Theme D: Remediation Concepts & Technologies
Ph.D. Work: yes
Title: **SIGNAL TREATMENT APPLIED TO ENVIRONMENTAL BIOLOGICAL SYSTEMS**
Cristina Vila; António Fiúza
Full mail address of authors: Research Center in Geoenvironment and Resources, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, PORTUGAL
Phone: +351 225081989
Fax: +351 22 5081448
E-mail: mvila@fe.up.pt; afiuza@fe.up.pt
Keywords: Biodegradation, Respirometry, Windowed Fourier Transform, Wavelets, System Identification

SUMMARY

The respirometry of soils contaminated by petroleum products offers a potential of research that has not been thoroughly exploited. During the last years we performed a long sequence of experiments in laboratorial reactors of different capacities, shifting from 1 L to 5L, where we measured continuously, on-line, the oxygen and carbon dioxide concentrations at the reactors inlet and in the immediate atmospheric vicinity of the contaminated soil submitted to biodegradation. The air flow was kept constant using a control system and the environmental properties, such as the temperature, were also simultaneously measured and recorded. The experimental conditions that favor the biodegradation, such as the moisture and the addition of nutrients, were periodically adjusted. Each experiments lasted, at least, around 30 days.

A single experiment creates an enormous quantity of data, with several millions of registers. We have been treating this enormous amount of information using several mathematical techniques. The first step is always the filtration of the data in order to eliminate anomalies strange to the process, such as voltage breakages and current losses. The length of the data is also reduced using conventional methodologies.

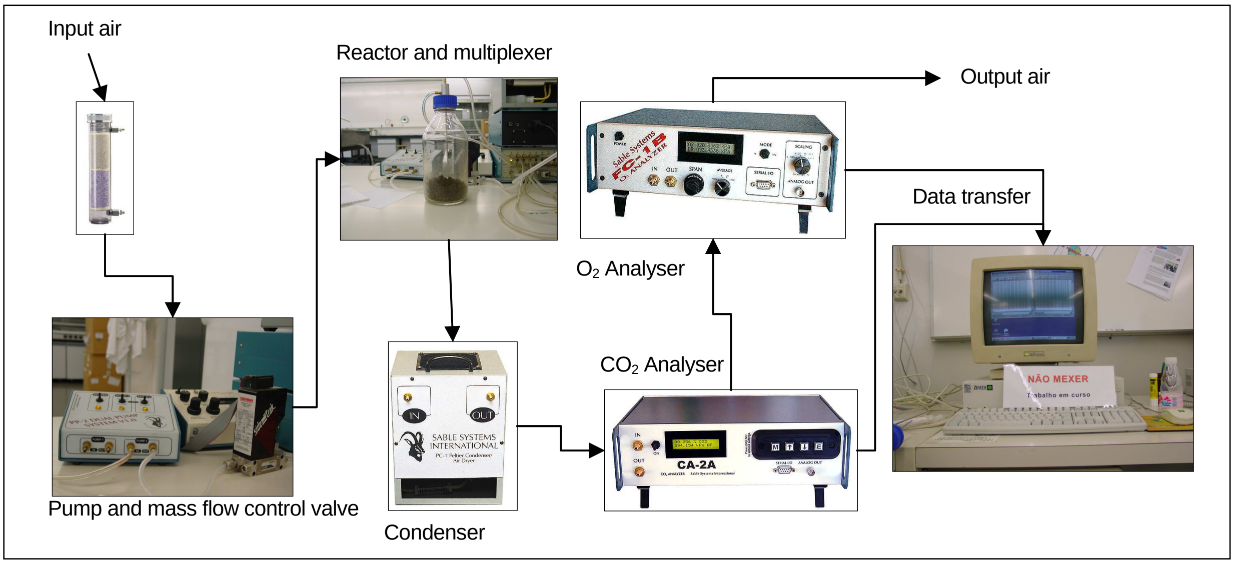
One of the methodologies we use is the signal treatment by Fourier Analysis. The periodograms allowed the detection of a daily cyclical behaviour in the activity of the microorganisms. Directional circular statistics allowed the establishment of circular correlations between the daily patterns of the temperature oscillation and the biological activity. Time series analysis is used to produce auto-correlograms as well as cross-correlation diagrams between the main variables involved.

The wavelet analysis is used to detect variation patterns at smaller scales and we investigated its usage as a method of filtration the data that keeps the inner core structure of the information without aliasing. We have been using the system identification theory to create digital data-driven models, either single input-single output or multiple input-multiple output. These models were applied to data filtered by conventional methods as well as using wavelets. It is also possible to create phenomenological relationships between the different measures allowing the determination not only of the kinetics but also of the stoichiometry of the biodegradation reactions.

INTRODUCTION

Soil respirometry has been seldom used for studying the biodegradation of organic contaminants in soils in opposition to its broad application for similar studies in aquatic environment. A soil respirometer measures continuously the concentrations in oxygen and carbon dioxide, sometimes other gases as methane, and some environmental variables such as temperature, in the atmosphere in the vicinities of a soil under biodegradation in a reactor.

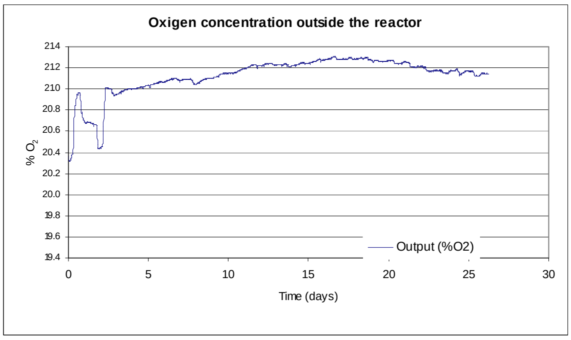
The respirometer used on this work was the model TR-RM8 Respirometer Multiplexer from Sable Systems International. The system, as it was initially conceived, consists of instruments that capture, direct and control a continuous air flow through a succession of interrelated devices, in order to analyse the composition of the air in the biodegradation reactor, and to communicate this information to a computer for storage and subsequent analysis. The schematic diagram represents the air path through the experimental devices. The feeding air is carefully collected outside the laboratory in a non-polluted zone, it is scrubbed in a drierite column and it is pushed by a pump with a mass flow controller (to maintain a constant volume of air) to the rest of the system. It first enters one of the maximum eight possible soil sample chambers (reactors) selected in the multiplexer, where biodegradation is occurring, and goes through the condenser and another drierite column to lose eventual existent humidity. Then, the dried air enters into the carbon dioxide analyser, passes through a small column of ascarite to remove the CO_2 content and is analysed in the oxygen probe, returning the air to the laboratory atmosphere. Let us examine the corresponding flux of data. Although the system was provided with an analogical/digital converter AD-201, with a Universal Interface in order to record and treat the data using software specially developed for respirometric studies, we decided to connect each analyser directly to the computer using a multi-port serial interface. With this procedure we collect the data directly, avoiding two data transformations as well as the data aliasing operated by the original software. The computer receives two records per second from each analyser, which results in an intense amount of data. The soil samples subjected to biodegradation inside the reactors were periodically moisted and nutrients were added once or twice during each experiment.

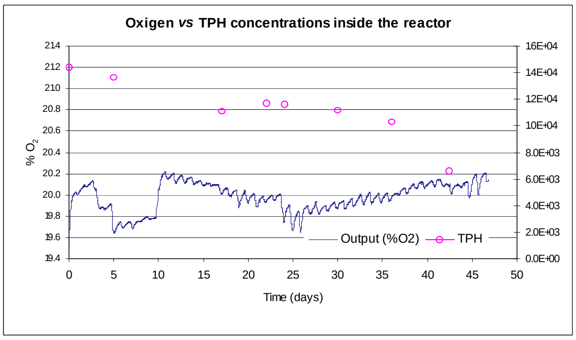


Several respirometric tests were made with different experimental conditions: using reactors either with a volume of 1 l or 5 l, varying the air flow rate and adding mineral medium and/or a nitrogen source. The biodegradation was also chemically quantified through the periodical collection of samples where the TPH (acronym for Total Petroleum Hydrocarbons) was chemically analysed. In some experiments the biomass was also estimated through plate count agar. When a full experiment is accomplished the following variables are recorded continuously: oxygen and carbon dioxide concentrations at the inlet, measured using an empty parallel reactor, O_2 and CO_2 concentrations at the outlet measured in the reactor where biodegradation is performed and the inner temperature. The TPH was analysed weakly (in average) and the biomass estimation was only performed in select experiments. From this data it is possible to estimate the instantaneous and accumulated oxygen consumption, the instantaneous and accumulated CO_2 production, as well as the ratio between the O_2 consumption and the CO_2 production. It is also possible to acquire a detailed determination of the kinetics and a complete time differentiation of all the evolution phases involved (adaptation, degradation and closing stage). When a full experiment is performed it is also possible to relate the previous variables to the actual degradation of TPH in the soil and, in certain experiments, to the evolution of biomass. As an example the operating conditions for a single experiment are exemplified:

Empty Reactor		**Reactor with contaminated soil**	
Chamber volume	5 l	Reactor volume	5 l
Air flow rate	5 ml/min	Air flow rate	6 ml/min
Residence time	22,6 days	Residence time	22,6 days
Soil Characteristics			
Soil sample weight	893 g	Initial TPH concentration	13.36 g/Kg soil
Soil humidity	22 %	Native microorganisms	6 x 10^6 CFU/g of soil
		Average soil grain size	61% between 2 mm and 50 µm

An example of measurements is represented in the next figure.



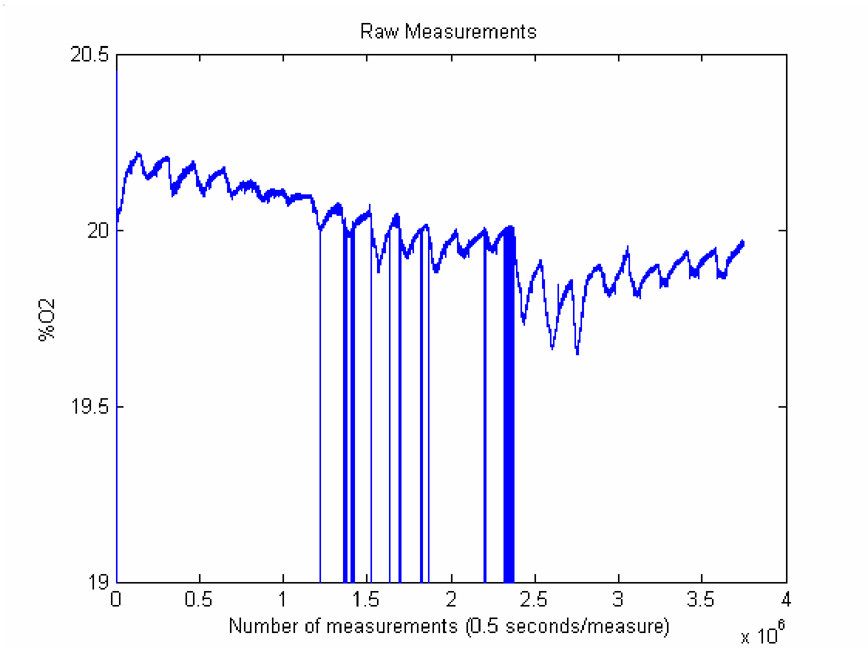


The left graph shows the input to the system (oxygen in the atmosphere) while the right one exhibits the main output variables (oxygen concentration at the outlet and TPH decay). It is possible to clearly differentiate the evolution: an initial instable phase of adaptation, the active biodegradation phase and the slow closing stage.

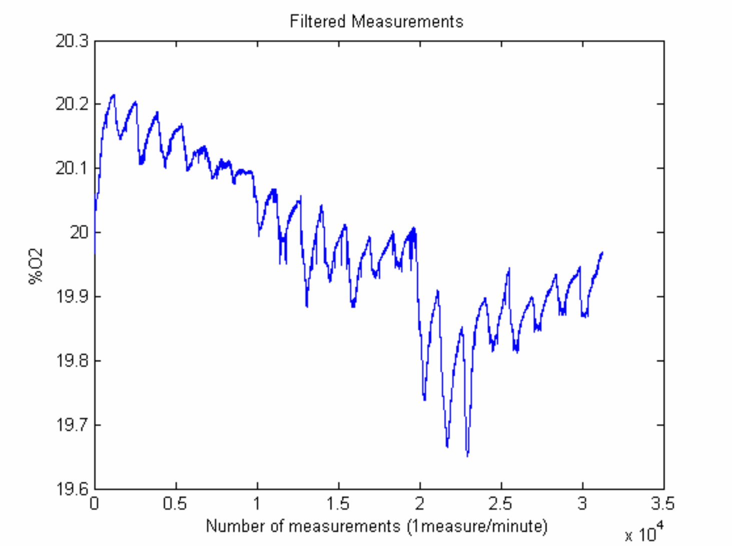
RAW DATA

The main data collected in all the experiments was the oxygen concentration at the inlet and at the outlet. As stated, in some other experiments the carbon dioxide concentrations were also measured; these values were filtered using moving averages in order to reduce the size of the available information.

As an example we show a file which has a net phase of active biodegradation. It has 3 747 609 registers of oxygen concentration inside the reactor where soil biodegradation is occurring, corresponding to 2 measurements per second and 21.7 days for duration of the experiment.



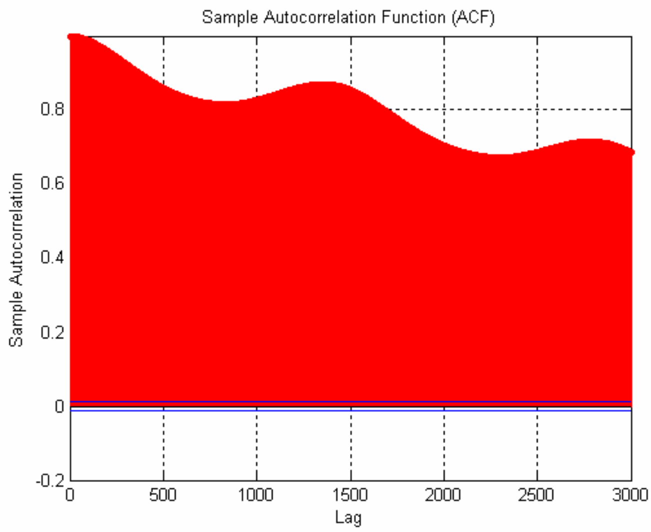
It is possible to detect immediately the existence of sinusoidal cycles, in a total number of 22, each corresponding to one day. It is also possible to detect easily through the graphic that variations in tension of the electric power, of short, medium and long duration, are occurring with a previously unsuspected frequency. The raw data is then filtered in order to annihilate these type of perturbations and also as a mean of reducing the total amount of data to one measurement each 10 or 20 minutes, thus reducing largely the enormous amount of data. As an example next figure shows the same after filtration. The number of register was reduced 120 times.



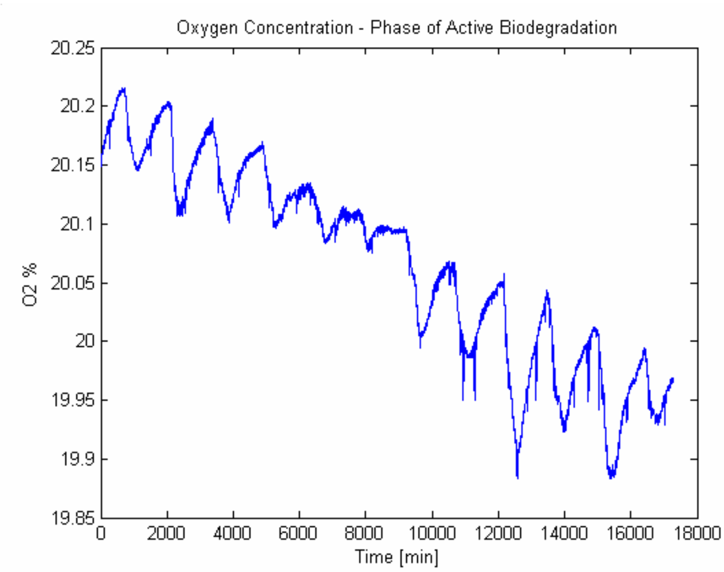
It also can be concluded that the output concentrations exhibit an obvious pattern: a daily cycle of microbial activity with maximum intensity during daytime and a perceptible reduction during night. It is also easily discernible three phases in the evolution of the biodegradation: a long initial active phase where oxygen concentrations are dropping, a relatively short steady-state phase with constant concentrations and the beginning of the decreasing phase of biological activity where oxygen concentrations are increasing.

PRELIMINARY ANALYSIS OF DATA

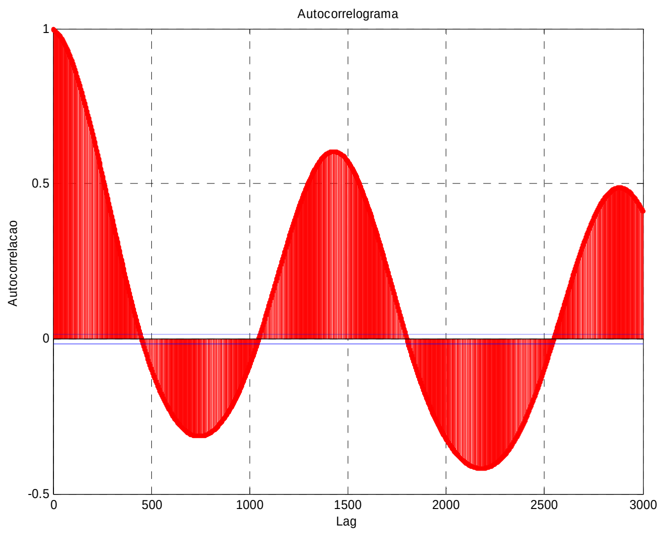
First an autocorrelogram is built for the filtered data. As an example for the previous filtered data, the following correlogram was obtained



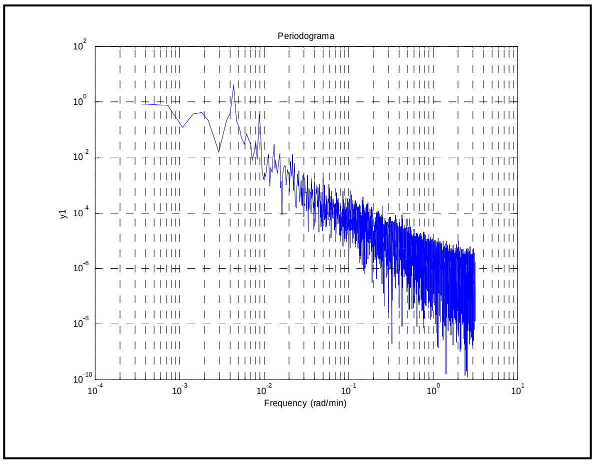
A maximum lag of 3000 was adopted, corresponding to around ten percent of the total length of the register. It can clearly be noticed a sinusoidal shape with decreasing amplitude. The period of the sinusoid is 1440 minutes, which corresponds to one day. The amplitude of the sinusoid decreases corresponding to a decreasing in the oxygen concentration and thus to a biological active phase of the degradation. Let us slash the file by sole considering a component with active biodegradation. We considered 1702 registers, corresponding to twelve days of activity, taken from the initial component of the global filtered results.

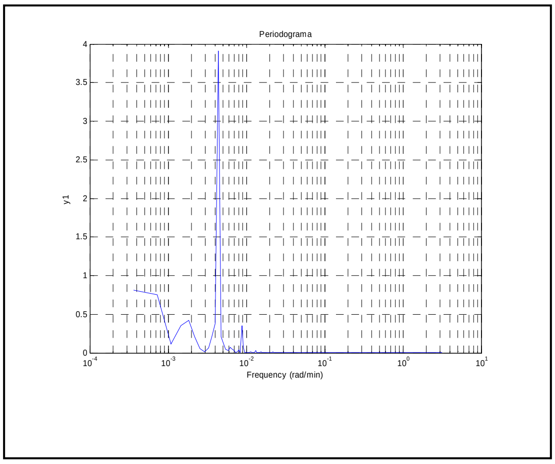


As next step the linear decreasing tendency was taken off the data. A new correlogram is built using this new information.



The cyclic component is kept but the tendency disappeared. A periodogram for this data is built and is shown in next figure as well as an amplification of it.





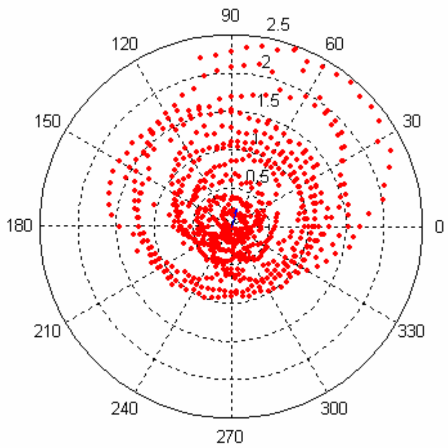
The periodicity of the biological activity becomes even more evident – the major peak corresponds to a 24 hours period and the minor to a period of 12 hours.

THE DAILY CYCLE OF BIOLOGICAL ACTIVITY – CORRELATION WITH TEMPERATURE

The main reason for the daily cycles of biological activity should be induced by variations in the temperature. However to establish a sound correlation between two seasonal phenomena involves a non-conventional statistical approach, due to this particular nature of the variables. The directional circular statistics can become a solution with certain particular assumptions. The sampling was done at regular intervals and the time was transformed into a circular variable with a cycle of 24 hours. The methodology is the following:

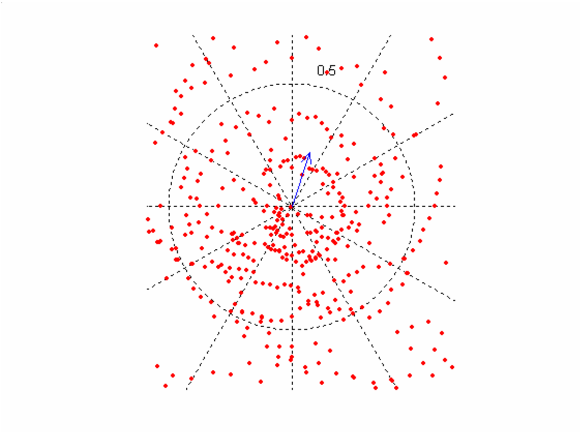
1. Transformation of the measured variable (concentrations, temperature) from nominal values to reduced form values.
2. Transformation of time into circular data with an amplitude of 1 day; each instant in a day corresponds to an angle between 0 and e 2π;
3. Each measure was considered as a vector in a bi-dimensional space, having as module the reduced value of the measured variable and orientated in the space in an angle corresponding to the instant when the measure was taken;
4. The resulting mean vector ($\bar{r}$) was calculated and corresponds to the vector sum of all the measures weighted by the number of vectors used in the estimation;
5. Its module, that we call *average resulting length* (ρ) and the angle ($\bar{\theta}$) of its orientation in space, that we call *direction of the mean vector*, where calculated.
6. The average resulting length is a point estimate of the average length of the measures and can be used as a measure for the concentration of the distribution. If the intensities are uniformly distributed around the circle, a preferential direction does not exist and the average resulting length would be near zero.
7. As well as in the conventional circular statistics we can estimate a *circular variance* defined by
$$v = 2(1-\bar{r}) \in [0,2]$$
8. And we can also define a *circular standard deviation* in the following way
$$s = \sqrt{-2.\ln(\bar{r})}$$

As an example we show the results obtained from one experiment where the variable is the oxygen concentration in the reactor where bioremediation is occurring.



In the figure each point represents a measure and the vector near the origin represents the resulting vector. In this example the preferential direction is 1.2454 radians, or 72,355º; the average reduced length is 0,2321 corresponding to a nominal value of 20,097 % O_2; the circular variance is 1,5359 and the standard circular variation is 1,7092.

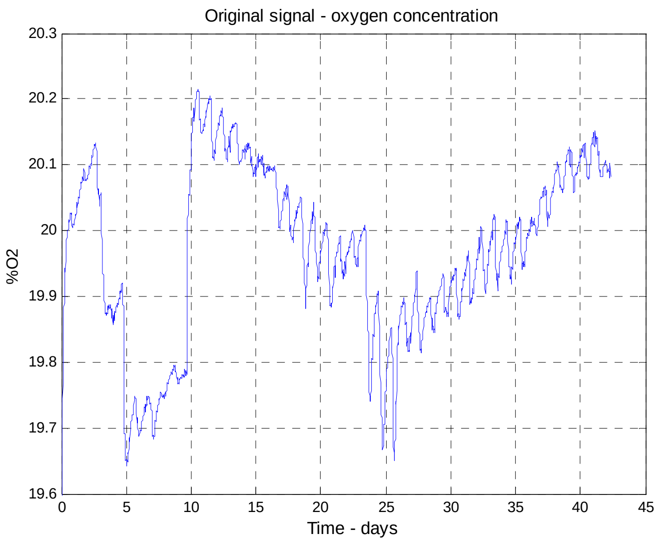
Next figure shows an amplification of the resulting vector which occurs 4.76 hours after the initial reference time (zero).

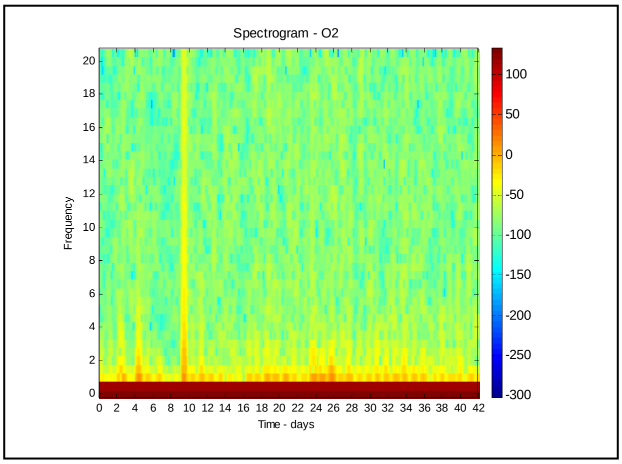


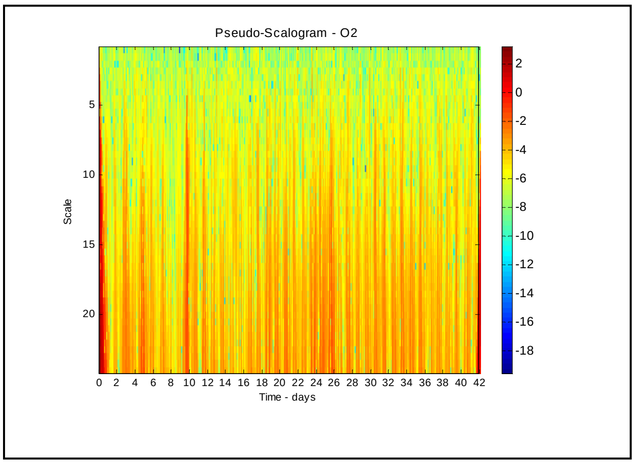
As known the von Mises distribution is the circular equivalent to the Gaussian normal distribution and can also be characterized by two parameters: the average direction $\bar{\theta}$ and a concentration parameter κ. The von Mises distribution is unimodal and symmetric around the average direction. As the concentration parameter increases the Vero similarity of observing a direction measure located near the mean direction increases. If κ is zero, all the directions are equally probable and the underlying distribution is uniform circular. The concentration parameter was estimated from the resulting average length $\bar{r}$. If we admit that the observations come from a von Mises distribution, the randomness test means that we want to prove there are no preferential directions or, in other words, the intensity of the phenomenon is the same for all the instants in time. This hypothesis is equivalent to consider that the concentration parameter $\kappa = 0$. We used the test initially developed by Lord Rayleigh, in the beginning of the 20th century, and later improved by Mardia (1972). The tests applied to the variables temperature, oxygen and carbon dioxide concentration inside the reactors where bioremediation is occurring, showed that the statistical hypotheses of uniformity should be rejected at the 95% level, while the same hypothesis should be accepted for the empty the oxygen and carbon dioxide concentrations in the empty reactors. The location of the resulting average vectors, or concentration parameters, also proved undoubtedly that there is a correlation between the temperature and the air concentration in oxygen and carbon dioxide when bioremediation is occurring.

APPLICATION OF THE WINDOWED FOURIER TRANSFORM

Applying windowed Fourier transform to the data acquired by respirometry it was possible to produce another characterization of these signals. As an example, for the oxygen concentrations inside a reactor where biodegradation was occurring (see next figure) it is possible to remark, both in raw data and in the spectrogram, around the tenth day an energy spot spread for all the frequencies which correspond, in the recorded data, to the jump from the instable initial period to the beginning of the active biodegradation period.







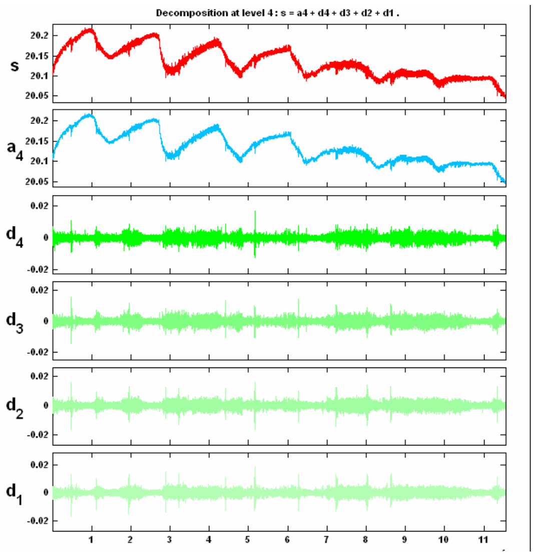
APPLICATION OF WAVELETS TO SIGNALS GENERATED BY BIOREMEDIATION

Wavelet theory has its source on the windowed Fourier transform, although the wave window is significantly different. Between 1960 and 1980, Morlet (seismologist) and Grossman (physician) developed some functions which constitute orthonormal basis and that can be scaled in order to be short in the high frequencies and long in the low frequencies, however maintaining its form – to such functions they called wavelets (Mallat, 1998).

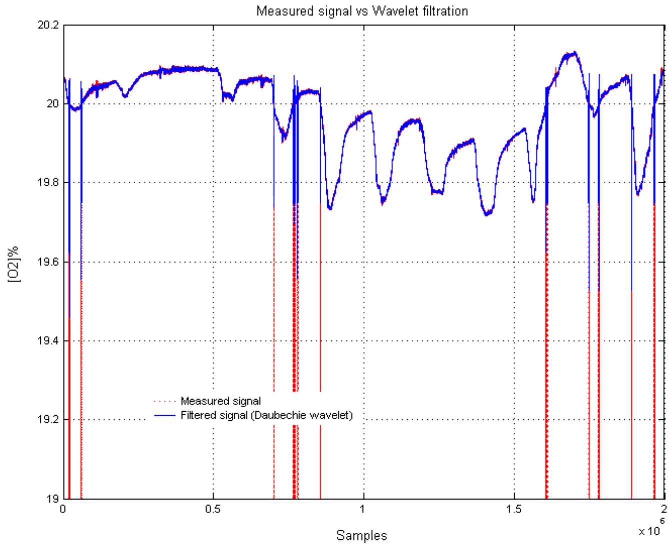
Starting from a mother-wavelet it is possible to cover all time-frequency domain through its successively dilatation and translation, which is mathematically translated by the following equation, where Ψ is the mother wavelet, t is time, s is the scale and u is time delay or translation unit:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$$

The scaling function exhibits some notorious advantages, especially when we are interested in details located in the high frequencies and consequently masked by white noise. Scaling function is also the responsible for the large application of wavelets, in digital signal processing, audio processing, compression data and image processing. The scale has an equivalent in the frequency and the spectrogram in the windowed Fourier transform has also its equivalent in terms of wavelet transform, which is designed by scalogram (Graps, 1995). Although, the scalogram is useful as an energy density function, its calculation is such a hard computational task that is rarely used. Nevertheless, we used a Daubechie wavelet to decompose the signals in study in a time-scale domain. In both signals, oxygen concentration and temperature, it was observed a straight concordance between the spectrogram and the wavelet transform that we called pseudo-scalogram. As an example next figure shows the decomposition of a signal to the 4^{th}. level.



The wavelets theory provides another powerful tool in denoising signals allowing them to be used in subsequent mathematical studies and simultaneously reducing the amount of data. It was applied in the filtration of oxygen concentration, generating an amount of data substantially reduced. This filtered data, either as input or output, can be used in stochastic models of biodegradation using the System Identification tools.



APPLICATION OF SYSTEM IDENTIFICATION AS A STOCHASTIC MODEL OF BIODEGRADATION

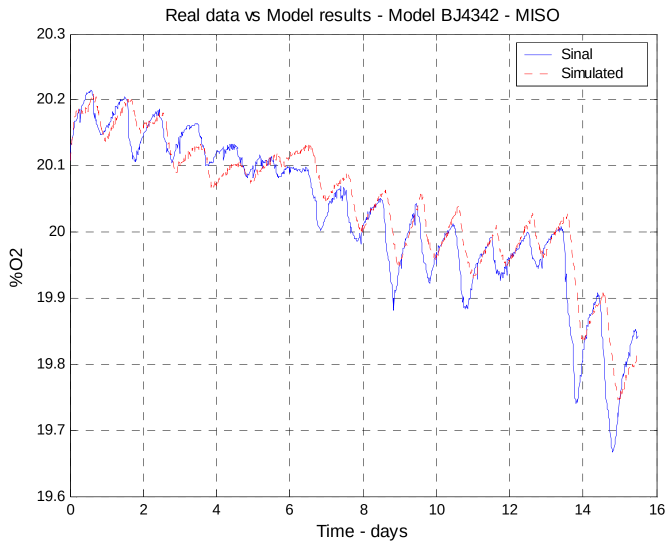
The different phases of biodegradation can be represented by the black box stochastic mathematical models used in System Identification Theory, either SISO, MISO or MIMO. The input variables can be the oxygen concentration in the external atmosphere or the temperature, and the output variables are oxygen or carbon dioxide concentrations.

Each SISO model has a general form

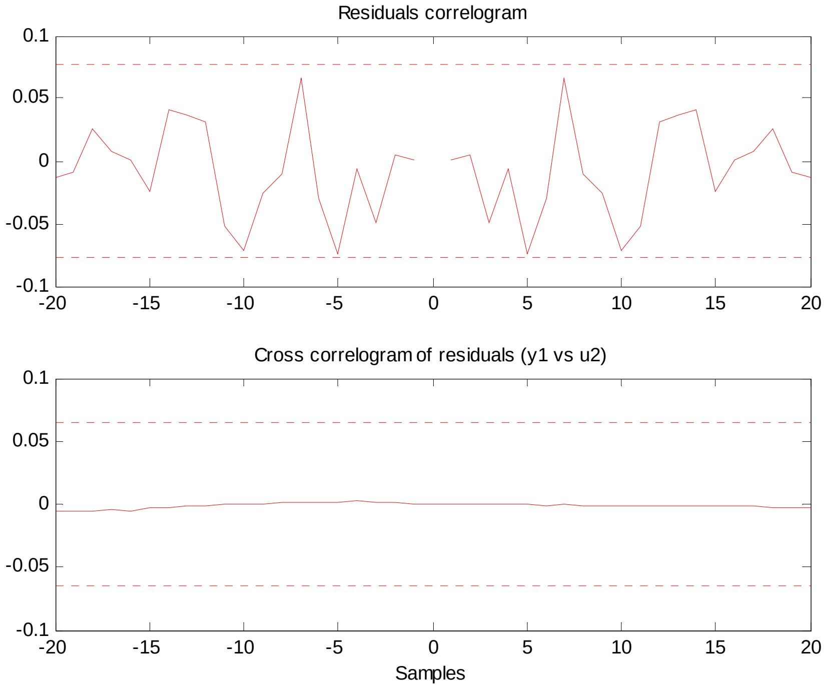
$$y(t) = \frac{B(q)}{F(q)} u(t-nk) + \frac{C(q)}{D(q)} e(t)$$

where y(t) is the output, u(t) the input, e(t) is a disturbance random variable, B(q), F(q), C(q) and D(q) are polynomials ordered in descending powers of the shift operator, represented as q. For instance $A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + ... + a_n q^{-n}$.

In the present example a MISO system (two inputs, one output) was chosen. The input data is the oxygen concentration at the inlet and the inner temperature and the output oxygen concentration at the outlet of the reactor; these values were filtered using moving averages in order to reduce the size of the available information. In other examples we used carbon dioxide as output variable.



The fitness of the model can be weighted by the residuals correlograma or by cross correlation.



CONCLUSIONS

We try to illustrate in this article the enormous potential of respirometry as a mean to study biodegradation in detail, although we only exemplify shortly part of the research we have conducting. From the data it is also possible to analyse not only the kinetics, but also the time evolution of the stoichiometry, subjects that we don't specify in this article.

References

Atlas, R.M., *Microbial Degradation of Petroleum Hydrocarbons: an Environmental Perspective, Microbiological Reviews*, vol. 45,1, 180-209, 1981

Fiúza, A., Vila, C.2003. *Bioremediation of a Sandy Soil Polluted by Petroleum Hydrocarbons – Respirometric Studies*, *Proceedings of the second European Bioremediation Conference*, Greece, 55-58.

Fiuza Antonio, Vila Cristina, *An insight into soil bioremediation through respirometry*, Environment International, 31 (2005), 179-183.

Graps, A., 1995. *An introduction to wavelets*, *IEEE Computational Science and Engineering*, 2, vol. 2.

Graves, D. A., Lang, C. A., Leavitt, M. E., 1991.Respirometric analysis of the biodegradation of organic contaminants in soil and water, *Applied Biochemistry and Biotechnology*, vol 28/29, 813-826.

Mallat, S., 1999. *A wavelet tour of signal processing*, 2nd ed. , Academic Press.

Mardia Kanti, Jupp Peter, *Directional Statistics*, John Wiley & Sons, 2000.

Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J-M., *Wavelet Toolbox User's Guide*, The MathWorks, 2000.

Shumway, R.H., Stoffer, D. S., 2000. *Time series analysis and its applications*, Springer-Verlag.

Bakken, George, 1991. *Time-Resolved Respirometry: Equations for the Simultaneous Measurement of All Respiratory Gases and the Calibration of Oxygen Consumption using Variable Inert Gas Flow Rates*, *J. Therm. Biol.* Vol.16, No. 6, pp 313-315.

Ljung Lennart, 1987. *System Identification Theory for the User*, Prentice Hall.

Vila Cristina, *Soil Remediation Observed through Respirometry: Signal Analysis in Biologic Systems*, Ph.D. Thesis, University of Porto, 2004.

Zimakowska-Gnoińska, D., Bech, J., *Estimation of the soil respiration using the constant pressure volumetric respirometer and the flow-through respirometer* UNI-RES10, *Environmental Monitoring and Assessment*, 64, 583-590., 2000.