Influence of temperature on synthetic data-based calibration models for low resolution open-path FTIR spectroscopy

S. CIĘSZCZYK*

Institute of Electronics and Information Technology, Lublin University of Technology, 38A Nadbystrzycka St., 20-618 Lublin, Poland

Abstract. This article presents the problem of determining the gas concentration under conditions of changing temperatures. The solution to this problem is based on spectral measurements using chemometric models. Such problems arise when measuring gases during various industrial processes, mainly the pyrolysis process. Chemometric models are built using synthetic calibration data. Line-by-line and statistical models were applied for spectra modelling. The effect of temperature on extracting the gas concentration from spectra of various resolutions was determined. Next, the global model was built and tested using calibration data for specific temperature ranges. The properties of the linear and nonlinear partial least-squares (PLS) models that were applied to the considered issues were demonstrated. CO and CH₄ were used as example gases.

Key words: FTIR, calibration model, temperature influence.

1. Introduction

Open-path Fourier transform infrared (OP-FTIR) spectroscopy is used to analyse atmospheric pollution [1–4], gases emitted from industrial processes [5], such as pyrolysis [6], and combustion [7, 8]. This type of measurement can be used to monitor various processes. CO₂ and H₂O are the main gases from typical combustion processes. The most commonly used spectral region for combustion diagnostics is the spectral range for carbon dioxide in the 4.3 μm region. This region is used for passive remote diagnostics [9] and for temperature determination [10]. Pyrolysis, in turn, is used in the analysis of organic materials (biomass) and for the production of high-energy gases, which are subsequently used for combustion. The main gases produced in the pyrolysis process are CO, CO₂ and CH₄. New area application of remote sensing is explosive detection [11] where promising results are obtained with OP-FTIR [12].

The most difficult part of extracting information from a measured spectrum is building a suitable calibration model. Typically, the calibration model is built using calibration data, which are generated from empirical processes in the laboratory under strictly controlled conditions with respect to the temperature and pressure, as well as the configuration of the spectrometer. The most commonly used models for gas content determination in OP-FTIR gas measurements are classical least-squares (CLS) and partial least-squares (PLS) regression [13]; a less frequently used method is principal component regression (PCR) [14]. Calibration is a laborious and often expensive process. If the instrumental (e.g., resolution) or environmental (e.g., temperature, pressure) conditions change during the measurement, the measurement must be repeated. In cases of indirect inference in metrology, modelling can be used instead of an expensive and difficult empirical process [15]. Data used to creating the inverse model can be generated with applying a forward model. In addition, the OP-FTIR spectroscopy models of gas spectra can be simulated [16, 17]. The model, which converts physical parameters to a measurable quantity, can be used as a direct model in an iterative comparison with measured data [18–20]. Due to the high computational complexity of gas spectra modelling methods, special acceleration techniques have been proposed [21]. Unfortunately, when comparing measured data to the synthetic data from the direct model by the least-squares methods, there are several input variables, which may lead to significant errors if not taken into account [22]. These errors may result from errors in the gas parameter database or from instrument misalignment. An experimenter does not provide a control of gas database, the parameters are validated and corrected by the measurements of different authors. Instrumental errors (ILS, instrumental line shape) are more important. An appropriate parameterisation of the data (spectra) can reduce ILS. One approach is the measurement and comparison of the integrated absorption of particular rotational lines [23, 24] or multiple lines (band, spectral range) [25–27]. For low resolution spectra, in which particular rotational lines of the gas are not visible, the ILS error is smaller than that of higher resolution measurements. Finally, very low resolution (25 cm⁻¹ average band absorption) statistical gas spectra models can be used for parameterisation [28].

In combustion processes and in biomass pyrolysis, significant changes in temperature occur in addition to the variability in gas concentration. In such cases, the least-squares approach is the most common, and this approach provides a solution to the so-called inverse problem [29, 30]. This problem arises from the difficulty in formulating a chemometric model that can determine the content of the ingredients based on the absorbance and its temperature-dependence. One of the most

*e-mail: s.cieszczyk@pollub.pl
frequently used methods that addresses this problem is the use of one selected mean temperature, which can cause large errors if the temperature varies from this value significantly [31]. It is known that the temperature of the measured gases can be employed to correct the extracted gas content [32] or can eventually be used as an additional variable in the inversion process [33, 34]. In resistive sensors [35,36], temperature value or its additional modulation [37] can result in increase of their sensitivity.

This article aims to investigate the possibility of building a calibration model that is based on gas spectra modelling and has low temperature cross-sensitivity. A low resolution measurement (0.5; 1; 2; 4 cm⁻¹) and a statistical narrow band (SNB) model are investigated.

2. Line-by-line modelling

The most popular method for modelling gas spectra uses HITRAN [38] database. This database contains individual parameters for calculating spectral lines, along with the parameters for the effect of various factors, such as temperature, pressure and the content of other components in the gas mixture, on the shape of the lines. The transmittance of a gas can be calculated as follows [39]:

\[
\tau = \exp \left[ -S(T) \cdot g(v - v_0) \cdot \frac{P_i}{k \cdot T} \cdot l \right],
\]

where \( S(T) \) – intensity of a spectral line, \( g(v - v_0) \) – normalised line shape function, \( v_0 \) – spectral line position, \( P_i \) – gas partial pressure, \( k \) – Boltzmann constant, \( l \) – length of the studied path.

The strength of the line changes with temperature [39]:

\[
S(T) = S(T_0) \cdot \frac{\exp \left( \frac{-h c E^\prime}{k T} - \frac{h c v_0}{k T} \right)}{Q(T)} \cdot \frac{1 - e^{-\frac{h c v_0}{k T}}}{1 - e^{-\frac{h c E^\prime}{k T}}},
\]

where \( c \) – speed of the light, \( h \) – Planck constant, \( T_0 = 293 \) K, \( E^\prime \) – the lower state energy, \( v_0 \) – spectral line position, and \( Q \) – the total internal partition sum.

Two main phenomena broaden spectral lines:

- collisional broadening, which causes the Lorentzian shape of spectral lines,
- Doppler broadening, which causes the Gaussian shape of spectral lines.

Lorentzian line shape:

\[
g_L(v - v_0) = \frac{1}{\pi} \frac{\gamma_L}{(v - v_0)^2}.
\]

Lorentzian half-width at half maximum (height):

\[
\gamma_L(v) = g_a \cdot \left( \frac{206}{T} \right)^n \cdot P,
\]

where \( g_a \) – air-broadened half-width, \( n \) – temperature coefficient, and \( P \) – total pressure.

Gaussian line shape caused by Doppler broadening:

\[
g_G(v - v_0) = \frac{1}{\gamma_G \sqrt{\pi}} \exp \left( -\frac{(v - v_0)^2}{\gamma_G^2} \right).
\]

3. SNB model

Statistical gas spectra models were developed to decrease the computational effort required to calculate radiative transfer. Such models are mainly used for calculating the average transmissivity of 25 cm⁻¹ bands (Fig. 2). Depending on the method for averaging the lines in the band, various types of models are formed. SNB model parameters are updated with the latest spectroscopic databases and empirical spectroscopic measurements [40, 41]. The average SNB band transmissivity of a gas is expressed as follows [42]:

\[
\tau = \exp \left[ -2 \frac{\gamma}{\delta} \left( \sqrt{1 + xplk \frac{\delta}{\gamma}} - 1 \right) \right],
\]

where \( l \) – length of path, \( p \) – total pressure, \( x \) – gas molar fraction, \( k \) [cm⁻¹ atm⁻¹] – mean line intensity to the spacing ratio, \( \delta \) [cm] – mean line spacing, and \( \gamma \) [cm⁻¹] – typical average line half-width.
The last parameter changes with temperature [42]:

\[
\gamma_{CO} = \frac{p}{p_s} \left[ 0.075 \cdot \frac{X_{CO2}}{T_s} + 0.12 \cdot \frac{X_{H2O}}{T_s} + 0.06 \cdot \left( \frac{T_s}{T} \right)^{0.7} \cdot (1 - X_{CO2} - X_{H2O}) \right],
\]

\[
\gamma_{CO2} = \frac{p}{p_s} \left[ 0.07 \cdot \frac{X_{CO2}}{T_s} + 0.1 \cdot \frac{X_{H2O}}{T_s} + 0.058 \cdot \left( 1 - X_{CO2} - X_{H2O} \right) \right],
\]

\[
\gamma_{H2O} = \frac{p}{p_s} \left[ 0.462 \cdot \frac{X_{CO2}}{T_s} + 0.079 \cdot \left( 1 - X_{CO2} - X_{O2} \right) + 0.106X_{CO2} + 0.036X_{H2O} \right],
\]

where \( p_s = 1 \) atm, and \( T_s = 296 \) K.

4. Gas content determination based on a measured spectrum

Methods for determining the constituents using OP-FTIR measurements can be divided into two groups: the chemometrics method and the method of minimisation of the difference between the measured spectrum and the spectrum from the direct model (iterative methods). Chemometrics methods are preferred in the spectroscopy community. Moreover, iterative methods often require the assistance of an expert. Two methods, PLS and RBF-PLS (Radial Basis Functions PLS) [43, 44], were used to build the chemometric model. Modelling was performed for methane and carbon monoxide. The simulation only takes into account the absorption of gases and assumes the absence of multiple scattering in the particle phase. These assumptions can cause additional measurement errors in the absorption coefficients, particularly for water vapour [45, 46].

4.1. Methane simulation. To test the effect of temperature on the determination of the methane content using the chemometrics method, an appropriate data set consisting of 441 points was produced.

The model built for creating spectra at 303 K was tested at 5 temperatures: 283 K, 293 K, 303 K, 313 K and 323 K. As shown in Figs. 3–6, errors for temperatures that differ from that of the calibration set are evident. Additionally, for PLS, the error increases with decreasing resolution. As a result, the error is nearly two times larger for the 4 cm\(^{-1}\) resolution than for the 0.5 cm\(^{-1}\) resolution. For RBF-PLS, the resolution did not change the temperature-dependent errors, and the temperature errors are comparable to those for PLS and the 0.5 cm\(^{-1}\) resolution.
To eliminate the effects of temperature on the determination of the methane content using chemometric models, an appropriate calibration set can be used. The set must include all the analysed temperatures (Fig. 7). In this case, only the 2 cm\(^{-1}\) and 4 cm\(^{-1}\) resolutions were considered (Figs. 8, 9). Better results were obtained for the nonlinear RBF-PLS. The temperature has a larger influence at the lower concentration of CH\(_4\); this influence can be observed as the wider spread in the points for a particular concentration. Each point for every concentration represents a different temperature (compare Fig. 7).
Influence of temperature on synthetic data-based calibration models for low resolution open-path FTIR spectroscopy

4.2. CO SNB model. The next stage of the research was to use the SNB model parameterisation and build an appropriate calibration model. In SNB, the spectra lines are integrated in $25 \text{ cm}^{-1}$ bands, resulting in 12 bands between 2000 and 2250 cm$^{-1}$ for CO. As a preliminary point, a model for the 300 K temperature was built and tested for tem-
temperatures from 270 to 330 K (Fig. 10). The PLS method yielded errors that were too large, even for the calibration data. For RBF-PLS, the temperature effects were similar to those found earlier for methane and can be approximated as 0.5%/K.

The next step was to generate a set of data (Fig. 11). It turned out, however, that a single model built from the calibration data for a temperature range of 270–490 K showed significant errors, particularly for small concentration ranges (Fig. 12).

Fig. 10. Error caused by temperature effects in the RBF-PLS model with SNB parameterisation, 25 cm$^{-1}$ resolution of CO

Fig. 11. CO data set, temperature span of 270–490 K

Fig. 12. Temperature errors for global calibration models: PLS (left) and RBF-PLS (right) with SNB spectra parameterisation at 25 cm$^{-1}$ for CO in the 270–490 K temperature range
Influence of temperature on synthetic data-based calibration models for low resolution open-path FTIR spectroscopy

Fig. 13. CO data set for the 400–500 K temperature range

Fig. 14. Temperature errors for global calibration models: PLS (left) and RBF-PLS (right), with SNB spectra parameterisation at 25 cm$^{-1}$, for CO, in the 400–500 K temperature range

Fig. 15. Temperature errors for global calibration models: PLS (left) and RBF-PLS (right), at 4 cm$^{-1}$ resolution for CO, for the 270–470 K temperature range
Fig. 16. Temperature errors for global calibration models: PLS (left) and RBF-PLS (right), 4 cm⁻¹ resolution for CO, for the 370–470 K temperature range.

A single model in the range of 270–490 K cannot be used for data based on such a small resolution. Thus, a set spanning 100 K from 400–500 K was created (Fig. 13). Unfortunately, the errors resulting from the temperature variation for the PLS model can reach up to 10%, and these errors are particularly significant for smaller concentrations (Fig. 14). For the RBF-PLS model, the error is less than 2%, and in the 500–2500 ppm·m range in particular, the error is smaller than 0.5%.

4.3. CO simulation and comparison with measured spectra. Then, in a similar manner, CO models at 4 cm⁻¹ resolution were created. In this case, the errors for the 270–470 K (Fig. 15) temperature range were two times larger than those for the narrower 370–470 K range model (Fig. 16). The PLS models are noticeably better than the RBF-PLS models.

Fig. 17. Comparison of the measured [49] and calculated spectra for concentration retrieved from the PLS 370–470 K model.

Literature reports about practical measurements in industrial environments where the temperature changes are rare. The same is true about the ability to produce these conditions in the laboratory. Such measurements require special equipment [47] and can be used to confirm the accuracy of synthetic spectra modelling [48]. The CO concentration was determined using measured spectrum [49] for the PLS 370–470 K model. The spectra are compared in Fig. 17. The given concentration values is 19200 ppm·m for the 180°C temperature. The calculated concentration is 18600 ppm·m with 3.1% error.

5. Conclusions

Most of the work to determine gas concentrations under varying conditions attempts to build a direct model and use iterative methods. Inverse models for such cases are created for typical average conditions. The maximum error, which may arise from changes in temperature, is only estimated in some cases. In this article, simulations at small resolutions between 0.5 and 4 cm⁻¹ and 25 cm⁻¹ resolution for SNB were performed. If the model is built for the specific conditions, the typical error due to temperature effects is 5% at 10 K. The resolution has a greater effect on the temperature influence for the PLS models than for the RBF-PLS models. Temperature has a greater effect on smaller gas concentrations. Therefore, it appears advisable to build separate models for small and large concentration ranges. A universal, temperature-robust model can be built for a specific temperature range. If chosen incorrectly, the model will not be able to reduce the effects of temperature below a certain error level.

REFERENCES

[3] M. Kastek, T. Piątkowski, and P. Trzaskawka, “Infrared imaging Fourier transform spectromter as the stand-off gas detection...
Influence of temperature on synthetic data-based calibration models for low resolution open-path FTIR spectroscopy


