

Environmental pollution detection: A novel chirped spectral modulation algorithm for a more accurate monitoring of gas pollutants in the atmosphere

Mohamed Shalaby*  and Fawzi S. Alorifi

Electrical Engineering Department, Faculty of Engineering, Imam Mohammad Ibn Saud Islamic University, Riyadh 11432, KSA

Received 24 September 2022 / Accepted 15 January 2023

Abstract. This work presents a new technique based on modulating the IR absorbance of each substance in a mixture in a chirped manner to reduce the effect of their partial spectral absorption overlap on the accuracy of determining their concentrations. This chirped spectral modulation CSM algorithm can deal with mixtures containing unknown substances rather than the substances whose concentrations are aimed. This novel algorithm, when compared to existing pattern recognition techniques, makes it easy to analyze the constituents of a mixture with high accuracy in the presence of traces of unknown components. It is found that the new algorithm can detect the presence of gas pollutants such as sulfur dioxide, carbon monoxide, carbon dioxide, nitrogen dioxide in a sample containing many other unknown polluting substances. This new algorithm is tested on air samples composed of predetermined percentages of air constituents and the results of calculations are compared with those of classical least squares CLS pattern recognition algorithm. The comparison showed that the new algorithm can detect down to very small traces of harmful gases such as NO₂, and SO₂, at least one order of magnitude less than those detected by the CLS approach. Finally, the new algorithm is used to examine collected air samples from an industrial zone, and in the middle and at the exit of a road tunnel in Riyadh area which showed that the percentages of sulfur dioxide, nitrogen dioxide, and carbon monoxide are well below the safe levels.

Keywords: Pattern recognition techniques, Environmental pollution monitoring techniques, Pollution detection, Fourier transform infrared spectroscopy, Gas pollutants in the atmosphere.

1 Introduction

Preserving the environment is one of the most important necessities of scientific research, which includes the atmosphere, water resources, and agricultural soil. With industrial progress, the importance of early detection of different types of pollutants increases, and this requires high accuracy and sensitivity [1–4].

The atmosphere in which we live, and all our activities take place, should be kept always clean. We should ensure that it is free from pollutants and substances harmful to human health. What we gain from exercises on the health of our brains is lost due to air pollution. There are still quite limited research results on this aspect to help protect the environment. Regular monitoring of atmosphere purity is very important; therefore, atmospheric air should be checked to control the level of pollutants in it. Developing more accurate techniques for the detection of harmful gases in the atmosphere below the safe levels recommended by

governmental agencies, leads to knowing the sources of pollution before the problem worsens and spreads from one region to another, causing harm to humanity, and containing the problem before it escalates. Harmful gases that may exist in the atmosphere result from human activities and sometimes by nature, such as volcanoes. Focusing on human activities, the sources, and bad effects of each of these gases are presented along with their recommended safe levels.

The recommended level of amount of NO₂ and SO₂ suggested by WHO “World Health Organization” is 25 µg/m³. It is known that up to 50 ppb (60 µg/m³) of carbon monoxide can be approved by government agencies in air [5].

Nitrogen dioxide is considered as one of the most dangerous air polluting substances. Burning of fuel in motor vehicles, power plants, and some industrial processes release NO₂. Human health is endangered due to exposure to NO₂. Inhaling air polluted with high levels of NO₂ affects the respiratory and circulatory systems and the brain, causing death. Motor vehicles and industries based on fossil fuel

* Corresponding author: myshalaby@imamu.edu.sa

consumption are the major drivers of NO₂ pollution in the most polluted cities. Decision makers must embark on a shift from private vehicles to public transportation that uses clean energy.

Sulfur dioxide is a toxic agent. Air polluted by SO₂ may cause stroke, cardiovascular disease, lung cancer, and eventually death. Burning of fossil fuels in vehicles, power plants, and some industrial processes release SO₂ in the atmosphere. Volcanoes and extraction of metals may cause production of SO₂. The conversion of sulfur dioxide into PM2.5 affects the health of millions of people. These suspended particulates are 25–100 times thinner than a human hair and therefore we cannot see it. The reason why these particles are very harmful to human health is their ability to go through respiratory organs and then to blood stream. The most efficient and easiest way to reduce PM2.5 levels is to install flue-gas desulfurization FGD units. Installed FGD units are critical in the process of reducing SO₂ emissions from power plants as they form a significant fraction of total PM2.5 pollution at different locations. Inhalation of PM2.5 particles causes eyes, nose, and throat irritation. It causes coughing, sneezing and shortness of breath. Long term exposure to PM2.5 can cause permanent respiratory problems such as asthma, chronic bronchitis and heart disease [6]. Children, elder people, and those with problems in heart and in the respiratory tract are more sensitive to health problems because of PM2.5 air pollution. These ambient particulate matter, cause rates of death more than alcohol and diabetes. Research shows that every 10 µg/m³ increase in PM2.5, increases all-cause mortality between 3% and 26%, chances of childhood asthma by 16%, chances of lung cancer by 36% and heart attacks by 44%.

High concentrations of sulfur oxides destroy vegetation and contribute to acid rain which can harm sensitive ecosystems. Sulfur oxides can react with other compounds in the atmosphere to form fine particles that reduce visibility, and damage statues and monuments.

Rules should be imposed to reduce emissions of SO₂ and pollutants that form sulfur oxides (SO_x) to help meet the air quality standards. Areas where the air quality does not meet these standards should be identified in order to develop plans to reduce the amount of SO₂ in the air.

Carbon monoxide is released from burning of fuels. Carbon monoxide high levels may cause poisoning leading to risks such as permanent brain damage, and even death due to the replacement of oxygen by carbon monoxide in red blood cells. Carbon monoxide causes also death of fetal of pregnant women.

On the other hand, carbon dioxide which is also released from industrial activities and burning of fuels causes greenhouse effect and ocean acidification.

Since the safe percentage of the presence of air pollutants in air is very small, a new, more accurate method for detecting these pollutants that differs from the existing methods is to be discovered.

It is a mandatory today to detect and determine the quantity of very small amounts of chemical contaminants down to or less than the safe levels of these substances. Moreover, this process of detection should be quick enough and cheap so that the state of the purity of the air can be

constantly monitored around the clock and in any spot on the planet. Along with the development in the measuring instruments used in the detection of substances, this must be accompanied by a development in the analysis to be more accurate and sensitive. The main feature in adopted pattern recognition techniques is their suitability for the accurate determination of mixtures components and their concentrations. The challenge facing these techniques is that these chemicals have similar spectra with some overlap. Therefore, these pattern recognition techniques should be developed to resolve the considerable spectral overlap and despite that, giving a correct result about the lowest possible limits for the amounts of environmental pollutants. Among the other advantages of these methods, in addition to accuracy, speed, and cheapness, is the determination of the components without previous separation. Least square methods have been used in several studies for the determination of different components in solutions. One of the adopted techniques to find the concentration of very small traces of oil in wastewater to a better accuracy is by transforming the absorption spectra of water and oil into vectors [7]. In this work, the absorption spectra are modulated by multiplying them with a chirped sinusoidal function in wavenumber. In this case, each absorption line is associated with a chirping frequency band different than other lines. This procedure alleviates the problem of the spectral overlap of the absorption spectra of substances in the atmosphere, hence the concentrations of these substances are obtained with better accuracy down to very small amounts less than or around the safe levels recommended by the WHO organization [8]. This technique is not only for laboratory purpose and is applicable for on ground and from the space measurements [9]. Its role comes after measuring absorption spectra. This approach is useful for various applications in different fields of pattern recognition such as people detection, fire detection, facial recognition, etc. [10].

The paper is organized as follows; first the analysis of the new algorithm is presented, then it is tested on samples of air contaminated with known percentages of harmful gases which are SO₂, NO₂, CO, and CO₂. The results of the new algorithm are compared with those of the CLS approach. Then the new algorithm is used to measure the level of polluting gases in an industrial zone, and at the exit and in the middle of a road tunnel. The bases of infrared spectroscopy are presented in [Appendix A](#), and the analyses of the Beer Lambert law, and the solution with the CLS approach are also given in [Appendix B](#) to help in comparing with the new proposed approach. Finally, [Appendix C](#) presents some data related to the studied gases in this work along with some useful relations to transform gases concentrations from µg/m³ to percentage in volume and ppm.

2 Methods

In this work, a new algorithm is designed that is capable of the determination of concentrations of the constituents in mixtures with better accuracy since it is based on mitigating the problem of the partial overlap between the absorption spectra of these substances. This study focuses on NO₂

and SO₂ due to their partial spectral overlap with other components of the air.

The idea is based on modulating the absorption spectrum of these substances in such a way to minimize the effect of this spectral overlap on the determination of each substance concentration in a mixture. The advantage of this proposed technique, in addition to the advantages of existing least squares methods, is its ability to find more accurately smaller traces of unknown components in presence of unexpected substances in the examined mixture. This new technique is suitable for the determination of environmental pollutants [11] and in other fields such as biomedical area for example to detect the amount of glucose in the blood in the near and mid-infrared range [12–16]. More clearly, using this approach does not necessitate any pre-preparation of the examined mixtures before being tested by the IR spectrometer [17].

This new algorithm is tested on samples to check its capability of detecting traces of carbon monoxide, carbon dioxide, nitrogen dioxide, and sulfur dioxide present in air besides other contaminants. Then the new algorithm is used to analyze air samples at an industrial zone during 8 weeks and in the middle and at the exit of a road tunnel in Riyadh area to ensure that their amounts are within the safety ranges.

3 Proposed chirped spectral modulation algorithm

The new method relies on modulating the absorption spectra of the substances in a mixture. Modulation is realized by multiplying the absorption spectra by a signal sinusoidally varying with wavenumber. The frequency of this modulating function is chirped and adjusted in order to accurately calculate the concentrations of these substances. The absorption spectrum of each substance is modulated in this manner and the same is carried out for the absorption spectrum of the mixture which is under analysis. A set of equations is constructed from which the unknown concentrations of the substances are determined.

Since the absorption spectrum of each substance resembles the fingerprint of a human being so that it is impossible for the absorption spectrum of a substance to be similar to that of another substance. Therefore, using this new algorithm, a modified absorption pattern after chirped spectrum modulation is obtained which is less likely to bear a resemblance to that of another substance. Moreover, identification of larger molecules is easier with absorption spectroscopy techniques. The larger the molecule, the more bonds it contains, and thus absorption occurs from this molecule at more wavenumbers.

The spectral absorption pattern is used to calculate a variable representing the absorption spectral summation of substance “*j*”, in a mixture, using the following equation as

$$b_j = \sum_{I=1}^N (I-1)^r k_j^2(I) \cos\left(\varphi(I-1)^2 \times \frac{2\pi}{N}\right), \quad (1)$$

where k_j is the absorption spectrum of constituent j , and j varies from 1 to the number of constituents tested in a mixture, and “*r*” and “ φ ” are two parameters. N is the number of sampling points in the wavenumber range, where $N = \frac{v_u - v_L}{\Delta v}$ and v_u and v_L are the upper and lower wavenumbers in the range of measurements respectively, and Δv is the resolution (wavenumber separation = 4 cm^{-1} in our experiment). The term $(I-1)^r$ in equation (1) represents v^r and plays an important role to distinguish between substances in a mixture if their absorption spectra partially overlap. By the summation of equation (1), the absorption spectrum of this substance is converted into a variable. Similarly, the joint spectral summation of substances “*j*” and “*p*” is given as

$$b_{jp} = \sum_{I=1}^N (I-1)^r k_j(I) k_p(I) \cos\left(\varphi(I-1)^2 \times \frac{2\pi}{N}\right). \quad (2)$$

In case of existing spectral absorption overlap between substances “*j*” and “*p*”, the calculated concentration of substance “*j*” or substance “*p*” will not be correct. The weighting function, v^r , where r is a parameter, and the parameter φ , in equations (1) and (2), are adjusted for more discrimination between the spectra of the studied substances of the atmosphere.

The summation is carried out over the whole spectrum ($448\text{--}3980 \text{ cm}^{-1}$) and this absorption spectrum of all the components in the mixture serves to find the concentration of each substance in this mixture.

After measuring the absorption spectrum of the mixture under study, a variable is calculated for each expected substance in this mixture as follows

$$d_j = \sum_{I=1}^N (I-1)^r A(I) k_j(I) \cos\left(\varphi(I-1)^2 \times \frac{2\pi}{N}\right), \quad (3)$$

where A is the absorption spectrum of the mixture under consideration.

We can construct the following matrix equation

$$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_L \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1L} \\ b_{21} & b_{22} & \dots & b_{2L} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ b_{L1} & b_{L2} & \dots & b_{LL} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_L \end{bmatrix} + U, \quad (4)$$

$$D = BC + U, \quad (5)$$

where matrix C represents the unknown concentrations of the expected substances in the mixture, and matrix U represents the contribution of the unexpected substances in the mixture which affects the accuracy of the calculated concentrations. The solution of equations (4) and (5) is obtained when matrix U equals zero and the unknown concentrations of the expected substances in the mixture is

$$C = B^{-1}D \quad (6)$$

where B^{-1} is the inverse of matrix B .

4 Test of the validity of the new algorithm

Initially, a focus was placed on some air pollutants produced by activities in industrial zones such as carbon dioxide, carbon monoxide, sulfur dioxide, and nitrogen dioxide. The absorption spectrum of these substances in the mid-infrared range is obtained from Spectra plot [18].

To test the ability of the new method to detect the proportions of the substances in a mixture, the spectrum of a sample of air containing percentage of these pollutants is prepared numerically, to know the lowest percentage that the new method can accurately detect in presence of a lot of other pollutants. An amount of water vapor (4×10^{-3} percentage in volume) is added to the mixture representing the unknown substance in the mixture denoted by matrix U in equations (4) and (5). Then this sample is tested using the vectorial algorithm applied on the IR absorption of this mixture.

Figure 1 indicates that the new approach can detect an amount of nitrogen dioxide down to $4.5 \times 10^{-7}\%$ before reaching the noise floor caused by the presence of other unexpected compounds in the mixture. The accuracy is indicated by the point at which the curve deviates from linearity (accuracy of detection is defined as the minimum detectable amount of a substance in an unknown mixture corresponding to a maximum percentage error of 10%). The lower the value of this point, the higher is the accuracy. Figures 2-4 show that new CSM algorithm produce correct values for SO_2 , CO, and CO_2 down to $7.8 \times 10^{-6}\%$, $1.7 \times 10^{-6}\%$, and $2 \times 10^{-3}\%$ respectively. Figures 3 and 4 show that levels of CO_2 and CO detected by the new CSM algorithm differ slightly compared to those obtained by the CLS approach. The level of CO is slightly underestimated whereas the value of CO_2 is slightly overestimated. These levels are below the safe percentages allowed by the World Health Organization. In other words, the CSM approach is not beneficial compared to CLS approach for detection of small amounts of carbon monoxide and carbon dioxide.

5 Experimental results and discussions

The analysis is carried out using the FTIR MB3000 spectrometer operating in the mid IR range with a gas cell of 14 m path length. Since the percentage of pollutants in the air sample in our case is very small, five samples are considered, at different instants, and the average of their spectral absorbance is calculated to minimize the error. Moreover, the absorption spectrum over the range from 448 to 3980 cm^{-1} is measured in the following wavenumber intervals 448-1200, 1200-1500, 1500-2700, 2700-3000, and 3000-3980 cm^{-1} . The gain of the detector circuit is adjusted in an interval in such a manner to increase sensitivity

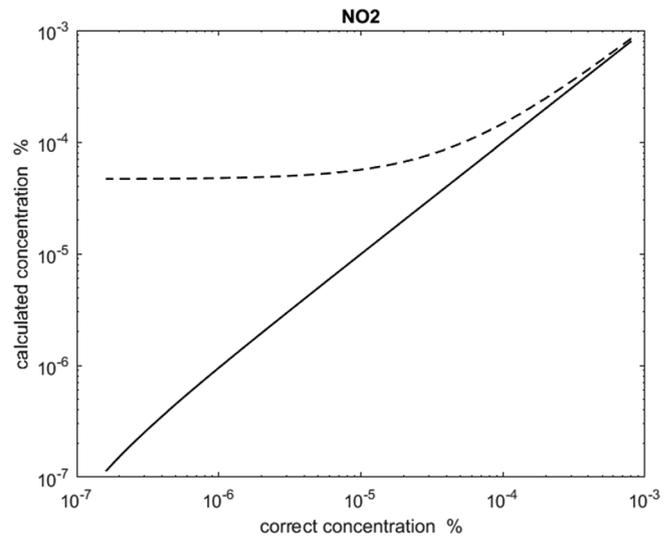


Fig. 1. The amount of nitrogen dioxide detected versus the amount previously added to the mixture, both by the CLS method (dashed) and by our new algorithm (solid). The curves show that the new algorithm shows a linear response down to better than $4.5 \times 10^{-7}\%$ whereas the CLS approach is linear down to about $4 \times 10^{-5}\%$.

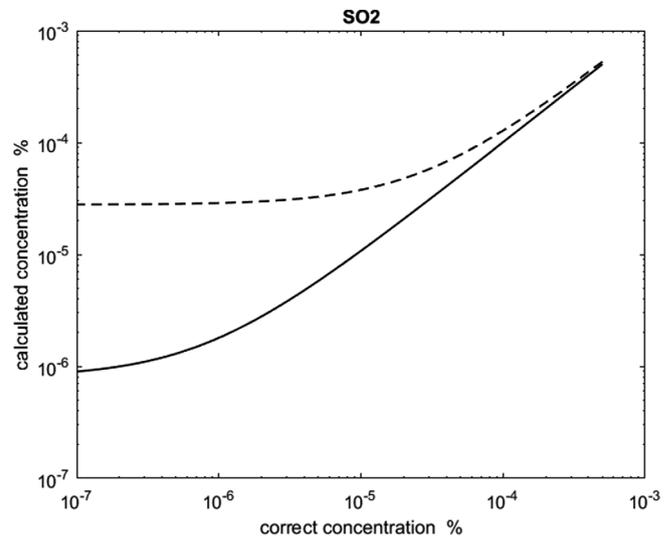


Fig. 2. The amount of sulfur dioxide detected versus the amount previously added to the mixture, both by the CLS method (dashed) and by CSM algorithm (solid). The curves show that the new algorithm reveals a linear response down to around $7.8 \times 10^{-6}\%$ whereas the CLS approach is linear down to about $2.5 \times 10^{-4}\%$.

without running into detector saturation. Finally, these measurements are grouped in one file prior to applying the new algorithm on it. It is relied on the dynamic range, the signal to noise ratio, and the sensitivity of the FTIR spectrometer used to detect the small traces of pollutants in the atmosphere.

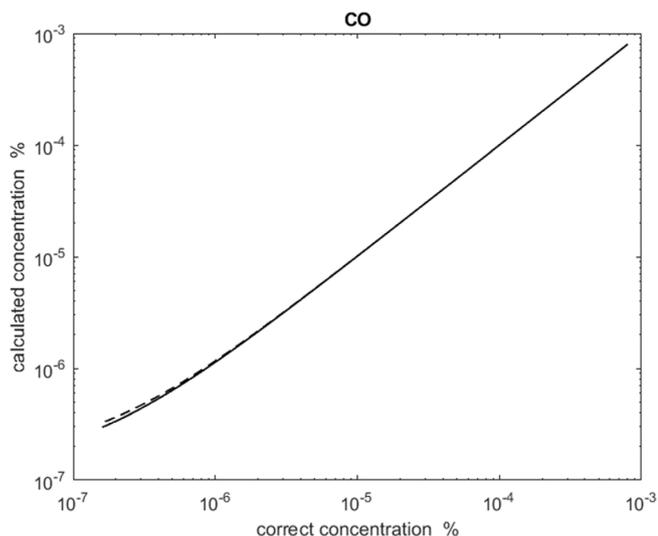


Fig. 3. The amount of carbon monoxide detected versus the amount previously added, both by the CLS method (dashed) and by CSM algorithm (solid). The two approaches show good behavior down to $1.7 \times 10^{-6}\%$.

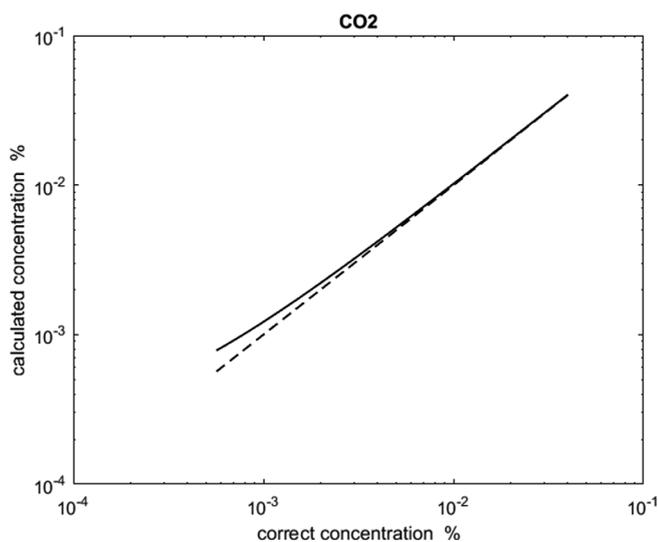


Fig. 4. The amount of carbon dioxide detected versus the amount previously added, both by the CLS method (dashed) and by CSM algorithm (solid). The two approaches show nearly the same behavior down to $2 \times 10^{-3}\%$.

(a) Monitoring air pollution in an industrial area

An industrial zone is chosen to examine the percentage of pollutants in air during 8 weeks. The instant of analyzing the air is at the peak of factories and workshops activities during the day. It should be noted that lifetime of NO_2 and SO_2 is less than 1 day since they are active gases.

The absorbance spectrum of the sample is analyzed using our new algorithm, it was possible to detect traces of carbon monoxide, carbon dioxide, and sulfur dioxide as shown in Figure 5. The error in these measured values,

according to Figures 1–4 is very small and cannot be determined and depends on unexpected constituents of the air sample other than the considered substances.

Moreover, the concentration of CO_2 during 8 weeks is calculated using CSM algorithm as shown in Figure 6.

(b) Monitoring air pollution in the middle and at the exit of a tunnel

The same procedure is carried out by examining air constituents searching for the concentrations of SO_2 and NO_2 in the middle and at the exit of a tunnel at the rush hour (peak of traffic clogging). This tunnel, in center of Riyadh area, is 3 lanes width, 830 m long and equipped with air ventilators distributed along the tunnel Figure 7. This road tunnel is away from any industrial activities to focus on air pollution caused by cars. The concentrations of SO_2 , NO_2 , CO , and CO_2 in the middle of the tunnel are 0.035, 0.041, 0.07, and 413 ppm respectively. The concentrations of SO_2 , NO_2 , CO , and CO_2 at the exit of the tunnel are 0.032, 0.038, 0.065, and 411 ppm respectively. These values are slightly higher than those registered in Figures 5 and 6 but still below the safe values for short intervals of exposure.

According to the analysis of this survey results, and even though the registered values of polluting gases are below the safe thresholds in case of short periods of exposure, the following recommendations are of great importance:

- Follow-up the periodic inspection of flue gas desulfurization filters installed on factors chimneys. Emission standards should be strengthened and flue gas pollution control technology on coal-fired power plants, smelters, and other major industrial SO_2 emitters should be implemented.
- Not licensing factories unless they submit proof of amount of emissions in the safe range.
- The gradual shift towards electric cars in order to reduce carbon dioxide emissions and to get rid of the global warming phenomenon that haunts the world and threatens to melt the ice in the polar regions and to reduce the levels of SO_2 and NO_2 pollution. It's very urgent that governments transfer investments in fossil fuels towards safer energy sources, such as wind and solar.
- Helping factories owners replace their old machines and install new environmentally friendly technologies through soft loans.
- Raise factories owners' awareness to adopt new technological methods reducing air pollutants.
- Periodic monitoring of level of pollutants in the atmosphere.
- The education of management and staff of factories is also a critical part of a successful health protection from dangerous effects of atmospheric pollutants.

In summary, it is believed that this new technique, compared to classical pattern recognition techniques, is able to determine the concentration of very small amounts of substances in a mixture containing the main constituents in addition to traces of other unexpected substances.

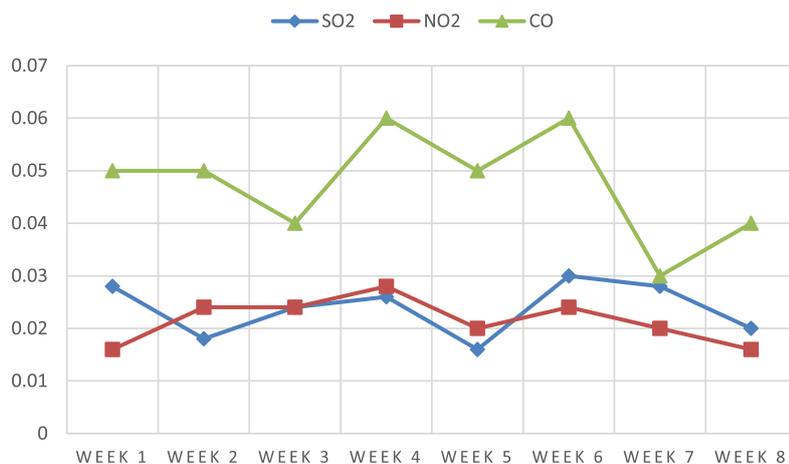


Fig. 5. The results of the field study which has been conducted by analyzing samples from an industrial zone in 8 weeks to monitor the level of air pollutants. The absorbance spectrum of these samples is measured using FTIR MB3000 spectrometer. Then, the CSM algorithm is applied to the obtained absorbance spectrum to determine the amount (ppm) of SO₂, NO₂, and CO in these samples.

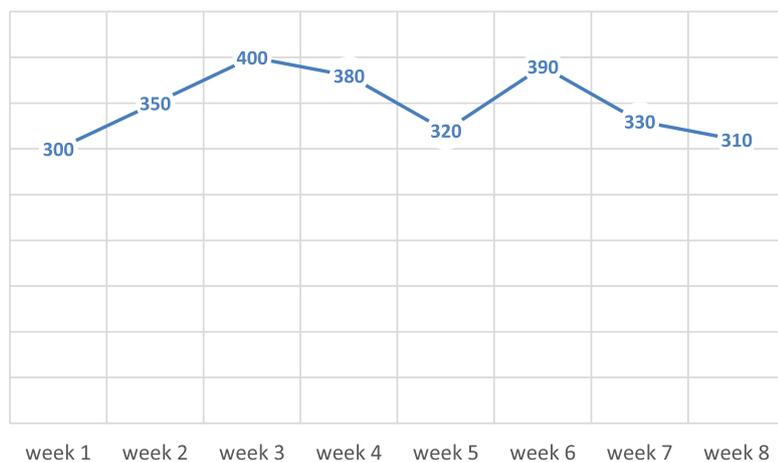


Fig. 6. The results of the field study which has been conducted by analyzing samples from an industrial zone in 8 weeks to monitor the level of air pollutants. The absorbance spectrum of these samples is measured using FTIR MB3000 spectrometer. Then, the CSM algorithm is applied to the obtained absorbance spectrum to determine the amount (ppm) of CO₂ in these samples.



Fig. 7. An aerial view showing the location of the road tunnel in Riyadh city.

The strength of this algorithm is attributed to the increased discrimination between substances, whose spectra is partially overlapping, by multiplying their spectra by an additional function. The function chosen is composed of a weighting function and a cosine function. The cosine function modulates the absorption spectrum with a frequency which will be tailored for more discrimination between the substance's absorption spectra. This algorithm reduces the effect of the other unidentified substances in the mixture.

6 Conclusions

The new algorithm validity is tested by comparing its results with other algorithms based on classical least

squares. The test is run on mixtures with different constituents. It was found that the novel algorithm is more sensitive to detect traces of pollutants in a mixture in presence of other components and gives more accurate concentration values.

The new spectrum modulation algorithm used allows us to detect very small traces of carbon monoxide, nitrogen dioxide, and sulfur dioxide in air in an industrial zone and at the exit of a road tunnel in Riyadh area. The amounts of these polluting substances are well below the internationally recommended levels.

Conflict of interest

The authors affirm that this research work has not any conflict of interests.

Availability of data and materials

The authors are willing to share all the data and infrared absorption spectra obtained during this research work upon request.

Funding

This work received a fund from Imam Mohammad Ibn-Saud Islamic University, Deanship of Scientific Research (research project number 19-12-14-004).

Authors' contributions

All authors contributed to the whole work. All authors read and approved the final manuscript.

Acknowledgments. The authors of this research work are thankful to the Deanship of Scientific Research-Imam Mohammad Ibn-Saud Islamic University who financially supported our research work (research project number 19-12-14-004).

References

- Baszanowska E., Otremba Z. (2014) Spectral signatures of fluorescence and light absorption to identify crude oils found in the marine environment, *J. Eur. Opt. Soc.* **9**, 14029.
- Haule K., Darecki M., Toczek H. (2015) Light penetration in seawater polluted by dispersed oil: results of radiative transfer modelling, *J. Eur. Opt. Soc.* **10**, 15052.
- Troia B., Passaro V.M.N. (2014) Investigation of a novel silicon-on-insulator Rib-Slot photonic sensor based on the vernier effect and operating at 3.8 mm, *J. Eur. Opt. Soc.* **9**, 14005.
- Kim Y.J., Platt U. (eds) (2008) *Advanced environmental monitoring*, Published by Springer.
- Jantrania A. (1991) Dealing with oil and grease in restaurant wastewater, *Small Flows J.* **5**, 1.
- Hadley M.B., Vedanthan R., Fuster V. (2018) Air pollution and cardiovascular disease: a window of opportunity, *Nat. Rev. Cardiol.* **15**, 4, 193–194.
- Shalaby A.R.M., AlMuhanna K.A., Shalaby M. (2020) Environmental pollution monitoring: a novel vectorial algorithm technique for oil detection in wastewater, *Spectrosc. Lett.* **53**, 10, 737–744. <https://doi.org/10.1080/00387010.2020.1832529>.
- Air Quality Guidelines for Europe, 2nd ed., WHO Regional Publications, *European Series*, No. **91**.
- Holloway T., Miller D., Anenberg S., Diao M., Duncan B., Fiore A.M., Henze D.K., Hess J., Kinney P.L., Liu Y., Neu J.L., O'Neill S.M., Odman M.T., Pierce R.B., Russell A.G., Tong D., West J.J., Zondlo M.A. (2021) Satellite monitoring for air quality and health, *Annu. Rev. Biomed. Data Sci.* **4**, 417–447. <https://doi.org/10.1146/annurev-biodatasci-110920-093120>.
- Martínez-Trinidad J.F., Carrasco-Ochoa J.A., Brants C., Hancock E.R. (eds), *Pattern Recognition, Third Mexican Conference, MCPR 2011, June 29–July 2, 2011*, Springer Proceedings, Cancun, Mexico.
- Griffiths P.R. (1977) Recent applications of Fourier transform infrared spectrometry in chemical and environmental analysis, *Appl. Spectrosc.* **31**, 6, 497–505.
- Yoon G., Kim Y.-J., Hahn S. (2003) Determination of glucose in whole blood samples by mid-infrared spectroscopy, *Appl. Opt.* **42**, 4.
- Yano T., Funatsu T., Suehara K.-I., Nakano Y. (2001) Measurement of the concentrations of glucose and citric acid in the aqueous solution of a blood anticoagulant using near infrared spectroscopy, *J. Near Infrared Spectrosc.* **9**, 43–48.
- Vonach R., Buschmann J., Falkowski R., Schindler R., Lendl B., Kellner R. (1999) Application of mid-infrared transmission spectrometry to the direct determination of glucose in whole blood, *Appl. Spectrosc.* **52**, 6, 820–822.
- Bhandare P., Mendelson Y., Peura R.A., Janantsch G., Kruse-Jarres J.D., Marbach R., Heise M.H. (1993) Multivariate determination of glucose in whole blood using partial least-squares and artificial neural networks based on mid-infrared spectroscopy, *Appl. Spectrosc.* **47**, 8, 1214–1221.
- Zeller H., Novak P., Landgraf P. (1989) Blood glucose measurement by infrared spectroscopy, *Int. J. Artif. Organs* **12**, 2, 129–135.
- Cirne I., Boaventura J., Guedes Y., Lucas E. (2016) Methods for determination of oil and grease contents in wastewater from the petroleum industry, *Chem. Chem. Technol.* **10**, 4, 437–444.
- Goldenstein C.S., Miller V.A., Mitchell Spearrin R., Strand C.L. (2017) SpectraPlot.com: integrated spectroscopic modelling of atomic and molecular gases, *J. Quant. Spectrosc. Radiat. Transf.* **200**, 249. <https://doi.org/10.1016/j.jqsrt.2017.06.007>.
- Haaland D., Easterling R., Vopicka D. (1985) Multivariate least-squares methods applied to the quantitative spectral analysis of multicomponent samples, *Appl. Spectrosc.* **39**, 73–84.
- Griffiths P.R., Bowie B.T. (2000) Measurement of the sensitivity and photometric accuracy of FT-IR spectrometers, *Appl. Spectrosc.* **54**, 8, 1192–1202.

- 21 Burns D.A., Ciurczak E.W. (2001) *Handbook of near-infrared analysis*, Marcel Dekker Inc., New York.
- 22 Oran Brigham E. (1988) *The fast Fourier transform and its applications*, Prentice Hall, Englewood Cliffs, New Jersey.
- 23 Griffiths P.R., de Haseth J.A. (2007) *Fourier transform infrared spectrometry*, 2nd ed., Wiley-Interscience.
- 24 Hirschfeld T. (1978) Dynamic range improvement in Fourier transform infrared spectrometry, *Anal. Chem.* **50**, 8, 1225–1226.
- 25 Antoon M.K., Koenig J.H., Koenig J.L. (1977) Least-squares curve-fitting of Fourier transform infrared spectra with applications to polymer systems, *Appl. Spectrosc.* **31**, 6, 518–524.
- 26 Martens H., Naes T. (1992) *Multivariate calibration*, John Wiley & Sons Inc., Chichester, pp. 180–202.
- 27 Arnold M.A., Geng L., Zhou X., Small G.W. (1997) Multivariate calibration models based on the direct analysis of near-infrared single-beam spectra, *Appl. Spectrosc.* **51**, 9, 1330–1339.

Appendix A

Theory of molecular detection

Principles of Fourier-transform infrared spectrometry

The key part of a Fourier transform infrared (FT-IR) spectrometer is the optical component termed an interferometer [19–22]. Commercial FT-IR spectrometers usually adopt the Michelson interferometer. The sample under test is placed at the exit of the interferometer. Infrared light coming from the source is split and traverses the sample under investigation and then is converted to an electrical signal with the aid of an infrared detector. The signal detected in relation to the relative position of one of the interferometer mirrors is called an interferogram. Then this interferogram is converted to sample absorbance versus wavenumber through applying the Fourier transform [23]. Theoretically one of the interferometer mirrors should move towards infinity but in practice, the movable mirror can only move for a certain distance. Therefore, the transformed interferogram has a finite resolution. The resolution or the difference between two neighboring frequencies ($\Delta\nu$) is inversely proportional to the maximum retardation caused by mirror movement δ_{\max} as $\Delta\nu = \frac{1}{\delta_{\max}}$ [24]. But

increasing the path difference caused by the mirror movement will increase the total noise of the interferogram. Doubling the length of the interferogram will typically increase the noise level by a factor of $\sqrt{2}$.

To avoid aliasing in sampling of any continuous signal, according to the Nyquist sampling theorem, the minimum sampling frequency should be twice the maximum frequency of this signal. A lower sampling rate than this limit will produce lower apparent frequency components which does not exist in the original signal. Hence, for a given resolution, $\Delta\nu$, and maximum frequency, ν_{\max} , the required number of points in the interferogram, N is

$$N = \frac{2\nu_{\max}}{\Delta\nu}.$$

Appendix B

Beer–Lambert law

The basic principles of least squares techniques are mentioned here to show the reason behind the lack of accuracy of these techniques when dealing with mixtures of some unknown constituents and to show the privilege of our novel technique over these least square techniques.

According to the Beer–Lambert law, the relationship between sample absorbance A , substance molar absorptivity ϵ , substance concentration c , and sample thickness t is given by [25–27];

$$A(\nu) = \epsilon(\nu)tc, \quad (\text{A-1})$$

where ν is the wavenumber.

Assuming that the absorbance of different components in a mixture at a given wavelength is linearly additive, the Beer–Lambert law relation is rewritten as;

$$A_i = \sum_{j=1}^l \epsilon_{ij}tc_j, \quad (\text{A-2})$$

where “ i ” refers to a wavenumber sampling point, “ j ” refers to a component, and the components in the mixture range from 1 to l . Then equation (A-2) can be rewritten as;

$$A_i = \sum_{j=1}^l k_{ij}c_j, \quad (\text{A-3})$$

where k_{ij} is the product of ϵ_{ij} and t .

The basis of absorption spectroscopy is to fit, with minimum error, the measured absorbance spectrum of a mixture to a linear summation of some spectral components expected in this mixture. The coefficients obtained from this fitting process express the concentrations of each of the known compounds’ spectra.

For a multi-wavelength and multi-component case, equation (A-2) can be written in a matrix form as;

$$A = KC + E. \quad (\text{A-4})$$

The matrix of residual spectra, E , called an error matrix, contains the part of the measured absorbance spectra in A of unexpected components in the studied mixture. Matrix E contains spectral substances that are not accounted for in matrix K such as absorption spectrum of substances that are not expected to be present in the mixture under examination or even any noise components. In fact, this E term, when some components in the mixture are not considered, is responsible for the reduced accuracy when it is required to search for the concentration of some components in a mixture. This causes the noise floor to increase and it will be shown that our new algorithm works better in these conditions, unlike the least-squares techniques.

The classical least squares CLS solution of equation (A-4) gives the concentrations of the components in the mixture as follows [25–27];

$$\hat{C} = \left(\hat{K}'\hat{K} \right)^{-1} \hat{K}'A, \quad (\text{A-5})$$

where \hat{C} is the estimated approximate concentration of the components in the mixture. In what follows, the algorithm of our proposed pattern recognition technique is proposed and its results are compared to those of the CLS approach.

Appendix C

Some data related to the studied gases in this work [8]

The level of carbon dioxide before the industrial age and the formation of the greenhouse effect was around 0.5 g/m³ (280 ppm).

Steps to change concentration “y” represented in μg/m³ to percentage in volume and ppm:

$$y \left[\frac{\mu\text{g}}{\text{m}^3} \right] = \frac{y}{\text{molar weight in grams}} \left[\frac{\mu \text{ mole}}{\text{m}^3} \right]$$

Since 1 mole of any gas occupies 24 dm³ (0.024 m³) at RTP (room temperature and pressure), then

Gas name	Molar weight (g)	Maximum recommended level in the atmosphere (μg/m ³)*
Carbon monoxide	28	60
Nitrogen dioxide	46	25
Sulfur dioxide	64	25

* Annual mean concentration.

$$\begin{aligned} y \left[\frac{\mu\text{g}}{\text{m}^3} \right] &= \frac{y}{\text{molar weight in grams}} 24(10^{-9}) \left[\frac{\text{m}^3}{\text{m}^3} \right] \\ &= \frac{y}{\text{molar weight in grams}} 24(10^{-7}) [\%] \\ &= \frac{y}{\text{molar weight in grams}} 24(10^{-3}) [\text{ppm}] \end{aligned}$$