Near-infrared spectroscopic assessment of contamination level of sewage
Tetsuya Inagaki, Yukari Shinoda, Mitsuhiro Miyazawa, Hitoshi Takamura and Satoru Tsuchikawa

ABSTRACT
We examined the use of near infrared (NIR) spectroscopy as a rapid technique for the evaluation of sewage quality. Influent water samples, primary sedimentation tank water samples, and final effluent water samples were collected from sewage treatment facilities in Nagoya, Japan and their NIR spectra obtained. Partial least squares (PLS) models for total phosphate (TP), total nitrogen (TN), biochemical oxygen demand (BOD), total organic carbon (TOC), and turbidity of sewage water were constructed from the NIR data. The models provided good correlation between measurements obtained conventionally and those predicted from spectroscopy. Spectral variation induced by background interference in samples affected accuracy. Loading plots and score plots derived from PLS regression analysis resolved the background interference and allowed highly accurate predictions. Spectral variation induced by contamination in the sewage was a main predictor of sewage quality. These results show that NIR spectroscopy shows potential for in-line, non-destructive measurement of sewage quality.

Key words | environmental monitoring, NIR spectroscopy, PLS regression, sewage analysis

INTRODUCTION
Destruction of the natural environment by human activities is increasing. Eutrophication of water is one of the more serious environmental problems. The degree to which treated sewage will cause eutrophication generally is indicated by the levels of total phosphorus (TP) and total nitrogen (TN), as well as biochemical oxygen demand (BOD) and total organic carbon (TOC). Therefore, it is important to monitor these parameters at sewage treatment plants. Raw sewage consists mainly of water, but also contains some solids produced by humans consisting of feces, toilet paper, laundry waste, and other materials that are disposed of down drains and toilets from households, offices, factories, and other industries.

Sewage water is a complex matrix, with many chemical characteristics. Conventionally, the degree of wastewater pollution is determined by chemical or biological analyses; however, these are not suitable for in-line monitoring because of the time required to conduct measurements. NIR radiation provides molecular vibrational information that can penetrate suspensions, allowing in-line monitoring of sewage (Osborne et al. 1993; Osborne & Fearn 1998; Siesler et al. 2002). NIR spectroscopic evaluation of TOC, COD, and BOD in drainage, partially purified drainage, purified drainage, and river water at various pollution levels has been previously investigated (Takamura et al. 2002; Omasa et al. 2004). These researchers employed two wavelength bands, 680–1,235 nm (14,700–8,100 cm⁻¹) and 1,100–2,500 nm (9,100–4,000 cm⁻¹), to predict pollutant concentrations. They concluded that good predictive accuracy was obtained only within the range of 680–1,235 nm, although good calibration was obtained using both wavelength ranges. The method is based on chemometrics, the art of processing data with numerical techniques that extract useful information from the spectra.
Chemometrics can be used not only for the prediction but also the understanding of spectral variation, thus it is useful to review the theoretical background of relevant spectroscopic methods.

In-line measurement of sewage contamination by UV–Vis radiation has been previously investigated (Langergraber et al. 2003, 2004; Rieger et al. 2004). Reports indicated that PLS regression analysis employing multiple UV-VIS wavelengths exhibited strong correlation between conventionally measured and spectroscopically predicted pollutant levels, which indicated a practical device based on this technique.

Theoretically, sensitivity of molecular vibration increases with a decrease in the wavenumber of light, indicating that NIR light from 10,000–4,000 cm\(^{-1}\) might be useful to monitor ppm-order variation in sewage quality (Ding et al. 2000). Fine chemometric analysis and a theoretical background investigation are important when investigating NIR spectroscopy as an environmental monitoring tool. In this study, the predictability of the contamination level in sewage was investigated using NIR spectroscopy, taking into consideration previous investigations.

**MATERIALS AND METHODS**

**Sample preparation**

Influent water samples, primary sedimentation tank water samples, and final effluent water samples were collected from several sewage factories in Nagoya, Japan, between July 2008 and October 2008 (six samples/day). A total of 55 samples were obtained from nine sampling locations (three or six samples were obtained from each location). Sampling time was 10:00–11:00 a.m. Sewage waters from these factories contained both industrial effluent water and human sewage. Several types of sewage parameters generally used as indices of pollution in sewage factories in Japan were employed here. Sewage water samples were immediately subjected to NIR spectral, TP, TN, BOD, and turbidity measurements. A subset of sewage water samples were frozen and stored, followed by TOC measurements when all samples had been collected.

**NIR spectral acquisition**

NIR spectra of sewage water were acquired in a transmission mode (path length: 1 mm) using a FT-NIR spectrometer (Bruker: MATRIX-F) with fiber optics. To improve the signal-to-noise ratio, 128 scans were accumulated at a spectral resolution of 8 cm\(^{-1}\) over the wavenumber range of 10,000–5,500 cm\(^{-1}\). Before each sample's spectroscopic data collection, the transmittance spectra of an empty quartz cell was obtained as a reference. Temperatures in the quartz cell were maintained at 25.0 ± 0.1°C using a water-jacketed cell holder and refrigerated water bath.

**Conventional chemical and biological measurements**

Conventional chemical and biological analyses were performed using standard methods employed at sewage treatment plants in Japan. A TOC analyzer (Shimadzu Corporation: TOC-VCSH) was used to determine TOC. BOD was calculated from the difference in oxygen concentration before and after a five-day incubation period. An oxygen electrode (HORIBA: OM-51) determined oxygen concentrations. TP and TN were measured by light absorbance at 880 nm and 220 nm, respectively, after samples were chemically dissolved by autoclave treatment with potassium peroxydisulfate (the “simultaneous resolve method” employed in the Nagoya sewage treatment plant). Turbidity was measured using a turbidity meter (HACH: 2100N turbidimeter).

**Multivariate analysis**

The spectral and chemical data were imported into multivariate software (CAMO software AS Unscrambler ver 9.6) for data analysis. Calibrations were developed for each measured sewage parameter by partial least squares (PLS) regression analysis (Kramer 1998). PLS relates the variance in data set \( Y \) block to the variance in independent parameters \( X \) block, using latent variable \( t \) block.

\[
X = \sum_{a=1}^{A} t_a p_a^T + E
\]

\[
Y = \sum_{a=1}^{A} t_a q_a^T + F
\]
where \( t_a \) is the \( a \)-th latent variable, \( p_a \) and \( q_a \) are the \( a \)-th loadings for the \( X \) and \( Y \) blocks, respectively, \( E \) and \( F \) blocks are residuals, and \( A \) is number of PLS components. Variable \( t \) is a linear combination of independent parameters \( X \) and PLS weight block \( w \), as shown in Equation (3).

\[
t = Xw
\]  
(3)

Loading and score plots provided significant information about each data set. Furthermore, the \( X \)-explained value and the \( Y \)-explained value also gave significant information. The \( X \)-explained value represents how well the PLS component explains the variation in the independent variable (i.e. NIR spectral data), while the \( Y \)-explained value represents how well the PLS component explains the variation in the dependent variable (i.e. TP, TN, BOD, TOC, and turbidity data).

Statistical pre-treatment of NIR spectral data was applied to provide a more accurate prediction model. Multiplicative scatter correction, which corrects measured spectral data using averaged spectra and the smoothing of 21 points, was used. Spectral pre-treatment reduced the noise due to sample conditions or spectral measurement conditions. The full inner cross-validation method ("leave one out") was used to determine the optimum number of PLS components (Kramer 1998). This method involves the use of a single value from the original sample as validation data, and the remaining values as training data. This procedure is repeated until each value in the sample is used once as validation data. Leave-one-out cross-validation often is used when the number of samples is small. Optimum number \( A \) is the number of PLS components which provide the greatest predictive value for validation. Calibration and validation were evaluated by correlation coefficient (\( R \) and \( r \)) between measured and predicted values, and root mean square errors of calibration and validation model predictions, \( RMSEC \) and \( RMSECV \), respectively.

**RESULTS AND DISCUSSION**

**Variation of NIR spectra**

Figure 1 shows an NIR spectrum of a sewage water sample collected from one of the sewage treatment plants in Nagoya. The spectrum is dominated by a strong absorption band due to the OH stretching vibration at 6,000–7,200 cm\(^{-1}\).

**Variation in measured sewage parameters**

Table 1 shows statistical results of measured sewage parameters, and Table 2 shows the high correlation between these parameters. The correlation between BOD and turbidity is especially high because light scattering particles in sewage consist mainly of organic matter.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Statistical results of measured sewage parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>SD</td>
</tr>
<tr>
<td>TP (mg/L)</td>
<td>2.60</td>
</tr>
<tr>
<td>TN (mg/L)</td>
<td>20.49</td>
</tr>
<tr>
<td>BOD (mg/L)</td>
<td>64.60</td>
</tr>
<tr>
<td>TOC (mg/L)</td>
<td>20.60</td>
</tr>
<tr>
<td>Turbidity (NTU)</td>
<td>30.90</td>
</tr>
</tbody>
</table>

Notes: SD: Standard deviation, NTU: Nephelometric turbidity units.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Correlation coefficients between measured sewage parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>TN</td>
</tr>
<tr>
<td>TP</td>
<td>1.00</td>
</tr>
<tr>
<td>TN</td>
<td>0.78</td>
</tr>
<tr>
<td>BOD</td>
<td>0.72</td>
</tr>
<tr>
<td>TOC</td>
<td>0.71</td>
</tr>
<tr>
<td>Turbidity</td>
<td>0.71</td>
</tr>
</tbody>
</table>
PLS analysis for sewage parameters

The relation between measured values and values predicted by PLS for each parameter is shown in Figure 2. Statistical results are summarized in Table 3. PLS analysis provided regression models between sewage parameter measurements and latent variables calculated from NIR spectra and component loadings. The complex prediction mechanism is discussed using the regression model for each measured sewage parameter.

The models provided good correlation between measured and predicted values for TP, TN, BOD, TOC, and turbidity with correlation coefficients for cross-validation of 0.79, 0.78, 0.83, 0.79, and 0.75, respectively. The RMSECV for each PLS regression model was adequately small at the ppm level. These calibrations demonstrate the ability of NIR spectroscopy to accurately predict multiple sewage parameters without any sewage pretreatment. The optimum number of PLS components was five for turbidity and seven for the other four parameters.

Good calibration results have been reported based on UV/VIS spectroscopy for nitrate, nitrite, soluble COD (chemical oxygen demand), and TSS (total suspended solids) with coefficients of determination of 0.993, 0.978, 0.905, and 0.848, respectively (Langergraber et al. 2003, 2004; Rieger et al. 2004). Thus it might be useful to develop a new multiple-measurement tool for sewage using a broad-spectrum spectroscopic technique (such as UV/VIS/NIR).

Loading plots and score plots derived from PLS analysis

Both loading and score plots provide significant information about the NIR spectra and were investigated to obtain the best prediction. Figure 3 shows the loading plots of the first, second, third, fourth, and fifth PLS components for each sewage parameter. The paired numbers in parentheses in these plots indicate the percent variation in the X and Y values, respectively, for each parameter. The X value represents how well the PLS component explains the variation in the independent variable (i.e. NIR spectra), while the Y value represents how well the PLS component explains the variation in the dependent variable (i.e. TP, TN, BOD, TOC, or turbidity) (Kramer 1998). Figure 4 shows score plots for BOD, where marks were separated according to (a) measurement date and (b) sample type.

For the first PLS component for all five parameters, the loading coefficients from 10,000 cm⁻¹ to 8,000 cm⁻¹

Table 3 | Statistical results of PLS regression analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number of samples</th>
<th>Optimum number of factors</th>
<th>RMSEC</th>
<th>R</th>
<th>RMSECV</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP (mg/L)</td>
<td>55</td>
<td>7</td>
<td>0.69</td>
<td>0.89</td>
<td>0.93</td>
<td>0.79</td>
</tr>
<tr>
<td>TN (mg/L)</td>
<td>55</td>
<td>7</td>
<td>3.27</td>
<td>0.91</td>
<td>5.10</td>
<td>0.78</td>
</tr>
<tr>
<td>BOD (mg/L)</td>
<td>55</td>
<td>7</td>
<td>20.70</td>
<td>0.90</td>
<td>29.40</td>
<td>0.80</td>
</tr>
<tr>
<td>TOC (mg/L)</td>
<td>55</td>
<td>7</td>
<td>5.39</td>
<td>0.89</td>
<td>7.60</td>
<td>0.79</td>
</tr>
<tr>
<td>Turbidity (NTU)</td>
<td>27</td>
<td>5</td>
<td>10.80</td>
<td>0.88</td>
<td>15.70</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Notes: RMSEC: Root mean square error for calibration; R: Correlation coefficient between measured and predicted value for calibration; RMSECV: Root mean square error for cross-validation; r: Correlation coefficient between measured and predicted value for cross-validation.
showed especially high positive values compared with other wavenumber ranges. The first PLS component score (see Figure 4(a)) shows several clusters by collection date. The first PLS component accounts for spectral baseline shift caused by variation in reference measurements. The first PLS component explains a small percentage of variance in the dependent variable (1% for BOD and 4% for turbidity), whereas explained variance in the independent variable (NIR spectra) is high (84% for BOD and 69% for turbidity).

Loading plots for the second and third PLS components show a different trend for turbidity than for all other sewage parameters. In the second PLS component for TP, TN, BOD, and TOC, and for the third component for turbidity, a characteristic negative peak in loading coefficient is evident near 7,000 cm\(^{-1}\). Although sample type corresponds to a decrease in contamination throughout the sewage treatment process, scores for the second PLS components, which express the variation of absorption due to water, (see Figure 4(b)) were not separated by sample type.
The variation in $Y$ explained by the second PLS component for BOD also was small (5.8%). These results suggest that the calculated second PLS component for BOD (and for TP, TN, and TOC) explains the spectral variation due to slight differences in setup of the quartz cell for NIR measurements. For the third PLS component for TP, TN, BOD, and TOC, and for the second component for turbidity, the loading coefficient also showed a characteristic peak due to water. The relatively high $Y$-explained value of 34% for BOD coupled with the low $X$-explained value of 1% for BOD indicates that the third PLS component significantly explains the variation in BOD. It is therefore suggested that the third PLS component include any chemical or physical change in sewage water with the variation of contamination level.

For prediction, the same NIR spectra used for calibration were employed. Thus, any difference in order of PLS component between turbidity and other sewage parameters is due to differences in variation of the independent variables. The $Y$-explained values were used to estimate the second PLS component for turbidity, and the third PLS component explains the variation in water absorption band for other sewage parameters caused by changes in turbidity.

The fourth and the fifth PLS components explain considerable variation in the dependent sewage parameters while independent variables were very small. We could not observe characteristic spectral feature in loading plot (see Figure 3(d,e)). Further research is needed to clarify this point and should attempt to reduce the effect of the spectral baseline shift and differences in the setup of the quartz cell, which affect mainly the first and the second PLS components, respectively.

The variation in NIR spectra was classified into several factor types by PLS regression analysis. Results indicate that NIR spectra can predict multiple sewage parameters simultaneously without pre-treatment of sewage if attention is paid to the light-scattering caused by turbidity.

**CONCLUSIONS**

The feasibility of using NIR spectroscopy for measurement of TP, TN, BOD, TOC, and turbidity in sewage was investigated. PLS regression models were constructed to show the relations between NIR spectra and measured sewage parameters. The models showed strong relations between measured and predicted values. Spectral variations were classified into several factors by PLS regression analysis, which were used to predict each sewage parameter. Spectral variation due to sample turbidity is a main component of the prediction. Prediction accuracy was sufficient, and evaluation of loading plots revealed the prediction mechanism. NIR spectroscopy presents the potential for the in-line measurement of sewage quality without sewage pretreatment.

**ACKNOWLEDGEMENTS**

We thank Nagoya City Water Works and Sewage Bureau for their cooperation during this research project and Professor Dr. Chisato Takenaka (Nagoya University) for measurement of TOC.

**REFERENCES**


Langergraber, G., Fleischmann, N. & Hofstädt, F. 2003
A multivariate calibration procedure for UV/VIS
spectrometric quantification of organic matter and
Langergraber, G., Fleischmann, N., Hofstädt, F. & Weingartner,
A. 2004 Monitoring of a paper mill wastewater treatment plan
Omasa, M., Takamura, H. & Matoba, T. 2004 Evaluation of
household drainage from food materials by near infrared
spectroscopy. *Near Infrared Spectroscopy: Proceedings of the
11th International Conference*, pp. 835–838.
Osborne, B. G. & Fearn, T. 1998 *Near Infrared Spectroscopy in
Food Analysis*. Longman Scientific and Technical, Harlow,
Essex, UK, p. 20.
Osborne, B. G., Fearn, T. & Hindle, P. H. 1993 *Practical
NIR Spectroscopy with Applications in Food and
Beverage Analysis*, Longman Scientific and Tech.,
Harlow.
Rieger, L., Langergraber, G., Thomann, M., Fleischmann, N. &
Siegrist, H. 2004 Spectral in-situ analysis of NO2, NO3, COD,
DOC and TSS in the effluent of WWTP. *Water Sci. Technol.*
**50**(11), 143–152.
Takamura, H., Miyamoto, H., Mori, Y. & Matoba, T. 2002
Evaluation of drainage by near infrared spectroscopy. *Near
Infrared Spectroscopy: Proceedings of the 10th International
Conference*, pp. 405–408.