Quantification of Carbon Dioxide (CO₂), Methane (CH₄), and Nitrous Oxide (N₂O) Using Near Infrared Spectroscopy and Multivariate Calibration in High Humidity Levels

Rafael L. Ribessi, Wilson F. Jardim, Jarbas J. R. Rohwedder and Thiago A. Neves

Departamento de Química Analítica, Instituto de Química, Universidade Estadual de Campinas, 13083-970 Campinas-SP, Brazil

Departamento de Engenharia Sanitária e Ambiental, Escola de Engenharia, Universidade Federal de Minas Gerais, 31270-010 Belo Horizonte-MG, Brazil

In this work we developed a promising analytical method combining Fourier transform near-infrared (FT-NIR) spectroscopic technique and first-order multivariate calibration using partial least-squares (PLS) model to simultaneously quantify the main greenhouse gases (GHG’s): methane (CH₄), carbon dioxide (CO₂), nitrous oxide (N₂O) and water vapor (H₂O). The models were built using 70 mixtures with different concentration of these gases, 0.25-32.0 ppm to CH₄ and N₂O, and 50-1100 ppm to CO₂ and different values of relative humidity (52-85%, 20 ºC) in synthetic air. After preparing each of the mixtures, they were analyzed by using FT-NIR and a reference analytical technique based on gas chromatography with mass spectrometric detection (GC-MS). The FT-NIR spectrometer was coupled with a long optical path cell, with 105.6 meters of optical path. In sequence, the spectra of all mixtures and its concentration values for each gas were used to build the multivariate calibration models, using PLS regressions. For this, the mixtures were grouped with Kennard Stone algorithm, 50 samples to calibration set and 20 samples to prediction set. The values of RMSEP (root mean square error of prediction) obtained for each model are 0.66, 28.7 and 0.66 ppm, respectively, for CH₄, CO₂, and N₂O. The limits of quantification (LOQ) for each PLS models are 0.26, 3.6, and 0.99 ppm, respectively, for CH₄, CO₂, and N₂O. The results show the potentiality of application of this system to monitoring emission sources in which the concentration of these gases are relatively high, as urban centers, industrial areas, and landfills.

Keywords: greenhouse gases, chemometrics, NIR spectroscopy, high humidity levels, PLS

Introduction

The continuous increase of greenhouse gases (GHG’s) emission in atmosphere is directly related to the global warming effect and the climate change.¹ The carbon dioxide (CO₂), methane (CH₄) and nitrous oxide (N₂O) have major influence on the greenhouse effect intensification, in conjunction with water vapor, since they are present in higher atmospheric concentrations and their averages are around 400, 1.85 and 0.33 ppm, respectively, in the atmosphere.¹ There is a global concern on monitoring sources and sinks of these gases, due to their influence on climate, for example, in agricultural areas, sewage treatment plants, hydroelectric reservoirs, landfill systems, urban and industrial areas.²⁻⁵

The monitoring of emission sources of GHG’s is important to better understand biogeochemical processes in which they are involved, and in the evaluation of recovery and sustainable energy processes. For this reason, several approaches are described in the literature, which employs different analytical techniques and types of sampling procedures, as well as, distinct ways of performing in situ analysis of GHG’s.⁶⁻⁹ Among the most used analytical techniques are gas chromatography (GC) and non-dispersive infrared (NDIR), due to their high accuracy and precision in gas sensing.¹⁰ Probably, GC is the technique most used in the monitoring (detect and quantify) of GHGs, such as CH₄, CO₂, N₂O, sulfur hexafluoride (SF₆), and hydrofluorocarbons (HFC’s).¹¹,¹² However, each of these gases needs a particular instrumental configuration to be quantified, from distinct stationary phases to individual detection systems, and it is necessary to perform a pretreatment in the samples to remove the water vapor before analyses.¹¹
NDIR is the most common spectroscopy technique used to detect and quantify the CO₂ present in the atmosphere. Measurements of GHG’s flux as used by techniques such as Eddy covariance, makes use of infrared spectroscopy to monitoring the variation of CO₂, H₂O, and CH₄ in different environmental matrices. The sensitivity of this technique may be increased to perform measurements of trace gases in the atmosphere by using optical long path cells, such as White’s cell, Hanst’s cell, and Herriott’s cell. The optical design of these systems allows a variable path ranging from centimeters to hundreds of meters and with long path configuration it is possible to quantify many gases in the concentration range of ppb.

Different spectroscopy techniques were also used to develop and perform measurement methods of GHG’s in different environmental matrices. For example, Christiansen et al., using cavity ring down spectroscopy quantified CH₄, CO₂, and N₂O exchanges between soil and atmosphere interface; tunable diode laser absorption spectroscopy (TDLAS) was used by Pataki et al. and Nadezhdinskii et al. to monitor CH₄ and CO₂ in an urban area, and quantum cascade laser spectroscopy (QCL) was used by da Silva et al. and Cui et al. for monitoring CH₄ and N₂O.

Near infrared spectroscopy (NIRS) is also used as an analytical technique for the quantification of gaseous species, although at high concentration levels, as in the case of natural gas. However, the determination of gases in ppm concentration range can be obtained using long optical path cells coupled with NIRS. NIRS is a type of vibrational spectroscopy that uses electromagnetic radiation in the range of 750-2,500 nm (4,000-13,300 cm⁻¹), in which the absorption of radiation is based on overtones and combination bands of vibrational modes in the molecules. At the NIRS region, it is possible to observe and discriminate fingerprints of many different compounds, due to NIRS relatively high signal to noise ratios. However, in the NIRS spectral range there are a large spectral overlap due to the absorption of different compounds. In this study, the high molar absorptivity of H₂O vapor and its great absorption region hinders GHG’s analyses in samples with high humidity levels. To overcome this, multivariate calibration model, as partial least square (PLS) regression, may be used to evaluate GHG’s in these types of samples, without the need of pre-processing steps or consumables to remove the humidity, which has been already employed in the mid-infrared (MIR) spectral region.

In this study, it was developed a multivariate calibration method using a Fourier-transform spectrophotometer (FT-NIR) coupled with a long optical path cell (Hanst’s, 105.6 m) for the determination of CH₄, CO₂ and N₂O in air samples with high humidity levels (52-85%). The results obtained from FT-NIR experiments were compared with those obtained by a reference method based on gas chromatography mass spectrometry (GC-MS).

**Experimental**

**Experimental procedure**

The multivariate calibration models were built from gas mixtures, which were prepared directly in the long optical path cell from Infrared Analysis, model 107-V (Anaheim, USA), which was coupled to a FT-NIR spectrophotometer from ABB Bomem, model MD 160D (Zürich, Switzerland). A scheme with the steps for preparing all the 70 mixtures is shown in Figure 1. In the first step, the optical cell was cleaned by purging synthetic air (99.99%, Air Liquide, São Paulo, Brazil) during 5 min with a flow rate of 10 L min⁻¹ and after this step, a background for FT-NIR was obtained, Figure 1A. In the next step, the humidity level was adjusted in the optical cell, with synthetic air (99.99%, Air Liquide, São Paulo, Brazil) by passing it through a bubble humidifier bottle before entering the cell and this purging step was interrupted when it was achieved the level of relative humidity desired, Figure 1B. A hygrometer by Cole Parmer (Illinois, USA) was used to control the relative humidity levels, which was previously calibrated. The relative humidity in the mixtures was between 52 and 86% at 20 °C.

![Figure 1](image-url)
In the next step, CH$_4$ (99.95%), CO$_2$ (99.95%) and N$_2$O (99.95%) supplied by White Martins (Rio de Janeiro, Brazil), previously diluted in synthetic air, were added to the cell. Different volumes from the gas mixtures of 4000 ppm (v/v), prepared in a Tedlar sampling bag of 1 L from SKC Inc. (Pennsylvania, USA) were injected into the optical cell using gastight syringes (Hamilton Company, Nevada, USA). After the gases injection, the optical cell was homogenized using a diaphragm pump (Gast Manufacturing, Michigan, USA) for 3 min, Figure 1C. The mixtures were prepared in the concentration range of 0.25-32 ppm to CH$_4$ and to N$_2$O, 50-1100 ppm to CO$_2$. The preparation of the mixtures was based on a mixture design, however, due to experimental difficulty to adjust the relative humidity levels, more points were added, ensuring that there was no correlation between the components. Finally, in the last step (Figure 1D), the spectra for each of the mixtures were obtained in the FT-NIR and 300 µL were collected from the optical cell and injected into the GC-MS model 17A/QP505A (Shimadzu Corporation, Kyoto, Japan) to obtain the reference values for each gas in the mixtures.

GC-MS method

The GC-MS method used a capillary column TG-BOND Q (Thermo Scientific, Massachusetts, USA), with the following dimensions: length of 30 m, internal diameter of 0.32 mm and film thickness of 10 micrometers. The carrier gas was helium at a flow rate of 2 mL min$^{-1}$ and the injector temperature was held at 100 °C. The samples were injected manually with gastight syringes. The injection volume was 300 µL per sample in splitless mode. An isotherm of 25 °C for 3 min was used for GHG's separation in the analytical column. The single ion monitoring (SIM) was used to improve sensitivity in the quantification and development of analytical curves for each GHG. The mass to charge ratios used for each gas were: CH$_4$ (16, 15, 14 m/z), CO$_2$ (44, 28, 16 m/z) and N$_2$O (44, 30, 14 m/z).

The analytical curves were constructed by diluting the pure gases on synthetic air (99.99%, Air Liquide, São Paulo, Brazil), and five concentration levels for each gas was prepared in triplicate. The concentration levels of CH$_4$ and N$_2$O were 0.1, 2.0, 10.0, 20.0 and, 40.0 ppm and to CO$_2$ were 50, 100, 400, 900 and 1200 ppm. In order to validate the analytical curves three certified reference standards (gravimetric preparation) were used with concentrations for CH$_4$, CO$_2$ and N$_2$O in synthetic air balance (Air Liquide, São Paulo, Brazil). For the calculation of the limit of detection (LOD) and the limit of quantification (LOQ) of each GHG it was used the following equations: LOD = 3 sd/b; LOQ = 10 sd/b, where sd is the standard deviation of the blank chromatogram (synthetic air) in the linear regression equation, and b is the slope from the calibration curve. The results obtained for the chromatographic method are presented in the Supplementary Information (SI) section.

Multivariate method

The spectra were obtained using a FT-NIR model MD 160D (ABB Bomem, Zürich, Switzerland) coupled with a long optical path cell model 107-V (Infrared Analysis, Anaheim, USA) with 105.6 m of optical path and an internal volume of 16 L. All spectra were obtained with 4 cm$^{-1}$ of spectral resolution, average of 100 scans in the range of 4000 to 6600 cm$^{-1}$. The background spectrum was obtained previously to the mixtures preparations, by purging the optical cell with synthetic air (99.99%, Air Liquide, São Paulo, Brazil).

Thus, it was applied in the obtained spectra the first derivative, by Savitzky-Golay algorithm (3 points window and 2º order polynomial) and mean-centered. The PLS models were developed using the NIPALS algorithm and leave-one-out cross validation. The samples were grouped in the calibration set and prediction set, respectively, 50 and 20 samples, using the Kennard-Stone algorithm. The software to perform the multivariate calibration was The Unscrambler X 10.4 (AspenTech, Massachusetts, USA). Limits of detection and quantification (LOD and LOQ) for the multivariate models were calculated via equations 1 and 2, respectively. As well as sensitivity (SEN) and analytical sensitivity (SEN anal.) (for the multivariate models were calculated via equations 3 and 4, respectively.

\[
\text{LOD} = 3 \sigma_B ||b|| \quad (1)
\]
\[
\text{LOQ} = 10 \sigma_B ||b|| \quad (2)
\]
\[
\text{SEN} = \frac{1}{||b||} \quad (3)
\]
\[
\text{SEN}_{\text{anal.}} = \frac{\text{SEN}}{\sigma_b} \quad (4)
\]

where $\sigma_b$ is the value of the standard deviation of the background, $\|b\|$ is the norm of the regression vector of the model.

Results and Discussion

Figure 2 shows the spectra for CO$_2$, CH$_4$ and N$_2$O gases and water vapor, in concentrations of 500, 10, 10 ppm and
80% relative humidity for a temperature of 20 °C. It is observed that there is a significant difference in the intensity of the absorbances of these species, which is reflected by the difference in the concentration levels at which these compounds are found in the atmosphere, which for the case of CH₄ and N₂O are at even lower levels of concentration. This demonstrates the difficulty in quantifying these three gases in the presence of high levels of humidity, using this spectral region. For this reason, the accuracy of univariate methods based on FT-NIR to quantify GHG’s are severely hampered due to the interference of water absorption, showing the importance of chemometric tools to extract analytical information in complex matrices, such as the atmosphere. The Figure S1 (SI section) shows the spectra of these gases obtained from HITRAN database.⁴⁰

The absorption bands that are presented in Figure 2 for CH₄, occurs in the spectral region between 4050 to 4750 cm⁻¹, with greater intensity, and at the region of 6000 cm⁻¹, which correspond, respectively, to the combination bands and first overtone regions.⁴⁰⁻⁴² It is worth noting, that the NIRS spectral region is well known for presenting a response to the absorptions of the C–H and C–C bonds, in this range of the spectrum there is also the absorption of NIRS radiation by the C=O bonds of CO₂ and N=O of N₂O. For N₂O there are two bands in the regions of 4450 and 4750 cm⁻¹, the first corresponds to the first overtone of the asymmetric stretch, and the second the combination between the asymmetric stretch and the first overtone of the symmetrical stretch of the connections present in the N₂O.⁴⁰⁻⁴³ As for CO₂, there are three bands in the region from 4750 to 5250 cm⁻¹ corresponding to the first overtone of the asymmetric stretch, and the combinations bands between the symmetrical and asymmetric stretches and the angular deformation of the O= C=O bonds.⁴⁰⁻⁴⁴ Finally, for H₂O there are two bands presenting intense absorbance between 4500-4000 cm⁻¹ and 5560-5000 cm⁻¹ due to the fundamental and the second overtone of O–H stretch of water, respectively.⁴⁰⁻⁴⁵

The spectra obtained for all the 70 mixtures are shown in Figure 3. The raw spectra, Figure 3A, show a typical baseline variation present in single beam instrument which is caused by variations in the radiation source and the detector’s temperature. Thus, first derivative was applied to remove the baseline fluctuation (Figure 3B).

After the baseline correction, the spectra were mean centered and the PLS regression was applied to the 70 samples. The presence of anomalous samples was verified for each gas, based on leverage and q-residual values after the PLS analysis.³⁰ The next step was applying the Kennard-Stone algorithm to group the samples in calibration and validation sets, containing 50 and 20 samples, respectively. This algorithm was used to ensure homogeneity of distribution of samples between sets.³⁴ The leave-one-out cross-validation was used to determine the number of latent variables (LV) used to build the models. This way, it was chosen the number of LV that presented the smallest root mean square error for cross validation (RMSECV) of each gas.⁴⁶

The results obtained with the PLS models are shown in Table 1. The RMSEP values obtained for the models were 0.66, 0.66 and 28.7 ppm for the CH₄, N₂O and CO₂, and 0.9% for the relative humidity (R.H.); and the values of determination coefficient (R²) were greater than 0.986.

**Figure 2.** (a) Spectra for CO₂, CH₄ and N₂O gases and water vapor in the same scale, and (b) zoom for highlight the CO₂, CH₄ and N₂O spectra in the range 4950-4225 cm⁻¹.
These results demonstrate a great correlation between the reference values obtained with the GC-MS and those obtained with the multivariate calibration model for the FT-NIR. The limit of detection and quantification (LOD and LOQ) are close to those obtained for the GC-MS method (Table S2, SI section). Comparing the LOD and LOQ values for CH$_4$ in the literature, the values obtained in this work are smaller than those found for the middle infrared region (MIR), due to the optical path used, which here was 105.6 meters.$^{32}$

Figure 4 shows the collection of analytical curves correlating the PLS models and the results obtained by the reference methods, in addition to the regression coefficients for each of the PLS models. In the Figures 4a, 4c, 4e and 4g, it is possible to observe a graphical representation for the correlation between the results obtained from the PLS models and those obtained from the reference methods. In the Figures 4b, 4d, 4f and 4h, as could be expected, the regression coefficients have higher values in the same range of the gases absorption spectra, shown in Figures 2, S1 and S2 (SI section). The fact of regression coefficients is quite similar to each GHG shows the model capability to extract the variance of each gas present in the data set, besides that, it is possible to see a minor intensity in the region of 5500-5000 cm$^{-1}$ for all gases, which can be associated to the residual humidity in the pure gases and also in the synthetic air used to prepare the mixtures.

The results presented by the models demonstrate their applicability for the determination and monitoring of different emission sources for these three gases. With adjustments to the experimental setup, as the optical path of the cell, the method presented here can be used to monitor the emission of methane by ruminants,$^{12,32}$ the emission of carbon dioxide in urban areas with large circulation of vehicles,$^{47}$ and monitor the emission of nitrous oxide in rice plantations.$^{48}$

**Conclusions**

A chemometric method based on Fourier-transform near infrared spectroscopy for measurements of carbon dioxide, methane, nitrous oxide, and water vapor was described. The method developed has shown to be a reliable tool in the

**Table 1.** Figures of merit for the FT-NIR multivariate method prepared in synthetic air with humidity ranging from 50 to 90%

<table>
<thead>
<tr>
<th></th>
<th>LV</th>
<th>LOD</th>
<th>LOQ</th>
<th>RMSEC</th>
<th>R$^2$</th>
<th>RMSEP</th>
<th>R$^2$</th>
<th>SEN</th>
<th>SEN$_{anal}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_4$/ppm</td>
<td>4</td>
<td>0.08</td>
<td>0.26</td>
<td>0.56</td>
<td>0.997</td>
<td>0.66</td>
<td>0.997</td>
<td>8.15 × 10$^{-4}$</td>
<td>39.93</td>
</tr>
<tr>
<td>N$_2$O/ppm</td>
<td>8</td>
<td>0.35</td>
<td>0.99</td>
<td>0.77</td>
<td>0.995</td>
<td>0.66</td>
<td>0.997</td>
<td>1.73 × 10$^{-4}$</td>
<td>8.47</td>
</tr>
<tr>
<td>CO$_2$/ppm</td>
<td>3</td>
<td>1.1</td>
<td>3.6</td>
<td>32.6</td>
<td>0.992</td>
<td>28.7</td>
<td>0.995</td>
<td>5.70 × 10$^{-4}$</td>
<td>2.79</td>
</tr>
<tr>
<td>R.H./%</td>
<td>3</td>
<td>0.01</td>
<td>0.03</td>
<td>1.3</td>
<td>0.986</td>
<td>0.9</td>
<td>0.996</td>
<td>6.05 × 10$^{-4}$</td>
<td>296.37</td>
</tr>
</tbody>
</table>

R.H.: relative humidity; LV: latent variable; LOD: limits of detection; LOQ: limits of quantification; RMSE (C and P): roots means square error of (calibration and prediction); R$^2$: determination coefficient; SEN: sensitivity; SEN$_{anal}$: analytical sensitivity.
quantification of GHG’s in concentrations from different sources of emission. LOQ values found for the chemometric method, using GC-MS as a reference, were 0.26, 0.99 and 3.6 ppm for CH₄, N₂O, and CO₂, respectively, and 1.3% for relative humidity. Despite the spectral region of the NIRS that is less sensitive than the spectral region of the MIR, the choice for using this region occurred due to the popularization of this spectral technology has experienced in recent years with the appearance of several portable devices appearing on the market. Such application has the potential to be used with the method described in this work, making it less costly to be used in the monitoring of different emission sources. Which is essential to government and civil society to take the best decisions on energy policy and on mitigating climate change.

The models demonstrated here may be used as a basis (backbone) for new models to monitoring different sources of GHG’s. The samples collected in these sources must be added to the models, ensuring that the variability of the local matrix is present in the model, without the need to collect hundreds of samples to build a new model for each monitoring point. Moreover, the models could also be improved and adapted to be used for monitoring different types of gaseous samples. For example, other gases could be added to this calibration model, as long as they have spectra in the NIR region, such as CO, O₃ and volatile organic compounds (VOC’s). Moreover, the variability presented in the multivariate model can be improved expanding the concentration range for these gases using a shorter optical path or being possible to decrease the values of LOD and LOQ by using a cell with a larger optical path.

Supplementary Information

Supplementary information is available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

This work was supported by CNPq (proc: 465768/2014-8), FAPESP (proc: 2014/50951-4), and INCTAA. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 (R.L.R. thanks CAPES for the fellowships, process 88881.154403/2017-01 Capes-WBI, and process 88887.195518/2018-00 INCT).
References


35. The Uncrambler X, 10.4; AspenTech, Massachusetts, USA, 2011.


Submitted: August 9, 2021
Published online: November 17, 2021

This is an open-access article distributed under the terms of the Creative Commons Attribution License.