Application of mathematical models and digital filters and their Processors of spectral analysis for aromatic compounds gas in a fluorescent chemical

AL-IBADI ZEYAD, MUTHANA ALBOEDAM, ILYA KATANOV, AL-ZUBAIDI SURA

Abstract

This work studies the effect of different methods of spectra processing on the aromatic compounds (benzene, toluene and xylene). Keeping up high-spatial objectives is progressively basic and various methodologies are utilized to check fragrant compounds that anticipate choosing a specific technique from research centre determinations. One notable part of all types of signal systems is the flexibility of adaptation. In addition, a spatial exactness is not fundamental to get a range of an expansive number of fragrant compounds where more prominent characterization and statistical mean are critical. Moreover, sufficiently low deviations of the expected values are achieved from the true values and the standard deviation to determine the properties of fragrant compounds compared with those of the aromatic compounds. A persistent baseline rectification is smoothed and performed followed by normalizing the rectified spectrum to their area. The auto fluorescence foundation is subtracted, for the pure range analysis, by utilizing scientific approaches: polynomial estimation (Poly Fit) and method Processors Gases Improved. The accuracy and reliability obtained are not complete and can be increased by developing algorithms, selecting other parameters and improving the quality of the training sample by eliminating the unwanted data. This could be done by increasing the sample size and studying it in more detail to avoid inaccuracies during the transition between concentrations Gas.

Keywords: Poly Fit, Polynomial, Processors Gases, Ultraviolet spectroscopy, sensors, fluorescence
1. Introduction

Aromatic compounds are broadly utilized in cutting edge industry. Thus, it is necessary to know their concentration in the air, particularly in industrial and populated ranges due to their harmfulness at parts per billion concentrations [14]. These compounds such as benzene, toluene and xylene, cause extraordinary harm to human health, for example, anaemia, sterility and a few sorts of cancers. They also cause harms to the environment. In spite of the health issues caused by these aromatic compounds, we could see the production of these compounds increments universally [18]. In the last twenty years, there have been used sensors to the analytical regarding chemical compounds in the environment in all diverse branches of industry and medicine. Different sensing is actually developed and it is used in numerous applications such as medical-related and biochemical industry and in safety systems and food-packaging sector [5]. To perform a dedicated look for optical-sensor materials and produce them, one should be capable of computing the spectral properties. These materials also have functional characteristics typically the sensor materials under adjustable and reproducible conditions [4]. Optic sensors based on fluorescent reagent-mediated principle usually are made up of sensor stuff and a sensor program. The sensor material includes indicator molecules immobilized over a polymer or other form of a matrix [9]. The sensor stage is closely related to the physical principle used to study the interaction of the sensor material and the technique of manufacturing. The reliable components of the fluorescence-reading platform are the source of the light to excite the fluorescence and to isolate the filtered fluorescent light from excitation light and photo detector. Yet, this method features certain limitations because of the need to make use of numerous chemical sensory elements concurrently, which makes calibration difficult. This requires getting alternative ways to overcome these types of difficulties [17]. In this study, the algorithm is developed in order to solve the problem of fluorescence spectrum analysis [8]. The study aims to use the particular information extracted from the spectral data. Furthermore, it attempts to typically increase the accuracy of calculating the particular concentration of aromatic hydrocarbons in the air. This efficiently allows the spectroscopy of the gas fluorescence determine the concentration [7].

2. MATERIAL AND METHOD

2.1. Experimental setup

In this work, (Dibenzoylmethanato) boron difluoride is prepared added to the method portrayed by Brown and Bladon [1]. We checked multicomponent blend concentrations; they are recorded employing a sensor comprising of a fabric blended on a glass slide. The complex optical fiber test is associated with the back of the chip. The glass slide with the sensor fabric is found in a fixed cell in which clean air or a mixture of air is provided with aromatic hydrocarbon vapors at a certain concentration. The sensor is associated with ultraviolet (LED Nichia, $\lambda = 375\, \text{nm}$) and detected with the Ocean Optics S2000 fiber spectrometer. In addition, the filtration of detected radiation with Lee Filters 226 LEE U.V. a 200-millisecond pulse signal detection scheme and a 185-second period are used.

2.2. Specimens Preparation

Sets of aromatic hydrocarbons are prepared from the laboratory in the Samara State University. Combinations of the concentrations of Barometers are Gas supply for Naphthalene (0.2 g / min), benzene (44.0 g / min), toluene (10.1 g / min), xylene (2.4g /min), and temperatures at which the IMP is maintained for naphthalene (50°C), benzene (110°C), toluene (100°C), xylene (90°C). Air purge flow, ml/min 200. They are introduced into the cell, and directly recorded the spectral
characteristics. We performed our study for more than 1300 samples. The experiments of the samples are performed within 185 seconds from the minute of testing. A standardized recording of the spectra is carried out. Here is vital to neglect the values of the taken tests and their estimations at which the concentration of all gases is zero in the study. The information and storage of major components compatible with the spectral characteristics information are pre-processed with MATLAB (R2013a, Math Works).

2.3. (Processors Gases) algorithms, filtering signals for removing the background of auto fluorescence

The whole signal processing process can be divided into six stages. The first starts with the help of the data logger. Here, we get real "raw" spectra. In this work, aromatic gas sensor spectra are used. In the next stage, we focused on a group of these spectra. The cut spectra are still very noisy. In the third stage of data processing, we remove the noise to make spectra more useful, and in the final stage, another non-informative parameter is deleted. After the background is successfully removed, the "pure" spectrum is obtained, consisting of peaks corresponding to the time (or wavelengths) in which molecular vibrations occur in the test sample. These processes are shown in Figure 1:

![Flowchart of data processing](image)

A monochromatic light occurrence on a sample is characterized by some phenomena such as reflection, scattering, or absorption. In the common case, the scattering process competes with its absorption process. When radiation is absorbed, the particle moves to a less energizing electronic state. The particle returns to the ground state can be either totally lighted or went with by light radiation at a lower recurrence. This radiation is called luminosity (Luminescence). Through mathematical models for recording signals, the signal received from the sensor can be spoken to by some
included scientific models that comprise the spectrometer signal and the foundation of self-sparkle and clamour (noise of all sorts) \[12\]. Because the noise filtering nature in signals is diverse, and to remove impulse noise at the preliminary stage, we apply a median filter. Its main advantage is that it reacts with less quality to readings that stand out sharply against the background of neighbouring ones. This allows it to be widely used to eliminate anomalous values in data arrays \[10\]. In addition, the median filter turns out to be more efficient than linear filters in cases when the differences in signal values are large compared to the variance of the Gaussian noise. This is because it gives a lower value of the root mean square error of the output signal in relation to the input noise-free signal when compared with optimal linear filters \[16\]. In general, as the arithmetic means of two averages are as follows. The output signal \(y_k\) of a sliding median filter with a width of \((2n+1)\) for the current sample \(k\) is formed from the input time series \(\ldots, x_{k-n}, x_k, x_{k+1}, \ldots\) in accordance with formula (Equation 1):

\[
y_k = Me(x_{k-n}, x_{k-n+1}, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_{k+n-1}, x_{k+n})
\]

Where \((x_1, \ldots, x_m, \ldots, x_{2n+1}) = x_{n+1}, x_m\) elements of the variation series i.e. ranked in ascending order of values \(x_m\):

\[
x_1 = min(x_1, x_2, \ldots, x_{2n+1}) \leq x_1 \leq x_2 \ldots \leq x_{2n+1} = max(x_1, x_2, \ldots, x_{2n+1})
\]

The width \((2n+1)\) of the median filter is chosen to suppress a pulse with a width of \(n\) samples. The smoothing procedure is carried out using the Savitsky-Golay filter. The Savitsky-Golay filtering method is better than the simple averaging because it retains the features of the spectrum such as the height and width of the peaks. The main advantage of this approach is that it tends to maintain local extrema, which are usually smoothed out by other averaging methods. The formula for the Savitsky-Golay filter is shown in Equation (2):

\[
S_i^* = \sum_{j=-n}^{n} c_j * S_{i+j} \quad (i = n+1, N-n)
\]

Here, \(S_i^*\) is the samples of the filtered signal; \(c_j\) - coefficients of Savitsky-Golay while \(n\) - is the half-width of the smoothing window. Savitsky-Golay coefficients are calculated by the formula (Equation 3):

\[
c_j = C_{1,j} \quad (j = 1, 2n+1)
\]

where \(C = (A^T.A)^{-1}.A^T\)

\[
\left\{
(i - n - 1)^{j-1}, i \in [1; n] \cup [n + 2, 2n + 1], j \in [2, K + 1] \\
i = n + 1; j = 1
\right\}
\]

The width \((2n+1)\) and the regression order \(k\) of the smoothing filter are chosen to avoid significant distortion or loss of informative signal components \[10\]. The overall filtering scheme is explained in Figure 2:

The background removal is one of the most difficult tasks for a qualitative analysis of experimental results. A variety of techniques encompasses either instrumental or software approaches to remove background fluorescence signals. The software methods have become the standard for correcting the contribution of fluorescence to the background. They do not require modification of the system and have no restrictions on the sample preparation. Among the mathematical methods, the first and
second derivative methods, frequency filtering, polynomial approximation, and wavelet transform are suggested as useful methods for background removal in certain situations [17]. Due to the difficulty of detecting peaks, the loss of some of them can cause the spectrum to match the background, leading to a poor estimate of the latter. Filtering based on the Fourier transform, one of the frequency filtering methods, requires the separation of the frequency components of the spectrum from the background and noise [13]. The polynomial approximation, due to its simplicity and convenience, has become the most popular method for removing fluorescence background in a wide variety of situations. However, a manual polynomial approximation requires user intervention to select areas where the curves coincide with the original data (to determine the upper and lower limits of the frequency range), not only does this take a lot of time, but also the limits change from case to case [2,3]. Although automatic polynomial approximation eliminates human intervention, their use is limited due to the high noise level. Wavelet transform methods can also be used with an automated curve fitting. However, choosing the appropriate wavelet transform threshold and the required level of resolution will cause the baseline display affect the background removal results [6, 15]. To obtain a “pure” spectrum, it is necessary to subtract the auto fluorescence background from the original filtered and smoothed signal. The background can be modelled by a polynomial (Equation 4), the order of which is selected on the basis of its effective removal minimizing the elimination of peaks in the spectrum signal [8] as follows:

\[ p(x) = p_1x^n + p_2x^{n-1} + \cdots + p_nx + p_{n+1} \]  

(4)

The modified polynomial algorithm is based on successive approximation. Thus, the first polynomial differs from polynomial \((i - 1)\) by less than some value of \(\epsilon\). It is generally accepted that spectra are identical if they differ by no more than 5% \((\epsilon \leq 0.05)\) (Equation 5):

\[ \sigma = |\sigma_{i-1} - \sigma_i|/\sigma_i < \epsilon \]  

(5)

In this equation, \(\sigma_{i-1}\) is the standard deviation of the \((i - 1)\) polynomial; \(\sigma_i\) is the standard deviation of the \(i\) polynomial, in more details in this method. It all starts with an approximation with a single polynomial \(p_1(v)\) according to formula (Equation 4), using the ”raw” signal of light \(O_0(v)\), where there is a displacement of the spectrum. Then the standard deviation is calculated by the formula (Equation 6):

\[ \sigma = \sqrt{(R(v_1) - \bar{R})^2 + (R(v_2) - \bar{R})^2 + \cdots + (R(v_2) - \bar{R})^2/n} \]  

(6)

The removal of peaks is organized according to the following principle. To minimize the distortion of the approximating polynomial, the main peaks are determined based on the inequality of the original signal which is greater than the polynomial and the standard deviation. Now the ”clean” spectrum is obtained by means of the difference between the polynomial approximation from the original signal [7] as in Equation (7).

\[ R(v) = O(v) - P(v) \]  

(7)
This equation is the difference between the approximation of the polynomial and the original signal, and \( \bar{R} \) is the average value of these differences. Then, the principal peaks are determined based on the inequality \( O_0(v) > p_1(v) + \sigma_1 \). Here the points correspond to the major peaks and are removed from the spectrum. These points are not followed in the subsequent steps. Peak removal is necessary to prevent unnecessary outliers in the data. The importance of this study is due to the need to study the spectral properties of aromatic compounds. Expected practical results in the course of the work are to develop an algorithm for processing spectral data with a low signal-to-noise ratio. Also, the methods of modelling the spectra of aromatic compounds, the comparative analysis and the developed algorithm will help researchers in the future in the completion of their work.

3. RESULTS

3.1. Spectral characteristics of Aromatic compounds

UV absorption spectrometry is a well-known analytical technique. In addition, it has been applied widely in analytical chemistry [11]. Absorption is definitely an attenuation of electromagnetic radiations at specific energy as it passes by means of an analyse. Electromagnetic radiation with high energy (lower wavelength) will enable the electron to excite by lower energy state to a high energy state. Beer-Lambert law is a constrained law valid for minimal concentrations of an analyse and for a monochromatic beam, provided that typically the physical and chemical attributes of the absorbed varieties tend not to change with the concentration. In absorption spectroscopy, the long optical route length and strong intake cross-sections are usually associated together with high sensitivity. Nevertheless, sensor saturation may occur, which often may limit sensitivity and even linearity [11].

3.2. Analysing Initial Data

The spectra of the sensor signal of the data logger are a set of readings, pairs of numbers, corresponding to the wavelength between time and the intensity of the background radiation from the sample. The time data in such as a range recorded not only auto fluorescence but also the phenomena that precede it as in Figure 3. For further data processing, we selected the range of interest after removing auto fluorescence. We selected distinct peaks. This work will be dedicated only to benzene, toluene, and xylene. The spectrum of aromatic compounds under study in the range of interest is shown in Figure 4.
3.3. Analysing the spectral properties of Aromatic compounds in gases using the improved method

As mentioned previously, the nature of noise filtering in signals is varied. Thus, we can use one of the two filters to remove the noise (mainly the pulse) such as an intermediate filter or noise filtering spectrum such as Savitsky-Golay filter, shown in Figure 5.
The signal from the output of the spectrograph $O(\lambda_n)$ is represented by some additive models. These models consist of the scattering signal $RS(\lambda_n)$ and the background of auto fluorescence $P(\lambda_n)$ (usually a smoothly changing line) dotted line, shown in Figure 6.

To obtain a "pure" spectrum, it is necessary to subtract the auto fluorescence background from the original filtered and smoothed signal. The background can be modeled by a polynomial (Equation 4). Its order is selected based on its effective removal while minimizing the subtraction of signal peaks. Based on our experiments, 7th and 9th order polynomials give the best approximation as in Figure 7.
we study and apply the proposed and developed method (Processors Gases) to each components of aromatic compounds separately. This is an iterative algorithm based on sequential approximation until the \((i-th)\) polynomial differs from the polynomial\((i-1)\) by less than some value of \(\epsilon\). When the points in the current spectrum exceed the corresponding polynomial approximation, the standard deviation are replaced by the polynomial points with the mean square error considering the noise effects avoiding artificial peaks. Otherwise, the spectrum points remain unchanged \([3]\). This method takes into account the distortion of the signal noise and the effect of the peaks of clean, large aromatic compounds on the polynomial approximation. However, the first and second iterations add extra parts, from removing the peaks and so on, to prevent unnecessary outliers in the data \([8]\). At the second and subsequent iterations, we perform the same procedure until the condition \((7th)\) is satisfied. The polynomial obtained at the last iteration will be considered as the auto fluorescence background. A clean spectrum emerged from subtracting this polynomial approximation in the original signal \([12]\).
Figure 8: Smoothed spectra with a highlighted autofluorescence background (top) and "pure" spectra (bottom) for the polynomial (7th) from the signal spectrum aromatic compounds for benzene.

Figure (8) shows that there is a very large discrepancy between the benzene related signal in the upper part that is pre-treated and the pure signal obtained through the use of the algorithm in lower part. This entails that the autofluorescence signal is identified and removed afterwards to obtain Pure signal.

Figure 9: Smoothed spectra with a highlighted auto fluorescence background (top) and "pure" spectra (bottom) for polynomial (7th) from the signal spectrum aromatic compounds for toluene.

However, Figure (9) reveals that there is only a slight difference between the signal related to Tulane in the upper part and the pure signal obtained through the use of the algorithm in the lower part. This is when auto fluorescence is detected and then removed to obtain a pure signal. The indication in Figure 10. of pure xylene is remarkably distinct.
Figure 10: Smoothed spectra with a highlighted autofluorescence background (top) and "pure" spectra (bottom) for the polynomial (7th) from the signal spectrum aromatic compounds for xylene.

Figure 11: The scattering of the deviations from the reference value prior to processing and using an average filter for each portion of aromatic compounds independently.
Figure 12: Scatter of deviations from the reference value (median filter) for each portion of aromatic compounds independently

Figure 13: Scatter of deviations from the reference value (Poly Fit) for each portion of aromatic compounds separately
4. DISCUSSION AND CONCLUSION

Our study took care of the effect of different methods of spectra processing on the principal components (Aromatic compounds), the spectral properties of the compounds and the characteristic peaks. The aim of this work is to identify the best method for preserving the fundamental components from scattering spectroscopy data. In this work, various methods of removing different distortions are investigated according to the initial model of signal observation (registration), such as registration noise and background auto fluorescence. The main task is to maintain an optimal balance to minimize the loss of the necessary information and at the same time to remove all unnecessary noise. Files with spectroscopy data from the scattering sensor are inserted into the program. The raw spectra are processed (procedures Med Filter, SGF, Poly Fit, Processors Gases), the result of which is “pure” spectra. Further, the obtained sets are compared with sets from a priori data. According to the comparison data, the influence of distortions is introduced by various algorithms for noise reduction and background removal of auto fluorescence. The scatter diagrams (whisker-plot) are constructed for the introduced absolute distortions of the principal components during the preprocessing (stages of noise reduction and removal of auto fluorescence). The Savitsky-Golay filter has a smaller error spread compared to the median filter, therefore it is easier to correct this spread. The methods of removing auto fluorescence include the improved method (Processors Gases) that has the smallest error spread compared to other methods. The polynomial points with the root mean square error could substitute the points in the current spectrum that outweighs the corresponding polynomial approximation in terms of the standard deviation and noise effects avoiding artificial peaks. Conversely, the spectrum points remain unchanged. In the subsequent iterations, the same procedure is repeated until the condition (Equation 5) is meet. This will help to obtain that the polynomial of last iteration as auto fluorescence background. From the original signal subtraction, a ”clean” spectrum emerged as in the polynomial approximation l.

Conflicts of interest: No conflict of interest is declared by the authors

ORCID:
ZEYAD AL-IBADI
https://orcid.org/0000-0002-6562-7176

acknowledgments: Sponsor acknowledgments Samara State University
References


