

A multivariate calibration procedure for UV/VIS spectrometric quantification of organic matter and nitrate in wastewater

G. Langergraber*, N. Fleischmann*, and F. Hofstädter**

* Department for Sanitary Engineering and Water Pollution Control (IWGA-SIG), University of Agricultural Sciences Vienna (BOKU), Muthgasse 18, A-1190 Vienna, AUSTRIA

(E-mail: langergaber@iwga-sig.boku.ac.at (corresp. author), fleischmann@iwga-sig.boku.ac.at)

** s::can Messtechnik GmbH, Herminengasse 10, A-1020 Vienna, AUSTRIA (www.s-can.at)

(E-mail: fhofstaedter@s-can.at, aweingartner@s-can.at)

Abstract

A submersible UV/VIS spectrometer for in-situ real-time measurements is presented. It utilises the UV/VIS range (200-750 nm) for simultaneous measurement of COD, filtered COD, TSS and nitrate with just a single instrument. A global calibration is provided that is valid for typical municipal wastewater compositions. Usually high correlation coefficients can be achieved using this standard setting. By running a local calibration improvements concerning trueness, precision and long term stability of the results can be achieved. The calibration model is built by means of PLS, various validation procedures and outlier tests to reach both high correlation quality and robustness. This paper describes the UV/VIS spectrometer and the calibration procedure.

Keywords

In-situ, multivariate calibration, PLS, real-time, on-line, UV/VIS spectrometry.

INTRODUCTION

A submersible UV/VIS spectrometer for in-situ real-time measurements is presented. The instrument proved to be a promising instrument to quantify wastewater pollutant loads for integrated management and control of municipal sewer networks and monitoring of the treatment plant itself. The submersible spectrometer utilises the UV/VIS range (200-750 nm) for simultaneous measurement of organic matter, suspended solids and nitrate with just a single instrument. The miniaturised UV/VIS spectrometer is a probe installed directly in the process. Therefore it requires no sampling, no sample preparation, and no reagents. The spectrometer is equipped with an auto-cleaning system.

A global calibration for typical municipal wastewater is provided as default configuration of the UV/VIS spectrometer. Usually high precision can be achieved using this standard parameter set. For many purposes such as plant control precision is more important than trueness and the global calibration often delivers sufficient results.

Due to the different composition of wastewaters, e.g. with significant industrial contributions, a second calibration step (local calibration) can be required to enhance the trueness. The local calibration correlates lab analysis with in-situ measurements of the actual wastewater. The local calibration is able

- ⇒ to take account of specific compositions of the wastewater and
- ⇒ possible matrix effects at the measurement location and/or,
- ⇒ to minimize cross sensitivities due to overlapping spectral shapes of target substances.

By running a local calibration improvements concerning trueness, precision and long term stability of the results can be achieved.

THE UV/VIS SPECTROMETER

The submersible UV/VIS spectrometer (Figure 1) is a spectrometric probe of 44 mm diameter and about 0.6 m length. It records light attenuation in the wavelength region between 200 nm and 750 nm and displays and/or communicates the result in real time. The measurement takes place directly in-situ without sampling or sample treatment. A single measurement typically takes about 15 seconds. The instrument is equipped with an auto-cleaning system using pressurised air. Due to the compact size the sensor can easily applied in 2" bore holes, for example for groundwater monitoring. An ex-proofed version is available for sewer applications.

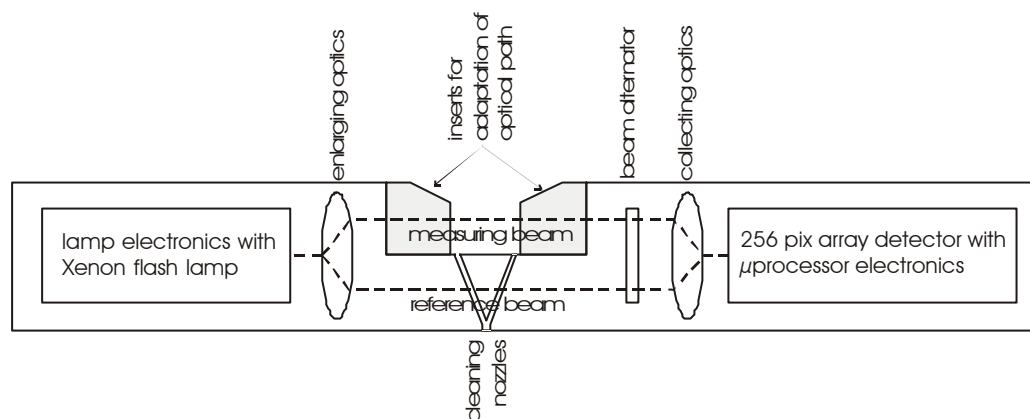


Figure 1. UV-VIS submersible spectrometer.

The instrument is a 2-beam 256 pixel UV/VIS spectrometer, with a Xenon lamp as a light source. On-board electronics control the entire measurement procedure. All of the controller electronics are included in the 44 mm tubular anodised aluminium housing and include a data logger and a water level meter. The communication is via RS232 or RS485 interface. The power supply can be by means of AC 220 V/50 Hz or DC 12-24 V, respectively. The low power consumption eases field application by means of battery or solar power supply. The probe has a data logger on board, capable to store e.g. complete absorption spectra of 1 month at a measuring interval of 30 minutes.

The path length can be adjusted from 2-100mm. This opens a wide range of applications from ultra pure waters ($\text{DOC} > 10 \mu\text{g/l}$) up to concentrated wastewaters with a COD of several 1000 mg/l. Usually a path length of 5 mm is used for wastewater applications. Spectrometric measurement methods are defined for single substances (e.g. Nitrate, Nitrite, Benzene, Xylene, Toluene) as well as for surrogate parameters (SAC, turbidity / suspended solids, CODEq, TOCeq, DOCEq).

The most important influence on in-situ absorption measurements is turbidity. Turbidity due to suspended substance causes light scattering, shading and thus influences absorption over the entire spectrum, whereas other interfering substances absorb light only in a limited wavelength region. It is well known that the spectral shape caused by suspended solids depends upon the wavelength with a factor λ^x , where x depends on the particle diameter. Therefore for turbidity compensation a mathematical equation was developed which describes the relationship between scattering intensity and wavelength as a function of the particle diameter based on the basic relationships given by Huber and Frost (1998). The turbidity compensation feature uses the original spectrum and estimates two parameters of the turbidity function. Turbidity compensation has two tasks: the measurement of turbidity/suspended solid and baseline compensation for the measurement of dissolved substances. The results of the turbidity compensation showed to be very sensitive to the initial values of these parameters.

The main advantage of the probe is, that it is applied directly in the process. Thus, measurement errors due to sampling, transport, storage, dilution etc. are not relevant. The broad range of available wavelengths allows high flexibility for the choice of the best correlating wavelengths for the calibration function and for the avoidance of cross sensitivities. This is an advantage compared to systems, which provide absorption measurements of a single or two wavelengths only.

SENSOR CALIBRATION

Principles

The linear relation between absorption and concentration of a single determinant is given by Lambert-Beer's law. Wastewater monitoring has to deal with a matrix of numerous dissolved and suspended compounds. The superposition of numerous single substance absorbances – sometimes even with overlapping peaks – can cause cross sensitivities and can lead to poor performance of the sensor. In this case the sensor can be calibrated with the water matrix of concern. Chemometric models are used for this purpose. These models formalise the procedure of correlating the required determinants to spectra.

Direct chemometric models can only be used if the spectra of all constituents are known and Lambert-Beer's law is valid. Therefore they cannot be used for wastewater where a great number of constituents is present and therefore surrogate parameters (e.g. COD, TSS) are commonly used. Indirect chemometric models are based on estimating the calibration parameters from calibration mixtures and have the following advantages (Otto,1999):

1. Interactions between constituents or between constituents and the sample matrix can be accounted for in the calibration (validity of Lambert-Beer's law is no longer a prerequisite).
2. Modelling of the background in a principal component becomes feasible.
3. Systems of even highly correlated spectra can also be used for multi-component calibration.

All types of calibration show better results if there is no or only poor correlation between the variables. Wastewater shows strong correlations between the various parameters (e.g. COD is strongly correlated with filtrated COD and TSS). Due to the co-linearities indirect calibration models are suitable for the problem. Additionally the basic assumption for standard correlation problems that reference measurements have to be error-free is not required (Danzer et al., 2001). The applied model – Partial-Least-Square regression (PLS) – especially accounts for concentration-spectra relationships and results in the most robust calibrations at present (Otto,1999).

The multivariate calibration procedure

Figure 2 shows a scheme of the applied multivariate calibration procedure. The entire evaluation of the spectra is used. The calibration model is built by means of PLS, various validation procedures and outlier tests to reach both high correlation quality and robustness. The result of the calibration procedure is a recovery function. The recovery function is obtained by plotting the reference measurements (actual targets) vs. the predicted values (estimated targets).

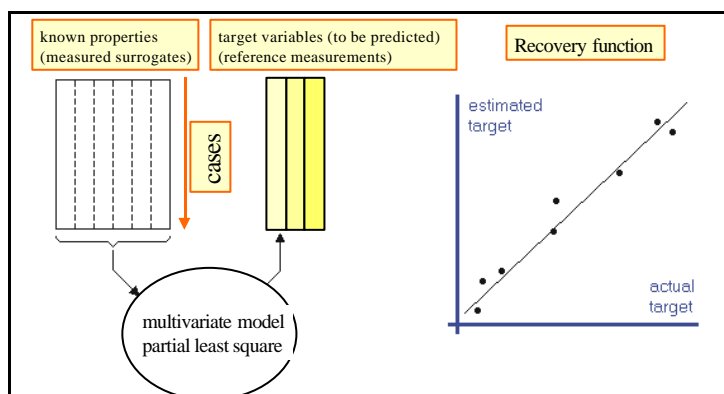


Figure 2: Scheme of the multivariate calibration procedure.

Table 1 shows the steps implemented in the multivariate calibration procedure. At the very beginning the spectral plausibility is validated to exclude wrong spectra. Usually calibration is done with turbidity compensated spectra. As an optional step, a 'local' calibration can be based as well on raw spectra. A recursive procedure includes the multivariate calibration with PLS, a multiple cross validation, and the detection and elimination of outliers and is repeated until a good calibration result is obtained. To perform PLS the spectra and the reference measurements have to be normalized.

Table 1: Steps of the multivariate calibration procedure.

Step	Task	Method
1	Plausibility control	Manually
2	Turbidity compensation of the raw spectra	curve fitting
3	Outlier elimination – Stepwise exclusion of suspected outliers	F- and T-Tests and Cook's distance
4	Building the calibration model	PLS (leave one out)
5	Evaluation of confidence intervals of the calibration function	PLS / leave one out
6	Evaluation of calibration function robustness	PLS / Cross validation

The most important wavelengths are identified automatically, usually 5 to 10 wavelengths are selected. Due to the multi dimensional calibration problem it is time consuming to calculate the global minimum. Therefore straight-forward decisions have to be made to choose the wavelengths and the selection of the wavelengths is checked manually. A plausibility control for the chosen wavelengths shall be included in the procedure.

The estimates of the measurement accuracy for unknown samples are calculated on the basis of a complete leave one out cross validation (Figure 3). Every set of data is removed as a test set from the training set once and a model is computed with the remaining data. Then the removed data are predicted and the sum of the square roots of the residuals over all removed objects is calculated ($PRESS_{CV}$ value – predictive residual sum of squares based on cross validation; Otto, 1999). The number of significant principal components is obtained from the minimum residual error (minimum $PRESS_{CV}$ value). For every constituent a $PRESS_{CV}$ value is calculated. For multiple constituents the mean value of the single $PRESS_{CV}$ values is used to determine the optimal number of principal components.

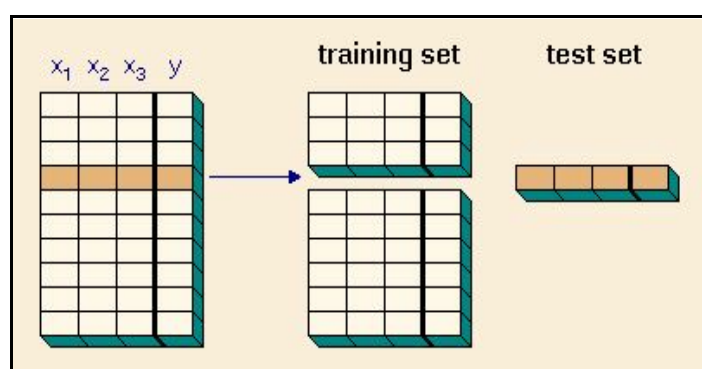


Figure 3: Leave one out cross validation – principle (Lohniger, 1999).

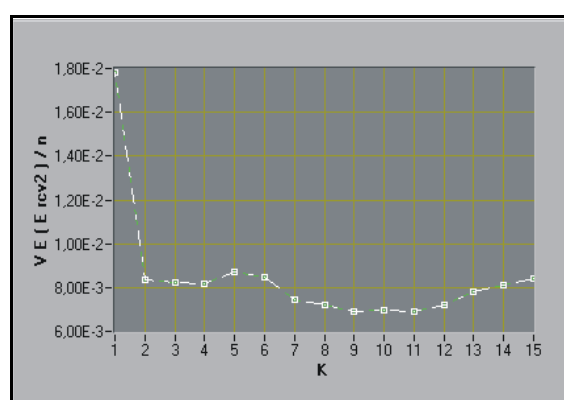


Figure 4: $PRESS_{CV}$ vs. principal components.

Figure 4 shows an example of a $PRESS_{CV}$ vs. principal components plot. The number of principal components is given on the x-axis and the $PRESS_{CV}$ value on the y-axis. It can be clearly seen that the 2nd principal component reduces the residuals significantly more than the subsequent ones. A global minimum of the $PRESS_{CV}$ value appears with 9 principal components. In this case the calibration would include the first 9 principal components.

The model fit within the training data set can be improved by integrating more principal components into the model. Up to a certain point this method will lead to a better chemometric model, after this point the model will become "over-sensitised". The cross validated residuals will increase when the model starts to become over-sensitised. Stochastic changes within the training data set are then included in the model without being of general validity.

Cross validation is also applied for outlier detection. When a certain test set is removed from the training data set F- and T-tests identify significant (95% level) changes of slope and residuals (Danzon et al., 2001). Discrimination between outliers and influential observations is based on COOK's distance (Otto, 1999). The maximum number of eliminated outliers can be defined by the user.

Global vs. local calibration

The UV/VIS spectrometer provides a global calibration for the parameters of concern, but recalibration with local reference samples will significantly improve the performance (trueness and precision) for quantitative measurements. The local calibration is based on reference grab samples analysed for the parameters of interest. The local calibration can be performed without demounting of the probe.

The whole calibration procedure can only be successful, if the following rules for the reference samples are taken into account:

- ⇒ validation of reference methods of the laboratory analysis,
- ⇒ minimisation of errors induced by sampling and sample storage,
- ⇒ the identity between sampled liquid and liquid measured in-situ has to be guaranteed,
- ⇒ the entire measurement range has to be covered by the reference samples and they have to be equally distributed over this range, and
- ⇒ randomisation of sampling in order to avoid correlation of subsequent samples.

RESULTS AND DISCUSSION

Influence of sampling on the calibration results

The "total error of measurement" consists of sampling, sample transport, storage, and analytical errors. Often comparison of field and lab methods only consider the usually well known analytical error and neglect other elements of the sample to lab chain. It is of interest to quantify the different possible reasons for the deviations between the model and the reference method. Unfortunately the major error contributions are unknown in most cases, nevertheless this fact has to be kept in mind.

Errors caused by sampling devices are well known and can lead to an over-estimation of e.g. COD up to a maximum value of 56% if vacuum sampling devices are used (Haider & Haider, 1998). This over-estimation is caused by sedimentation effects in the sample chamber.

Figure 5 shows the influence of the sampling procedure on the calibration results for COD. To calibrate the UV/VIS spectrometer for industrial wastewater reference samples were taken in two different ways: During the first sampling ('sampling 1') the spectrometer was measuring in-situ while samples for calibration were taken and the sampling times were recorded. Using these samples for calibration the correlation coefficient was only 60.6%. It was supposed that the in-situ and the lab measurements were not done on identical samples. During an additional sampling ('sampling 2') the spectra of the new samples were measured off-line in order to reduce the sampling error. The

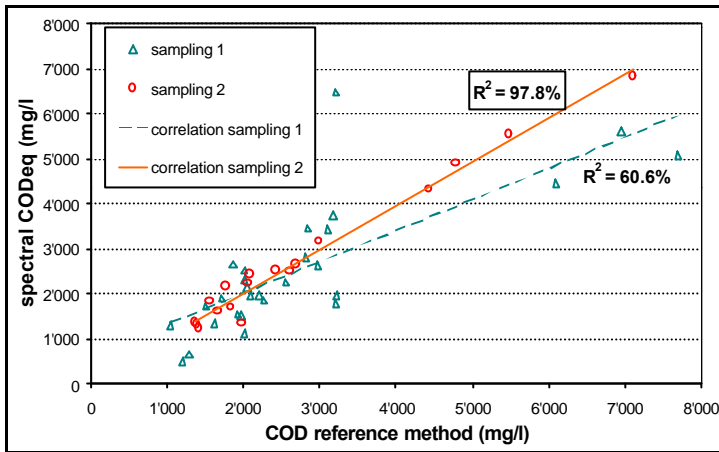


Figure 5: Different calibration results for different sampling strategies.

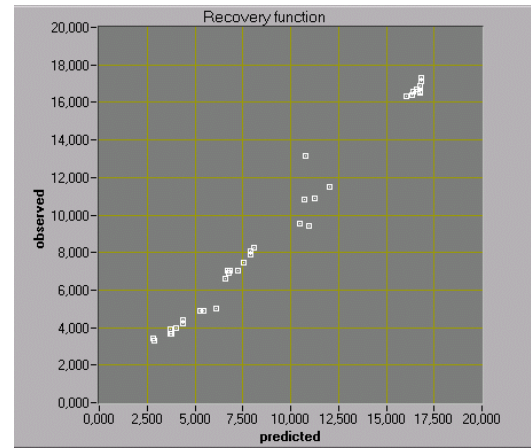


Figure 6: Recovery function - predicted vs. observed nitrate concentration.

calibration results in a very high correlation coefficient of 97.8 %. Taking the standard analytical errors of several percent (usually < 5 %) into account the reached correlation coefficient for 'sampling 2' is rather close to the maximum that can be achieved.

Global calibration for nitrate (effluent of a wastewater treatment plant)

The background matrix always exists in wastewater. Therefore single wavelength measurements for nitrate cannot be used without any calibration on site. Double wavelength systems try to compensate such interferences by the measurement of the absorbance outside of the nitrate peak wavelength range. This is an improvement, but cannot deal with the usual matrix changes that occur particularly by using one calibration on different wastewater treatment plants.

In order to be able to obtain a “global calibration”, effluent data of five different wastewater plants were used. The concentration ranges are 3 to 17 mg_{NO3-N/l} for nitrate and 9 to 300 mg_{COD/l} for the background matrix. Both ranges cover a representative area for effluents of municipal wastewater treatment plants. Figure 6 shows the recovery function for the global calibration of the UV/VIS spectrometer for nitrate. A global calibration for different wastewater plants could be achieved.

Local calibration for COD, filtered COD, TSS, and nitrate (primary clarifier effluent)

Figure 7 and Figure 8 show the calibration results for COD and filtered COD, and for TSS and NO₃-N, respectively. The data were measured over 4 month (Winkler et al., 2002). The first two month were used for calibration (28 data sets) and the last two month for validation (24 data sets).

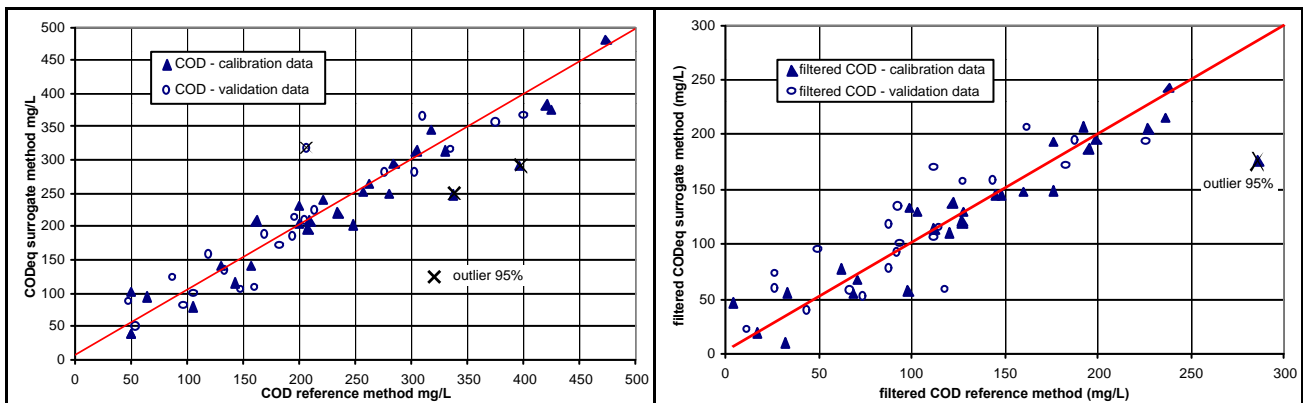


Figure 7: Calibration results for COD (left) and filtered COD (right).

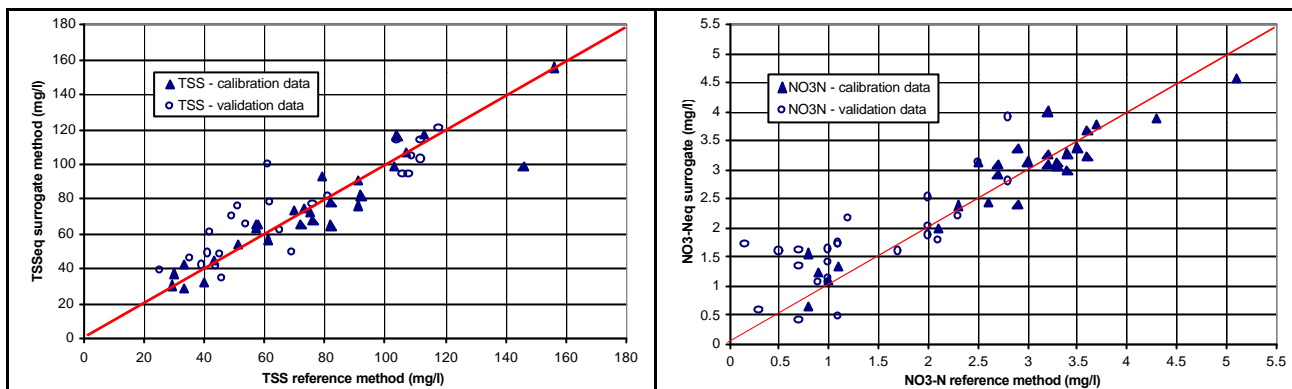


Figure 8: Calibration results for TSS (left) and NO3-N (right).

Table 2 summarises the calibration results. Comparing the obtained correlation coefficients for the validation data set a calibration with a ‘standard’ method using the ‘best’ wavelength for each parameter performs worse than the presented method using the whole spectrum (‘spectrum’ method = multivariate calibration). The correlation coefficients for all 4 parameters are lower for the ‘standard’ method. For COD, filtered COD and TSS the ‘spectrum’ method improves the prediction. However, nitrate can only be predicted using the ‘spectrum’ method. The results of the multivariate calibration procedure are also given in Table: The coefficients of determination, the upper and lower determination limits, and the results of the cross validation (standard error of prediction and mean residual).

Table 2: Calibration results.

Parameter	COD	filtered COD	NO3-N	TSS
Correlation coefficient for ‘standard’ method	0.88	0.72	0.14	0.83
Correlation coefficient for ‘spectrum’ method	0.90	0.91	0.68	0.95
Detailed results of the multivariate calibration procedure				
Coefficient of determination	0.91	0.86	0.90	0.92
Upper limit *	600 (568)	400 (344)	10 (7)	250 (188)
Lower limit *	0 (38)	0 (3)	0 (0)	0 (20)
Standard error of prediction	23.5	16.9	0.29	5.98
Mean residual	28.0	20.2	0.37	7.75

* Values in brackets are given by the calibration procedure

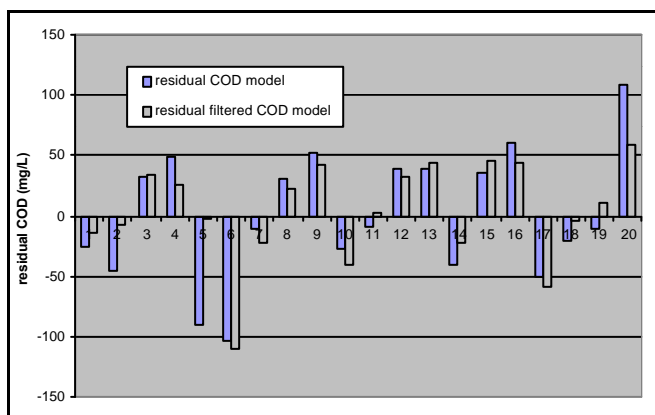


Figure 9: Residual values for COD and filtered COD.

In order to account for the other error contributions all minor residuals (below 15 %) are eliminated from the following graph. Figure 9 shows the residuals for all cases where they are above 15 % of the actual COD value. It is obvious that the residual of COD and filtered COD have the same sign 18 out of 20 cases and that the residuals are all in the same order of magnitude. This is a very strong indication that these deviations have the same source. Due to the same order of magnitude sampling errors as described by Haider & Haider (1998) did not occur.

Therefore for the shown example these deviations can be caused only by two reasons: changes in the matrix composition or a systematic error of the COD analysis. In this case the error for COD analysis was determined to be < 3 % (Winkler et al., 2002). It could be concluded that the main contribution is by changes in the matrix composition.

Whereas COD analysis includes almost 100 % of organic carbon, the surrogate method includes only determinants of the organic carbon matrix that absorb light. If the matrix composition is changing then the overall absorption coefficient changes. There are some important fractions of organic carbon that do not show any absorption in the UV/VIS spectra, like short chain fatty acids, sugars and starch. If the concentration changes of these are not proportional to the total COD substances change it would cause deviations between the chemometric model and the reference measurements. As the substances mentioned above are dissolved they cause deviations for the total as well as for the filtered COD and the deviation must have the same sign.

CONCLUSIONS

A chemometric analysis based on a spectrometric probe and a fully automated PLS calibration is presented. By using PLS it is relative simple to build calibration models that lead to very good coefficients of determination as there are a lot of independent variables provided by the continuous UV/VIS spectra. The presented method is capable to eliminate cross sensitivities of substances with similar spectral shapes or overlapping peaks as well as certain background signals and it can account for matrix effects of different substances (i.e. changes of the spectral shape trough reactions). The experience showed that most of the times the reference data are the critical part of the whole calibration procedure. Therefore it is essential to guarantee the quality of the reference measurements (regarding reference analysis method, measurement range, sampling errors, identity of samples and randomisation of sampling) to obtain good calibration results.

The available global calibration for the UV/VIS spectrometer valid for typical municipal wastewater composition usually provides sufficient measurement accuracy. Due to the different composition of wastewaters, e.g. with significant industrial contributions, a 'local' calibration step can be required to enhance the accuracy. The local calibration correlates datasets of parallel lab and in-situ measurements. By running a local calibration improvements concerning trueness, precision and long term stability of the results can be achieved especially when matrix effects occur.

REFERENCES

- Danzer K., H.Hobert, C. Fischbacher & K.-U. Jagemann (2001): Chemometrik – Grundlagen und Anwendungen; Springer-Verlag, Berlin.
- Haider R. & S. Haider (1998): Overestimated COD values for raw wastewaters extracted by vacuum sampling devices due to sedimentation effects; *Wat.Sci.Tech.* **37**(12), 293-300.
- Huber E. and Frost M. (1998): Light scattering by small particles; *J Water SRT – Aqua* **47**(2), 87-94.
- Lohninger H. (1999): Teach/Me – Data Analysis, Handbook; Springer-Verlag, Berlin
- Otto M. (1999): Chemometrics: Statistics and Computer Application in Analytical Chemistry; Wiley-VCH Verlag GmbH, Weinheim.
- Winkler S., L. Rieger, M. Thomann, H. Siegrist, C. Bornemann & N. Fleischmann (2002): In-Line Monitoring Of COD And COD-Fractionation: Improving Dynamic Simulation Data Quality; Accepted for oral presentation at the *3rd IWA International World Water Congress*, April 7-12, 2002, Melbourne, Australia.