

Calibration transfer for predicting lake-water pH from near infrared spectra of lake sediments

Paul Geladi,^{a*} Hans Bärning,^a Eigil Dåbakk,^a Johan Trygg,^a Henrik Antti,^a Svante Wold^a and Bo Karlberg^b

^a*Department of Organic Chemistry, Umeå University, S-90187 Umeå, Sweden.*

^b*Department of Analytical Chemistry, Stockholm University, S-10691 Stockholm, Sweden.*

With near infrared spectra of lake sediment samples, it is possible to predict lake-water pH using a partial least squares (PLS) calibration model. The calibration models have large model and prediction residuals and it is important to understand the residuals. All the sediment samples were measured on five different instruments. This allows the use of each of the five instruments as master with the other four as field instruments. When using more than one near infrared instrument, prediction becomes unreliable unless calibration transfer is used. A number of techniques including Savitzky–Golay Transform (SGT), finite impulse response (FIR), piecewise direct standardisation (PDS), orthogonal scatter correction (OSC) and wavelet transform (WT) were compared. The quality of the predictions was expressed as root mean square error of prediction (*RMSEP*), but the calibration transfer methods were also compared on practical usefulness. In these data, the OSC filtering worked best and gave adequate calibration transfer results.

Keywords: calibration transfer, environmental samples, orthogonal signal correction, wavelet transformation, piecewise direct standardisation, finite impulse response, Savitzky–Golay smoothing.

Introduction

For environmental monitoring of aquatic systems, lake water chemistry variables are measured repeatedly over time. Large monitoring programmes are expensive and demands from society for cost-reductions have limited these programmes. Therefore, cost-effective approaches are desirable. Promising attempts in this respect have been studies on the use of near infrared (NIR) spectrometry of lake sediment samples for prediction of important lake-water chemistry variables.^{1–5} Sediment is continually formed in all lakes with composition depending on the lake biogeochemistry as well as catchment

chemistry (catchment is the influx of water and solids from precipitation). Lake sediments, therefore, mirror the average biogeochemistry of lakes and allow integrated measures of lake-water chemistry variables to be estimated on the basis of properties of the sediment. Hence, lake sediments are very useful for studies of long-term water quality changes.

The basic assumption of this work is that the properties of the lake sediments are dependent on the chemical properties of the lake water, such as pH. It is further assumed that the properties of the organic material are reflected in the NIR spectra and that the spectra are related to lake-water chemistry. Depending on sediment accumulation rate, the samples

in this study typically cover between two and five years of sediment deposition.

NIR spectrometry is widely used in industry for quality control and process monitoring. The uses of NIR in biological and environmental contexts are rare but slowly increasing.^{4,6} Earlier, calibration (= training) models have been made for prediction of total organic carbon (TOC), total phosphorous and pH^{1,2} in lake water from NIR spectra of sediments from the same lakes. As is the case with many other environmental applications, the calibration models often produce substantial unexplained residual variances for the prediction of the responses. There have been attempts to study the sources of these unexplained variances in NIR-based lake sediment models^{2,7} and to split the residual variances into contributions from sampling error, noise in the response variable and lack of fit.

The use of NIR on solid samples is advantageous in time and cost compared with wet-chemical methods. NIR is also non-destructive, so the solid samples are not consumed. In this paper, dried and ground lake sediment is measured by NIR reflectance spectrometry and the NIR spectra are used to predict the pH of the lake water, obtained by wet-chemical techniques, in a PLS-based multivariate calibration model. The model is given as:

$$\text{pH} = f(\mathbf{x}) \quad (1)$$

with \mathbf{x} the NIR spectrum vector. For a PLS, or other linear regression model, this first Equation is linearised as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{f} \quad (2)$$

with \mathbf{y} , a vector of pH values, \mathbf{X} , a matrix of NIR spectra, \mathbf{b} , the regression coefficients and \mathbf{f} , the residual vector. Both measurement noise in \mathbf{X} and in \mathbf{y} influence the values of \mathbf{b} and \mathbf{f} in Equation 2. This will be discussed below.

Equation 2 works reasonably well, but large prediction errors (the residual \mathbf{f}) still occur, which is not unusual for environmental measurement.^{1,2,4,5} This is just a manifestation of natural variation. As an example of this, Figure 1 shows NIR-predicted pH values plotted against measured values for a group of representative test samples. All points should fall on a straight line and deviations from this line should be

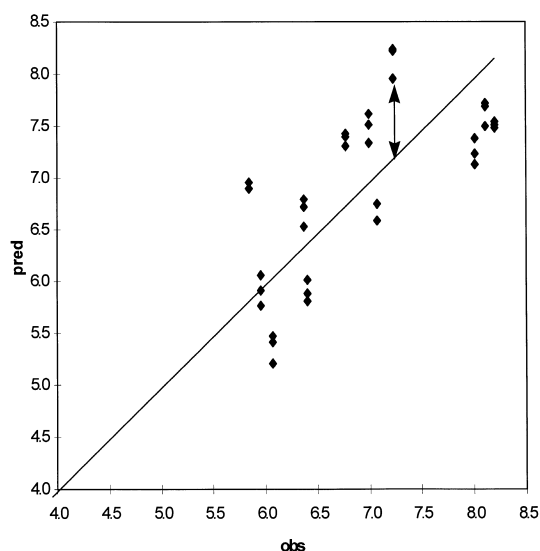


Figure 1. A typical predicted/observed plot for pH of lake-water samples. The horizontal axis is observed pH. The vertical axis is pH predicted from the NIR spectra. The plot shows deviations from an ideal straight line. Given the nature of the samples, some residual (see the double arrow) is expected, but reducing the residual as much as possible is desirable. The samples are triplicates for 12 lakes in the test set.

explained and brought under control as well as possible. A better knowledge of variability and its components allows better planning for obtaining more precise and accurate models. The lowest error obtained in this way for a large data set is a standard deviation of around 0.35 pH units.⁷

Because of the large number of samples to be measured and the long-term nature of a monitoring programme, it is clear that calibration transfer is an important issue. A calibration transfer is a transformation of the data measured on one instrument (the **field instrument**, also termed the slave instrument or remote instrument) to serve the calibration model of another instrument (the **master instrument**, also main instrument) better. This is shown schematically in Figure 2.

In order to study calibration transfer, sediment samples from a number of lakes were measured on five NIR instruments and calibration models were made against the lake-water pH values measured by wet-chemical methods. The sample set consists of 24 lakes \times 3 replicates per lake \times 700 wavelengths,

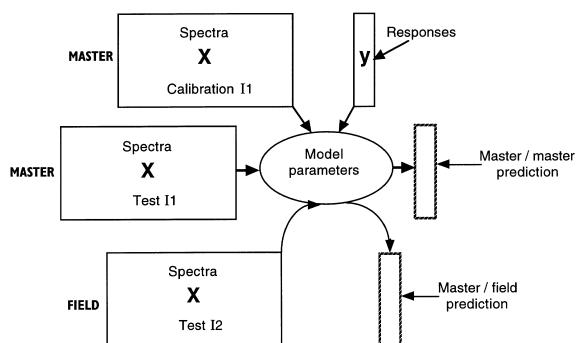


Figure 2. The calibration set with spectra and responses (pH) for instrument 1 (the master instrument) gives model parameters. When the model parameters are used on the test set spectra measured on the same instrument, a master / master prediction is obtained. When the test set measured on instrument 2 (the field instrument) is used with the model parameters of instrument 1, a master / field prediction is obtained which is not always good unless individual instrument parameters are removed from the spectra.

using every 2 nm from 1100 to 2498 nm. The samples came from a larger set of 71 lakes, sampled in triplicate, but, unfortunately, most samples were used up by wet-chemical analysis, so there was not enough to fill the sampling cup. Figure 3 gives a graphical representation of the data sets, called UMU, Röbbäck, Skogis, Foss1 and Foss2. With five instruments, one instrument can be used as the master and the other four as field instruments. The master instrument role can also be rotated between the five instruments, giving a large number of possible

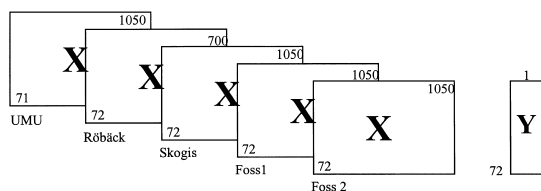


Figure 3. A representation of the data sets used. There are 1050 wavelength variables from 400 to 2498 nm for four of the instruments and 700 wavelengths from 1100 to 2498 nm for one older instrument with a different monochromator. One data set has only 71 objects since one spectrum was lost. Lake-water pH forms the response data Y .

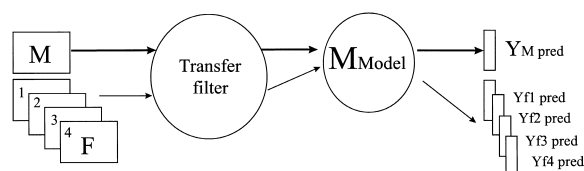


Figure 4. With all samples measured on five different instruments, it is possible to rotate which instrument is the master and which are the four field instruments. This gives a number of possible combinations of calibration transfer.

model and prediction sets, as is shown in Figure 4. It is possible to create a 5×5 matrix of prediction results with master/master predictions on the diagonal and master/field predictions outside the diagonal (Figure 5). A separate study, with only four instruments that have the extended wavelength range 400–2498 nm, is not shown in this paper.

Sometimes calibration transfer is only used to make spectra look identical when they come from different instruments. With the spectra used for this paper, visual inspection was not considered and pre-

	Master 1	Master 2	Master 3	Master 4	Master 5
Field 1	●				
Field 2		●			
Field 3			●		
Field 4				●	
Field 5					●

Figure 5. Rotation of the master instrument role, gives a 5×5 matrix of results. The master training / master test data are on the main diagonal and the master training / field test results are off the diagonal. A good calibration transfer method gives comparable results in every cell of the matrix. The filled circles indicate situations where no calibration transfer is necessary since master and field instrument are the same.

diction error in the response variable, expressed as root mean square error of prediction (*RMSEP*) was used. It is not the spectra themselves, but the quality of the predicted pH values, that is the goal. A test of this quality is the difference between predicted and measured response for the test set. This difference is squared and averaged over all test set samples and the square root is taken to make it look like a standard deviation, expressed in the same units as the response.

$$RMSEP = [1/J \sum_{j=1}^J (y_{pj} - y_{mj})^2]^{1/2} \quad (3)$$

with J the number of test samples, y_{pj} the predicted pH for test sample j and y_{mj} the measured pH for test sample j .

Calibration transfer has been used in the food, pharmaceutical and polymer industries,^{8–16} but rarely for environmental samples. A difference is that the number of samples and the amount of sample available are usually limited in environmental cases. The major test of the quality of a calibration model is the use of a test set. In order to do this, the 24 samples were split into two groups of 12 each.

Calibration transfer

During the first half of the 1990s, a number of publications were devoted to the problems of transfer of calibration models.^{8–16} The reason for the need of transfer is that the classical method of calibrating the instrument in the morning and doing all measurements during one day is not always practical. This is especially true for NIR, where large calibration sets are often needed and where large numbers of samples are measured over extended periods of time. The environmental samples used in this paper are no exception.

Calibration transfer techniques can be divided into hard and soft. The hard techniques are based on control and understanding of the hardware used. Each instrument is considered as an individual with specific, well-defined properties. The soft techniques are based on using measured spectra. The individual properties of the instruments are assumed to be unknown, but expressed by the spectra. This means that each spectrum reflects not only the chemical composition of the samples, but also the individ-

ual properties of the instruments on which it was measured. Calibration transfer aims at eliminating spectral differences due to individual properties of the instruments by separating those properties from the sample properties. This can be given conceptually by the equation:

$$\mathbf{X} = \mathbf{X}_{\text{chem}} + \mathbf{X}_{\text{instr}} + \mathbf{E} \quad (4)$$

where \mathbf{X} is the matrix of NIR spectra, \mathbf{X}_{chem} is the part of it related to the composition of the sample, $\mathbf{X}_{\text{instr}}$ is the part related to instrument-specific systematic variation and \mathbf{E} is random noise. Equation 4 can have many different forms: additive, multiplicative and so on. The calibration transfer functions used in this paper are techniques for removing $\mathbf{X}_{\text{instr}}$ from Equation 4 in order to make Equation 2 work better. The techniques are all based on soft modeling. The real mechanisms that produce the instrumental differences are assumed unknown, but the effects are estimated by trial and error from the available spectra and removed.

Stable and well-determined calibration samples are best, but the calibration standards should, ideally, also be as close as possible in composition and properties as the samples to be measured. For environmental samples, it is difficult to make such artificial standards and a subset of the samples to be measured is used as calibration standards. This subset is the calibration set for each master instrument and also the set that is used for the calibration transfer.

A number of techniques for calibration transfer have been reported in the literature. In some cases, existing spectrum smoothing and cleaning methods, such as multiplicative scatter correction (MSC),^{17,18} Savitzky–Golay transformation (SGT)¹⁹ and derivation,²⁰ are sufficient. For the present work, methods like finite impulse response (FIR) and piecewise direct standardisation (PDS), as found in the literature, were supplemented with the newer orthogonal signal / scatter correction (OSC) and with wavelet transformation (WT) of the data. Many of the above mentioned methods have a number of parameters to choose from and this requires a steep learning curve and subjective decisions. The goal of the study is to find how damaging a change of instrument is and how good the calibration transfer methods are.

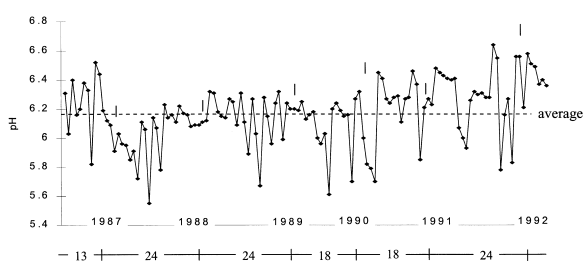


Figure 6. Variation over time of lake pH for one example: lake Fiolen.

Experimental

Sediment sampling and pH measurement

Lake sediment samples were taken in triplicate from 73 lakes in southern Sweden in 1993.^{2,21–23} The samples were taken from the deepest point in each lake, using a HON–Kajak gravity corer and frozen in the field immediately after coring. The upper 1 cm of the sediments was used for NIR analysis. Dependent on accumulation speed, this corresponds to two to five years. The samples were freeze-dried and ground in a mortar before the NIR measurement.

As the dependent variable, average measures of lake pH were used. For some of the 73 lakes, pH values had been recorded at least once a year over the period that the upper 1 cm in the sediment corresponded to. These values were then averaged to give a more precise estimate of the mean pH in the lake. An example of how the pH may vary over time is given in Figure 6. The pH data were obtained from national, as well as regional, water monitoring programmes.

Spectrometer and sampling cups

The NIR spectrometers were all from Foss NIRSystems. Four were NIRSystems 6500 (400–2498 nm) and one was an NIRSystems 5000 (1100–2498 nm). Readings were taken every 2 nm, giving 1050 wavelength variables for the 6500 and 700 for the 5000. For comparison of the five instruments, the wavelength range was limited to 1100–2498 nm. A comparison of four instruments in the wavelength range 700–2498 nm was also carried

out and gave similar results, but this is not reported here. The sediments were used in rotating sampling cups for reflectance measurement. The reflectance data were converted to apparent absorbance and stored as MATLAB files. The samples were triplicates for 24 lakes, giving 72 spectra for each instrument, except for UMU, where one replicate spectrum was lost, giving 71 spectra.

The NIRSystems spectrometers used had a number of hard internal calibration procedures that were repeated before every measurement, such as wavelength calibration and measurement of the internal “total reflectance” standard. However, these internal hard calibrations are not sufficient and, because of different construction and age of the instruments, soft calibration transfer, as in Equation 4, was still needed.

The samples were measured with calibration and test samples in randomised order. This avoided block effects because the calibration samples were subject to all drift and noise in the instruments during the whole measuring session.

Software

Software used was Unscrambler 6.0,²⁴ SIMCA P 7.0²⁵ and MATLAB 5.0.²⁶ Also the PLS_Toolbox²⁷ from Eigenvector Research was used. 3D loading plots were made with the program 3DROT.²⁸

Results and discussion

Selection of calibration and test set

With environmental data, it is hard to make designs that span the sampling space (the subset of multivariate space that encompasses the whole population) properly. The sampled part of the sampling space is often irregular, with an uneven spread of the samples, making it difficult to select calibration and test samples properly. It was decided not to separate the triplicates; all three samples from a lake were put either in the test set or in the calibration set.

The available data were split into 12 (\times 3) calibration and 12 (\times 3) test samples. There were two demands on calibration and test sets: they should span each other's spectral space, but they should also have the same range in the responses (pH). The spectral

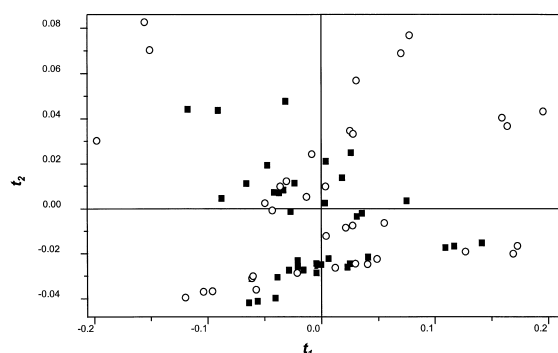


Figure 7. The t_1 – t_2 scoreplot used for the selection of calibration and test sets. Open round symbol = calibration, square symbol = test. t_1 = first principal component, t_2 = second principal component.

space selection was based on a principal component analysis after mean-centring of the data set for one instrument, leading to a score plot of the first and second component. The two selected components explained more than 90% of the sum of squares of the data. Both calibration and test sets should span these two components, but the test set should not be an extrapolation of the calibration set. The score plot is shown in Figure 7. The split in calibration and test sets was not ideal but acceptable. Another con-

sideration is that the responses of both calibration and test sets should have a reasonable range. It was assumed that the present choice of calibration and test set was one of the best possible ones, considering both criteria. The pH mean, minimum and maximum were: for the calibration set: 6.72, 4.54 and 8.45; for the test set: 6.92, 5.85 and 8.20.

Raw data results

For comparison it is important to do PLS predictions without calibration transfer. Table 1 gives the *RMSEP* results of prediction on the test set. The calibration model was made after mean-centring and with the number of components determined by cross-validation (Unscrambler). The table clearly shows the best values along the diagonal, where calibration transfer is not needed. The off-diagonal results were sometimes reasonable, but very often they were unacceptable. This indicated the need for calibration transfer. Also, the small differences in the number of cross-validated components pointed to individual differences resulting from the instruments.

The matrix in Table 1 can be considered as a distance matrix. The results of the singular value decomposition of this matrix are shown in Figures 8 and 9. Figure 8 shows the scree plot, indicating that three important components existed. They explained 71.6, 16.6 and 11.5% of the total sum of squares.

Table 1. Prediction results as *RMSEP* expressed in pH units. The cross-validated number of components for the model is given as XVAL. Raw data are used. Values of *RMSEP* that are more than twice the “master” *RMSEP* are given in bold. * means master-master prediction (no transfer needed).

Instrument	Master				
Field	UMU	Röbäck	Skogis	Foss1	Foss2
XVAL	9	10	8	10	7
UMU	0.67*	5.03	1.78	0.90	2.82
Röbäck	5.44	0.69*	1.84	5.07	1.02
Skogis	1.26	3.49	0.73*	1.12	5.79
Foss1	0.71	4.23	0.95	0.58	3.77
Foss2	0.93	7.37	5.61	2.43	0.79*

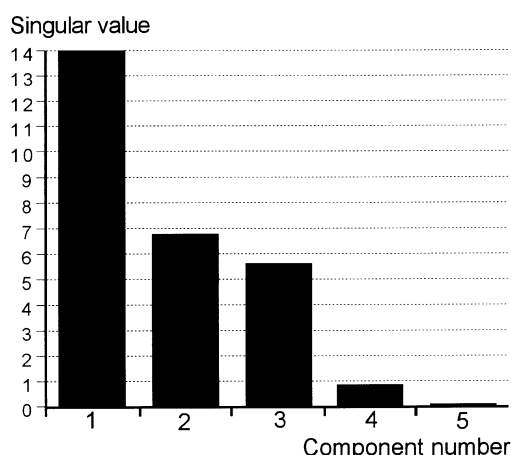


Figure 8. The scree plot for the data matrix in Table 1. The singular values (vertical axis) are plotted against component number (horizontal axis).

Figure 9 shows the 3D loading plot, with the master instruments indicated. Except for Foss1 and UMU, all the master instruments were a long way from each other and, therefore, produced different spectra for the same samples. It is clear that the instruments had individual contributions to the spectra that should be determined and removed.

When judging calibration transfer methods, the prediction quality is very important, but when differences in prediction quality are small, ease of use should be taken into consideration. Some calibration

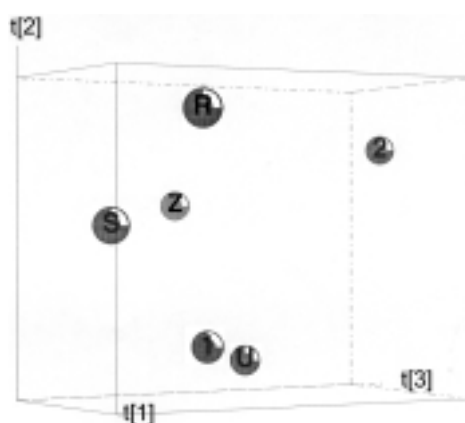


Figure 9. The 3D loading plot for the singular value decomposition of the data in Table 1. The master instruments are UMU, R, Skogis, Foss1 and Foss2. Z is the origin.

transfer methods have a large number of parameters that can be varied, making their use in practical situations awkward.

Savitzky–Golay smoothing

Sometimes simple transformations of the data remove individual instrument differences. This is the case with offsets and slopes that can be removed by multiplicative scatter correction (MSC) or by derivation. A classical smoothing / derivation technique is the Savitzky–Golay transformation (SGT).¹⁹ This

Table 2. Prediction results as *RMSEP* expressed in pH units. The cross-validated number of components for the model is given as XVAL. Transfer function: was Savitzky–Golay (SGT). Values of *RMSEP* that are more than twice the “master” *RMSEP* are given in bold. * means master–master prediction (no transfer needed).

Instrument	Master				
	UMU	Röbäck	Skogis	Foss1	Foss2
Field	7	10	10	10	6
XVAL	7	10	10	10	6
UMU	0.50*	1.79	0.53	0.59	1.19
Röbäck	1.01	0.66*	1.11	0.78	1.12
Skogis	0.48	2.00	0.65*	0.75	1.56
Foss1	0.49	1.64	0.61	0.54*	1.27
Foss2	1.63	2.85	2.12	0.91	0.52*



Figure 10. The 3D loading plot for the principal component analysis of the data in Table 2. The master instruments are UMU, Röbbäck, Skogis, Foss1 and Foss2. Z is the origin.

technique was applied to all the spectra with one derivative, a window width of 11 and polynomial fit in the order of 3. The results of this are shown in Table 2. The derivation / smoothing showed improvements over using the raw spectra and, in some cases, the results were very good, but this was not so for every master / field pair. It was interesting that derivation / smoothing gives better results for the master / master predictions. Slightly worse results were obtained by MSC correction. They are not shown.

Singular value decomposition on the data matrix of Table 2 gave three significant components, but with much smaller singular values than those for the raw data. They explained 90.7, 6.6 and 2.7% of the total sum of squares. The loading plot is shown in Figure 10. This plot shows some distance from the other master instruments for Röbbäck and Foss2, but one should realise that the singular values are much smaller than the ones for Figure 9. Plots like these are easier to understand than the data tables.

The fact that smoothing / derivation gives better master / master predictions indicates that the raw data were noisy and contained regions that were irrelevant for the prediction.

The Savitzky-Golay smoothing / derivation method requires the selection of quite a number of parameters: derivative order, derivative window size

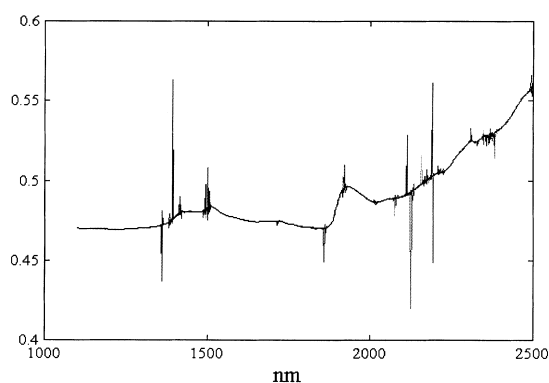


Figure 11. Artifacts in FIR-filtered spectra. Reference spectrum: mean of calibration spectra of master instrument UMU. Spectra to be corrected: test data from field instrument Röbbäck. Window size: 401. The corrected spectra contained spikes caused by unstable linear regression solutions.

and order of the interpolation function. It is quite impractical to test all possibilities and a subjective decision, that requires a learning curve, has to be made.

Finite impulse response

In this work, the FIR¹¹ technique was used as provided with the PLS_Toolbox.²⁷ FIR is based on a simple linear regression in a moving window. The reference spectrum is an average of some or all reliable master spectra, containing all features of the master instrument. The spectra in the matching window for the field instrument are used to calculate correction coefficients for the wavelength to be corrected, usually the wavelength in the middle of the window. This is very much like using a moving MSC filter. A good, recent, description of FIR is available in the Reference 29.

Here, the FIR^{11,29} technique gave artifacts in the transferred spectra. This was due to the flat shape of the spectra and the fact that the calculation of the coefficients tended to give singular results. An example of these artifacts is shown in Figure 11. The figure shows that the FIR transformation gave rather good results, except for the regions where spikes appear. It is not possible to foresee the appearance of these spikes for the present data. This means that for every window size tested, all resulting spectra

Table 3. Prediction results as *RMSEP* expressed in pH units. Transfer function used was FIR transform. The process window size is 5. * means master–master prediction (no transfer needed).

Instrument	Master				
Field	UMU	Röbäck	Skogis	Foss1	Foss2
XVAL	7	7	7	6	7
UMU	0.56*	0.87	0.57	0.68	0.62
Röbäck	0.82	0.53*	1.05	0.50	0.57
Skogis	0.91	0.46	0.55*	0.78	0.57
Foss1	1.58	1.31	1.11	0.47*	0.93
Foss2	0.52	0.73	0.65	0.62	0.52*

should be visually checked for spikes. For other (less noisy) data, FIR calibration transfer may work very well.

Knowing that the artifacts appear at window sizes, ranging from very small to really large, is the reason for being very restrictive about using the FIR method on flat spectra. However, a complete list of results using the FIR method is given in Table 3 and the *RMSEP* values turned out reasonably. As will be seen below, PDS and OSC gave lower *RMSEP* results.

Because of the difficulty in selecting window sizes and the often unexpected occurrence of spikes, using FIR requires operator interaction and is time-consuming. Furthermore, it did not work as well as

some other calibration transfer techniques. There is the possibility of using smoothing to remove spikes, but a quick test showed that this only removed the spikes and did not improve the *RMSEP* values.

Piecewise direct standardisation

A classical technique for calibration transfer is piecewise direct standardisation (PDS).¹³ PDS is based on regression between a window in the master spectra and the corresponding window in the field spectra. Principal component regression is used to calculate the regression coefficients. The version of PDS supplied by Eigenvector Research in the PLS Toolbox²⁷ was used. The PDS function was studied

Table 4. Prediction results as *RMSEP* expressed in pH units. Transfer function used was PDS transform. The process window size is 1 except where indicated between parentheses. * means master–master prediction (no transfer needed).

Instrument	Master				
Field	UMU	Röbäck	Skogis	Foss1	Foss2
UMU	0.67*	0.58 (9)	0.68	0.58 (5)	0.83
Röbäck	0.67	0.69*	0.73	0.55	0.73
Skogis	0.76	0.56	0.73*	0.53	1.10
Foss1	0.69	0.56	0.72	0.58*	0.79
Foss2	0.82	0.64	0.81	0.55	0.79*

Table 5. Prediction results as *RMSEP* expressed in pH units. Transfer function used was OSC transformation with 2 OSC components and 1 PLS component. * means master–master prediction (no transfer needed).

Instrument	Master				
Field	UMU	Röbäck	Skogis	Foss1	Foss2
UMU	0.79*	0.78	0.78	0.75	0.81
Röbäck	0.61	0.61*	0.60	0.59	0.62
Skogis	0.76	0.75	0.74*	0.72	0.77
Foss1	0.64	0.63	0.62	0.61*	0.65
Foss2	0.70	0.69	0.69	0.68	0.70*

over a window range from 1 to 13. The minimum *RMSEP* for each case is disclosed in Table 4. It is clear that PDS gave a strong improvement over using raw data, but also an improvement over the FIR results in Table 3. On the diagonal are the results when PDS was used on a test set from the same instrument. Note that these results are the same as when using raw data. This means that the PDS transform did not change the transferred spectra, which is expected.

The fact that the best results, in most cases, were found using window size of one, implies that for most master / field combinations only offset correction is needed between the instruments at each wavelength. This property seems to conform pretty well with the features of flat spectra mentioned earlier about the FIR technique. However, this also suggests that the instruments are very stable in terms of wavelength shift.

Orthogonal signal correction

Orthogonal signal correction^{30,31}(OSC) is a recently developed method for spectral filtering, done in order to simplify and enhance multivariate calibration models. Unlike other filtering methods such as MSC^{17,18} and standard normal variate (SNV),³² it uses the response variable \mathbf{y} to orthogonalise the removed part of the predictor variable block \mathbf{X} . This is done by calculating a number of principal components of \mathbf{X} and orthogonalising them to the responses \mathbf{y} . The new score vector is now orthogonal to the responses and the corresponding loading (called OSC component) is used to perform filtration of \mathbf{X} . If nec-

essary, the procedure is repeated. This takes away parts of the spectra that are irrelevant to the responses and drastically reduces the number of PLS components to be used in the final calibration model. This is advantageous in NIR, since several PLS components often only describe instrument specific phenomena. It was shown for other NIR data sets that two OSC components suffice to remove portions of NIR spectra not useful for prediction. In all models used, two principal components were used for OSC correction and only one PLS component was needed to get a good PLS model on the corrected data.

The OSC filtering was done for the calibration set of each instrument separately and also on the corresponding test set. The OSC filtered data sets were then used for master / master and master / field predictions. The results are given in Table 5. They were slightly better than PDS results. A special feature is that the results were good for all entries of Table 5. Principal component analysis of the data of Table 5 only gave one significant component explaining almost all the sum of squares. This is a sign that calibration transfer has succeeded.

Wavelet transform

A new calibration technique that utilises the wavelet transform is also presented here. For detailed description of the Wavelet transform, reference is given to more dedicated literature.^{33,34} The wavelet transform is an expansion of an arbitrary function (signal) into localised building blocks, labeled by their scale and position. This simply means

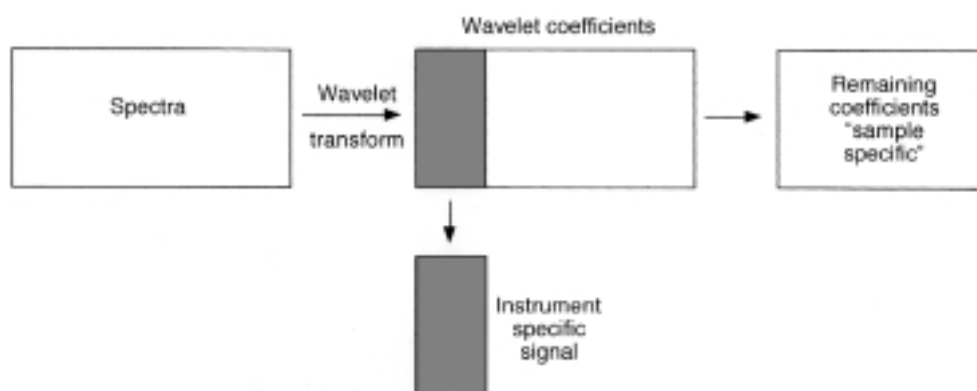


Figure 12. The Wavelet transform replaces spectra by Wavelet coefficients. These are split in low and high frequency coefficients. For low frequency coefficients are assumed to be instrument-specific and the remaining ones (sample-specific) are used for calibration and test sets.

that a signal that is wavelet-transformed is chopped up in bandpasses of frequencies ranging from zero Hz up to the Nyquist frequency. For a spectrum the Nyquist frequency would be half the number of wavelengths.

In recent years, the use of the wavelet transform has focused on compression and noise reduction of separate signals. Not as much attention has been given to the analysis of multiple signals, where the resulting wavelet coefficients are further used to build multivariate calibration models.³³ In this work, the wavelet transform is used as a calibration transfer technique. The reason is simple. In calibration transfer, information in signals that is instrument-dependent is removed, and the information

that is sample-dependent is kept (see Figure 12). In this work, a sound assumption is made about the signals. It is assumed that the parts of the signal that are due to individual instrument properties are in the low-frequency regions. This assumption is not too unreasonable, considering base line and slope drifts in NIR signals that often are machine-dependent.

In wavelet analysis, one must choose a mother wavelet, which is the basic function used to create all other basic functions used in the analysis. The orthogonal Daubechies-4 wavelet was employed.³⁵ It is probably the most used mother wavelet and it has suitable mathematical properties. A method description of how the wavelet transform is implemented in calibration transfer can be found in Figure 12. After

Table 6. Prediction results as *RMSEP* expressed in pH units. Transfer function used was wavelet transform (WT) with removal of low frequency wavelet coefficients. * means master–master prediction (no transfer needed).

Instrument	Master				
Field	UMU	Röbäck	Skogis	Foss1	Foss2
UMU	0.43*	0.97	0.41	0.59	0.64
Röbäck	0.52	0.49*	0.43	0.61	0.60
Skogis	0.77	0.95	0.41*	0.41	0.76
Foss1	1.02	1.08	0.51	0.41*	0.74
Foss2	0.44	0.41	0.38	0.42	0.57*

performing the wavelet transform on each NIR spectrum, the resulting wavelet coefficients produce a matrix of the same size as the original matrix. As mentioned, the assumption about the signals is that the lower frequencies of the signals can be influenced by instrument properties and, therefore, should be removed. This is done by simply excluding the first part of the wavelet coefficient matrix, corresponding to lower frequencies. The calibration model is built from the reduced wavelet coefficients matrix and the test set is also predicted from the reduced wavelet coefficients.

The results of the wavelet-based calibration transfer are shown in Table 6. They were very promising, considering the novel approach used in this paper. The wavelet transform calibration method does not require any standard samples to calibrate against, nor does it even need to know the different instrument's different spectra. All that was assumed was that the instrument-dependent variations which influence the NIR spectra are mainly in the lower frequency regions, a sound assumption. Also, the method is superior in speed, compared to the other techniques mentioned in this report. The selection of the cut-off point between low-frequency and high frequency wavelet coefficients is not very critical. A slightly wrong choice does not make the calibration models unstable.

Conclusion

Calibration and test sets of 12×3 samples each (three replicates) were measured on five different in-

struments in the wavelength region 1100–2500 nm. It was shown that the instrumental properties of the instruments are reflected in the spectra in such a way that calibration transfer is necessary. All calibration transfer methods should remove individual differences between the spectrometers and keep the chemical information intact. There are a number of principles available for this and some methods that are commonly found in the literature were tested. These included smoothing, FIR, PDS, OSC and wavelet transform. Of the methods used, OSC and PDS worked very well, while wavelets were slightly worse and FIR substantially worse. The lowest obtained *RMSEP* was around 0.5 pH units. Considering the small sizes of test and calibration sets, this is a reasonable result. Other studies with more samples showed a *RMSEP* of 0.35, which is the lowest possible value in this type of study. Table 7 shows an overview of the results in Tables 1 to 6. The total sum of squares indicates the extent to which the predictions for each transfer method were wrong. The distribution of the sum of squares over the eigenvalues indicates how well the instruments matched each other. The best situation is when only one eigenvalue is different from zero and the total sum of squares is small. One should be aware that the conclusions presented are only valid for environmental data of the type studied in this paper and that pharmaceutical, food, forestry and other NIR spectra may give different results.

An additional problem for these studies is that the reference values for the response variable are not exactly known. They can be estimated in a number of ways from available pH data for each lake. The prob-

Table 7. Total sum of squares (SS) and eigenvalues for the Tables 1 to 6 (rounded).

	SS	1	2	3	4	5
Raw data	273	196	45	32	0.7	0
SGT	39	36	2.6	1	0	0
FIR	15.6	14.8	0.4	2	2	0
PDS	12.4	12.3	0.1	0	0	0
OSC	12.0	12.0	0	0	0	0
WT	10.0	9.7	0.2	0.1	0	0

lem is that some lakes have many pH values and some lakes only have a few and that some lakes have a large variation in pH over the years studied and some lakes have a smaller variation. These are some of the inherent problems in environmental investigations.

The data pretreatment and calibration transfer methods can be used alone or in combination. Most pretreatment methods also have a number of choices in parameters. This creates an infinite number of ways of pretreating spectral data for calibration transfer purposes. Only a fraction of these can be explored. With the data studied in this paper, it seems that smoothing / derivation improved master / master prediction slightly and that calibration transfer methods improved master / field prediction substantially. Combining calibration transfer methods gave only small improvements and increased the risk of unexpected instabilities because of the huge number of parameters involved. It was found that OSC was easiest to use with just the number of OSC components to be removed as a choice of parameters. PDS required a test of a number of windows. The wavelet transform required subjective removal of coefficients, but seemed to be stable and easy to use otherwise. FIR and the Savitzky–Golay smoothing did not work well and required a number of parameters to be tested and were, therefore, not very practical.

The calibration transfer study was carried out after completed sampling and on stored samples. Any future sampling campaign would have to take into account the findings in this study. One of them is that more calibration and test samples should be available to span the calibration space more completely and to give more accurate *RMSEP* values. Another one is that more accurate pH measurements over time are needed. The present study used pH values and sediment samples from different agencies that were not supposed, from the origin, to be used together.

Acknowledgements

The owners of the NIR spectrometers are acknowledged for letting us access their instruments and the Department of Environmental Health of Umeå University is acknowledged for providing the

samples. The authors also thank Barry Wise and Neal Gallagher of Eigenvector Research for useful comments.

References

1. T. Korsman, M. Nilsson, J. Öhman and I. Renberg, *Env. Sci. Technol.* **26**, 2122 (1992).
2. M. Nilsson, E. Dåbakk, T. Korsman and I. Renberg, *Env. Sci. Technol.* **30**, 2586 (1996).
3. D. Malley, P. Williams, M. Stainton and B. Hauser, *Can. J. Fish Aquat. Sci.* **50**, 1779 (1993).
4. D. Malley and M. Nilsson, Environmental applications in near infrared spectroscopy: seeing the environment in a different light, *Spectrosc. Eur.* **6**, 8 (1995).
5. E. Dåbakk, M. Nilsson, P. Geladi, S. Wold and I. Renberg, *Water Res.* **34**, in press (1999).
6. W. Foley, A. Mc Ilwee, I. Lawler, L. Argones, A. Woolnough and N. Berding, *Oecologia* **116**, 293 (1998).
7. E. Dåbakk, P. Geladi, M. Nilsson, S. Wold and I. Renberg, *J. Near Infrared Spectrosc.* **7**, 241 (1999).
8. O. De Noord, *Chemometrics and Intelligent Laboratory Systems* **25**, 85 (1994).
9. M. Forina, G. Drava, C. Armanino, R. Boggia, S. Lanteri, R. Leardi, P. Corti, P. Conti, R. Giangiacomo, C. Galliena, R. Bigoni, I. Quartari, C. Serra, D. Ferri, O. Leoni and L. Lazzeri, *Chemometrics and Intelligent Laboratory Systems* **27**, 189 (1995).
10. E. Bouveresse and D. Massart, *Anal. Chem.* **67**, 1381 (1995).
11. T. Blank, S. Sum and S. Brown, *Anal. Chem.* **68**, 2987 (1996).
12. I. Adhihetty, J. McGuire, B. Wangmaneerat, T. Niemczyk and D. Haaland, *Anal. Chem.* **63**, 2329 (1991).
13. J. Shenk and M. Westerhaus, *Crop Sci.* **31**, 1694 (1991).
14. Y. Wang and B. Kowalski, *Appl. Spectrosc.* **46**, 764 (1992).
15. Y. Wang, D. Veltkamp and B. Kowalski, *Anal. Chem.* **63**, 2750 (1991).

16. Z. Wang, T. Dean and B. Kowalski, *Anal. Chem.* **67**, 2379 (1995).
17. P. Geladi, D. MacDougall and H. Martens, *Appl. Spectrosc.* **39**, 491 (1985).
18. T. Isaksson and B. Kowalski, *Appl. Spectrosc.* **47**, 702 (1993).
19. A. Savitzky and M. Golay, *Anal. Chem.* **36**, 1627 (1964).
20. T. O'Haver and G. Green, *Anal. Chem.* **48**, 312 (1976).
21. G. Pettersson, I. Renberg, P. Geladi, A. Lindberg and F. Lindgren, *Journal of Paleolimnology* **9**, 195 (1993).
22. I. Renberg, *Journal of Paleolimnology* **6**, 167 (1991).
23. I. Renberg, *Boreas* **10**, 255 (1981).
24. *The Unscrambler 6.0 User's Guide*. Camo AS, Trondheim, Norway (1996).
25. *User's Guide to SIMCA 7.0*. Umetri AB, Umeå, Sweden (1998).
26. *Using MATLAB Ver. 5*. The Mathworks, Natick MA, USA (1996).
27. B. Wise and N. Gallagher, *PLS_Toolbox 2.0 for use with MATLAB*. Eigenvector Research, Lake Manson, WA, USA (1998).
28. *3DROT Ver. 1.5*. Prosoft, Umeå, Sweden (1994).
29. S. Sum and S. Brown, *Appl. Spectrosc.* **52**, 869 (1998).
30. S. Wold, H. Antti, F. Lindgren and J. Öhman, *Chemometrics and Intelligent Laboratory Systems* **44**, 175 (1998).
31. J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg and S. Wold, *Chemometrics and Intelligent Laboratory Systems* **44**, 229 (1998).
32. R. Barnes, M. Shanoa and S. Lister, *Appl. Spectrosc.* **43**, 772 (1989).
33. J. Trygg and S. Wold, *Chemometrics and Intelligent Laboratory Systems* **42**, 209 (1998).
34. G. Strang and T. Nguyen, *Wavelets and filter banks*. Wellesley-Cambridge Press, Wellesley, MA, USA (1997).
35. I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, USA (1992).

Received: 14 February 1999

Revised: 11 May 1999

Accepted: 20 July 1999

Web Publication: 11 October 1999