Wavelet Transform Smoothing Filters for Metal Oxide Gas Sensor Signal Cleaning

Enobong Bassey*, Jacqueline Whalley*§, Philp Sallis* and Krishnamachar Prasad†.
* Geoinformatics Research Centre/ † School of Engineering
Auckland University of Technology
Auckland, New Zealand
§jwhalley@aut.ac.nz

Abstract— **This paper reports on a series of experiments to evaluate the methods for feature extraction and denoising the digital signal from thin film zinc oxide-tin dioxide composite gas sensor devices. The aim was to find a method that not only cleaned the signal but also maintained the shape, precision and resolution of the signal. It was found that the Savitzky–Golay smoothing filter method gave the best, smoothed and cleaned, approximation of the sensor response regardless of the thin film composition, target gas concentration or operating temperature.**

Keywords- denoising; gas sensor devices; signal processing

I. INTRODUCTION

Many industrial and domestic activities require more precise and reliable measurements of trace gases such as methane, carbon monoxide, and carbon dioxide. International, national, and local legislation frequently requires continuous monitoring of air quality and rate of emissions. As a result there has been considerable effort to fabricate low cost, accurate, and practical gas sensors. Recent efforts have focused on the fabrication and usefulness of metal oxide gas sensors; for a useful review see [1]. Typically, raw signals acquired from these gas sensor devices are contaminated by noise and outliers. As a result, the signal is occluded to a significant degree making accurate measurement of a sensor's response impossible.

Noise in sensor systems has several possible sources and may be introduced at various points in the measurement process. Several forms of noise, including thermal and shot noise, are irreducible because they are inherent to the underlying physics of the sensors or electronic components. Other forms of noise which could be avoided originate from processes and include 1/f noise, transmission, and quantization noise [2]. Noise introduced in the early stages of measurement is considered to be the most harmful as it propagates and may be amplified through subsequent stages in the signal pathway.

While physical filters have been found to be successful in producing a cleaner signal they do not cover the full resolution and shape of the curve. In order to improve the interpretability, sensitivity and selectivity, of gas sensor array signals it is preferable to use the full resolution and coverage of the signal. Several signal processing approaches have been investigated as methods for reducing noise levels. However, these methods are typically static or steady state approaches and therefore do not encompass the full temporal signal. A useful summary of statistical and optimization methods that have been used to process gas sensor array signals is provided by Gutierrez-Galvez [3]. Our work focuses on improving existing pre-processing techniques to eliminate noise, smooth and filter data, enhance sensor signals, and improve measurement and response features.

II. SIGNAL PRE-PROCESSING

The process of signal pre-processing facilitates noise elimination, data smoothing/filtering and signal enhancement, with the sole aim of increasing the signal-to-noise ratio without greatly distorting the signal. The choice of signal pre-processing method has a significant impact on the result and performance of the signal conditioning system [2]. According to Johnson *et al*. [4] any signal pre-processing method should:

- preserve the chemical selectivity differences between different profiles.
- limit run-to-run retention/migration time shift.
- be fast and less memory-demanding to deal with large numbers of data sets in a short period of time.
- significantly improve the precision of retention/ migration time estimation .

Wavelet transform smoothing filters, for this purpose, should meet all these criteria. Wavelets are a time-frequency representation for continuous-time (analog) signals [5] which easily represent the different features of a signal, especially sharp signals and discontinuities. When applied to analytical signal processing, wavelet transforms provide a simple procedure with short operation time, low memory requirements, high precision, and good reproducibility [6]. For this work we chose the Savitzky–Golay smoothing filter (SGF) for evaluation. SGF is a finite impulse response (FIR) wavelet transform smoothing filter which is superior to other adjacent averaging FIR filters because it preserves the features of the data in the signal, such as peak height and width. While SGFs are more effective at preserving the pertinent high frequency components of the signal, this improvement is at the expense of lower noise reduction. For this reason SGFs are generally considered to be less effective than standard averaging FIR filters at eliminating noise. However, SGFs are useful for minimising the least-squares error in fitting a polynomial to frames of noisy data. Therefore, it is likely that an SGF might be a good choice for gas sensor signal cleaning where it is important to preserve the height, width, amplitude, and overall profile of the signal while also reducing noise. In this study we evaluated the use of an SGF, for gas sensor signal pre-processing using zinc oxide and tin dioxide composite (ZnO–SnO_2) thin film sensor devices, by comparing SGF performance with a moving average filter and local regression methods.

III. EXPERIMENTS

Five sensor devices were fabricated by depositing thin films of zinc oxide and tin dioxide composites (ZnO–SnO_2) using a radio frequency (RF) sputtering process for 30 minutes on an Si/SiO_2 substrate (Table 1) [8] . Each of these sensor devices were exposed to different concentrations of methanol [100, 150, and 200 ppm] at three different temperatures [150, 250, and 350ºC]. A constant voltage of 5V was applied to the sensing elements while recording the sensor response to the target gas as a function of time of exposure to target gas.

TABLE I. THIN FILM COMPOSITIONS (mol%)

ZnO	100%	75%	50%	25%	-
SnO_2	-	25%	50%	75%	100%

In order to extract relevant key features from the data in terms of the static change of the sensor parameter (*i.e.* resistance or conductivity), we used a fractional difference model [9] where $X_{ij}^{g}\mathrm{X}_{\mathrm{ij}}^{\mathrm{g}}$ is the response of the sensor i to the target gas j, and X_i^0 is the response of the sensor to the synthetic air (1).

$$X_{ij} = \frac{X_{ij}^{g} - X_i^0}{X_i^0} \quad (1)$$

The experimental process can be represented by the response vector $R_q = \left(R_{1q}, R_{2q}, R_{pq}, \ldots, R_{Cq}\right)^T$ where R is the response current, C is the number of concentrations, T is the number of temperatures, p is the concentration and q is the operating temperature.

For each sensor device a temporal response curve (current *vs.* time) was depicted and smoothed using several different fitting algorithms. After residual analysis the parameters of each response curve were evaluated, fits generated and the curve reconstructed to determine the accuracy of the models. Subsequently, an optimal model was selected for generating the best polynomial model. The performance of the SGF was compared with a moving average filter (MA), local regression smoothing (lowess and loess), and robust local regression (rlowess and rloess) methods. Details of these methods and the results of the experiments are provided in the following sections.

A. Moving Average Filtering (MA)

An MA, equivalent to low pass filtering, was used to smooth data by replacing each data point with the average of the neighbouring data points within a specified span of data points as described by the difference equation (2) where $y_s(i)$ is the smoothed value for the i^{th} data point, N is the number of neighbouring data points on either side of $y_s(i)$, and $2N+1$ is the span.

$$y_s(i) = \frac{1}{2N+1}\left(y(i+N-1) + \cdots + y(i-N)\right) \quad (2)$$

B. Local Regression Smoothing: Lowess, Loess, rLowess & rLoess

The lowess and loess smoothed values are determined by considering neighbouring data points. The process is weighted using a regression weight function that is defined for all the data points contained within the specified span. Lowess and loess are differentiated by the model used in the regression: lowess uses a linear polynomial, while loess uses a quadratic polynomial.

The robust local regression methods (rlowess and rloess) differ from lowess and loess in that a lower weight is assigned to outliers in the regression, and a zero weighting is given to data outside six mean absolute deviations. This robust approach typically gives results that are more resistant to outliers.

C. Savitzky–Golay Smoothing Filter (SGF)

The SGF is a generalization of a moving average filter [10]. SGF coefficients are determined by an unweighted linear least-squares regression and a polynomial model of specified degree. The procedure consists of replacing the central point p of a window (2p+1) with the value obtained from the polynomial fit. The window is moved one data point at a time until the entire signal is scanned creating a new smoothed value for each data point. The smoothed signal $g(t)$ is calculated by convolving the signal $f(t)$ with a smoothing (or convolution) function $h(t)$ [11] for all observed data points p where $f(m)$ is the curve function at point m and $h(m-t) \neq 0$ (3). The convolution function $h(t)$is defined for each combination of degree of the polynomial and window size.

$$g(t) = f(t) * h(t) = \frac{\Sigma f(m)h(m-t)}{\Sigma h(m)} \quad (3)$$

In SGF each data point f_i is replaced with a linear combination of g_i (4) and a number of nearby neighbours n where nL is the number of neighbouring points prior to the data point i, nR is the number of neighbours after data point i, and the coefficients c_n are the weights of the linear combination [12].

$$g_i = \sum_{n=-nL}^{nR} c_n f_{i+n} \quad (4)$$

The moving window average (5) is computed as the average of the data points from *fi-nL* to *fi+nR*, for some fixed $nL = nR = M$ and the weights $cn = 1/(nL + nR + 1)$ [13]:

$$g_i = \sum_{n=-M}^{M} \frac{f_{i+n}}{2M+1} \quad (5)$$

The weights c_n are chosen in such a way that the smoothed data point g_i is the value of a polynomial fitted by least-squares to all $(nL + nR + 1)$ points in the moving window. That is, for the group of $2M+1$ data centered at $n = 0$ the coefficient of the polynomial is obtained by (6).

$$c_n = p(n) = \sum_{k=0}^{N} a_k n^k \quad (6)$$

This minimises the mean-squared approximation error (7) for the group of input samples centred on $n = 0$:

$$\varepsilon_{\mathrm{N}} = \sum_{\mathrm{n=-M}}^{\mathrm{M}}\left(\mathrm{p(n)\text{-}x[n]}\right)^2 = \sum_{\mathrm{n=-M}}^{\mathrm{M}}\left(\sum_{\mathrm{k=0}}^{\mathrm{N}} \mathrm{a_k n^k}\text{-}\mathrm{x[n]}\right)^2 \quad (7)$$

Therefore g_i the smoothed data point [13] is given by (8).

$$g_i = \frac{\sum_{n=-nL}^{nR} c_n f_{i+n}}{\sum_{n=-nL}^{nR} c_n} \quad (8)$$

IV. RESULTS

Local regression (lowess and loess) and robust local regression (rlowess and rloess) smoothing were carried out using a span of 10% of the data points. The results from smoothing the raw data using local regression smoothing and

robust local regression were found to give essentially the same shape resolution as the raw data, for all the sensor devices employed in this study, regardless of sensor composition, operating temperature or target gas concentration (see Fig. 1, Fig. 2 and Fig. 3).

MA and SGFs were used to smooth the data using a span of 5 and 55. For both methods it was found that using a span of 55 gave better smoothing/shape resolution than using a span of 5 as shown in Fig. 2 and Fig. 3. However, in the case of moving average, although smoothing was improved the approximation of the curve was poorer because less raw data points were fitted (Fig. 2). It was also found that using SGF, the height and width of narrow peaks is accurately captured by higher degree polynomials, but wider peaks are poorly smoothed. For optimality, a polynomial degree of three was applied for the implementation of the SGF smoothing phase.

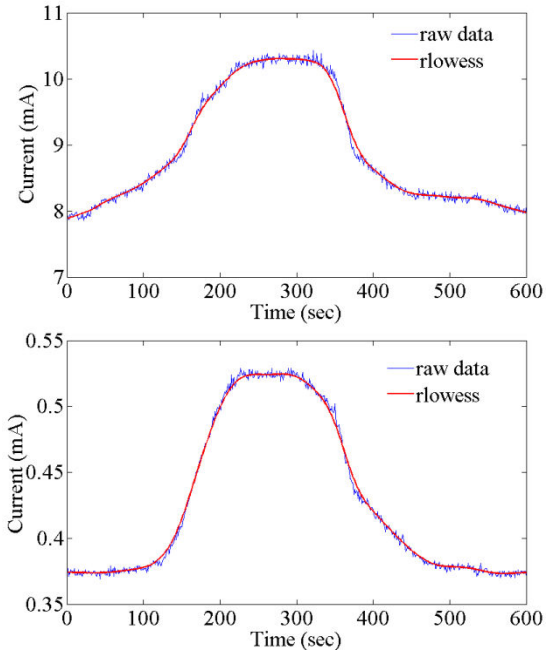

Figure 1. Example of rlowess smoothed signals for ZnO:SnO_2 composites both with 150 ppm methanol at 250 ºC, (top) 50:50, (bottom) 75:25.

Curve fitting was undertaken for each of the smoothing processes, and the coefficient of determination (R-squared (R^2)) was calculated using a polynomial of three (Fig. 4). R-squared indicates how well data points fit a statistical model, and provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model. In other words, R^2 is proportional to the variability of the response signal in the polynomial model.

In our case, R^2 indicates the proportionate amount of variation in the response signal explained by the independent variables t in the polynomial model where SSE is the sum of squared error, SSR is the sum of squared regression, and SST is the sum of squared total.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad (9)$$

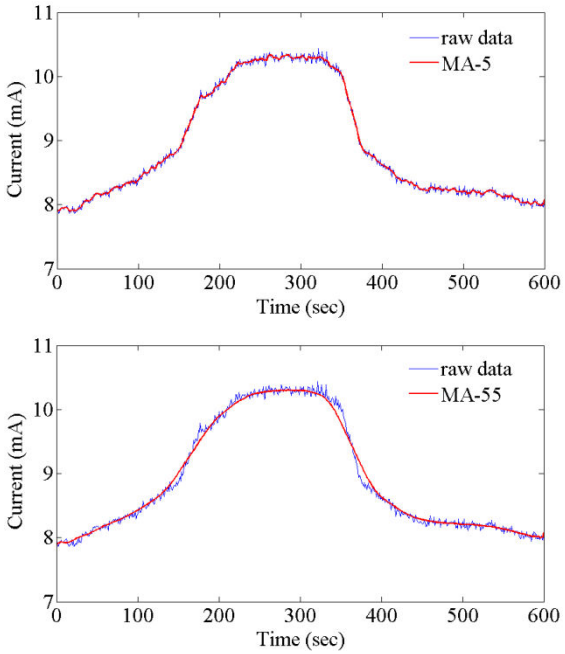

Figure 2. Example of Moving Average method, composite 50:50 with a window span of, (top) 5, (bottom) 55, both with 150 ppm methanol at 250 ºC.

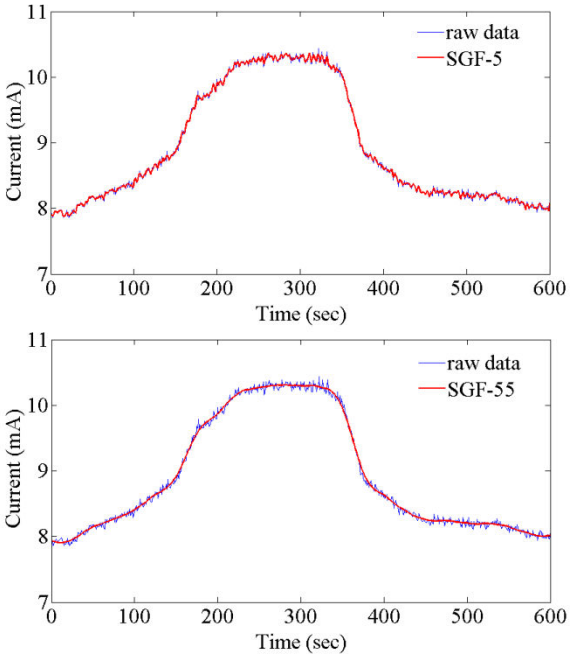

Figure 3. Example of SGF, 50:50 composite with a window span of, (top) 5, (bottom) 55, both with 150 ppm methanol at 250 ºC.

Table 2 gives the R^2 values obtained using a polynomial of 3. The R^2 value of the 100% SnO_2 device was greater than that of the 100% ZnO device and the higher the proportion of ZnO in the composite thin film the lower the R^2. The higher the R^2 value the better the fit. The best value of R^2 was observed for all devices when a MA with a window span of 55 was employed. However, considering R^2 alone is not sufficient as it is also important to understand the data.

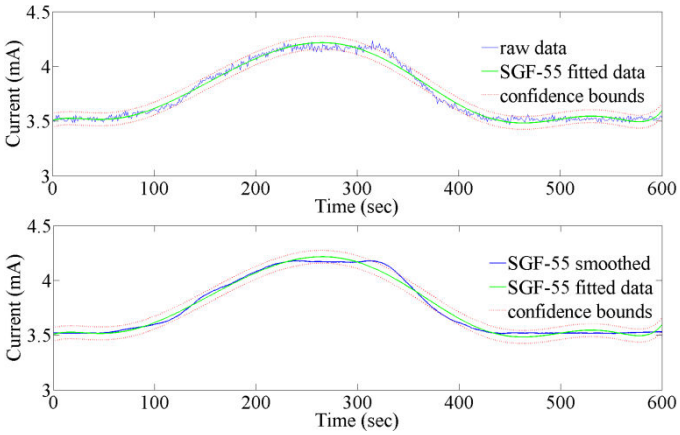

Figure 4. Fitting the confidence bounds 25:75 ZnO:SnO_2 device 150 ppm methanol at 250 ºC, SGF-55; (top) raw data, (bottom) smoothed data.

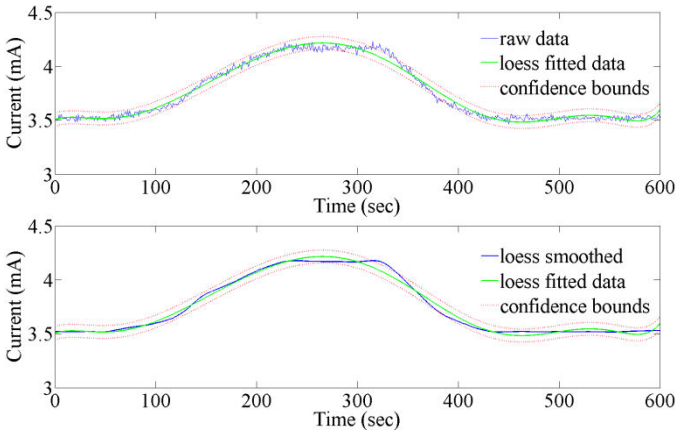

Figure 5. Fitting the confidence bounds 25:75 ZnO:SnO_2 device 150 ppm methanol at 250 ºC, loess; (top) raw data, (bottom) smoothed data.

On the basis of the depictions of the smoothing results, the raw data, and the confidence bounds both the SGF and the loess methods gave better shape resolution than the other methods evaluated for all the sensor devices tested (Fig. 4 and Fig. 5). However, overall SGF was determined to be the best approximation of the sensor response for all the sensor devices and configurations tested because the best resolution was achieved and shape of the signal was maintained.

SGF was then evaluated using various polynomial models. It was found that the ninth degree of the polynomial model provided the best fit to the raw data.

TABLE II. THE COEFFICIENT OF DETERMINATION FOR EACH METHOD AND SENSOR DEVICE (150 PPM METHANOL AT 250 ºC).

Smoothing method	**R^2**				
	Composite – ZnO:SnO_2				
	100:0	**75:25**	**50:50**	**25:75**	**0:100**
lowess	0.84224	0.70462	0.770123	0.74419	0.94374
rlowess	0.83671	0.69671	0.760713	0.73663	0.93501
loess	0.84256	0.70464	0.770499	0.74597	0.94386
rloess	0.83697	0.69630	0.760816	0.73790	0.93419
MA-5	0.83581	0.69651	0.759855	0.73596	0.93173
MA-55	0.84678	0.71173	0.776481	0.75046	0.94906
SGF-5	0.83512	0.69595	0.759182	0.73490	0.92765
SGF-55	0.83683	0.69685	0.761543	0.73707	0.93603

V. CONCLUSION

This study has explored signal response and methods for extracting the desired digital signal while maintaining the shape and resolution of that signal.

A simple procedure to test different polynomial models, with confidence bounds, on the raw data was developed for easy application to quantised gas sensor response data.

Curve fitting approaches were used to validate the results of three possible methods. Of the methods investigated, and as expected, it was found that an SGF gives the best approximation of the sensor response for all the sensor devices tested. This is a promising result and future work will involve testing the SGF signal pre-processing method on signals produced from an array of sensors.

It should be noted that the degree of the polynomial used was a compromise in which we chose the one which gave the best result when considering all the variables of the experiments. This approach was taken in order to be able to compare the results so that a generalizable data approximation approach could be developed.

A generalizable method should be able to be applied to the pre-processing of sensor signal data without specific tailoring for each device composition and its operating conditions. If we did tailor the approach to each individual scenario, it is likely that the results reported here could be improved but this would be at the expense of developing a standard procedure for signal pre-processing. Because of the possible generalizability of our approach, it is also likely that the method developed here might be applied to other types of signals such as those obtained from other types of sensors (for example, rainfall gauges). Further work is required to evaluate the degree of generalizability of our method.

REFERENCES

[1] C. Wang, L. Yin, L. Zhang, D. Xiang, and R. Gao, “Metal Oxide Gas Sensors: Sensitivity and Influencing Factors,” Sensors, vol. 10, 2010, pp. 2088-2106.

[2] I. García-Pérez, M. Vallejo, A. García, C. Legido-Quigley, and C. Barbas, "Metabolic fingerprinting with capillary electrophoresis," J. Chromatogr. A, vol. 1204, issue 2, 2008, pp.130-139.

[3] A. Gutierrez-Galvez, “Coding and learning of chemosensor array patterns in a neurodynamic model of the olfactory system,” Ph.D., Texas A&M University, United States – Texas, 2006.

[4] K. J. Johnson, B. W. Wright, K. H. Jarman, and R E.. Synovec, "High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis," Journal of Chromatography A, vol. 996, issues (1–2), 2003, pp. 141-155.

[5] C. Perrin, B. Walczak, and D. L. Massart, "The Use of Wavelets for Signal Denoising in Capillary Electrophoresis,” Analytical Chemistry, vol. 73, issue 20, 2001, pp. 4903-4917.

[6] L. Bao, J. Mo, and Z. Tang, "The Application in Processing Analytical Chemistry Signals of a Cardinal Spline Approach to Wavelets," Analytical Chemistry, vol. 69, issue 15, 1997, pp. 3053-3057.

[7] A. Savitzky, and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," Analytical Chemistry, vol. 36, issue 8, 1964, pp. 1627-1639.

[8] E. Bassey K. Prasad, and P. Sallis, “Analysis of metal oxide nanostructures prepared for gas sensing,” Workshop on Smart Sensors Instrumentation and Measurement, May 2013. [Online]. Available from:

http://www.geo-informatics.org/publications/Analysis of metal Bassey _GRC 20IM Workshop 2013.pdf

[9] E. L. Hines, P. Boilot, J. W. Gardner, and M. A Gongora, "Pattern Analysis for Electronic Noses," in Handbook of Machine Olfaction: Electronic Nose Technology. T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, Eds. Wiley-VCH Verlag GmbH & Co. KGaA, 2004, pp. 133-160.

[10] A. Savitzky, and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," Analytical Chemistry, vol. 36, issue 8, pp. 1627-1639, 1964.

[11] C. Perrin, B. Walczak, and D. L. Massart, "The Use of Wavelets for Signal Denoising in Capillary Electrophoresis," Analytical Chemistry, vol. 73, issue 20, pp. 4903-4917, 2001.

[12] R. Schafer, "What Is a Savitzky-Golay Filter?," Signal Processing Magazine, IEEE, vol. 28, issue 4, pp. 111-117, 2011.

[13] M. Zuppa, C. Distante, P. Siciliano, and K. C. Persaud, "Drift counteraction with multiple self-organising maps for an electronic nose," Sensors and Actuators B: Chemical, vol. 98, issues 2–3, 2004, pp. 305-317.