Benefits, limitations and uncertainty of *in situ* spectrometry

S. Winkler, J.-L. Bertrand-Krajewski, A. Torres and E. Saracevic

ABSTRACT

In situ spectrometers are comparable robust instruments and can be operated long term with relatively low maintenance demand. The decisive factor for their applicability is the development of a medium and installation location specific correlation model, which estimates the concentration of the requested target parameter from the measured absorption values. In some cases it might turn out, that the development of a single site specific correlation model is not sufficient, due to frequent and substantial variations of the (waste) water composition.

In order to assess the total uncertainty of *in situ* spectrometry, a comprehensive lab test was carried out. It includes a detailed investigation of the uncertainty of CODreference methods (DIN and small tube tests), a precision analysis of the absorption spectra and the impact of those two factors on the total uncertainty of the COD_{EQ} correlation model.

Key words | COD, optical sensors, PCA, PLS, spectrometry, turbidimetry, uncertainty

S. Winkler
E. Saracevic
Institute of Water Quality and Waste Management,
Vienna University of Technology,
Karlsplatz 13/E 226, A - 1040 Vienna,
Austria
E-mail: *swinkler@iwag.tuwien.ac.at*

J.-L. Bertrand-Krajewski
A. Torres
LGCIE, INSA-Lyon, 34 avenue des Arts,
F-69621 Villeurbanne cedex,
France
Email: *jean-luc.bertrand-krajewski@insa-lyon.fr*;
andres.torres@insa-lyon.fr

INTRODUCTION

While *in situ* spectrometers show application advantages compared to traditional analysers (in situ, multiparameter, no permanent consumption of chemicals, etc.) their applicability is implicitly limited by their indirect measurement principle. Absorption is, by definition, measured in a particle free medium. If a spectrometer is applied in a medium with particles, the measured signal is the total light attenuation due to absorption of the medium at the applied wavelength range and the scattering of light due to particles. This already highlights the requirement to consider the water matrix properties and even more important its changes when deriving target parameters from the measured absorption signal. The decisive factor is whether suitable correlation models can be established which calculate the requested target parameters from the measured absorption and which are (i) accurately enough to fulfil the application requirements and (ii) robust enough to compensate for disturbances (mainly matrix changes). The following main sources of uncertainty can be named:

(1) *Concerning the spectrometer*:

- Mixing conditions at the measurement location; e.g. impact of the sample presentation on the measured absorption and its precision
- Magnitude of matrix changes at the measurement location
- Precision of *in situ* spectrometers
- Longterm stability of the instrument (light source and detector, S:N ratio, etc.)

(2) *Concerning the reference analysis*:

- Sampling uncertainty at the measurement location (taking grab samples)
- Storage and transport of the reference samples
- Sample processing at the laboratory
- Uncertainty of the reference method

(3) *Concerning the derived 'target' parameter*:

- The combined effect of all the above mentioned factors on the total uncertainty of the correlation

model including the uncertainty of the model structure.

An experimental setup was chosen to investigate the influence of some of the key elements impacting the total uncertainty of the target parameter (in this paper: total COD of raw sewage) derived from an UV-(VIS)-measurement; the elements indicated in *italic* were chosen. The sampling uncertainty at the monitoring location ('on site') and the impact of changes of the wastewater composition ('matrix') were not considered, besides the influence of dilution of the raw sample to yield the five dilution samples.

A detailed investigation of matrix effects would have required a much more complex (and subsequently expensive) experimental setup, which was not feasible within the scope of these investigations. Nevertheless, the chosen approach can be extended to include more sources of uncertainty if required within a specific context.

MATERIALS AND METHODS

A comprehensive study assessing the uncertainties of different methods for determination of COD has been carried out using laboratory standard solutions and raw sewage samples (Bertrand-Krajewski *et al.* 2007). Besides laboratory (DIN 38409–H43 1980) and field methods (small tube tests ISO 15705 2002), two different *in situ* spectrometers have been included in the investigations. In order to cover a broad COD concentration range, the raw sewage sample taken from a large municipal treatment plant was processed according to the scheme in Figure 1. The methods applied within the study are summarised in Table 1.

According to Figure 1, each dilution sample was analysed by both the COD_{DIN}–laboratory method and the small tube test field COD_{STT}–method. In order to investigate the impact of sub-sampling on the total uncertainty, the analyses were based on three sub-samples from each dilution sample. For each sub-sample multiple analyses were carried
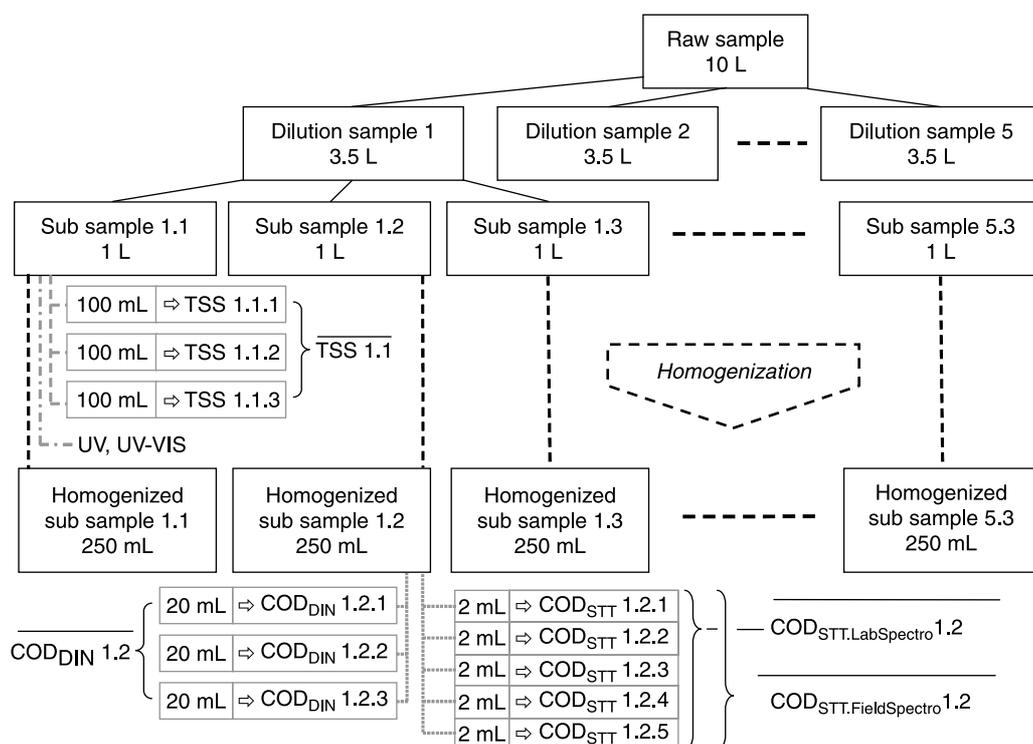

Figure 1 │ Sampling and measurement scheme.

Table 1 | Methods used for COD-determination

Method	Symbol	Norm/Type	Range	Spectral bandwidth	Light source	Optical pathlength
COD-lab method	COD_{DIN}	DIN 38409-H43	15,300 mg/L			
COD-field small tubes tests method	COD_{STT}	Hach-Lange LCI 400 Hach-Lange LCI 500	0–1,000 mg/L 0–150 mg/L			
UV-spectrometry		TRIOS ProPS CW	190–380 nm	0.75 nm	Deuterium lamp	10–60 mm
UV-VIS-spectrometry	COD_{EQ}	SCAN spectrolyser	200–750 nm	2.50 nm	Xenon flashlamp	5 mm

out. For the COD_{DIN}–method, three replicates of 20 mL were analysed for each sub-sample, yielding a total of nine COD_{DIN}–measurement values for each dilution sample (3 sub-samples × 3 replicates). For the COD_{STT}–method, five replicates of 2 mL in five small tubes were analysed for each sub-sample. In order to also include the influence of the photometric optical reading, each tube was processed five times with the photometer (i.e. five repeated readings for each replicate). This yields a total of 75 measurement values for each dilution sample with the COD_{STT}method (3 sub-samples × 5 replicates × 5 readings).

RESULTS AND DISCUSSION

COD-laboratory and-field method: variance analysis

A standard ANOVA (analysis of variance, e.g. Neuilly & CETAMA 1998) was carried out, considering both independent and interacting factors contributing to the total variance of the results: (i) the effect of diverting the dilution sample into three sub-samples (*sub-sampling*), (ii) the effect

of multiple analyses of a specific sub-sample (*replication*), (iii) the combined effect of sub-sampling and replication of these sub-samples (*sub-sampling × replication*) and (iv) the repeated readings with the photometer in the case of the COD_{STT}–method (*repeated readings*). For the COD_{DIN}–method a 1-factor ANOVA was carried out to evaluate the effect of sub-sampling, while for the COD_{STT}–method a 2-factors ANOVA allowed evaluating the contributions of both sub-sampling and replication.

For the five dilution samples, COD mean value, standard deviation and total variance results are given in Table 2. Absolute values of total variance vary greatly for each method and each dilution sample. Dilution sample n°2 shows a very low total variance for the COD_{DIN}–method compared to the other dilution samples, due to a better homogeneity of its three sub-samples. For both COD_{DIN} and COD_{STT}methods, dilution sample n°4 shows the highest variance, due to a very high heterogeneity of its sub-samples.

Table 2 highlights the utmost importance of a high quality set of reference data for later development of the

Table 2 | COD mean value, standard deviation and total variance for dilution samples 15

Dilution sample n°	Approx. dilution ratio	COD_{DIN} mean value (mg/L)	COD_{DIN} standard dev. (mg/L)	COD_{DIN} total variance (mg²/L²)	COD_{STT} mean value (mg/L)	COD_{STT} standard dev. (mg/L)	COD_{STT} total variance (mg²/L²)
1	1/32	69	11	113	59	4	15
2	1/16	117	6	34	105	9	81
3	1/8	249	19	345	248	18	334
4	1/4	573	75	5,676	656	95	8,939
5	1/2	883	28	793	953	37	1,396

Raw data, no outliers removed.

correlation model. Despite all laboratory procedures were carried out with great care, significant uncertainties in the results were observed. Building a large data base of high quality reference values, including repetitive analyses, is the only way to detect such uncertainties.

Figure 2 shows the relative contributions of i) sub-sampling, ii) replication, iii) combination of sub-sampling and replication, and iv) multiple readings to the total variance in a very synthetic way for both methods. For the COD_{DIN}–method, ANOVA-results indicate that, except for dilution sample n°2, sub-sampling is the main contribution to total variance, ranging from 51 to 97% of the total variance; the residual variance, corresponding to replication and random errors, is less significant. For dilution sample n°2, the three sub-samples were very homogeneous (see Table 2) resulting in a significantly lower total variance, compared to the other samples. Due to the comparably low total variance, the contribution of replication and random errors is significantly higher than for all other samples. For the COD_{STT}–method, the main contributions (55 to 91%) to the total variance are i) sub-sampling and ii) its interaction with replication; in comparison, the contribution of replication only is

smaller, ranging from 5 to 40% of the total variance. The residual variance, corresponding to repeated readings with the photometer and random errors, represents the smallest fraction (47%).

Summarizing the above, it can be stated that the COD_{DIN} and the COD_{STT}–method provide similar results in terms of absolute value as well as total uncertainty. The COD_{DIN}–method tended towards underestimation of concentrations above its upper detection limit (300 mg_{COD}/L); it is assumed that this is due to the introduction of the additional dilution step. The most important contribution to the total variance appears to be (sub-)sampling and its interaction with multiple analyses.

UV-spectrometry: precision analysis

In the study outlined above, a large number of raw sewage samples have been analysed inter alia by means of an UV-spectrometer. Two different ways of sample presentation were applied: (i) a 10 mm-quartz cuvette inserted into a special fitting at the UV-spectrometer ('*cuvettemethod*') and (ii) the UV-spectrometer entirely submersed in a 2 L-graduated cylinder placed on a magnetic stirrer ('*cylinder-method*'). Within this setting, each dilution sample (see Figure 1) as well as three sub-samples thereof have been analyzed. The UV-spectrometer provides the 'raw' measurement signal, i.e. the light counts of the UV-detector for each specific wavelength. The measurement mode was set up such that five single measurements were carried out per measurement cycle; the instrument combines these five single raw spectra and outputs one absorption spectrum per measurement cycle. Five measurement cycles have been carried out for each sample; from that an average raw spectrum and its uncertainty for each measurement cycle were derived.

Figure 3 shows the average raw UV-spectra (five repetitions) of two different sub-samples of the same dilution sample. The repetitions introduce another sub-sampling step, since the quartz-cuvette has been refilled prior each measurement. The repeatability of the sub-sequent spectra is good; some differences between the two groups of raw spectra can be detected due to inequalities of the two sub-samples introduced by diverting the dilution sample. Assuming a normal distribution of the average raw
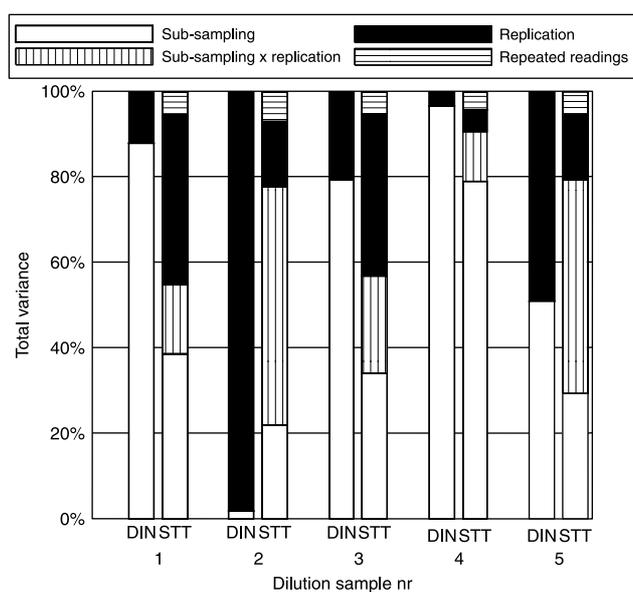

Figure 2 │ Synthetic results of the variance analysis for COD_{DIN} measurement (1 factor ANOVA for sub-sampling) and COD_{STT} measurements (2-factors ANOVA for sub-sampling and replication).
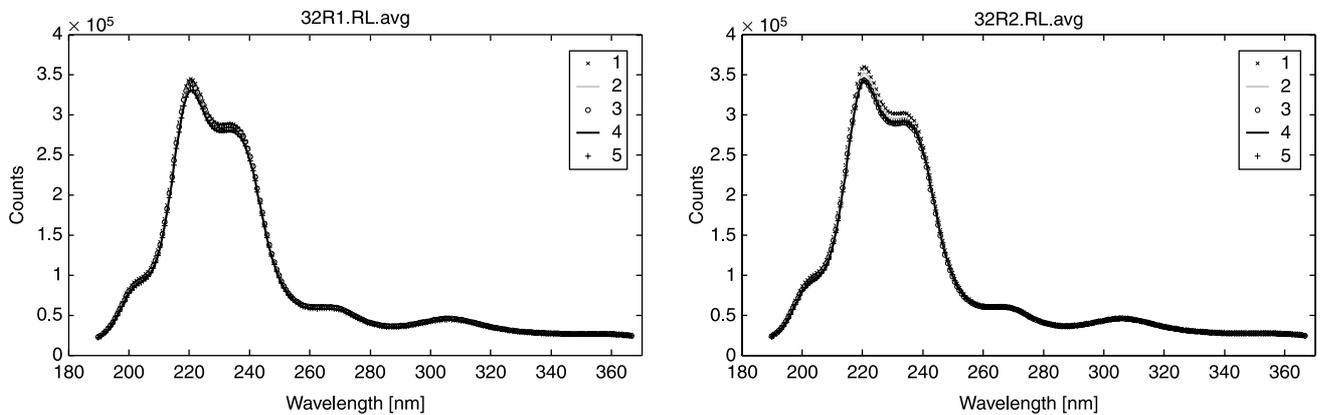
Figure 3 │ Average raw UV-spectra of two sub-samples of the same dilution sample.

UV-absorption spectra, the 95%confidence interval was calculated as twice the standard deviation. This resulted in a maximum width of the confidence interval of the average raw spectra of less than ±3% for 14 out of the 15 measurement cycles (3 sub-samples, 5 measurements per sub-sample) investigated here.

The one measurement cycle showing a lower precision was analysed in detail; it was found that one of the five single measurements of that cycle could be considered as an outlier (measurement 2, see Figure 4 left). Eliminating this specific measurement from that measurement cycle has a significant impact on the width of the 95%confidence interval of the average raw UV-spectrum of that measurement cycle (Figure 4 right); its maximum width is reduced from ±11% to ±2.4%.

Figure 5 shows a comparison of the mean raw UV-spectra and their respective 95%-confidence intervals

of the UV-measurements of the three sub-samples; for each sub-sample five measurement cycles have been carried out. The confidence intervals of the three mean UV-spectra overlap – the confidence interval of sub-sample 3 is confined within the upper limit of the confidence interval of sub-sample 2 and the lower limit of the confidence interval of sub-sample 1 – it was therefore depicted transparent with black patch edges. In a statistical sense, the three mean UV-spectra of the different sub-samples can be regarded as equivalent. The mean raw UV-spectra of sub-sample 1 and sub-sample 3 are almost identical, the mean raw UV-spectra of sub-sample 2 shows higher values – on average +2.5% relative to sub-sample 1. These results are well in accordance with the TSS-laboratory analysis of these sub-samples, which gave a TSS-concentration of $50 \pm 4\,\mathrm{mg_{TSS}/L}$, $54 \pm 4\,\mathrm{mg_{TSS}/L}$ and $52 \pm 4\,\mathrm{mg_{TSS}/L}$ for sub-samples 1–3, respectively.
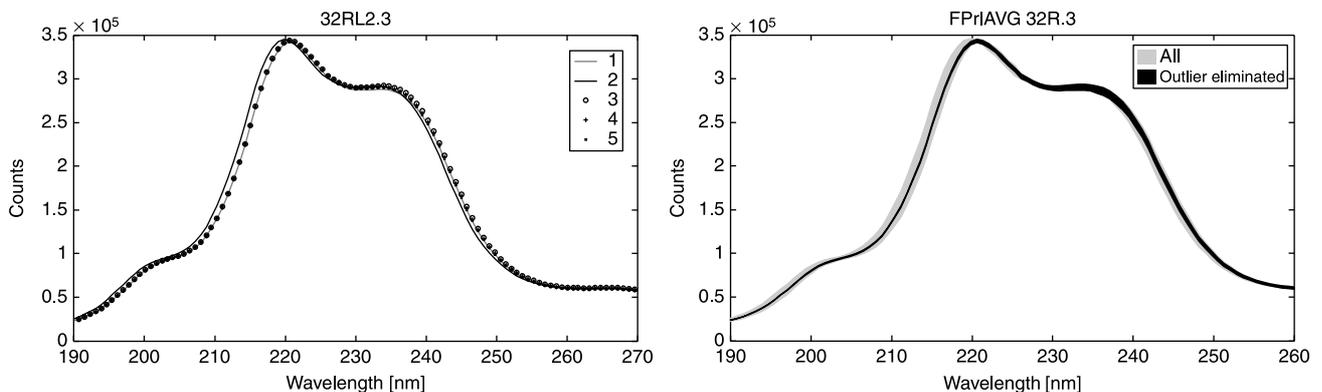

Figure 4 │ Left: Single raw spectra of one measurement cycle. Right: 95%-confidence interval of the average raw UV-spectrum of that measurement cycle without/with elimination of measurement 2.
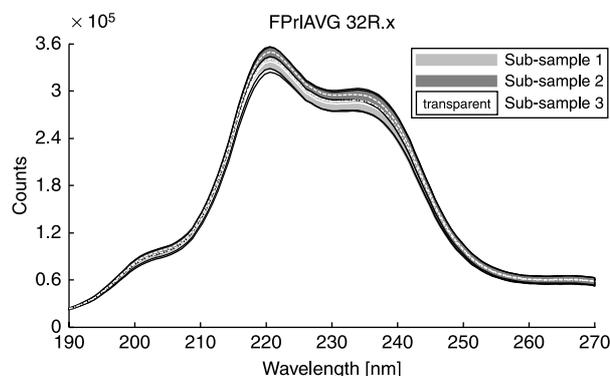
Figure 5 | Average raw UV-spectra (white lines) and respective 95%-confidence intervals of three sub-samples of the same dilution sample.

Due to the comparably large sample volume required for the 'cylinder'-method, only the dilution sample (see Figure 1) was used in these tests. In order to compare the two sample presentation methods, the average UV-absorption and its 95%-confidence interval was calculated for both methods. Nine measurement cycles have been carried out with the 'cylinder'-method. Each minute a measurement cycle was triggered automatically, while the sample was continuously stirred during the entire measurement period. Of the measurements with the 10 mm-quartz cuvette, 15 measurement cycles have been considered (five measurement cycles per sub-sample) to calculate the average UV-absorption and its 95%-confidence interval.

Figure 6 shows that the average UV-absorption derived from the 'cuvette-method' is lower than the one derived from the 'cylinder-method'. This maybe partly due inevitable settling within the cuvette or due to the formation of
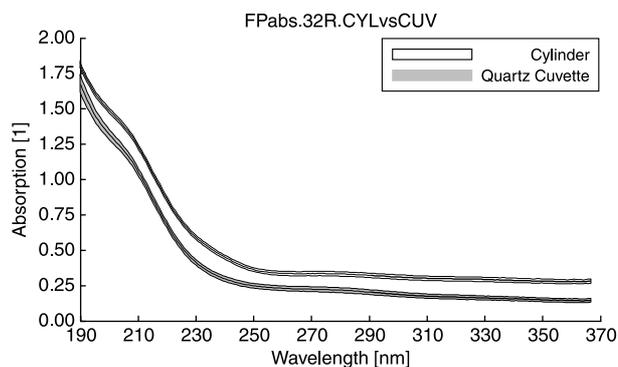

Figure 6 | Average UV-absorption and respective 95%-confidence interval of a dilution sample. Sample contained in a 2 L graduated cylinder and a 10 mm-quartz cuvette.

density layers within the cylinder despite continuous stirring.

The important aspect is that – obviously – the spectrometer is presented with different samples despite all sub-samples (and sub-sub-samples) were taken with great care from the same dilution sample.

The effect of sub-sampling can also seen by the fact that the width of the 95%-confidence interval of the 'cuvette'-method is larger ($\leq \pm 10\%$) than the one of the cylinder-method ($\leq \pm 5.5\%$). The 'cuvette'-method includes two sub-sampling steps, which is not the case with the 'cylinder'-method. Obviously, the sub-sampling introduces variance which is not compensated by the higher number of measurement repetitions.

Summarizing the above, it can be stated that the UV-spectrometer shows high precision given that sample composition continuity can be maintained. It was also shown that the mode of sample presentation has a significant impact on the measurement results. This corresponds to the selection of the probe installation location for a field application. As long as a suitable probe installation location is selected (completely mixed, no air bubbles) a high precision of the measured UV-absorption spectra can be expected. This provides a good basis for development of suitable correlation models for the required target parameters. On the other hand, it clearly highlights the necessity for developing a unique correlation model for each application – the two average UV-spectra in Figure 6 represent the same sample, but obviously two different correlation models would be required for deriving the same target parameter (COD = 63 $\pm$ 3 mg/L after removal of outliers, see Table 3).

UV-VIS-spectrometry: uncertainty of the target parameter

The above described tests were also carried out with the UV-VIS-spectrometer. From these tests a mean UV-VIS-spectrum and its 95%-confidence interval for each of the five dilution samples has been derived. Additionally, the mean COD_{DIN} concentration (after removal of outliers based on the Anova analysis, in order to obtain the best available estimation of the 'true' concentration for all dilution samples - details not shown in this paper) was used as a *reference* COD

Table 3 | Mean COD$_{DIN}$ values and COD$_{EQ}$ values and their respective 95% confidence intervals

		Dilution sample #				
		1	**2**	**3**	**4**	**5**
COD$_{DIN}$	[mg/L]	$63 \pm 3^*$	117 ± 4	$238 \pm 6^*$	$524 \pm 16^*$	883 ± 22
COD$_{EQ}$	[mg/L]	69 ± 10	111 ± 9	237 ± 14	525 ± 15	889 ± 44

*Outliers removed.

concentration for each dilution sample, with its 95% confidence interval. The corresponding PLS (Partial Least Squares) correlation model COD$_{EQ}$ = f(Abs(λ_i)) was established to estimate COD equivalent concentrations from both the spectra and the reference *COD* values (Torres & Bertrand-Krajewski 2008). Moreover, 1,000 Monte Carlo simulations were run in order to evaluate the uncertainties in this model and in the equivalent concentrations due to uncertainties in both spectra and reference COD values. The results are illustrated in Figure 7: in the upper left triangle, the 1,000 runs are illustrated, showing the dispersion of the results. In the lower right triangle, both mean COD$_{DIN}$ and COD$_{EQ}$ concentrations are represented with their 95% confidence intervals (see also data in Table 3).

Applying the PLS correlation model, the COD$_{DIN}$ and the COD$_{EQ}$ methods deliver almost identical mean concentration values for all five dilution samples. However, it has to be considered that in total only five data points were available to develop the correlation model and that all of them were included in the calibration data set. No additional data sets were available for validation. Since all dilutions were prepared from the same raw sample, also the variation of the wastewater matrix was relatively low.

CONCLUSIONS

In-situ spectrometers are comparable robust and low maintenance tools for continuous measurement of a number of target parameters such COD, NO$_3$—N or TSS. Due to the indirect measurement principle, the development of a suitable correlation model, describing the matrixspecific relationship between the measured absorption and the concentration of the target parameter is essential. High quality reference analysis is of utmost importance for building a calibration and a validation data set.

The uncertainty of the reference method (including sampling and sample preparation) has a direct impact on the uncertainty of the target parameter correlation model. An analysis of variance (ANOVA) showed that subsampling and its interaction with replication are the largest sources for the total uncertainty of the reference method. In order to minimize the uncertainty of the calibration data set this aspect has to be considered; this means that building a large reference data set based on multiple analyses is the only way to establish a solid base for later development of the target parameter correlation model.

An analysis of the raw absorption spectra showed that sample presentation has a significant impact on the measured absorption values. This clearly highlights the requirement for developing an installation location and medium specific target parameter correlation model. The precision of
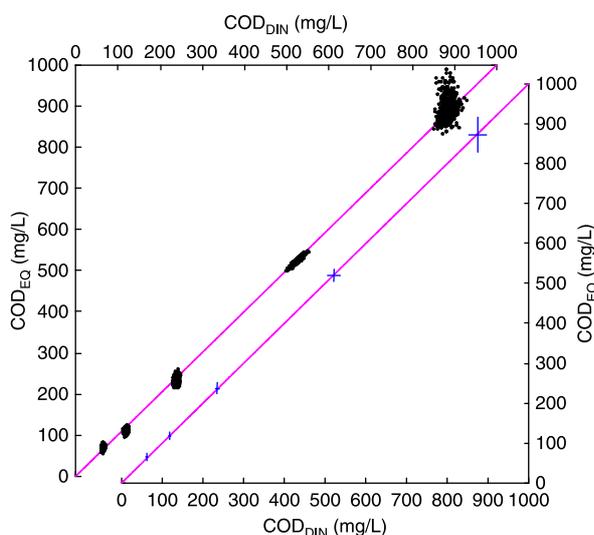

Figure 7 | Upper left: COD$_{EQ}$ vs. mean COD$_{DIN}$ for 1000 MC simulations considering the uncertainty in both the absorption spectra and the mean COD$_{DIN}$−reference concentration. Lower right: Results of the correlation model COD$_{EQ}$ = f (Abs(λ_i)) and mean COD$_{DIN}$ and their respective 95%-confidence intervals.

the tested spectrometers was good, independent of the sample presentation mode; the maximum width of the 95% confidence interval was typically smaller than $\pm 2.5\%$ of the measured absorption.

Applying a suitable PLS correlation model between mean absorption spectra and mean COD_{DIN} values (without outliers) used as *reference* COD values, the average concentration of the COD_{EQ} values given by the model were almost identical to the mean COD_{DIN} values.

While the uncertainty might be higher under usual field conditions, the results of the experiments show that in situ spectrometry can, for example, be a suitable method for accurate estimation of average (daily or hourly) loads utilising a possible high measurement frequency (typically up to $1 \, min^{-1}$).

ACKNOWLEDGEMENTS

REFERENCES

Bertrand-Krajewski, J.-L., Winkler, S., Saracevic, E., Torres, A. & Schaar, H. 2007 Comparison of and uncertainties in raw sewage COD measurements by laboratory techniques and field UV-visible spectrometry. *Water Sci. Technol.* **56**(11), 17–25.

DIN 38409-H43 1980 *Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung - Gruppe H: Bestimmung des chemischen Sauerstoffbedarfs (CSB) – Kurzzeitverfahren (H 43).* Berlin (Deutschland): Deutsches Institut für Normung. [in German].

ISO 15705 2002 *Water quality - Determination of the chemical oxygen demand index (ST-COD) - Small-scale sealed-tube method.* Geneva (Switzerland): ISO, p. 18.

Neuilly, M. & CETAMA 1998 *Modélisation et estimation des erreurs de mesure.* Paris (France): Technique et Documentation. 704 p ISBN 2-7430-0272-7.

Torres, A. & Bertrand-Krajewski, J.-L. 2008 Partial Least Squares local calibration of a UV-visible spectrometer used for *in situ* measurements of COD and TSS concentrations in urban drainage systems. *Water Sci. Technol.* **57**(4), 581–588.