Position Papers of the of the 17th Conference on Computer Science and Intelligence Systems pp. 27–33
DOI: 10.15439/2022F125
ISSN 2300-5963 ACSIS, Vol. 31

Prototypical investigation of the use of fuzzy measurement data in a case study in water analysis

Stefanie Penzel, Mathias Rudolph
Leipzig University of Applied Science
Karl-Liebknecht-Str. 132
04277 Leipzig, Germany
Email: {stefanie.penzel, mathias.rudolph}@htwk-leipzig.de

Helko Borsdorf
Helmholtz Centre for Environmental Research - UFZ, Department of Monitoring and Exploration Technologies, Permoserstr. 15 Leipzig, Germany

Olfa Kanoun
University of Technology Chemnitz, Professorship Measurement and Sensor Technology, Reichenhainer Str. 9 Chemnitz, Germany

***Abstract*—A common problem when using real data is the fact that the values usually exhibit some degree of uncertainty. Measurement uncertainties therefore represent a major challenge when trying to interpret and draw conclusions from real data. This is especially true in on-site analysis in the environmental sector where the uncertainty in sample plays such a large role. An approach for the modelling and analyze of data for polluted water and the inclusion of measurement uncertainties is presented. This approach is based on fuzzy modelling, in which the uncertainty of the parameters is represented by so-called fuzzy numbers and thus reflect a possible blurred range of these parameter values. The result is a fuzzy pattern classifier, which allows a fuzzy and thus realistic characterization of unknown water samples. The procedure is exemplified using the extinction spectra taken using a UV/Vis spectrometer.**

I. Introduction

THE conservation of water resources and the need to continuously monitor the quality of these water resources (e.g., in watercourses, wastewater, bathing lakes, etc.) is of increasing importance nowadays. The determination of sufficient characteristic values to describe the water quality and the subsequent characterization represent a significant challenge. Various parameters play an important role in this. Polycyclic aromatic hydrocarbons (PAHs) such as benzene or naphthalene are a priority substance in water policy. In addition, other significant indicators may be relevant for the determination of pollutants. [1] These substances can be determined using a variety of standardized analytical methods. However, many methods have limitations, particularly when investigating very low concentrations in water. In addition, these measurement methods are traditionally performed in the laboratory mostly after water sampling at different locations at different times. These approaches are no longer considered efficient [2–4]. To detect and analyze the formation of pollutants directly at the source, an on-site sensor system is required. Continuous and unbiased measurements of this type can then be used for the optimal control and verification of water quality. For this reason, the Helmholtz Centre for Environmental Research (UFZ) is working on an experimental setup that can analyze water samples directly on-site using ultraviolet (UV) / visible (Vis) spectroscopy. Each sensor-based measurement has an objective uncertainty, which essentially depends on the measuring method and instrument. For example, for measuring instruments this uncertainty can be specified by an accuracy or error class according to DIN 1319-2, DIN 1319-3. Unstable operating conditions which occur especially in the environmental sector, and here with mobile on-site analysis, lead to additional uncertainties. For instance, seasonal changes in temperature and humidity may also contribute to the uncertainty of a measurement.

One of the advantages of using fuzzy classification methods is that such uncertainties can be characterized. The assignment to a pollutant substance is not crisply defined but is categorized according to a grade of membership. These are in the range between zero and one. The underlying fuzzy is based on a theory published by Zadeh in 1965 [5] and since then, it has been used and further developed in many areas, current such as the selection process for outsourcing users [6] or for the description of transportation problems through the extension of fuzzy sets [7]. The basic idea is to extend the classic binary classification, in this case the pollutant is present or not, to allow a gradual change. In our case, this allows the model to output that the pollutant may be present and further analysis is necessary. This is shown by the membership function to a fuzzy set. Such fuzzy forms of description, in which the crisp values are included as special cases, represent a new optimized meaning in the characterization of water quality. They have the advantages of greater flexibility and proximity to reality compared to the crisp forms of description, and moreover allow the adequate implementation of expert knowledge.

The procedure is demonstrated using the measurement data of water samples recorded by a UV/Vis test setup. The characteristic properties of different water samples (here measured in the form of extinction spectra) with different substance concentrations are to be derived from the data in a

so-called learning phase and to be used to model a fuzzy pattern classifier. With this classifier, a statement is to be made as to whether a certain substance is present with uncertainty, present or present with a certain degree of certainty. If a substance is not present, the membership of these three classes will be zero. This serves as the basis for a fuzzy and, thus, realistic characterization of current (unknown) water samples in the next working phase. Depending on the results, recommendations for action can be made afterwards.

II. Methodical Basics

A. The UV/Vis measuring setup

The measuring device used is a UV/Vis measuring setup. UV/Vis spectroscopy uses electromagnetic radiation to detect substances in water. 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 to a blank measurement. By means of the absorbed wavelength and the level of absorption, the type and concentration level of the substance can be concluded. The experimental setup was realized at the UFZ Leipzig and will be integrated into a mobile submersible probe for future on-site data acquisition. The setup consists of a UV/Vis light source with a deuterium and a tungsten lamp, a measuring cell, and a spectrometer. The connection of the three components is used via optical fibers. The measuring cell consists of two collimator adapters with optical windows, a stainless-steel flow cell and two 90° collimators each. The cuvette containing the dissolved sample substance can be inserted into the flow cell. The control and data acquisition of the measurement setup are carried out on a laptop using Python software, which automatically compares the measured data with a prerecorded blank spectrum to create an extinction spectrum from the two transmission spectra according to the Beer-Lambert law. Measurement uncertainties already arise during the testing in the laboratory, e. g. from the lamp due to fluctuations in energy supply or due to the noise caused by the spectrometer. For later mobile use, the results can also be affected by e. g. temperature or humidity. Which would be reflected in the noise behavior of the spectra, or it can lead to a rise in the baseline.

B. Fundamentals of the Fuzzy Pattern Classification

The fuzzy system used here is based on the fuzzy pattern classifier introduced by Bocklisch [8]. This methodology is widely used in pattern recognition for object classification. Here, a set of fuzzy membership functions $\mu: x \rightarrow [0, 1]$ are created per class, which model characteristic features of this class. Through the membership functions, the feature values $x \in IR$ of an object can be mapped to the unit interval, which represents the membership to a feature of an ideal class member. All the memberships are then merged and classified into the appropriate classes. Then the object is assigned to the class that has the highest aggregated value. This procedure is already used in many areas such as in signal processing applications and automation systems [9, 10] or in the field of neuronal statements and medical diagnostic reasoning [11, 12]. In addition, this fuzzy modeling is also used for data-inherent structures [13] or for online recognition of fuzzy time series patterns [14].

The exact procedure is divided into a ***learning phase*** and a ***working phase***. In the learning phase, a fuzzy classification model is constructed in a multidimensional feature space. This can be achieved by choosing between a data-driven or expert-based approach. In the data-driven procedure, several measurement runs are performed for predefined prototypical dilution series. First, the recorded object data sets (learning data) are divided into crisp groups. Two strategies are possible:

1) A cluster analysis (e.g., hierarchically agglomerative) is performed for the object data. This is a mathematical method, which creates corresponding groups through the accumulation of certain similar objects (in the sense of a small distance measure) as a result.

2) An a priori division of the objects into groups based on expert knowledge is carried out. This can be done by dividing the objects, here e.g., dividing the extinction maxima according to before or above the detection limit.

Subsequently, these crisp groups are transferred to fuzzy groups. The description of each group in the one- or multi-dimensional feature space is achieved here by a highly flexible, parametric membership function of the AIZERMAN potential function type. This function is described and illustrated in simplified form for the symmetrical one-dimensional case (Fig. 1).

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

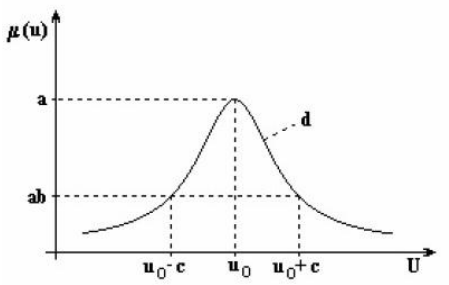

Fig. 1 One-dimensional AIZERMAN potential function [4]

Meaning of the parameters (see also Fig. 1):

- Local information u0 (crisp): representative of the fuzzy quantity (special case)

- Broadening c (to the left and right side of u0): precisely observed range
- Border membership b ϵ [0,1]: determines the membership values at the borders of the strictly area
- Maximum value of membership a (usually normalized with a = 1)
- d describes the continuously decreasing course of the membership function (d → ∞: crisp (binary) description)

The potential function can be used to describe both one-sided open intervals and closed intervals in a fuzzy way. In addition, the differentiation of the left and right-sided branch increases the adaptability. In the multivariate case, the characteristic dimension of the membership function is expanded accordingly, whereby each group can be represented by an analytically closed membership function. By using a closed analytical membership function to describe each group in the one- or multi-dimensional feature space, the method used here also differs from the rule-based fuzzy logic [15]. In a data-driven procedure, the parameters are calculated automatically from the recorded (learning) data sets by means of supervised learning [8].

The abovenamed AIZERMAN potential function may be applied to each axis of a multidimensional space. Thus, even information about high-dimensional groups can be described efficiently by a few parameters. A further advantage of the AIZERMAN potential function approach is that trapezoidal and triangular attribution functions as well as the so-called fuzzy singletons (crisp description as a special case of fuzzy case) can be converted into such a uniform description form, thus enabling a highly flexible and universal application with the possibility of modelling.

As an alternative to this data-driven approach, the parameters can be determined by expert knowledge, i. e for each of the characteristics fuzzy areas are defined manually and the fuzzy groups are then formed. This approach is typically used for linguisdetertic characteristics. Their values are not exactly defined, but colloquially defined by certain expressions (e. g. “small”, “medium”, “large”).

In the working phase, the classification model (Fuzzy Pattern Classifier) created in the learning phase is used for fuzzy identification of the current water sample (rep-resented by corresponding working data). The result is an membership or sympathy vector, whose components indicate the memberships to all declared classes. The current water status can be determined in a precise way from the maximum attribution values. The security (or risk) of this decision can be determined by the differences in the membership values.

III. Research Results and Discussion

A. Structure of the data base

Several dilution series with different concentrations of benzene, naphthalene, uranine and rhodamine B were prepared for the compilation of different data sets. By means of the measurement setup, extinction spectra were recorded for each substance at different concentrations (see Fig. 2).

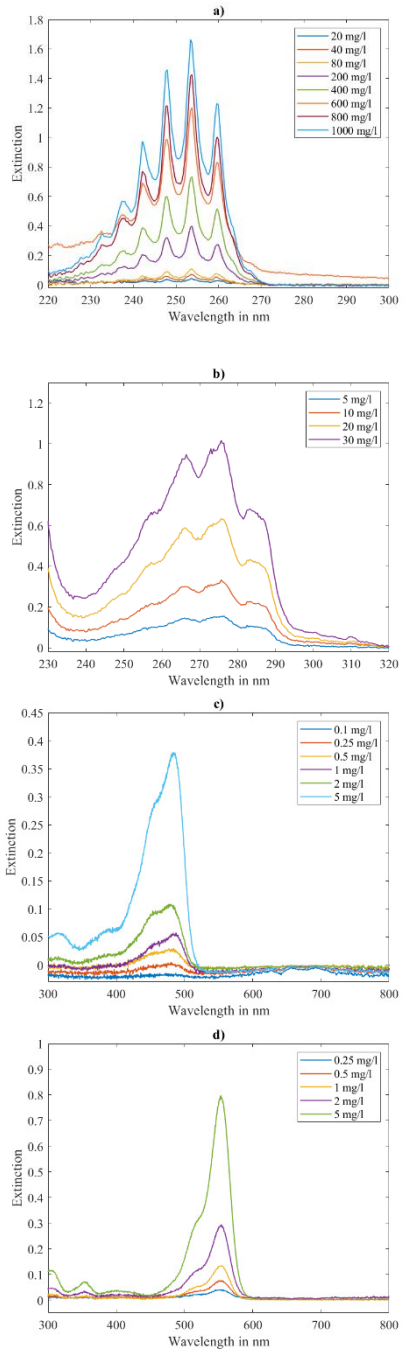

Fig. 2 Concentration-dependent extinction spectra of (a) benzene, (b) naphthalene, (c) uranine and (d) rhodamine B

Each substance was subjected to several measurement runs. Table 1 presents the selected dilution samples or

their respective datasets and explains the abbreviations contained therein. The classification into this groups is based on the limit of detection or limit of quantification of the recorded data and is used for the subsequent classification into sharp classes.

TABLE I.
WATER SAMPLES USED WITH ABBREVIATED TITLE

Short description	Description of the samples with classification into the different existing classes
BUP	Benzene is unsurely present
BP	Benzene is present
BSP	Benzene is surely present
NUP	Naphthalene is unsurely present
NP	Naphthalene is present
NSP	Naphthalene is surely present
UUP	Uranine is unsurely present
UP	Uranine is present
USP	Uranine is surely present
RUP	Rhodamine B is unsurely present
RP	Rhodamine B is present
RSP	Rhodamine B is surely present

The extinction spectra of each compound were first described mathematically with an algorithm. For the mathematical description, several Gaussian functions were added to a total function and the parameters were each adapted to a spectrum of a substance. This characteristic overall function is then overlayed on all spectra and fitted to the spectra using the method of least squares. The R-squared is calculated. As the overall function is characteristic for each substance, the R-squared basically indicates the probability with which a certain substance is present. The extinction maximum gives a statement about the concentration content of the substance. These data were then stored in an overall dataset. Subsequently, the data sets were selected on a random basis and then divided into so-called learning and work data (see Section 2.1). The R-squared and extinction maxima were stored in an object file (.OTX). Fig. 3 shows a section of the created object file. This consists of a header with the necessary information about the data and then lists the object number, the corresponding class (here the assignment to the respective water sample) and the measured values for each of the two characteristics.

```
Objects: Learningdata_wateranalysis
2 Characteristics with following designation:
R-squared
Extinctionsmaxima
Obnr.   Class   Characteristics...
1       2       0.939083807     1.265447611
2       2       0.978008981     1.426726876
3       2       0.979550916     1.66513819
4       2       0.977382282     1.472939489
5       2       0.977817292     1.670339799
```

Fig. 3 Extract of learning data in OTX format

B. Classifier Development

In the ***learning phase***, a fuzzy classification model was first constructed in the two-dimensional characteristic space. The data-driven approach was combined with an expert-based approach by dividing the learning data into sharp groups and then building up the fuzzy pattern classifiers. A priori grouping is used to divide the learning data into crisp groups (see Fig. 4).

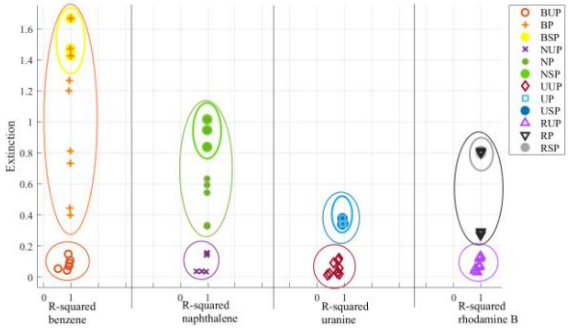

Fig. 4 Object distribution of data according to a priori grouping

The parameters for the fuzzy pattern classifier were transferred from crisp groups to fuzzy ones based on the parametric belonging function. For each characteristic of an object the membership function was described with the parameter values. Here, c is the elementary uncertainty of the objects and can be regarded as the measurement uncertainty of the respective measured values. Subsequently, the objects are first unified in one-dimensional sets and then transformed into multidimensional fuzzy pattern classes (in this case two-dimensional) using an N-fold compensatory Hamacher intersection operator [16]:

$$\cap_{Ham}^{N} \mu_i = \frac{1}{\frac{1}{n}\sum_{i=1}^{N}\frac{1}{\mu_i}} \qquad (2)$$

Here n describes the total numbers of dimensions and i present the index of the basis functions. If this is applied to all sets, the result is 12 classes in the two-dimensional feature space. (See Fig. 5).

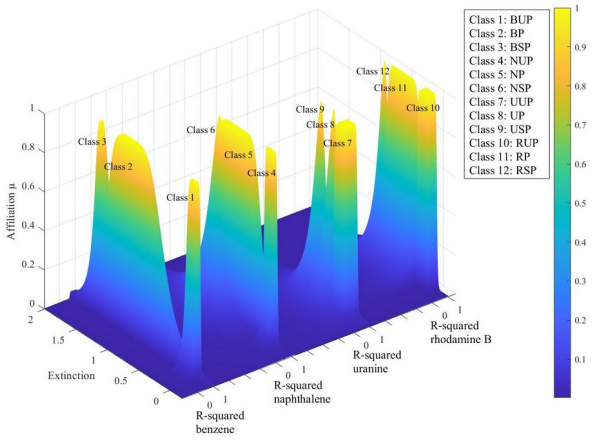

Fig. 5 Result of the classifier development

For this fuzzy pattern classifier, only the measured concentrations were considered. To ensure that very high concentrations can also be automatically included in the evaluation, the fuzzy pattern classifier was adapted using expert knowledge. The expert-based procedure offers a supplement to the data-driven approach, with which an adaptation of the constructed classifier can be carried out. Since no very high concentrations were measured in the present test runs, but it is known from Beer-Lambert's law that these also increase with increasing extinction maxima, the respective classes were enlarged in the direction of the feature "extinction maxima" (see Fig. 6).

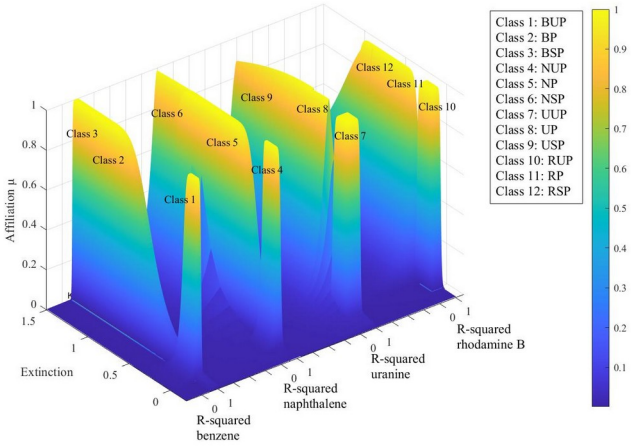

Fig. 6 Expert-based adaptation of the built classifier

In the ***working phase***, the classification model created in this way was used for the fuzzy identification of the current water samples. The work data are represented here by “artificial” work data, since they were generated by means of the original learning data set. This means that the total function was again overlayed on the spectra and the R-squared and the extinction maximum were determined. For a given substance or mixture of substances with the corresponding characteristic values for R-squared and extinction maxima, a characterization can now be carried out by determining the membership to the fuzzy groups described by the fuzzy pattern classifier. The selection of the group can typically be made according to the highest membership value. As an example, this is demonstrated for a total of six water samples or their object data sets, see Table 2 and Fig. 7.

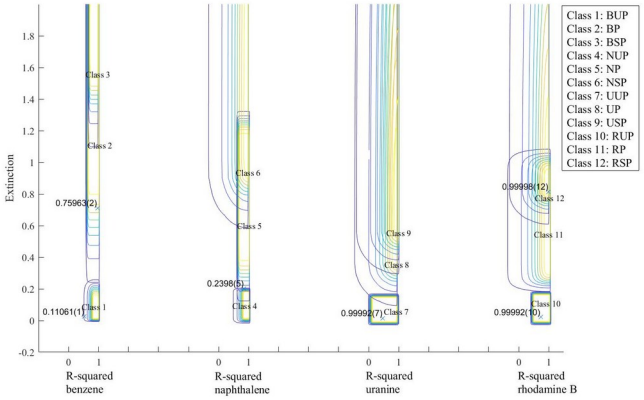

Fig. 7 Graphical representations of the assignment of test data

In Table 2, the largest value has been marked to illustrate the accuracy of the possible assignment. A partial superimposition of classes does not always allow a clear assignment. Nevertheless, interpretations can be made based on the calculated class membership.

Point 1 is only assigned to BUP with a very low membership value, which already indicates that benzene can only be present here with uncertainty. It can therefore be assumed that benzene is hardly present in this measurement. Point 2 is clearly assigned to BP with a value above 0.75. Point 6 is also clearly assigned to classes RP and RSP. Therefore, it can be concluded that benzene is present at point 2 and that rhodamine B is present for sure at point 6. At point 3, there is a low allocation to NP, whereby a tendency towards NUP is also recognisable. This fact can be explained by the same substance, but with different concentrations of these two substance mixtures, which is also reflected in the strong superposition of the corresponding classes. Here, the substance should be further observed to see in which direction it develops. In the case of points 4 and 5, both points are clearly allocated to a specific class with point 4 belonging to UUP and point 5 belonging to RUP. Since the concentrations here are very low, both are assigned to the classes that represent an uncertain presence of the respective substance.

IV. Conclusion

An approach to the characterization of water samples for on-site methods using fuzzy classification was presented. Measurement uncertainties during data acquisition and the

TABLE II.
MEMBERSHIP DATA OF THE TEST DATA FROM THE WORKING PHASE

No.	R-squared	Extinction-maxima	$\mu_{BUP,1}$	$\mu_{BP,2}$	$\mu_{BSP,3}$	$\mu_{NUP,4}$	$\mu_{NP,5}$
1	0.5323070	0.0252742	0.1106125	0.010408	0.0003607	0.000563	0.000000
2	0.9625610	0.7112262	0.0001511	0.759628	0.0297332	0.000000	0.000000
3	0.8480867	0.2022218	0.0000000	0.000000	0.0000005	0.154456	0.239803
4	0.4689550	0.0144004	0.0000000	0.000000	0.0000001	0.000000	0.0000000
5	0.7336823	0.0212848	0.0000000	0.000000	0.00000000	0.000000	0.0000000
6	0.9948569	0.8143327	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
No.	**R-squared**	**Extinction-maxima**	$\mu_{NSP,6}$	$\mu_{UUP,7}$	$\mu_{UP,8}$	$\mu_{USP,9}$	$\mu_{RUP,10}$
1	0.5323070	0.0252742	0.0044855	0.0000000	0.0019588	0.0018526	0.000000
2	0.9625610	0.7112262	0.0072945	0.0000000	0.0021951	0.0021690	0.000000
3	0.8480867	0.2022218	0.0227007	0.0000000	0.0081697	0.0071840	0.000000
4	0.4689550	0.0144004	0.0000006	0.9991880	0.0558700	0.0236306	0.000000
5	0.7336823	0.0212848	0.0000001	0.0000000	0.0000007	0.0000005	0.999922
6	0.9948569	0.8143327	0.0000001	0.0000000	0.0000006	0.0000004	0.0000000
No.	**R-squared**	**Extinction-maxima**	$\mu_{RP,11}$	$\mu_{RSP,12}$			
1	0.5323070	0.0252742	0.0004723	0.00071276			
2	0.9625610	0.7112262	0.0009161	0.00084021			
3	0.8480867	0.2022218	0.0019905	0.00158935			
4	0.4689550	0.0144004	0.0007182	0.00320964			
5	0.7336823	0.0212848	0.0000001	0.0000000			
6	0.9948569	0.8143327	0.9992831	0.9997920			

associated fluctuations in the measured characteristic values can be modelled much more flexibly and more realistically than with conventional methods due to the fuzzy group description. The data necessary to obtain the membership functions can be obtained both by real measurements and by a linguistic description of different states by a human expert. Alternatively, a combined approach is possible. In summary, the consideration of uncertainties in the detection and evaluation of water samples is of great benefit. Firstly, the data can be modelled in a much more flexible and realistic manner by means of the implementation of fuzzy information. Secondly, data-based and/or expert-based modelling can be used (applicability to numerical or linguistic characteristics including mixed combinations of characteristics) which also offers an advantage for such methods. Finally, the modelling of states with fuzzy/incomplete description and the applicability to high-dimensional characteristic spaces can be realized.

Overall, the presented methodology offers a suitable approach for automatic classification of water sample data in on-site analysis. Successful field deployments for future applications require a more extensive data base with an increased number of characteristics for more detailed characterization of the water samples.

REFERENCES

[1] Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (OJ L 327 22.12.2000 p. 1). (2006). In P. Sands & P. Galizzi (Eds.), Documents in European Community Environmental Law (pp. 879-969). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511610851.056.

[2] S. Zhuiykov, "Solid-state sensors monitoring parameters of water quality for the next generation of wireless sensor networks", Sens. Actuators B Chem., Bd. 161, Nr. 1, S. 1-20, 2012. DOI: https://doi.org/10.1016/j.snb.2011.10.078

[3] T. P. Lambrou, C. G. Panayiotou and C. C. Anastasiou, "A low-cost system for real time monitoring and assessment of potable water quality at consumer sites", Proc. IEEE Sensors, pp. 1-4, Oct. 2012. DOI: 10.1109/ICSENS.2012.6411190.

[4] T. P. Lambrou, C.C. Anastasiou, C. G. Panayiotou und M.M. Polycarpou, "A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking Water Distribution Systems", in IEEE Sensors Journal, Bd. 14, Nr. 8, S. 2765-2772, Aug. 2014. DOI: 10.1109/JSEN.2014.2316414.

[5] L. A. Zadeh, "Fuzzy Sets." - In: Information and Control 8, 338 – 353, 1965. DOI: https://doi.org/10.1016/S0019-9958(65)90241-X

[6] V. Traneva, S. Tranev and D. Mavrov" Interval-Valued Intuitionistic Fuzzy Decision-Making Method using Index Matrices and Application in Outsourcing" In: Proceedings of the 16th Conference on Computer Science and Intelligence Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, D. Ślęzak (eds). ACSIS, Vol. 25, pages 251–254, 2021. DOI: http://dx.doi.org/10.15439/2021F77

[7] V. Traneva, S. Tranev, "Two-Stage Intuitionistic Fuzzy Transportation Problem through the Prism of Index Matrices" In: Position and Communication Papers of the 16th Conference on Computer Science and Intelligence Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, D. Ślęzak (eds). ACSIS, Vol. 26, pages 89–96, 2021. DOI: http://dx.doi.org/10.15439/2021F76

[8] S. F. Bocklisch, “Prozessanalyse mit unscharfen Verfahren.” - Verlag Technik, Berlin, 1987. ISBN: 3-341-00211-1.

[9] V. Lohweg, C. Diederichs, D. Müller, “Algorithms for hardware-based pattern recognition.” EURASIP Journal on Applied Signal Processing 2004(12), 1912–1920, 2004. DOI: https://doi.org/10.1155/S1110865704404247

[10] U Mönks, V. Lohweg, and H. L. Larsen, “Aggregation Operator Based Fuzzy Pattern Classifier Design, Machine Learning in Real-Time Applications (MLRTA 09),” in KI 2009 Workshop, Paderborn | September 15th, 2009, accepted for Publication, 2009.

[11] S. F. Bocklisch, F. Bocklisch, M. Beggiato, J. F. Kremsa “Adaptive fuzzy pattern classification for the online detection of driver lane change intention, Neurocomputing,” Volume 262, 148-158, 2017. DOI: 10.1016/j.neucom.2017.02.089.

[12] F. Bocklisch, D. Hausmann, “Multidimensional Fuzzy Pattern Classifier Sequences for Medical Diagnostic Reasoning” In: Applied Soft Computing, Volume 66, 297-310, 2018. DOI: https://doi.org/10.1016/j.asoc.2018.02.041

[13] A.-J- Hempel, S. F. Bocklisch, “Parametric Fuzzy Modelling Framework for Complex Data-Inherent Structures” In: Proc. IFSA-EUSFLAT 2009, pp. 885–890, 2009. Corpus ID: 16501413.

[14] G. Herbst, “Online Recognition of Fuzzy Time Series Patterns” In: Proc. IFSA-EUSFLAT 2009. Corpus ID: 10151850.

[15] H. J. Zimmermann, “Fuzzy Set Theory and its Applications.” - Springer-Verlag, Berlin, 2001. ISBN: 978-94-015-7153-1.

[16] U. Scheunert, “Neue Verfahren der Fuzzy-Mengen-Verknüpfung und Fuzzy-Arithemtik und ihrer Anwendung bei der Sensor-Daten-Fusion. PhD thesis, TU Chemnitz, 2001.