MULTI SENSOR SYSTEM FOR FAST ANALYSIS IN ENVIRONMENTAL MONITORING WITH AN APPLICATION IN WASTE WATER TREATMENT

Kai Hennig, Thorsten de Vries, Ralf Paetzold, Kai Jantos,
Eberhard Voss and Angelika Anders

University of Hanover, Institute of Biophysics, Herrenhäuserstr. 2, D-30419 Hanover
Email: khennig@pobox.com

ABSTRACT

The presented mobile multisensor system consists of a LIF spectrometer, a reflection spectrometer, and sensors for temperature, pH value, redox potential, and conductivity for *in situ* investigations. Additional sensors can be integrated easily. This system is suitable for a fast analysis of a great number of samples (environmental monitoring). The very short measuring time and the low costs of measurements are the main advantages of the system in comparison with the chemical standard analysis. One example of a wastewater analysis shall demonstrate the characteristics of the system. A neural network shall be used as a model for the evaluation. The system can be adapted to various tasks. The margin of error varies between 5% and 25% in dependence on the application.

Keywords: fast analysis, environmental monitoring, *in situ*, wastewater analysis, laser-induced fluorescence (LIF), mobile device, artificial neural network

INTRODUCTION

Nowadays a great deal of new demands is made for the sample analysis in environmental monitoring. These demands cannot be fulfilled sufficiently by using conventional chemical methods. For this reason a new method for fast analysis has been developed at the Institute of Biophysics, University of Hanover. This method is relatively simple and can be performed at low costs.

The main conditions for environmental monitoring are a simple performance of *in situ* measurements and reliable online analysis of the achieved data. In this area the conventional chemical analysis often proves to be too complicated concerning the temporal as well as the financial aspect. Therefore, special emphasis was laid on a good implementation of *in situ* and online measurements. Furthermore, it was found that single examination methods are often not sufficient to deliver clear results in a complex environment. Only the inclusion of several methods enables a useful interpretation of the complex interactions. Therefore, a multisensor system was developed which can basically deal with facts which do not have to be completely understood and which therefore (or for other reasons) cannot be pressed into a set of strict analytical rules. Physical parameters such as temperature, pH value, and conductivity and also chemical parameters such as the specific fluorescence and absorption of molecules are involved in the measuring process.

Taking this system into consideration it is not a measurement in a strict sense but rather a simulation of a measurement. Evaluation models which enable such a system can be generally named metaheuristic. Mathematically spoken they are methods and algorithms for the solution of nonlinear problems. The most important representatives are artificial neural networks (ANN), genetic algorithms (GA), and the fuzzy logic. All of these methods have already been used for several decades. The evaluation model which is used in this measuring system is the ANN, a multilayer perceptron (MLP), whose training algorithm is a GA besides the commonly used back-propagation algorithm (1, 2).

Concerning its precision, the demands which can be made on such a measuring system have proven to be very variable. So the margin of error varies between 5% and 25% in dependence on the application. In some cases this system can only be used for the estimation of orders of magnitude and in extreme cases it only allows for a yes or no answer regarding a certain state of the examined object. Other applications, however, deliver relatively exact results. Because of these circumstances the main application of the introduced system is in the ranges for which no proven process model exists or the financial or temporal advantages predominate the disadvantages of this system.

The presented measuring system only shows the basic construction; it can be varied. Therefore, measuring devices with a computer interface can be integrated rather easily. Measuring results from an apparatus without such an interface can be transferred manually. Thus, only an automation of the measuring process is excluded.

The adaptation of the measuring system (hard- and software) to new tasks has to be done manually. At the moment the oxygen concentration can be added to the above-mentioned physical parameters. These parameters can be used directly as input values after having been normalized specifically for the ANN.

The second class of parameters which can be used are spectroscopic data. Due to the large amount of some measuring data a reduction and a pre-processing of the data are necessary. The average amount of data of a spectroscopic measurement comes to about 1,000 floating point numbers (intensity over wavelength). Normally such an amount is too large for the modelling system so that chemometric methods have to be used to reduce the amount of data.

In principle data can be used for the evaluation which are no measuring parameters in a strict sense. Such data could be the day- or the process time. Taking periodic processes into consideration one can easily imagine that time is a useful factor to describe such a process.

Besides the adaptation of the evaluation process also a mechanical adaptation of the sensors has to be made according to the task. Due to the great variation of physicochemical sensors it is normally no problem to adapt them. Those sensors are available for nearly all ranges. For the spectroscopic evaluation of liquid samples of reactors a bypass can be installed which is activated by a peristaltic pump and which makes the medium pass the sensor of the spectrometer. An optical fibre can be used for the investigation of solid material which e.g. leads the laser beam to the measured object when measuring laser induced fluorescence and which leads the fluorescence signal to the spectrometer by using a second channel.

METHODS

Adaptation of the evaluation model

The simulation is mainly based on a training algorithm and its capability of optimising the inner structure of the MLP; a process which is called teaching. Therefore, an extensive training has to be performed before the measurements are started and the network is adapted to its appropriate task. During this training procedure the input data are faced with the output data of the network. The training algorithm now tries to minimize the error between the expected results (taken from a parallel measurement with a conventional method) and the current output (supervised learning). From the mathematical point of view the output data become a representation of the input data. A simpler expression is that the network - when the training is completed - does a kind of pattern recognition on its input and tries to generate the desired output from it. Therefore, it is very important for a successful simulation that there is a sufficient correlation between input and output data. If such a correlation exists, it will normally be used by the network. The MLP is strongly related to the classic multivariate regression (3).

The most remarkable characteristic of this process is certainly the fact that the user does not have to know anything about the correlation between input and output data. The choice of the input data can

therefore be optimised by the training success. Process models of such a kind will be classified as non-parametric models, as the input parameters in comparison to the output parameters are non-specific from the user's point of view.

There are two main reasons for an excessively high output error. Configuration problems of the ANN may be one reason for that. Another reason may be an insufficient correlation between the input and the output data. One can improve the result by taking more parameters into consideration which allows for a better characterization of the measuring object. In this connection it should be mentioned that input data which are not part of the correlation can be recognized and excluded by the network during the training phase. Up to a certain number of input parameters it is conducive to add new ones. If too many parameters are used, the effect may be reversed and the error of the output may increase again. Therefore, it can be said that good results demand an optimum of selection and amount of input parameters.

Experimental set-up

The sensors which are used here for the determination of physical parameters such as temperature, pH value, conductivity, and oxygen are standard sensors (Consort). Therefore, they shall not be described in more detail.

The inclusion of spectroscopic data for the evaluation of molecular specific parameters is a more complex problem. For that purpose laser-induced fluorescence spectroscopy (LIF) is used; a method which has already proven to be suitable for *in situ* measurements (4, 5). Besides a nitrogen laser (Lasertechnik Berlin) for the excitation at 337 nm we also installed a tuneable dye laser (Lasertechnik Berlin) including a frequency doubling crystal for the range between 200 nm and 700 nm. The fluorescence spectrum is recorded by a spectrometer and a CCD camera (Oriel Instaspec IV) which can pass its data directly to the evaluating computer. Further spectroscopic data are gained from reflection and absorption measurements. A tungsten halogen lamp (Mikropack) serves as a light source. In order to make the handling easier for *in situ* measurements all the beams are led through fibre optic components. The samples are fixed in such a way that the spectroscopic data can be gained without an additional darkening.

All measurements described, the physical parameters as well as the LIF spectra, can be performed simultanously to utilize the rate advantage of the evaluaion model. The whole software is implemented in Sun Microsystem's Java, thus it is runnable on any Java supporting platform.

Data pre-processing and evaluation

Normalization of the physical parameters to the specific size of 1 for the ANN is the only necessary pre-processing of these input data. The pre-processing of the spectroscopic data is more complex. A great number of efficient methods for a multivariate data analysis are suitable for performing the necessary reduction of a large amount of data (6). The most important methods are the factor analysis, the K-means algorithm for cluster analysis, and various smoothing algorithms for suppressing the background noise of the CCD camera.

The amount of data can also be minimized manualy. The plot of the spectroscopic data may show some characteristics which can specifically be separated by simple methods such as the determination of the derivative or the mean value, see Figure 1 and a comment on it below. Also the selection of the wavelength range can be limited with basic background knowledge about the examined object.

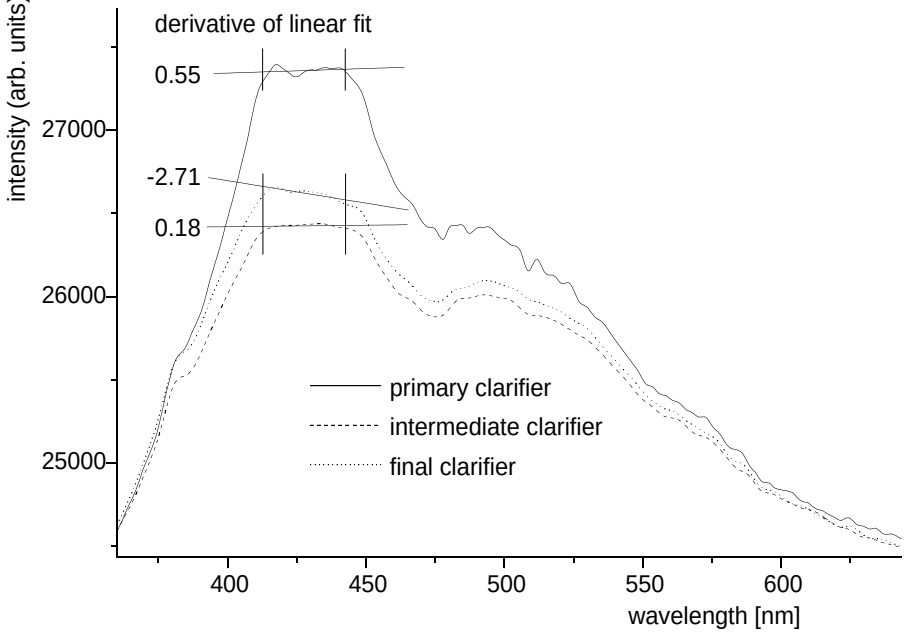

Figure 1: Fluorescence spectrum of yellow substance. Samples were taken from three different clarifiers of a communal purification plant

RESULTS

In one demonstrated example three input values will be used as the basis for a simulation of one output value. As input parameters the temperature, pH value, and conductivity of a primary clarifier in a purification plant will be used, see Figure 2. The ammonium concentration of this clarifier is the value which should be simulated (Figure 3). The ammonium content was determined using a conventional method and measured every fifteen minutes together with the input data.

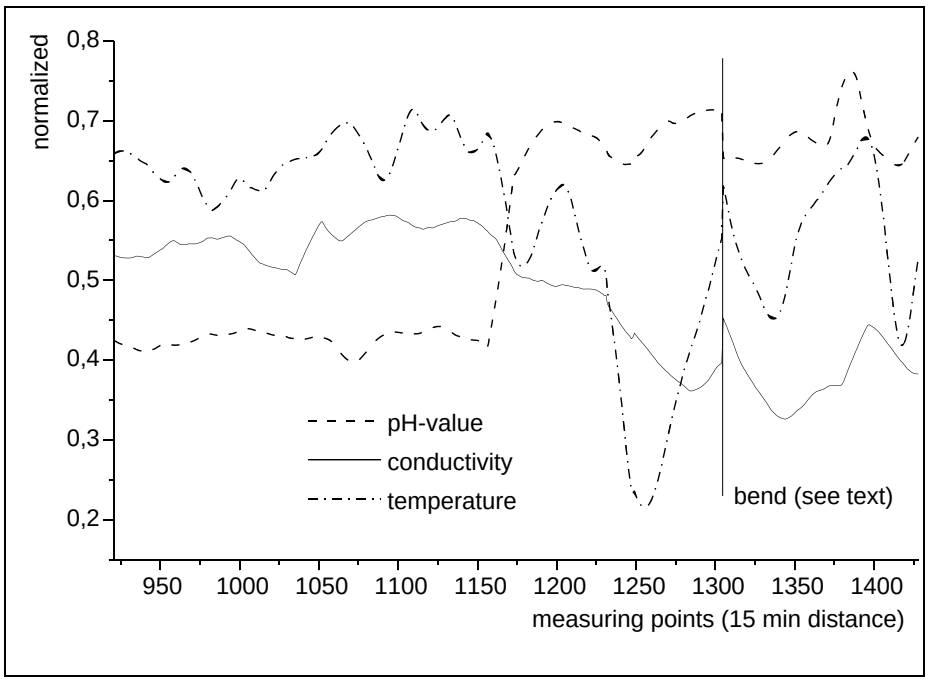

Figure 2: Input parameters for the neural network, measured in a primary clarifier of a purification plant. All values were normalized to 1.

Due to the fact that the used training algorithm does not take any time relation between the measuring points into account, they have no temporal dependency. Thus the sequence of the input of the measuring points has no influence on the result of the training. Single measuring points and also ranges of measuring points can be excluded. Records out of log files used for the training phase can be combined freely with each other. Due to disturbances during the measuring process, ranges had to be excluded from the training data quite

often. As a result of this, sharp bends can be seen in the curves of Figures 2 and 3.

The first layer of the configuration of the MLP in this experiment consists of three input neurons, followed by a hidden layer with tweny neurons and an output layer with one single neuron (Figure 4).

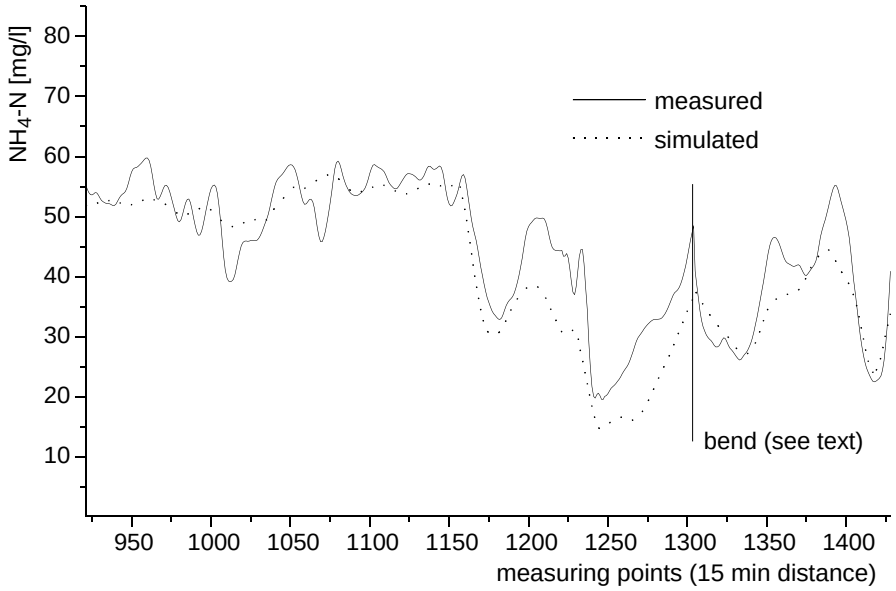

Figure 3: Measured and simulated ammonium concentrations in a primary clarifier of a purification plant.

The result of this is a number of 80 internal network parameters which can be optimised by the training algorithm. The training data set consists of about 3,400 measuring points. The relation of these two numbers excludes that the MLP is able to learn the assignment of input and output data by heart. Under appropriate conditions ANN can manage that. The configuration in this experiment, however, does not have enough network parameters to do so. According to this fact and the success of the training which can be seen in Figure 3 one can conclude that the training process uses a correlation between the input and the output data in order to deliver the present results.

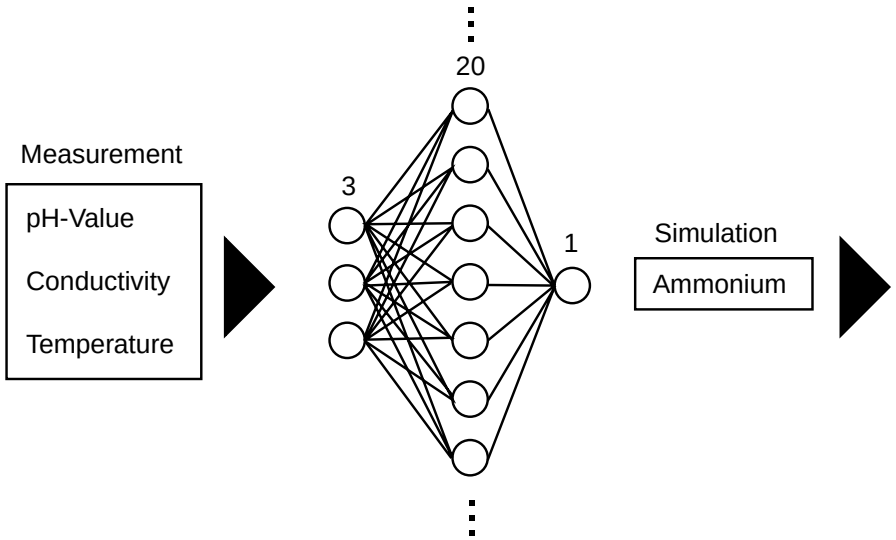

Figure 4: Configuration of the used multi-layer perceptron (MLP). Input layer with 3 neurons, a hidden layer with 20 neurons and an output layer with 1 neuron.

The actual success of the training can be checked by carrying on measuring conventionally after the training phase is completed. If archival data are used for the training, a certain amount of data are held back and when the training has been completed, the check can be done with these data. The training data used for this experiment originated from log files of a communal purification plant which can be regarded as archival data in this case. Therefore, 25% of the overall data were held back. The subsequent

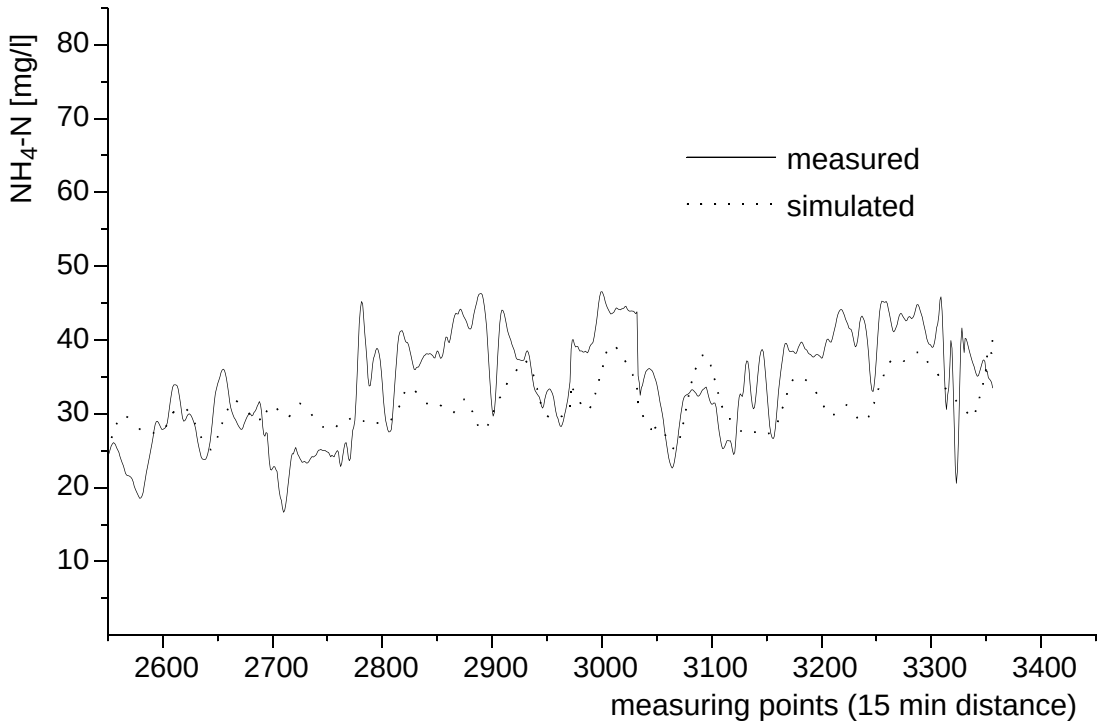

Figure 5: Check for the actual success of the training process. Measured and simulated ammonium concentrations in a primary clarifier of a purification plant. The measured values were excluded from the training process.

check of the training success with them is shown in Figure 5. The results yielded by the check show the same errors as the training data.

DISCUSSION AND CONCLUSION

Due to the high error of up to 25% it is recommendable to add more measuring parameters in order to better characterize the measuring object. The correlation between the three given input values and the output is possibly not sufficient in order to simulate the exact course of the ammonium curve in detail, which is a principal reason for an excessively high output error, see above.

The wastewater purification process mentioned here is an aerobic process for wastewater treatment. Besides the living micro-organisms which eliminate the biologically degradable waste by consuming oxygen also ammonium salts have an effect on the purification process due to their oxygen-consuming character (7). Thus, it is conceivable to include the oxygen content of the clarifier as a possible sensible parameter. Additional parameters, used for solving similar tasks with better results (8, 9), are the turbidity of the samples and the redox-potential mentioned above.

Further experiments shall show that it also makes sense to include spectroscopic data. Preliminary experiments already give a hint that the information in the existing spectroscopic data can be easily included in the evaluation by the mentioned chemometric methods. Figure 1 shows that already a simple linear fit function allows for some statements about the purification level of the sample. In the course of the fluorescence spectrum which has been excited with 337 nm one can notice the broad spectrum of yellow substance in a primary clarifier. The expression yellow substance means all the dissolved degradation products of organic substances. It is noticeable that the derivation of the linear fit through the top of the plateau of the peak on the left side in Figure 1 changes from positive to negative values during the purification process.

Due to the fact that all the results of this system are only simulations of measurements with a possible error of about 20%, it can be stated that this method and its implementation of the MLP should be carefully compared with conventional analytical methods. Therefore, further investigations shall show whether the high output error can be reduced by optimising the MLP and its training algorithm. The algorithms used at present can be regarded as straightforward implementations and the assumption is justified that a significant improvement of the present performance can be achieved by optimising the MLP.

REFERENCES

1. Brause, R. 1995. Neuronale Netze. (Teubner)
2. Rumelhart, D.E. 1986. Parallel Distributed Processing; Vol I,II,III. (MIT Press)
3. Kropp, J. 1999. Neuronale Netze und qualitative Wissensbasen in der integrativen Umweltsystemanalyse. (dissertation.de Verlag im Internet)
4. de Vries, T. et al, 1999. Mobile Sensor System for Environmental Monitoring, Proceedings of Spie Environmental Monitoring and Remediation Technologies. Edited by Vo-Dinh, T., Spellicy, R.-L. Vol. 3853: 249-253
5. Loehmannsroeben, Th. K. 1997. Laserfluoreszenzspektroskopie als extraktionsfreies Nachweisverfahren fuer PAK und Mineraloele in Bodenproben, Analytiker Taschenbuch Bd. 15 (Springer Verlag)
6. Kachigan, S.K 1986. Statistical Analysis - An Interdisciplinary Introduction to Univariate & Multivariate Methods. (Radius Press)
7. Mudrack, K., Kunst, S. 1994. Biologie der Abwasserreinigung. (Gustav Fischer)
8. Becher, T. 1994. Simulation, Ueberwachung & Steuerung v. Abwasserbehandlungsanlagen mit neuronalen Netzen. Abwassertechnik. 5/1994
9. Rosenwinkel, K.-H. 1998. Dynamische Simulation und kuenstl. neuronale Netze im Abwasserbereich. (ISAH-Publication)