# Evaluation of air quality in urban areas using a statistical model for data analysis

FRANCISC POPESCU (1), SORIN LUGOJAN (2), OLIVIA BUNDAU (2), NICOLAE LONTIS (1), LIVIO BELEGANTE (3), ADRIAN EUGEN CIOABLA (1)
(1) Department of Mechanical Machines, Technology and Transportation and (2) Department of Mathematics, "Politehnica" University of Timisoara, Bv. M. Viteazu 1, RO-300222 Timisoara
(3) Laser Remote Sensing Department, National Institute of R&D for Optoelectronics, 1 Atomistilor Street, P.O. Box MG-5, RO-77125, Magurele, Ilfov
ROMANIA
ingfrancisc@gmx.net, lbelegante@inoe.inoe.ro

*Abstract:* The paper is an attempt to describe and develop a statistical method for evaluating the accuracy and quality of measured values of pollutants concentration during a typical air quality monitoring campaign in urban areas. The paper  and the statistic evaluation method is based on results obtained during a case study conducted for the International Airport "Traian Vuia" of Timisoara, during different representative episodes of several days in 2008.

*Key-Words:* air quality, urban environment, pollutant, statistical model.

## 1 Background

Air pollution is a major concern for all nations, with a higher or lower development level. The rapid increase of the industry sector and urban development had generated substantial quantities of substances and poisonous materials, which are, mostly evacuated in the atmosphere, in addition the traffic and mobility necessity determined the developing of infrastructures that allow the different fleet, mostly equipped with internal combustion engines, to act. The human society is not recognizing that the environment has only a limited capacity to process all its waste, without major changes. Each of us is a polluter but also a victim of pollution.

Urban (traffic and industry) activities result in the emission of air pollutants that adversely affect public health and the environment, including nitrogen oxides (NOx), hydrocarbons (HC), particulate (PM), carbon monoxide (CO), and toxics. NOx and HC are precursor emissions of ground-level ozone, which causes lung irritation and aggravates diseases such as asthma, chronic bronchitis, and emphysema. Recent health effects studies have shown an association between existing levels of fine particles (size, concentration) and health effects such as increased respiratory illness, cardio-pulmonary morbidity, and premature mortality. For example, a link between air pollution and mortality was demonstrated in two studies using data collected by the American Cancer Society. [1] The study tracked over 500,000 adults in 51 cities over an 8-year period.

Transportation systems, that plays an essential role in economic and social development as well as in the creation of wealth and standards for our modern societies in development or transition, ensures access to jobs, housing, goods and services and answer to the need of people for mobility. [2] However, the continuing expansion of urban road transport in particular the increase of urban automobile fleet raises serious concerns about long-term sustainability of urban environment. As a consequence, environmental issues became serious constraints for the growth and development of European cities.

Notably, significant improvements have been made over the past two decades regarding aircraft fuel efficiency and other technical improvements to reduce emissions. However, these advancements may be offset in the future by the forecasted growth of airport operations and other aviation activities. Because aircraft are only one of several sources of emissions at an airport, it is also considered essential to effectively manage emissions from terminal, maintenance and heating facilities; airport ground service equipments; and various ground transport traveling around, to and from airports. Optimizing airport design, layout and infrastructure; modifying operating practices for greater efficiencies; retrofitting the GSE fleet to "no-" or "low-" emitting technologies; and promoting other environmentally-friendly modes of ground transport are some of the current opportunities airports and the rest of the aviation industry can adopt or apply to help meet these goals and encourage sustainable development in commercial air transportation. [4]

The measurements have been conducted in a large regional airport, the Traian Vuia Airport from Timisoara. Major air pollutants have been measured. For the measurements two mobile laboratories were used, equipped with reference instruments, meteorological instruments and open path instruments. Both laboratories have been placed near airport apron. The measurements have been conducted with equipment and specialists from "Politehnica" University of Timisoara and National Institute of R&D for Optoelectronics, in 2008.

## 2 Experimental

The experimental setup consists in two mobile air quality monitoring laboratories (figure 1). Each laboratory is equipped with reference point instruments for major pollutants ($SO_2$, $O_3$, NOx, CO, $CH_4$, NMHC, THC and PM10), HORIBA AP370 type instruments and two DOAS instruments. The path direction of the DOAS instruments was sited along with airport taxing lane and one DOAS path length was 60 meters and the other ~ 300 meters, oriented in the same direction. Meteorological sensors (wind speed and direction, air temperature, pressure and humidity) are mounted around the mobile laboratories. [7] The following pollutants have been continuously measured, with 10 second resolution, over the entire measuring episode with high precision equipment [3] from Horiba AP370 family: $SO_2$ measured with two APSA370 instruments ( EN 14212:2005); NO, $NO_2$ and NOx measured with two APNA370 instruments ( EN 14211:2005); $O_3$ measured with two APOA370 instruments (EN 14625:2005), CO measured with two APMA370 instruments and one Siemens Hawk DOAS instrument (N 14626:2005); $CH_4$, NMHC and THC measured with two APHA370 instruments (EN 12619:2002), PM10 measured with two Lekel LSV3 instruments (EN 12341:2002). [8]

The detailed flights schedule was obtained; all international and national/regional flights are counted. In addition, due to the summer period, all charter flights to/from Greek Islands have been considered.

Figure 1. View of the experimental setup. "Politehnica" University of Timisoara (UPT) and National Institute for Optoelectronics (INOE) mobile AQM laboratories

Figure 2 shows the carbon monoxide CO measurements, conducted with 3 different instruments: two reference NDIR point measurement instruments and one DOAS-IR Siemens-Hawk instrument. The length of the DOAS-IR open path was 60 meters, in line with the airport apron. A very good correlation of the measured values can be observed, especially for the CO-UPT and CO-DOAS instruments. The CO-INOE measurements are in the same trend but the measured values are with ~ 0.4 mg/m3N lower then the other instruments. This could be caused by an error in span gas calibration. On top of the figure 2 are drafted the departures and arrivals of national/regional, international and charters corroborated with the carbon monoxide measured values. The dependency between aircraft traffic on the apron and the CO measured values is, as expected, obvious. However, figure 2 is important because it shows that the selected placement of the mobile air laboratories near airport facilities and apron is representative and the measured values are valid. [9]

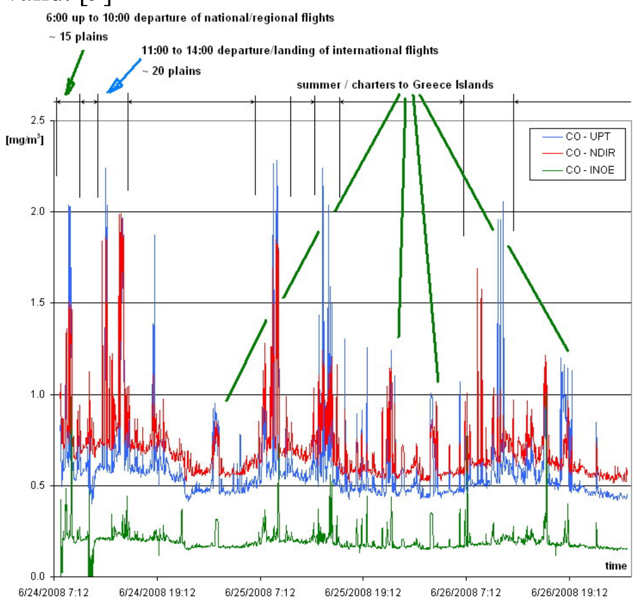


Figure 2. CO measurements and correlation with airplanes traffic.

Figure 3 shows the measured values for sulphur dioxide $SO_2$, with two instruments. Because the SO2-INOE instrument has given high variations (in short term intervals) two 4th degree polynomial trend lines are added to the graph, one for the recorded values of each instrument. The correlation is almost perfect, just differences of about 5 µg/m3N are recorded, differences that can easily be considered equivalent with the measurement uncertainty.

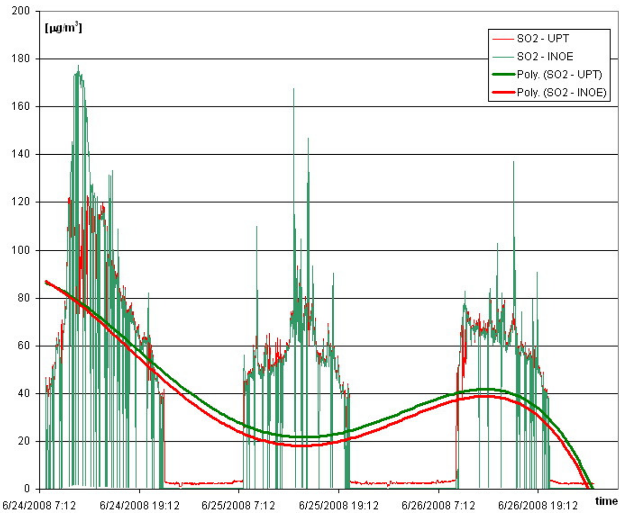


Figure 3. $SO_2$ measurements and polynomial (4 degree) trend lines.

The recorded values for $SO_2$ are not higher then the 150 µg/m3N limit value regulated by 1999/30/EC Directive but they are about 10 times higher then the background values (7 µg/m$^3$). The only possible responsible for the $SO_2$ high values is the airplane fuel.

## 3 The statistical model

In the followings it is presented a comparative investigation using a single sample of one air pollutant. [4]

Let's suppose that in one session of measuring the air pollution, it is obtained a set of $N$ result. Theoretically, that is given by a selection of volume $N$ :

$$\xi = (\xi_1, \xi_2, ...., \xi_N).$$

The random variables $\xi_1, \xi_2, ...., \xi_N$ may be considered independent, due to the detection technique. But what is the type of those random variables?

Since the air pollution is brought about by some regular factors, as well as some irregular ones (in predictable, accidental), it is not possible to specify a certain theoretical type of random variable for the general case. However, of the intensity of the pollution is approximately constant in a time interval, then the measurement during that interval may be considered as an imperfect measuring of a constant value. Therefore, one may consider a normal random variable $X \to N(m, \sigma^2)$ which approximates the pollution intensity during that time interval. That would allow a better theoretical approach of the pollution phenomenon. From the practical point of view, it is also important to know the time constancy interval, especially the ends of them, which signify changes in the pollution dynamic.

The aim of the present study is to give a practical way of finding out such intervals. Here are the steps required for that purpose:

1.One considers a small positive value $\varepsilon$, which is an upper bound of the admissible error.

2.A cycle begins here for $k = 1, 2..., [N/2]$

3.The row matrix of $N$ data is decomposed in blocks of $k$ date, respecting the initial order. The last block may have less then $k$ variables

4.It is computed the mean of each block.

5.If the difference of the means of the first two blocks is less than $\varepsilon$, then the two blocks are considered a single new block (keeping the order). If the difference of the means of the first and the third block is less than $\varepsilon$, then the first three old blocks become a single one, and so on, till the difference between the first block and another one is greater or equal to $\varepsilon$.

6.The procedure 5. is repeated beginning with the second new block, and so on, till the end of the data row matrix.

7.The end of the cycle about $k$.

8.The number $b$, of the final blocks, generally decreases, as $k$ increases. One takes into consideration a value of $k$, such that for $k+1$ the number $b$ stabilizes itself or increases.

9.For the chosen $k$, only the long final blocks are considered as time constancy intervals. For each of those time constancy intervals, it is considered a normal random variable having the mean and the variance of the sample corresponding to the interval. That approximates the pollution intensity during the time interval.

### 3.1 Example

Let's suppose, for the sake of simplicity, that a measurement, independently performed each 3 minutes, gives the following results:

$50, 51, 52, 53, 51, 53, 54, 55.$

The admissible error must be less than $\varepsilon = 1.1$.

a)For $k = 1$, the new blocks are:

$(50,51),(52,53,51,53),(54,55)$.

b)For $k = 2$, the initial blocks are:

$(50,51),(52,53),(51,53),(54,55)$

having the means:

$m_1 = 50.5; m_2 = 52.5; m_3 = 52; m_4 = 54.5.$

The new blocks are:

$(50,51),(52,53,51,53),(54,55)$.

c) For $k = 3$, the initial blocks are:

$(50,51,52),(53,51,53),(54,55)$,

having the means (using only 2 decimals):

$m_1 = 51; m_2 = 52.33; m_3 = 54.5.$

There are no new blocks.

d)For $k = 4$, there are no new blocks.

Hence, for $k = 1,2,3$, it results $b = 3$, and for $k = 4$, it results $b = 2$.

The first stabilization of $b$ is at $k = 1$, and one considers only the second final block as rendering a time constancy interval:

$I = [6,15]$ (in minutes).

During that time interval, the pollution intensity is approximated by a normal random variable of type $N(52.25,1.09)$. Of course, the approximation is not a good one, due to the small data volume, required by simplicity.

## 3.2 Remarks

In real conditions, the measurement of pollution intensity furnishes a much large volume of data (over one thousand). A practical criterion in choosing the time constancy interval is that its volume exceeds 50.

The normal random variable $X$, of (2), represents a better approximation of the pollution intensity, as $k$ is increasing, due to the Central Limit Theorem, see, for example,( [5], pp.372).

The peaks, coming from rapid variations of the pollution, may also be studied by the present method, taking small values for $k = 1,2,\ldots$, n.

## 3.3 The programming code

The following program, written in Mathlab, will be used to find the time constancy intervals. The data row matrix is denoted by $A$, and it is supposed that it is loaded into the interval memory of the computer. The variables $\varepsilon$ and $k$ are those of the previous paragraph.

```
function  r = conc(A, ε, j)
  i = j;
  while(abs(A(i+1) - A(j)) < ε) & (i < length(A) - 1)
  i = i + 1
  end
r = i
endfunction
function  IB = bloc(A, ε)
  i = 1; j = 1; IB(1) = 1;
  while j < length(A)
  i = i + 1; j = conc(A, ε, j) + 1;
  IB(i) = j
  end
endfunction
function  Med = media(A, k)
  j = 1; n = length(A);
  for  i : k : (n - k + 1)
  Med(j) = mean(A(i : i + k - 1));
  j = j + 1
  end
endfunction

function  R = const(A, ε, k)
  M = media(A, k);
  I = bloc(M, ε);
  p = length(I);
  R = k * (I(2 : p) - 1) + 1
endfunction
% main program
ε = ...;
const(A, ε, 1)  [enter]
const(A, ε, 2)  [enter]
const(A, ε, 3)  [enter]
```

and so on.

The function *conc* realizes the concatenation of two or more sequential elements of $A$, beginning with the $j$ - the element, if the difference between the $j$ - the element and each other is less than $\varepsilon$.

The function *bloc* applies the concatenation to the first element of $A$, then to the first non-concatenated element, and so on.

The function *media* computes the means of successive blocks of $k$ elements of $A$.

The function $const$ divides the input matrix in blocks of $k$ elements (the last block) may contain elements). It forms a matrix containing the means of the blocks, and concatenates the successive blocks of $A$, provided the difference between the first mean and the other means is less than $\varepsilon$. Eventually it gives the ordinal number of the beginning element of each concatenated block.

### 3.4 An application

Let's figure out the time intervals mentioned in section 3, for one pollutant measured in the experiment described in section 2, that is CO. The programming code of section 4 is applied for a sample volume of

$N = 1492$ and $\varepsilon = 0.1$. According to Remark 1, we are interested in $k > 50$. The number of new blocks is:

$$b = 13, \text{ for } k = 51$$
$$b = 15, \text{ for } k = 52$$
$$b = 14, \text{ for } k = 53$$
$$b=14, \text{ for } k = 54$$

We choose $k = 53$, due to the stabilization of $b$. The interval (block) ends are: (1, 53), (54,106), and so on. It results that the time intervals are: (0,159), (160,318), and so on (in minutes). Within a certain error, the pollution may be considered constant during each of those intervals, having a normal law (2), with parameters estimated by the mean and the variance of the corresponding block. For example, during the first time interval, the normal law is N (0.52, 0.006).

## 4 Conclusion

The statistical and computational model presented here may be used in order to simplify the study of pollution, post factum or even in real time. It renders some time intervals during which the pollution may be considered constant, and the probabilistic law governing the detection may be considered normal. There are two parameters which gives the necessary finesse:

- the admissible error $\varepsilon$.
- the volume of the initial blocks $k$.

The proper values of those two parameters must be chosen, for each particular case, by several simulations.

### Acknowledgement

This work was partially supported thru Romanian Atmospheric Research 3D Observatory (RADO), Ref no: STVES – 2008/115266 with the financial support of Norwegian Cooperation Programme and partially by the strategic grant POSDRU/89/1.5/S/57649, Project ID 57649 (PERFORM-ERA), co-financed by the European Social Fund / Investing in People, within the Sartorial Operational Program for Human Resources Development 2007 – 2013, and “National Program PNII 2007 – 2013” trough the “Supply control and reduction of the pollution degree by micro-CHP systems driven by spark-ignition engines” TE ID 39 – 2009 grant.

*References:*

[1] C.A. Pope., M.J. Thun, M.M. Namboodiri., D.W. Dockery, J.S. Evans, F.E. Speizer, C.W. Heath, Particulate air pollution as a predictor of mortality in a prospective study of US adults, *American Journal of Respiratory Critical Care Medicine*, Vol. 151, 1995, pp. 669-674.

[2] F. Popescu, I. Ionel, C. Ungureanu, Air quality monitoring in the city of Timisoara, *Journal of Environmental Protection and Ecology*, Vol. 10, 2009, pp. 1-13

[3] HORIBA GmbH Process & Environmental, AP 370 User manuals, Tulln, Austria, 2009 version, http://www.horiba.com/process-environmental/products/ambient/

[4] M. Cames, O. Deuber, *Emission Trading in International Civil Aviation*, Report, Öko-Institut e.V., Institute for Applied Ecology, Berlin, 2009.

[5] G Blom, *Probability and Statistics,* Springer-Verlag, 1989.

[6] D. Wackerly, W Mendenhall, R. Scheaffer; *Mathematical Statistics with application*, Thomson, 2009

[7] F. Popescu, I. Ionel, C. Talianu, Evaluation of air quality in airport areas by numerical simulation, *Environmental Engineering and Management Journal*, Vol. 10(1), 2011, pp. 115-120

[8] F. Popescu, Advantages in the use of Biodiesel in an urban fleet. Case study: major cross-roads in the Timisoara city, *Journal of Environmental Protection and Ecology*, Vol 10(1), 2009, pp. 182-191

[9] D. Nicolae, C. Talianu, E. Carstea, C. Radu, Using classification to derive aerosol number density from lidar measurements, *Journ. Optoelectron. Adv. Mater*. Vol. 9(11), 2007, pp. 3518-3521