

DATA ACQUISITION SYSTEM FOR BIOREMEDIATION LABORATORY

Ioan MOGA¹, Istvan KISS², Tiberiu VESSELENYI³, Balázs FEHÉR⁴

¹University of Oradea, imoga@uoradea.o

²Bay Zoltán Nonprofit Ltd. for Applied Research
Institute for Biotechnology (BAY-BIO), balazs@bay.u-szeged.hu

³University of Oradea, vtiberiu@uoradea.ro

⁴Bay Zoltán Nonprofit Ltd. for Applied Research
Institute for Biotechnology (BAY-BIO), balazs@bay.u-szeged.hu

Abstract—In order to optimize bioremediation processes, these processes have to be studied in controlled environments. Such an environment is developed using test tanks in which underground liquid flow, temperature and pressure parameters can be controlled. In order to obtain reliable results these parameters must be monitored with state of sensors and data acquisition systems. This paper describes such a system which had been developed in the framework of cross-border cooperation program CBC-HURO, 0802/100_AF project.

Keywords—bioremediation, data acquisition system, temperature sensors, sensor calibration.

I. INTRODUCTION

BIODEGRADATION is a key process in the natural attenuation of contaminants at hazardous waste sites.

Bacteria can metabolize hydrocarbons and other contaminants, converting them to less toxic products. Some live deep underground, some live in the absence of oxygen. During a bioremediation process specific organisms are injected into the groundwater, and in some cases, special nutrient are injected with the microbes. The method is especially useful for remediation of hydrocarbons in groundwater.

Due to human activities, organic pollutants (including oil) are spilled to the environment where they threaten public health, often as contaminants of soil or groundwater. Petroleum hydrocarbons are pollutants because these compounds are frequently spilled, readily water-soluble, partly toxic and carcinogenic, and therefore undesirable in soils and drinking water supplies. The petroleum hydrocarbons influence the microbial community in under grounds. Living organisms are able to transform or mineralize many organic pollutants, and bioremediation techniques have been developed to remove pollutants from a contaminated site. However, fast and easy methods to document both the efficacy of bioremediation and the changes in soil microbial communities during bioremediation are not well developed.

In order to optimize bioremediation processes, these processes have to be studied in controlled environments.

Such an environment is developed using test tanks in which underground liquid flow, temperature and pressure parameters can be controlled. In order to obtain reliable results these parameters must be monitored with state of sensors and data acquisition systems..

II. DATA ACQUISITION SYSTEM FOR THE BIOREMEDIATION LABORATORY

In the framework of cross-border cooperation program CBC-HURO, 0802/100_AF project, a data acquisition system has been developed which is capable to read 400 analog channels in order to monitor temperature and CO₂ sensor signals. The laboratory installation in which the measured process takes place consists of two test tanks (show in figure 2.c), which are equipped with temperature and CO₂ sensors. The acquired data is used as control parameter for the bacterial process or can be useful as validation data for the modeling and simulation system. In order to develop the data acquisition system the National Instruments NI PXI 1042 base had been chosen, which satisfies all the system requirements. Figure 1 shows the structure of the data acquisition system, figure 2.a presents the components of the data acquisition system core and figure 2.b shows the picture of the ready to use system placed in the research laboratory.

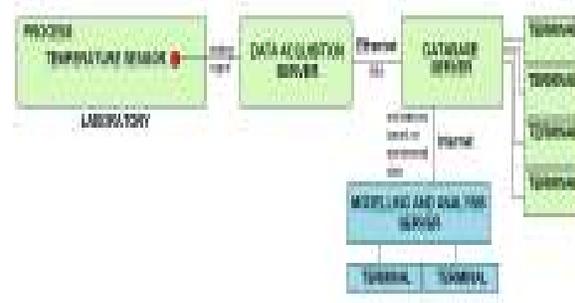


Fig. 1. Data acquisition system structure.

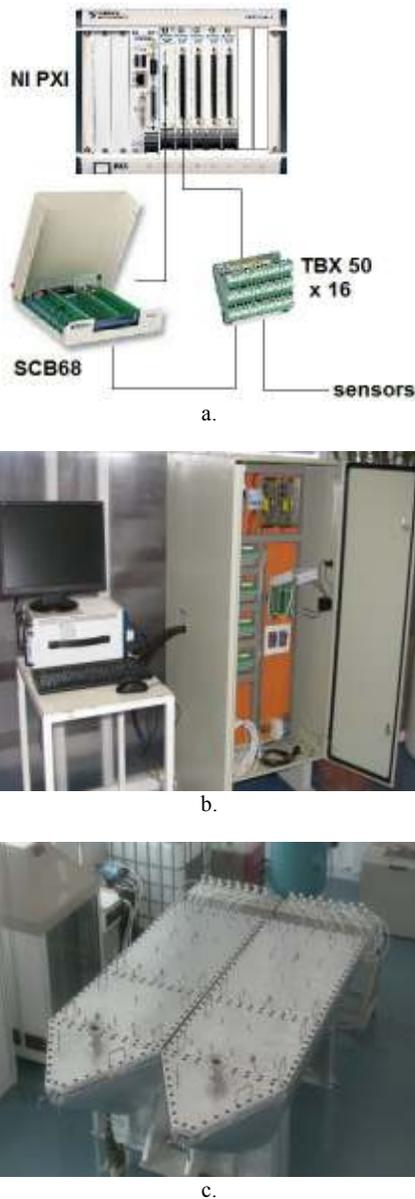


Fig.2. Data acquisition system core (a), developed system (b), test tanks.

The NI PXI system is composed by one NI-PXI-1042 base equipped with NI PXI-6221 data acquisition card with 16 Analog Inputs (on 16 bit), 24 Digital I/O, 2 Analog Outputs and four pieces of NI PXI-2575 Multiplexer with 196 channels each.

The system is also equipped with the NI LabVIEW Database Connectivity Toolkit, having the following characteristics:

- quickly connect to local and remote databases;
- implement many common database operations without having to perform SQL programming;
- connects to popular databases, such as Microsoft Access, SQL Server,

Oracle, postgresSQL, MySQL;
 - uses the ODBC interface to access the database management systems.

As a result the data read by the acquisition system is available via Internet in the modeling and simulation facility at the University of Oradea. In order to work properly the sensors must be calibrated. The developed calibration procedure is described in the followings.

II. TEMPERATURE SENSOR CALIBRATION PROCEDURE

Temperature sensors for multipoint measurement system had been tested at Bay Zoltan in Szeged and in the University of Oradea. The calibration had been made placing the sensors in a controlled environment in water and using a high precision Digital Thermometer (HPDT).

The data was read by Internet at the modeling laboratory at the University of Oradea.

In the first step the offset value (bias) of every sensor is computed by first computing the mean value for the 10 sensors and then adding the difference of this value to each measurement.

After bias correction for each sensor, in the second step an approximation function is searched, based on sensors mean value in points that were also measured with the HPDT instrument.

The calibration measurements for 9 temperature sensors are given in table 1.

TABLE I
 CALIBRATION MEASUREMENTS

Measurement index	Sensor mean values [degree C]	HPDT values [degree C]
0	1.54860	1
1	4.89000	5
2	9.77710	10
3	14.7500	15
4	19.5643	20
5	24.2071	25
6	29.1929	30
7	34.2929	35
8	38.9500	40

The diagram representing sensor and HPDT data is given in figure 3. It can be observed that the two diagrams are not fitted.

In order to fit the sensor data to the HPDT data an approximation function had been searched using CFTOOL module of the MATLAB programming environment.

The statistical data analysis functions were used to acquire knowledge on measured values distribution in order to have an image on accuracy and precision of measurements. After that, curve fitting algorithms were employed to find a minimal error function which complies with the measured data and which can be a good candidate for calibration function. We also present the results of statistical analysis and curve fitting obtained and conclusions that can be drawn from these results.

Generally, curve fit algorithms determine the best-fit parameters by minimizing a chosen merit function. In order to optimize the merit function, it is

necessary to select a set of initial parameter estimates and then iteratively refine the merit parameters until the merit function does not change significantly between iterations. One of the most used algorithm for nonlinear least squares calculations is the Levenberg - Marquardt Algorithm (LMA).

The LMA interpolates between the Gauss–Newton algorithm and the method of gradient descent and is more robust than the Gauss–Newton algorithm, which means that in many cases it finds a solution even if it starts very far off the final minimum. For well-behaved functions and reasonable starting parameters, the LMA tends to be a bit slower than the Gauss–Newton algorithm. LMA can also be viewed as Gauss–Newton using a trust region approach.

A disadvantage of the LMA is that it finds only a local minimum, not a global minimum.

In order to understand curve fitting algorithms there are two important topics that has to be defined: parametric fitting and the least square method.

A. Parametric fitting.

Parametric fitting involves finding parameters for one or more models that can be fitted to data. The data is assumed to be statistical in nature and is divided into two components: *deterministic and random component*.

The deterministic component is given by a parametric model and the random component is often described as *error* associated with the data. The model is a function of the independent variable and one or more coefficients. The error represents random variations in the data that follow a specific probability distribution, usually a normal or Gaussian distribution. In case of optical density measurements the variations can come from many different sources, but we assume that they are present in the measured data. Systematic variations can also exist, but they can be eliminated with a proper calibration of the instrument and careful handling of the instrument.

There are several different models available for curve fitting. Some of the most used are presented below.

1. *Straight Line model*. Is made choosing the line that minimizes the least square sum of the vertical distance d , of all predictor data;

2. *Logarithmic model*. Calculates the least squares fit through points by using the following equation: $y = a + b \cdot \ln x$, where a and b are constants and \ln is the natural logarithm function;

3. *Exponential model*. Generally one-term and a two-term exponential models are applied, as given by

$$\begin{aligned} y &= a \cdot e^{bx} \\ y &= a \cdot e^{bx} + ce^{dx} \end{aligned} \quad (1)$$

where a, b, c and d are constants, and e is the base of the natural logarithm;

4. *Power model*. The power model is implemented by using the following equation: $y = ax^b$, where a and b are constants;

5. *Polynomial model*. Is implemented by using the following equation: $y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \dots$, where a_0, a_1, a_2, \dots , are constants. In general form this can be written as:

$$y = \sum_{i=1}^{n+1} a_i x^{n+1-i} \quad (2)$$

where $n + 1$ is the *order* of the polynomial, n is the *degree* of the polynomial.

Polynomials are often used when a simple empirical model is required. The polynomial model can be used for interpolation or extrapolation, or to characterize data using a global fit.

6. *Power Series Models*. For power series models one-term and a two-term power series model can be used as given by:

$$\begin{aligned} y &= a \cdot x^b \\ y &= a + bx^c \end{aligned} \quad (3)$$

Power series models describe a variety of data like for example the rate at which reactants are consumed in a chemical reaction is generally proportional to the concentration of the reactant raised to a certain power.

7. *Rational Models*. Rational models are defined as ratios of polynomials as for example, given by:

$$y = \frac{\sum_{i=1}^{n+1} p_i x^{n+1-i}}{x^m + \sum_{i=1}^m q_i x^{m-i}} \quad (4)$$

where: n is the degree of the numerator polynomial, while m is the degree of the denominator polynomial.

B. Least square method

The majority of algorithms use the method of least squares when fitting data. Fitting requires a parametric model that relates the response data to the predictor data with one or more coefficients. The result of the fitting process is an estimate of the model coefficients.

To obtain the coefficient estimates, the least-squares method minimizes the summed square of residuals. The residual for the i -th data point r_i is defined as the difference between the observed response value y_i and the fitted response value \hat{y}_i , and is identified as the error associated with the data:

$$r_i = y_i - \hat{y}_i \quad (5)$$

The summed square of residuals is given by:

$$S = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

where n is the number of data points included in the fit and S is the sum of squares error estimate. There are a large number of types of least-squares fitting which includes: linear least squares, weighted linear least squares, robust least squares, nonlinear least squares.

When fitting data that contains random variations, there are two important assumptions that are usually made about the error:

1. the error exists only in the response data, and not in the predictor data;

2. the errors are random and follow a normal (Gaussian) distribution with zero mean and constant variance, σ^2 .

As it has been shown in the previous chapter, the errors are assumed to be normally distributed because the normal distribution often provides an adequate approximation to the distribution of many measured quantities. Although the least-squares fitting method does not assume normally distributed errors when calculating parameter estimates, the method works best for data that does not contain a large number of random errors with extreme values. The normal distribution is one of the probability distributions in which extreme random errors are uncommon.

However, statistical results such as confidence and prediction bounds do require normally distributed errors for their validity.

If the mean of the errors is zero, then the errors are purely random. If the mean is not zero, then that model is not the right choice for the data, or the errors are not purely random and contain systematic errors.

A constant variance in the data implies that the "spread" of errors is constant. Data that has the same variance is sometimes said to be of equal quality.

The curve fitting function is given in figure 2.

This function has the following expression and fit characteristics:

$$f(x) = a \cdot x^b + c \quad (7)$$

Coefficients (with 95% confidence bounds):

$$a = 0.7866$$

$$b = 0.2817$$

$$c = -1.287$$

Goodness of fit:

$$SSE: 0.1154$$

$$R\text{-square: } 0.9383$$

$$\text{Adjusted } R\text{-square: } 0.9178$$

$$RMSE: 0.1387$$

The diagram of sensor values and HPDT data after fitting is presented in figure 3.

As it can be seen from diagram in figure 3 the calibration procedure is complete.

During experiments the calibration procedure will be applied as necessary and is recommended to be made every 3 month or if observed sensor data is out of range.

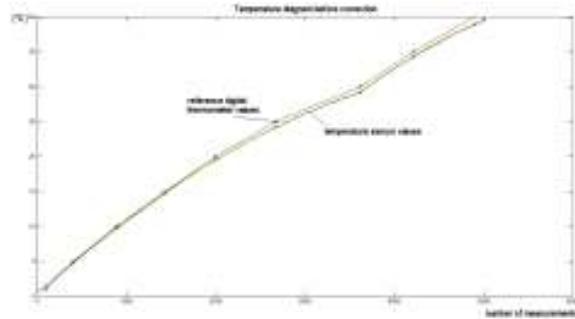


Fig. 3. HPDT and sensor data before fitting.

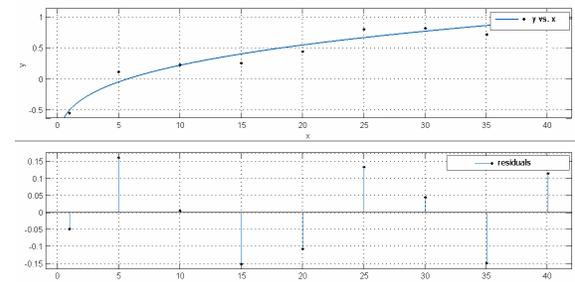


Fig.4. Fitting curve and approximation error curve.

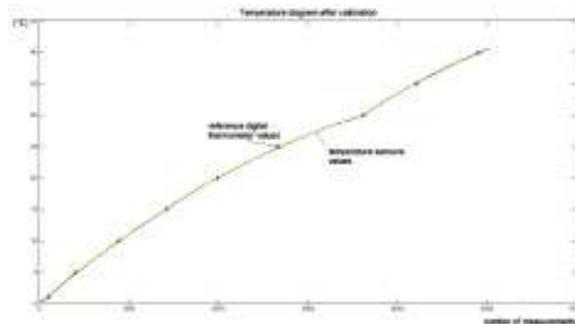


Fig. 5. HPDT and sensor data after fitting.

III. CONCLUSION

The acquisition system developed as a result of CBC-HURO, 0802/100_AF project, and installed at the BayBio bioremediation laboratory in Szeged is a great capacity (over 400 analog channels), flexible system, programmable via the Labview software. It has the capacity to transmit the data measured in the experimental laboratory in Szeged to the modeling and simulation facility at the University of Oradea. The data acquisition application software can also be modified or adjusted from the University of Oradea via the same Internet connection.

The system is capable to also read data from other sensors like for example CO₂ or other gas sensors. This data collected via experiments are then used in bacterial process modeling, in order to optimize the process of bioremediation.

ACKNOWLEDGMENT

This research was supported by the European Union and co-financed by the States of Hungary and Romania in the framework of cross-border cooperation program

CBC-HURO, 0802/100_AF, acronym “Micromodel”
project.

REFERENCES

- [1] National Instruments, NI-PXI User manual, 2012,
<http://sine.ni.com/np/app/main/p/ap/global/lang/ro/pg/1/sn/n24:PXI-FSLASH-CompactPCI/>
- [2] Labview , Programmer’s Manual, National Instruments, 2012,
<http://www.ni.com/pdf/manuals/321296b.pdf>
- [3] MATLAB, Programming Manual, Mathworks, 2012.
- [4] MATLAB CFTOOL User manual, Mathworks, 2012.
- [5] LM35A, Precision Centigrade Temperature Sensor, National Semiconductor, 2000.