20

Statistical Tools and Optoelectronic Measuring Instruments

Ionel Sabin and Ionel Ioana
Universitatea "Politehnica" Timişoara
Romania

1. Introduction

In the frame of European research projects, several air quality measuring campaigns in cross roads, streets, parks as well in a non ecological waste deposit were realized. The analyzed signals, representing CO, NO_2, O_3, SO_2 and HC concentrations, were measured with several optoelectronic instruments. Two of the utilized optoelectronic devices are shortly presented at the beginning of the chapter.

Due to their random character, pollutant concentrations signals can be analysed using statistical processing methods. The main statistical functions and parameters taken into account within this chapter are histograms, correlation coefficients, correlation and covariance functions (Ionel et al., 2009). Actually, statistical tools are usually utilized in analysing ecological data (Zuur et al., 2007) but, as far we know, it is not common to imply statistics in a comparative analysis of optoelectronic devices (Ionel et al., 2007).

Specific pre-processing procedures must be used for signal conditioning. Thus, „ideal" low-pass filtering based on fast Fourier transform can be implemented for the rejection of measurement noise and artefacts from the pollutant concentration signals. On the other hand, „ideal" high-pass filtering allows the extraction of the variable component of the pollutant level signals. In order to avoid redundant measurements, one can use interpolation for increasing the number of samples, especially in the case of slowly varying meteorological parameters.

Computer experiments with real pollutant concentration signals lead to some practical recommendations concerning acquisition parameters like data size and sampling frequency. The most important practical rules are as follow: assure the temporal length of the measured signal, assure the necessary resolution on the time axis, and make interactive verifications of the acquisition parameters during the measuring campaign. Guidance on MATLAB software for calculating statistical functions and parameters are provided.

As a particular application, the correlative comparison of two carbon monoxide (CO) measuring instruments is presented. The point source device and the open path optical remote sensing instrument do actually not measure the same quantity but a statistical comparison of the two instruments is still possible. The correlative analysis leads to the expected conclusion that the open path instrument is more suitable for monitoring the pollution level in a large area than the classical point source device.

2. Pollutant concentrations measured with optoelectronic instruments

2.1 The optoelectronic measuring instruments

One of the utilized instruments was the specialized HORIBA APMA-350E CO monitor, which furnishes the local pollution level. Fig. 1 presents a bloc diagram of this instrument working on the classical Non-Disperse Infrared (NDIR) method. The APMA-350E represents a generation of ambient CO monitors designed to eliminate routine calibration cycles and to provide long-term stable measurements and unattended continuous operation. It features a newly developed cross-flow modulation (CFM) technique which results in remarkable improved zero drift performance and sensitivity. The cross-flow modulated analyzer incorporates the basic design features of the conventional NDIR analyzer.

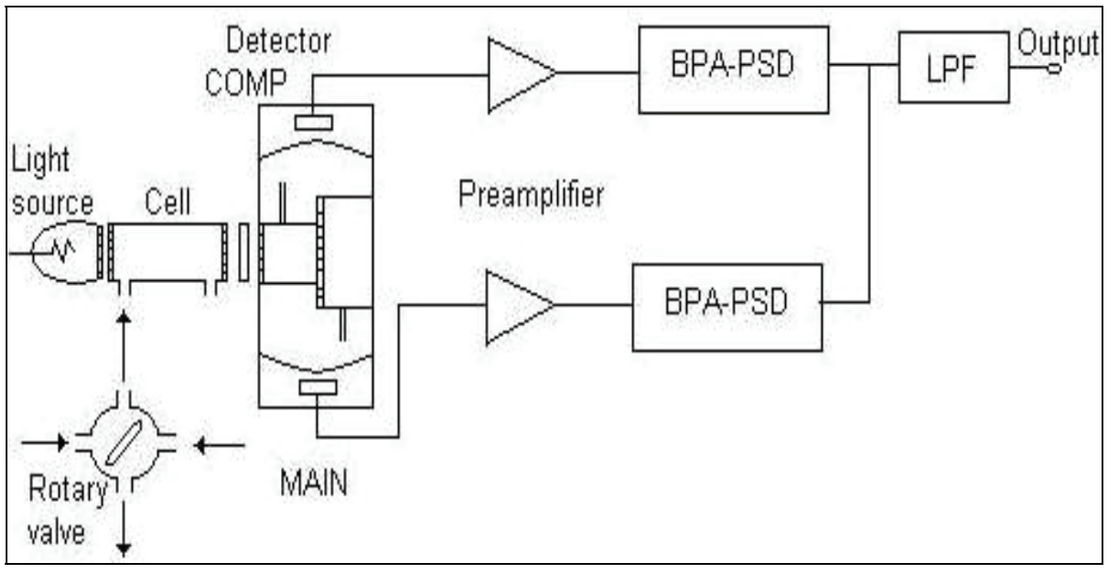


Fig. 1. Bloc diagram of the HORIBA APMA-350E CO monitor

The essential new element in this design, according to Fig. 1, is a rotary valve that alternately directs the sample gas and a reference gas to the one cell of the analyzer. By this method, the distinction between the sample and the reference optical path is eliminated and each path alternately functions as a reference and a sample path. The requirement for an optical chopper to modulate the detector output is thereby eliminated. In the cross-flow analyzer design, sensitivity is inherently increased because the amount of IR (infrared) energy absorbed and translated into the output signal is theoretically doubled for any concentration at the given modulation frequency. In addition, the signal-to-noise ratio is significantly better because the optical chopper which tends to introduce noise in the conventional NDIR instrument is removed in this CFM design. In the CFM scheme, gas flow rates and cell configuration can be selecting providing very smooth modulation. To minimize interference, dual detector system employing a compensating detector located behind the main detector is adopted in this instrument. The two detectors are charged in such a way that response to the interference gas in the second detector is compared to that of the measured gas. The signal from this detector is amplified and subtracted from the main detector signal, in the electronic part of the analyzer.

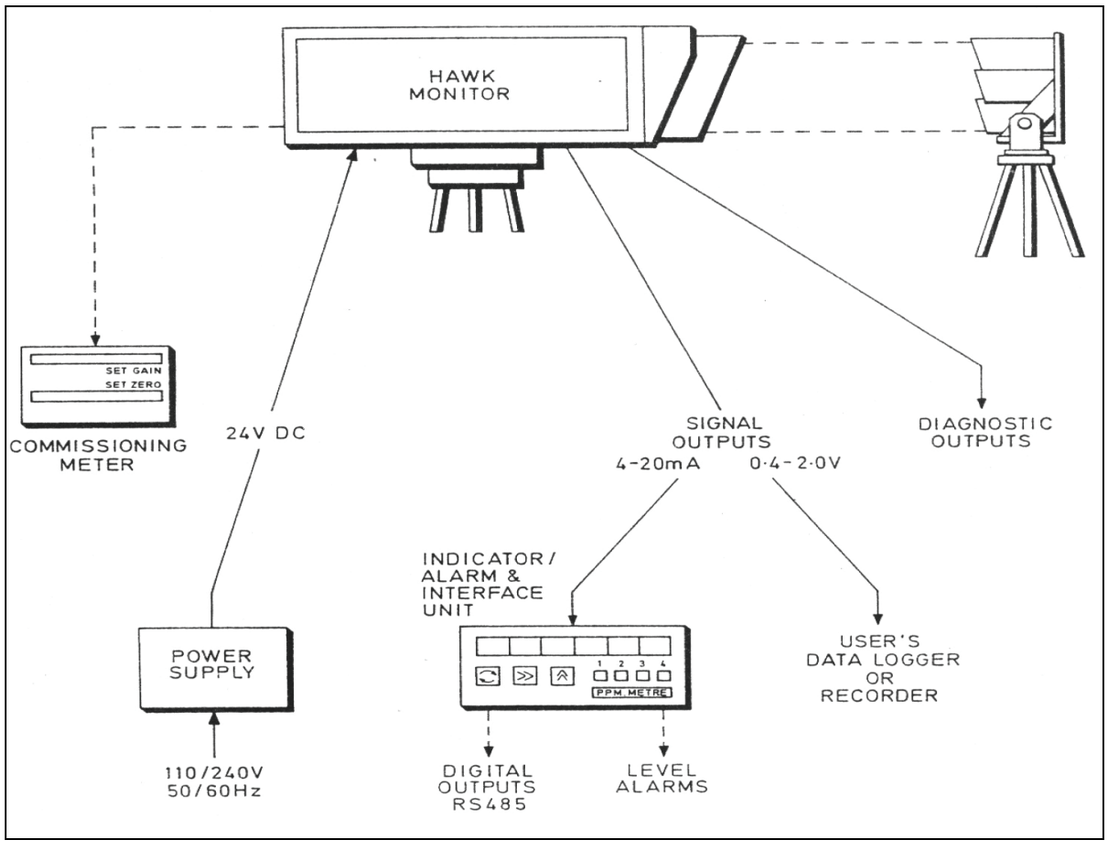


Fig. 2. A schematic diagram of the IR DOAS HAWK instrument

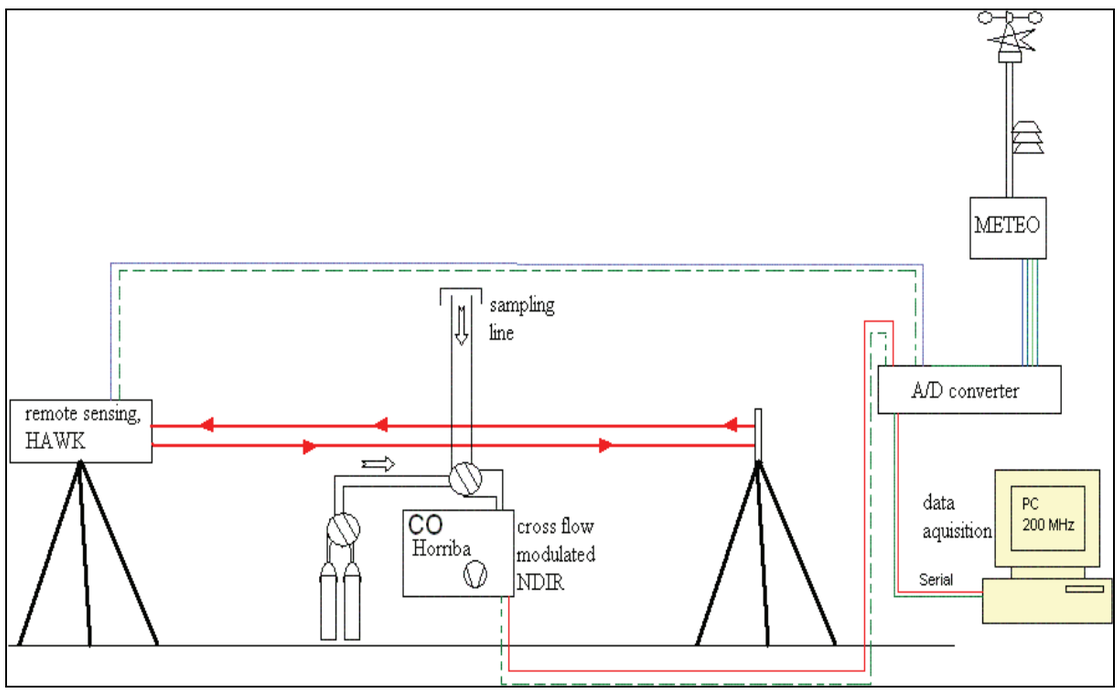


Fig. 3. Measuring setup with SIEMENS-HAWK and HORIBA-APMA 350E instruments

The second utilized instrument was an IR HAWK system from Siemens Environmental Systems, with the schematic diagram presented in Fig. 2. This instrument is an IR DOAS (Differential Optical Absorption Spectroscopy) apparatus, which can be configured to detect several species of pollutants including carbon monoxide. The beam path can be up to 400m and detection is typically better than 50ppb. The HAWK system works by measuring the absorption of infrared radiation passing along the instrument beam path by the gas to be measured. The system consists of a monitor, which contains the source and the detector unit, and a reflector. The total path length is therefore twice the distance between monitor and reflector. The source emits over a range of wavelengths and the beam is modulated after generation. The beam is reflected back to the monitor where it is filtered at a wavelength specific to gas of interest. The filtered beam is focused onto a detector which compares filtered and unfiltered reflected light in order to measure the concentration of the target gas.
Open path techniques have an advantage over the point source detectors: the sample volume is mach greater, the non-uniformity of the sample is eliminated and a more representative value of the concentration to be measured is obtained. Under field conditions, the degree of mixing is affected by the local environment, primarily, wind and thermal gradients.
Fig. 3 shows a typical relative setup for the HAWK and HORIBA analysers. One should observe also the meteorological mast, which continuously sent data (15 minutes mean values) to the general data acquisition system.

2.2 Measured and conditioned signals

The CO-concentration signals were measured with both HAWK open path monitor and the HORIBA point monitor at a sampling period of 6 seconds. This corresponds to a sampling frequency $f_S = 600$ cycles/hour and a maximal (Nyquist) frequency of $f_{\max} = 300$ cycles/hour. Each signal contains 4950 samples expressed in $[\,\mathrm{mg/m^3N}\,]$ as represented in Fig. 4. The total registration length covers 8 hours and 15 minutes. Obviously, both signals have a non-stationary random character.

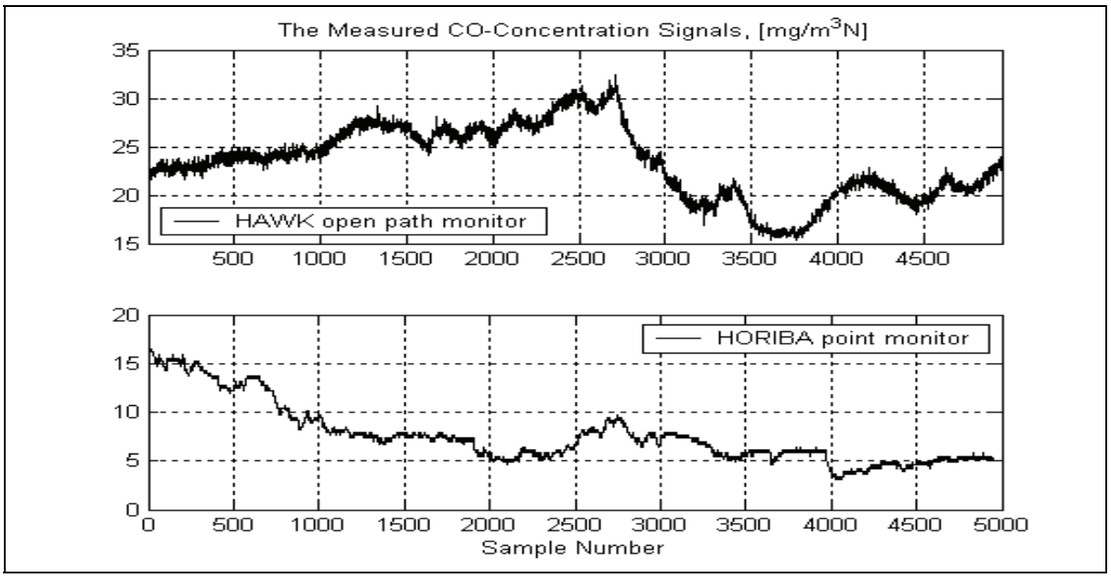


Fig. 4. CO-concentration signals measured with the HAWK open path and the HORIBA point monitor, respectively

During the same time interval of more than 8 hours, some meteorological parameters were also measured. Two of them, namely temperature, in [°C] , and wind velocity, in [m/s], are represented in Fig. 5. The direction of the wind were measured and utilized in the determination of the wind component parallel to the optical axis of the HAWK open path monitor as well the component perpendicular to this axis. The two wind components, in [m/s], are also represented in Fig. 5.

The sampling period for the meteorological parameters was 15 minutes, so that each of the meteorological parameters is determined through 33 values. The corresponding sampling frequency of 2 cycles/hour is a good choice for a slowly varying quantity like temperature, but can become critical in a turbulent environment with rapid changing wind direction or intensity. One can appreciate that, in our case, the sampling frequency was great enough for temperature and for wind velocity as well.

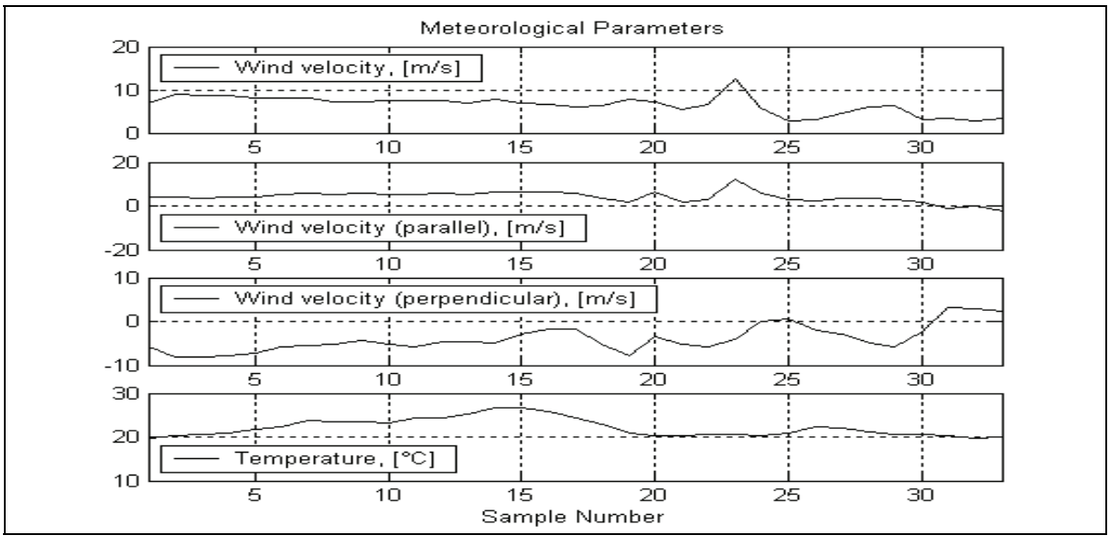


Fig. 5. Meteorological parameters: wind velocity and temperature

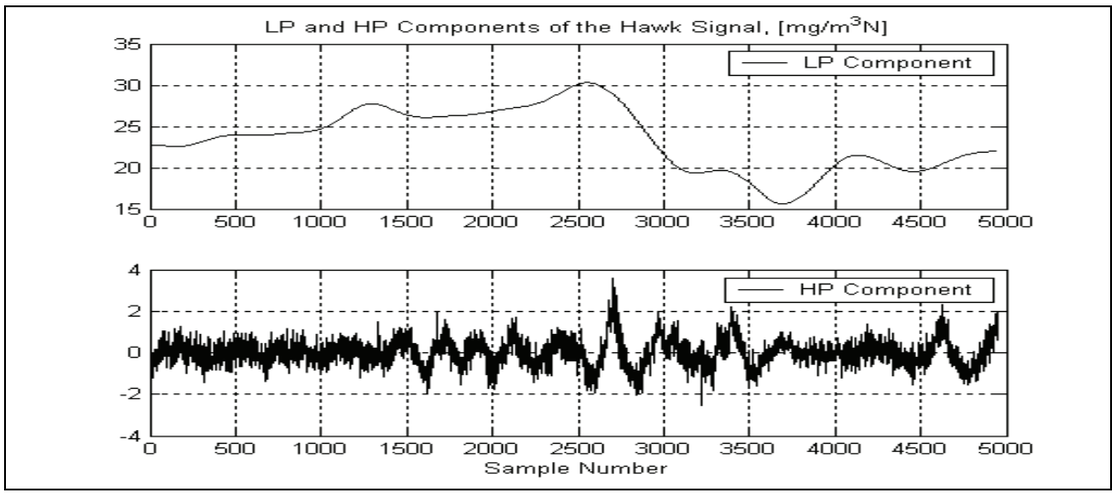


Fig. 6. LP and HP components of the CO-concentration signal measured with the HAWK instrument

Certainly, in a routine measurement choosing very different sampling frequencies for CO-concentrations and meteorological factors is not justified. In our case, the high sampling frequency of the CO-concentrations allows a characterization of the noise associate with these measurements. On the other side, choosing a higher sampling rate for the slowly varying meteorological parameters could be equivalent with storing a large amount of redundant data.

In order to separate the stabile local mean value of the CO-concentration from the measurement noise, an ideal low-pass (LP) and high-pass (HP) filtering of the signals were performed. The two components of the signal measured with the HAWK instrument are represented in Fig. 6. By „ideal" filtering we mean the infinitely sharp frequency characteristic at the cut-off frequency so that the sum of the HP and LP components gives exactly the values of the original unfiltered signal.

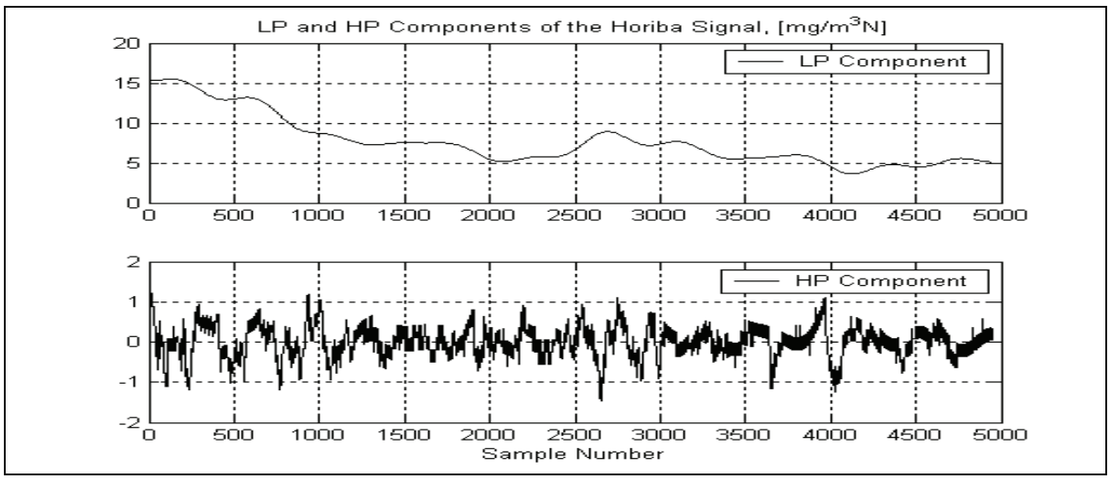


Fig. 7. LP and HP components of the CO-concentration signal measured with the HORIBA instrument

Practically, the ideal filtering was implemented through a direct fast Fourier transform (FFT) followed by a windowing the obtained spectrum with the desired filter characteristic and finally a reverse FFT to obtain the time domain representation of the LP or HP component of the CO-concentration signal.

The cut-off frequency of the LP filter representing also the corner frequency of the HP filter, was empirically chosen to be 2 cycles/hour. So the LP component contains frequencies between 0 and 2 cycles/hour while the HP component covers the range from 2 cycles/hour to 300 cycles/hour. Similar observations are valid for the LP and HP components of the CO-concentration measured with the HORIBA instrument. These are represented in Fig. 7.

Other signal processing techniques, like wavelets or short-time spectra can be used to separate the HP and LP components of the Pollution level signals. However ideal filtering based on Fourier transformation is simpler and very efficient.

The power of the measurement noise, calculated as the mean value of the HP components of the CO-concentration proves to be greater in the case of the HAWK instrument, than for the HORIBA device: $P_{HAWK} = 0.4151$ in comparison with $P_{HORIBA} = 0.1455$. This relation can be observed also in the graphical representation of the unfiltered signals, in Fig. 4. Experiments show that measurement noises may introduce an up to 5% error in the determination of the

correlation coefficients. This effect could be neglected. However, the separation of LP and HP components of the CO-concentration signals can be a very useful signal conditioning step which allows a simple elimination of possible artifacts in the measured pollution levels.

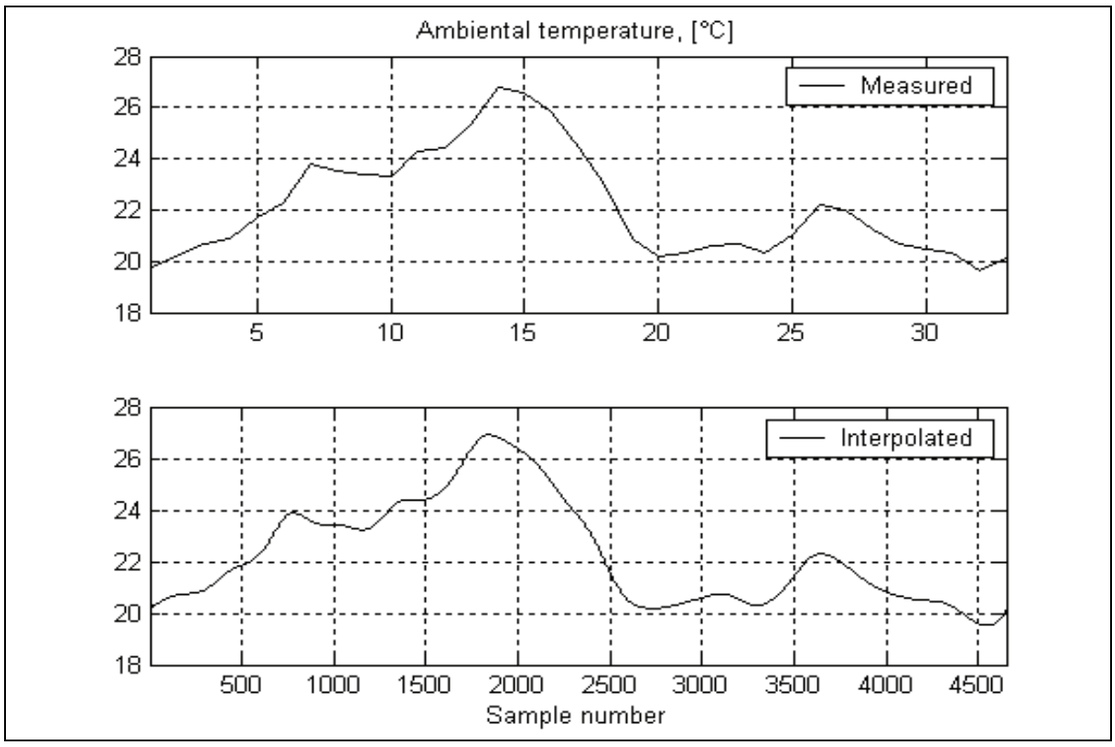


Fig. 8. Ambient temperature: measured and interpolated

Another pre-processing step is concerning the measured meteorological parameters. In order to calculate the correlation coefficients between CO-concentration signals on one side, and the meteorological parameters on the other, we must have the same number of samples in every signal. Therefore, the temperature and wind velocity signals were interpolated using a cubic spline procedure; the number of samples was increased from 33 to 4950. Due to the great interpolating errors at the beginning and the end of the signals, the first 150 and the last 150 samples of the interpolated signals were rejected. For example, Fig. 8 presents the ambient temperature signal in both forms, measured and interpolated. The first and the last 150 samples from the LP components of the CO-concentration signals were also eliminated. Finally, all signals, representing the pollution level, as well as the meteorological factors have the same length: 4650 samples.

3. Histograms and related parameters

The histogram is a representation of the signal amplitudes according to their values, regardless of the time variable: on the pollutant concentration axis a number of equally spaced containers are defined and the histogram returns the number of signal samples in each container (Montgomery & Montgomery, 2006; Navidi, 2008; Peck et al., 2007). This function can be easily implemented using the MATLAB software (Hoffmann & Quint, 2007; Martinez, & Martinez, 2002). Namely, *N=hist(Y)* puts the elements of *Y* into ten equally

spaced containers and returns the number of elements in each container while *N=hist(Y,M)*, where *M* is a scalar, uses *M* bins.

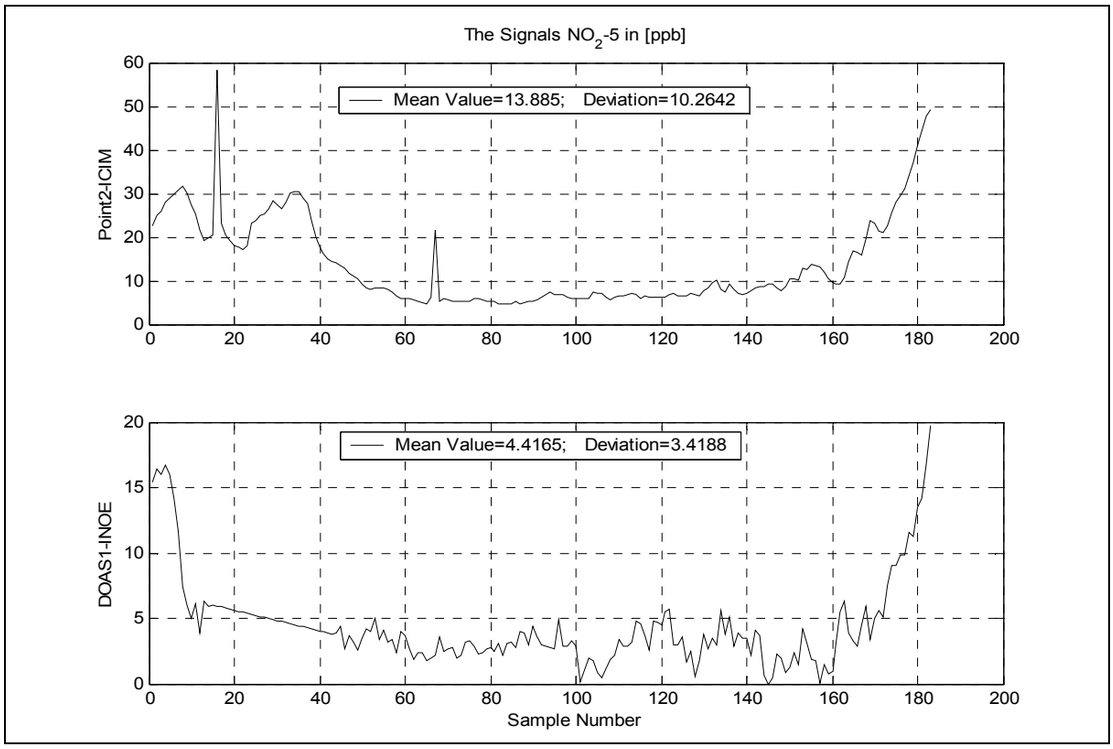


Fig. 9. Two NO_2 signals measured simultaneously

The histogram can be a useful tool for a preliminary characterisation of the pollutant concentration signals, complementary to there usual temporal representation. For example, Fig. 9 presents two NO_2 concentration signals, each containing 183 measured samples. The corresponding histograms of the same signals can be seen in Fig. 10. The realisations represented in Fig. 9 contain the full information about the measured concentrations. Particularly, one can observe the similarity of there evolution. The signal shown in the upper panel has two artefacts which must be eliminated during the pre-processing step. However, the soft limitation of this signal appears with greater clarity in the corresponding histogram from Fig.10. At the same time, the quasi normal (Gaussian) character of the signal represented in the lower panel is more evident in the corresponding histogram then in the temporal representation. The good continuity of both histograms suggests a sufficient size of the measured signals.

The normalized histogram (with area equal to one) is an approximate representation for the probability density function of the analysed signal. Some statistical parameters related to the histogram (through the probability density function) can be utilized for a quantitative or qualitative characterisation of the pollutant concentration signals. Thus, the mean value (Martinez, & Martinez, 2002; Montgomery & Montgomery, 2006; Papoulis, 1991; Therrien, 2007)

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

and the standard deviation

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \tag{2}$$

usually reported together, as in Fig. 9 and Fig. 10, offer a quantitative description for the location of the pollutant concentration values. Two other statistical parameters can be utilized for a rather qualitative characterisation of the data set in comparison with a normal distribution. Namely, the skewness (the third central moment divided by the cube of the standard deviation)

$$s = \frac{1}{N-1}\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3}{\sigma^3} \tag{3}$$

describes the lack of symmetry of the histogram. One can mention that for any symmetrical distribution (e.g. the normal distribution), the skewness is zero.

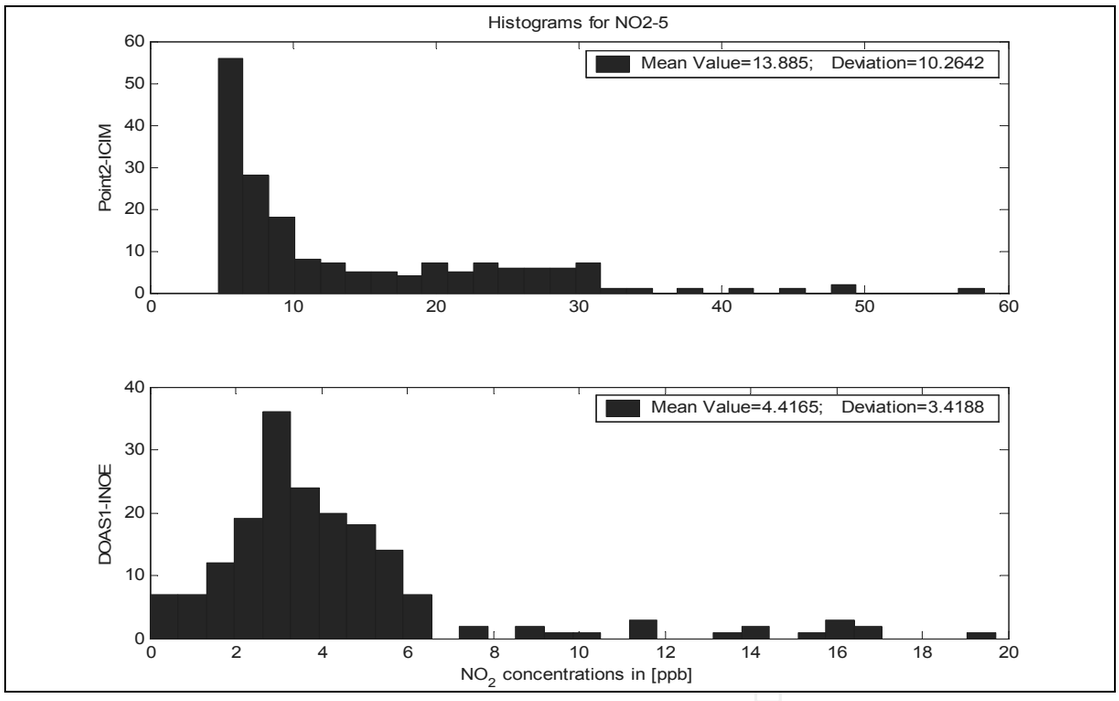


Fig. 10. Histograms of two NO_2 signals

The kurtosis (the fourth central moment divided by fourth power of the standard deviation)

$$k = \frac{1}{N-1}\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^4}{\sigma^4} \tag{4}$$

offers a characterization of whether the histogram is peaked or flat relative to a Gaussian distribution. The kurtosis of a normal distribution equals three. Pollutant concentration signals with high kurtosis have a sharp peak near the mean value while signals with low

kurtosis have a rather flat top near the mean value. Obviously, in (1), (2), (3) and (4), x_i is the current sample from the N samples of the pollutant concentration signal.
The MATLAB functions *mean(X)*, *std(X)*, *skewness(X)* and *kurtosis(X)* return the sample mean, sample standard deviation, sample skewness and sample kurtosis, respectively (Hoffmann & Quint, 2007; Martinez, & Martinez, 2002). For matrices, the results of this functions are row vectors containing the sample mean, standard deviation, skewness respectively kurtosis of each column. The commands *mean(X,DIM)* and *std(X,FLAG,DIM)* takes the mean respectively the standard deviation along the dimension *DIM* of *X*. If *FLAG=0*, *std* normalizes by (*N-1*), otherwise *std* normalizes by *N*.

4. The correlation coefficient

4.1 Definitions and practical rules for acquisition parameters

Taking into account the usual lack of stationarity of pollutant concentration signals, the correlation coefficient is, probably, the most useful statistical tool revealing possible influences between such data sets. Particularly, the (Pearson) correlation coefficient (Papoulis, 1991; Therrien, 2007; Zuur et al., 2007) is defined by the following relation:

$$r = \frac{\frac{1}{N}\left[\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \cdot \sum_{i=1}^{N} y_i\right]}{\sigma_x \sigma_y} \tag{5}$$

where $\left[x_i\right]$ and $\left[y_i\right]$, $i = 1,2,...,N$ are two simultaneously measured pollutant concentration signals, with standard deviations σ_x and σ_y, respectively. One can prove that $|r| \le 1$. The variability of the two data sets can be interpreted according to the range including the particular value of the correlation coefficient. Thus, $-1 \le r \le -0.33$ denotes negative correlation, while $0.33 \le r \le 1$ signifies that the two signals are positive correlated. If $-0.33 \le r \le 0.33$, the two pollutant concentration signals are said to be uncorrelated (independent). Obviously, the defined domains for the correlation coefficient are, to some extent, arbitrary. More than this, in our interpretation of the estimated value of r, one must take care of the confidence interval of this parameter (Martinez, & Martinez, 2002; Papoulis, 1991; Shen & Lu, 2006). Practically, not only the value of the correlation coefficient but also his confidence interval must be in one of the three ranges in order to conclude that the analyzed signals are positive correlated, negative correlated or independent.
The MATLAB *corrcoef* function calculates the correlation coefficients. Thus, in the command *R=corrcoef(X)*, *R* is the matrix of correlation coefficients for the array *X* (each row of *X* is an observation and each column of *X* is a variable). This function has also the possibility to calculate the lower and upper bounds for a 95% confidence interval for each correlation coefficient (Hoffmann & Quint, 2007; Martinez, & Martinez, 2002).
Two acquisition parameters of the pollutant concentration signals are important for a reliable determination of the correlation coefficient: the sampling frequency and the number of samples in the measured data sets. In order to establish some practical rules for the acquisition phase of future pollutant concentration measuring campaigns, several signal processing experiments were carried out. Thus, Fig. 11 presents the evolution of the correlation coefficient calculated for a moving segment of data, containing only 30 samples measured with a sampling period of five minutes. (It is meaningless for this analysis that the two signal pairs were designated as Point1UPT-DOASINOE and Point1UPT-POINT3INOE,

respectively). However, the huge variation of the correlation coefficients shows that the size of the data is totally insufficient, at least for this sampling frequency.

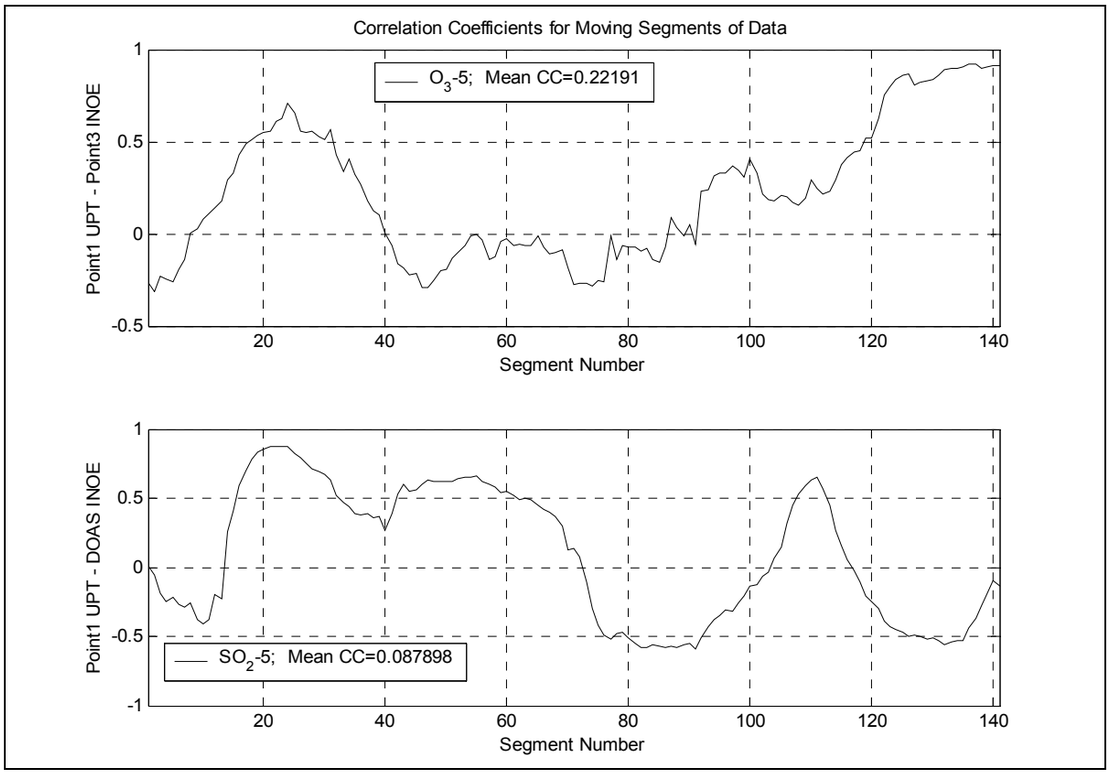


Fig. 11. Correlation coefficients for moving segments of data

Fig.12 shows the dependence of the correlation coefficient on size of the measured pollutant concentration signals. The correlation coefficient shown in the upper panel has a transient behaviour. The final value of the correlation coefficient proves that the two signals are strongly positive correlated. Just 60 measured samples of the signals could be sufficient for a good determination of the correlation coefficient for these signals, achieved at a sampling period of five minutes. Fig. 12 shows also the oscillatory behaviour of the correlation coefficient between Point1UPT and DOASINOE signals measured with a sampling period of five minutes. The low values of the correlation coefficient, proves that the two pollutant concentration signals are practically independent. Taking the confidence interval of the analysed data sets into account, the relative high values of r for segment length between 100 and 140 are not high enough to draw the conclusion that the two signals are positive correlated. The final decay of the correlation coefficient for segment length greater than 140 definitively proves the independence of the two concentration signals.

A practical conclusion can be drawn from this experiment: the calculation of a correlation coefficient with transient behaviour is practically completed as soon as the value of r reaches the quasi-steady state; if the evolution of the correlation coefficient is characterized by small amplitude variations, an optimal segment length is difficult to establish; however, since the duration of a pollutant concentrations measuring campaign is measured in days while the calculation of the correlation coefficient takes a few seconds, the interactive signal processing, during the measuring campaign is highly recommended.

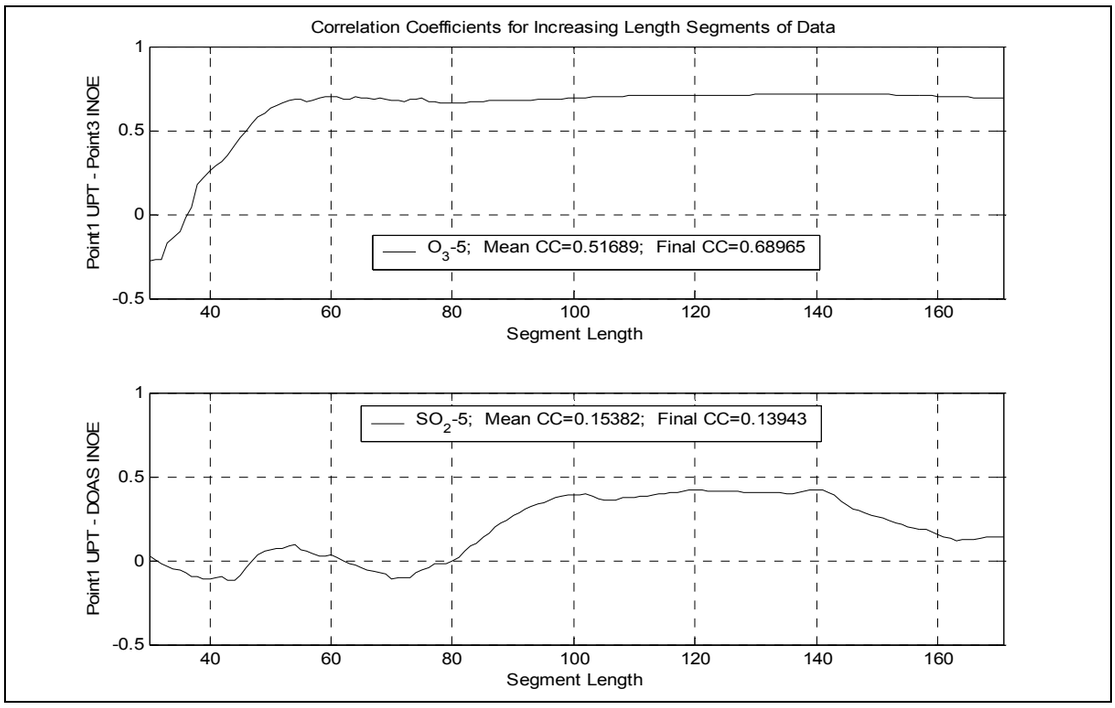


Fig. 12. The dependence of the correlation coefficient on the length of data segments

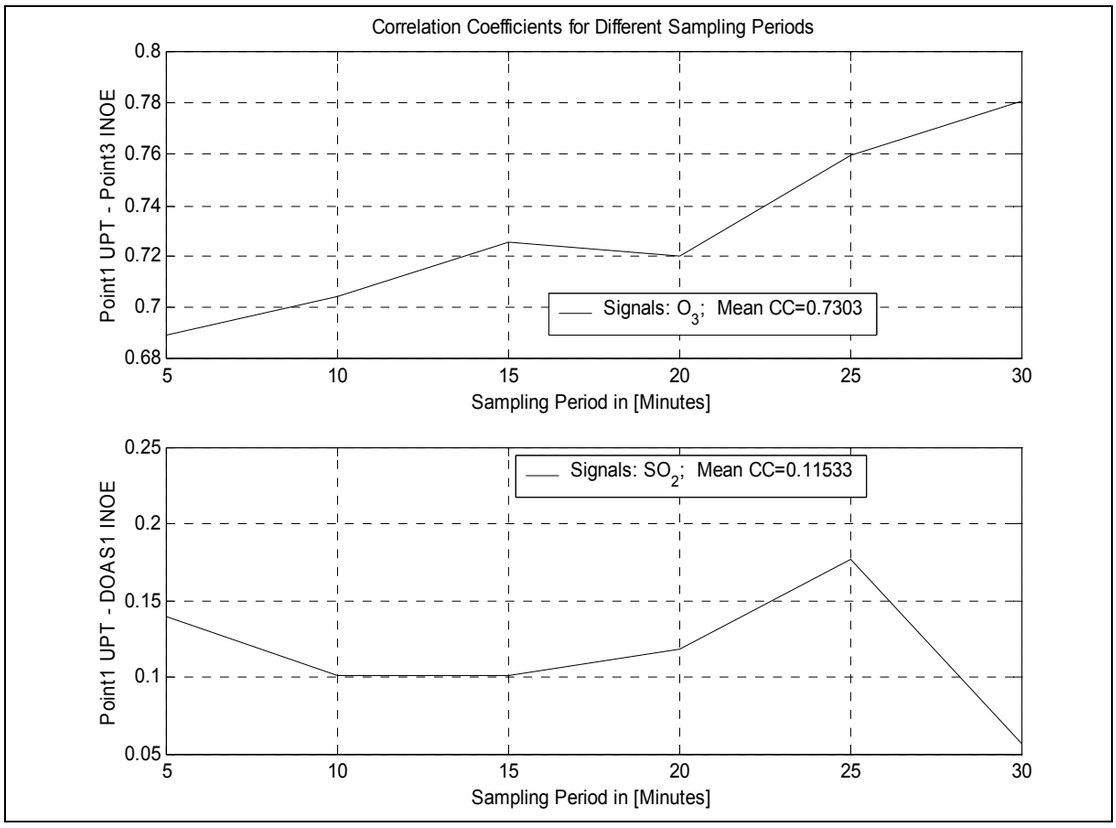


Fig. 13. Correlation coefficients calculated for different sampling periods

The dependence of the correlation coefficient on the sampling period is illustrated in Fig.13. The correlation coefficients are presented for two signal pairs achieved at a sampling period of 5, 10, 15, 20, 25 and 30 minutes, respectively. Corresponding to the increasing sampling period, the number of samples utilized in the calculation of the correlation coefficient was 172, 86, 57, 42, 34 and 28, respectively, so that the temporal signal length is maintained approximately constant (about 850 minutes). Comparing to the whole range [-1 to +1] of possible values for the correlation coefficient, the variations of the results shown in Fig.13 are small. This proves a high positive correlation between Point1UPT and Point3INOE signals (correlation coefficient about 0.73) and statistical independence between Point1UPT and DOAS1INOE signals (correlation coefficient about 0.11).
The practical conclusion of this experiment is that a reduced number of samples in the data sets can be compensated by a corresponding higher sampling period. This conclusion differs from the usual statement found in books on statistics, where the confidence interval of the correlation coefficient is determined by the number of sample in the analyzed signals. Our interpretation of this fact is that pollutant concentration signals are not artificial random signals generated by a certain mathematical rule, but physical signals where the generating mechanism can change in time. Consequently, one must give the signal enough time to show his features, and this can be done assuring a minimal temporal signal length, i.e. a minimal value of the product (Segment Size) · (Sampling Period). Certainly, the condition of a minimal number of measured samples (about 30 samples) must be also fulfilled.

4.2 Correlation coefficients Interpretation: a case study

The sample Pearson product-moment correlation coefficient, *r*, was computed for the following signals: LP component of the CO-concentration signal measured with the HAWK instrument (HAW, shown in Fig. 6), LP component of the CO-concentration signal measured with the HORIBA device (HOR, shown in Fig. 7), the wind velocity (W, shown in Fig. 5), the wind velocity component parallel with the optical axis of the HAWK instrument (Wpar, shown in Fig. 5), the wind velocity component perpendicular to the optical axis of the HAWK instrument (Wper, shown in Fig. 5), and the ambient temperature (T, shown in Fig. 8). Table 1 presents the approximate values (with only two decimals) of the determined correlation coefficients.

HAW	HOR	W	Wpar	Wper	T
1.00	0.23	0.45	0.33	-0.40	0.62
0.23	1.00	0.56	0.11	-0.68	-0.09
0.45	0.56	1.00	0.75	-0.73	0.23
0.33	0.11	0.75	1.00	-0.35	0.49
-0.40	-0.68	-0.73	-0.35	1.00	0.15
0.62	-0.09	0.23	0.49	-0.15	1.00

Table 1. Approximate correlation coefficients

In order to facilitate the interpretation, a diagram of the correlation coefficients is represented in Fig. 14. One can see a positive but small correlation coefficient between the two CO-concentration signals ($r \cong 0.23$). Related to the maximal possible value ($r \cong 1.00$), the actual correlation coefficient reflects on the fact that the two instruments do actually not measure the same quantity. The poor correlation proves the fundamental difference between a local pollution level (measured with HORIBA instrument) and spatial averaged CO-concentrations (from the HAWK system), not only as absolute values but also as variations tendencies.

The averaged signal measured with the open path remote sensing instrument is strongly positive correlated with the temperature ($r \cong 0.62$), while this meteorological parameter has practically no influence on the local pollution level ($r \cong -0.09$). This can be explained by the intensification of the activities in the non ecological waste deposit by daytime, fact clearly sensed by the open path instrument but not by the Horiba device which measures the pollution in a certain point, at soil level. The CO-concentrations measured with both instruments are moderate positive correlated to the magnitude of the wind velocity ($r \cong 0.45$ and $r \cong 0.56$ with the HAWK and HORIBA signals respectively). However, the component of the wind velocity perpendicular to the optical axes is negative correlated with both measured pollution levels. Due to the geometrical arrangement of the measuring systems this wind component tends to clean the air. As a global conclusion, the HAWK CO-concentration signal is better correlated with the meteorological parameters than the HORIBA signal.

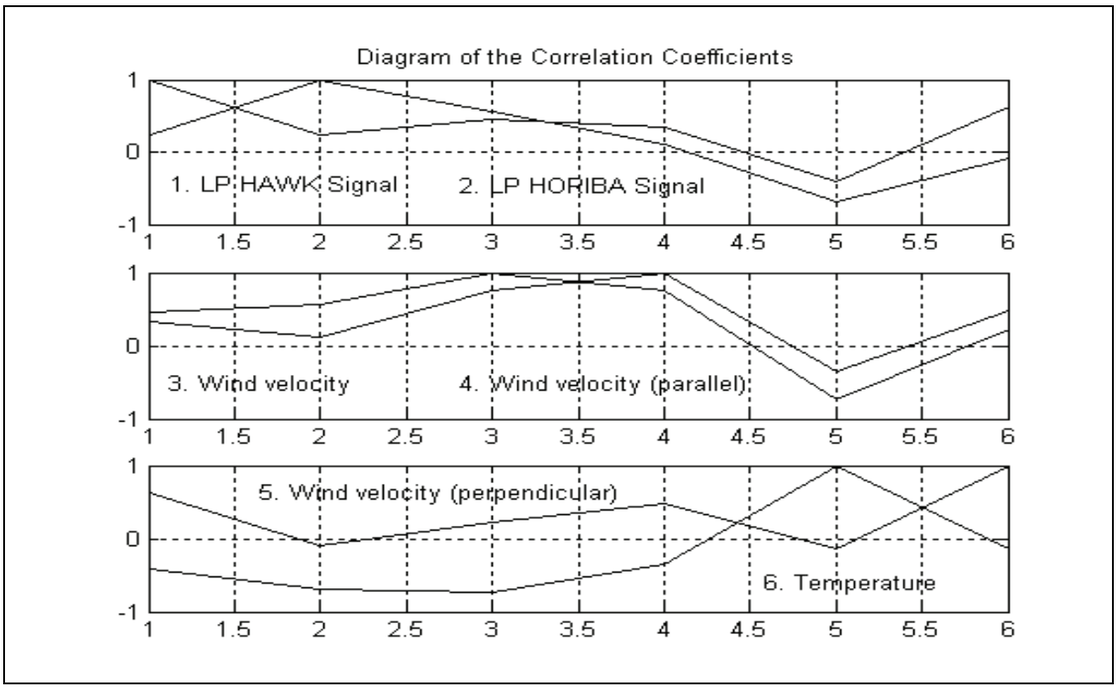


Fig. 14. Graphical representations of correlation coefficients

The 95% confidence intervals of the correlation coefficients were computed using a procedure based on the Fisher transformation (Shen & Lu, 2006). Exact values of the correlation coefficients between HAWK CO-concentration and the other five signals are

given in Table 2, together with their confidence intervals. Due to the great number of samples in each signal, the confidence intervals are narrow around the calculated correlation coefficients. Fig.15 gives a graphical image of the narrowness of these intervals.

HAW -95%	HAW	HAW +95%
1.0000	1.0000	1.0000
0.2039	0.2313	0.2583
0.4260	0.4492	0.4719
0.3085	0.3343	0.3596
-0.4281	-0.4043	-0.3800
0.6006	0.6187	0.6361

Table 2. Exact values of the correlation coefficients and their confidence intervals

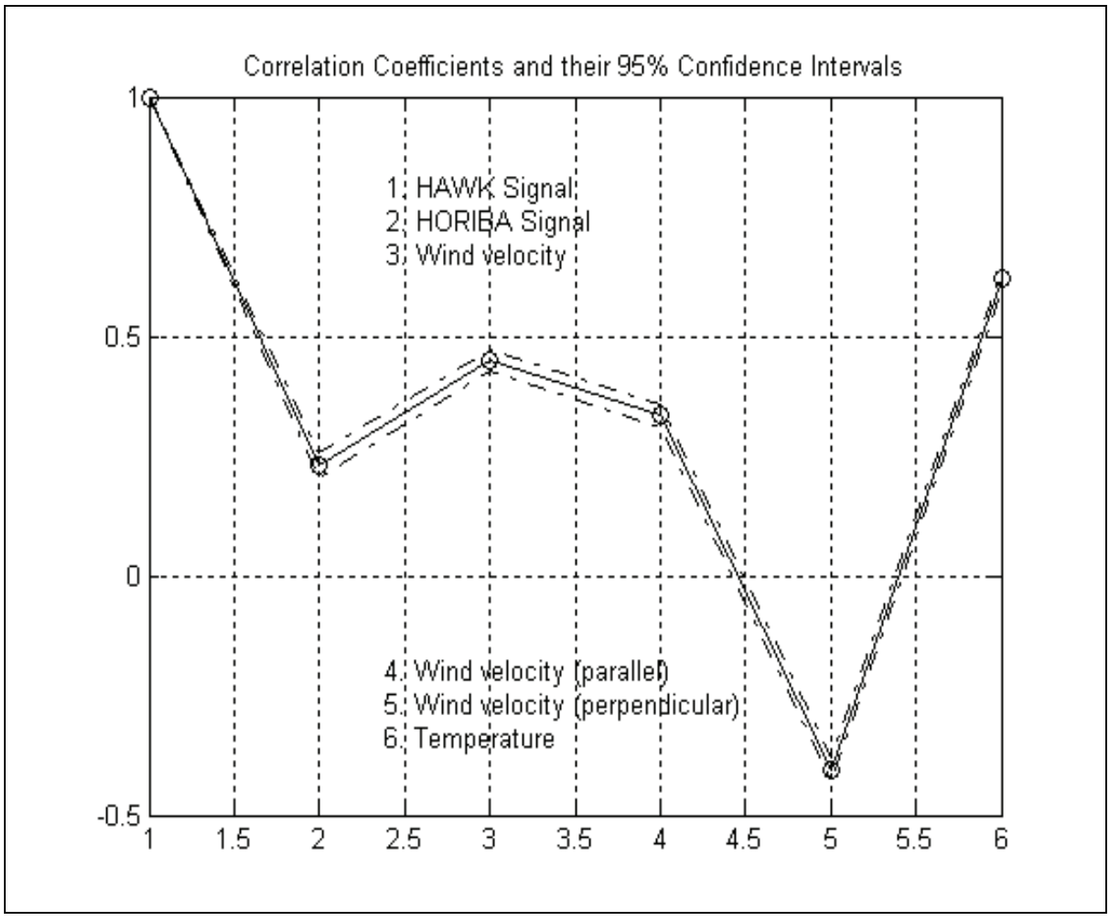


Fig. 15. Correlation coefficients of the HAWK signal (-o) and their 95 % confidence intervals (- ·)

5. Correlation and covariance functions

The correlation and covariance functions can put into evidence power and temporal relations between pollutant concentration signals. The sample crosscorrelation function is defined by the formula (Therrien, 1992):

$$R_{xy}[l]=\begin{cases}\dfrac{1}{N-l}\cdot\sum\limits_{i=0}^{N-1-l}x_{i+l}y_i\,;\ for\ 0\le l<N\\ \dfrac{1}{N-|l|}\cdot\sum\limits_{i=0}^{N-1-|l|}x_i y_{i+|l|}\,;\ for-N<l<0\end{cases} \tag{6}$$

can be calculated if the measured pollutant concentration signals are stationary, but this is rarely the case. However, for $[y_i]\equiv[x_i]$ we obtain the autocorrelation $R_{xx}[l]$ of the sequence $[x_i]$. Fig. 16 shows the autocorrelation functions for Point1UPT signal (in the upper panel) and Point3INOE (in the lower panel). The value of the autocorrelation at lag zero, $R_{xx}[0]$, gives the power of the corresponding signal: approximately 1024 ppb^2 for Point1UPT and 1078 ppb^2 for Point3INOE signals. After some experience, one can appreciate the spectral content of the signal from the shape of the autocorrelation function.

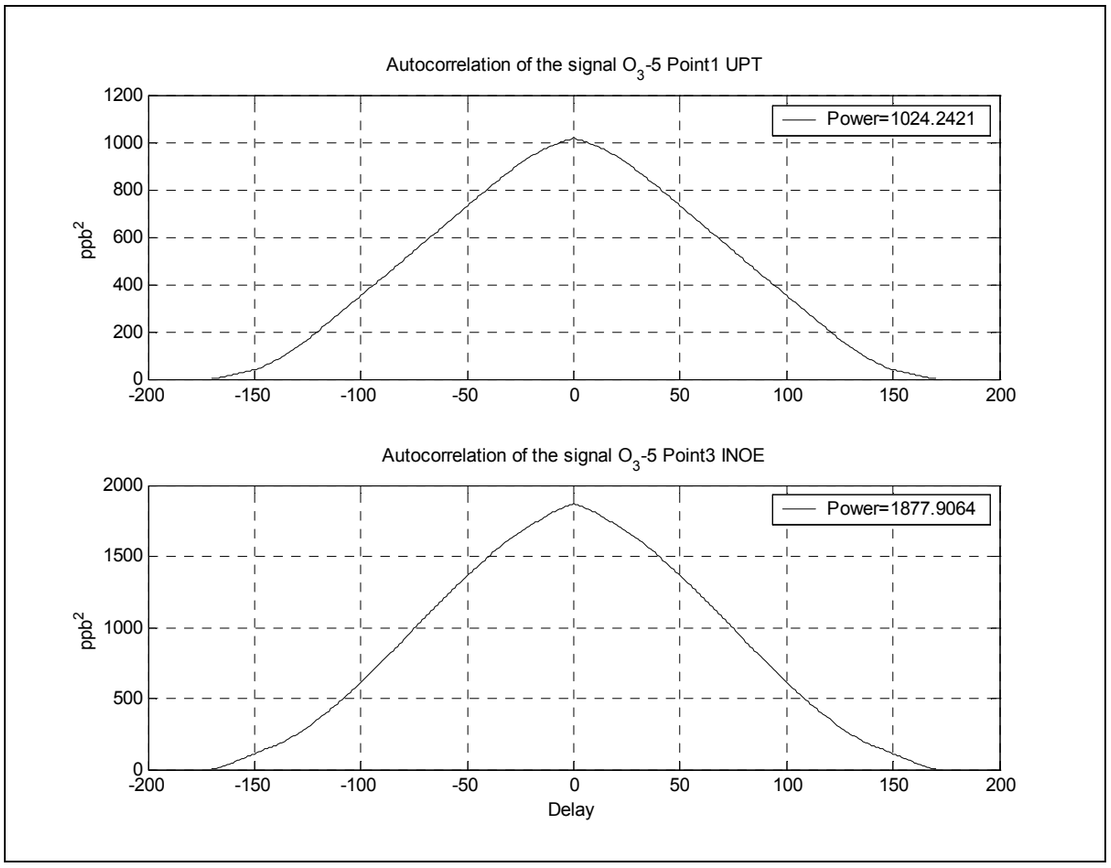


Fig. 16. Typical autocorrelation functions

The sample auto covariance, $C_{xx}[l]$, is the sample autocorrelation of the centred sequence $[x_i]$ (the sequence after removing the mean) while the sample cross covariance, $C_{xy}[l]$, is the cross correlation of the centred sequences $[x_i]$ and $[y_i]$. Thus, Fig. 17 presents the auto covariance functions for the same signals, namely Point1UPT and Point3INOE pollutant concentrations. The auto covariance at zero lag, gives the power of the variable part of the analysed signal: approximately 74 ppb^2 for Point1UPT and 131 ppb^2 for Point3INOE signals. The shapes of the auto covariance functions show that Point1UPT signal has a highly random character while in the Point3INOE signal the spectral components are concentrated in a small range near 0.016 mHz.

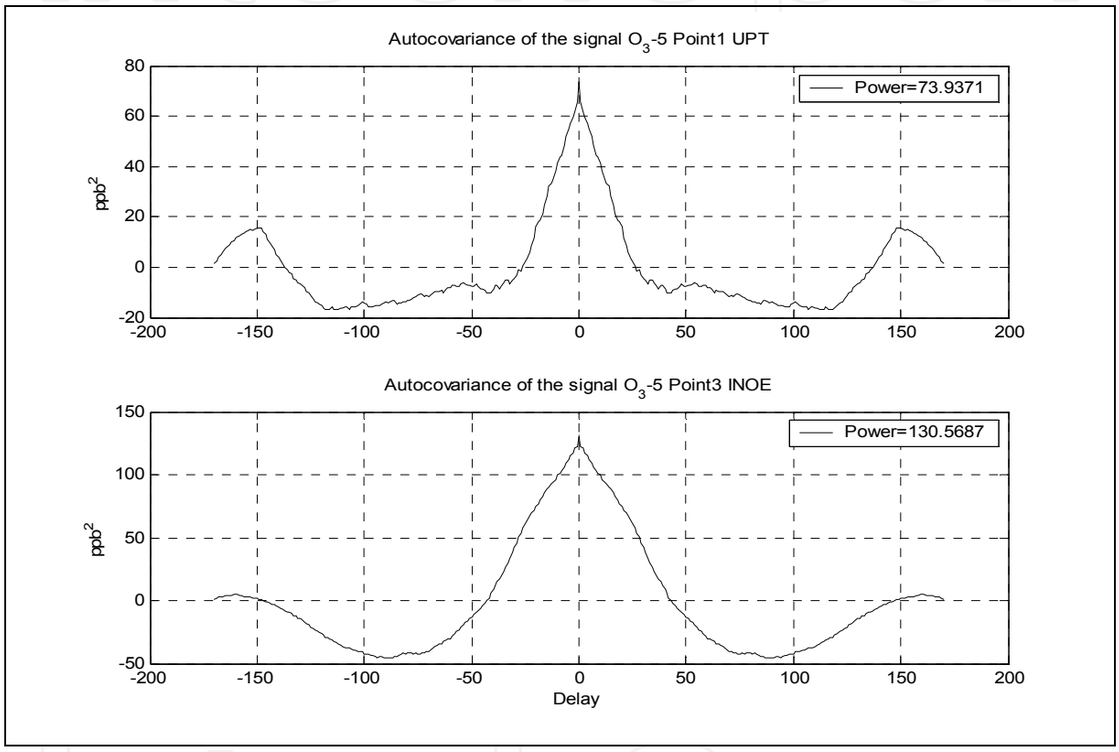


Fig. 17. Typical auto covariance functions

The cross-covariance functions offer an interesting possibility to determine delays between pollutant concentrations measured simultaneously. The position of the peak in the covariance function around zero gives the temporal delay between two signals. Depending on the wind direction and intensity the pollutants can be transported from one place to the other in the experimental area. However such temporal relations can be put into evidence only if the resolution on the time axis is sharp enough to allow the proper localization of the crosscovariance peak. This condition was not fulfilled during the related measuring campaign: the sampling period should be in the range of seconds while the signals were achieved with sampling periods of 5 to 30 minutes. The consequence is that the peak of the autocorrelation function appears in the origin (zero lag) or they have a flat maximum around zero, as shown in Fig.18.

Summarizing, the autocorrelation and auto-covariance functions are useful tools in establishing power relations between pollutant concentration signals measured with

optoelectronic instruments. For reliable results, a sufficient temporal length of measured data sets must be assured, so that the signals can manifest there features. Temporal relations i.e. delays between certain signals measured simultaneously can be revealed using the crosscovariance function. But, for this purpose a supplementary condition must be fulfilled: a small sampling period must assure a good resolution on the time axis.

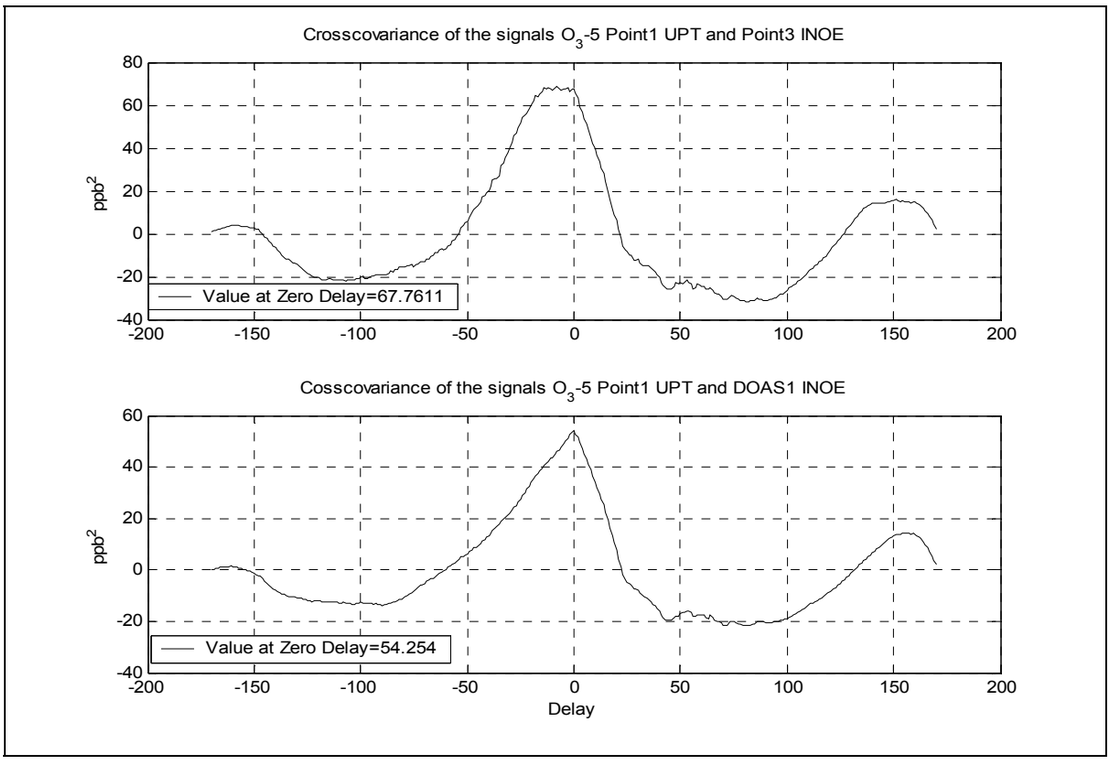


Fig. 18. Experimental cross covariance functions

The MATLAB *xcorr* function produces estimates of the cross-correlation function. For example, the command *C=xcorr(A,B)*, where *A* and *B* are length *M* vectors, returns the length 2**M*-1 crosscorrelation sequence *C*. Particularly, *C=xcorr(A)*, where *A* is a vector, returns the autocorrelation sequence. One can limit the range of lags in the (auto/cross) correlation function to (*-Maxlag*, *Maxlag*), using the command form *xcorr(...,Maxlag)*. Similarly, the MATLAB function *xcov* produces estimates of the (auto/cross) covariance function (actually, correlation functions of sequences with their means removed).

6. Conclusion

Due to the random character of the pollutant concentrations measured with optoelectronic instruments, statistical signal processing methods are recommended. Histograms, correlation coefficients, (auto/cross) correlation, (auto/cross) covariance functions or statistical parameters like mean, standard deviation, skewness and kurtosis can be useful tools in analyzing such signals. However, according to the purpose of the measuring campaign, the experiment must be carefully designed in order to obtain reliable results. This chapter reveals some practical rules for setting acquisition parameters like data

(segment) size and sampling frequency. As a first rule, for reliable correlation coefficient determination one must assure a sufficient temporal length of the concentration signals, i.e. the product between segment size and sampling period must be large enough in order to obtain stable values of the correlation coefficients. This rule is also valid for the calculation of autocorrelation or auto covariance functions. However, if we are interested to use cross correlation or cross covariance function to reveal delays between pollutant concentration and/or meteorological signals, another rule must also be taken into account: assure the necessary resolution on the time axis i.e. the sampling period must be small enough in comparison with expected delays. In any case, interactive verification and setting of the acquisition parameters during the measuring campaign, according to the purpose of every particular experimental research, are recommended.

Signal conditioning procedures must be implemented before determining the statistical functions and parameters. Thus, ideal filtering based on fast Fourier transform is an useful pre-processing step allowing a simple rejection of measurement noise and possible artefacts of the pollution level signals. Interpolation can be used to increase the number of samples of the slowly varying meteorological parameters, avoiding redundant measurements.

The purpose of one measuring campaign was the correlative comparison of two CO-concentration optoelectronic measuring instruments, working on different principles. Within this research the correlation coefficient proved to be the most useful tool in analyzing dependencies between pollution levels and the meteorological factors. The open path remote sensing instrument measures spatial averaged values which show better correlation to the meteorological parameters. Thus, the open path instrument is better suited for monitoring the pollution level in a large area than the classical NDIR device.

7. References

Hoffmann, J. & Quint, F. (2007). *Signalverarbeitung mit MATLAB® und SIMULINK®. Anwendungsorientierte Simulationen*, Oldenbourg Verlag, ISBN 978-3-486-58427-1, München

Ionel, I.; Ionel, S. & Nicolae, D. (2007). Correlative comparison of two optoelectronic carbon monoxide measuring instruments. *Journal for Optoelectronics and Advanced Materials*, Vol. 9, No. 11, pp. 3541-3545 ISSN

Ionel, I.; Ionel, S. & Lie, I. (2009). Statistical Tools in the Analysis of Pollutant Concentrations Measured with Optoelectronic Instruments, *Proceedings of the 11th WSEAS International Conference on Sustainability in Science and Engineering (SSE '09),* Vol. II, pp.293-298, Timişoara, Romania, May, 2009, Published by WSEAS Press

Martinez, L. W. & Martinez, R. A. (2002). *Computational Statistics Handbook with MATLAB®*, Chapman & Hall/CRC, ISBN 1-58488-229-8, Boca Raton

Montgomery, C. D. & Runger, C. G. (2006). *Applied Statistics and Probability for Engineers*, 4th Edition, John Wiley & Sons, Inc., ISBN 0-471-74589-8, New York

Navidi, W. (2010). *Statistics for Engineers and Scientists*, McGraw-Hill, Inc., 3rd Edition, ISBN-13: 978-0071222051, New York

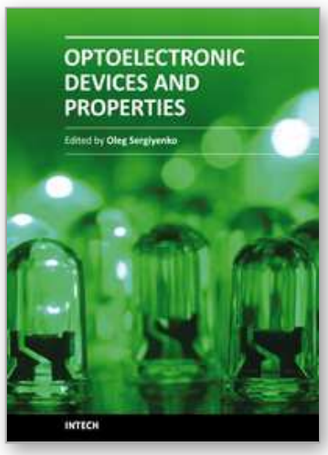
Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Inc., 3rd Edition, ISBN 0-07-100870-5, New York

Peck, R.; Olsen, C. & Devore, J. (2008). *Introduction to Statistics and Data Analysis*, Duxbury Press, 3rd Edition, ISBN-13: 978-0-495-11873-2, Pacific Grove, CA

Shen, D.; Lu, Z. (2006). Computation of Correlation Coefficient and Its Confidence Interval in SAS, www2.sas.com/proceedings/sugi31/170-31.pdf

Therrien, C. W. (1992). *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall International, Inc., ISBN 0-13-217985-7, Englewood Cliffs

Zuur, A. F.; Ieno, E. N. & Smith, G. M. (2007). *Analyzing Ecological Data*, Springer Science + Business Media, ISBN-13: 978-0-387-45967-7, New York



Optoelectronic Devices and Properties

Edited by Prof. Oleg Sergiyenko

ISBN 978-953-307-204-3
Hard cover, 660 pages
Publisher InTech
Published online 19, April, 2011
Published in print edition April, 2011

Optoelectronic devices impact many areas of society, from simple household appliances and multimedia systems to communications, computing, spatial scanning, optical monitoring, 3D measurements and medical instruments. This is the most complete book about optoelectromechanic systems and semiconductor optoelectronic devices; it provides an accessible, well-organized overview of optoelectronic devices and properties that emphasizes basic principles.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen