

## Near-Infrared Reflectance Spectroscopy of Sediments: A Potential Method To Infer the Past pH of Lakes

Tom Korsman,<sup>\*,†</sup> Mats Nilsson,<sup>‡</sup> Jerker Öhman,<sup>§</sup> and Ingemar Renberg<sup>†</sup>

Department of Ecological Botany, University of Umeå, S-901 87 Umeå, Sweden, Department of Forest Site Research, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden, and Center for Peat Research, Box 5097, S-900 05 Umeå, Sweden

■ We show that there is a pH-related fraction of lake sediments that can be recorded by near-infrared (near-IR) reflectance spectroscopy. Relationships between near-IR spectra of surface sediment samples and measured lake-water pH values, and between near-IR spectra of sediment cores and historical pH values inferred by diatom analysis, are modeled by partial least squares regression. The standard errors of prediction of these models are comparable to those obtained by modeling of diatom and lake-water pH data. We believe that near-IR reflectance spectroscopy has the potential to become a fast and simple method for inferring lake-water pH from lake sediment cores and thus a complement to diatom analysis, which is the best but time-consuming technique available to study lake acidification history.

### Introduction

Surface water acidification is a major environmental problem in large areas of northern Europe and North America (1, 2). Studies of lake sediments have contributed significantly to the understanding of acidification processes, where diatom analysis of sediment cores has been used to study lake acidification history (3, 4). Diatoms are microscopic algae abundant in all aquatic environments (5). Recent diatom pH reconstruction techniques give predictions with error estimates of 0.25–0.5 pH unit (6, 7). There are, however, some problems involved in using diatom analysis for pH reconstruction as a routine method, since diatom identification requires considerable expertise and counting is very time-consuming. It is, therefore, desirable to develop simpler and more rapid methods for reconstructing past lake-water pH on a routine basis. These faster methods would be particularly important in applied limnology and lake management. For example, to decide whether or not lime should be applied to a lake, it is necessary to know whether and to what extent the lake has acidified in recent times, or whether, alternatively, it is naturally acidic.

We here begin to explore the potential of near-infrared (near-IR) reflectance spectroscopy as a method to study the pH history of lakes. We test the hypothesis that there is a pH-related fraction of lake sediment that can be detected with near-IR spectroscopy and modeled by multivariate calibration methods. We use two approaches to assess relationships between near-IR and pH data: modeling of (1) near-IR spectra from surface sediment samples and measured lake-water pH values (the modern calibration data set) and (2) down-core near-IR spectra from sediment cores and corresponding historical pH values derived from diatom analysis (the fossil calibration data set). A tentative quality comparison between prediction of pH from surface sediments using near-IR analysis and

diatom analysis is also made. Finally, a first attempt to reconstruct the pH history of two lakes, using sediment cores and the calibration model derived from the modern calibration data set, is presented and discussed.

### Materials and Methods

**Near-Infrared Reflectance Spectroscopy.** Near-IR spectroscopy has well-established industrial applications (8), e.g., for analyzing the moisture, protein, fat, and carbohydrate content in food products (9). Other applications include determination of the chlorophyll content of plants (10) and the detection of diseases and defects in plants (11–13). The near-IR spectra reflect the molecular composition of the samples analyzed. The absorption bands observed in near-IR spectroscopy derive from fundamental vibrational modes, which give rise to overtones and combinations observable in the near-infrared spectrum (700–2500 nm). Overtones of, for example, O–H, N–H, and C–H stretching vibrations occur in this spectral region. Also, hydrogen bonding to electron-donating atoms such as oxygen, nitrogen, and sulfur results in vibrations with large transition moments (9). Furthermore, the overtone regime has provided evidence for the presence of simultaneous vibrational transitions in hydrogen-bonded adducts (14). Simultaneous transitions may even have greater intensity than that for an overtone.

The main problem with near-IR spectroscopy is the complex nature of near-infrared spectra. The dependence of the reflectance on the scattering properties of the sample add to this complexity (9). Most peaks of interest are overlapped by interfering peaks, which means that multivariate calibration methods are very convenient to gain quantitative information from the spectra. Once a calibration model has been developed, the analyses can, however, be made by nonscientific personnel, and there are additional advantages of the technique. Sample preparation comprises only drying and grinding, and the analysis of sediment samples can be carried out at a rate of one sample per 3–5 min. Furthermore, apart from the possible effects of grinding, the technique is nondestructive.

**Modern Calibration Data Set.** This data set consists of near-IR data from surface sediments and measured lake-water pH from 21 forest lakes in southern Sweden, with a pH range from 4.4 to 7.0. The lakes, which are part of a monitoring program conducted by the Swedish Environmental Protection Agency, have been sampled for water chemistry analyses two to four times each year. The pH values used in this study are the mean values of the hydrogen ion concentrations for the period 1984–1987. Surface sediment samples of the top 0.5 cm were taken in 1987 with a HON-Kajak gravity corer (15) from the deepest point in each lake.

The sediment samples were dried for 24 h at 50 °C and ground in a mortar. Approximately 2 mL of the powdered sediment was used for the near-IR analysis on a Technicon InfraAnalyzer 400 DR (16). The absorbance was measured

<sup>†</sup> University of Umeå.

<sup>‡</sup> Swedish University of Agricultural Sciences.

<sup>§</sup> Center for Peat Research.

as  $\log 1/\text{reflectance}$  for 19 wavelengths between 1445 and 2348 nm. All statistical analyses were made on non-scatter-corrected data.

Partial least squares (PLS) regression (17–19) was used for calibration and prediction of the pH values from the near-IR spectra, using the computer program SIRIUS (Version 2.3, April 1991; Pattern Recognition Systems A/S, 5015 Bergen, Norway). With PLS the absorbance values for all wavelengths are used simultaneously to model the dependent variable (pH). The optimal number of PLS components was determined by minimizing the mean-squared error between the measured and predicted pH. One sample at a time was removed from the data set and the remaining samples were used in the calibration model, from which the pH value for the removed sample was predicted. This cross-validation procedure (jack-knifing) was repeated until the pH for all samples was predicted (20). The error estimates presented in the study are standard errors of prediction (SEP) defined as

$$\text{SEP} = \left( \sum_{i=1}^n (y_i - \hat{y}_i)^2 / n \right)^{1/2} \quad (1)$$

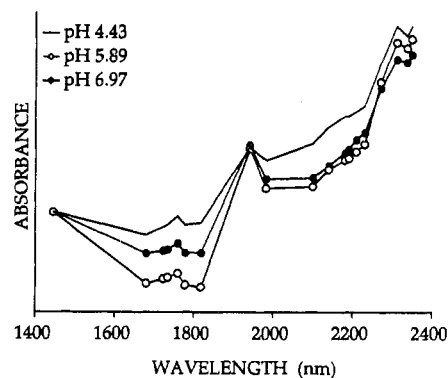
where  $y_i$  is the observed pH value for sample  $i$  and  $\hat{y}_i$  is the predicted pH value for sample  $i$ . All SEP values presented have been calculated by jack-knifing.

**Fossil Calibration Data Set.** It can be hypothesized that if near-IR spectra do contain pH-related information, it should be possible to model down-core near-IR changes with a substitute measure of pH, i.e., diatom inferred pH (DpH). The sediment cores for modeling relationships between near-IR spectra and diatom inferred pH were taken with a freeze corer (21) from Lilla Öresjön (57°33' N, 12°20' E) and Örvattnet (59°44' N, 12°45' E). Both lakes are acidified forest lakes in southwest Sweden (22–25). Lilla Öresjön has a present pH of 4.6, and Örvattnet a pH of 4.8. The core from Lilla Öresjön was cut into 40 contiguous subsamples (0.5 cm thick between 0 and 15 cm and 1 cm thick between 15 and 25 cm below the sediment surface). The core from Örvattnet was cut into 30 contiguous subsamples (0.5 cm down to 10 cm and 1 cm between 10 and 20 cm). Samples for near-IR analysis were prepared and analyzed as for the surface sediments. Diatom slides were made using a preparing technique with no decanting and centrifuging (26), and 500 valves per subsample were counted. Diatom inferred pH was calculated by weighted averaging (WA) (7), using the surface sediment diatom calibration set produced by the SWAP project (27). The DpH range for the Lilla Öresjön core is between 4.7 and 6.3, and between 4.5 and 6.2 for Örvattnet.

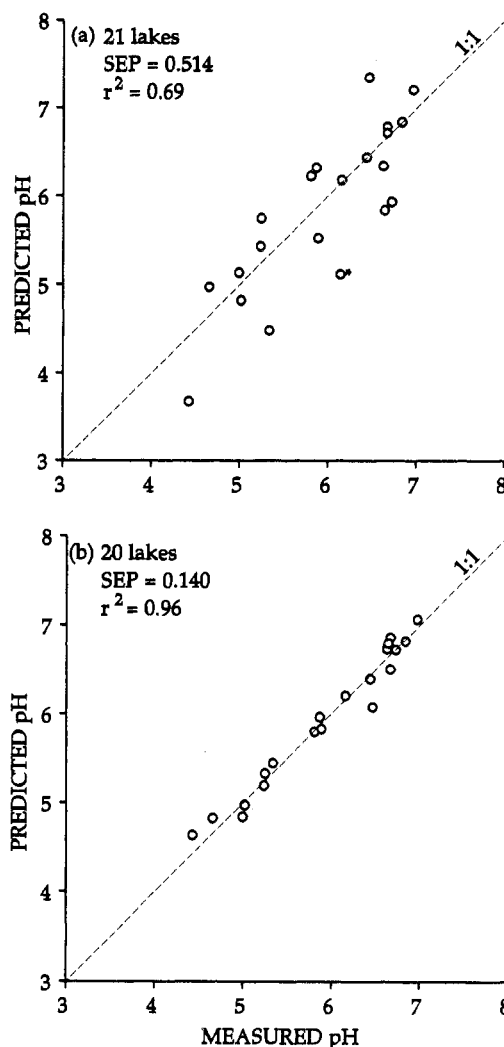
**Comparison of Near-IR and Diatom Models.** Fifteen of the lakes in the modern calibration data set were also included in the SWAP diatom calibration set. This subset of lakes has been used for comparison between near-IR inferred pH (NIRpH) and diatom inferred pH in surface sediments.

## Results

To illustrate the character of near-IR data, spectra for surface sediment samples from three lakes are shown in Figure 1. Although the absolute absorbance values vary considerably, the overall pattern of the spectra is rather similar with two major peaks, at ~1950 and ~2300 nm. These three spectra show that there is no simple relationship between the near-IR spectra and the lake-water pH, because of the complex nature of near-IR spectra, the different sediment types, and the scattering properties of the sediment samples. Scatter correction is often used in connection with near-IR analysis (28), but is not used in



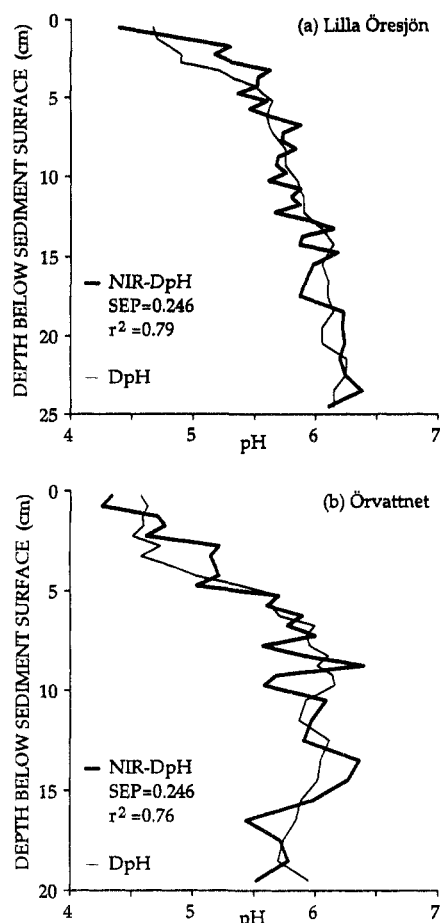
**Figure 1.** Near-infrared absorbance spectra for surface sediment samples from three lakes included in the modern calibration data set. The curves presented are for lakes with the lowest, median, and highest water pH values. The spectra have been normalized with respect to the absorbance value for 1445 nm.



**Figure 2.** Predicted lake-water pH from near-IR spectroscopy of surface sediments plotted against measured lake-water pH, using data from (a) 21 and (b) 20 lakes (\*, excluded lake). The predictions were calculated by jack-knifing.

this study since PLS can overcome the scatter effects and prediction using scatter-corrected data did not improve the results.

The relationship between near-IR predicted pH from surface sediment samples and measured lake-water pH is shown in Figure 2. The prediction of lake-water pH from the near-IR spectra gave a standard error of prediction (SEP) of 0.514 (20% of the pH range) and an  $r^2$  value of

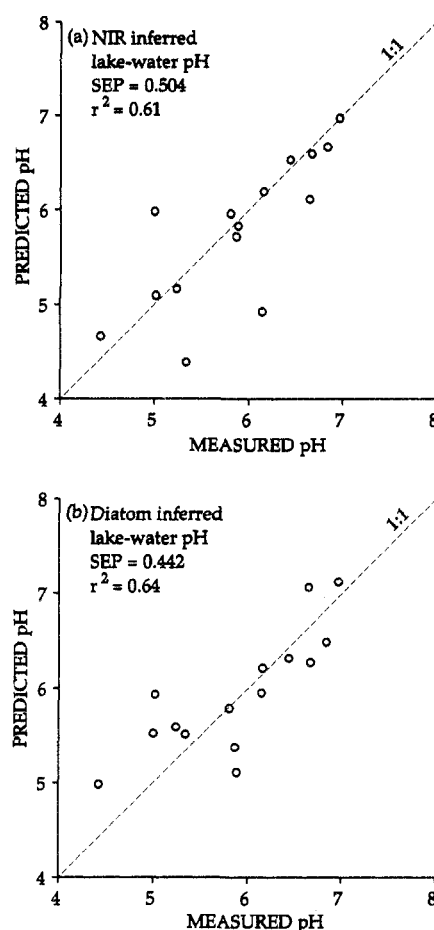


**Figure 3.** Predicted historical lake-water pH from the sediment cores from (a) Lilla Öresjön and (b) Örvattnet. The thin lines represent diatom inferred pH (DpH), and the thick lines near-IR predicted DpH (NIR-DpH). The NIR-DpH curves were calculated by jack-knifing, with data from both lakes included in the calibration model.

0.69. When data from one of the lakes (the one with the highest residual between measured and predicted pH values) were excluded, a SEP of 0.140 (5% of the pH range), and an  $r^2$  value of 0.96, were obtained. Regression models using spectra weighted to unit variance gave a somewhat poorer fit between predicted and measured pH, with a SEP of 0.522 when all 21 lakes were used in the model.

The second approach used to assess relationships between near-IR spectra of sediments and lake pH was prediction of historical pH values using near-IR spectra of sediment cores from Lilla Öresjön and Örvattnet. The results demonstrate good agreement between diatom inferred pH (DpH), which is a measure of historical pH values, and near-IR predicted DpH (NIR-DpH) for both cores (Figure 3). When both cores were used in the regression model, the SEP was 0.246 (14% of the DpH range). The  $r^2$  value for Lilla Öresjön was 0.79, and for Örvattnet 0.76. If only data from the Lilla Öresjön core were used to predict the DpH for that core, a SEP of 0.16 (10% of the DpH range) was obtained ( $r^2 = 0.86$ ). The corresponding value for Örvattnet was 0.247 (15% of the DpH range,  $r^2 = 0.52$ ). The poorest fit to DpH was obtained for the topmost part in both cores, but further down-core the fit is good, particularly for Lilla Öresjön. Moreover, the slight trend of increasing DpH between 20 and 10 cm below sediment surface in Örvattnet could be inferred from the NIR-DpH curve.

The results of the comparison between near-IR and diatom calibration models for the 15-lake subset show that



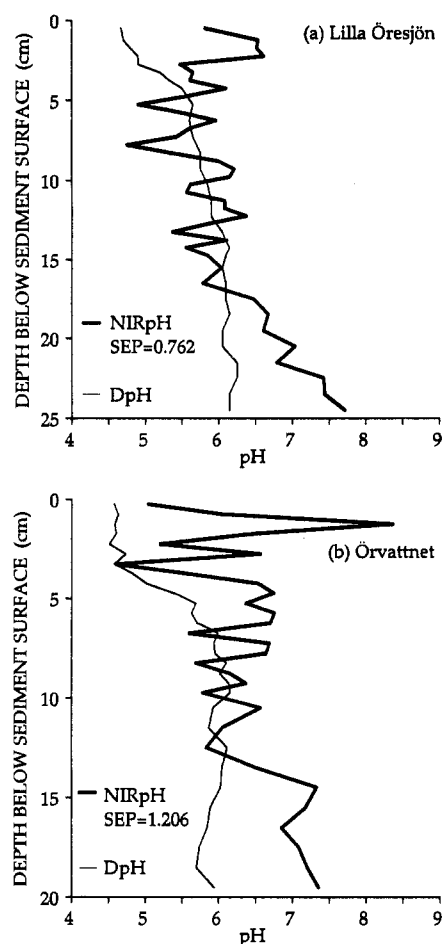
**Figure 4.** Predicted lake-water pH from surface sediment samples from 15 lakes: (a) partial least squares (PLS) predictions using near-IR spectra and (b) weighted averaging predictions with tolerance downweighting using diatom data. The predictions were calculated by jack-knifing.

these models produce rather similar predictions of lake-water pH. A SEP of 0.504 was obtained when pH was predicted by PLS regression using near-IR spectra ( $r^2 = 0.61$ ) (Figure 4a). Weighted averaging regression with tolerance downweighting [Wa(tol)] (7), using diatom data, gave a SEP of 0.442 ( $r^2 = 0.64$ ) (Figure 4b), and without tolerance downweighting 0.530 ( $r^2 = 0.55$ ). For PLS regression of lake-water pH using diatom data, the SEP was 0.467 ( $r^2 = 0.59$ ).

Encouraged by the good predictions of lake-water pH from near-IR spectra of the surface sediments, we used the surface sediment calibration model to predict past pH using the sediment cores from Lilla Öresjön and Örvattnet. Considerably poorer predictions were achieved; the SEP for Lilla Öresjön was 0.762, and for Örvattnet 1.206 (Figure 5) (the DpH values were used as observed pH values for the calculation of SEP). Moreover, all near-IR inferred pH values have large residual variance, suggesting that all core samples are outliers in prediction [i.e., the ratio between the residual variance in the spectral data and the residual variance for the individual sample is larger than the value in the  $F$  distribution for a probability level of 5% (19)].

#### Discussion

The good correlations obtained between near-IR predicted and measured pH for the modern calibration data set (Figure 2), and between near-IR and diatom predicted pH for the two sediment cores (Figure 3), show that the sediment contains information about lake-water pH that



**Figure 5.** Predicted historical lake-water pH from the sediment cores from (a) Lilla Öresjön and (b) Örvattnet. The thin lines represent diatom inferred pH (DpH), and the thick lines near-IR predicted pH (NIRpH). The NIRpH curves were obtained by the surface sediment calibration model.

can be recorded with near-IR and modeled with PLS regression. These results support our suggestion that near-IR analysis might become a useful method to infer past pH conditions from lake sediment cores. The initial comparison between near-IR and diatom predictions (Figure 4), using data from the 15 lakes that were included in both the modern near-IR and diatom data sets, indicates that near-IR reconstruction could become as accurate as diatom pH reconstruction, once the method is developed.

The attempt to use the reconstruction model from the modern data set to infer pH from the sediment cores from the two lakes was less successful (Figure 5). There are two possible explanations: (i) the modern calibration data set is too small (21 lakes), and (ii) degradation processes change the composition of the sediment during the sediment consolidation process.

Clearly the development of one standard prediction model from a modern data set would require more than 21 lakes. To achieve predictions with good accuracy, PLS regression, like all regression models, can only be used for interpolations. It is, therefore, necessary that the full range of variation in the predicted samples is covered in the calibration set. A more or less even distribution of the calibration samples is also important. The limitations of our modern calibration data set are clearly illustrated by the fact that when only one lake which deviated from the rest of the sample population was excluded, the SEP value improved from 0.51 to 0.14.

The near-IR spectrum reflects the molecular composition of the lake sediment and can derive from both organic

and mineral matter. The organic content of these sediments is between 40 and 60% (the main part of the rest is diatom valves). It is therefore likely that the organic matter is most important for the near-IR spectra in this study. There is considerable biological activity in the recently deposited lake sediment. The organic material at the sediment surface is exposed, for example, to bacterial activity and is not "stabilized" before further burial. It is, therefore, possible that diagenetic processes during burial affect the ability to make accurate reconstructions of historical pH trends using a prediction model derived from a surface sediment data set. However, this is not necessarily the case since the sediment fractions that contain the pH information might not be changed to a large extent during aging. The close agreement between NIR-DpH and DpH (Figure 3) might indicate that the failure to accurately reconstruct past pH in the cores (Figure 5) using the modern calibration data set model was caused by the restricted modern calibration data set rather than by diagenetic processes.

If further investigations show that diagenetic processes result in information loss or distortion, one solution might be to create a separate calibration model for consolidated sediments, or ultimately a combined unconsolidated/consolidated calibration model. A model for consolidated sediments would require that diatom inferred pH values instead of measured lake-water pH values have to be used. This is, however, not a major problem as a large number of diatom pH reconstructions, made in several countries, have demonstrated a very good agreement between measured and diatom inferred pH. One might argue that the diatom inferred pH gives a better integrated record of the lake-water pH than do ordinary pH measurements.

The results presented in this study were obtained using an instrument mainly developed for food analysis, with optical filters for only 19 wavelengths. Preliminary studies have been made with a scanning instrument measuring at every fourth nanometer between 400 and 2500 nm. These results seem to improve the pH calibrations considerably, with the implication that spectral information important for predicting lake-water pH also occurs at wavelengths not measured by the filter instrument used in this study.

For further improvements in the statistical analysis, mathematical transformations of the data could also be tested, using, for example, the derivative of the spectra or Fourier transforms. Different algorithms for nonlinear PLS calibration might also improve the predictions. Other future approaches could involve improvement of sample preparation. Freeze-drying, instead of oven-drying, might preserve information lost by heating. To obtain a more homogenous particle size distribution, grinding could be made using more sophisticated methods, as more homogenous samples should give less noise in the calibration model.

### Conclusions

Acidification of lakes results in major changes of flora and fauna, which are reflected in the sediment composition. Palaeolimnologists have used detailed microscopical analyses of a variety of biological remains from sediment cores to study lake acidification trends. More than 50 years of research has made diatom analysis the most accurate technique for inferring past pH from sediments, but it is, unfortunately, a very time-consuming method that also requires specially trained analysts.

We have presented here a new approach to infer past pH by using near-IR analysis, which is a fast and nondestructive spectroscopic technique. The near-IR spectrum of a sediment sample is a function of the properties of the

sedimentary material, and relationships between the spectral data and environmental parameters such as pH can be established by multivariate calibration methods.

PLS predictions of pH from the near-IR spectra of surface sediments show good agreement with measured lake-water pH. The same applies for the prediction of diatom inferred pH by analysis of sediment cores from two acidified lakes. There is clearly pH-related information in the lake sediment, which can be detected by near-IR analysis. We believe that near-IR analysis can become useful for inferring past pH, as well as several other lake-water parameters, from sediment cores.

Although near-IR analysis is a widely used technique in industry, its application to lake sediments and environmental issues is novel. The possibilities of the method are extensive and its application to other issues (e.g., eutrophication and pollution) need to be considered.

#### Acknowledgments

We thank Erik Renberg and Carina Jonsson for help with the near-IR analyses, Allan Holmgren and Göran Wikander for discussions about near-IR theory, and John Anderson and Frank Oldfield for valuable comments on the manuscript.

#### Literature Cited

- (1) Mason, B. J., Ed. *The surface waters acidification programme*; Cambridge University Press: Cambridge, England, 1990.
- (2) Norton, S. A.; Lindberg, S. E.; Page, A. L., Eds. *Acidic precipitation. Vol. 4: Soils, aquatic processes, and lake acidification*; Springer-Verlag: New York, 1990.
- (3) Battarbee, R. W.; Mason, J.; Renberg, I.; Talling, J. F., Eds. *Palaeolimnology and lake acidification*; Cambridge University Press: Cambridge, England, 1990.
- (4) Charles, D. F.; Battarbee, R. W.; Renberg, I.; van Dam, H.; Smol, J. P. In *Acidic precipitation. Vol 4: Soils, aquatic processes, and lake acidification*; Norton, S. A.; Lindberg, S. E., Page, A. L., Eds.; Springer-Verlag: New York, 1990; pp 207-276.
- (5) Dixit, S. S.; Smol, J. P.; Kingston, J. C.; Charles, D. F. *Environ. Sci. Technol.* **1992**, *26*, 23-32.
- (6) Ter Braak, C. J. F.; van Dam, H. *Hydrobiologia* **1989**, *178*, 209-223.
- (7) Birks, H. J. B.; Line, J. M.; Juggins, S.; Stevenson, A. C.; ter Braak, C. J. F. *Philos. Trans. R. Soc. London, B* **1990**, *327*, 263-278.
- (8) Wetzel, D. L. *Anal. Chem.* **1983**, *55*, 1165A-1176A.
- (9) Osborne, B. G.; Fearn, T. *Near infrared spectroscopy in food analysis*; John Wiley & Sons: New York, 1986.
- (10) Tkachuk, R.; Kuzina, F. D. *Can. J. Plant Sci.* **1982**, *62*, 875-884.
- (11) Asher, M. J. C.; Cowe, I. A.; Thomas, C. E.; Cuthbertson, D. C. *Plant Pathol.* **1982**, *31*, 363-371.
- (12) Porteous, R. L.; Muir, A. Y.; Wastie, R. L. *Agric. Eng. Res.* **1981**, *26*, 151-160.
- (13) Nilsson, M.; Carlsson, U.; Elmqvist, T. *Phytopathology*, in press.
- (14) Burneau, A.; Corset, J. *Can. J. Chem.* **1973**, *51*, 2059.
- (15) Renberg, I. *J. Paleolimnol.* **1991**, *6*, 167-170.
- (16) Day, M. S.; Fearn, F. R. B. *Lab. Pract.* **1982**, *31*, 439-443.
- (17) Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Espensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. In *Chemometrics: Mathematics and statistics in chemistry*; Kowalski, B. R., Ed.; D. Reidel Publishing Co.: Dordrecht, The Netherlands, 1984; pp 17-95.
- (18) Öhman, J. Ph.D. Dissertation, University of Umeå, Sweden, 1988.
- (19) Martens, H.; Naes, T. *Multivariate calibration*; John Wiley & Sons: New York, 1989.
- (20) Osten, D. W. *J. Chemom.* **1988**, *2*, 39-48.
- (21) Renberg, I. *Boreas* **1981**, *10*, 255-258.
- (22) Battarbee, R. W.; Renberg, I. *Philos. Trans. R. Soc. London, B* **1990**, *327*, 227-232.
- (23) Renberg, I.; Brodin, Y.-W.; Cronberg, G.; El-Daoushy, F.; Oldfield, F.; Rippey, B.; Sandøy, S.; Wallin, J.-E.; Wik, M. *Philos. Trans. R. Soc. London, B* **1990**, *327*, 391-396.
- (24) Grahn, O.; Hultberg, H.; Landner, L. *Ambio* **1974**, *3*, 93-94.
- (25) Grahn, O. *Water, Air, Soil Pollut.* **1977**, *7*, 295-305.
- (26) Renberg, I. *J. Paleolimnol.* **1990**, *4*, 87-90.
- (27) Stevenson, A. C.; et al. *The surface waters acidification project palaeolimnology programme: Modern diatom/lake-water chemistry data-set*; ENSIS Publishing: London, 1991.
- (28) Geladi, P.; MacDougall, D.; Martens, H. *Appl. Spectrosc.* **1985**, *39*, 491-500.

Received for review December 4, 1991. Revised manuscript received May 20, 1992. Accepted July 3, 1992. This study was supported by grants from the Centre for Environmental Research, University of Umeå, and Swedish Natural Science Research Council.