ELSEVIER

0308-8146(95)00118-2

Food Chemistry, Vol. 55, No. 3, pp. 305–311, 1996

0308-8146/96 $15.00+.00
Analytical, Nutritional and Clinical Methods Section

Non-invasive and non-destructive percutaneous analysis of farmed salmon flesh by near infra-red spectroscopy

Gerard Downey

Teagasc, The National Food Centre, Dunsinea, Castleknock, Dublin 15, Ireland

(Received 9 May 1995; accepted 12 June 1995)

Near infra-red (NIR) reflectance spectroscopy (700–1100 nm wavelength range) has been applied to the percutaneous measurement of oil and moisture in farmed salmon carcasses. Spectra were recorded through the skin and scales by means of a fibre-optic probe. Six selected sites were used on the dorsal and ventral surfaces of each fish side; 294 sample sites were utilized. Reference chemical values for oil and moisture were determined on these sites after excision. Calibrations were developed and evaluated separately for dorsal and ventral sites. The best dorsal calibration produced a standard error of prediction (SEP) for oil and moisture of 2.0 and 1.45%, respectively; corresponding figures for the best ventral calibration were 2.4 and 1.9%, respectively.

INTRODUCTION

Salmon has always been considered as a high-value gourmet fish. Traditionally, the market for this commodity has been supplied by wild fish but during this century the practice of fish farming has developed as a means of increasing sales and supplying the market at times of the year when wild fish are not normally available.

From the consumer's perspective, fresh salmon low in oil will be 'dry' and rather 'poor' in flavour; in contrast, fish with a high oil content is often characterized as possessing a 'rich' mouthfeel and having a 'fuller' flavour. The salmon smoker has better defined quality requirements; for optimum uptake of smoke, flesh oil content should ideally be between 8 and 12% (w/w). If oil contents drop below 8%, flavour uptake will be poor; in the case of fish with a higher than desirable lipid content, the finished product will be 'oily' and may lose oil after processing with a consequent deterioration in appearance in retail packs.

In farming situations, feed formulation and usage rate are the principal regulators of flesh composition. Feed is also the single largest cost input into a commercial salmon farming operation (Heen, 1992). Thus, the drive to optimize fish oil content arises from the need to meet processor and consumer requirements as well as from considerations of cost-control and, ultimately, profitability.

Chemical techniques currently available for the determination of fish flesh composition suffer from a number of disadvantages. Among these may be mentioned the fact that they require the destruction of the fish being analysed, are expensive, generate chemical residues which present disposal problems and are time-consuming.

This study describes an investigation into the utility of near infra-red spectroscopy for the routine analysis of fish flesh composition in a rapid and cost-effective manner; moisture and fat were the proximate constituents selected for investigation on the grounds of their pre-eminent importance in salmon processing and marketing. Given the irregular shape and dimensions of the fish, the most practical option for spectral collection was the use of a fibre-optic probe. A major interest of fish farmers is to monitor the fish flesh composition during growth and post-slaughter. In order to maximize the relevance of this work to such commercial enterprises, spectra were collected through the skin and scales.

Some previous work on the analysis of fish flesh by near infra-red (NIR) spectroscopy has been published (Gjerde & Martens, 1987; Mathias *et al.*, 1987; Valdes *et al.*, 1989; Rasco *et al.*, 1991; Lee *et al.*, 1992; Sollid & Solberg, 1992). In three of these reports (Gjerde & Martens, 1987; Mathias *et al.*, 1987; Valdes *et al.*, 1989) water was removed from the fish by lyophilization and a high correlation was reported between laboratory and spectral values for crude lipid and moisture contents on dried material. In another (Rasco *et al.*, 1991), NIR reflectance in the 900–1800 nm range was used on fresh

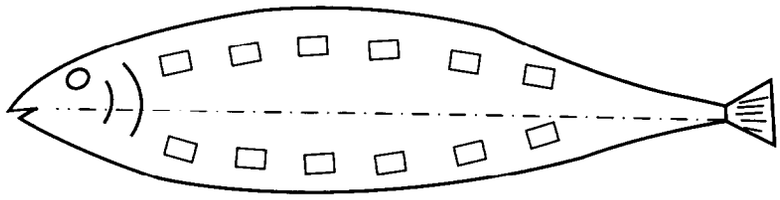

Fig. 1. Sampling site distribution on salmon side.

and frozen rainbow trout muscle tissue while Sollid and Solberg (1992) described initial work using near infra-red transmission spectroscopy. Lee *et al.* (1992) have reported work on the analysis of trout by NIR spectroscopy using a bifurcated fibre-optic probe; these authors also made their spectral measurements through skin and scales but used a different probe construction to that utilized in the present work and only estimated flesh oil content.

MATERIALS AND METHODS

Salmon selection

Farmed salmon were purchased in a retail outlet. In order to maximize information describing the natural variation in salmon composition and condition, individual fish were purchased at different times over a period of 1 year and were selected to include a range of maturity and quality. Salmon were obtained from several commercial farm sources although no information was recorded linking individual fish to specific farms. Fish were purchased during the morning and transported to The National Food Centre where they were maintained at 4°C until spectra were collected later the same day.

Experimental protocol

Several sites were selected and marked on each side of every salmon purchased; normally there were six along the dorsal surface of each side and six along the ventral surface (Fig. 1). Sites were approximately equally spaced along each surface and were slightly larger than the surface area of the fibre-optic probe head. Spectra were recorded at each site; when the entire set of spectra for each salmon had been acquired, the sites were excised and stored at −20°C prior to chemical analysis. This storage period was normally less than 14 days.

Reference analyses

Salmon were analyzed for content of both moisture and fat using semi-automated procedures which integrated microwave drying and methylene chloride extraction. Both analyses were performed using equipment manufactured by the CEM Corporation (P.O. Box 200, Matthews, NC 28106, USA); moisture was determined on the CEMAVAC-80 while oil estimations were made on the FES-80. Accuracy and precision data for this procedure were obtained using 25 cans of a Certified Reference Material (Statens Livsmedelsverk, P.O. Box 622, S-751 26 Uppsala, Sweden); precision figures for oil and moisture estimation were 0.14 and 0.22, respectively.

NIR spectra collection

Near infra-red spectra were recorded using a NIR-Systems 6500 scanning spectrophotometer (NIRSystems Inc., Maryland, USA); this equipment was fitted with a surface interactance fibre-optic probe and a transmittance detector module. Spectra were recorded over the wavelength range 400–1100 nm at 2 nm intervals; above 1100 nm, the fibre material became a significant absorber of near infra-red radiation. Operation of the spectrophotometer and the collection and manipulation of spectra was performed using a software package supplied by the instrument manufacturer—NIRS3 (ISI Inc., Port Matilda, USA).

Sample spectra were recorded at defined sites on fish by placing the surface interactance probe head on the site manually and holding it in position during scanning. The surface of the probe was wiped with a lens tissue after contact with the fish surface; excess moisture was removed from the surface of the fish by wiping with a tissue immediately before spectral acquisition. Each sample site was scanned six times with the mean of these six spectra being stored for further processing.

Data treatment

Prior to calibration development, all of the dorsal spectra were examined mathematically to detect the presence of any outlier samples, i.e. samples which are significantly different spectrally from the others. The procedure involved used the CENTER option in NIRS3. Spectra were converted to their first derivatives in order to minimize scatter effects and to emphasize small absorption peaks. Derivatized spectra were then subject to principal components analysis (PCA) and their spectral scores projected into PC space. Next the distance of each score from the average score was calculated as a Mahalanobis distance and a histogram of the frequency of these distances was plotted. Those samples with a Mahalanobis distance greater than 3.0 were, somewhat arbitrarily, highlighted as outlying samples. Such samples may unduly influence wavelength selection and regression coefficient size during calibration development, and

Table 1. Summary compositional data for total, dorsal and ventral sample sets

Samples	Constituent	Mean %	SD	Range %	*n*
Total	Moisture	68.0	3.7	57.9–74.7	294
	Oil	10.0	4.7	2.3–23.0	294
Dorsal	Moisture	69.7	3.9	62.1–74.7	147
	Oil	7.5	3.2	2.3–16.7	147
Ventral	Moisture	66.3	3.8	57.9–73.5	147
	Oil	12.4	4.6	3.3–23.0	147

the preference is to remove them unless some independent evidence exists which confirms them to be members, albeit extreme, of the spectral cluster.

Calibration development and evaluation sample sets were composed using the SELECT option in NIRS3 (Shenk & Westerhaus, 1991). This is an extension of the clustering algorithm used in CENTER and identifies those samples which describe the total variation in any given collection. Samples thus identified form a calibration development set with the remainder being used for evaluation purposes.

Calibrations were developed and evaluated using stepwise multiple linear regression (SMLR), partial least squares (PLS) and modified PLS procedures (Shenk & Westerhaus, 1991); in each case, the calibration and evaluation procedures were performed on raw (log $1/R$), first derivative and second derivative spectra. All data were scatter-corrected by a multivariate algorithm, Detrend (Barnes *et al.*, 1989), provided in the NIRS3 software.

Twelve calibrations were developed for each of the three mathematical procedures; in each case, four models were produced with the maximum number of terms allowed in each being restricted to five, six, seven and eight, respectively. Each was examined for the magnitude of the correlation coefficient and, more importantly, the size of the standard error of calibration. On the basis of these criteria, the optimum number of terms in each model was determined and these equations evaluated on the evaluation sample set.

RESULTS AND DISCUSSION

Reference chemical data on salmon samples

A total of 294 salmon sites were scanned, excised and analysed. Summary data for this total sample set are shown in Table 1. These values describe a wide variation in fish maturity as was desired at the outset. Histograms of the data in Table 1 are shown in Figs 2 and 3; it may be seen from these graphical representations that distributions of both parameters are quite significantly skewed. In the case of oil, only small numbers of samples exhibit values above 16%; for moisture, most samples lie above 63%.

Equal numbers of dorsal and ventral samples are represented in the above dataset. Initial examination of the collected spectra revealed two phenomena:

(i) some spectra exhibited significant noise levels below 850 nm; and
(ii) spectra fell into two types according to the slope of the spectral trace from low to high wavelengths.

Mathematical inspection of the spectra using CENTER revealed that two distinct populations did, in fact, exist; these corresponded to dorsal and ventral site spectra, respectively. The difference in the two types of spectra is clearly evident in Fig. 4 which shows the mean dorsal and ventral traces. Reasons for this difference must arise from the composition of tissues at these body locations; such compositional variations will affect

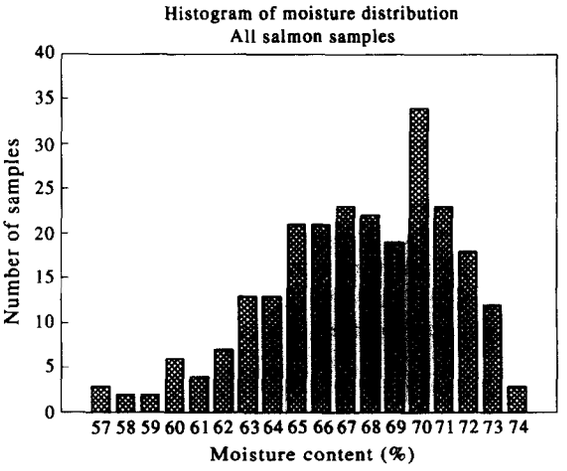

Fig. 2. Moisture distribution—total sample set.

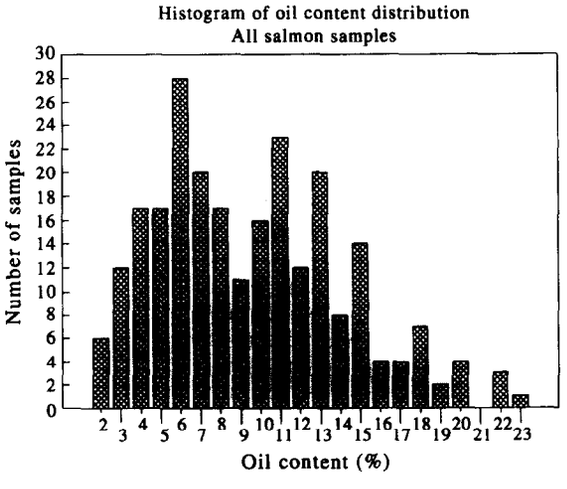

Fig. 3. Oil distribution—total sample set.

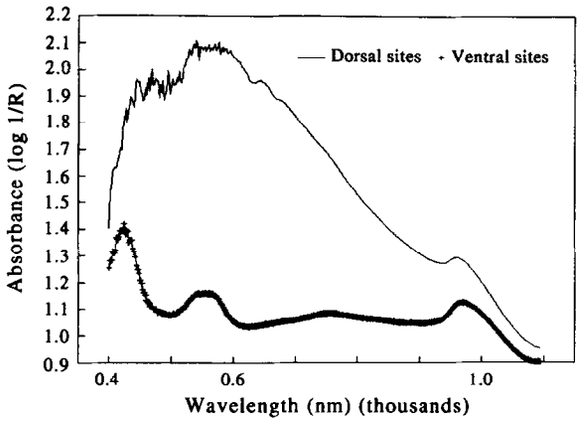

Fig. 4. Mean spectra—dorsal and ventral sites.

spectral shape through the mechanism of light absorption and scatter in an unpredictable way. Whatever the reason for this phenomenon, it precluded the development of a single set of calibrations on all salmon samples; it became necessary to divide the total sample collection into two containing only dorsal or ventral spectra, respectively. Calibration development and evaluation exercises were performed on each of these collections separately. Compositional data for both dorsal and ventral sites are also shown in Table 1.

As would be expected, the main difference in the chemical composition of ventral and dorsal sites relates to oil content. In the case of the ventral sites, the mean oil value of 12.4% is significantly higher than the corresponding value of 7.5% for the dorsal sites. Ventral sites also exhibited a greater range in oil content (3.3–23.0%) than was the case for dorsal material (2.3–16.7%). As a consequence, moisture contents for these sites differed in the opposite direction, i.e. ventral sites contained less moisture on average than the dorsal sites.

Calibration development—dorsal sites

Because of the problem with noise below 850 nm found in some dorsal site spectra, calibration development and evaluation procedures used the restricted wavelength range of 850–1100 nm. Examination of the dorsal spectra collection revealed the presence of two outlying samples—these were deleted. Composition of the dorsal site development and evaluation sets utilized is summarized in Table 2. Their prediction performance is summarized in Tables 3 and 4. Examination of the moisture prediction statistics revealed that all of the squared correlation coefficient values were between 0.63 and 0.72 (correlation coefficients between 0.79 and 0.85 respectively); values of this magnitude are acceptable for this type of sampling presentation and tissue matrix. Standard error of prediction (SEP) values ranged from 1.45 to 1.75. These tended to be lower with calibrations derived by PLS and MPLS; with these latter techniques, derivative spectra generally produced slightly lower prediction errors. Bias values (most around 0.2%) are not of any practical importance, even if they are of statistical significance, while the poorest slope values (i.e. those which exhibit greatest deviation from the theoretically perfect value of 1.0) were obtained with log $1/R$ data. Derivative spectra, whether treated by PLS or MPLS, gave rise to those values closest to 1.0.

Overall, the best calibration for the prediction of moisture in dorsal sites was that produced by partial least squares treatment of second derivative spectra (Table 3).

With regard to oil prediction, squared correlation coefficients varied in magnitude between 0.62 and 0.70, while the standard error of prediction (SEP) values ranged from 1.89 to 2.16; as was the case for moisture prediction, these values tended to be consistently lowest for MPLS calibration strategies and highest when

Table 2. Dorsal calibration development and evaluation set composition

	Constituent	Mean %	SD	Range %	*n*
Development	Moisture	69.9	2.6	63.6–74.7	80
	Oil	7.3	3.2	2.3–15.6	80
Evaluation	Moisture	69.5	2.8	62.1–73.9	53
	Oil	7.8	3.3	3.1–16.7	53

Table 3. Dorsal calibration performance data—moisture

Calibration	Signal	SEP	R^2	Bias	Slope
SMLR	log $1/R$	1.67	0.66	0.24	0.93
	First derivative	1.74	0.63	0.21	1.12
	Second derivative	1.75	0.63	0.09	0.85
PLS	log $1/R$	1.53	0.70	0.21	0.99
	First derivative	1.60	0.69	0.21	1.08
	Second derivative	**1.45**	**0.69**	**0.21**	**1.05**
MPLS	log $1/R$	1.59	0.71	0.21	0.95
	First derivative	1.54	0.71	0.21	1.09
	Second derivative	1.55	0.72	0.21	1.00

Table 4. Dorsal calibration performance data—oil

Calibration	Signal	SEP	R^2	Bias	Slope
SMLR	log $1/R$	2.06	0.64	−0.28	0.84
	First derivative	1.98	0.69	−0.73	0.93
	Second derivative	2.16	0.62	0.26	0.79
PLS	log $1/R$	1.89	0.69	0.31	0.95
	First derivative	**2.04**	**0.70**	**0.31**	**1.01**
	Second derivative	1.93	0.69	0.31	0.89
MPLS	log $1/R$	1.93	0.69	0.31	0.91
	First derivative	1.95	0.68	0.31	1.08
	Second derivative	1.96	0.69	0.31	0.93

SMLR was the procedure utilized. Both PLS and MPLS produced a constant bias value equal to 0.31; log $1/R$ and second derivative SMLR calibrations produced bias values which were slightly lower than this. The highest bias value (0.73) was obtained using SMLR on log $1/R$ spectra. With the exception of two results, slope values for the regression line fitted to reference and predicted oil values lay between 0.91 and 1.08.

In selecting the optimum calibration, the desire is for that which produces the lowest SEP consistent with evidence of long-term stability. High bias values and regression slopes which are significantly different from 1.0 are indications of possible instability. For these reasons, the calibration of choice for the prediction of oil in dorsal fish sites is PLS on first derivative spectral data (Table 4).

Calibration development—ventral sites

While the spectral region utilized for the dorsal site work was restricted to 850–1100 nm owing to noise problems, the range used for ventral site examination was wider at 700–1100 nm. The composition of calibration development and evaluation sample sets is summarized in Table 5. As was the case for the dorsal site sample sets, these two approximate well in the composition distribution of the two analytes. This is supporting evidence for the efficacy of the mathematical clustering approach used in their selection.

The predictive performance of the calibrations developed for oil and moisture prediction is summarized in Tables 6 and 7. In the case of moisture, squared correlation coefficient values ranged from 0.54 to 0.77; it was notable that the R^2 values obtained for the SMLR calibration techniques were the lowest. Standard error of prediction values varied quite widely from 1.90 to 2.79. The worst results were consistently obtained with SMLR treatments while the most consistent behaviour was shown by MPLS. The lowest value for this parameter (SEP = 1.90) was found with PLS treatment of the first derivative spectra. The lowest bias value observed was produced by MPLS of second derivative data, although the most consistently low figures emerged following PLS treatment (0.14–0.19). Bias values produced by SMLR were consistently the highest. The lowest bias value (0.14) was that obtained following PLS treatment of the first derivative spectra. Slope values closest to 1.0 resulted from PLS and MPLS treatments.

Table 5. Ventral calibration development and evaluation set composition

Set	Constituent	Mean %	SD	Range %	*n*
Development	Moisture	66.1	3.6	57.9–73.4	65
	Oil	12.7	4.5	4.2–23.0	65
Evaluation	Moisture	66.5	3.9	57.9–73.5	68
	Oil	12.2	4.7	3.3–22.8	68

Table 6. Ventral calibration performance data—moisture

Calibration	Signal	SEP	R^2	Bias	Slope
SMLR	log $1/R$	2.79	0.54	0.79	0.95
	First derivative	2.78	0.54	0.54	0.82
	Second derivative	2.51	0.62	0.55	0.89
PLS	log $1/R$	2.25	0.69	0.19	1.01
	First derivative	**1.90**	**0.77**	**0.14**	**1.04**
	Second derivative	2.11	0.71	0.15	1.01
MPLS	log $1/R$	2.08	0.72	0.19	1.03
	First derivative	2.12	0.71	0.28	1.02
	Second derivative	2.14	0.70	0.09	1.01

Table 7. Ventral calibration performance data—oil

Calibration	Signal	SEP	R^2	Bias	Slope
SMLR	log $1/R$	3.59	0.48	−0.97	0.89
	First derivative	3.35	0.51	−0.49	0.94
	Second derivative	3.74	0.39	−0.49	1.03
PLS	log $1/R$	2.58	0.71	−0.17	1.02
	First derivative	**2.41**	**0.74**	**−0.16**	**1.04**
	Second derivative	2.65	0.69	−0.15	1.02
MPLS	log $1/R$	2.65	0.69	−0.21	1.03
	First derivative	2.67	0.69	−0.32	1.03
	Second derivative	2.68	0.68	−0.08	1.02

The optimum calibration for the prediction of moisture in ventral sites on salmon was that produced following PLS treatment of the first derivative spectra.

In the case of oil prediction (Table 7), R^2 values ranged from a low of 0.39 to a high of 0.74. MPLS calibrations were most consistent with regard to this parameter while SMLR treatments produced poor R^2 data. The maximum value for this parameter was produced by PLS treatment of the first derivative spectra ($R^2=0.74$, $r=0.86$). The observed range in standard error of prediction values was from a minimum of 2.41 to a maximum of 3.74. Again, SMLR methods consistently produced the worst SEP figures (3.35–3.74) while MPLS produced SEP values which varied least from one data pre-treatment to the next. The lowest SEP value was produced by PLS on first derivative spectra. Bias values ranged from 0.08 to 0.97. In this instance, the lowest value (0.08) resulted from MPLS treatment of second derivative spectra although PLS treatments resulted in the most consistent results (0.15–0.17). Slope values varied from 0.89 to 1.04. With the exception of SMLR treatment of log $1/R$ or first derivative data, all slope values were good and showed little variation (1.02–1.04).

The most successful calibration for the prediction of oil contents in ventral sites of salmon is that produced by PLS treatment of first derivative spectra.

A number of observations may be made about the optimum prediction equations described above:

(a) the best prediction of either moisture or oil was obtained at dorsal sites, despite the act that the range in both analytes on this surface is lower than for ventral sites;
(b) correlation coefficients for moisture and oil were lower for dorsal sites than for ventral but the magnitude is probably not statistically significant; this behaviour stems directly from the fact that the range in analyte concentration in dorsal sites is lower than in ventral;
(c) bias values obtained by linear regression were higher for dorsal calibrations—this need not be a cause of undue concern but is a caveat as regards future use of these calibrations;
(d) without exception, the best mathematical procedure for the prediction of moisture and oil in salmon flesh was partial least squares (PLS); in three cases out of the above four, the use of first derivative was optimal.

CONCLUSIONS

Of the published work related to the topic of this report (Gjerde & Martens, 1987; Mathias *et al.*, 1987; Valdes *et al.*, 1989; Rasco *et al.*, 1991; Lee *et al.*, 1992; Sollid & Solberg, 1992), only Lee *et al.* (1992) used a similar approach to that reported herein.

These latter workers confined their investigation to the measurement of oil in trout muscle at a number of sites on the dorsal surface of the fish; in their case, SMLR and PLS (both using second derivative spectra) produced comparable standard errors of prediction of 2.4%. Results reported above for the dorsal surface reveal greater accuracy (SEP = 2.04%) by using PLS on first derivative spectra. The correlation coefficient between reference and predicted oil values obtained (0.9) was somewhat higher than the result shown in this work (0.84) although the ranges in oil content found in both fish samples was similar. This difference may therefore be a function of the different fish species investigated.

Prediction accuracies reported by the other publications are all significantly better than the above, i.e. in the range 1.0–1.5% for oil; this shows the dramatic improvement in accuracy which is possible by the removal of water from the sample matrix.

One factor not investigated in the present work but referred to in Lee *et al.* (1992) is the variation in fat composition at different sites on a fish carcass. These workers sought to optimize site selection on the basis that overall accuracy may be correspondingly optimized and concluded that, of three sites studied, the most suitable (i.e. that which produced equations of the highest predictive accuracy) was mid-way between the posterior insertion of the dorsal fin and the anterior insertion of the adipose fin. In the present work, effort was concentrated on accumulating a large sample set with a view to permitting spectral collection and accurate prediction virtually anywhere on the dorsal or ventral surfaces. The results obtained reflect the utility and validity of this approach but raise the possibility of

more accurate calibrations if these were to be developed for specific sites on each surface.

An accuracy level of that obtained in the present work should have significant practical possibilities in the salmon-rearing industry. While this accuracy will be unsuitable for fine control of fish fat levels, it will be able to detect very high and very low levels. This may permit better consistency of product through the removal of extreme carcasses. It should also be valuable for charting body composition during growth. A strength of the technique utilized is in the speed of the analytical method and the fact that it is non-destructive. Difficulties mitigating against better accuracy levels relate particularly to the high scattering effects of fish scales and the possible low light levels returning to the instrument detector, thus reducing the signal-to-noise ratio significantly.

Another complicating factor is the heterogeneous nature of fish flesh composition. Even in a sample of the surface area utilized in this work, it may be that considerable variation in oil and moisture distribution occurs. From the viewpoint of the chemical analysts this is of no consequence given that a cube of flesh is excised and thoroughly homogenized prior to analysis. Spectrophotometrically, however, such inhomogeneities will reduce the achievable prediction error because the chemical value for a constituent will not correspond exactly to that which the NIR instrument sees. A more refined tissue excision protocol, especially with regard to depth, may improve the apparent NIR performance.

No consideration of prediction accuracy is complete without reference to the precision of the reference method used to produce chemical data. Given that a figure for the CEMAVAC instrument used in this work is in the region of 1%, the near infra-red prediction (which can only hope to approximate but never exceed the reference technique) appears favourable for an application in which an approximation of oil content is required.

ACKNOWLEDGEMENTS

G. Downey wishes to acknowledge financial support from the European Commission Operational Programme for Rural Development.

REFERENCES

Barnes, R. J., Dhanoa, M. S. & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.*, **43**, 772–7.

Gjerde, B. & Martens, H. (1987). Predicting carcass composition of rainbow trout by near infrared reflectance spectroscopy. *J. Anim. Breed. Genet.*, **104**, 137–48.

Heen, K. (1992). Comparative analysis of the cost structure and profitability in the salmon aquaculture industry. In *Salmon Aquaculture*. Blackwell Scientific, Oxford, pp. 220–38.

Lee, M. H., Cavinato, A. G., Mayes, D. M. & Rasco, B. A. (1992) Noninvasive short-wavelength near infrared spectroscopic method to estimate the crude lipid content in the muscle of intact rainbow trout. *J. Agric. Food Chem.*, **40**, 2176–81.

Mathias, J. A., Williams, P. C. & Sobering, D. C. (1987). The determination of lipid and protein in freshwater fish using near infrared reflectance spectroscopy. *Aquaculture*, **61**, 303–11.

Rasco, B. A., Miller, C. E. & King, T. L. (1991). Utilisation of NIR spectroscopy to estimate the proximate composition of trout muscle with minimal sample pre-treatment. *J. Agric. Food Chem.*, **39**, 67–72.

Sollid, H. & Solberg, C. (1992). Salmon fat content estimation by near infrared transmission spectroscopy. *J. Food Sci.*, **57**, 792–3.

Shenk, J. S. & Westerhaus, M. O. (1991). Population definition, sample selection and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci.*, **31**, 469–74.

Valdes, E. V., Atkinson, J. L., Hilton, J. W. & Leeson, S. (1989). Near infrared reflectance analysis of fat, protein and gross energy of chicken and rainbow trout carcasses. *Can. J. Anim. Sci.*, **69**, 1087–90.