

A SMART-AUTOLEARNED ARCHITECTURE AND PROTOCOL FOR VOLATILE ORGANIC COMPOUND MEASUREMENT

*Aimé Lay-Ekuakille*¹, *Salvatore Gioffreda*², *Amerigo Trotta*³

¹ Dipartimento d'Ingegneria dell'innovazione, University of Lecce, Via Monteroni, 73100 Lecce, Italy
Phone +39.0832.297207, Fax +39.0832.324005, aime.lay.ekuakille@unile.it, http://smaasis-misure.unile.it

² Infotel S.r.l., Viale Magna Grecia 215, 74100 Taranto, Phone +39.099.7303711, Italy, gioffreda@isy.net.it

³ Dipartimento di Elettrotecnica ed Elettronica, Faculty Polytechnic of Bari, Via Orabona 4, 70100 Bari, Italy, trotta@misure.poliba.it

Abstract: Emissions monitoring of volatile organic compounds (VOC) from industrial facilities is complex for both economical, computational and practical reasons. A new wide areas monitoring approach based on 'few sensors' wireless network supported by a predictive learning algorithm and an air modeling software can obtain both local and wide VOC estimation.

Keywords: Volatile organic compounds (VOC), learning algorithm, air monitoring, wireless sensors.

1. INTRODUCTION

Concerns, with emissions of volatile organic compounds (VOC), from waste industrial facilities are increasing, but monitoring waste facilities is complex for both economical, computational and practical reasons. Conventional emission estimations focus on statistical treatment and standardization of emission sources with consequent collection data strategies on measuring concentrations. These methods, developed from EPA (U.S. Environmental Protection Agency) and sorted by increasing refinement, are: *Average Emission Factor Approach*, *Screening Ranges Approach*, *EPA Correlation Approach*, and *Unit-Specific Correlation Approach*. All of the approaches require an accurate recognition of equipment components by type (i.e., valves, pumps, connectors, etc.) and by service (i.e., heavy liquid, light liquid, and gas)[1]. EPA approaches describe the state of art in the area, other private methods implement one or more of them because of their standardization and completeness and their costs tolerance. On the other hand EPA approaches compute total facilities emission results as a complex product of static emission source inputs and cannot offer flow estimation or simulation in the facility area. This limit can be found in the measurement methodology that is centered statically and statistically on the emission sources.

2. PURPOSE

To overcome static limits of EPA emission estimates a new measurement methodology will be realized. New method provides Hardware and Software solutions. Hardware solution implements a sensors network that will cover the whole or the interested facility area, for obvious practical problems the network must be wireless connected.

The software solution has two main functions: the first is the collection of network emission data (for both computational needs and inventory reasons) and the data filling of unchecked areas emission trend with a predictive learning algorithm; the second is the implementation of acquired and predicted data with a reliable air model to offer a wide area emission diffusion view and accurate local measurement.

At present we can resume EPA Methods with the scheme in figure 2-1:

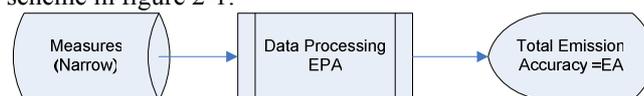


Figure 2-1 EPA Methods summarizing scheme

This process produces total emission estimation with accuracy EA, but it involves two problems:

1. It consider few really measured data
2. It adopts statistical analysis instead of actual survey

To overcome this limits we aim to carry out a new predictive Algorithm obtaining the whole data set of measures. The complete dataset can be processed with a new EPA-Based Method to obtain total emission estimation with accuracy LE likely better than EA figure 2-2. This method will be called LEG.



Figure 2-2 EPA-Based Method summarizing scheme

The prediction process works computing new measured data, that represent an incomplete data set for the whole period to reckon, with historical data to obtain the whole data set for the period, see figure 2-3

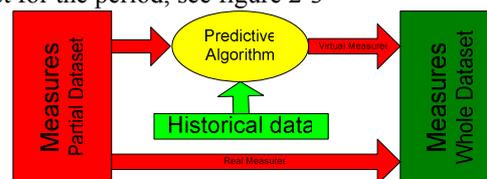


Figure 2-3 EPA-Based Method summarizing scheme

The goal is yielding predictive algorithm and process to produce reliable virtual measures.

To Analyze the accuracy of LEG (*Lay-Ekuakille Gioffreda*) method in comparison with the Standard EPA

Methods, we will compare LEG Method and EPA Correlation Equation Approach Method results.

3. METHODS

3.1. Hardware Implementation

The hardware implementation concerns projecting a sensors grid that must cover the whole facility area or the facility interested area. Because of huge extension network, they cannot be wired but new wireless technologies provides optimal solutions. Sensor technology is based on Photo-Ionization Detection (PID) Transmitters.

For application purpose the position of sensor nodes needs to be predetermined and well-structured to report final area emission trend.

Network design must observe two constraints: sensors area detection range and wireless range area. The largest limit concerns wireless area while the strictest one concerns sensor detection area. With these constraints a network with links no more than 20m can be designed, but further considerations related to the employ of a predictive algorithm allow to revise network design. Prediction estimation provides virtual measurement nodes named Points of Computation (PC). PC can be used to enlarge network links extension and reduce the number of 'real'

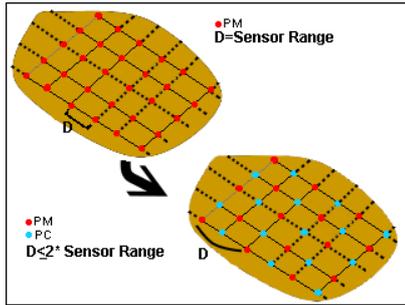


Figure 3.1.1 Wireless network design

sensor nodes named Points of Measure (PM). Actual link extension is subject to the prediction reliability and is:

$$D_{Link} \leq 2 \cdot S_{Range} \quad (1)$$

where S_{Range} is sensor detection range. Figure 3.1.1 summarizes network links measuring. Furthermore there are no constraints for asymmetric grid implementation, indeed it is wise to set sensor nodes closest to the emission sources than in vast areas.

3.2. Air Pollution models

Different scenarios affect the air quality model. Models of air quality need input data that characterize emissions, meteorology, air chemistry in the considered area [2].

We use mathematical models through statistical functions to predict the amount of air pollution from industrial uses [3]. The Gaussian plume model is the most common air pollution model [4]. It is based on a simple formula that describes the tree-dimensional concentration field generated by a point source under stationary meteorological and emission conditions

$$C = \frac{Q}{2\pi\sigma_y\sigma_z\bar{u}} \exp\left[-\frac{1}{2}\left(\frac{y_r}{\sigma_y}\right)^2\right] \exp\left[-\frac{1}{2}\left(\frac{h_e - z_r}{\sigma_z}\right)^2\right] \quad (2)$$

in which, u is the average horizontal wind speed, h_e is the effective emission height (i.e., $h_e = z_s + \Delta h$), and plume sigmas that depend upon Taylor's statistical theory of diffusion. Coefficients σ_y and σ_z are also called respectively lateral and vertical dispersion coefficients. Sigmas can be computed [5] as

$$\sigma_y = \sigma_\theta x (1 + 0,0308 x^{0,4548})^{-1} \quad x \leq 10.000 \quad (3)$$

$$\sigma_z = \sigma_\phi x \left[1 + 0,945 \left(t / T_0 \right)^{0,8} \right]^{\frac{1}{2}} \quad z \geq 50 \text{ m} \quad (4)$$

$$\text{in which } \sigma_\theta = b \quad \text{with } b = 1,92 \pm 0,5 \quad (5)$$

$$\text{And } \sigma_\phi = c \quad \text{with } c = 1,25 \pm 0,03 \quad (6)$$

and the diffusion time t . Elements σ_θ and σ_ϕ are called the standard deviations of the crosswind and vertical wind vector components. In any tri-dimensional figure, x-axis, y-axis and z-axis represent respectively Average Wind Velocity Vector u , Aperture and Concentration C .

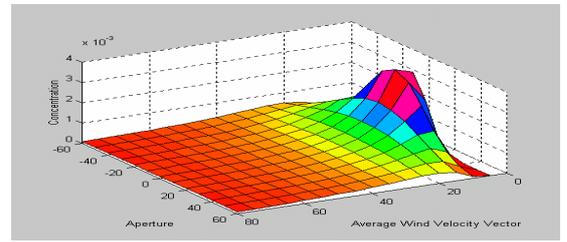


Figure 3.2.1. Spatial distribution of concentration for $z=53 \text{ m}$ and $b=1.92$

3.3. Predictive Algorithms

Concentration distribution model is a dynamical system described by delay differential equation (DDE), then a predictive algorithm must deal with this mathematical model. Known predictive algorithms which examines DDE are Genetic Algorithms (GA), other typical prediction algorithms are neural networks (NN) and their specializations. Three algorithms have been processed: Packard-Mayer [6-7] Genetic Algorithm (PMGA), Multi-Layer Perceptrons (MLPS) and Genetic Based Self Organizing Algorithm (GBSON) [8]. These algorithms have different approaches to the prediction problem: PMGA has a statistical approach to the problem while MPLS has typical system identification approach, GBSON is a hybrid method of the Neural Network and Genetic Algorithms.

4. RESULTS

4.1 Predictive Algorithms

A known example of a dynamic process is the so called Mackey-Glass process, which originates from a model of the amount of CO_2 in the blood. An often used variant of the Mackey-Glass time series is described by the DDE

$$dx = \frac{ax(t - t_{del})}{1 + x^c(t - t_{del})} - bx(t) \quad (8)$$

Typical values are: $a=0,2$; $b=0.1$ $c=10$ and $t_{del}=17$.

This process will be used to test presented predictive algorithms. Figure 4.1 plots Mackey-glass process while figure 4.2 shows same process split into a training set and a testing set of data. For each algorithm same series and same

data are processed. For each testing, 1200 values are taken (one per second) but to train and test only two sets of 500 values are considered.

Error estimation is based on Mean Square Error

$$mse = \frac{1}{Q} \sum_{k=1}^Q e^2(k) \quad (9)$$

the results are shown from figure from 4.1.1 to 4.1.7. Table 4.1.1 resumes most important results.

Algorithm	Training Error	Testing Error	CPU Time
MLPS	0,021	0,21	19,7
PMGA	0,002	0,002	0,8
GBSON	0,220	0,22	8,5

Table 4.1.1. Main Results for inspection of 1000 points (1 per second) with 500 points for training and 500 points for Prediction and test.

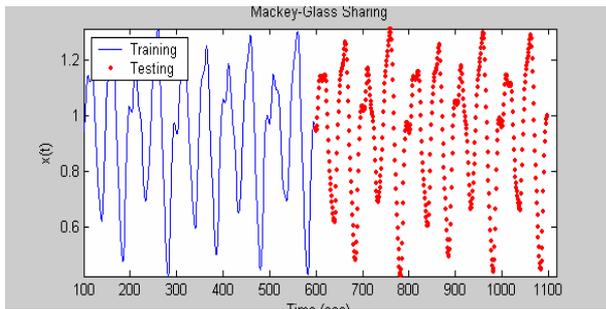


Figure 4.1.1. Subdivision of training and testing data set

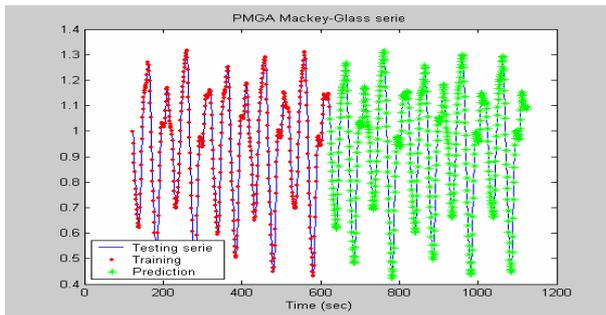


Figure 4.1.2. PMGA Reconstruction

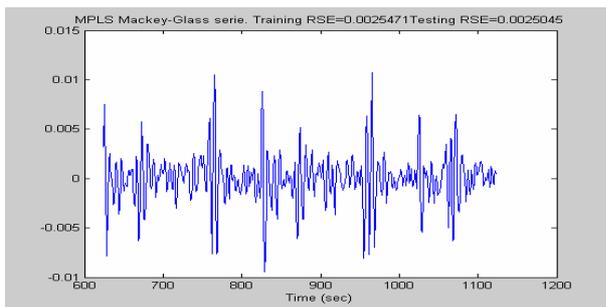


Figure 4.1.3. PMGA Error Plot

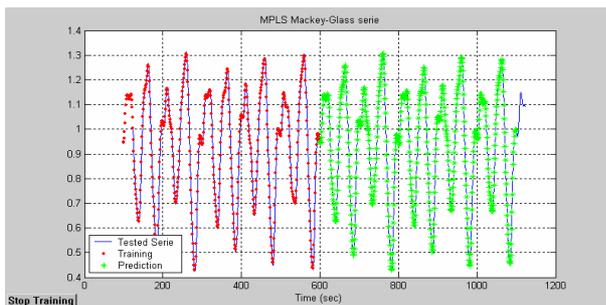


Figure 4.1.4. MLPS Reconstruction

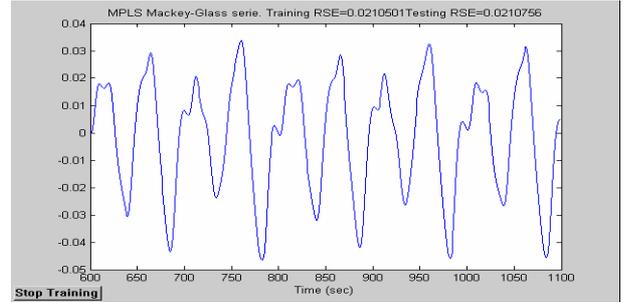


Figure 4.1.5. MPLS Error Plot

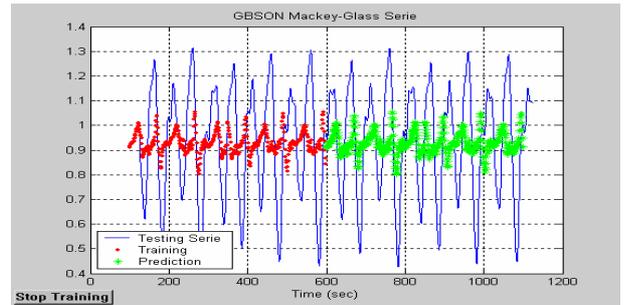


Figure 4.1.6. GBSON Reconstruction

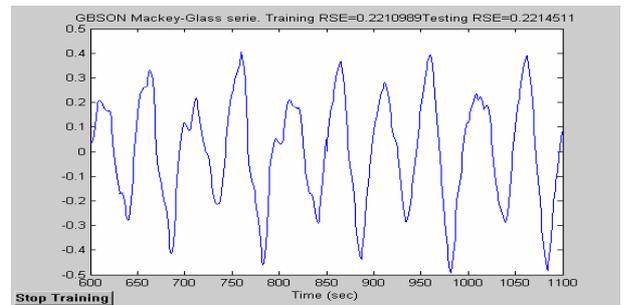


Figure 4.1.7. GBSON Error Plot

4.2 EPA-LEG comparison

Since we produced three predictive algorithms, LEG application has been performed for each of them. Figure 4.2.1. is an overview of the data collection and analysis required to apply each of the methods. The comparison has been executed for a plan consisting of:

Equipment Type	Number	Hours of operation
Ligh Liquid Pumps	12	4380
Ligh Liquid Pumps	15	8760
Gas Valves	40	8760

Table 4.2.1. Plan configuration

and Butil Acetate as VOC, in the following ways:

Comparison	Annual	Weekly
Emission	✓	✓
Standard Deviation		✓
Medium square Error		✓
Variance		✓
Average		✓

Table 4.2.2. Comparisons definition

The following experiments have been executed:

Experiment	Measured Weeks	Measure Frequency	Week Measures	Weeks
1	2	Daily	2	23
2	2	Daily	5	23
3	2	Daily	7	23
4	6	Daily	3	23
5	6	Daily	7	23

Experiment	Measured Weeks	Measure Frequency	Week Measures	Weeks
6	12	Daily	3	23
7	12	Daily	7	23

Table 4.2.3. Experiments definition

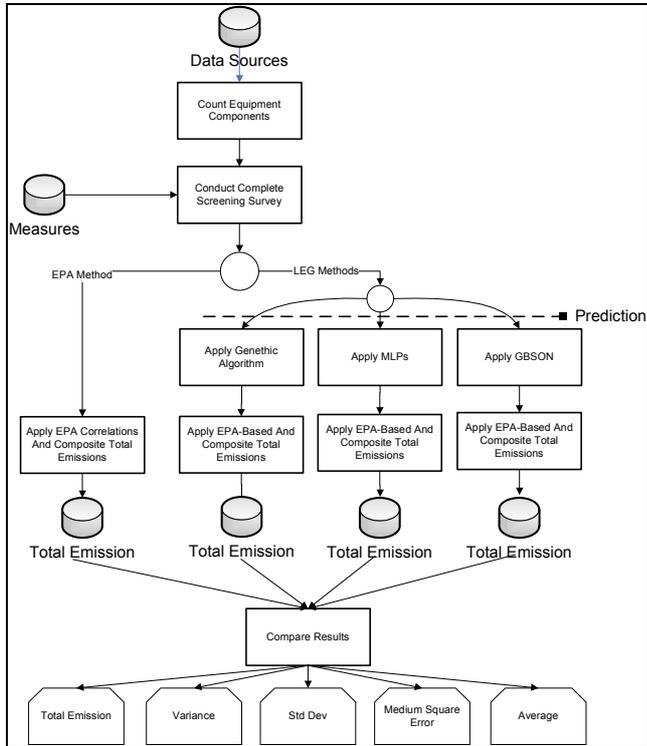


Figure 4.2.1: Overview of Data Collection and Analysis Approaches

In order to evaluate the behaviors of the methods related to data dispersion, for each experiment datasets with the following specialty has been analyzed:

Specialization	Standard Deviation
A	<4
B	<14
C	<30
D	<50

Table 4.2.4. Datasets Dispersion

4.2.1 Annual Comparison

From figures 4.2.2 to 4.2.5 represent graphical results for experiment 1, while tables from 4.2.5 to 4.2.8 represent the whole results of annual comparison.

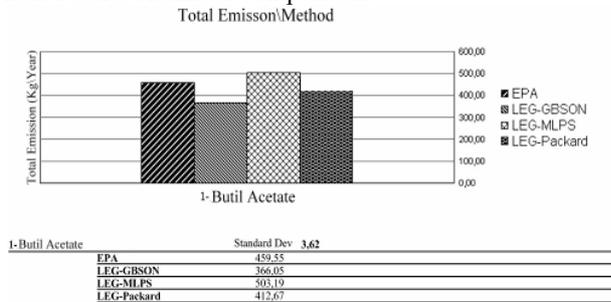


Figure 4.2.2. Annual results for Experiment 1.A

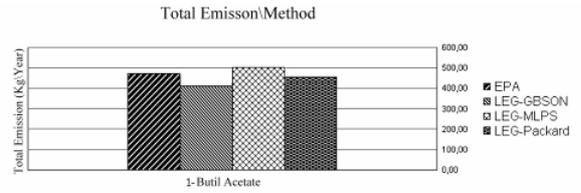


Figure 4.2.3. Annual results for Experiment 1.B

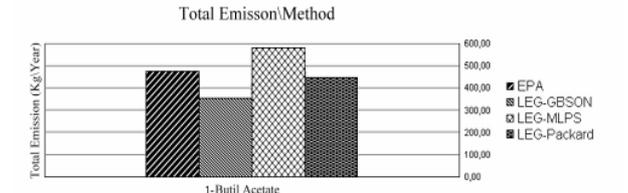


Figure 4.2.4. Annual results for Experiment 1.C

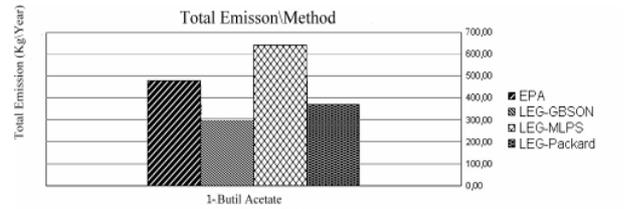


Figure 4.2.5. Annual results for Experiment 1.D

Experiment 1				
Specialization	EPA Std	LEG-GBSON	LEG-MLPs	LEG-Packard
A	459.55	366.05	503.19	412.76
B	477.12	401.85	501.75	449.19
C	472.34	330.65	573.89	425.12
D	479.78	296.35	643.71	376.75

Table 4.2.5: Annual results for Experiment 1

Experiment 2				
Specialization	EPA Std	LEG-GBSON	LEG-MLPs	LEG-Packard
A	444.42	414.645	471.09	432.425
B	467.23	431.255	493.865	449.945
C	476.14	372.82	517.565	433.29
D	494.16	454.63	542.095	405.215

Table 4.2.6: Annual results for Experiment 2

Experiment 3				
Specialization	EPA Std	LEG-GBSON	LEG-MLPs	LEG-Packard
A	461.75	389.26	511.16	444.67
B	481.12	386.34	547.035	453.7
C	476.34	372.975	565.42	428.71
D	488.78	324.065	585.07	415.465

Table 4.2.7: Annual results for Experiment 3

Experiment 4				
Specialization	EPA Std	LEG-GBSON	LEG-MLPs	LEG-Packard
A	450	415.35	468	432
B	486.71	439.5	519.32	457.51
C	491.46	384.815	548.965	437.4
D	498.16	408.495	606.265	398.53

Table 4.2.8: Annual results for Experiment 4

4.2.2 Weekly Comparison

Weekly comparison has been taken to inspect emission trend for each applied method.

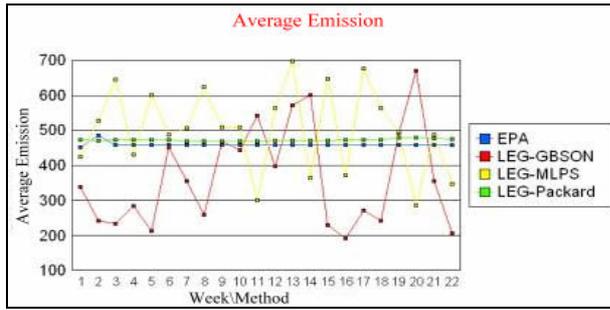


Figure 4.2.6: Average Emission Estimation for Experiment 1.A

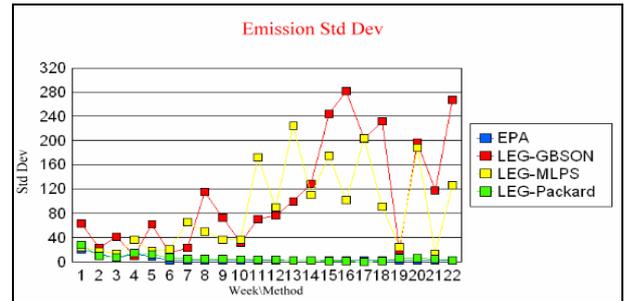


Figure 4.2.11: Std Dev of Emission Estimation for Experiment 5.B

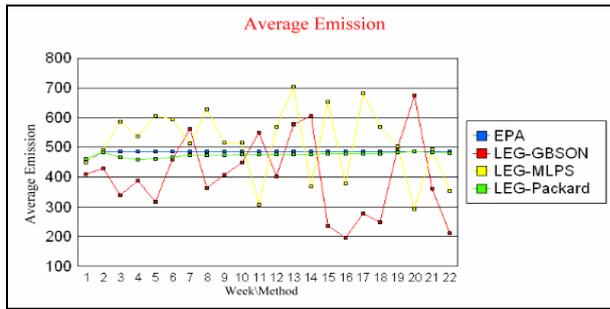


Figure 4.2.7: Average Emission Estimation for Experiment 2.C

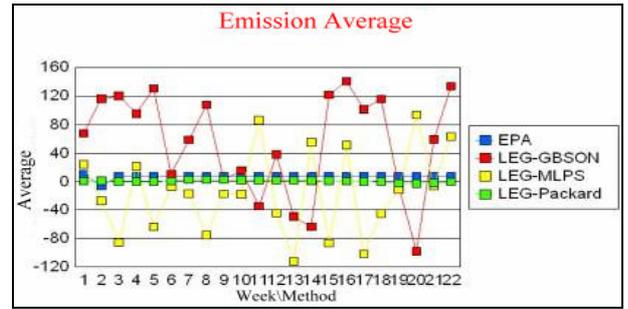


Figure 4.2.12: Average of Emission Estimation for Experiment 1.A

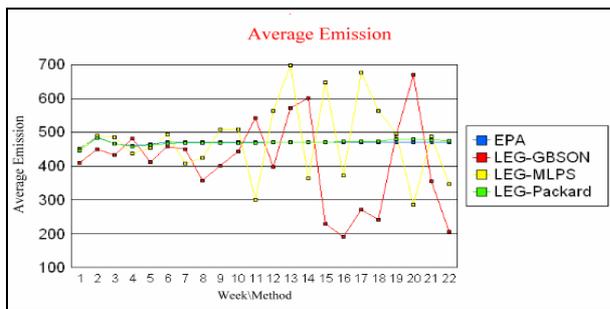


Figure 4.2.8: Weekly Average Emission Estimation for Experiment 5.B

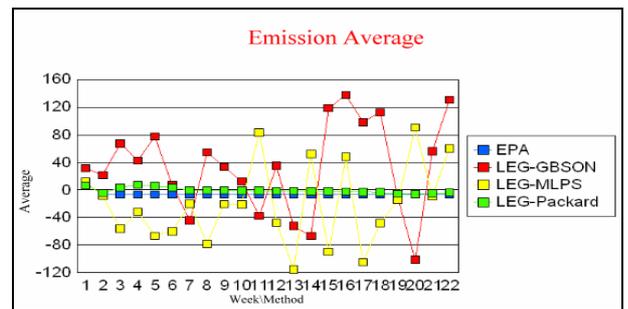


Figure 4.2.13: Average of Emission Estimation for Experiment 2.C

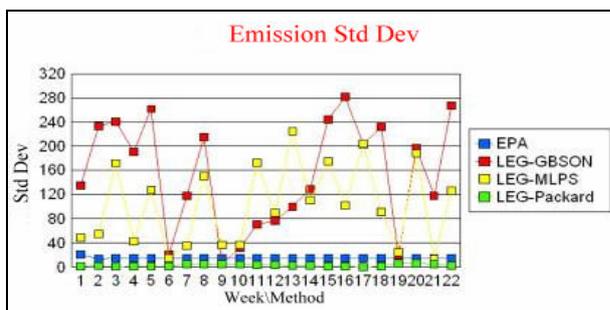


Figure 4.2.9: Std Dev of Emission Estimation for Experiment 1.A

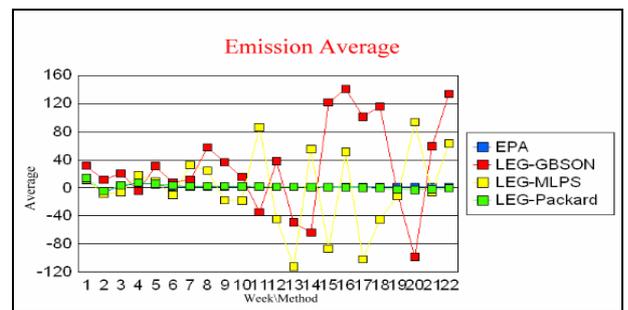


Figure 4.2.14: Average of Emission Estimation for Experiment 5.B

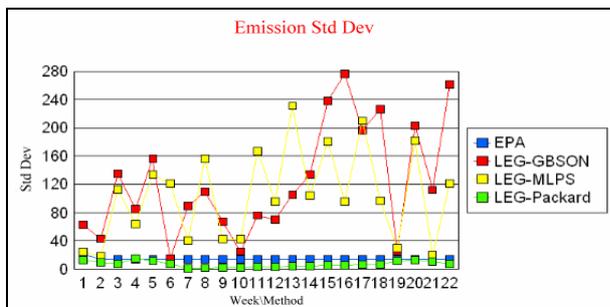


Figure 4.2.10: Std Dev of Emission Estimation for Experiment 2.C

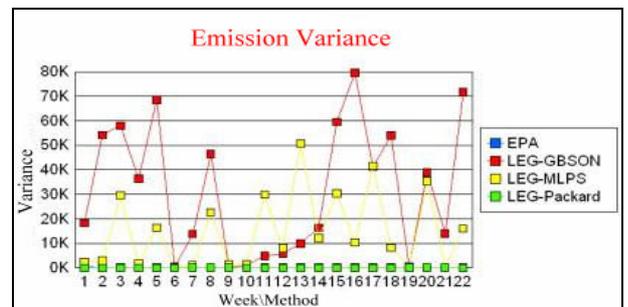


Figure 4.2.15: Variance of Emission Estimation for Experiment 1.B

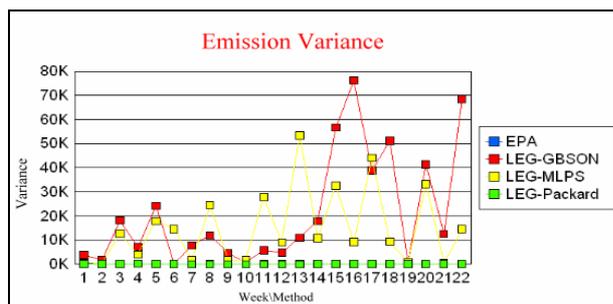


Figure 4.2.16: Variance of Emission Estimation for Experiment 2.C

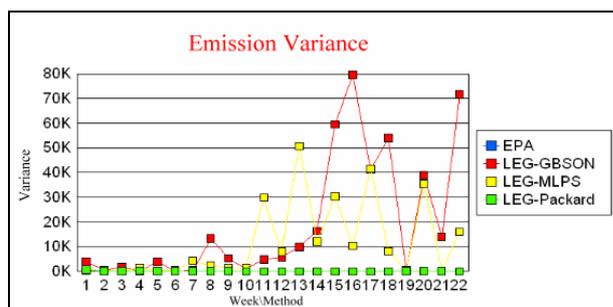


Figure 4.2.17: Variance of Emission Estimation for Experiment 5.B

4. CONCLUSION

A VOC emission estimation approach, named LEG, has been described. This new method involves Hardware Network to measure concentrations in wide areas such as facility fields, mathematical Models for concentration predictions and air quality modeling, Software interface to report VOC evolutions and estimation. Performed tests demonstrate that Packard Mayer Genetic Algorithm (PMGA) represents the best model to predict chaotic dynamic systems, it offers best prediction results either computational or reliability aspects (Table 4.1.1.). Further tests demonstrate that, even, when the number of points and values range improve, PMGA offers the same performance while Multilayer Perceptrons (MLPS) and Genetic Based Self Organized Network (GBSON) need adjustments in training epochs and in networks layers to work reliably.

Furthermore Standard EPA Correlation Equation Approach and LEG comparison has been performed. In order to evaluate predictive algorithms behavior with actual data, LEG Method has been implemented with any presented predictive algorithm (Figure 4.2.1), the results relate same consideration.

Comparisons between LEG-PMGA, LEG-MPLS, LEG-GBSON reports that LEG-PMGA (see figures 4.2.2-4.2.4 and tables 4.2.5-4.2.8) produce the best accuracy both in analytical and real experiments. In many cases (see figures 4.2.5 and table 4.2.7) LEG-MPLS and LEG-GBSON produce unstable results, and need adjustment parameters to work reliably. This happens especially with small measure sets and when measures dispersion increase.

Comparison between LEG-PMGA and standard EPA Correlation Equation Approach report that methods produce comparable results but LEG-PGMA is more accurate when measures dispersion increase. In fact, figure 4.2.2 and tables 4.2.5-4.2.8 show that with low Standard Deviation LEG Total Annual emission is between 90% and 100% than EPA

Total annual emission, with higher Standard Deviation (see figures 4.2.3-4.2.5 and tables 4.2.5-4.2.8) LEG accuracy increase jointly, reaching in the best case 30%. This is due to the different LEG emission Estimation that considers dynamically the measure fluctuation in the entire period versus EPA emission estimation that considers statically the only collected measures.

That is better shown in figures 4.2.6-4.2.17 where, considering weeks with collected measure, we can notice a similar behaviour of week-emission while, considering weeks without collected measures, EPA results remain static and LEG results evolve with predicted measure fluctuation. The gap depends on measures dispersion. Results inspection suggest the following:

Standard Dev. of Measures	LEG vs EPA Accuracy
<4	4%-10%
<14	6%-10%
<30	10%-20%
<50	10%-30%

With Standard Deviation of measures higher than 50; further considerations have to be contemplated with data dispersion. If measures are estimated as low-fluctuation (EPA Method considers it constant in many cases) we can assume that LEG-Method, after a startup process and parameter configuration, automatically reports emission Estimation with better accuracy than EPA Correlation Equation Approach. Otherwise it needs an in-depth inspection to determinate the correct parameterization.

So we can assert that LEG Method is easiest to use than EPA Correlation Equation approach Method because it is self-organized and it can produce VOC emission estimation at any moment characterizing it dynamically with the available measures collection. For the same reason LEG method is more accurate than EPA and produces emission estimations based on the actual feature of the considered Plant.

REFERENCES

- [1] EPA "Protocol for Equipment Leak Emission Estimate" EPA-453/R-95-017, Chap. 2 pp 3, November 1995
- [2] R.M. Srivastava, H.M. Parker, "Robust Measures of Spatial Continuity in Geostatistics", Proc. Of the Third International Geostatistics Congress, September 5-9, 1988, Avignon, France.
- [3] G. Andria, A. Lay Ekuakille, M. Notarnicola "Mathematical models for atmospheric and industrial pollutant prediction" Dipartimento di Elettrotecnica ed Elettronica, Politecnico di Bari, Bari
- [4] P. Zannetti, *Air Pollution Modeling - Theories, Computational Methods and Available Software*, Van Nostrand Reinhold, New York, 1990
- [5] R. Draxler, *Determination of Atmospheric diffusion parameters*, *Atmos. Environ.*, N.10: pp.99-105
- [6] T.P. Meyer & N.H. Packard, Local forecasting of high dimensional chaotic dynamics., *Non linear Modelling and forecasting*, Addison- Wesley, 1992.
- [7] N.H. Packard, A genetic learning algorithm for the analysis of complex data. *Complex Systems* 4, n.5, pp 543- 572, 1990.
- [8] Amalia Foka "Time Series Prediction Using Evolving Polynomial Neural Networks", Control Systems Centre Department of Electrical Engineering & Electronics, University of Manchester Institute of Science and Technology, 1999