Development of a methodology for water monitoring in on-site analytics

Stefanie Penzel[1], Thomas Mayer[2], Mathias Rudolph[1]

1 Leipzig University of Applied Sciences, Karl-Liebknechtstraße 132, 04277 Leipzig
2 Helmholtz Centre for Environmental Research - UFZ, Permoserstraße 15, 04318 Leipzig

Abstract

The increasing presence of harmful substances in waters and the associated health and economic impacts require the development of methods for rapid and accurate detection of these hazardous substances. The determination of sufficient parameters to describe the water quality and the subsequent characterization represent a major challenge from both a practical and scientific point of view. Up to now, such water tests are usually carried out in the laboratory after sampling. In the context of the studies presented here, it is shown how a methodology for water measurements can be carried out directly on site. For this purpose, realistic spectra were recorded on a UV/VIS fluorescence measurement set-up developed at the Helmholtz Centre for Environmental Research Leipzig for on-site analysis, and an evaluation method based on fuzzy logic was developed that considers the measurement uncertainties that arise.

1. Introduction

Water monitoring involves testing and analyzing water quality for various contaminants and pollutants to ensure that the water is in safe condition. Some of the key parameters that are typically measured in water monitoring are Chlorophyll a and b, pH, dissolved oxygen, temperature and many more. These substances can be analyzed by a variety of standardized analytical methods, such as MPN (Most Probable Number) methods (ISO 7899-1:1998, n.d.) or HPLC (High-performance Liquid Chromatography) methods (ISO/IEC 17025:2017, 2018). The work presented here focused on the possible contamination by algae.

However, many methods have their limitations, especially when studying very low concentrations in water. In addition, these are traditionally carried out in the laboratory, usually after water sampling at various places and at different times. These procedures are therefore no longer considered efficient in many cases (Lambrou et al., 2012; Lambrou et al., 2014; Zhuiykov, 2012). To detect and analyze the generation of pollutants directly at the source, an on-site sensor system is required. In the environmental field, the corresponding on-site analytics make an excellent contribution to the analysis of dynamic and heterogeneous systems as well as to the characterization of running processes.

Currently, a submersible probe for water monitoring is being developed at the Helmholtz Centre for Environmental Research (UFZ) Leipzig. With this probe a simultaneous recording of transmission and fluorescence spectra of the water samples can be done. Despite the many advantages of on-site analysis, there is a central disadvantage compared to analyzes carried out in the laboratory due to higher measurement uncertainty, caused by the ambient conditions such as varying temperature. In the context of the studies presented here, additionally to the recording and analysis of realistic water data, an evaluation methodology with integration of measurement uncertainties is addressed. The experimental findings are used to develop a mathematical model to capture and describe the uncertainties in the measurements.

For this purpose, the methods of fuzzy pattern classification were chosen, which are based on the theory of Zadeh (Zadeh, 1965). Fuzzy pattern classification represents a highly flexible approach to provide a fuzzy and thus realistic characterization of current water conditions. Fuzzy pattern classification already has several applications, e.g., in automation and signal processing systems (Lohweg et al., 2004; Mönks et al., 2009) or in medical applications (Bocklisch et al., 2017; Bocklisch & Hausmann, 2018). As a result, after feature extraction into this classifier, unknown data can be evaluated, and classes can be assigned to the presence of a substance.

2. Materials and Methods

2.1 Experimental Setup

An experimental setup designed at the Helmholtz Centre for Environmental Research - UFZ was used to record the data. With this setup, transmission and fluorescence spectra of water samples can be recorded. The setup consists of a UV/VIS light source equipped with a deuterium and a tungsten lamp for transmission measurements and an LED array for the excitation of the fluorescence, which is directly attached to the measuring cell at an angle of about 90°. The UV/VIS light source is connected to the measuring cell by optical fibers. The basic body of the measuring cell is made of stainless steel. Two opposing 90° collimators and the LED array are attached to it. The UV/Vis light source and the spectrometer are connected to the collimators by fiber optic cables. This is a Broadcom QMINI Wide UV with a resolution of 1.5 nm FWHM. The control and recording of the data is implemented in each case with a software. The schematic diagram of the experimental setup is shown in figure 1a. For the recordings on site, the experimental setup is integrated into a submersible probe (see figure 1b).

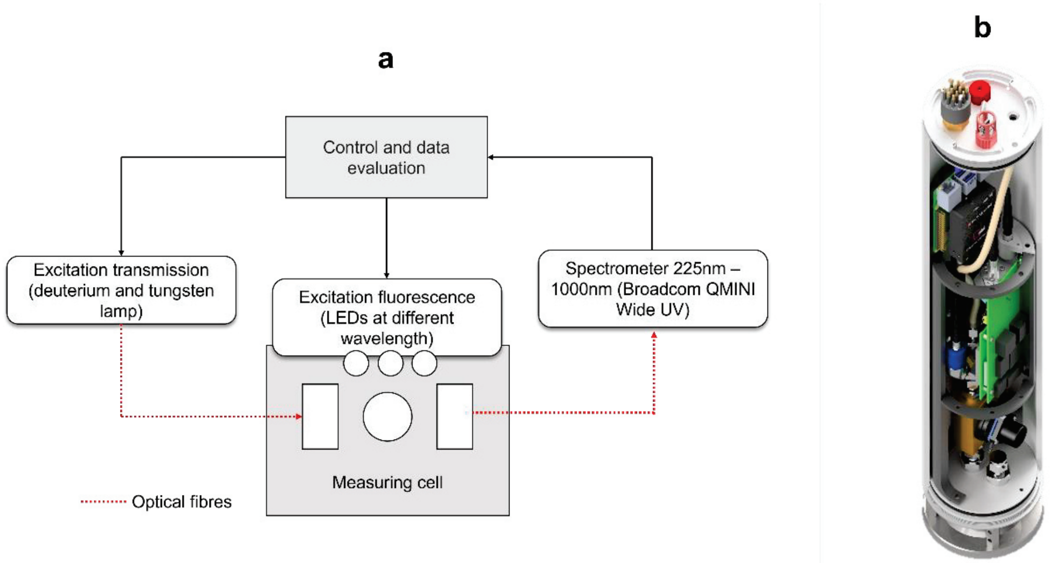

Figure 1: a) Schematic representation of the experimental setup and b) submersible probe for data acquisition on site

The transmission measurements are generated by UV/VIS spectroscopy, which bases on the detection of changes in the electromagnetic spectrum to generate information about the water sample in question. During transmission, electron transitions occur between different states in the molecule and, in the case of pollutants in water, for example, part of the radiation is absorbed by the pollutants. This can be seen in the absorption/extinction spectrum by comparing the measurement with a blind measurement. Based on the absorbed wavelength and the degree of absorption, the type and concentration level of the substance can be inferred. Fluorescence is the spontaneous emission of light after a substance or material is excited with electromagnetic radiation. Due to vibrational relaxation, some of this energy is re-emitted. When the molecule returns to its ground state, it fluoresces at a wavelength different from that it was illuminated. This process does not always occur because not every substance fluoresces. Usually, fluorescence is measured at 90° to the excitation source. This way the stray light can be minimized. (Hinderer, 2020)

The detection of algae was selected as the application. For realistic data, a reference substance is produced in the laboratory using spinach extract and water in accordance with DIN EN 17899 Annex C.3. Algae are identified by chlorophyll A and turbidity, for example. These are also contained in the spinach extract.

2.2 Methodical Approach

Several significant features can be generated from the spectra of spinach extract with water. For this purpose, the spectra are mathematically described with a calculated function according to the method of least squares. The descriptive function results from the addition of several Gaussian functions $g_{total,i}$, which was calculated for a certain concentration.

$$g_{total,i} = a1 \cdot e^{-\frac{(x-b1)^2}{2 \cdot c1^2}} + a2 \cdot e^{-\frac{(x-b2)^2}{2 \cdot c2^2}} + \ \dots \ + ai \cdot e^{-\frac{(x-bi)^2}{2 \cdot ci^2}} \tag{1}$$

Figure 2 shows this as an example for an extinction spectrum at 40 % dilution series value.

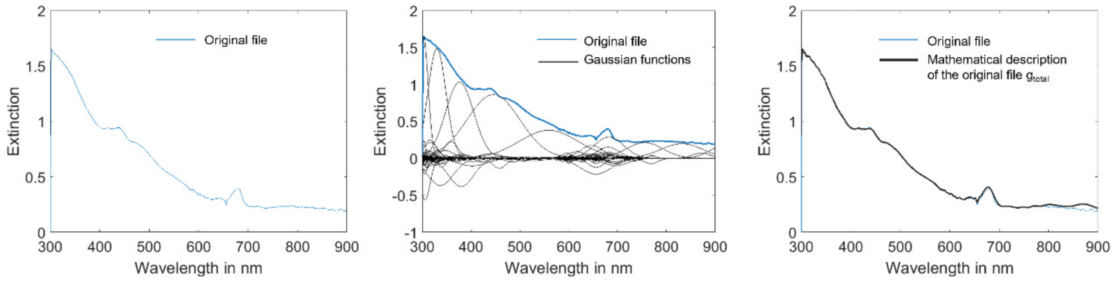

Figure 2: Mathematical description of an extinction spectrum at 40 % diluted solution series value

Subsequently, this mathematical description is fitted over all spectra with different concentrations using the least squares method. This results in two features. One is the specification of the R-square, which stands for the presence of a substance. And on the other hand, the specification of a z-parameter, which represents the shift of the function upwards or downwards and thus generates a statement about the level of the substance present.

With these features, the fuzzy classification can be performed. A distinction is made between a learning phase and a working phase. In the learning phase, a fuzzy classification model is built in a multidimensional feature space. In the first step, the object data sets of the features are divided into sharp groups.

The method used here is a cluster analysis with a Ward method, in which the cluster centroid is calculated from the arithmetic mean of the forming objects and then the squared Euclidean distances of the objects to the centroids are formed. Subsequently the distances are summed and the clusters that have the smallest increase in total are linked (Backhaus et al., 2018). Thus, groups with similar objects (in the sense of a small distance measure) are obtained. The classification can also be done via expert-based approach.

In the following, these sharp groups are transformed into fuzzy groups. For this purpose, a special methodology of fuzzy classification, the fuzzy pattern classification according to S. F. Bocklisch (1987b), is used. The description of each group in the one- or multi-dimensional feature space is done by the AIZERMAN membership function, which is a highly flexible parametric potential function type. The simplified one-dimensional form can be represented as follows:

$$\mu(\mathrm{u}) = \frac{a}{1+\left(\frac{1}{b}-1\right)\cdot\left(\frac{|u-u_0|}{c}\right)^d} \tag{2}$$

The parameters are automatically calculated from the recorded learning data by supervised learning, with the classifier structure optionally unrotated or rotationally adjusted. For the multivariate case, the feature dimension of the membership function can be extended accordingly and thus each group can also be described by an analytically closed membership function. In the working phase, unknown objects are read into the fuzzy pattern classifier resulting from the learning phase and membership values to the respective classes are calculated. The assignment of the objects to the respective class is done by the highest membership value.

3. Results and Discussion

The evaluation methodology will be shown in this article using the transmission/extinction spectra measurements as an example. The procedure for the fluorescence spectra can be done analogously. For the learning data set, spectra were recorded at 15 dilution series values each between 0.5 % and 50 % and converted to extinction spectra. In table 1 the information about the recorded data and in figure 3 the extinction spectra are shown. For these spectra, the mathematical description was used (see Equation 1 with calculated parameters for the extinction spectrum at 40 % diluted solution series value) to calculate an R-squared and a z-parameter as a feature for each spectrum.

Measuring method	Solutions	Repeats
Transmission / Extinction	(0.5, 1, 2.5, 5, 7.5, 10, 15, 20, 22.5, 25, 30, 35, 40, 45, 50) %	6

Table 1: Information about the recorded water spectra with the algae-like substance

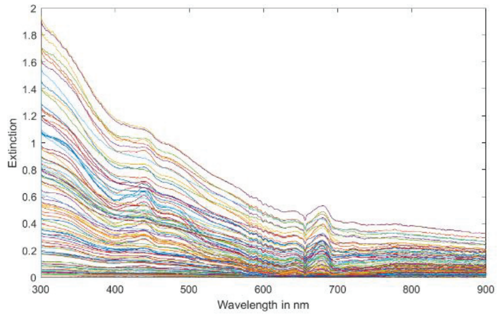

Figure 3: Illustration of all extinction spectra at different dilution series values

In the next step of the learning phase, the features are grouped into sharp classes performing a cluster analysis using the procedure described in Section 2.2. Here, class 1 represents a low presence of the algae-like substance, class 2 algae-like substance is present, and class 3 a high amount of the algae-like substance is present. The transition to the fuzzy classes is done using the above calculations according to Bocklisch et al. (2017) or Penzel et al. (2022) in the three-dimensional feature space. Figure 4a represents the clustered groups and in 4b the resulting fuzzy class are shown.

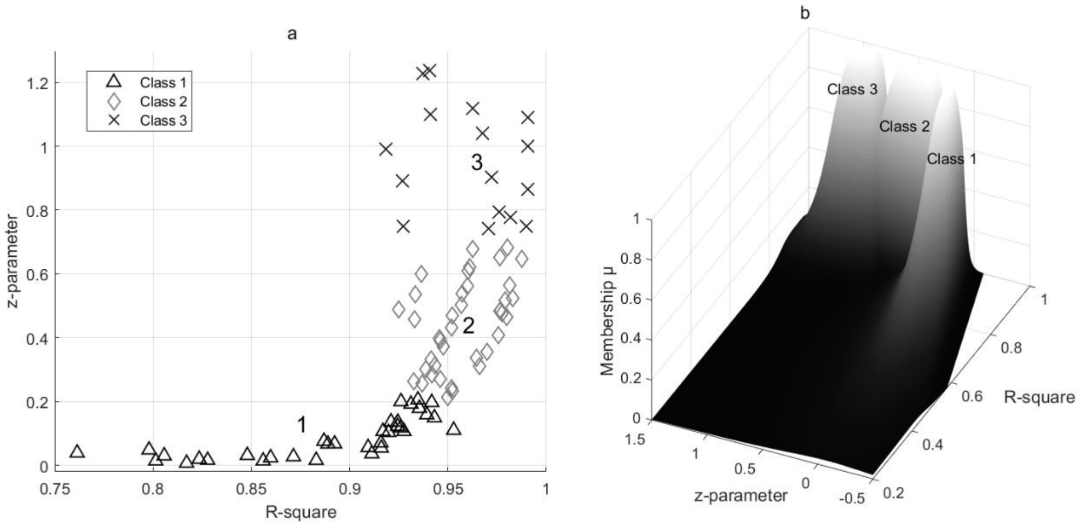

Figure 4: a) Sharp grouping of the characteristics by means of cluster analysis and b) Result for the rotationally matched fuzzy pattern classifier

The classifier structure was rotationally adjusted, as this reproduces the learning data more advantageously than the unrotated version. The correspondingly used parameters for the functions are listed in table 2. Figure 4 shows that at a lower z-parameter (representing the concentration level), the R-squared is also low. This can be attributed to the fact that lower concentrations of the substance are more difficult to detect, which is reflected in the R-squared. Therefore, this class is described with a larger fuzziness than the other two. Also, overlapping of the classes is evident, as each class presents the presence of the algal-like substance and differs only in quantity.

number	u_0	angle	b_L	b_R	c_L	c_R	d_L	d_R
Class 1								
1	0.883	1.000	0.248	0.484	0.200	0.131	2.273	3.231
2	0.085	0.783	0.672	0.279	0.063	0.124	6.538	2.120
Class 2								
1	0.957	1.000	0.531	0.703	0.253	0.265	10.230	3.807
2	0.443	1.512	0.398	0.453	0.028	0.042	9.599	7.030
Class 3								
1	0.961	1.000	0.086	0.240	0.238	0.306	8.036	3.275
2	0.955	1.615	0.320	0.289	0.042	0.050	11.412	3.613

Table 2: Parameters for the parametric membership function according to AIZERMAN for the creation of the fuzzy pattern classifier

For the working phase, spectra of spinach extract in water in different dilutions were again recorded and a working data set was generated by extracting the features R-squared and z-parameter. The objects of the working dataset were read into the fuzzy pattern classifier to calculate membership values of each class. Figure 5 represents the fuzzy pattern classifier from the top view. In this figure, the objects are marked by the red crosses and the assignment to the classes – displayed in parentheses – with the respective membership value is already shown.

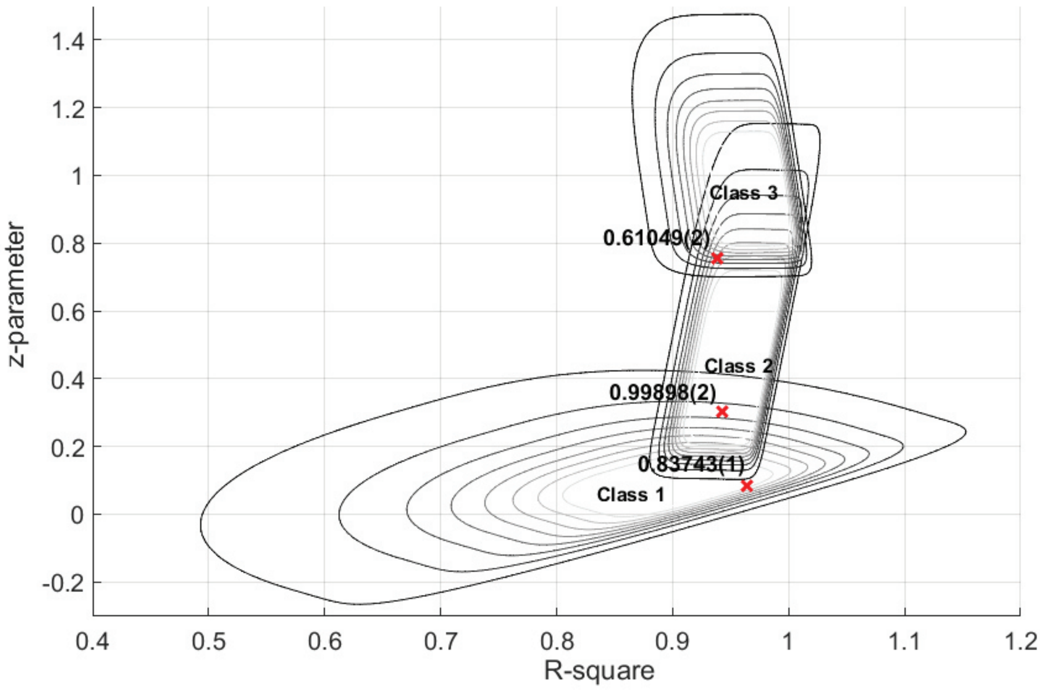

Figure 5: Graphical representation of the Membership of the working data

All objects of the working data set can be assigned to classes. The assignment to class 1 with the membership value of 0.83743 reflects the statement that an algae-like substance is present to a very small extent. With the membership value of 0.99898, a clear assignment to class 2 is evident, so that the presence of an algae-like substance can be assumed. The last object has a comparatively lower membership value to class 2.

Here the tendency goes rather to class 3, which predicts a high quantity. In this case, recommendations for action should be given for pollutant treatment.

Conclusion and outlook

Water monitoring is essential for protecting public health, maintaining healthy ecosystems, and ensuring sustainable water use for agriculture and industry. The investigations presented here focused primarily, in addition to the recording of spectra by means of light sensors, on the generation of an evaluation methodology for on-site analysis. For this purpose, after a pre-processing of the water spectra, a fuzzy pattern classifier was generated. This considers the uncertainties in the detection and evaluation of the water data and represents a much more flexible and realistic modeling by allowing the implementation of fuzzy information.

In follow-up work, more data will be included to optimize the evaluation methodology on the one hand, and data will also be recorded on site with the submersible probe on the other hand. The methodology can be adapted to specific applications since the generation of features by mathematical description and the following transfer to the fuzzy pattern classifier can also be done for other spectra and thus also for other substances. Overall, it is still necessary to determine the exact limit values of the concentrations for the respective classes in order to be able to generate more precise recommendations for action.

Acknowledgement

Freistaat Sachsen

Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

References

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2018). Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Springer Gabler.

Bocklisch, F., Bocklisch, S. F., Beggiato, M. & Krems, J. F. (2017). Adaptive fuzzy pattern classification for the online detection of driver lane change intention. Neurocomputing, 262, 148–158. https://doi.org/10.1016/j.neucom.2017.02.089

Bocklisch, F. & Hausmann, D. (2018). Multidimensional fuzzy pattern classifier sequences for medical diagnostic reasoning. Applied Soft Computing, 66, 297–310. https://doi.org/10.1016/j.asoc.2018.02.041

Bocklisch, S. F. (1987b). Prozessanalyse mit unscharfen Verfahren.

Hinderer, F. (2020). UV/Vis-Absorptions- und Fluoreszenz-Spektroskopie. Essentials. https://doi.org/10.1007/978-3-658-25441-4

ISO 7899-1:1998. (o. D.). ISO. https://www.iso.org/standard/14852.html

ISO/IEC 17025:2017. (2018, 1. März). ISO. https://www.iso.org/standard/66912.html

Lambrou, T. P., Anastasiou, C. C. & Polycarpou, M. M. (2014). A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking Water Distribution Systems. IEEE Sensors Journal, 14(8), 2765–2772. https://doi.org/10.1109/jsen.2014.2316414

Lambrou, T. P., Polycarpou, M. M. & Anastasiou, C. C. (2012). A low-cost system for real time monitoring and assessment of potable water quality at consumer sites. IEEE-Sensors. https://doi.org/10.1109/icsens.2012.6411190

Lohweg, V., Diederichs, C. & Müller, R. D. (2004). Algorithms for Hardware-Based Pattern Recognition. EURASIP Journal on Advances in Signal Processing, 2004(12). https://doi.org/10.1155/s1110865704404247

Mönks, U., Lohweg, V. & Larsen, H. L. (2009, 15. September). Aggregation Operator Based Fuzzy Pattern Classifier Design. https://www.researchgate.net/publication/229035282_Aggregation_Operator_Based_Fuzzy_Pattern_Classifier_Design

Penzel, S., Rudolph, M., Borsdorf, H. & Kanoun, O. (2022). Prototypical investigation of the use of fuzzy measurement data in a case study in water analysis. Computer Science and Information Systems (FedCSIS), 2019 Federated Conference on. https://doi.org/10.15439/2022f125

Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338–353. https://doi.org/10.1016/s0019-9958(65)90241-x

Zhuiykov, S. (2012). Solid-state sensors monitoring parameters of water quality for the next generation of wireless sensor networks. Sensors and Actuators B-chemical, 161(1), 1–20. https://doi.org/10.1016/j.snb.2011.10.078