Computational complexity and length of recorded data for fluctuation enhanced sensing method in resistive gas sensors

L. Lentka, J Smulko
Department of Metrology and Optoelectronics, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland
E-mail: lukasz.lentka@gmail.com

Abstract. This paper considers complexity and accuracy of data processing for gas detection using resistance fluctuation data observed in resistance gas sensors. A few selected methods were considered (Principal Component Analysis – PCA, Support Vector Machine – SVM). Functions like power spectral density or histogram were used to create input data vector for these algorithms from the observed resistance fluctuations. The presented considerations are important for proposing relatively cheap and mobile gas detection devices of limited computations abilities and utilizing fluctuation enhanced gas sensing method.

1. Introduction
Resistive gas sensors are commonly used in various practical applications [1-3]. Their most important task is to detect gas of interest or determine its concentration at sufficiently high accuracy because this feature can be often crucial for living standard and health of human beings. A single resistance gas sensor is often insufficient to reach high concentration accuracy or high detection rate. Therefore a set of gas sensors is applied with selected data processing algorithm. There are many well-known algorithms and methods [4-5] which can be used to detect various gases utilizing resistive sensors or other, even more expensive, gas sensors types (e.g. electrochemical or optical [6, 7]). Nowadays there is strong pressure on proposing inexpensive and miniaturized gas sensors to detect various gases in portable systems, accessible for wide range of customers (e.g. in mobile phones). Thus, the system which applies a set of resistive gas sensors cannot be used in portable applications because of high energy consumption (each resistive gas sensor has to be heated up). Another proposed method can utilize fluctuation phenomena observed in resistive gas sensors (fluctuation enhanced sensing – FES) at low frequency range to gather more data from much limited number of gas sensors. Unfortunately, saving energy for heating requires more energy for data recording and processing. That issue has been addressed in the paper. Selected algorithms are considered due to their computational complexity and accuracy. These algorithms are applied in gas sensing area and can use the FES method as well. The conclusions should help us to decide what kind of statistical parameters characterizing resistance fluctuations would be most efficient in practical applications of the FES method.

Power spectral density and histogram characterizing noise phenomena together with the selected algorithms of data processing were taken into account. Both functions characterizing fluctuations were
considered by analysing their random error due to limited number of recorded noise samples used for their estimation.

2. Power spectral density
Fluctuations can be characterized by power spectral density (PSD), determining their power distribution versus frequency. The power spectral density \( S(f) \) of random signal can be defined by equation:

\[
S(f) = \frac{1}{2\pi N} \sum_{i=1}^{N} |X_i(f)|^2
\]

where \( X_i(f) \) is a Fourier transform (spectrum) of time series being a part of the analysed signal \( x(t) \). Power spectrum density is estimated by recording noise samples, computing Fourier transform for the consecutive time series and averaging over the set \( N \) of the estimated spectra to reduce its random error. The FES method utilizes low frequency noise which depends on ambient atmosphere of the applied gas sensor [2, 3]. Resistance fluctuations are observed as voltage fluctuations across the sensor biased by DC voltage [8-11]. It has been confirmed that low frequency noise power spectrum is often much more sensitive to any changes of ambient atmosphere of gas sensor than a change of its DC resistance only [12]. Moreover, the PSD is a function of frequency and gathers information about presence of the investigated gases by changing its intensity and slope versus frequency. Usually, the product \( f S(f)/U^2 \) of voltage power spectral density multiplied by frequency \( f \) and normalized by squared DC voltage bias of the sensor is used. That product is independent from actual measurement circuit and depends on the sensor and measured gas only. Thanks to multiplying the measured spectra by frequency, visibility of any spectra shape changes is improved [12].

Averaging over the spectra set reduces variance of the estimated PSD. When \( N \) independent spectra are averaged a random error \( \varepsilon_r \) of the PSD is given by equation:

\[
\varepsilon_r = \left( \frac{1}{N} \right)^{1/2}
\]

For example, if we average \( N = 40 \) spectra, the random error is reduced to \( \varepsilon_r \approx 16\% \) which can be accepted for noise measurements. Strong increase of the averaged spectra number is often not recommended because it requires much longer noise records. Assuming a sampling frequency \( f_s = 4 \) kHz necessary to observe \( 1/f \) noise in the gas sensors and 1024-point spectrum we need to record noise in time of 10.24 s to estimate \( N = 40 \) independent spectra. That time can be accepted for practical applications. Twice lower random error means four time longer time of noise recording. It means that further increase of noise recording time cannot be often accepted.

The PSD function requires application of fast Fourier transform algorithm which is very popular and can be implemented even in a cheap portable gas sensing devices optimized for signal processing operations, like FPGA arrays or signal processors. The estimated PSD can be used as input data vector for regression or detection algorithms to determine gas concentration or its presence only. Such algorithms require additional computation. Another possibility is to reduce dimension of the PSD function by estimating its slope only to use it for gas detection. Unfortunately, the gas detection based on the PSD slope only gives much worse detection results than PCA or SVM algorithms using the PSD function [13].

3. Histogram
Histogram of noise record is one of the statistical methods characterizing the observed fluctuations and is independent from the presented PSD function. Unfortunately, in numerous cases the observed noise has Gaussian distribution only. Then, we cannot get any additional information about the present gases. Histogram is a good tool when we observe non-Gaussian noise component, characteristic for the investigated gas. Such noise was observed in some resistive gas sensors and characterized by various statistical measures [2, 14]. We can claim that such situation should be more common for new
gas sensor materials (e.g. nanoparticles functionalized by organic ligaments) because of low dimension of the grains (about a few nm only) composing the gas sensing layer.

In order to achieve an acceptable random error $\varepsilon_r$ of estimated histogram we have to record sufficient number of noise samples. When the investigated signal time series $x(t)$ has probability density $p(x)$ it is necessary to collect sufficient number $K$ of noise samples. The random error depends on probability of occurring the observed signal value $x(t)$. That probability is determined by the product of probability density $p(x)$ and range $\Delta x$ where the histogram was estimated. Thus, the random error $\varepsilon_r$ is defined by:

$$
\varepsilon_r = \left( \frac{1}{Kp(x)} \right)^{1/2}
$$

We can suppose that any differences from normal distribution should be visible only at relatively high signal values $x(t)$, typically 2-3 times greater than its standard deviation $\sigma_x$ as reported in literature [14]. When we consider that $p(x)$ is Gaussian distribution and that we divide the range $\pm 3\sigma_x$ into 20 equal intervals we have to collect about $2 \times 10^6$ of noise samples to limit random error $\varepsilon_r$ of the extreme ranges to a few percent only. It means that we have to prolong time of noise recording about 50 times $(2 \times 10^6/(40 \times 10^2))$ when compared with the data collected for estimating PSD function at the same sampling frequency. Such increase can be difficult in practical applications. Even much longer recording time was used when instead of histogram a bispectrum function was estimated to determine intensity of non-Gaussian component [14].

Intensity of non-Gaussian component can be determined by other measures, like skewness or kurtosis [14]. Both measures can be used to determine asymmetry (skewness) or peakedness (kurtosis) of the observed distribution from the normal one. These measures can be estimated using much shorter noise records than histogram. For example, a skewness $\gamma_1$ of the signal $x(t)$ having standard deviation $\sigma_x$ and mean value $\mu_x$ is defined by formulae:

$$
\gamma_1 = E \left[ \left( \frac{x - \mu_x}{\sigma_x} \right)^3 \right]
$$

where the operator $E$ means averaging. The estimated value $\gamma_1$ is a parameter only instead of function as it was in a case of histogram. It means that its usefulness for gas detection algorithms will be certainly more limited but does not require so long noise recording.

### 4. Principal Component Analysis

Very popular method applied for data processing in electronic nose applications is Principal Component Analysis (PCA). The method considers data collected from the gas sensors array as an input vector and performs their linear transformation to expose the most informative components (principal components). That method reduces size of the input data and makes further data processing much easier. The model created by the PCA method can be used to predict gas concentration (regression algorithm) or detect its presence (detection algorithm). Both algorithms require linear algebra operations only and can be performed by relatively cheap and popular embedded systems like smartphones or dedicated microcontrollers. The same method can be applied when the input data is a data vector composed of PSD function or histogram.

In general the regression algorithm is performed during three steps. Firstly, the PCA method transforms the sensors data matrix to determine the most informative principal components for further analysis. Secondly, a model of linear regression is prepared using the input data vector obtained in the first step and gas concentration measurements using the reference method. The model comprises of a set of regression coefficients having the same dimension as the number of applied principal components. The same method is used in various chemometric applications [4,15]. The created model can be used in the third step to predict gas concentration by the input data representing answer of the gas sensors (e.g. the estimated PSD of resistance fluctuations in a gas sensor) to their ambient atmosphere.
If a low number of principal components, typically limited to 5-10 only, is selected we can assure relatively low computational complexity and accuracy of gas concentration estimation equal to a few ppm. Such conditions were observed for the FES method for some gases [13]. Thus, it is possible to use the PCA algorithm in a portable device.

5. Support Vector Machine

Gas sensors response to presence of various gases is often nonlinear (e.g. dependence between DC resistance of gas sensor and gas concentration is logarithmic). Therefore we have to consider other algorithms than the presented PCA as a more advanced alternatives. Such algorithms should be more efficient in gas concentration prediction by the cost of more complicated computations. Support vector machine (SVM) algorithm is a developed method of the mentioned PCA algorithm because it includes a nonlinear case which should be more efficient for gas detection using the FES method [16]. The SVM algorithm for regression or detection aims was proposed by Vapnik [16] and further developed into its least square version (LS-SVM) by Suykens to reduce necessary computations when compared with the original SVM algorithm [5]. Again the LS-SVM regression algorithm requires three steps as was mentioned for the PCA algorithm.

The LS-SVM algorithm transforms an input data into a new space using nonlinear function (figure 1). When this function is linear the algorithm simplifies to the PCA method. The LS-SVM algorithm finds the optimal solution by solving the system of linear equations and using least square criterion. The applied nonlinear function provides the conversion of the input feature space into the transformed space where data are linearly separable to assure the highest margin between the data.

![Nonlinear kernel function](image)

**Figure 1.** Illustration of support vector machine algorithm using nonlinear transformation.

The most intense computations in the LS-SVM algorithm are performed during model creation using a set of input vectors \( x_k (k = 1, \ldots, N) \) recorded during \( N \) measurements at different ambient gas atmospheres. When the necessary computations are performed by PC computer the LS-SVM algorithm requires at least a few times longer time computing when compared with the PCA algorithm.

The established model for the LS-SVM method estimates a vector of coefficients \( a_k (k = 1, \ldots, N) \) defining the set of hyperplanes which maximizes the margin between the input data after nonlinear transformation. Then we can predict gas concentration \( y(x) \) by utilizing the input data vector \( x \) (e.g. data vector of PSD or histogram when the FES method is applied) and the created model, defined by coefficients \( a_k \) and parameter \( b \):

\[
y(x) = \sum_{i=1}^{N} a_k K(x_i, x) + b
\]

where \( K(x_i, x) \) is a nonlinear kernel function applied in the LS-SVM algorithm. The most popular kernel function is polynomial, sigmoidal or Gaussian radial basis function (RBF). Equation (5) does not require intense computations. Thus, data processing of similar complexity has to be performed as in a case of previously mentioned PCA algorithm. The more intense computations are performed only during model creation. That step has to be done only once at the first stage before applying regression method. It means that we can use regression method based on the LS-SVM algorithm than on the PCA even in portable gas detection systems.
The main privilege of the LS-SVM method is its better accuracy when compared with the PCA. It has been observed that the same input data can give a few percentage better accuracy of gas concentration prediction when the LS-SVM algorithm and Gaussian radial basis function were used instead of the PCA method.

6. Conclusions
Problems of data processing for gas detection using fluctuation enhanced sensing method were considered. It has been underlined that power spectral density of the observed resistance fluctuations requires lower number of recorded noise samples to be estimated at given accuracy (random error) than histogram. Both functions can be uses as input data vectors for gas detection or gas concentration regression algorithms. Next, two popular algorithms for gas sensing were presented and compared (PCA and LS-SVM). It was concluded that the slightly better results are observed when the LS-SVM algorithm is applied due to better prediction of gas concentration. Unfortunately, that algorithm means more computations during creation of the regression model.

7. References
[1] Heszler P, Ionescu R, Llobet E, Reyes L F, Smulko J, Kish L B, and. Granqvist C G 2007 On the selectivity of nanostructured semiconductor gas sensors *Physica Status Solidi B* 244(11) 4331–4335
[2] Kwan C, Schmera G, Smulko J, Kish L B, Heszler P, and Granqvist C G 2008 Advanced agent identification with fluctuation-enhanced sensing *IEEE Sens. J.* 8(6) 706–713
[3] Kotarski M and Smulko J 2010 Hazardous gases detection by fluctuation-enhanced gas sensing *Fluct. Noise Lett.* 9 359–371
[4] Barman I, Dingari N C, Singh G P, Soares J S, Dasari R R, and Smulko J M, 2012 Investigation of noise-induced instabilities in quantitative biological spectroscopy and its implications for noninvasive glucose monitoring *Analytical Chemistry* 84.19 8149–8156
[5] Suykens J, Van Gestel T and De Brabanter J 2002 *Least squares support vector machines* (Singapore: World Scientific)
[6] Wojtas J, Rutecka B, Popiel S, Nawala J, Wesolowski M, Mikołajczyk J, Bielecki Z 2014 Explosives vapors–concentrating and optoelectronic detection *Metrol. Meas. Syst.* 21 177–190
[7] Dunst K, Jasinski G, Jasinski P 2014 Potentiometric oxygen sensor with solid state reference electrode *Metrol. Meas. Syst.* 21 205–216
[8] Kotarski M and Smulko J 2009 Noise measurement set-ups for fluctuations-enhanced gas sensing *Metrol. Meas. Syst.* 16 457–464
[9] Ederth J, Smulko J, Kish L B, Heszler P and Granqvist C G 2006 Comparison of classical and fluctuation-enhanced gas sensing with Pd x WO 3 nanoparticle films *Sens. Actuators B* 113 310–315
[10] Kish L B, Li Z, Solis J L, Marlow W H, Vajtai R, Granqvist C G, Lantto V, Smulko J and Schmera G 2005 Detecting harmful gases using fluctuation-enhanced sensing with Taguchi sensors *IEEE Sens. J.* 5 671–676
[11] Kubersky P, Sediłák P, Hamácek A, Nespurek S, Kuparowitz T, Sikula J, Syrový J T 2014 Quantitative fluctuation-enhanced sensing in amperometric NO 2 sensors *Chem. Phys.*, http://dx.doi.org/10.1016/j.chemphys.2014.10.021
[12] Kish L, Vajtai R and Granqvist C 2000 Extracting information from noise spectra of chemical sensors: single sensor electronic noses and tongues *Sens. Actuators B* 71.1 55–59
[13] Lentka L, Smulko J, Ionescu R 2015 *Efficiency of Gas Detection Algorithms Using Fluctuation Enhanced Sensing* The 23rd Int. Conf. on Noise and Fluctuations, 2–6 June 2015, Xi’an China
[14] Smulko J, Granqvist C G, Kish L B 2001 On the statistical analysis of noise in chemical sensors and its application for sensing *Fluctuation and Noise Letters*, 1,03, L147-L153
[15] Lorber A, Kowalski B R Estimation of prediction error for multivariate calibration 1988 *Chemom.* 2, 93-109
[16] Vapnik V 2000 *The nature of statistical learning theory* (Springer Science & Business Media)
Acknowledgments
This work was supported by the National Science Center, Poland, the grant decision: DEC-2012/06/M/ST7/00444 “Detection of gases by means of nano-technological resistance sensors.”