boldsymbol{\beta}}_k$ and the table $\boldsymbol{X}$.
#2 *Perturbation*: The configuration is perturbed by applying Equation (5), leading to a new vector of the coefficients $\hat{\boldsymbol{\beta}}_m$. A new value of the energy, E_m, is then assessed from $\hat{\boldsymbol{\beta}}_m$ and $\boldsymbol{X}$.
#3 *Possible selection of the new configuration*: Two values of energy, E_k and E_m, obtained from $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\beta}}_m$, respectively, are available. If $E_m \geq E_k$, the new configuration is considered as preferable to the current one. In this case, $\hat{\boldsymbol{\beta}}_m$ is selected and normalised and becomes the current estimation $\hat{\boldsymbol{\beta}}_k$. E_m also replaces E_k. On the contrary, if $E_m < E_k$, the new configuration is selected with a probability $prob = \exp(\frac{(E_m - E_k)}{T}$. This rule, which is the more fundamental aspect of SA, means that a new configuration which is less performing than the current one may sometimes be selected. It allows the algorithm to avoid false solutions such as local maximum. As the temperature T decreases, the probability of finding a new solution less efficient than the current one also decreases and becomes negligible for very small values of T.

Steps #2 and #3 are repeated until a pre-defined number of iterations is reached. The current vector $\hat{\boldsymbol{\beta}}_k$ is eventually the final estimation of $\boldsymbol{\beta}$. During the SA algorithm, the temperature T is regularly lowered, following the pre-defined annealing schedule.

Experimental

Sample collection and spectral acquisition

The sample collection is the same as the one used in another work.[3] Some complementary details on the nature of the samples can be found in this study. Four varieties of spring barley, named PYRAMID, FERMENT, ALEXIS and CELINKA and one variety of winter barley, named TULIP, were studied. The samples were collected daily during the malting process performed during the pilot scale, from the end of steeping until the end of germination. As the germination process lasted five days, the time sequence comprised six malting steps numbered from 1 (end of steeping), to 6 (end of germination). Thirteen independent batches were performed in this way, leading to 13 (batches) × 6

© IM Publications Open LLP 2000

(steps) = 78 samples. Two aliquots of each wet sample were processed in a laboratory grinder. The Visible-NIR spectra of the ground materials were acquired in the reflection mode between 400 to 2500 nm at 2 nm intervals, with a monochromator spectrometer (model 6500, NIRSystems, Silver Spring, MD, USA). Three spectra were acquired for each aliquot. The spectra collection, therefore, included 468 spectra (2 aliquots × 3 spectra × 13 batches × 6 malting steps).

Data processing

In order to reduce the effect of the specular reflection, the spectra were first processed by the multiple scatter correction method[4] taking the average spectrum of the whole collection as reference. SA was then applied on the corrected spectral data. For each test, the maximum number of iteration was set to 10,000. The temperature was initially set to 0.02 and decreased by 10% every 100 iterations. In the perturbation procedure [Equation (5)] α ranged from –1000 to 1000 and σ was set to 10. These values were found to be relevant in preliminary tests. In order to estimate the accuracy of SA, a cross-validation procedure was applied. The 36 (2 aliquots × 3 spectra × 6 steps) spectra belonging to one among the 13 batches were removed from the calibration set and SA was applied on the remaining spectra. The SA procedure was repeated 13 times, by removing the spectra belonging to each batch one after the other. Each test led to an estimation of the β coefficients. The values of the Spearman 's coefficients obtained on the batches which were not in the calibration set (cross-validation set) were an indication of the performance of the model. The software for performing SA on NIR spectra has been written in C++ language in the authors' laboratory.

Results

Performance of the simulated annealing procedure

Table 1 gives the Spearman's coefficient obtained between the observed and the predicted ranks after the 13 independent tests. For each trial, the coefficients obtained during the calibration phase were close to 0.97. The results obtained in cross-validation were less stable and ranged from 0.90 to 0.99. The results are also presented in the form of allocation matrices (Table 2). In these matrices, the rows indicate the actual rank (i.e, the malting step) and the columns, the rank found from the assessed degree of progression of the samples. Each number inside the matrix gives the corresponding percentage. For example, the number at the third row and the fourth column of the matrix indicates the percentage of spectra of samples collected at the third malting step wrongly ranked in the fourth position. The numbers on the diagonal of the allocation matrices, therefore, represent the percentage of spectra correctly ranked. The results in calibration and cross-validation were obtained by summing up the results of the 13 independent tests. On average, 83.8 and 76.9 % of the spectra were correctly allocated, for the calibration and the cross-validation set, respectively. It must be noted that the errors mainly appeared between neighbouring malting steps. For example, in cross-validation, 25.6 % of the spectra acquired at the 6th malting step were attributed to the 5th step. However, the confusions did not occur between the 6th actual step and the

Table 1. Spearman's coefficient between observed and predicted ranks of the maling steps.

Batches[a]	Calibration	Cross-validation
Pyramid 321	0.96	0.97
Pyramid 322	0.97	0.99
Pyramid 323	0.97	0.94
Ferment 324	0.97	0.96
Ferment 326	0.97	0.96
Alexis 327	0.97	0.95
Alexis 330	0.97	0.90
Celinka 328	0.96	0.92
Celinka 329	0.97	0.98
Celinka 331	0.97	0.93
Celinka 332	0.97	0.98
Tulip 333	0.97	0.95
Tulip 334	0.96	0.97

[a] The name of the batch is formed by the name of the barley variety followed by a numeric code (same codes as Allosio *et al*.[3]

© IM Publications Open LLP 2000

Table 2. Observed v. Predicted malting steps.

Calibration set						
	1	**2**	**3**	**4**	**5**	**6**
1	83.7	16.1	0.2			
2	15.0	81.1	3.7	0.2		
3	1.4	2.7	89.4	6.4	0.1	
4		0.1	5.7	88.4	4.8	1.1
5			0.1	5.0	78.1	16.8
6			0.9	0.0	17.0	82.2
Cross-validation set						
	1	**2**	**3**	**4**	**5**	**6**
1	80.8	17.9	1.3			
2	16.7	74.4	9.0			
3	2.6	7.7	78.2	11.5		
4			10.3	84.6	5.1	
5			1.3	3.8	69.2	25.6
6					25.6	74.4

Rows: actual malting steps; columns: predicted malting steps.
Results expressed as percentage.

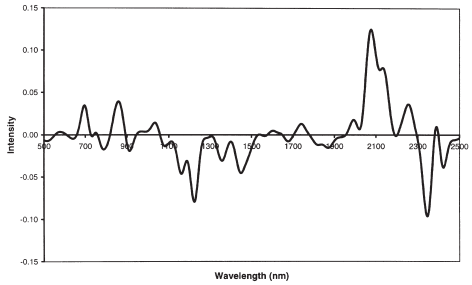

Figure 1. Estimated coefficients of the model as a function of the wavelengths (example).

predicted steps numbered 1 to 4. Figure 1 shows the curve of an estimation of the coefficients (elements of $\hat{\boldsymbol{\beta}}$ as a function of the wavelengths obtained in an experiment. According to the perturbation procedure [Equation (5)], this curve was the sum of Gaussian curves, which is visible on the graph. It was not possible to find a direct spectral interpretation of these coefficients. Moreover, the general shape of the values of $\hat{\boldsymbol{\beta}}$ was very different from one test to another. This was not surprising: in NIR spectroscopy, it is commonly observed that the absorbances at different wavelengths are highly correlated. In these condi-

Table 3. Correlation matrix between variables assessed from 13 independent models.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1	1.00												
2	0.99	1.00											
3	0.99	1.00	1.00										
4	0.99	0.98	0.99	1.00									
5	0.98	0.99	0.98	0.97	1.00								
6	0.98	0.96	0.98	0.97	0.93	1.00							
7	0.99	0.99	0.98	0.96	1.00	0.95	1.00						
8	0.99	0.98	0.99	0.97	0.96	0.98	0.97	1.00					
9	0.99	1.00	0.99	0.99	0.99	0.96	0.99	0.97	1.00				
10	0.98	0.98	0.98	0.99	0.95	0.97	0.96	0.98	0.97	1.00			
11	0.97	0.99	0.98	0.97	0.97	0.95	0.98	0.96	0.98	0.97	1.00		
12	0.99	0.99	0.99	0.97	0.97	0.98	0.98	0.99	0.98	0.96	0.98	1.00	
13	0.97	0.97	0.98	0.96	0.94	0.98	0.96	0.97	0.96	0.97	0.96	0.98	1.00

© IM Publications Open LLP 2000

tions, the estimations of $\hat{\beta}$ obtained in independent tests are actually infinite. However, many very different estimations of the coefficients of the model might lead to similar estimations of the degrees of modification. In order to verify this hypothesis, the values of the degree of modification obtained from the 13 independent tests were compared. As each test gave an estimation of $\boldsymbol{\beta}$,13 estimations $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, ..., \hat{\boldsymbol{\beta}}_{13}$ were obtained. It was possible to assess the corresponding estimated degrees of modification $\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2, ...\hat{\boldsymbol{y}}_{13}$ for the wholecollection (468 spectra). The correlation matrix between these different estimations of $\boldsymbol{y}$ is given on Table 3. The correlation coefficients were always high and ranged from 0.93 to 1. From this observation, it was concluded that $\boldsymbol{y}$ could actually be seen as a latent variable, representative of the degree of modification of the samples in the course of the process.

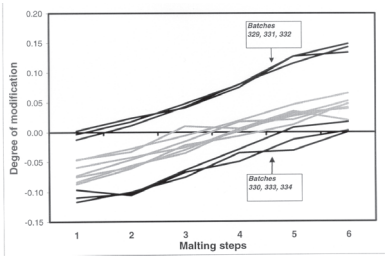

Figure 2. Degree of modification as a function of the malting steps (13 batches). Upper dark curves: batches 329, 331 and 332 showing rapid modifications. Lower dark curves: batches 330, 333 and 334 showing slow evolutions.

Degree of modification as a function of time

Figure 2 shows the evolution of the assessed degree of modification of the samples as a function of the malting steps for the 13 batches. These curves were obtained by averaging the results obtained from the repetitions. As expected, the degree of modification generally increased with time. Some of the curves presented a sigmoidal shape that is typical of growth curves. The curves of the degree of modification as a function of time were in accordance with the results of Allosio *et al.*,[3] obtained from the same set of spectral data. By using the PARAFAC model, these authors have identified three groups of samples according to their behaviour during malting. The batches coded *330* (variety ALEXIS), *333* and *334* (two batches of variety TULIP) in the mentioned work formed a single group and were identified as weakly modified during the malting process. In Figure 2, the three corresponding curves are placed at the bottom of the graph, with small values of the degrees of modification. In the same way, another group was formed from three batches of the variety CELINKA coded 329, 331 and 332, which was found to be highly modified. In the present study, the corresponding samples showed high degrees of modification and were represented by the upper three curves.

Conclusion

Simulated annealing has rarely been used in NIR for the establishment of predictive model. This method is easy to use and is adapted to data presenting non-linear characteristics. In the present application, it has given a non-expensive method for the calibration of NIR instruments, which did not require the assessment of laboratory reference values. The presented procedure can be applied in the numerous industrial applications that involve continuous time-dependent transformations.

References

1. S. Kirkpatrick, C.D. Gelatt and MP Vecchi, *Science* **220**, 671 (1983).
2. W.H. Press and S.A. Teukolsky, W.T. Vettering and B.P. Flannery, in *Numerical Recipes in C, 2nd Edition*. Cambridge University Press, Cambridge, UK, p. 444 (1994).
3. N. Allosio-Ouarnier, P. Courcoux, D. Bertrand and P. Boivin, in *Near Infrared Spectroscopy: Proceedings of the 9th International Conference*, Ed by A.M.C. Davies and R. Giangiacomo. NIR Publications, Chichester, p. 273 (2000).
4. P. Geladi, D. MacDougall and H. Martens, *Appl. Spectrosc.* **39**, 491 (1985).

© IM Publications Open LLP 2000