

energy&fuels

Subscriber access provided by Heriot-Watt | University Library

Article

FTNIR spectroscopy and PLS algorithm for monitoring compositional changes in hydrocarbon gases under in-situ pressure

Reza Khodaparast Haghi, Jinhai Yang, and Bahman Tohidi

Energy Fuels, **Just Accepted Manuscript** • DOI: 10.1021/acs.energyfuels.7b01677 • Publication Date (Web): 02 Aug 2017Downloaded from <http://pubs.acs.org> on August 4, 2017

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.



ACS Publications

Energy & Fuels is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

1 FTNIR spectroscopy and PLS algorithm for
2 monitoring compositional changes in hydrocarbon
3 gases under in-situ pressure

4 *Reza K. Haghi, Jinhai Yang*, Bahman Tohidi*

5 *Centre for Gas Hydrate Research, Institute of Petroleum Engineering, School of Energy,*
6 *Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh EH14 4AS,*
7 *United Kingdom.*

8 * *Corresponding Author: petjy@hw.ac.uk*

9 **Keywords:** FTNIR spectroscopy; PLS; Hydrocarbons; Natural gas; High pressures.

10 **ABSTRACT**

11 The ability of Fourier Transform Near-Infrared (FTNIR) spectroscopy and chemometric
12 method was investigated to determine the concentration of major hydrocarbon components of
13 natural gases at pressures from 3.44 to 13.78 MPa and temperatures from 278.15 to 313.15 K.
14 Various PLS models were developed to determine the concentration of methane, ethane,
15 propane, i-butane, and n-butane simultaneously in gas phase at different pressures and
16 temperatures using the acquired FTNIR spectra. Several pre-processing techniques were
17 tested prior to the construction of the calibration models. The first Savitzky-Golay derivative
18 with smoothing over 5 points plus orthogonal signal correction (OSC) was found to be the
19 best method for FTNIR data pre-processing. Good agreement was obtained between the
20 predicted data by PLS models and the measured values with a standard error of prediction
21 (SEP) of 0.184 – 0.217, 0.165 – 0.209, 0.136 – 0.181, 0.098 – 0.154 and 0.096 – 0.142
22 cmol/mol for methane, ethane, propane, i-butane and n-butane respectively at different

1 temperature and pressure conditions. The developed PLS models were evaluated for a real
2 natural gas, and a good agreement between the PLS model prediction and the GC analysis
3 was gained at different pressures. Finally, the sensitivity of the FTNIR spectroscopy
4 technique to the system pressure and temperature was investigated. It was verified that
5 changes in pressure and temperature within a certain range affect the accuracy of the PLS
6 models. The results suggest that FTNIR spectroscopy in association with chemometric
7 method based on the PLS algorithm is a viable approach for monitoring changes in the
8 concentration of major components in gas phase.

9 **1. INTRODUCTION**

10 Nowadays, natural gas is one of the most efficient and popular sources of energy in the
11 world, and it plays an essential role in the manufacturing industry and transportation,
12 commercial and residential sectors.¹ Methane is the main component of natural gases and is
13 typically between 87 to 97 cmol/mol, and ethane, propane, butanes and pentanes are others
14 main hydrocarbon components in natural gases.²⁻³ Also, natural gas composed of non-
15 hydrocarbon gases such as nitrogen and carbon dioxide.⁴ Monitoring the hydrocarbon
16 compositions of natural gas is important to evaluate the quality of gas, and it has various
17 application in oil and gas industry during hydrocarbon production, transportation and
18 processing. For instance, in oil and gas sector, it is important to monitor and characterise the
19 composition of natural gas products precisely and continuously to ensure the quality of
20 natural gas. Moreover, monitoring the composition of produced gas could provide an early
21 indication of hydrate formation based on the reduction of some hydrocarbon components in
22 the gas phase.⁵⁻⁸ Gas chromatography (GC) is by far the most promising method for
23 measuring the concentration of hydrocarbons in gas phase.⁹ However, this approach has some

1 certain drawbacks, e.g., the measurement time. The time taken for the analysis of gas sample
2 using GC is varied from 5 minutes to 30 minutes depending on the length and temperature of
3 the columns. Longer column, and lower temperature will increase the retention time (required
4 time for a compound to pass through the column). Furthermore, GC analysis is pricey as
5 carrier gas and maintenance are needed for running GCs. Over the last decades, spectroscopic
6 methods have become more and more favourable in food, pharmaceutical, and petroleum
7 industries.¹⁰⁻¹⁴ The major advantages of these methods compared to other analytical methods
8 are that they can be used for at-line and in-line measurement to determine the selected species
9 of interest inside the sample with fast response and high accuracy.¹⁵ Moreover, there is no
10 need for carrier gas such as helium or nitrogen that are required by a GC, which further
11 reduces the operational cost. In particular, vibrational spectroscopy methods such as near-
12 infrared (NIR), middle-infrared (MIR) and Raman spectroscopy are well-suited for the
13 determination of hydrocarbon species inside the gas phase.^{9, 16} However, each method has its
14 advantages and disadvantages.¹⁷ The infrared region is based on the interaction of
15 electromagnetic radiation with the compounds and is described by the energy transfer
16 between the light and the matter. The infrared region is split into three subsets, near-infrared
17 (NIR), middle-infrared (MIR) and far-infrared (FIR). MIR and NIR are employed
18 commercially in the context of natural gas analysis, where MIR takes the measurement
19 wavelength between 2500 to 25000 nm (wavenumber 4000 to 400 cm^{-1}), in which
20 fundamental vibrational bands can be found. NIR covers 780 to 2500 nm (wavenumber
21 12500 to 4000 cm^{-1}) range and represents the vibrational overtone and combination bands
22 that are derived by the fundamental vibrational observed in the MIR region.

1 A number of research projects have been carried out for the development and improvement of
2 MIR and NIR techniques to monitor and determine the selected species of interest in the gas
3 mixtures.¹⁷ Danta *et al.*¹⁸ employed NIR spectroscopy to monitor the changes in the
4 concentration of methane in natural gas for quality control where the pressure for calibration
5 and test samples were fixed at 0.4 MPa at room temperature. They selected the spectral range
6 from 9100 to 4800 cm^{-1} (1099 to 2083 nm) since it carries the relevant spectroscopic
7 information to monitor the methane content in natural gas. A comparison was made between
8 NIR and MIR spectroscopies to measure methane, ethane and propane contents in natural
9 gases using various chemometric algorithms by Makhoukhi *et al.*¹⁹ They found that NIR
10 spectroscopic method is more accurate than MIR to measure methane, ethane, and propane
11 contents in natural gases. NIR spectroscopy has found broader applications in the industrial
12 processes than MIR spectroscopy. One of the reasons is that transmitting materials for NIR
13 are less expensive than MIR.²⁰ Also, due to high absorption of materials in MIR region and
14 low amount of energy produced by the MIR sources, samples need to be analysed through a
15 very efficient and tiny path length.²¹ In 2014, Rohwedder *et al.*²² reported the use of a
16 MicroNIR spectrometer to determine concentrations of methane, ethane, propane and butane
17 in synthetic gas mixtures at atmospheric pressure. Partial least squares (PLS) was employed
18 to develop the calibration model to relate the spectrum of synthetic gas mixtures and their
19 methane, ethane, propane and butane contents. The results revealed that this instrument could
20 be employed as an optical hydrocarbon analyser with good accuracy and fast response.

21 Literature survey shows that there are limited NIR spectroscopy works reported for
22 measuring the concentration of hydrocarbon species (methane through butanes) in the gas
23 phase at high pressures. For field applications where pipeline systems are operating at high

1 pressures, there is a strong demand for fast response, user-friendly, and cost-effective
2 methods for composition analysis of hydrocarbon gases at in-situ pressures. GC analysis just
3 can be performed close to atmospheric pressure, and it is necessary to regulate the pressure of
4 gas from high pressures to low pressures. This reduction in pressure can cause changes in the
5 composition of the hydrocarbon phases as heavy hydrocarbons drop out from the gas phase
6 into the liquid phase. Dong *et al.*²³ developed a downhole NIR analyser to predict the
7 concentration of hydrocarbons in gas phase using chemometric methods under in-situ
8 pressure up to 137.89 MPa and temperature up to 423.15 K. It should be mentioned that in
9 that study propane, butanes and pentanes were grouped together in order to improve the
10 accuracy of gas analyser and these components cannot be measured separately in gas samples
11 by the analyser.

12 In this communication, the capability of the FTNIR spectroscopy and PLS algorithm was
13 investigated for measuring methane, ethane, propane, i-butane and n-butane contents at
14 various pressures and temperatures. Different pre-processing techniques were carried out
15 prior to the construction of PLS calibration models to investigate the model prediction
16 capability. Moreover, we examined the influence of a change in pressure and/or temperature
17 in the accuracy of one of the developed PLS regression models that were calibrated at a
18 known condition of temperature and pressure.

19 **2. METHOD**

20 The basic of NIR spectroscopy measurement is based on the Beer-Lambert's law given in
21 Eq.1:

$$22 \quad A = \log\left(\frac{I_0}{I}\right) = \xi lc \quad (1)$$

1
2
3 1 Where A represents the absorbance of the beam, I_0 is the intensity of incident light, I is the
4 intensity of incident light after passing through the sample, ξ is the molar absorptivity, l is the
5 sample path length (the length that light travels), and c is the sample concentration. Major
6 absorption bands for hydrocarbons in the NIR region normally occur in the ranges 1100-1200
7 nm, 1350 to 1450 nm, and 1600 to 1850 nm.³ The absorption bands in the ranges from 1100
8 to 1200 nm are related to the second overtone of the hydrocarbons, whereas the absorption
9 bands between 1350 to 1450 nm belongs to the first overtone of the combination modes of
10 the hydrocarbons which are associated with CH bonding. The main features of the spectra are
11 the absorption in the range of 1600 to 1850 nm that is associated with the first overtone
12 stretching of CH, CH₂ and CH₃ bands that are related to the methane, ethane, propane, i-
13 butane and n-butane in the hydrocarbons gas mixtures.

14 FTNIR spectra of pure methane, ethane, propane, i-butane, and n-butane at atmospheric
15 pressure, at a temperature of 293.15 K and spectral range of 1100 to 2000 nm are depicted in
16 Figure 1. The FTNIR spectra of hydrocarbon components in this region are different which is
17 due to the differences in their molecular structures. As it is evident, there are some
18 overlapping bands throughout the NIR regions that are referring to the first overtone, second
19 overtone and first overtone combination band of pure hydrocarbon components. Hence, it is
20 required to employ multivariate methods to extract the desired information from spectral data
21 to measure the concentration of each single component.

22 **2.1 Partial least square (PLS) analysis**

23 Principal components analysis (PCA) and partial least squares (PLS) are the most commonly
24 multivariate mathematic techniques to develop the calibration models using the spectra
25 obtained from calibration samples. The PCA goal is to extract the information from a data

1 matrix (X) by explaining the variation in the data, whereas PLS models the relation between
2 the spectra data and the components concentration using latent variables. The details about
3 PLS and PCA algorithms as well as their pros and cons can be found in other studies²⁴⁻²⁵. In
4 the current study, the PLS algorithm is used to perform the calibration and the regression to
5 determine the concentration of methane, ethane, propane, i-butane, and n-butane in the gas
6 phase. The PLS analysis can be done by using all the calibration spectra in the region of
7 interest. These latent variables capture the maximum covariance between the reference data
8 and the recorded spectrum. The equation of PLS models are derived as follows:

$$9 \quad X = TP^T + E \quad (2)$$

$$10 \quad Y = UQ^T + F \quad (3)$$

11 Where X is the spectral data, Y represents the content of hydrocarbons (methane through
12 butanes) in synthetic gas mixtures. P and Q resemble the loadings matrix of the X and Y,
13 respectively. T and U correspond to score matrices of X and Y. E and F are both residual
14 errors of the X and Y that represent the noise or irrelevant variable.²² In this method, the
15 latent variables of the dependent variable (X) are correlated to the latent variables of the
16 independent variable (Y) [Eqs.2 and 3]. In other words, the aim of using a PLS regression
17 model is to decompose both X and Y into two loadings and score matrices and then find a
18 regression model between the score matrices of X and Y with a maximum covariance [Eq.4].

$$19 \quad U = TB + G \quad (4)$$

20 Where B is a diagonal matrix that represents the regression coefficients between T and U and
21 G is the residual value. This regression coefficient can be used for future content prediction
22 of the component of interest.²⁶⁻²⁷

1
2
3 1 One of the important factors to develop an appropriate PLS model using the FTNIR spectra
4
5 2 of calibration data is selecting the optimum number of latent variable (LV).²⁸ There are
6
7 3 several approaches to find the optimal number of latent variables, which is described in the
8
9 4 literature.²⁹ In this work, the leave-one-out cross-validation was employed to develop the
10
11 5 PLS model. In this method, after removing of one sample from the calibration data set a
12
13 6 developed PLS model is used for the remaining samples to predict the concentration of the
14
15 7 removed sample. This procedure was repeated for all the calibration data set, and then the
16
17 8 root mean square errors of cross-validation (RMSECV) was calculated using Eq.5:

20
21
22 9
$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_{ci} - \hat{y}_{ci})^2}{n}}$$
 (5)
23
24
25

26 10 Where y_{ci} and \hat{y}_{ci} are the actual and predicted value of the sample that was left out from the
27
28 11 validation set and n is the number of validation samples. After selecting the appropriate
29
30 12 calibration model, a series of raw data that were not part of the calibration data set was used
31
32 13 to determine the accuracy of the developed model. Therefore, the root mean square error of
33
34 14 prediction (RMSEP) was measured for the predicted concentration of each individual
35
36 15 component. Standard error of prediction (SEP) was calculated using Eq. 6 to examine the
37
38 16 significance of bias in each PLS model.

39
40
41
42 17
$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_{pi} - \hat{y}_{pi} - bias)^2}{n}}$$
 (6)
43
44
45

46 18

47
48 19 The bias can be calculated through Eq. 7:

49
50
51 20
$$bias = \frac{\sum_{i=1}^n (y_{pi} - \hat{y}_{pi})}{n}$$
 (7)
52
53
54

1 Student's "t" statistic test was also carried out with 95% confidence level and n-1 degrees of
2 freedom³⁰, to investigate the existence of systematic error for each model.

$$3 \quad t_{value} = \frac{|bias|\sqrt{n}}{SEP} \quad (8)$$

4 Relative prediction deviation (RPD) was calculated to investigate the predictive performance
5 of the developed model. RPD is the ratio of the standard deviation of all the prediction set
6 (SD) to SEP.

$$7 \quad RPD = \frac{SD}{SEP} \quad (9)$$

8 The greater the RPD value, the higher ability of the PLS model to predict the concentration of
9 species. An RPD value below 1.5 reveals that the calibration model is poor and cannot be
10 used as a reliable model for further prediction.³¹

11 **2.2 Pre-processing of near-infrared spectra**

12 The main concept behind the pre-processing techniques is to eliminate unwanted background
13 data and reduce noise level before construction of the calibration model.^{20, 32} Various pre-
14 processing techniques can be applied to the spectral data to improve the accuracy of the
15 developed calibration model. In this study, the spectra were pre-treated using multiplicative
16 scatter correction (MSC); standard normal variate (SNV); first and second Savitzky-Golay
17 derivatives (SGD1 and SGD2)³²; first Savitzky-Golay derivative followed by the orthogonal
18 signal correction (SGD1 + OSC) and second Savitzky-Golay derivative followed by the
19 orthogonal signal correction (SGD2 + OSC). OSC was applied after the SGD1 and SGD2 to
20 eliminate the spectral data that are unrelated to the dependent variable to reduce the data
21 variance in the spectra. Detail theory of the OSC is described in elsewhere.³³ It should be
22 highlighted that, in all cases, the pre-treated spectra were also mean-centred before

1 developing the PLS model. The SGD1-OSC method produced the best results in terms of root
2 mean square error of cross validation (RMSECV) compare to other techniques and was used
3 for data pre-treatment. This paper, therefore, considers the results of PLS model using this
4 technique.

5 The pre-processing techniques including MSC, SNV, first and second Savitzky-Golay
6 derivatives and OSC plus PLS analyses were all performed in Unscrambler® X10.3 (CAMO,
7 Oslo, Norway).

8 **3. EXPERIMENTAL**

9 **3.1 Apparatus**

10 Figure 2 shows a schematic of the experimental apparatus. An FTNIR spectrometer
11 (Arcoptix) was used for spectra acquisition in the working range between 900 to 2600 nm. A
12 20 watt halogen light sources with attenuator (HL-2000-FHSA, Ocean Optics) was guided to
13 the FTNIR spectrometer via a high-pressure cell containing the test sample through fibre
14 optic cables. The sample cell used for the FTNIR measurements is made of titanium with an
15 effective optical path length of 1.4 cm, internal diameter of 5.0 cm (inner volume of about 28
16 cm³), and operating temperature from 273.15 to 323.15 K, and maximum operating pressure
17 of 35 MPa. Each end of the cell are equipped with a sapphire window that allows the light
18 pass through the sample. A lens holder was positioned at one end of the cell window to house
19 the collimating lens. The collimating lens (74-UV, Ocean Optics) was employed to convert
20 the divergent beam of the light source into a parallel beam to improve the signal intensity.
21 The FTNIR spectrometer has USB connectivity for control and data acquisition. For
22 measurements of test samples, an average of thirty spectra was recorded, and the spectrum

1 measurement was repeated three times. The analysis for each measurement typically took
2 about 90 seconds.

3 The temperature of the cell was controlled using a jacket connected to a temperature-
4 controlled bath that circulates coolants through the jacket. A high precision Platinum
5 Resistance Temperature (PRT) probe was used to measure the temperature of the sample. A
6 precision thermometer (Prema 3040) was employed to calibrate the temperature probe. The
7 uncertainty of the calibrated PRT probe is estimated to be within ± 0.1 K. A piezo resistive
8 silicon pressure transducer (Druck PDCR 4060) was used to measure the pressure inside the
9 sample cell. The pressure transducer was regularly calibrated against a dead weight pressure
10 balance. This calibration procedure ensures that the pressure transducer is accurate to ± 0.01
11 MPa.

12 **3.2 Materials**

13 The specification and suppliers of the materials used in this study are listed in Table 1.
14 Thirty-one calibration gas mixtures and twenty prediction gas mixtures were prepared in the
15 laboratory with composition of methane varying between 80 and 100 cmol/mol, ethane
16 between 1 to 12 cmol/mol, propane between 1 to 6 cmol/mol, i- butane between 0.5 to 2
17 cmol/mol and n-butane between 0.5 to 2 cmol/mol. The concentration of the calibration
18 mixtures was designed to cover the typical concentration ranges of natural gas components
19 (see Table 2). All the synthetic gas mixtures were prepared inside the high-pressure vessel by
20 combining pure components on a weight basis using a two digit electronic balance
21 (Sartorius, Cubis MSA8201S-0CE-D0, accuracy ± 0.01 g). After injecting all the pure
22 components into the vessel, the composition of the mixtures was calculated for each sample.
23 All the synthetic gas mixtures were then analysed using a gas chromatography (Varian model

1 CP-3800) to validate their compositions. A Flame Ionisation Detector (FID) was used to
2 detect the concentration of hydrocarbons in the gas samples, and the Thermal Conductivity
3 Detector (TCD) was employed to detect the concentrations of nitrogen and carbon dioxide. A
4 small difference was observed between the experimental value and measured value by GC
5 (see Table 2). Average of three measurements by GC was chosen as the true value for the
6 calibration and prediction data set.

7 **3.3 Procedures**

8 Before starting the experiments, the high-pressure cell, sample cylinders, and the sapphire
9 windows were cleaned thoroughly by n- heptane and then compressed air was passed inside
10 the line, valve, and cell to dry the whole system. To ensure there is no leakage in the entire
11 system, a pressure test was performed by injecting nitrogen into the cell, and the system was
12 left for few hours. Once no leakage was observed in the system, the system was depressurised
13 and vacuum was applied to the high-pressure cell, as well as all fluid loading lines. A
14 spectrum of empty cell (vacuumed) with an average of thirty scans was recorded as the
15 reference spectrum before each measurement.

16 A typical spectrum of the gas mixture was measured at various pressures and temperatures
17 following the procedure below. Firstly, the FTNIR cell was set at desired temperature, and
18 then the cell was connected to a pressurised cylinder containing the synthetic gas mixture and
19 filled. A floating piston in the middle of the pressure cylinder was driven using pressurised
20 nitrogen. Hence, the pressure of cell can be adjusted by injection/withdrawal of nitrogen
21 behind the moving piston. The gas samples were injected/withdrawn slowly into/from the
22 FTNIR cell to avoid damage to the sapphire windows. Once the desired equilibrium pressure
23 had been reached the line was disconnected from the pressure cylinder and the spectrum was

1 recorded three times by the averaging of thirty scans to reduce the noise level. The pressure
2 of the cell was kept constant during FTNIR measurements. The averaged spectrum was then
3 used for data pre-processing and developing the (PLS) regression models. This procedure
4 was repeated for all the samples at three different temperatures.

5 **4. RESULTS AND DISCUSSION**

6 **4.1 Influence of temperature and pressure on FTNIR spectrum of hydrocarbons** 7 **mixtures**

8 In order to examine the influence of temperature on FTNIR spectra at constant pressure, the
9 pressure of the synthetic gas mixture inside the cell was kept constant at 6.89 MPa, and the
10 FTNIR spectra were recorded at three different temperatures (278.15 K, 293.15 K, and
11 313.15 K). As it clear in Figure 3, the absorption of the gas mixture increased slightly with
12 decreasing temperature at regions those hydrocarbons absorb the NIR radiation. These
13 changes with temperature designate that the FTNIR spectra of hydrocarbons mixtures are
14 dependent on temperature. Therefore, it is necessary to record the FTNIR spectrum at
15 different temperatures within a certain range to expand the range of applicability of this
16 method.

17 Spectra of one gas mixture at constant temperature and four different pressures are presented
18 in Figure 4. It is evident that the absorbance spectrum of the gas mixture is highly dependent
19 on pressure, and an increase in pressure leads to an increase in absorbance spectrum at three
20 different NIR regions that carry the relevant information for interested hydrocarbon species
21 that compose the gas mixture. Moreover, to avoid optical saturation at different pressures, the
22 detector gain level or the light intensity has been set properly.

1
2
3 1 All this analysis confirms that the spectroscopic signals are temperature and pressure
4
5 2 dependence and variation in pressure and temperature may affect the accuracy of
6
7 3 chemometric models that is constructed in a single operational temperature and pressure.
8
9
10 4 Thereby, Individual PLS models were built at various temperatures and pressures to extend
11
12 5 the application of this method since the temperature and pressure of gas mixtures may vary in
13
14 6 natural gas pipelines. The operation range of pressure was chosen for the construction of the
15
16 7 calibration model according to the path-length of the FTNIR cell. Path length is the distance
17
18 8 that the light travels through the gas samples, and this length can be carefully chosen in
19
20 9 accordance with the amount of an absorber along the distance that light travels. Hence,
21
22 10 selection of optimum optical path length is the first important step in the development of
23
24 11 FTNIR spectroscopy method. One criterion is to achieve sufficient signal-to-noise ratio in the
25
26 12 interested spectra region. Very strong absorbance ($AU > 2$) increases the error of
27
28 13 measurement because the sample absorbs most of the lights and only a small amount of light
29
30 14 can be detected by the detector. On the other hand, very low absorbance provides not so
31
32 15 much information about the evolution of the species concentration. Consequently, to decrease
33
34 16 the error and uncertainty of the measurement and to achieve to a good signal to noise ratio,
35
36 17 the operating pressure is chosen to be in a range between 3.44 MPa to 13.78 MPa where the
37
38 18 absorbance unit varies from around 2 to 0.05 AU.

19 **4.2 Spectral range selection**

20 Selection of the right wavelength region plays an important role to improve the performance
21
22 21 of the calibration model. The idea behind the optimum wavelength selection is the
23
24 22 identification of an appropriate subset that will provide lower error values for the validation
25
26 23 and prediction data set.³⁴ To arrive at the true wavelength region, the region with lower

1
2
3 1 RMSECV value and latent variable numbers should be selected as an optimal subset of
4
5 2 wavelengths for a given data set.
6
7

8 3 Figure 5 shows the FTNIR spectra of all thirty-one synthetic gas mixtures containing
9
10 4 methane, ethane, propane, i-butane and n-butane that are used for quantitative analysis at
11
12 5 pressure of 10.34 MPa and temperature of 293.15 K. After inspecting the whole NIR spectral
13
14 6 region, the spectral range lower than 1100 nm and higher than 2100 nm were excluded from
15
16 7 data set due to poor and extreme absorption of hydrocarbons in these areas, respectively.
17
18

19
20 8 In order to select the appropriate spectral range for generation of the calibration models, a
21
22 9 PLS model pre-treated with first Savitzky-Golay derivative (SGD1) was developed for one of
23
24 10 the given data set at a specific temperature and pressure condition ($T = 293.15$ K and $P =$
25
26 11 6.89 MPa). The plots of weighed regression coefficients for the developed PLS model in the
27
28 12 NIR range between 1100 to 2000 nm were plotted for all the components together in order to
29
30 13 discover the regions that are mostly contributed in the generation of the calibration model
31
32 14 (Figure 6). The significance of a variable in developing a PLS model can be find out
33
34 15 according to the magnitude of the PLS regression coefficients.³⁵ It is apparent that the
35
36 16 magnitude of the regression coefficients for all the components are very small at the second
37
38 17 overtone (1100-1200 nm) and the first overtone of the combination modes of hydrocarbons
39
40 18 (1350 - 1400 nm) and consequently these regions cannot provide valuable information for the
41
42 19 construction of the calibration models. As can be seen, the first overtone region of
43
44 20 hydrocarbons (1600 - 1850 nm) is contributed most to the construction of the calibration
45
46 21 model. Significant deviations were associated with all the components with the PLS models
47
48 22 developed using regions between 1100 - 1200 nm, 1350 - 1450 nm which directed that the
49
50 23 predictions of methane, ethane, propane, i-butane and n-butane cannot be trusted for models
51
52
53
54
55
56
57
58
59
60

1 that developed across these regions. The first overtone region provides better accuracy for the
2 gas species with lower molar absorptivity (low concentration). Lower RMSECV values and
3 few numbers of latent variables were observed for all the individual components in this
4 region for all calibration pressures and temperatures. To find the optimum wavelength region
5 for quantitative determination of interested hydrocarbon components, the NIR spectrum from
6 1600 - 1850 nm was split into several intervals. The results revealed that the restricted region
7 between 1670 - 1800 nm provides lower RMSECV and RMSEP values compared to other
8 regions (See Supporting Information).

9 **4.3 Construction of the calibration models (PLS)**

10 After recording the FTNIR spectra of each sample (synthetic gas mixture) at specified
11 pressure and temperature conditions, firstly the reference intensity (I_0 , vacuumed cell) and the
12 sample spectrum (I) were both normalized (maximum method) and then used to convert to
13 absorbance unit. Various pre-processing techniques were then applied to calibration data set
14 to evaluate the effect of these techniques on the accuracy of the developed PLS model. As
15 mentioned earlier, the best results obtained when the first Savitzky-Golay derivative (SGD1)
16 with smoothing over 5 points plus orthogonal signal correction (OSC) were applied to the
17 FTNIR spectra data. In order to find the optimum number of the smoothing points, different
18 smoothing points were implemented to the spectra data using the Savitzky-Golay algorithm.
19 The best results in terms of RMSECV were observed when the Savitzky-Golay with five
20 smoothing points was applied to the spectra data. A slight improvement was noted in the
21 RMSECV value for all the components while orthogonal signal correction applied to the
22 spectra data after applying the Savitzky-Golay algorithm. The presence of outliers in dataset
23 was investigated during construction of the PLS models by computing Q residuals and

1
2
3 1 Hotelling's T^2 with 95% confidence level. Hotelling's T^2 is defined as the sum of normalised
4
5 2 squared scores³⁶ and was employed to measure the variation between the sample and the
6
7 3 model. Q residuals are the sum of square errors between the sample and the model, in other
8
9 4 word, Q residuals present the amount of variation that is not captured to build up the model
10
11 5 for each sample.³⁷ Samples with large residual and high Hotelling's T^2 values were identified
12
13 6 as anomalous samples and removed from the dataset. The existence of outlier in the dataset
14
15 7 could be due to the error in FTNIR measurements or the error in sample preparation.³⁸
16
17 8 Therefore, it is required to remove the outliers from the data set during construction of the
18
19 9 calibration models and develop new PLS models with remaining calibration data set because
20
21 10 the presence of outliers affects the accuracy and performance of the regression model. In this
22
23 11 work, one sample in calibration data set was found as outliers. Furthermore, the optimum
24
25 12 number of the latent variables for each component was selected based on the RMSECV
26
27 13 values of different latent variables. With an increase in a number of latent variable, the
28
29 14 RMSECV value starts to decrease. Basically, while no significant variation is observed in the
30
31 15 RMSECV value from one number of latent variable to the next, the subsequent latent
32
33 16 variables are not considered for developing of the PLS models since they may just provide
34
35 17 some noise or irrelevant information. The optimal number of latent variables for the PLS
36
37 18 model was selected to be 2 for methane, 4 or 5 for ethane, propane and i-butane, and 5 or 6
38
39 19 for n-butane according to the target temperature and pressure of the system. The results of the
40
41 20 developed PLS models are tabulated in Table 3. Low RMSECV values and high R^2 values
42
43 21 were found for all the final PLS models, indicating created models can be trusted for
44
45 22 monitoring the composition of main hydrocarbons in the gas mixtures.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 4.4 Evaluation of the PLS models

2 4.4.1 Synthetic gas mixtures

3 To evaluate the accuracy of the developed PLS models twenty independent synthetic gas
4 mixtures within the range of the calibration were prepared in the laboratory and analysed with
5 GC. The results of PLS models that were obtained for methane, ethane, propane, i-butane and
6 n-butane at various pressure and temperature conditions are presented in Table 4.
7 Additionally, the performance of each PLS model at various temperatures and pressures for
8 independent samples is represented graphically through plots of FTNIR-predicted data
9 derived from the PLS models versus the values determined by the GC (Figure 7 – Figure 9).
10 As can be seen, there is no significant difference between FTNIR predictive values and
11 measured values, and the PLS-determined methane, ethane, propane, i-butane and n-butane
12 concentrations are so adjacent to the 0-error line which indicates the good predictive
13 capability of the developed PLS models at different pressures. Regarding RMSEP, results for
14 methane, ethane, propane, i-butane and n-butane varied from 0.176 to 0.202, 0.164 to 0.204,
15 0.133 to 0.176, 0.094 to 0.153, and 0.094 to 0.141 cmol/mol, respectively, indicating a small
16 difference in comparison to the measured values at various pressure and temperature
17 conditions. Furthermore, the R^2 statistic was calculated for individual samples, and results
18 show good agreement between predicted and measured values. It can be seen all the
19 developed PLS models have a good predictive statistics in terms of low RMSEP and SEP
20 values and high R^2 value that indicate the capability of the proposed PLS models to
21 determine the composition of samples. However, in some cases, higher RMSEP values were
22 observed for some of the studied components at a pressure of 3.44 MPa compared to others.
23 This may be attributed to the low absorbance of hydrocarbons at respective pressure
24 compared to other calibration pressures. As one can see from Table 4 the values of SEP and

1
2
3 1 RMSEP are very similar, showing that biases do not affect the accuracy of PLS models. One-
4
5 2 tailed t-test was also carried out for all the developed PLS models using Eq. 8 for the
6
7 3 prediction samples to test the significance of bias that was included in the model. It was
8
9 4 noticed that the relevant bias does not produce significant systematic errors since the $t_{\text{calculated}}$
10
11 5 for all the components was less than t_{critical} at a level of 95% confidence.³⁰ All of the models
12
13 6 developed in this study had RPD value higher than 2.77, indicating created models can be
14
15 7 trusted for monitoring the composition of main hydrocarbons in gas samples. The RPD
16
17 8 values ranged from 20.85 to 24.59, 13.07 to 16.56, 11.24 to 14.96, 2.78 to 4.37 and 3.27
18
19 9 to 4.83 for methane, ethane, propane, i-butane and n-butane models, respectively, at
20
21 10 different T&P conditions. Limit of detection (LoD) was also calculated for the final
22
23 11 multivariate calibration models to define the minimum concentration of a species that can be
24
25 12 measured using the FTNIR spectrometer. The replicate spectra of 10 samples without any
26
27 13 solute (pure nitrogen) and the spectra of 10 samples with the highest value of solute were
28
29 14 measured at different T&P conditions. The corresponding final PLS regression models were
30
31 15 employed to predict the concentration of each component. Then, the average standard
32
33 16 deviation of the predicted values was calculated for each model and was multiplied by 10/3 to
34
35 17 roughly calculate the LoD³⁹.

18 **4.4.2 Natural gas and certified gas mixtures**

19 The prediction capability of the developed PLS models at 293.15 K and various pressures
20 were also examined for one unknown natural gas mixture and one certified gas mixture that
21 contains hydrocarbon components (methane through pentanes), carbon dioxide, and nitrogen
22 against the gas chromatography as a conventional and reliable method for monitoring the
23 concentration of hydrocarbons in gas mixtures. It is well-known that nitrogen as well as

1 oxygen are not NIR active and do not absorb the emitted light. Moreover, carbon dioxide
2 absorbs the NIR light at another region different from the region which was used to construct
3 the calibration models. Figure 10 illustrates carbon dioxide, methane, and nitrogen absorption
4 spectra in the range between 1600 to 2100 nm that was obtained by the FTNIR spectrometer
5 at room temperature and 3.44 MPa. It is clear that carbon dioxide and methane
6 (representative of hydrocarbons) absorb the NIR lights at two different regions, and there is
7 no interference between absorption spectra of hydrocarbons and carbon dioxide. It is
8 expected that existence of these components in the sample do not affect the accuracy of the
9 results. The objectives of these tests were to compare the NIR values obtained with those of
10 GC analysis and to investigate the influence of the presence of nitrogen, CO₂, and other
11 heavier hydrocarbons such as pentanes in the natural gas to confirm the accuracy and
12 reliability of the developed PLS models. The spectra were recorded at 293.15 K and four
13 different pressures. Then, the withdrawn samples were injected into the GC by means of gas
14 syringes with the volume of 500 µl. Measurement time for GC and FTNIR analysis were
15 about 15 minutes and 90 seconds, respectively. For both methods, the average value of three
16 measurements was used as the final value. The results are summarised in Tables 5 and 6. It
17 should be noted that the measured values obtained by GC for nitrogen, carbon dioxide, i-
18 pentane and n-pentane was removed from the results and the rest of the measured values by
19 GC were normalised to 100 cmol/mol. It can be seen that there is a good agreement between
20 the predicted and measured values of hydrocarbons using NIR (at all considered pressures)
21 and GC for the same components. As expected, the presence of nitrogen, carbon dioxide in
22 the natural gas do not affect the accuracy of the chemometric models. The difference between
23 NIR and GC measurements can be explained by the accuracy of their measurements. These

1
2
3 1 results suggest that this method is insensitive to the components that are NIR inactive, as well
4
5 2 as components that do not contain carbon-hydrogen molecular bonds in their structures.
6
7

8 3 To summarise, FTNIR spectroscopy is a fast, accurate and robust method to monitor and
9
10 4 measure the composition of main hydrocarbons species simultaneously in the gas phase.
11
12 5 Moreover, because of its capability to operate at high pressures, the FTNIR analyser can be
13
14 6 employed at in-situ pressure beside the gas outlet of the pipelines for online monitoring of
15
16 7 hydrocarbon gas compositions without requiring any gas carrier and skilled operators in
17
18 8 comparison with gas chromatography, which makes it easy to use with significant reduction
19
20 9 in the operation costs.
21
22
23

24 **4.4.3 Temperature and pressure sensitivity analysis**

25
26
27 11 The effect of a shift in pressure and temperature was examined for a specific PLS regression
28
29 12 model. A series of experiments were conducted to characterise the sensitivity of the
30
31 13 constructed PLS regression model at a pressure of 6.89 MPa and temperature of 293.15 K. To
32
33 14 investigate the effect of pressure shift, firstly one synthetic gas sample containing methane,
34
35 15 ethane, propane, i-butane and n-butane was introduced into the FTNIR cell, and the pressure
36
37 16 and temperature of the cell were then set at 6.89 MPa and 293.15K respectively. The
38
39 17 concentration of each component was predicted using the developed PLS models, and the
40
41 18 average of predicted values for three measurements for all the five components was utilised
42
43 19 as the reference value. Then, the pressure of the gas sample in the cell was varied from 6.89
44
45 20 MPa to 7.58 MPa and from 6.89 to 6.20 MPa with a specific interval at the fixed temperature.
46
47 21 The spectra were recorded in triplicate for each pressure to calculate the standard deviation.
48
49 22 The components concentration of the sample was predicted using acquired spectra by using
50
51 23 the developed PLS model at 293.15 K and pressure of 6.89 MPa and were subtracted from
52
53
54
55
56
57
58
59
60

1 the reference values to calculate the deviation in cmol/mol. The results are presented in
2 Figure 11. These results specified that the variation of the pressure affects the performance of
3 the PLS model at the calibrated pressure. As shown, a negative systematic error was observed
4 for propane, ethane, and n-butane while the pressure for the validation set was lower than the
5 pressure of the calibration set whereas, a positive systemic error was found for i-butane and
6 methane at pressures lower than calibration set. A contrariwise trend was observed for all the
7 components while the measurement pressure was higher than the pressure of the calibration
8 set. It is worthy to note that our findings indicate that the error of PLS models to predict the
9 concentration of individual components is negligible when the shift is less than 0.14 MPa in
10 the measurement pressure, which is just related to the uncertainty of measurement.

11 The accuracy of one of the developed PLS model ($P = 6.89$ MPa, $T = 293.15$ K) was also
12 evaluated regarding the influence of sample temperature variation on the FTNIR predicted
13 values. At this time, the pressure of the sample inside the cell was kept constant ($P = 6.89$
14 MPa) and the spectra were recorded at various temperatures. In Figure 12, the deviation of
15 the measured values for methane, ethane, propane, i-butane, and n-butane with respect to the
16 actual value of the synthetic gas mixture (values measured at $P = 6.89$ MPa, $T = 293.15$ K)
17 using the chemometric models is shown as a function of temperature together with error bars.
18 No significant variation in the concentration of methane, ethane and propane had been
19 observed while the shift in temperature was less than 2 K. It should be noted that small errors
20 in this range of temperature are within the uncertainty measurements of the models.
21 Furthermore, negligible errors were noticed for i-butane and n-butane while the temperature
22 was changed from 278.15 K to 288.15 K with 1 K interval. All these results confirmed that
23 variation in pressure and temperature of the sample affect the accuracy of PLS models that

1
2
3 1 created at specific pressure and temperature. Therefore, it is necessary to maintain the
4
5 2 temperature and pressure of the system to an accuracy of ± 2 K and ± 0.14 MPa to achieve
6
7 3 accurate results.
8
9

10 11 4 **5. CONCLUSIONS**

12
13 5 The results of this study demonstrate that FTNIR spectroscopy methods associated with
14
15 6 chemometric techniques can be used to determine the composition of methane, ethane,
16
17 7 propane, i-butane and n-butane in the gas phase at in-situ pipeline pressure. FTNIR
18
19 8 calibration models were developed using PLS regression with the first Savitzky-Golay
20
21 9 derivative plus orthogonal signal correction (OSC) pre-treatment in the spectral range
22
23 10 between 1670 to 1800 nm at various pressures (3.44, 6.89, 10.34 and 13.78 MPa) and
24
25 11 temperatures (278.15 K, 293.15 K, and 313.15 K). Overall, the PLS models yielded
26
27 12 reasonably low deviations from the GC analysis for all the components. The sensitivity of the
28
29 13 FTNIR spectroscopy technique to the test pressure and temperature was investigated. It was
30
31 14 concluded that the developed PLS models could provide sufficient measurement accuracy if
32
33 15 the shift is less than 2 K in the measurement temperature and 0.14 MPa in the measurement
34
35 16 pressure.
36
37
38
39

40 41 17 **ACKNOWLEDGMENT**

42
43 18 This work was part of a Joint Industrial Project (JIP) at the Institute of Petroleum
44
45 19 Engineering, Heriot-Watt University. The JIP was financially funded by BP and TOTAL
46
47 20 which is gratefully acknowledged. The author would also like to thank Jim Allison for his
48
49 21 assistance.
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
1 REFERENCES

- 2 1. Kirk, J. L.; Bristow, A. L.; Zanni, A. M., Exploring the market for Compressed Natural Gas light
3 commercial vehicles in the United Kingdom. *Transportation Research Part D: Transport and*
4 *Environment* **2014**, *29*, 22-31.
- 5 2. Dai, Q.; Lastoskie, C. M., Life cycle assessment of natural Gas-powered personal mobility
6 options. *Energy & Fuels* **2014**, *28* (9), 5988-5997.
- 7 3. Burns, D. A.; Ciurczak, E. W., *Handbook of near-infrared analysis*. CRC press: 2007.
- 8 4. Baker, R. W.; Lokhandwala, K., Natural gas processing with membranes: an overview.
9 *Industrial & Engineering Chemistry Research* **2008**, *47* (7), 2109-2121.
- 10 5. Glénat, P.; Munoz, J.; Haghi, R.; Tohidi, B.; Mazloum, S.; Yang, J., Field Test results of
11 monitoring Hydrates formation by gas composition changes during gas/condensate production with
12 AA-LDHI. In *Proceeding of the 8th International Conference on Gas Hydrates, Beijing, China*, 2014.
- 13 6. Tohidi, B.; Anderson, R.; Chapoy, A.; Yang, J.; Burgass, R. W., Do we have new solutions to
14 the old problem of gas hydrates? *Energy & Fuels* **2012**, *26* (7), 4053-4058.
- 15 7. Kondo, W.; Ohtsuka, K.; Ohmura, R.; Takeya, S.; Mori, Y. H., Clathrate-hydrate formation
16 from a hydrocarbon gas mixture: Compositional evolution of formed hydrate during an isobaric
17 semi-batch hydrate-forming operation. *Applied Energy* **2014**, *113*, 864-871.
- 18 8. Glénat, P.; Saha, P.; Macpherson, C.; Yang, J.; Tohidi, B., Field Tests Results of Production
19 Inside Hydrate P&T Zone Using New Hydrate Monitoring Instruments. In *Offshore Mediterranean*
20 *Conference and Exhibition*, Offshore Mediterranean Conference: 2015.
- 21 9. Eichmann, S.; Weschta, M.; Kiefer, J.; Seeger, T.; Leipertz, A., Characterization of a fast gas
22 analyzer based on Raman scattering for the analysis of synthesis gas. *Review of Scientific Instruments*
23 **2010**, *81* (12), 125104.
- 24 10. Balabin, R. M.; Lomakina, E. I.; Safieva, R. Z., Neural network (ANN) approach to biodiesel
25 analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near
26 infrared (NIR) spectroscopy. *Fuel* **2011**, *90* (5), 2007-2015.
- 27 11. Mueller, J. J.; Baum, S.; Hilterhaus, L.; Eckstein, M.; Thum, O.; Liese, A., Simultaneous
28 determination of mono-, di-, and triglycerides in multiphase systems by online Fourier transform
29 infrared spectroscopy. *Analytical chemistry* **2011**, *83* (24), 9321-9327.
- 30 12. Cervera-Padrell, A. E.; Nielsen, J. P.; Jønch Pedersen, M.; Müller Christensen, K.; Mortensen,
31 A. R.; Skovby, T.; Dam-Johansen, K.; Kiil, S.; Gernaey, K. V., Monitoring and control of a continuous
32 Grignard reaction for the synthesis of an active pharmaceutical ingredient intermediate using inline
33 NIR spectroscopy. *Organic Process Research & Development* **2012**, *16* (5), 901-914.
- 34 13. Balabin, R. M.; Syunyaev, R. Z.; Karpov, S. A., Quantitative measurement of ethanol
35 distribution over fractions of ethanol-gasoline fuel. *Energy & fuels* **2007**, *21* (4), 2460-2465.
- 36 14. Ge, L. N.; Zhan, H. L.; Leng, W. X.; Zhao, K.; Xiao, L. Z., Optical characterization of the
37 principal hydrocarbon components in natural gas using terahertz spectroscopy. *Energy & Fuels* **2015**,
38 *29* (3), 1622-1627.
- 39 15. Alves, J. C. L.; Poppi, R. J., Simultaneous determination of hydrocarbon renewable diesel,
40 biodiesel and petroleum diesel contents in diesel fuel blends using near infrared (NIR) spectroscopy
41 and chemometrics. *Analyst* **2013**, *138* (21), 6477-6487.
- 42 16. Eichmann, S. C.; Kiefer, J.; Benz, J.; Kempf, T.; Leipertz, A.; Seeger, T., Determination of gas
43 composition in a biogas plant using a Raman-based sensor system. *Measurement Science and*
44 *Technology* **2014**, *25* (7), 075503.
- 45 17. Kiefer, J., Recent Advances in the Characterization of Gaseous and Liquid Fuels by Vibrational
46 Spectroscopy. *Energies* **2015**, *8* (4), 3165-3197.

- 1
2
3 1 18. Dantas, H. V.; Barbosa, M. F.; Nascimento, E. C.; Moreira, P. N.; Galvão, R. K.; Araújo, M. C.,
4 2 Screening analysis of natural gas with respect to methane content by near-infrared spectrometry.
5 3 *Microchemical Journal* **2014**, *114*, 210-215.
6 4 19. Makhoukhi, N.; Péré, E.; Creff, R.; Pouchan, C., Determination of the composition of a
7 5 mixture of gases by infrared analysis and chemometric methods. *Journal of molecular structure*
8 6 **2005**, *744*, 855-859.
9 7 20. Manley, M., Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis
10 8 of biological materials. *Chemical Society Reviews* **2014**, *43* (24), 8200-8214.
11 9 21. Fujisawa, G.; Van Agthoven, M. A.; Jenet, F.; Rabbito, P. A.; Mullins, O. C., Near-infrared
12 10 compositional analysis of gas and condensate reservoir fluids at elevated pressures and
13 11 temperatures. *Applied Spectroscopy* **2002**, *56* (12), 1615-1620.
14 12 22. Rohwedder, J.; Pasquini, C.; Fortes, P.; Raimundo, I.; Wilk, A.; Mizaikoff, B., iHWG- μ NIR: a
15 13 miniaturised near-infrared gas sensor based on substrate-integrated hollow waveguides coupled to a
16 14 micro-NIR-spectrophotometer. *Analyst* **2014**, *139* (14), 3572-3576.
17 15 23. Dong, C.; O'Keefe, M. D.; Elshahawi, H.; Hashem, M.; Williams, S. M.; Stensland, D.;
18 16 Hegeman, P. S.; Vasques, R. R.; Terabayashi, T.; Mullins, O. C. In *New downhole fluid analyzer tool for*
19 17 *improved reservoir characterization*, Offshore Europe, Society of Petroleum Engineers: 2007.
20 18 24. Kramer, R., *Chemometric techniques for quantitative analysis*. CRC Press: 1998.
21 19 25. Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N., A review of near
22 20 infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of pharmaceutical*
23 21 *and biomedical analysis* **2007**, *44* (3), 683-700.
24 22 26. Wold, S.; Sjöström, M.; Eriksson, L., PLS-regression: a basic tool of chemometrics.
25 23 *Chemometrics and intelligent laboratory systems* **2001**, *58* (2), 109-130.
26 24 27. Abdi, H., Partial least square regression (PLS regression). *Encyclopedia for research methods*
27 25 *for the social sciences* **2003**, 792-795.
28 26 28. Centner, V.; Verdú-Andrés, J.; Walczak, B.; Jouan-Rimbaud, D.; Despagne, F.; Pasti, L.;
29 27 Massart, D.-L.; De Noord, O. E., Comparison of multivariate calibration techniques applied to
30 28 experimental NIR data sets. *Applied spectroscopy* **2000**, *54* (4), 608-623.
31 29 29. Wehrens, R.; Linden, W. v. d., Bootstrapping principal component regression models. *Journal*
32 30 *of Chemometrics* **1997**, *11* (2), 157-171.
33 31 30. Standard, A., E1655 (2005) Standard practices for infrared multivariate quantitative analysis,
34 32 ASTM International, West Conshohocken, PA, 2005. doi: 10.1520/E1655-05R12.
35 33 31. Williams, P. C.; Sobering, D., How do we do it: a brief summary of the methods we use in
36 34 developing near infrared calibrations. *Near infrared spectroscopy: The future waves* **1996**, 185-188.
37 35 32. Rinnan, Å.; van den Berg, F.; Engelsen, S. B., Review of the most common pre-processing
38 36 techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* **2009**, *28* (10), 1201-1222.
39 37 33. Wold, S.; Antti, H.; Lindgren, F.; Öhman, J., Orthogonal signal correction of near-infrared
40 38 spectra. *Chemometrics and Intelligent laboratory systems* **1998**, *44* (1), 175-185.
41 39 34. Spiegelman, C. H.; McShane, M. J.; Goetz, M. J.; Motamedi, M.; Yue, Q. L.; Cote, G. L.,
42 40 Theoretical justification of wavelength selection in PLS calibration: development of a new algorithm.
43 41 *Analytical Chemistry* **1998**, *70* (1), 35-44.
44 42 35. Adedipe, O. E.; Johanningsmeier, S. D.; Truong, V.-D.; Yencho, G. C., Development and
45 43 validation of a near-infrared spectroscopy method for the prediction of acrylamide content in
46 44 French-fried potato. *Journal of agricultural and food chemistry* **2016**, *64* (8), 1850-1860.
47 45 36. Bro, R.; Smilde, A. K., Principal component analysis. *Analytical Methods* **2014**, *6* (9), 2812-
48 46 2831.
49 47 37. Ballabio, D.; Consonni, V., Classification tools in chemistry. Part 1: linear models. PLS-DA.
50 48 *Analytical Methods* **2013**, *5* (16), 3790-3798.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 38. Fernandes, H. L.; Raimundo Jr, I. M.; Pasquini, C.; Rohwedder, J. J., Simultaneous
2 determination of methanol and ethanol in gasoline using NIR spectroscopy: effect of gasoline
3 composition. *Talanta* **2008**, *75* (3), 804-810.

4 39. Lutz, O. M.; Bonn, G. K.; Rode, B. M.; Huck, C. W., Reproducible quantification of ethanol in
5 gasoline via a customized mobile near-infrared spectrometer. *Analytica chimica acta* **2014**, *826*, 61-
6 68.

7

8

9

10

11

12

13

14

15

16

17

18

Table 1. Purity and source of samples used in this study.

Chemical name	Source	Mole fraction purity	Analysis method
Methane	BOC	0.999	GC
Ethane	BOC	0.995	GC
Propane	BOC	0.995	GC
i-butane	BOC	0.995	GC
n-butane	BOC	0.995	GC
Carbon dioxide	BOC	0.999	GC
Nitrogen	BOC	0.999	GC

Table 2. Calibration samples used for the construction of the PLS models, cmol/mol.

No.	Methane		Ethane		Propane		i-butane		n-butane	
	GC	Calculated	GC	Calculated	GC	Calculated	GC	Calculated	GC	Calculated
1	79.89	79.78	12.04	12.11	5.51	5.55	2.01	1.94	0.55	0.62
2	80.02	79.90	11.03	10.98	6.01	6.09	0.97	1.02	1.97	2.01
3	80.00	80.11	10.54	10.51	6.41	6.38	1.07	1.01	1.98	1.99
4	80.11	80.02	9.11	9.09	8.12	8.22	1.15	1.20	1.51	1.47
5	84.05	83.89	11.06	11.11	4.03	4.12	0.45	0.51	0.41	0.37
6	84.01	83.90	10.02	10.09	5.01	5.11	0.49	0.45	0.47	0.45
7	84.0	84.18	9.02	9.08	4.52	4.43	0.41	0.35	2.05	1.96
8	84.03	84.22	7.05	6.96	6.01	6.02	1.99	1.95	0.92	0.85
9	84.05	83.89	11.95	12.05	3.31	3.35	0.38	0.45	0.31	0.26
10	87.05	86.88	8.01	7.99	2.01	2.08	1.41	1.48	1.52	1.57
11	86.50	86.66	5.40	5.39	4.89	4.82	1.11	1.05	2.10	2.08
12	87.05	86.99	6.94	6.96	4.03	4.11	1.53	1.55	0.45	0.39
13	87.11	87.02	7.92	7.85	2.53	2.63	0.84	0.89	1.60	1.61
14	87.06	87.01	10.91	11.05	1.35	1.25	0.35	0.40	0.33	0.29
15	87.05	87.11	5.28	5.18	5.45	5.47	1.10	1.05	1.12	1.19
16	90.11	90.29	5.91	5.96	3.05	2.97	0.41	0.34	0.52	0.44
17	90.05	90.16	5.40	5.30	2.45	2.49	0.89	0.94	1.21	1.11
18	90.01	90.15	7.98	7.89	1.01	1.11	0.43	0.38	0.57	0.47
19	90.02	90.11	7.02	6.89	1.95	2.01	0.46	0.43	0.55	0.56
20	93.07	92.95	5.09	5.18	1.08	1.12	0.28	0.22	0.48	0.53
21	93.04	93.24	4.05	4.01	0.98	0.91	0.89	0.87	1.04	0.97
22	92.98	93.11	4.44	4.40	1.45	1.41	0.61	0.55	0.52	0.53
23	93.01	93.02	3.01	3.06	2.01	1.94	1.01	1.11	0.96	0.87
24	93.0	93.11	6.48	6.39	0.52	0.50	0	0	0	0
25	93.01	93.15	5.90	5.99	0.49	0.39	0.29	0.22	0.31	0.25
26	96.12	96.01	2.12	2.05	1.11	1.19	0.35	0.39	0.30	0.36
27	96.0	96.20	3.05	2.95	0.95	0.85	0	0	0	0
28	98.05	97.84	1.45	1.55	0.50	0.61	0	0	0	0
29	99.38	99.54	0	0	0	0	0.29	0.21	0.33	0.25
30	99.05	98.89	0.43	0.50	0.52	0.61	0	0	0	0
31	100.0	100.0	0	0	0	0	0	0	0	0

Table 3. Performance of the developed PLS models

Components	Temperature (K)	Pressure (MPa)	LV*	RMSECV (cmol/mol)	R ²
Methane	278.15	3.44	2	0.201	0.997
	293.15	3.44	2	0.206	0.996
	313.15	3.44	2	0.205	0.996
	278.15	6.89	2	0.193	0.996
	293.15	6.89	2	0.192	0.996
	313.15	6.89	2	0.191	0.997
	278.15	10.34	2	0.195	0.997
	293.15	10.34	2	0.198	0.997
	313.15	10.34	2	0.191	0.998
	278.15	13.78	2	0.189	0.997
	293.15	13.78	2	0.192	0.998
	313.15	13.78	2	0.198	0.998
Ethane	278.15	3.44	5	0.190	0.996
	293.15	3.44	5	0.181	0.996
	313.15	3.44	5	0.183	0.996
	278.15	6.89	4	0.174	0.997
	293.15	6.89	5	0.186	0.996
	313.15	6.89	4	0.176	0.996
	278.15	10.34	4	0.196	0.996
	293.15	10.34	4	0.166	0.997
	313.15	10.34	4	0.162	0.997
	278.15	13.78	4	0.191	0.996
	293.15	13.78	4	0.173	0.997
	313.15	13.78	4	0.184	0.996
Propane	278.15	3.44	4	0.132	0.997
	293.15	3.44	5	0.151	0.996
	313.15	3.44	5	0.142	0.997
	278.15	6.89	4	0.154	0.996
	293.15	6.89	5	0.123	0.996
	313.15	6.89	4	0.124	0.996

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

		278.15	10.34	5	0.123	0.995
		293.15	10.34	4	0.123	0.995
		313.15	10.34	4	0.116	0.996
		278.15	13.78	4	0.125	0.995
		293.15	13.78	5	0.114	0.996
		313.15	13.78	5	0.142	0.995
		278.15	3.44	4	0.062	0.988
		293.15	3.44	4	0.073	0.983
		313.15	3.44	4	0.075	0.982
		278.15	6.89	4	0.093	0.980
		293.15	6.89	4	0.091	0.979
		313.15	6.89	5	0.117	0.982
	i-butane	278.15	10.34	4	0.084	0.982
		293.15	10.34	4	0.106	0.981
		313.15	10.34	4	0.094	0.981
		278.15	13.78	5	0.101	0.985
		293.15	13.78	5	0.073	0.989
		313.15	13.78	5	0.099	0.981
		278.15	3.44	5	0.082	0.979
		293.15	3.44	5	0.074	0.982
		313.15	3.44	5	0.075	0.981
		278.15	6.89	6	0.073	0.984
		293.15	6.89	5	0.072	0.985
		313.15	6.89	5	0.073	0.985
	n-butane	278.15	10.34	5	0.062	0.989
		293.15	10.34	5	0.074	0.985
		313.15	10.34	5	0.077	0.984
		278.15	13.78	5	0.095	0.982
		293.15	13.78	5	0.078	0.983
		313.15	13.78	6	0.083	0.983

* Latent variables

Table 4. Prediction of the PLS models for methane, ethane, propane, i- butane and n-butane in synthetic gas mixtures (independent samples)

Components	Temperature (K)	Pressure (MPa)	RMSEP (cmol/mol)	R ²	SEP (cmol/mol)	RPD	LoD (cmol/mol)
Methane	278.15	3.44	0.182	0.998	0.185	24.46	0.91
	293.15	3.44	0.202	0.998	0.199	22.74	0.90
	313.15	3.44	0.216	0.998	0.217	20.85	0.90
	278.15	6.89	0.184	0.998	0.188	24.07	0.89
	293.15	6.89	0.190	0.998	0.188	24.07	0.89
	313.15	6.89	0.189	0.998	0.191	23.69	0.88
	278.15	10.34	0.176	0.998	0.179	25.28	0.88
	293.15	10.34	0.184	0.998	0.187	24.20	0.89
	313.15	10.34	0.181	0.998	0.184	24.59	0.89
	278.15	13.78	0.181	0.998	0.187	24.20	0.88
	293.15	13.78	0.191	0.999	0.192	23.57	0.89
	313.15	13.78	0.180	0.998	0.184	24.59	0.88
Ethane	278.15	3.44	0.204	0.994	0.209	13.07	0.38
	293.15	3.44	0.198	0.995	0.201	13.59	0.39
	313.15	3.44	0.201	0.995	0.206	13.26	0.39
	278.15	6.89	0.183	0.997	0.185	14.77	0.39
	293.15	6.89	0.178	0.997	0.180	15.18	0.38
	313.15	6.89	0.166	0.996	0.171	15.98	0.39
	278.15	10.34	0.185	0.997	0.188	14.53	0.38
	293.15	10.34	0.177	0.997	0.178	15.35	0.38
	313.15	10.34	0.166	0.996	0.173	15.79	0.37
	278.15	13.78	0.187	0.997	0.190	14.38	0.38
	293.15	13.78	0.173	0.998	0.174	15.70	0.38
	313.15	13.78	0.164	0.997	0.165	16.56	0.37
Propane	278.15	3.44	0.176	0.992	0.181	11.24	0.33
	293.15	3.44	0.167	0.994	0.173	11.76	0.33
	313.15	3.44	0.172	0.995	0.177	11.49	0.34
	278.15	6.89	0.143	0.995	0.149	13.65	0.33

1								
2								
3		293.15	6.89	0.138	0.996	0.142	14.32	0.33
4		313.15	6.89	0.131	0.996	0.139	14.63	0.33
5								
6		278.15	10.34	0.144	0.995	0.149	13.65	0.32
7		293.15	10.34	0.133	0.996	0.136	14.96	0.33
8		313.15	10.34	0.136	0.996	0.139	14.63	0.33
9								
10		278.15	13.78	0.136	0.995	0.141	14.43	0.32
11		293.15	13.78	0.135	0.995	0.139	14.63	0.33
12		313.15	13.78	0.133	0.995	0.137	14.85	0.32
13								
14								
15								
16		278.15	3.44	0.153	0.924	0.154	2.78	0.14
17		293.15	3.44	0.143	0.934	0.149	2.87	0.14
18		313.15	3.44	0.139	0.935	0.141	3.04	0.13
19								
20		278.15	6.89	0.124	0.949	0.126	3.40	0.13
21		293.15	6.89	0.106	0.951	0.111	3.86	0.14
22		313.15	6.89	0.124	0.934	0.126	3.40	0.14
23	i-butane	278.15	10.34	0.108	0.949	0.109	3.93	0.13
24		293.15	10.34	0.112	0.954	0.119	3.60	0.12
25		313.15	10.34	0.112	0.943	0.112	3.82	0.13
26		278.15	13.78	0.108	0.945	0.108	3.96	0.13
27		293.15	13.78	0.094	0.952	0.098	4.37	0.13
28		313.15	13.78	0.107	0.942	0.108	3.96	0.13
29								
30								
31								
32								
33								
34		278.15	3.44	0.136	0.939	0.141	3.29	0.15
35		293.15	3.44	0.128	0.934	0.129	3.60	0.16
36		313.15	3.44	0.141	0.929	0.142	3.27	0.16
37		278.15	6.89	0.112	0.940	0.114	4.07	0.16
38		293.15	6.89	0.101	0.949	0.102	4.55	0.15
39		313.15	6.89	0.129	0.946	0.133	3.49	0.14
40	n-butane	278.15	10.34	0.117	0.941	0.119	3.90	0.15
41		293.15	10.34	0.106	0.951	0.108	4.30	0.14
42		313.15	10.34	0.103	0.951	0.106	4.38	0.14
43		278.15	13.78	0.126	0.949	0.128	3.63	0.14
44		293.15	13.78	0.094	0.959	0.096	4.83	0.15
45		313.15	13.78	0.117	0.958	0.118	3.93	0.14
46								
47								
48								
49								
50								
51								
52								
53								
54								
55								
56								
57								
58								
59								
60								

Table 5. Comparison of the NIR results at 293.15K and various pressures with GC for one certified gas mixture, in cmol/mol.

Components	GC		FTNIR*							
	Absolute	normalized	3.44 MPa		6.89 MPa		10.34 MPa		13.78 MPa	
			Com.	Dev.	Com.	Dev.	Com.	Dev.	Com.	Dev.
Methane	88.11	91.27	91.13	0.14	91.21	0.06	91.19	0.08	91.19	0.08
Ethane	5.91	6.12	6.05	0.07	6.09	0.03	6.11	0.01	6.12	0.0
Propane	1.92	1.99	2.05	-0.06	1.95	0.04	1.97	0.02	1.96	0.03
i-butane	0.35	0.36	0.42	-0.06	0.43	-0.07	0.42	-0.06	0.40	-0.04
n-butane	0.25	0.26	0.35	-0.09	0.32	-0.06	0.31	-0.05	0.33	-0.07
Carbon dioxide	1.37	0.00	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Nitrogen	2.09	0.00	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

* Com. and Dev. denote composition and deviation, respectively.

Table 6. Comparison of the NIR results at 293.15K and various pressures with GC for an unknown natural gas sample, in cmol/mol.

Components	GC		FTNIR							
	Absolute	normalized	3.44 MPa		6.89 MPa		10.34 MPa		13.78 MPa	
			Com.	Dev.	Com.	Dev.	Com.	Dev.	Com.	Dev.
Methane	90.29	92.53	92.12	0.41	92.04	0.49	92.04	0.49	92.08	0.45
Ethane	5.48	5.61	5.75	-0.14	5.78	-0.17	5.78	-0.17	5.73	-0.12
Propane	1.35	1.37	1.46	-0.09	1.48	-0.11	1.46	-0.09	1.49	-0.12
i-butane	0.20	0.22	0.33	-0.11	0.32	-0.10	0.34	-0.12	0.35	-0.13
n-butane	0.25	0.27	0.34	-0.07	0.38	-0.11	0.38	-0.11	0.35	-0.08
i-pentane	0.05	0.00	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
n-pentane	0.03	0.00	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Carbon dioxide	1.02	0.00	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Nitrogen	1.32	0.00	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

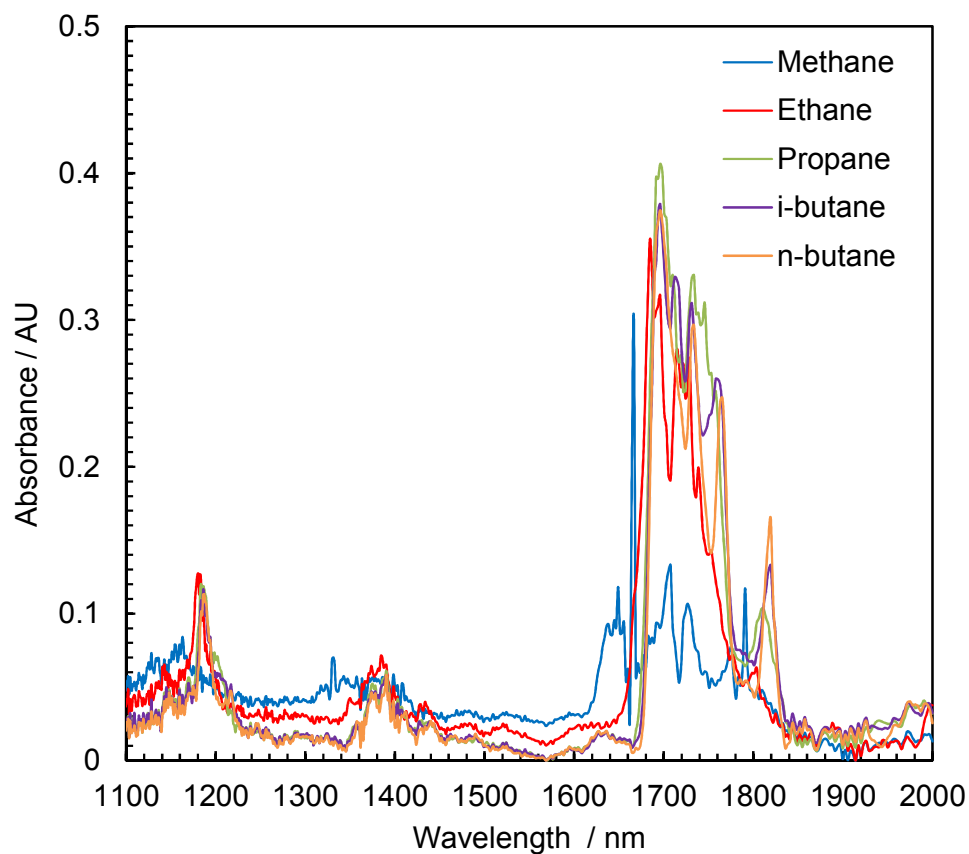


Figure 1. FTNIR Spectra of the pure hydrocarbons at atmospheric pressure and room temperature that captured by the FTNIR spectrometer.

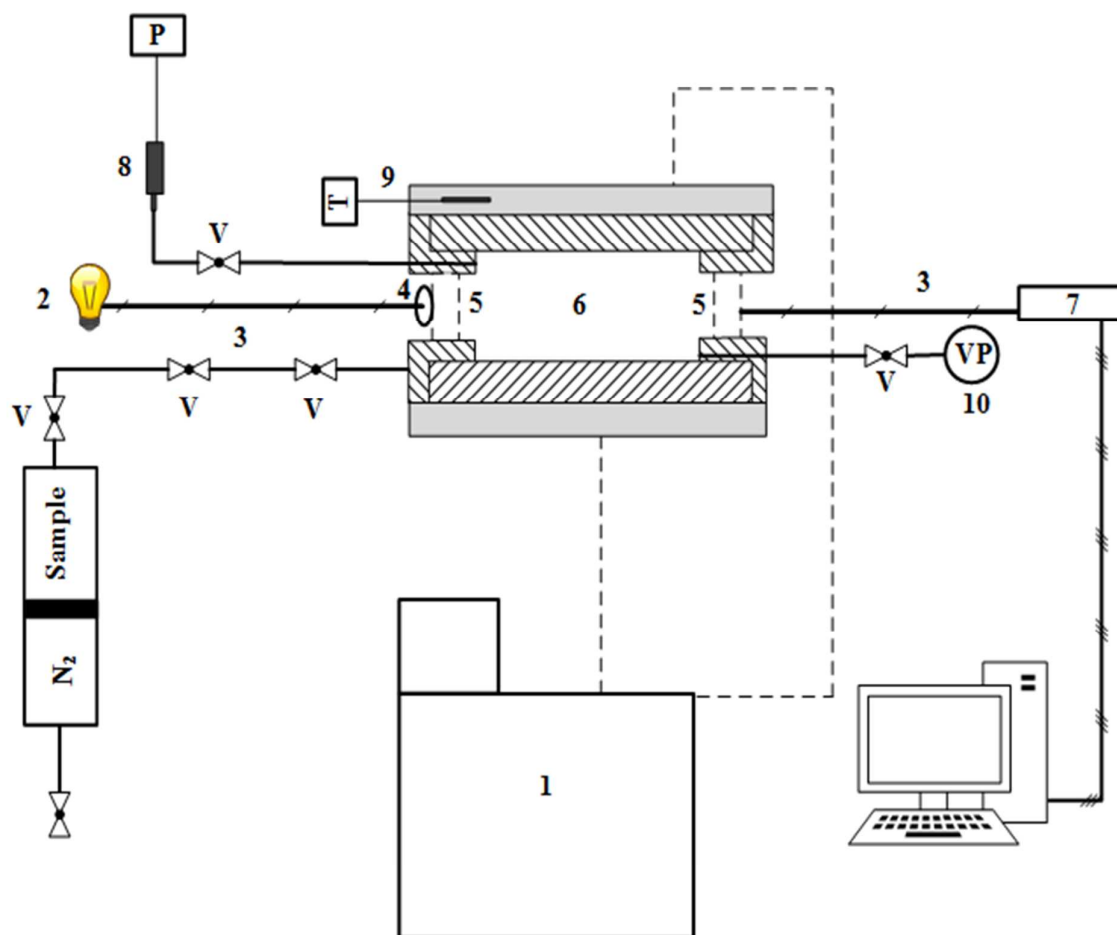


Figure 2. Schematic diagram of the FTNIR setup, 1: Cooling / Heating bath, 2: NIR light source, 3: Fibre optic, 4: Collimating lens, 5: Sapphire windows, 6: High-pressure cell, 7: Spectrometer, 8: Pressure transducer, 9: Temperature probe, 10: Vacuum pump and V: valve.

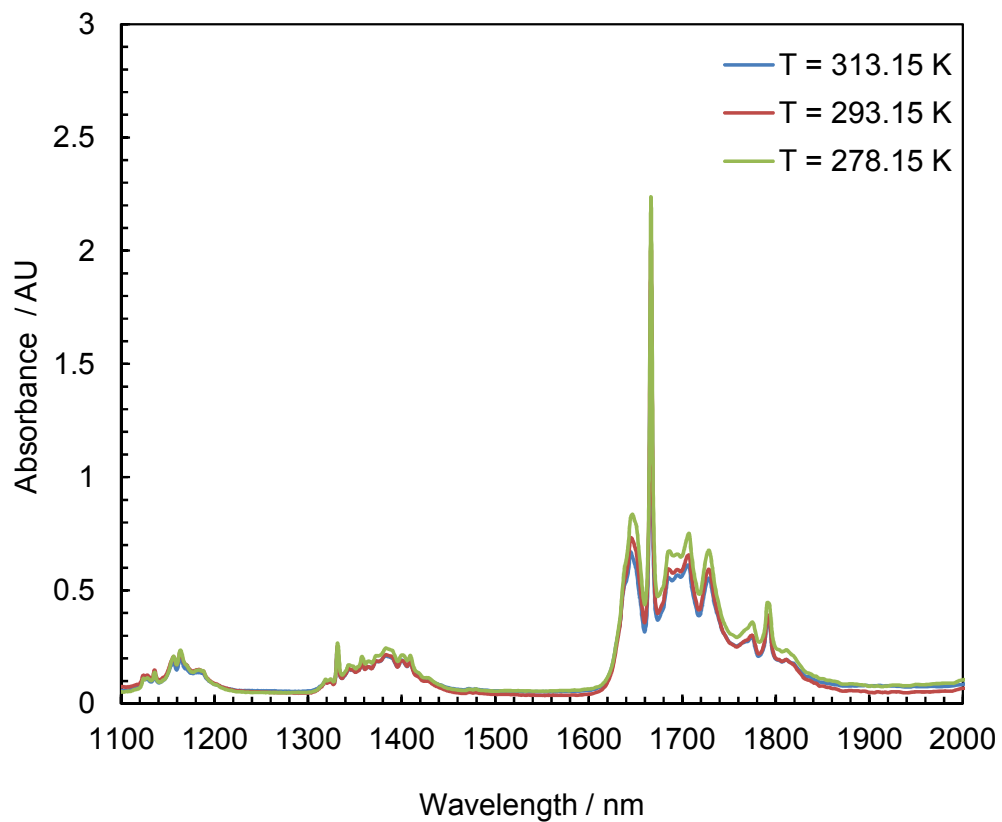


Figure 3. FTNIR Spectra of one synthetic gas mixture at 6.89 MPa and three different temperatures: 278.15 K, 293.15 K, and 313.15 K.

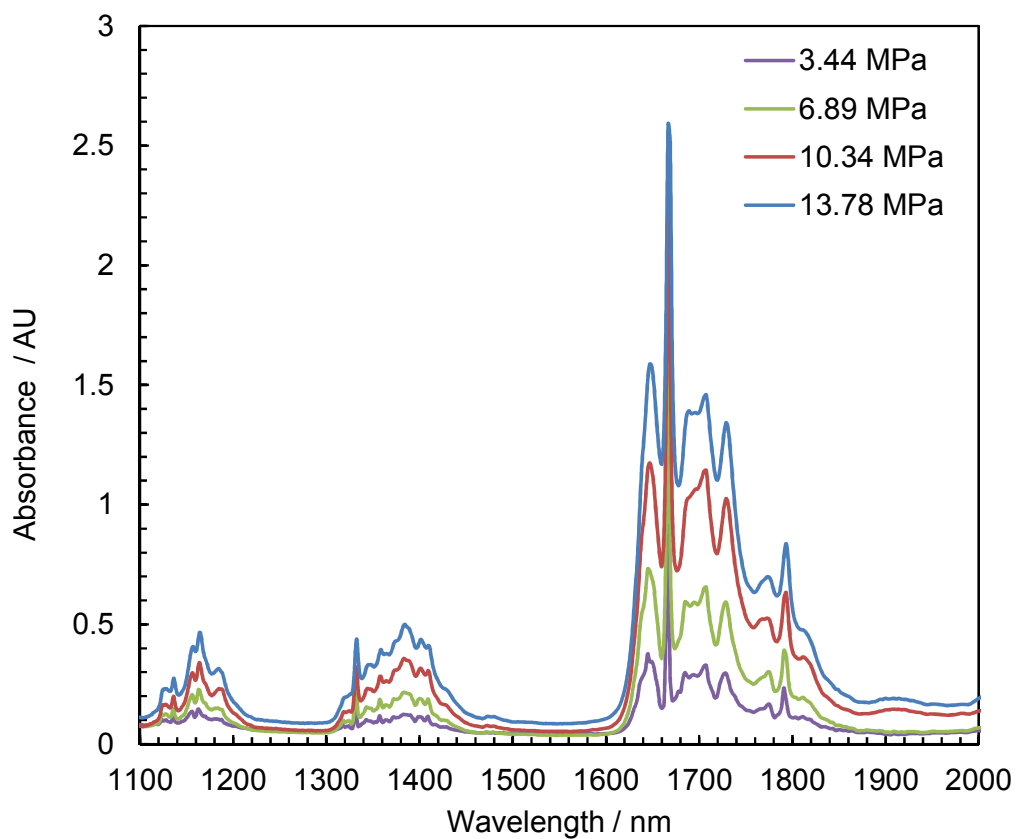


Figure 4. FTNIR Spectra of one synthetic gas mixture at 293.15 K and four different pressures: 3.44 MPa, 6.89 MPa, 10.34 MPa and 13.78 MPa.

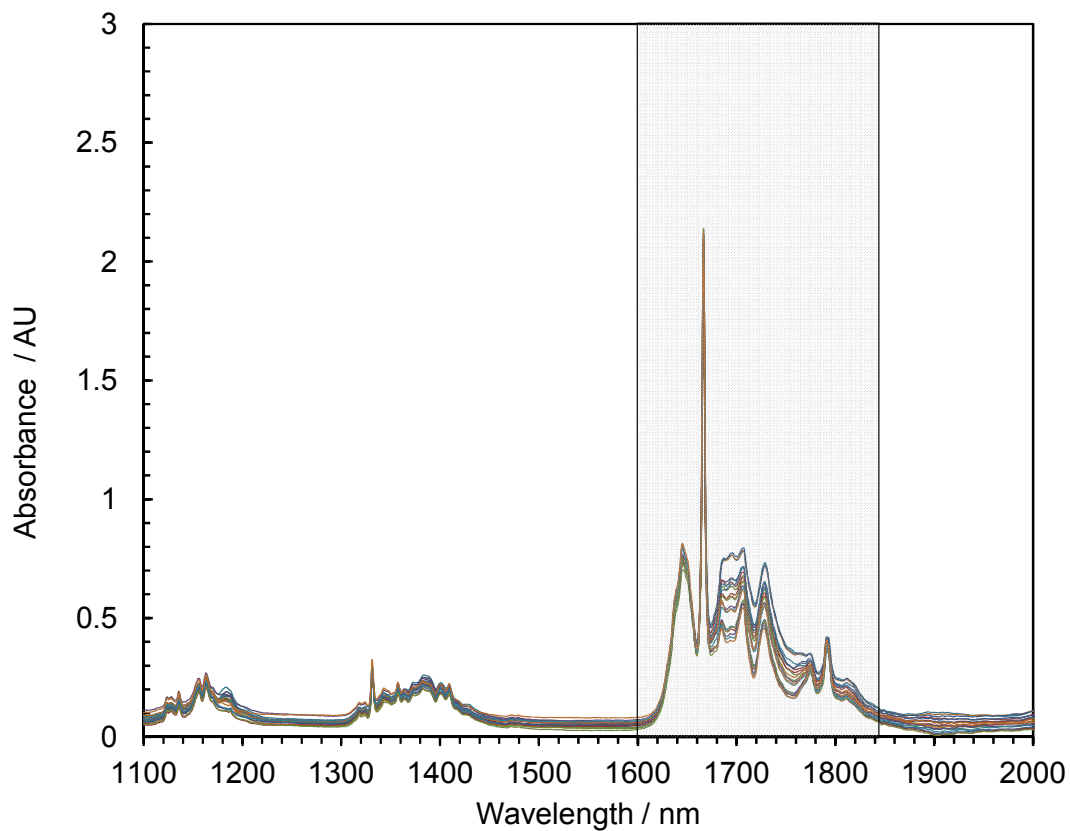


Figure 5. FTNIR Spectra of 30 synthetic gas mixtures with different methane, ethane, propane, i-butane and n-butane contents at pressures of 10.34 MPa and temperature of 293.15 K.

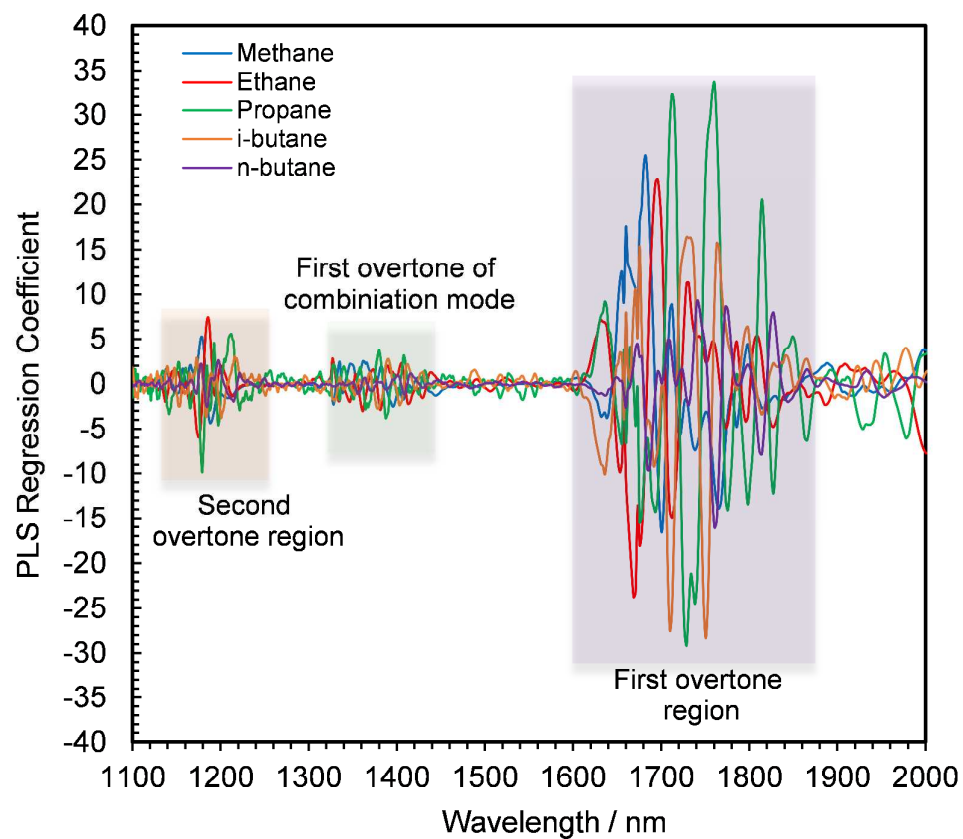


Figure 6. Weighted regression plot for each component.

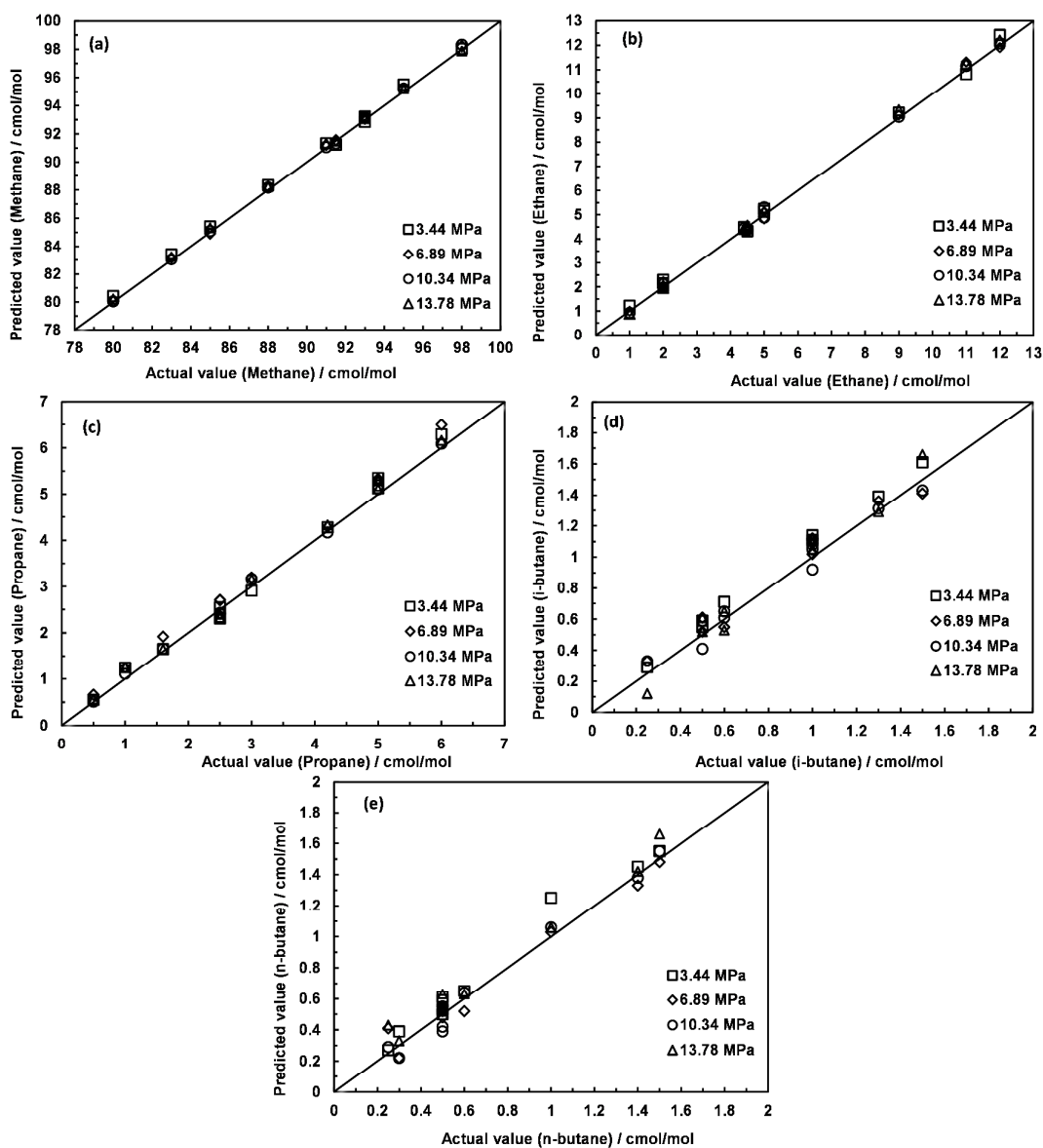


Figure 7. PLS regression plot of predicted versus actual concentration of methane (a), ethane (b), propane (c), i-butane (d) and n-butane (e) in synthetic gas mixtures (independent samples) at temperature of 278.15 K and various pressures (3.44 Δ , 6.89 \square , 10.34 \diamond , and 13.78 \times MPa).

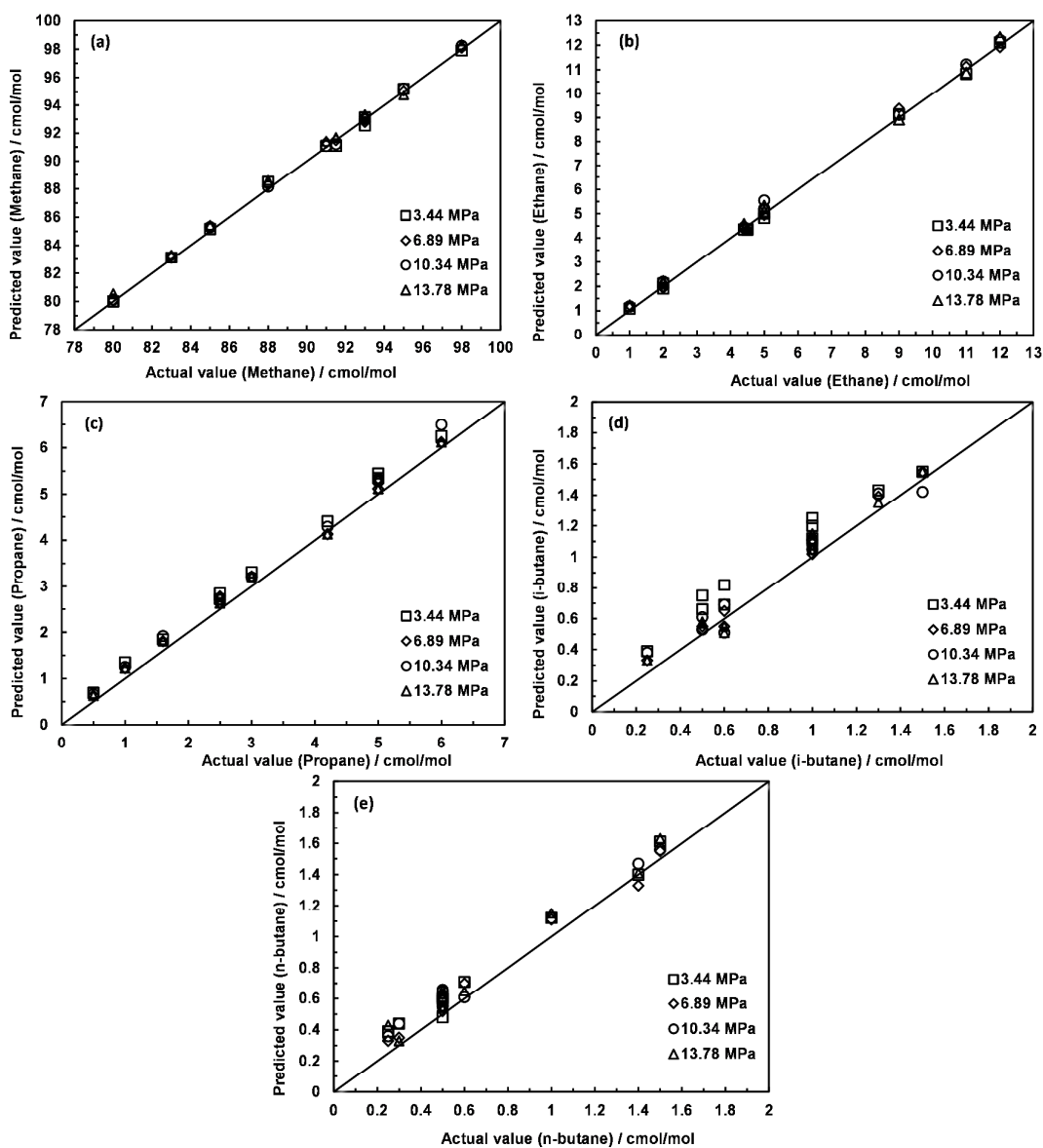


Figure 8. PLS regression plot of predicted versus actual concentration of methane (a), ethane (b), propane (c), i-butane (d) and n-butane (e) in synthetic gas mixtures (independent samples) at temperature of 293.15 K and various pressures (3.44 Δ , 6.89 \square , 10.34 \diamond , and 13.78 \times MPa).

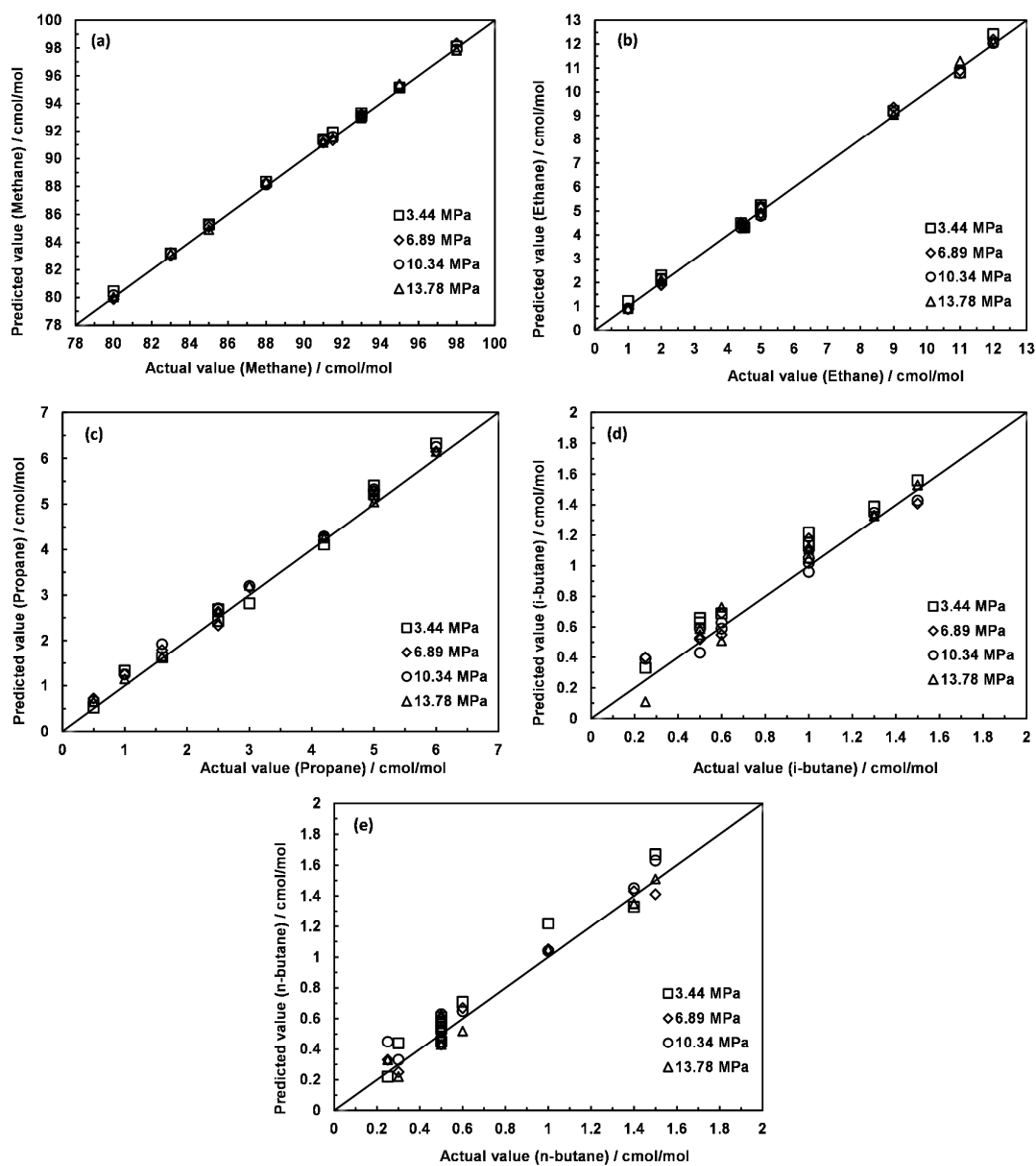


Figure 9. PLS regression plot of predicted versus actual concentration of methane (a), ethane (b), Propane (c), i-butane (d) and n-butane (e) in synthetic gas mixtures (independent samples) at temperature of 313.15 K and various pressures (3.44 Δ , 6.89 \square , 10.34 \diamond , and 13.78 \times MPa).

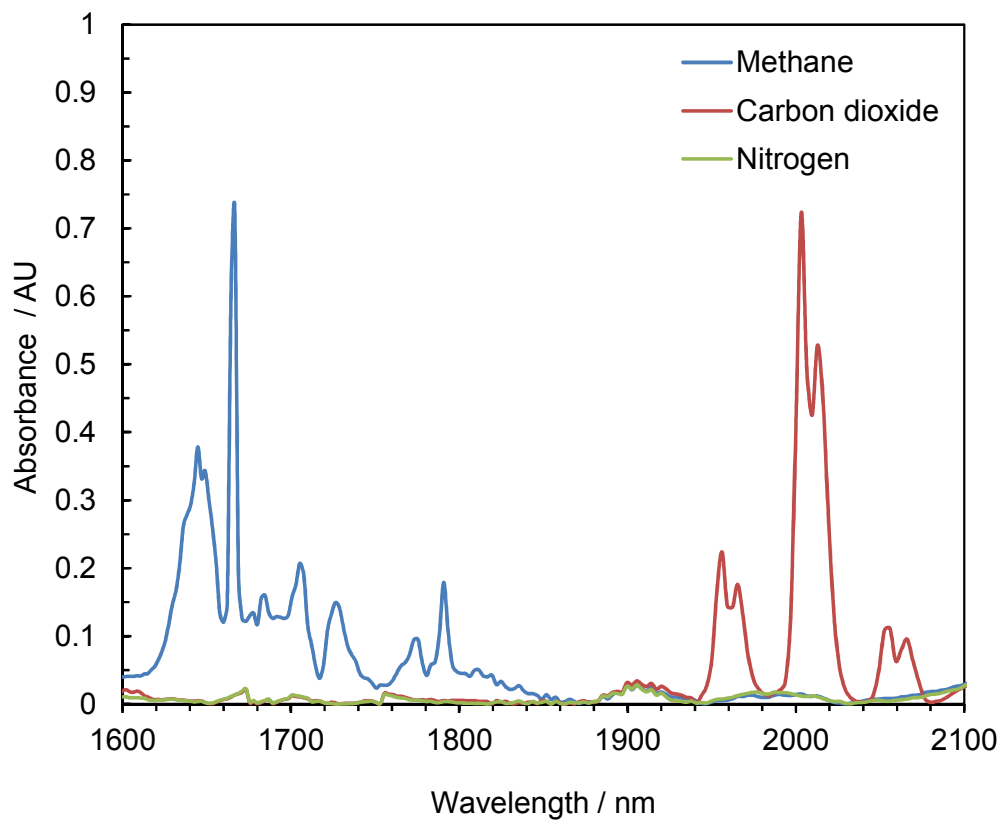


Figure 10. FTNIR spectra of pure carbon dioxide, methane and nitrogen at a pressure of 3.44 MPa and room temperature from FTNIR spectrometer.

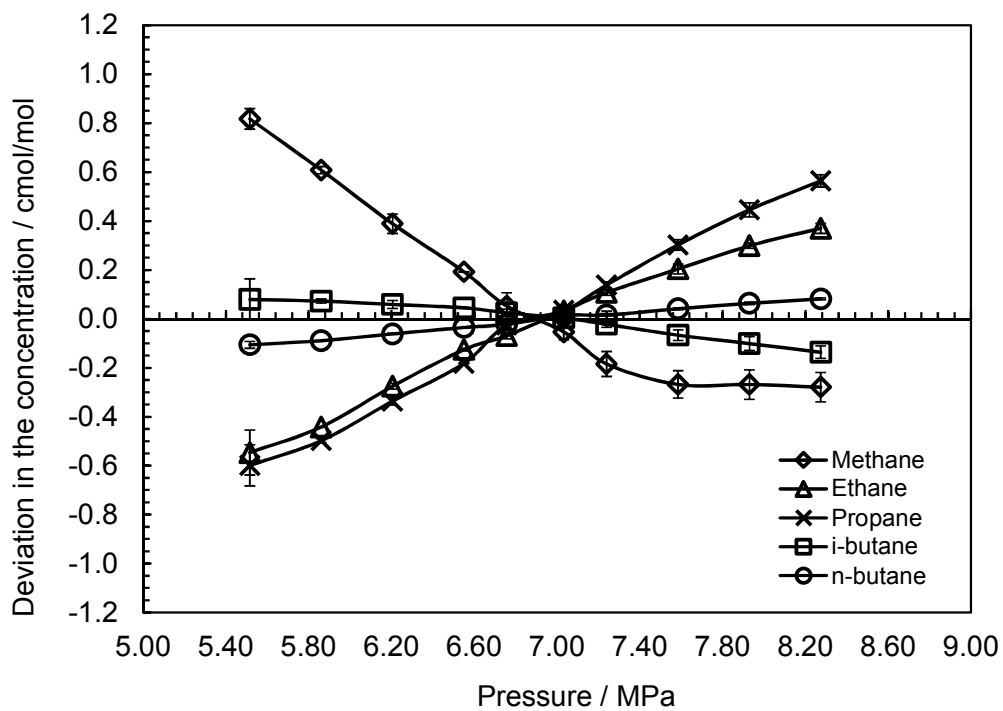


Figure 11. Effect of pressure variation on the FTNIR predicted values on the 293.15 K isotherm. The error bars characterise the standard deviation between three measurements.

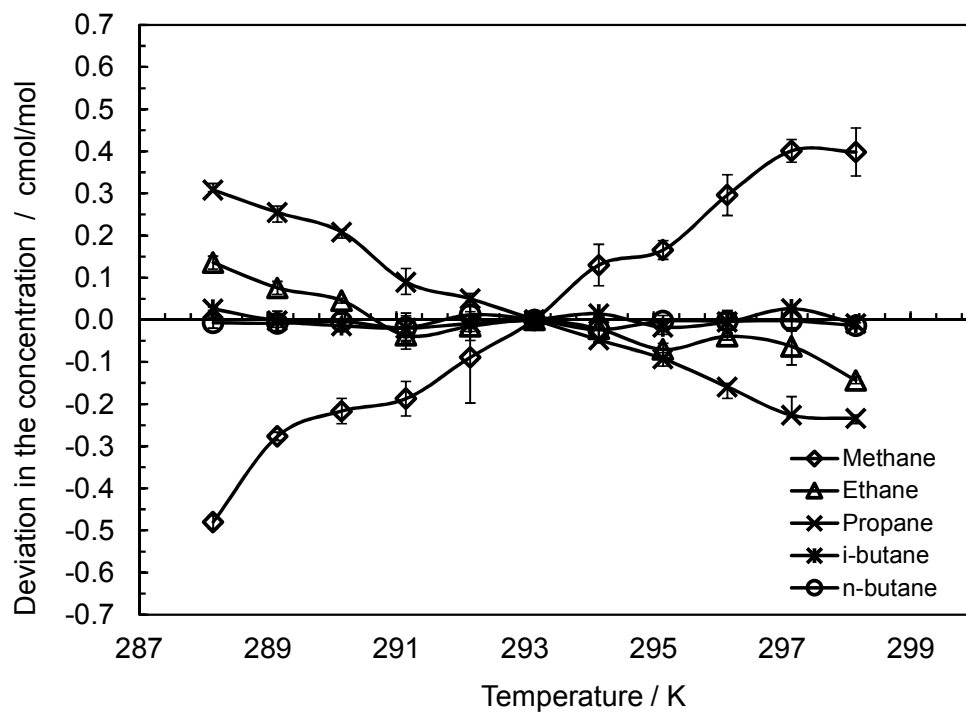


Figure 12. Effect of temperature variation on the FTNIR predicted value at a pressure of 6.89 MPa. The error bars characterise the standard deviation between three measurements