Vis-NIR Measurement of Soluble Solids in Cherry and Apricot by PLS Regression and Wavelength Selection

Paolo Carlini, Riccardo Massantini, and Fabio Mencarelli*

Istituto di Tecnologie Agroalimentari, Facoltà di Agraria, Università degli Studi della Tuscia, Via S. C. De Lellis snc, I-01100, Viterbo, Italy

Experimental results are presented on the use of partial least squares (PLS) regression and wavelength selection for the definition of models for visible-near-infrared (Vis-NIR) evaluation of soluble solids content in fruits. First, the relatively easy to deal with—but still not studied in the literature—case of cherry fruit is presented in detail. By using a very simple selection scheme, involving the subsampling of the spectral interval from 600 to 1100 nm with a fixed step, accurate models were found, consistently showing very favorable combinations of SEC and SEP values, in the 0.50 °Brix range for a total variation of about 15 °Brix. Apricot fruit represented a more difficult species, and wavelengths to be included in the calibration had to be individually selected for the best results. Nevertheless, parsimonious models could be found, including a total of 38 spectral lines and leading to SEP values at the 0.75 °Brix level.

Keywords: *NIR; soluble solids content; nondestructive analysis; cherry; apricot; wavelength selection; fruit quality*

INTRODUCTION

Visible-near-infrared (Vis-NIR) spectroscopy is an established technique for determining chemical constituents in agricultural products (Williams and Norris, 1987; Osborne et al., 1993) which is gaining increased attention in the field of postharvest quality evaluation of fruits, comparable to that devoted to different physical methods. Most recently, McGlone and Kawano (1998) determined with very good accuracy both dry matter and soluble solids content (SSC) of kiwifruit, while firmness was not satisfactory predicted; Peiris et al. (1997; 1998) determined soluble solids content of peaches and processing tomatoes. Carlini et al. (1999) presented preliminary results on the use of wavelength selection methods for the accurate evaluation by PLS regression of soluble solids in cherries, apricots, loquat fruits, and peaches.

Experiments reported here were primarily concerned with Vis-NIR interactance measurement of SSC on fruits, highly correlated to total sugar content, one of the most important quality parameters (Reid, 1992).

Fresh fruits are invariably characterized by a very high moisture content, and spectral regions in the Herschel-infrared, from 700 to about 1100 nm, assume a particular relevance. This interval presents a few distinctive advantages: water absorption peaks are less strong and broad and do not risk to mask spectral information correlated to low concentration constituents; light can penetrate much farther in fruits of many different species. Wavelengths belonging to the visible part of the electromagnetic spectrum are sometimes also included—chlorophyll and anthocyanins absorption bands, among others, belong to this spectral range. Sugiyama (1999) found on both "Andes" and "Earl's" melon cultivars a high (inverse) correlation between absorbance at 676 nm and sugar content (°Brix) and furthermore the spectral range between 450 and about 690 nm consistently showed an high (inverse) correlation.

Partial least squares (PLS) regression (Geladi and Kowalski, 1986) was used early in the development of models for fruits and vegetables, and many applications are still reported today (see, e.g., Slaughter et al., 1996; McGlone and Kawano, 1998). Recently however, many researchers have closely scrutinized the working hypothesis that processing the full spectrum (hundreds of wavelengths), or at least a large subset of contiguous lines, can always lead to the best results: suited models may often involve few tens of carefully selected wavelengths in PLS-like schemes. For example, Centner et al. (1996) developed an elimination method, called uninformative variable elimination-PLS (UVE-PLS), and tried it on chemical data sets: it outperformed in every instance standard full spectrum PLS, chiefly when many nonuseful wavelengths were known to exist. Osborne et al. (1997; 1999) proposed a forward selection technique for the automatic identification of wavelengths to be included in a PLS regression model. In the evaluation of kiwifruit SSC, the best results were obtained beginning with a set of four or eight useful wavelengths plus some casually chosen and utilizing leaps 23 wavelengths long in the search. The calibrations obtained improved on full spectrum PLS 99% of the time in a "Monte Carlo" setup. Effects of preprocessing operations have not been discussed in detail, with results on standardized and unstandardized data grouped together.

The study reported here was particularly aimed at an assessment of the usefulness of some forms of wavelength selection in the building of PLS models for SSC in cherry and apricot. In the first part, simple subsampling schemes will be discussed in detail and compared to reference full spectrum models for cherry. The last part of the paper will be concerned with

* Author to whom correspondence should be addressed (fax +39-0761357498; e-mail mencarel@unitus.it).

10.1021/jf000408f CCC: $19.00
Published on Web 11/01/2000

calibrations for apricot fruit, probably more interesting from the point of view of prospective practical applications, but which need a less straightforward selection process in order to obtain satisfying performance.

The species studied have never been considered before for Vis-NIR nondestructive quality determination, with or without the adoption of modern regression techniques or wavelength selection methods (Kays, 1999).

MATERIALS AND METHODS

Fruit Sample Sets. Cherry fruits (*Prunus serotina* L. cv. "Ravenna") for the experimentation were hand-harvested from a small orchard close to Viterbo, Italy, at the beginning of Summer 1998, selecting first class samples, uniform in size and color, then immediately brought to the laboratory, and evaluated at room temperature (20 ± 1 °C). The same procedure was repeated many times, for about 3 weeks. Apricot fruits (*Prunus armeniaca* L. cv. "Boccuccia Spinosa" and "Errani") were hand harvested close to Rome during the months of June and July 1998 and were then processed similarly to cherries.

Vis-NIR Method and Constituent Measurement. Absorption spectra were measured on each intact fruit using a fast (1.8 scans/s) Vis-NIR (400−2500 nm) spectrophotometer *NIRSystems* (Silver Spring, MD) model 6500, with 2 nm spectral resolution. A fiber optic probe, about 1.2 m long and working by interactance, was fitted to the system, consisting of the following: a central bundle of fibers returning the light and an outer ring, about 0.8 cm in diameter, emitting the light interacting with the sample. For system management and calibration *NIRS-2* Version 4.00 package by *Infrasoft International*, running under the *Microsoft MSDOS* operating system, was adopted. Spectra were measured by hand-placing the interactance probe against the fruit at a random position along the equator. Fifteeen individual scans were averaged for the recording of each spectrum

SSC readings were taken for flesh cut from the same location on the fruit where the optical scans were conducted. To precisely evaluate each reference SSC, fruit pulp was slightly comminuted and centrifuged for 5 min by an *ALC micro centrifugette 4204*. The supernatant was then analyzed by a laboratory refractometer built by *Officine Galileo* (Florence, Italy) model RG701. For the ripest cherry samples, rich in interfering red-black pigments, a small quantity of *PVPP* (polyvinylpoly-pyrrolidone) was added to the *Eppendorf* vial containing the pulp during the centrifugation in order to facilitate the ensuing refractometer reading.

Data Analysis Generalities. Prior to model building, a randomized procedure, not followed by any manual manipulation, split the whole spectrum/reference SSC data sets into a calibration set, used for model optimization, and a smaller prediction set, used for validation on independent samples.

Preliminary experiments led to the adoption of a slightly modified form of PLS—involving the normalization of the residuals at the end of each iteration—as our multivariate regression algorithm. It was systematically able to produce better final results than plain PLS on the considered data set, both cherries and apricots, irrespective of the other modeling choices. Calibrations have been further studied only if their complexity in terms of number of PLS factors was below a maximum order, established case by case by an 8-fold cross-validation strategy, that is the minimum of standard error of cross validation (SECV) versus number of factors, the latter varying from 1 to 10.

The most commonly used statistics, i.e., SEC, SEP, and R^2, were computed together with prediction bias, that is the systematic component (the mean value) of the errors when the model is applied to prediction set samples. Finally, the adimensional ratio called standard deviations ratio, $SDR \triangleq Data\ Std\ Dev/SEP$ (Chang et al., 1998), is also reported. Additionally, to better characterize apricot models quantile-quantile plots (Venables and Ripley, 1997) have been computed by using *R* statistical system (available free of charge through http://www.r-project.org) running under *Linux* operating system.

Table 1. Soluble Solids Level in Calibration and Prediction Cherries and Apricots (°Brix)

	samples	mean	std dev	min.	max.
		Cherries			
total	158	13.69	2.96	8.42	23.55
calibration	100	13.71	3.04	8.42	23.55
prediction	58	13.65	2.86	9.60	20.85
		Apricots			
total	162	11.40	1.95	7.45	17.00
calibration	100	11.40	1.85	7.80	16.70
prediction	62	11.41	2.12	7.45	17.00

Preprocessing Steps. These were very helpful in reducing the adverse effects of the physical (vs *chemical*) sample structure on the quantitative evaluation of the constituents and were adopted in various combinations.

Derivation. Numerical derivation eliminated the spectra of offset (I der.) and offset+slope (II der.) noninformative components. Unfortunately, as a secondary unwelcome effect, measurement noise could be greatly amplified. Smoothing operations were therefore used, consisting of a preliminary averaging over a fixed number of points (*segment* option). The beneficial effect of derivativizing was often sensitive to the *gap* chosen, i.e., the lag between each pair of data points processed in the course of the computation.

Detrending correction (Barnes et al., 1989). In this case a fit of each individual spectrum with a quadratic polynomial was performed and the residual utilized in the calibration.

MSC (Multiplicative Scatter Correction) (Isaksoon and Naes, 1988). This is one of the most widely used methods. Each spectrum was linearly regressed against the mean spectrum and the fitted constants used to compute the corrected spectrum.

SNV (Standard Normal Variate) correction (Barnes et al., 1989). Each spectrum separately was normalized to null mean value and unit variance.

Wavelength Selection. In the first place simple spectral selections were studied, with wavelengths equally spaced in the restricted segment 600−1100 nm and not individually picked. Therefore, after the computation of various combinations of preprocessing steps, each spectrum was subsampled, keeping only a sequence of wavelengths spaced from a minimum step equal to 2 nm (corresponding to the full acquired spectrum) to a maximum of 20 nm. Sequences shifted forward half the corresponding step were also constructed: for instance, for an 8 nm step, 604 nm, 612 nm, 620 nm, and so on, in addition to the standard 600 nm, 608 nm, 616 nm, ... The first part of each acquired spectrum, from 400 nm up to 600 nm, was completely discarded since it was affected by measurement noise.

For apricot suited wavelength intervals were found by way of trial and error procedures: starting with the full Vis-Herschel interval from 600 to 1100 nm, possibly subsampled with a fixed step as for cherry, noninformative segments were tentatively purged and the resulting performance evaluated.

RESULTS AND DISCUSSION

Cherry Fruit. Calibration and validation sets included 100 and 58 samples, respectively, and had the levels of soluble solids shown in Table 1. A second derivative pretreatment computed with a 10 point gap invariably led to the best performances in predicting SSC.

The first calibrations, using 2 and 10 nm selection steps, clearly indicated that utilizing a maximum of 10 PLS factors, as suggested by the 8-fold cross-validation procedure, was indeed reasonable, albeit slightly unexpected for the Vis-Herschel-Infrared interval, usually characterized by broad and "smooth" peaks. We checked

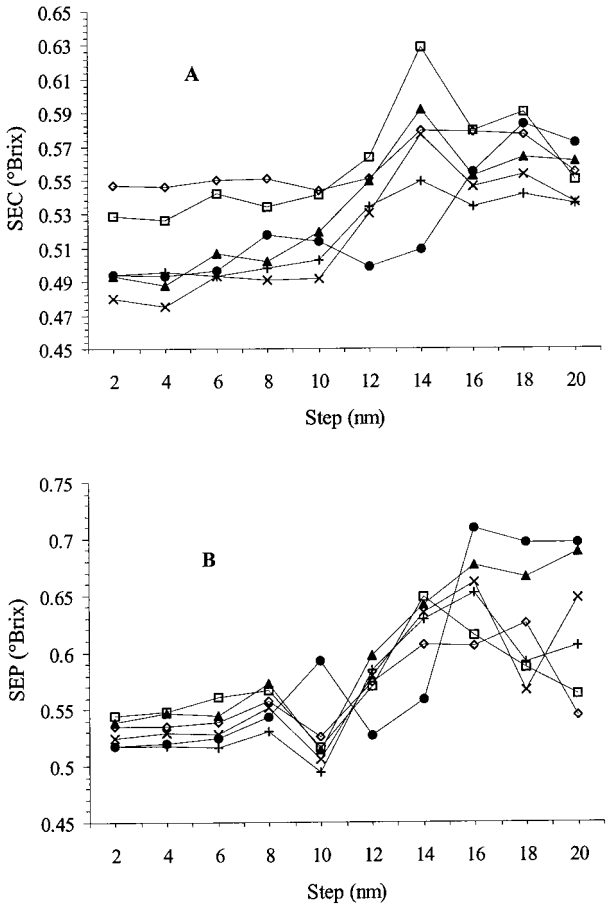

Figure 1. SECs (A) and SEPs (B) for cherry fruit as a function of the subsampling step and for six different preprocessing combinations (10 PLS factors): (◇) SNV&Det, Seg = 5; (□) SNV&Det, Seg = 1; (▲) MSC, Seg = 1; (×) MSC, out, Seg = 1; (+) MSC, out, Seg = 3; and (●) MSC, out, Seg = 3, shifted.

this number by varying the number of cross-validation groups in the range from 4 to 10, and in all the cases the same value ±1 was found by the automatic software procedure. As a rule, SECV(10 factors) values were equal to about 0.75 × SECV(6 factors) and about 0.8 × SECV(8 factors). In fact, both Osborne et al. (1997) on kiwifruits and Slaughter (1996), on tomatoes, employed a large number of factors—in the range from 10 to 30—while dealing with the same spectral interval. Osborne et al. (1997), in particular, used as many as 19 and 25 factors for unstandardized and standardized data, respectively, on the basis of a preliminary full spectrum calibration. In contrast with this approach, Bangalore et al. (1996) tried to jointly optimize the number of PLS factors and the choice of the individual wavelengths. Garrido Frenich et al. (1995), however, had found no evidence that the additional computational effort was really worth implementing.

For 10 PLS factors the first five plots of Figure 1 depict the resulting model accuracy, in terms of SEC (A) and SEP (B), as a function of the subsampling step and for five different combinations of the remaining pretreatments. The possible elimination of a couple of outliers is labeled *out*. A sixth plot, labeled (●), is discussed below.

For steps increasing from 2 to 10 nm and a corresponding large drop in the number of retained wavelengths (from 240 to 48), SECs did not grow appreciably, the largest increase taking place for MSC, Seg = 1 preprocessing options (Figure 1, (▲)): from a minimum of 0.49 °Brix for a 4 nm step to a maximum of 0.52 °Brix for a 10 nm step. For the combination labeled (+) in Figure 1 the increase is as small as 0.01 °Brix and definitely not statistically significant, from a minimum value of 0.49 °Brix, for 6 nm, to 0.50 °Brix, for 10 nm. More generally, except a "crisis" for the 14 nm step, SECs grow very slowly and, for the (+) series, and for a 20 nm step the SEC was still 0.54 °Brix.

By reducing the amount of spectral data processed by the regression algorithm while not considering at all the chemical assignment of each particular wavelength one would rather expect a sizable accuracy decrease on the calibration set, easily attributable to the elimination of useful information. The present findings suggest that the adopted PLS algorithm works better when spectral data is less correlated, thanks to the use of a moderately large step in the selection process. For steps as large as 10 nm, any possible information loss, probably small indeed due to the high collinearity, appears well compensated by a better working of the algorithm.

The foregoing considerations are at least qualitatively consistent with the findings of Osborne et al. (1997) who obtained the best results by using a leap 23 wavelengths long (corresponding to 76 nm for a 3.3 nm-spectral resolution apparatus) in their circular search scheme. The experimental setup was however different from ours, involving a different sampling technique, based on a fast photodiode array detector (Osborne et al., 1999). Furthermore, at variance with McGlone and Kawano (1998), the authors did not make use of derivative preprocessing operations, expected to appreciably reduce the correlation between neighboring data points. It should also be noted that the specific search scheme used does *not* exclude at all the possibility that many adjacent wavelengths could end up selected by the algorithm.

Also, Lammertyn et al. (1998) found useful a 10-fold subsampling/averaging of the spectra (acquired with 0.5 nm resolution) prior to the subsequent preprocessing operations. This approach bears some resemblance to that studied here, especially so if the preprocessing and wavelength selection operations were interchanged. Although, Lammertyn and co-workers did not discuss the effect of varying the subsampling step on models for soluble solids.

Notably, every sequence of models attained the minimum SEP for a 10 nm step, that is only one wavelength in every five acquired, half the size of the spectral interval corresponding to the gap (i.e. 2 nm by 10 points divided by 2). The most favorable set of preprocessing options, leading to the lowest SEP, is that including MSC scattering correction, with the elimination of two outliers and a smoothing over a three-point segment (Figure 1, (+)). Alternatively, very similar results—only slightly worse in terms of SEP—could be obtained without any form of preliminary smoothing (Figure 1, (×)).

Figure 1, series (●), depicts summarizing figures of merit of calibrations involving the same preprocessing steps as for (+) series but including wavelengths forward shifted half the subsampling step, that is "maximally different" from the former but equally spaced. Interestingly, whereas very similar trends can be observed on the SEC plot, the computed SEP value for a 10 nm step, 0.59 °Brix, is as much worse than the average typical of 2−8 nm step calibrations, ~0.54 °Brix, as the corresponding SEP in series (+), 0.49 °Brix, is

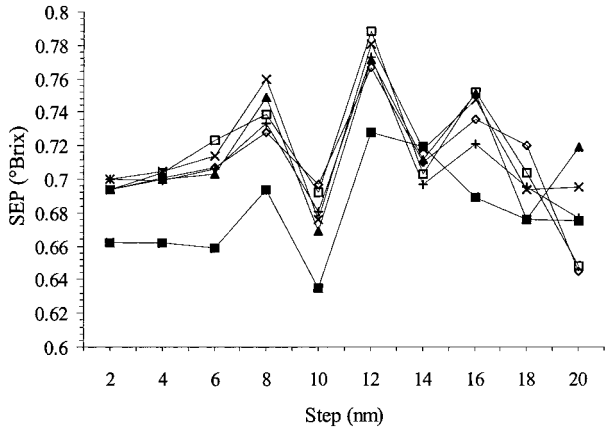

Figure 2. SEPs for cherry fruit as a function of the subsampling step and for seven different preprocessing combinations (eight PLS factors): (◇) SNV&Det, Seg = 5; (□) SNV&Det, Seg = 1; (▲) MSC, Seg = 1; (×) MSC, out, Seg = 1; (+) MSC, out, Seg = 3; and (■) MSC, out, Gap = 9, Seg = 3.

Table 2. Statistics of Models for Cherry as the Number of PLS Factors Changes: SEC, SEP, and Bias in °Brix

factors	SEC	R^2	SEP	Bias	SDR
		Step 2 nm			
6	0.74	0.94	0.86	0.06	3.44
7	0.66	0.95	0.74	0.10	4.00
8	0.64	0.95	0.70	0.15	4.23
9	0.52	0.97	0.56	0.04	5.29
10	0.49	0.97	0.52	0.04	5.70
		Step 10 nm			
6	0.70	0.94	0.79	0.06	3.72
7	0.66	0.95	0.73	0.10	4.07
8	0.64	0.95	0.68	0.15	4.35
9	0.53	0.97	0.53	0.05	5.57
10	0.50	0.97	0.49	0.04	5.99

better. This result is a strong indication that the five series discussed above are characterized not only by reduced collinearities but also by combinations of especially favorable wavelengths if compared with different equally spaced choices. It can also per interpreted as recommending the individual selection of spectral lines to be included in the model, when dealing with more difficult species.

In Table 2 we report complete statistics of calibrations involving from 6 to 10 PLS factors, 2 and 10 nm selection steps, in the pretreatment conditions labeled as (+) in Figure 1 (MSC, out, Seg = 3). In particular, 9 and 10 factor calibrations performed very well in every respect, both in terms of SEC and SEP values, small and similar to each other, in terms of Bias, in fact as low as one tenth of the mean errors, and finally in terms of SDR, larger than 5 and close to 6. The second series of models consistently attains better SDR figures, up to 8% for 6 factors.

To further investigate some of the points discussed above we also analyzed calibrations involving only eight factors (Figure 2), that is two less than the number suggested by the cross-validation procedure. A trend already visible in Figure 1, toward one more especially favorable condition for a 20 nm step, becomes much more evident here. However, only for a 10 nm step the whole set of plots simultaneously shows a minimum SEP, exactly as happens in Figure 1 for 10 factors. Overall, it is a little puzzling that a selection step exactly half the gap in the derivative led to the lowest prediction errors, even if we argued above that this effect is partly due to an especially favorable initial

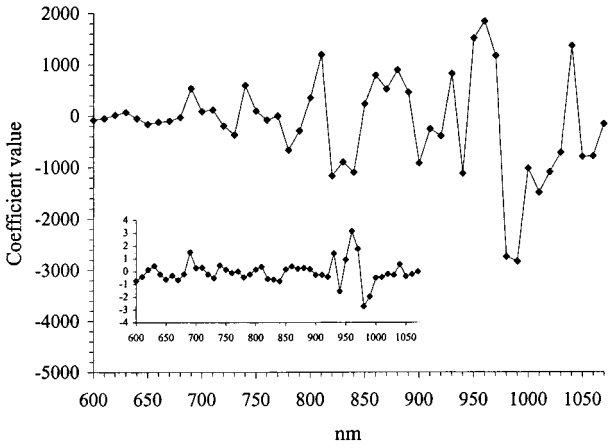

Figure 3. Coefficient vector for a 10 PLS factors model for cherry (step = 10 nm, MSC, out, Seg = 3), constant term = 16.59. Inset, the same multiplied by standard deviation at each wavelength.

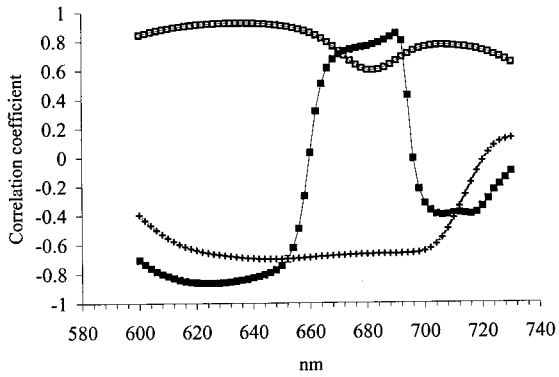

Figure 4. Correlation coefficient between SSC and absorption at each wavelength: (□) cherry, raw spectra; (■) cherry, second derivative spectra; and (+) apricot, raw spectra.

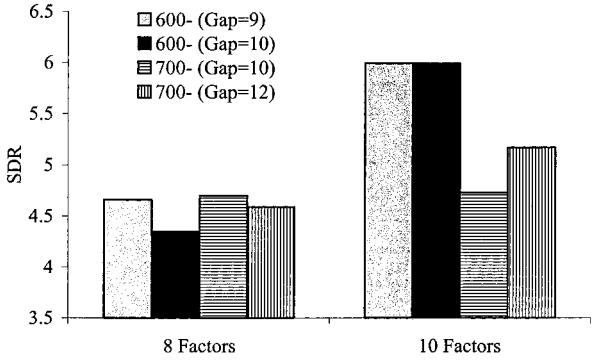

Figure 5. Representative models for cherries on the interval 700−1100 nm, compared in terms of SDR values to results presented before (step = 10 nm, MSC, out, Seg = 3).

offset in the subsampling of the spectra. Therefore, we investigated if our set of preprocessing options was still optimal for eight factors, to understand if the relation found was indeed robust to a variation in the number of PLS factors. SEP statistics obtained by using a nine point gap are appreciably lower (Figure 2, (■)), albeit characterized by overall trends confirming those discussed above. More generally, changing the gap below 8 or above 10 points led to plots either qualitatively similar to those displayed in Figures 1 and 2 or flattened out, with neither step favored over the others and definitely higher values.

Figure 3 shows the coefficient vector of the model labeled as (+) in Figure 1, for 10 factors and a 10 nm

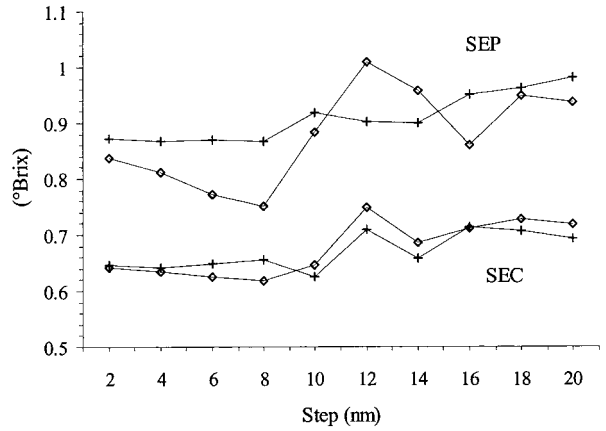

Figure 6. SECs and SEPs (°Brix) for apricot fruit as a function of the subsampling step (MSC, Gap = 10, Seg = 3): (+) simple subsampling and (◇) subsampling followed by individual selection.

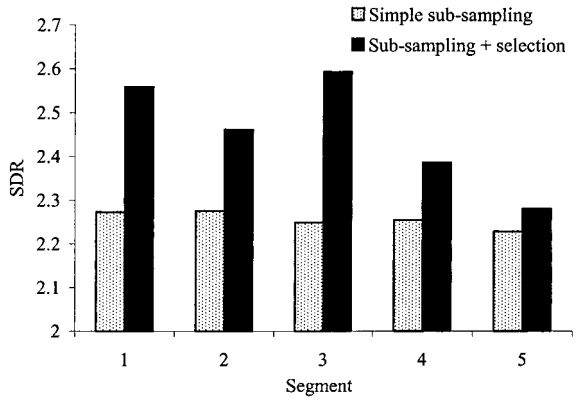

Figure 7. SDR statistics for apricot as a function of the segment used in the computation of the numerical derivative (step = 8 nm, MSC, Gap = 10).

step, characterized by a typical oscillatory-type behavior and remarkably consistent with spectroscopic studies of model solutions (Williams and Norris, 1987; Workman, 1996). Wavelengths assigned a small regression coefficient can have a sensible effect on the prediction due to a large range of variation. In fact, a much more uniform distribution of amplitudes is evident in the Figure 3, inset, which displays the same vector with each point multiplied by the respective data standard deviation.

Quite unexpectedly, raw spectra were moderately correlated, $R \cong +0.6$, with SSC in the chlorophyll band, probably due to the combined effect of different constituents (Figure 4, (■)). It should be stressed, however,

Table 3. Statistics of Models for Apricot as the Number of PLS Factors Changes: SEC, SEP, and Bias in °Brix

factors	SEC	R^2	SEP	Bias	SDR
Simple Subsampling					
6	0.75	0.84	1.01	0.08	1.94
7	0.68	0.86	0.91	0.03	2.14
8	0.66	0.87	0.87	0.02	2.24
Subsampling + Selection					
6	0.73	0.85	1.00	0.03	1.95
7	0.68	0.86	0.88	0.03	2.22
8	0.62	0.89	0.75	0.05	2.60

that at variance with other fruits studied by our group (by comparison, Figure 4, (+), shows the behavior of apricot) cherry is very low in chlorophyll content in all the considered ripening stages (Looney et al., 1996). On the other hand, a correlation plot of preprocessed spectra (MSC, second derivative), Figure 4, (■), displayed a broad negative peak centered at about 620 nm. Comfortingly, Delwiche et al. (1987) found a reflectance increase at about 600 nm during maturation of other stone fruits, notably peaches.

Figure 5 compares calibrations involving only wavelengths beyond 700 nm with a few noteworthy models already discussed. For 10 factors the elimination of the Vis interval led to a reduced accuracy, whereas for eight factors no significant differences could be observed. In the former case, slightly better statistics were obtained by the use of a 12 points gap. Overall, the contribution of the visible portion appears to be not negligible. A better feeling of its entity can be obtained considering that calibrations including this restricted interval alone are characterized by SEP values at the 1.1 °Brix level, a remarkable result, especially if compared with SSC range of variation.

Apricot Fruit. Calibration and validation sets included 100 and 62 samples, respectively, and had the levels of soluble solids shown in Table 1.

As for cherries, a second derivative pretreatment, computed with a 10 point gap and a three point segment, invariably led to the best performances in predicting SSC. Likewise, since the beginning scatter effects were treated by MSC for consistency with some of the well performing models for cherry, but in fact different schemes (i.e., SNV, detrendizing) led to very similar results, with figures of merit typically differing by not more than 3−4%, even without any correction at all. The usual 8-fold cross validation procedure suggested in this case the use of a maximum of eight PLS factors, a value which was checked by the same means already discussed above. As a rule, SECV(8 factors) values were equal to about 0.75 × SECV(4 factors) and about 0.9 × SECV(6 factors).

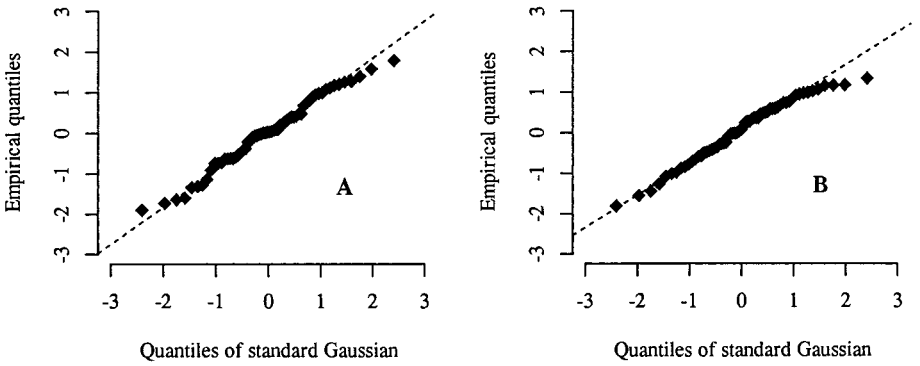

Figure 8. Quantile-quantile plots of model errors (°Brix) for prediction set apricot samples (eight PLS factors): (A) simple subsampling (step = 8 nm) and (B) subsampling followed by individual selection.

Calibrations optimized on the spectral segment going from 600 to 1100 nm led to both SECs and SEPs appreciably worse (Figure 6, (+)) if compared to the easier to deal with cherry fruit, also due to the lower number of factors reliably usable, that is eight instead of 10. Consistently, SEC values were comparable for eight factors (~0.6 °Brix), whereas SEP values were definitely higher in the apricots case, signaling a more pronounced tendency toward overfitting, counteracted by the cross-validation procedure. Remarkably, however, when studying the effect of the subsampling operation a general trend could be confirmed once more: even for steps larger than 10 nm values did not grow rapidly and for steps varying from 2 nm (240 wavelengths) up to 8 nm (60 wavelengths) essentially equivalent results could be obtained.

To investigate if these results could be somewhat improved by a more careful selection of wavelengths included in the model, a few subsegments were tentatively purged from the full Vis-Herschel interval, while keeping both the spectral range from 800 to 900 nm and a window around 1010−1030 nm, known from the literature to be highly informative (see, e.g., Workman, 1996).

Eventually, four disjoint segments remained—630−680 nm, 720−750 nm, 800−940 nm, and 950−1020 nm—a 50 nm wide interval in the chlorophyll band included, which in the case of apricots displayed a much more conventional behavior (Figure 4). Plots labeled as (◇) in Figure 6 present summarizing statistics of the latter calibrations (eight PLS factors), compared for a varying subsampling step to those discussed above, Figure 6, (+). The best results were evidently obtained for 8 nm, and a corresponding total of 38 (vs 60) wavelengths kept in the model, where SEP could be lowered from 0.87 to 0.75 °Brix and SEC could be slightly reduced too, from 0.66 to 0.62 °Brix. Unfortunately, the trend manifested for steps larger than 10 nm appear not easy to understand in detail. Arguably, some very general assumptions are no longer valid here, consider, f.i., that less than 26 wavelengths end up included in the model, a number becoming very close to that of PLS factors.

In Table 3 we report complete figures of merit of calibrations involving from 6 to 8 PLS factors and a 8 nm step, with or without the subsequent selection of most suited wavelength intervals. In Figure 8, quantile−quantile plots of prediction errors of eight factor calibrations are shown, in general very useful to diagnose possible diseases affecting the regression model. Indeed, only minor differences can be remarked, among which a reassuring slightly more linear ("more Gaussian") behavior in plot (B) of the central tract between −1 and +1 quantiles.

By varying the segment used in the computation of the derivative the soundness of the additional selection made could be confirmed, in particular its robustness to variations in the pretreatment options: Figure 7 shows consistently better SDR values, up to +0.34 for the reference choice of three points. A different selection, tentatively tried during the calibration development process, which excluded a spectral segment between 910 and 930 nm to obtain marginally improved SEP figures (about 0.73 °Brix for a 8 nm step), could not pass this test and was not further studied.

On the other hand, changing the gap option nullified by and large (data not shown) the beneficial effect of every tried wavelength selection strategy, thus confirming the high sensitivity of the models to such parameter, already pointed out above in relation to eight factor calibrations for cherry. Comfortingly, our experience clearly suggests that a suitable gap can be defined very early in the calibration development process, not changed anymore during the ensuing refinements, involving forms of wavelength selection or whatever.

Not including the visible portion of the spectrum led for apricot to worse models, characterized by SEPs at the 0.87 °Brix level for a 8 nm step and 31 wavelengths included, albeit still improving on the simply subsampled case, at the 0.90 °Brix level in the same conditions.

CONCLUSION

Short wave Vis-NIR reflectance (600−1100 nm) could be used to accurately predict °Brix on whole cherry. Common data pretreatments were effective for removal of baseline variation. Processing only a subset of the spectrum, down to 48 data points, did not worsen the predictive accuracy, which for a particular choice of the subsampling offset could even be slightly but consistently improved. For apricot, wavelengths to be included in the model had to be individually selected for the best results. However, in this case too sufficiently accurate models could be constructed on the restricted Vis-Herschel interval, readable by affordable spectrophotometers.

In the future it will be important to extend the investigation to calibrations including many varieties for each species, possibly coming from different orchards and multiple harvest years. In the past this general concern was deemed very critical for the prospective application of such techniques: the foreseeable cost of online sorting and grading systems was so high to believe that the first machines available would be installed only at big packinghouses, collecting produce coming from a lot of very unlike orchards. Today this point of view begins to change thanks to the development of cheaper systems, which could well be used on a local basis (Osborne et al., 1999). In this new perspective it will be of the utmost importance to define truly effective automatic or semiautomatic wavelength selection procedures, optimized for the concerned application field and usable by lightly trained operators.

ACKNOWLEDGMENT

This research was founded by MURST (ex 40%) project "Fisiologia postraccolta e aspetti qualitativi dei prodotti ortofrutticoli" and supported by the Italian Postharvest Working Group.

LITERATURE CITED

Bangalore, A. S.; Shaffer, R. E.; Small, G. W. Genetic algorithm based method for selecting wavelengths and model size for use with partial least-squares regression: application to near-infrared spectroscopy. *Anal. Chem.* **1996**, *68*, 4200−4212.

Barnes, R. J.; Danoa, M. S.; Lister, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772−777.

Carlini, P.; Massantini, R.; Mencarelli, F. Wavelength selection methods for PLS-based Vis-NIR evaluation of SSC in fresh fruits. To appear In *Proceedings of the NIR'99, 9th Inter-*

national Conference on Near-Infrared Spectroscopy; Verona, Italy, 13−18 June 1999.

Centner, V.; Massart, D.-L.; de Noord O. E.; de Jong, S.; Vandeginste, M. B.; Sterna, C. Elimination of uninformative variables for multivariate calibration. *Anal. Chem*. **1996**, *68*, 3851−3858.

Chang, W. H.; Chen, S.; Tsai, C. C. Development of a universal algorithm for use of NIR in estimation of soluble solids in fruit juices. *Trans. ASAE* **1998**, *41*, 1739−1745.

Delwiche, M. J.; Tang, S.; Rumsey, W. J. Color and optical properties of clingstone peaches related to maturity. *Trans. ASAE* **1987**, *30*, 1873−1879.

Garrido Frenich, A.; Jouan-Rimbaud, D.; Massart, D. L.; Kuttatharmmakul, S.; Martinez Galera, M.; Martinez Vidal, J. L. Wavelength selection method for multicomponent spectrophotometric determinations using partial least squares. *Analyst* **1995**, *120*, 2787−2792.

Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.

Isaksoon, T.; Naes, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc.* **1988**, *42*, 1273−1284.

Kays, S. J. Non destructive quality evaluation of intact, high-moisture products. *NIR News* **1999**, *10*, 12−15.

Lammertyn, J.; Nicolaï, B.; Ooms, K.; De Smedt, V.; De Baerdemaeker, J. Nondestructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Trans. ASAE* **1998**, *41*, 1089−1094.

Looney, F.; Webster, A. D.; Kupperman, E. M. Harvest and Handling Sweet Cherries for the Fresh Market. In *Cherries: Crop Physiology, Production and Uses*; Webster, A. D., Looney, D. E., Eds.; CAB International: 1996.

McGlone, V. A.; Kawano, S. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biol. Technol.* **1998**, *13*, 131−141.

Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd ed.; Longman Scientific and Technical: U.K., 1993.

Osborne, S. D.; Jordan, R. B.; Künnemeyer, R. Method of wavelength selection for partial least squares. *Analyst* **1997**, *122*, 1531−1537.

Osborne, S. D.; Künnemeyer, R.; Jordan, R. B. A low-cost system for the grading of kiwifruit. *J. Near Infrared Spectrosc.* **1999**, *7*, 9−15.

Peiris, K. H. S.; Dull, G. G.; Leffler, R. G.; Kays, S. J. Nondestructive determination of soluble solids content of peach by near-infrared spectroscopy. In *Proc. Sens. Nondestr. Test. International Conference*; Northeast Reg. Agric. Eng. Serv.: Ithaca, NY, 1997; pp 77−87.

Peiris, K. H. S.; Dull, G. G.; Leffler, R. G.; Kays, S. J. Near-infrared (NIR) spectrometric technique for nondestructive determination of soluble solids content in processing tomatoes. *J. Am. Soc. Hortic. Sci.* **1998**, *123*, 1089−1093.

Reid, M. S. Maturation and maturity indices. *Postharvest technology of horticultural crops*; Kader, A. A., Ed.; University of California, Division of Agriculture and Natural Resources: Publication 3311, 1992.

Slaughter, D. C.; Barrett, D.; Boersig, M. Nondestructive determination of soluble solids in tomatoes using near-infrared spectroscopy. *J. Food Science* **1996**, *61*, 695−697.

Sugiyama, J. Visualization of sugar content in the flesh of melon by near-infrared imaging. *J. Agric. Food Chem*. **1999**, *47*, 2715−2718.

Venables, W. N.; Ripley B. D. *Modern Applied Statistics with S-PLUS*, 2nd ed.; Springer-Verlag: New York, 1997.

Williams, P. C.; Norris, K. H. *Near-infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists: St. Paul, MN, 1987.

Workman, J. J. Interpretive spectroscopy for near-infrared. *Appl. Spectrosc. Rev*. **1996**, *31*, 251−320.

Received for review March 29, 2000. Revised manuscript received August 28, 2000. Accepted August 28, 2000.

JF000408F