

Analysis of natural gas by gas chromatography Reduction of correlated uncertainties by normalisation

Andrew S. Brown^{a,*}, Martin J.T. Milton^a, Chris J. Cowper^b, Gavin D. Squire^b,
Wolfram Bremser^c, Richard W. Branch^a

^a Analytical Science Group, National Physical Laboratory, Queens Road, Teddington, Middlesex TW11 0LW, UK

^b EffecTech Ltd., Unit 2, Bridge Street Industrial Estate, Trinity Road, Uttoxeter, Staffordshire ST14 8SP, UK

^c Bundesanstalt für Materialforschung und Prüfung, Richard-Willstätter-Strasse 11, 12489 Berlin, Germany

Received 26 January 2004; received in revised form 5 April 2004; accepted 8 April 2004

Abstract

The results of gas chromatographic analysis of natural gas mixtures reveal strong correlations (Pearson correlation coefficient of >0.96) between the uncertainty of each component and variations in the ambient pressure. Although correction for ambient pressure variations can reduce this variability, normalisation of the results of each analysis using the assumption that the sum of all component amount fractions is unity provides significantly greater reductions in the uncertainty of each measured component. We show that the uncertainty in normalised components can be estimated approximately using the correlation coefficient as a measure of the correlation present in the measurements, or exactly using a full calculation of the variance/covariance (V/C) structure of the data.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Uncertainty; Normalisation; Variance; Covariance; Natural gas; Gases; Alkanes

1. Introduction

Natural gas is a highly traded and valuable commodity. Consequently, there is a substantial demand for high quality analysis of the composition of real and synthetic mixtures of natural gas. The major component in all but extreme cases is methane, which is also one of the species of greatest interest, together with nitrogen, carbon dioxide and those hydrocarbons with up to six carbon atoms. Hydrocarbons with 12 or more carbon atoms are routinely found in natural gases, but the concentrations of these species are sufficiently low that they have little effect on most of the important physical properties [1] (e.g. calorific value [2], density, Wobbe index and compression factor). An exception to this is the hydrocarbon dew point, which is highly dependent on the concentration of these heavier hydrocarbon species.

A widely used method for the analysis of natural gas uses gas chromatography (GC) [3–7] with a dual detec-

tion system consisting of thermal conductivity detection (TCD) and flame ionisation detection (FID) [8,9]. TCD is typically used to measure nitrogen, carbon dioxide and lower-molecular-mass hydrocarbons, whereas FID is more suited to the measurement of higher hydrocarbons, whose quantities diminish rapidly with increasing carbon number. To combine the responses from the two detectors, a ‘bridging’ component [10] (typically propane) is selected and measured on both detectors. The ratio of these responses at each injection is then used to ‘scale’ the data from one detector to the other, effectively transforming the analysis into a single operation method.

The GC response is calibrated using a suite of reference gases [11], with certified values for each component. A calibration curve is established for each component from an appropriate statistical fit to the data. In many cases, a working reference gas mixture or some other type of quality control procedure is used to demonstrate the on-going validity of the calibration curve without performing the full calibration procedure.

An important feature of the analysis of natural gas is that all components or groups of components with concentrations

* Corresponding author. Tel.: +44-20-8943-6831;
fax: +44-20-8614-0448.

E-mail address: andrew.brown@npl.co.uk (A.S. Brown).

above the instrumental limit of detection are quantified. It is usual to take advantage of this by introducing the assumption that the complete mixture has been measured as an explicit constraint during the processing of the data. Such methods of analysis are therefore termed ‘complete mixture’ methods. Their use is particularly important when the analysed composition is used to calculate the value of an extrinsic physical property, because it is essential that the component compositions (expressed as amount fractions) sum to unity in order to obtain a valid result. The use of the complete mixture assumption also has a significant influence on the estimated uncertainty of the results. In some routine natural gas analyses, all compounds containing six or more carbon atoms are quantified as a single (‘C₆₊’) peak. This approach can introduce extra uncertainty into the calculation of extrinsic physical properties.

Two methods are in widespread use for reducing measured data to a complete mixture. The first is normalisation, which involves scaling the sum of all the measured components to unity. For a mixture with q components ($w = 1, \dots, q$) the normalised amount fraction of component i is calculated using:

$$x_i = \frac{x_i^*}{\sum_{w=1}^q x_w^*}, \quad (1)$$

where x_i^* is the amount fraction of component i before normalisation.

The second is the component-by-difference method, in which all of the components apart from one (usually the major component, methane) are measured. The major component is then calculated by:

$$x'_1 = 1 - \sum_{i=2}^q x_i^*, \quad (2)$$

and all other components remain unchanged, i.e.

$$x'_i = x_i^* \quad \text{when } i \neq 1. \quad (3)$$

In this paper, we investigate the uncertainties in the analysis of a set of synthetic natural gases by GC. This shows that errors in natural gas analysis are dominated by variations in the size of the sample measured by GC. These variations may be caused by a number of factors, such as changes in the injection volume, atmospheric pressure, atmospheric temperature (although thermal control of the GC sample loop may reduce these) and detector drift. We hypothesise that these changes cause the uncertainty of the measurement of an ‘unknown’ natural gas mixture to be substantially over-estimated because they are largely correlated between all measured components. We then compare the performance of the normalisation and component-by-difference methods by application to these experimental data. Following an analysis of correlations in the data, we apply the two methods (normalisation and component-by-difference) in different ways. The study quantitatively assesses how the uncertainty

Table 1

Range of amount fractions contained in the seven reference gases studied in this work

Analyte	Amount fraction (mmol/mol)	
	Minimum	Maximum
Nitrogen	2.1	117
Carbon dioxide	0.51	78.8
Methane	648	983
Ethane	1.0	136
Propane	0.47	78.1
Isobutane	0.073	11.8
<i>n</i> -Butane	0.11	11.7
Neopentane	0.043	3.5
Isopentane	0.043	3.5
<i>n</i> -Pentane	0.054	3.4
<i>n</i> -Hexane	0.045	3.4

of the measurement is affected by the application of pressure correction (to remove the effect of atmospheric pressure variations) and the subsequent normalisation procedures (to remove the effect of other ‘sample size’ variations).

2. Experimental

A series of seven reference gas mixtures, each of which contained 11 components typically found in natural gas (nitrogen, carbon dioxide, methane, ethane, propane, isobutane, *n*-butane, neopentane, isopentane, *n*-pentane and *n*-hexane)¹ with known composition and uncertainty, were measured by gas chromatography. The gas mixtures were prepared gravimetrically in internally-passivated aluminium cylinders; quality assurance checks have shown that natural gas mixtures are stable in such cylinders for many years. The range of compositions for each component used in this work is shown in Table 1. For each component the seven amount fractions were approximately equally distributed amongst the reference gases.

All measurements were performed on a Daniel Model 500 process gas chromatograph system (Daniel, UK) using a TCD system in accordance with ISO 6974-5:2001 [12]. The method used a column switching/backflush arrangement (a helium-actuated heated six-port diaphragm valve) whereby the sample was initially injected onto a packed boiling point column divided into short and long sections. The long section (25% SF-96 on 80–100 mesh Chromosorb P, 7 ft. length × 1/16 in. external diameter; 1 ft. = 30.48 cm, 1 in. = 2.54 cm) provided separation of C₃ to C₅ hydrocarbons, while C₆ and heavier hydrocarbons (when present as impurities in the reference gases) were retained on the short section (20% oxydipropionitrile on 80–100 mesh Porasil C,

¹ The nomenclature used throughout this paper is consistent with that used in the natural gas industry. The IUPAC-recommended nomenclature for isobutane, *n*-butane, neopentane, isopentane, *n*-pentane and *n*-hexane are, respectively: 2-methylpropane, butane, 2,2-dimethylpropane, 2-methylbutane, pentane and hexane.

12 in. length \times 1/16 in. external diameter) from which they were backflushed to the detector as a single peak. Methane, ethane, nitrogen and carbon dioxide passed rapidly and unresolved through the boiling point columns onto a packed porous polymer column (80–100 mesh HayeSep N, 7 ft. length \times 1/16 in. external diameter) for separation. The GC system operated isothermally (70–80 °C), using helium carrier gas (99.999%, 6 bar column head pressure) and a total measurement cycle time of 240 s run-to-run.

The seven reference gases were attached to an autosampler and analysed in turn. A single representative, fully purged, measurement was taken of each reference gas before proceeding to measure the next. After a single measurement was obtained for all seven reference gases, the entire procedure was repeated 10 times such that 10 independent measurements were taken on each gas mixture. Hence, a response data set collected within a single daily period consisted of instrument responses for 10 independent measurements of 11 components contained in seven reference gases. Hence, for each component, 77 measured responses were recorded. Similarly, 110 measurements were obtained per reference gas over a total of 70 instrument cycles. The calibration experiment described here was performed on five separate 24 h periods to give five daily data sets. For simplicity, these are referred to as Days A (3–4 January), B (4–5 January), C (9–10 January), D (4–5 March) and E (7–8 March).

The gas chromatograph was operated in a stopped flow injection mode, the flow through the sample loop being stopped immediately prior to injection. The vent side of the sample loop was open to atmospheric pressure in order to ensure there were no sample flow effects on the amount of sample injected onto the columns. To enable pressure corrections to be applied to the data, the ambient pressure was logged at the time of each injection. Any variation in atmospheric pressure alters the physical size of the sample by changing the number of molecules of each component in the sample loop of the GC system.

3. Analysis of data

3.1. Statistics of pressure correlation

The extent to which the results of GC analysis are correlated with the atmospheric pressure is of great importance in selecting the best method for processing them. The correlation between two variables, x_i and y_i can be quantified by evaluating the Pearson correlation coefficient r [13]:

$$r^2 = \frac{[\sum(x_i - \bar{x})(y_i - \bar{y})]^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2} \quad (4)$$

The Pearson correlation coefficient varies from 1 (complete correlation) through zero (no correlation) to -1 (complete anti-correlation) and measures the proportion of the total

Table 2

Pearson correlation coefficients between atmospheric pressure and the results for each analyte in cylinder number 860 on Day D

Analyte	Pearson correlation coefficient
Nitrogen	0.992
Carbon dioxide	0.973
Methane	0.997
Ethane	0.997
Propane	0.997
Isobutane	0.722
<i>n</i> -Butane	0.966
Neopentane	0.442
Isopentane	0.512
<i>n</i> -Pentane	0.967
<i>n</i> -Hexane	0.994
Total	0.997

variance (s_{total}^2) that is accounted for by the correlation

$$s_{\text{predicted}}^2 = r^2 s_{\text{total}}^2 \quad (5)$$

Since the total variance can be expressed as the sum of a correlated component ($s_{\text{predicted}}^2$) and a random component (s_{random}^2):

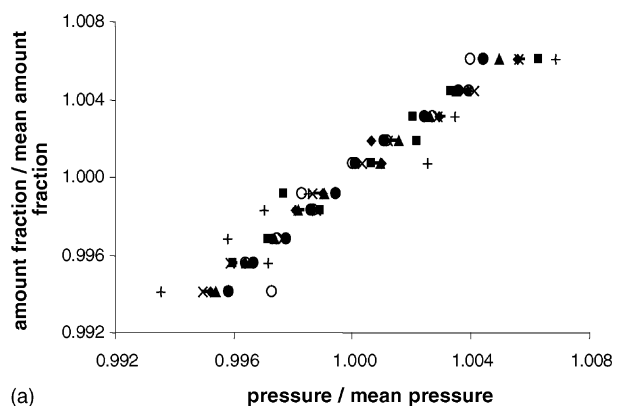
$$s_{\text{total}}^2 = s_{\text{predicted}}^2 + s_{\text{random}}^2 \quad (6)$$

it follows that:

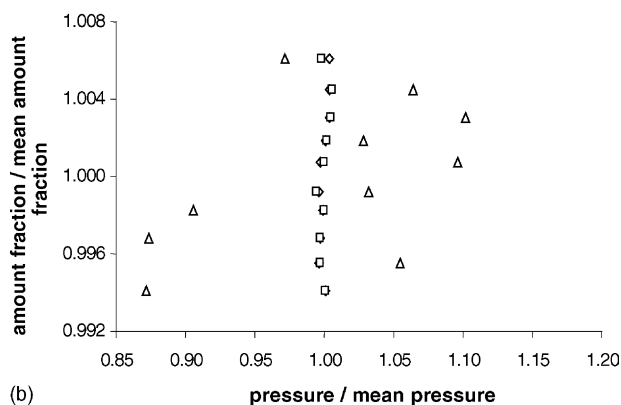
$$s_{\text{random}}^2 = s_{\text{total}}^2 (1 - r^2) \quad (7)$$

We present an analysis of the results obtained on Day D, during which the variation in atmospheric pressure was approximately 13 mbar. The Pearson correlation coefficient between the data for each component in one reference gas mixture (cylinder number 860) and the recorded pressure was calculated and the results are shown in Table 2. It can be seen that the correlation is greater than 0.96 for all components except isobutane, isopentane and neopentane. Fig. 1a and b show the replicate measurements (normalised to the daily mean) for each component plotted against the recorded pressure. Fig. 1c is an expansion of Fig. 1b and illustrates the behaviour of isobutane and neopentane, which is significantly different from the other components.

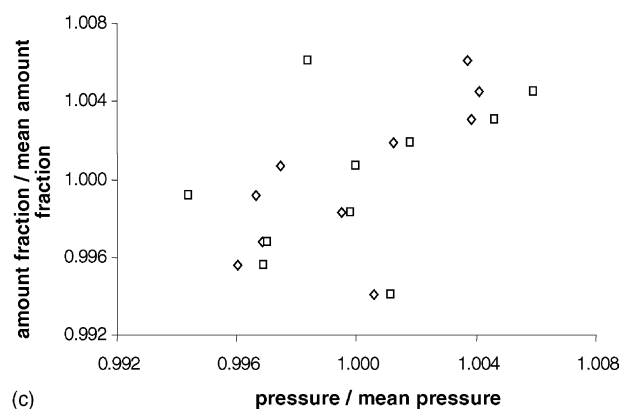
Two of the least strongly correlated components, isobutane and isopentane have the lowest amount fractions in this reference gas mixture. At these low amount fractions, random variations in the data caused by detector noise dominate the variation due to pressure, resulting in a much smaller value for the Pearson coefficient. This conclusion is confirmed by Fig. 2, which plots the Pearson coefficient against the recorded atmospheric pressure for all 11 components in each of 8 gas mixtures (the seven outlined above together with another in the same concentration range) yielding a total of 88 data points. A very high degree of correlation is observed for all components with an amount fraction greater than 0.01 mol/mol, confirming



(a)



(b)



(c)

Fig. 1. Measured amount fraction on Day D normalised to the mean for Day D vs. atmospheric pressure for: (a) nitrogen (\blacklozenge), carbon dioxide (\blacksquare), methane (\blacktriangle), ethane (\blacksquare), propane (\bullet), *n*-butane (\circ), *n*-pentane ($+$) and *n*-hexane (\times); (b) isobutane (\diamond), neopentane (\square) and isopentane (Δ); and (c) isobutane (\diamond) and neopentane (\square) plotted on an expanded y-axis. All data are from analyses of cylinder 860.

that the correlation with atmospheric pressure decreases substantially when random noise shows a relative increase.

The other compound exhibiting poor correlation is neopentane, which has a relatively large amount fraction of 2.3 mmol/mol, but is partially co-eluted with *n*-butane in the chromatogram. The extent of this co-elution varies with the relative peak size (and therefore amount fraction) of each component and is likely to be the dominant source of variation.

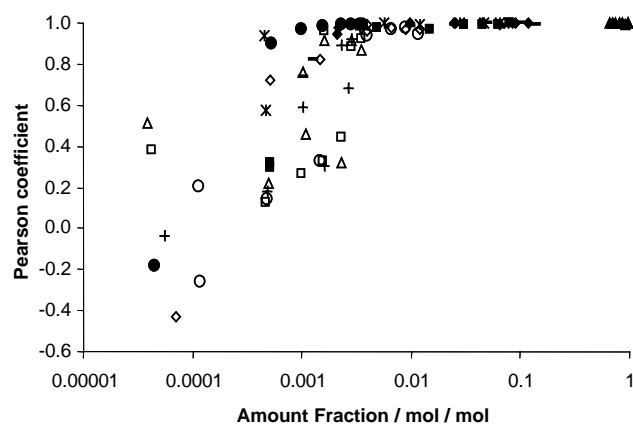


Fig. 2. Pearson coefficient for all components in all reference gases vs. amount fraction on Day D. The symbols refer to the same components as in Fig. 1.

3.2. Standard deviation of results

The results in Figs. 1a and 2 suggest that variations in the atmospheric pressure cause substantial variations in the sample size. As discussed in the introduction, the method of normalisation can be used to reduce these variations.

The influence of normalisation on the variability in the measured data can be analysed by evaluating the standard deviation of 10 consecutive runs acquired on each day. Fig. 3a shows the standard deviation of the measured amount fraction of each component in each reference gas mixture plotted against the amount fraction. Fig. 3b shows the same data after correction for the atmospheric pressure at the time of injection. It can be seen that the pressure correction reduces the standard deviation by a factor of between 2 and 5 for all components with an amount fraction greater than 5 mmol/mol. However, it leads to a small increase in variability at lower amount fractions.

Normalisation of this data (on a run-by-run basis) according to Eq. (1) leads to a further reduction in the correlation (Fig. 3c) of between 1.5 and 3 for all components with amount fraction greater than 5 mmol/mol. In the case of methane measurements, the reduction is up to a factor of 100, giving a total reduction of over 400 between Fig. 3a and c. The normalisation process therefore reduces correlation to a greater extent than atmospheric pressure correction, which suggests that pressure variation, whilst accounting for much of the correlation, is not the only mechanism that gives rise to correlated variability in the data during the day. It is noticeable from the slight trend in Fig. 3c that even after normalisation, some correlation between the standard deviation and amount fraction is still present.

The analysis presented in Figs. 1–3 is all based on results from Day D. In the following section, we show how the normalisation procedure performs in reducing variability between days.

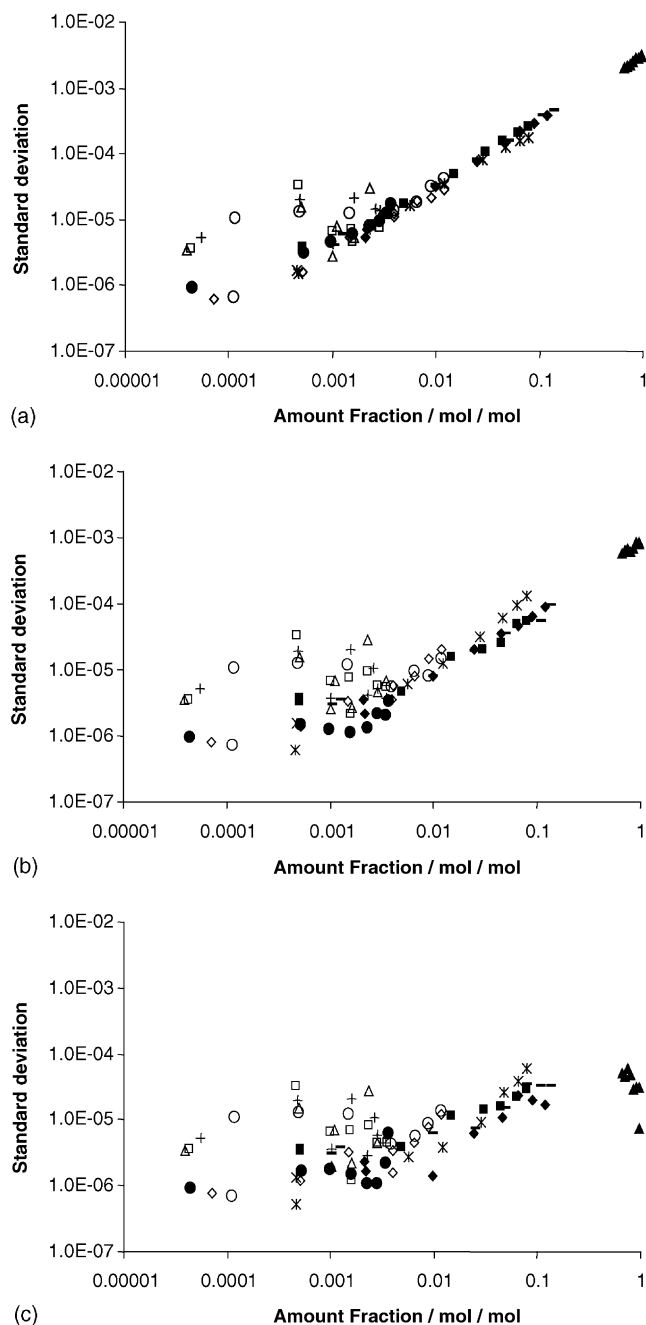


Fig. 3. Standard deviation of the amount fraction (a) before pressure correction; (b) after pressure correction; and (c) after pressure correction and normalisation vs. amount fraction for all components in all reference gas mixtures on Day D. The symbols refer to the same components as in Fig. 1.

3.3. Normalisation (run-to-run and day-to-day)

The normalisation procedure used in the previous section was implemented by applying Eq. (1) independently to each run. We refer to this as normalisation on a run-by-run basis. In this section, we determine whether this is the optimum way for normalisation to be employed. In order to do this, the three normalisation procedures outlined schematically in

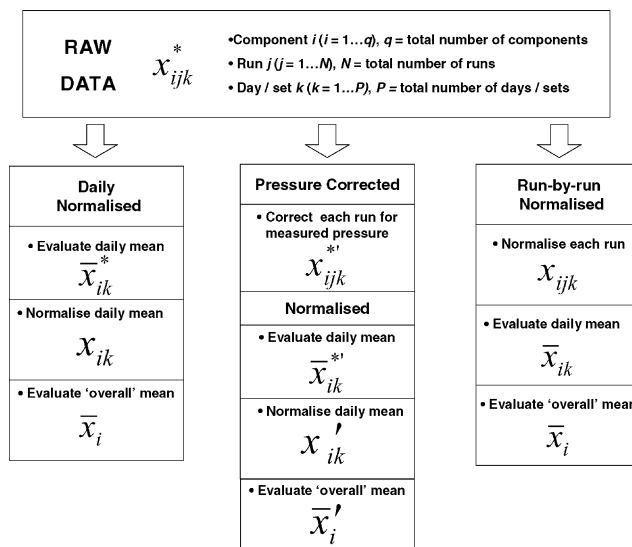


Fig. 4. The three procedures used for data processing. The mathematical operations are outlined in greater detail in Appendix A.

Fig. 4 (and expressed in further detail in Appendix A) were compared:

- Daily normalisation: The 10 daily analyses for each component were averaged to form a daily mean. The daily mean values for each component were then normalised. The result was then calculated from the mean of these.
- Pressure correction and daily normalisation: Each raw data point was pressure corrected, before following the same procedure as for daily normalisation ((i), above).
- Run-by-run normalisation: Each of the 10 daily runs were normalised independently, then averaged to calculate the daily mean. The result was then calculated by averaging the five daily mean results.

Since a large set of data was available for this study, it has been possible to compare the performance of these procedures on the basis of the standard deviation of repeated application to the data from the five different days. The results for the methane component in one of the reference gas mixtures (cylinder number 809) are typical of all of the mixtures and are described in detail here. In order to provide a visual representation of the inter-day and intra-day variability, the results for methane on each day of analysis are represented in Fig. 5a–e using the equation:

$$b = \exp - \left(\frac{x - \bar{x}_k}{2s_k} \right)^2 \quad (8)$$

where \bar{x}_k is the mean methane amount fraction measured on Day k and s_k is its standard deviation. This equation is used to provide a convenient means of displaying the result graphically and is not intended to imply that the data are normally distributed.

Fig. 5a shows the distribution of the measured methane amount fractions over the 5 days of analysis. A wide

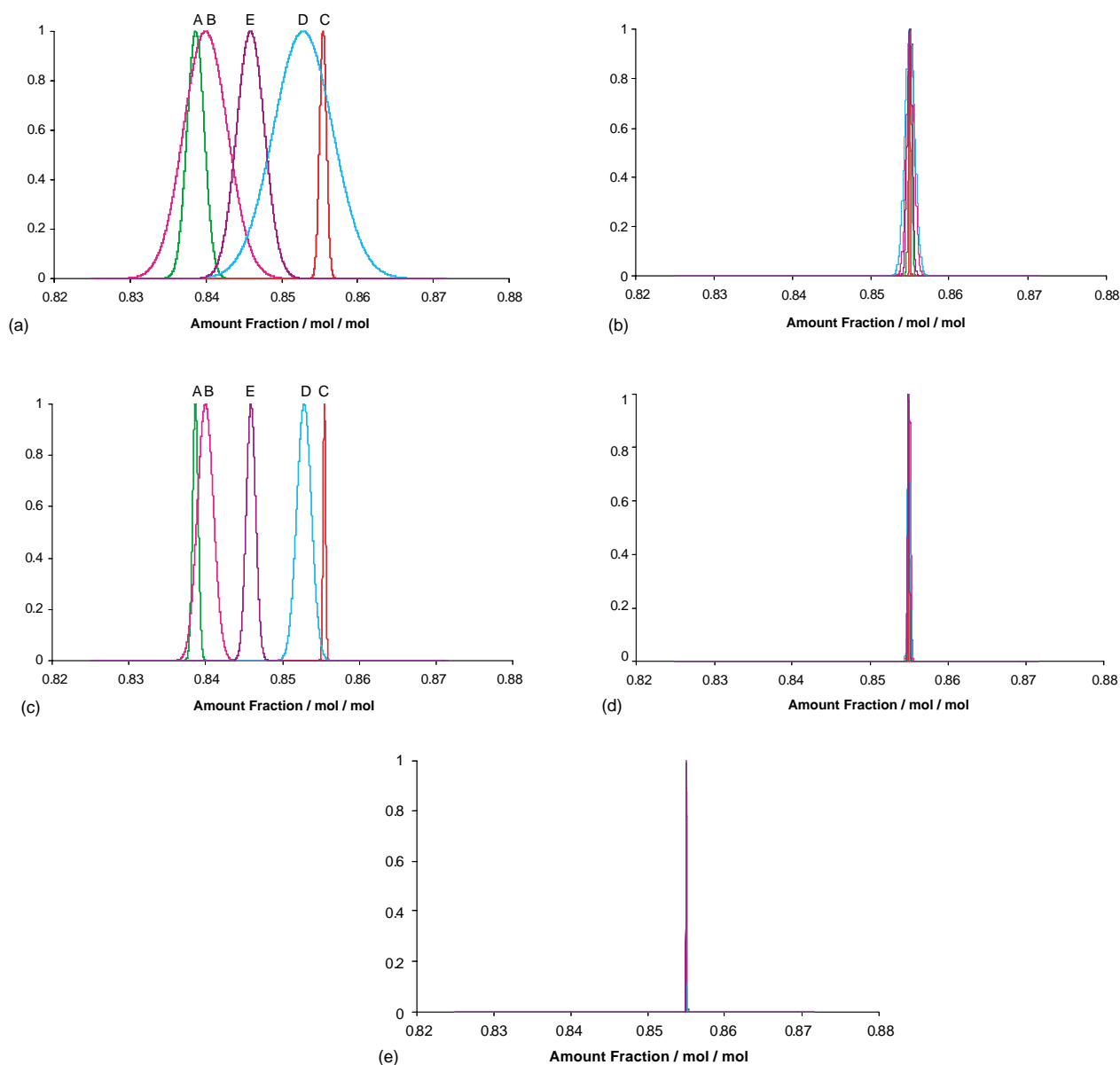


Fig. 5. Distribution of methane amount fractions (for cylinder 809). Labels correspond to data collected over the following days: (A) 3–4 January, (B) 4–5 January, (C) 9–10 January, (D) 4–5 March, (E) 7–8 March. Data shown: (a) as raw data; (b) after daily normalisation (order of peaks from broadest to narrowest: D, B, E, A, C); (c) after pressure correction; (d) after pressure correction and normalisation and; (e) after run-by-run normalisation.

variation in the data can be seen in the daily mean amount fractions, which range from 0.8387 to 0.8555 mol/mol, corresponding to a spread of approximately 2% relative to the mean value. There is also a wide variation in the daily standard deviations, the largest and smallest being recorded on Days D and C, respectively, with the greatest and least variations in atmospheric pressure, once again demonstrating the significant influence of pressure variations on the reproducibility of GC data. These raw data were then treated by each of the three normalisation procedures.

The results of normalising the data on a daily basis according to Procedure (i) are plotted in Fig. 5b. This procedure has the effect of reducing the variability dramatically;

the daily amount fractions now span a range that is approximately 180 times smaller than Fig. 5a. The daily standard deviations also decrease, indicating a significant reduction in the sample size-effect. However, the order of the daily standard deviations from largest to smallest (D, B, E, A, C) remains the same, suggesting that there is some variability in the data that normalisation on a daily timescale does not remove.

The application of pressure correction according to the first step in Procedure (ii) also improves the raw data further, as shown in Fig. 5c. Pressure correction applied to the mean value on each day has little effect on the daily amount fractions, but reduces the daily standard deviations. The

Table 3
Comparison of normalisation procedures for methane in cylinder 809

Procedure	\bar{x} (mol/mol)	$u(\bar{x})$ (mol/mol)
Raw data	0.84656	0.00076
Day-by-day normalisation	0.85499	0.00012
Pressure correction and normalisation	0.85499	0.00004
Run-by-run normalisation	0.85499	0.00001

Uncertainties are calculated by use of Eq. (9).

subsequent normalisation of this pressure-corrected data (Fig. 5d) produces a larger change, reducing the plotted data to a series of daily standard deviations with less variability than the results on any single day when processed by Procedure (i).

The use of run-by-run normalisation according to Procedure (iii) shows the largest effect. The results of using this procedure are shown in Fig. 5e. The plotted data are reduced to a series of curves with very similar daily mean values and extremely small daily standard deviations. This procedure is expected to perform well with data of this type and is most effective in reducing correlated variations in the data.

The effect of each of the procedures is summarised in Table 3, which gives the mean amount fraction of the five daily runs (\bar{x}) together with its associated uncertainty. This uncertainty ($u(\bar{x})$) was calculated using:

$$u(\bar{x}) = \sqrt{\frac{\sum_k u(x_k)^2}{N^2}} \quad (9)$$

where $u(x_k)$ is the uncertainty in the measured amount fraction on Day k and N the number of sets of data (five in this case, one for each day of analysis). Table 3 shows that the values for the ‘mean amount fraction’ for each of the three normalisation procedures are the same. These three values differ from the raw data because the total measured amount fraction was not equal to unity in any of the repeated measurements.

The application of the normalisation approach to other components produces similar but less dramatic results. The results obtained for ethane from the same gas mixture (cylinder number 809) are shown in Table 4. Ethane is present in this reference gas at an amount fraction approximately 100 times smaller than that of methane. The results follow the same pattern as those for methane (Table 3) but the reduction in the standard deviation of the mean values is much smaller, the run-by-run value reducing to 31% of the raw

Table 4
Comparison of normalisation procedures for ethane in cylinder 809

Procedure	\bar{x} (mmol/mol)	$u(\bar{x})$ (mmol/mol)
Raw data	9.65327	0.00961
Day-by-day normalisation	9.74929	0.01216
Pressure correction and normalisation	9.74929	0.00398
Run-by-run normalisation	9.74929	0.00300

Uncertainties are calculated by use of Eq. (9).

Table 5
Comparison of normalisation procedures for isopentane in cylinder 860

Procedure	\bar{x} ($\mu\text{mol/mol}$)	$u(\bar{x})$ ($\mu\text{mol/mol}$)
Raw data	40.7542	1.8331
Day-by-day normalisation	41.0845	1.8497
Pressure correction and normalisation	41.0839	1.8461
Run-by-run normalisation	41.0838	1.8470

Uncertainties are calculated by use of Eq. (9).

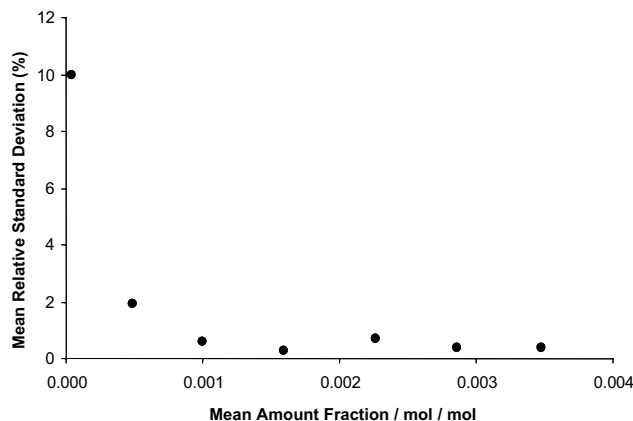


Fig. 6. Relative standard deviation vs. amount fraction for isopentane. The plotted values are the standard deviations and amount fractions averaged over the 5 days of analysis.

data value for ethane, compared to only 1.7% in the case of methane. This pattern can also be seen to continue in the results for isopentane in cylinder 860 (Table 5), which has an amount fraction approximately one hundred times smaller than that of ethane in cylinder 809. In this case, no significant improvement is seen with normalisation because at the low mole fraction of isopentane in this cylinder, random variations are much more significant than sample size-effects (as shown in Fig. 1b). This is further illustrated by comparing the standard deviation of the 10 repeat measurements of isopentane in each of the seven cylinders (Fig. 6). Here, the general trend towards increasing relative standard deviation with decreasing amount fraction can be observed.

3.4. Methane-by-difference

The methane dataset was also analysed by the methane-by-difference method in order to establish a comparison. The results are shown in Table 6. The methane-by-difference

Table 6
Comparison of methane-by-difference (MBD) procedures for methane in cylinder 809

Procedure	\bar{x} (mol/mol)	$u(\bar{x})$ (mol/mol)
Raw data	0.84656	0.00076
Day-by-day MBD	0.85642	0.00015
Pressure correction and MBD	0.85642	0.00005
Run-by-run MBD	0.85642	0.00030

Uncertainties are calculated by use of Eq. (9).

method is not well suited to processing data with very strong correlations and it therefore gives results with different characteristics from those produced by normalisation. In particular, the general improvement of analytical data observed during normalisation is not observed when using the methane-by-difference method. In all cases, the standard deviation of the mean is reduced from that of the raw data, however, it is the daily calculation (with or without prior pressure correction) that reduces the variability to the greatest extent.

4. Comparison of measured variability with calculated uncertainty

In the analysis presented above, we have characterised the reduction in uncertainty in the measured quantities due to normalisation by evaluating the standard deviation of repeated measurements of each of the quantities involved. This is possible with such a large data set. However, it is more common to estimate the uncertainty according to the standard procedure described in the ISO Guide to the Expression of Uncertainty in Measurement [14]. This shows that the standard uncertainty ($u(x_i)$) of a measurand (x_i) can be calculated as:

$$u^2(x_i) = \sum_{w=1}^{w=q} \left(\frac{\partial f_i}{\partial x_w^*} \right)^2 \cdot u^2(x_w^*) + 2 \sum_{a=1}^{a=q-1} \sum_{b=a+1}^{k=q} \left(\frac{\partial f_i}{\partial x_a^*} \right) \cdot \left(\frac{\partial f_i}{\partial x_b^*} \right) \cdot u(x_a^*, x_b^*) \quad (10)$$

where f represents the measurement equation expressed as a mathematical function. Applying this to the normalisation procedure described by Eq. (1), the partial derivatives required to evaluate Eq. (10) are:

$$\left(\frac{\partial x_i}{\partial x_i^*} \right) = \frac{\sum_{w=1, w \neq i}^{w=q} x_w^*}{T^2} = \frac{T - x_i^*}{T^2}$$

and $\left(\frac{\partial x_i}{\partial x_w^*} \right) = \frac{-x_i^*}{T^2} \forall x_w^*, w \neq i \quad (11)$

where the sum of all component amount fractions is given by:

$$T = \sum_{w=1}^{w=q} x_w^* \quad (12)$$

In the case where the data is completely uncorrelated, the covariance term in Eq. (10) is equal to zero. Substitution of Eq. (11) into Eq. (10) and rearranging gives an expression for the fractional uncertainty in x_i :

$$\frac{u^2(x_i)}{x_i^2} = \left(1 - \frac{2x_i^*}{T} \right) \cdot \frac{u^2(x_i^*)}{x_i^{*2}} + \left(\frac{1}{T^2} \right) \cdot \sum_{w=1}^{w=q} u^2(x_w^*). \quad (13)$$

Eq. (13) is in widespread use [15], but is only valid when the uncertainties in the un-normalised components are completely uncorrelated. It predicts that the normalised uncertainty will be less than the un-normalised uncertainty for any component present with an amount fraction greater than 0.5 mol/mol. When the uncertainties in the un-normalised components are positively correlated, the uncertainty in the normalised component will be smaller than that calculated from Eq. (13) because the covariance term in Eq. (10) is less than zero. In the limiting case where the uncertainties are positively correlated with a correlation coefficient of unity, the uncertainties in the normalised amount fraction tend to zero. As we have shown in the previous section, uncertainties in this type of analysis are generally positively correlated, so we expect Eq. (13) to over-estimate the true uncertainty. Despite this, Eq. (13) is in widespread use because it is usually thought that the process of quantifying the correlations is impractical.

In the component-by-difference method, the measurement equation is given by Eq. (2). When the resulting uncertainties are uncorrelated, the uncertainty of each component is given by:

$$u(x'_i) = u(x_i^*) \quad \text{when } i \neq 1 \quad (14)$$

and

$$u^2(x'_1) = \sum_{i=2}^q u^2(x_i^*) \quad \text{for methane.} \quad (15)$$

Eqs. (14) and (15) are also only valid when the uncertainty in the un-normalised amount fractions ($u(x^*)$) are completely uncorrelated. A feature of both of these methods is that their use establishes correlations in the $u(x_i)$. The importance of determining these covariance terms correctly is shown in Appendix B. However, it is common practice in routine natural gas analysis to ignore them.

The data described in the previous section can be used to determine the effect of using Eq. (13) without quantification of the correlations. Throughout the following discussion, we again analyse data recorded on Day D during which the widest variation in atmospheric pressure was recorded. For each component in cylinder number 860 (typical of all of the mixtures measured), the mean amount fractions and uncertainty of the repeat measurements were calculated before and after normalisation (Table 7). The uncertainty in this instance was taken to be represented by the standard deviation of the repeated measurements. This approach effectively ignores any contribution to the uncertainty from the reference values of the gas mixtures and is not best practice. However, since the uncertainties in the reference values are unaffected by the normalisation procedure, this approximation does not alter the conclusions of this work.

Table 7 also shows the results of using Eq. (13) to estimate the uncertainty. It leads to a lower estimate of the uncertainty of the major component, methane, but increases the uncertainty of all components with an amount fraction

Table 7
Comparison of amount fractions and uncertainties before and after normalisation for cylinder 860 on Day D

Analyte	Amount fraction (mmol/mol)		Uncertainty (% relative)				
	x^*	x	[S.D.] _a	[S.D.] _b	Eq. (13)	Pearson	V/C
Nitrogen	64.61	64.68	0.341	0.036	0.411	0.033	0.036
Carbon dioxide	4.85	4.85	0.347	0.077	0.433	0.075	0.076
Methane	800.04	800.98	0.321	0.006	0.078	0.005	0.006
Ethane	108.58	108.71	0.350	0.029	0.405	0.015	0.031
Propane	5.58	5.58	0.289	0.050	0.388	0.032	0.050
Isobutane	0.51	0.51	0.318	0.237	0.411	0.217	0.236
<i>n</i> -Butane	6.60	6.60	0.282	0.085	0.383	0.068	0.086
Neopentane	2.34	2.35	0.355	0.354	0.440	0.315	0.354
Isopentane	0.04	0.04	8.858	8.704	8.861	7.529	8.690
<i>n</i> -Pentane	3.41	3.41	0.423	0.128	0.496	0.093	0.127
<i>n</i> -Hexane	2.26	2.27	0.349	0.048	0.435	0.042	0.049

x^* and x represent amount fractions before and after normalisation, respectively and five estimates of uncertainty are given by: [S.D.]_a: standard deviation before normalisation; [S.D.]_b: standard deviation after normalisation. Eq. (13): results of Eq. (13). Pearson: results of Eq. (13) corrected for correlation using Pearson correlation coefficients. V/C: results of a full variance/covariance calculation.

less than 0.5 mol/mol. The uncertainties calculated from Eq. (13) are up to an order of magnitude larger than the standard deviation of repeated measurements. As discussed above, it is expected that Eq. (13) overestimates the uncertainty of the compositions because it assumes them to be completely uncorrelated, which is not the case for these data.

A more exact estimate of the uncertainty is given by deriving an expression similar to Eq. (13) that includes the full correlation terms. However, it is possible to improve the results from Eq. (13) by using Pearson correlation coefficients to estimate the random contribution of the uncertainty. This is achieved by evaluating the Pearson coefficient (r) for the correlation between the measured response and the pressure. This enables the component of the variance due to this correlation ($s_{\text{predicted}}^2$) to be subtracted from the total (s_{total}^2) to give the remaining random variance (s_{random}^2) according to Eq. (6). Since we expect all of the variance due to correlation to be removed by the normalisation procedure, s_{random}^2 should give a good estimate of the uncertainty after normalisation. The results of this approximate calculation using Pearson correlation coefficients are also shown in Table 7. This approach produces uncertainty values similar to, but in every case smaller than, those from the standard deviation of repeated measurements. This confirms that the use of Eq. (13) substantially overestimates the uncertainty by incorrectly assuming that no correlation exists in the dataset. Such calculations can be improved by the use of Pearson coefficients to estimate the random component in the variability.

A more thorough approach which accounts for the correlations caused by ‘sample size’ variations is to calculate a full variance/covariance (V/C) matrix for the data. The V/C method has the benefit that all of the covariances inherent in the data (including those from the normalisation procedure) are calculated automatically. The use of the V/C method on the data discussed here yields estimates for the standard de-

viation (Table 7) that are very close to the value calculated for the validity of the normalised data.

5. Summary and conclusions

The analysis of a large set of GC data recorded from a range of natural gas reference mixtures has allowed a detailed study of the statistics of the variations. There is very strong evidence for a ‘sample-size’ effect dominating the variability of the data. The Pearson correlation coefficient for the variation with pressure is greater than 0.96 for all components with amount fractions greater than 3 mmol/mol. For components with lower amount fractions, random noise (isobutane and isopentane) or peak co-elution (neopentane) dominate these systematic sample size-effects.

Correction for this pressure effect reduces the standard deviation of the data by a factor of between 2 and 5 for all components except the least abundant. However, normalisation reduces the standard deviation by a further factor of between 1.5 and 3. The largest reduction in standard deviation is observed for methane, which can be improved by up to 100 times, reducing the standard deviation of the measurements from 0.3% (relative) to less than 0.01% (relative). It is important to select the optimum timescale (either daily or run-by-run) over which to normalise the data. The best choice may obviate the need for pressure correction completely.

The results reported here made use of a very large set of data so that detailed analysis of the variability was possible. It is also possible to calculate the uncertainty in this type of results using a variance/covariance matrix or an approximate estimate can be based on a correction calculated from the Pearson correlation coefficient together with Eq. (13). As expected, when Eq. (13) is used without such a correction for the extent of correlation, it leads to an over-estimate of the uncertainty and incorrectly predicts that normalisation only

reduces the uncertainty for components with amount fractions greater than 500 mmol/mol. Our analysis shows that, under the correct conditions, the uncertainty of all components can be improved.

Acknowledgements

The work carried out at NPL was supported by the UK Department of Trade and Industry under the National Measurement System's Valid Analytical Measurement (VAM) programme. The authors would like to acknowledge useful discussions with Courtney Sayer and Gergely Vargha.

Appendix A

This Appendix describes the mathematical operations carried out to the raw data for each of the three normalisation procedures outlined in Section 3 and Fig. 4. In all instances, the raw data takes the form x_{ijk}^* where x^* is the un-normalised amount fraction of component i (number of components = q) for run j (total number of runs = N) on Day l (number of days of analysis = P). The operators, \mathbf{N} and \mathbf{P} represent the normalisation and pressure correction procedures respectively.

A.1. Procedure (i)—daily normalisation

Evaluation of daily mean:

$$\bar{x}_{ik}^* = \frac{\sum_{j=1}^N x_{ijk}^*}{N} \quad (\text{A.1})$$

Normalisation of daily mean:

$$x_{ik} = \mathbf{N}\bar{x}_{ik}^* \quad (\text{A.2})$$

Evaluation of 'overall' mean:

$$\bar{x}_i = \frac{\sum_{k=1}^P x_{ik}}{P} \quad (\text{A.3})$$

A.2. Procedure (ii)—pressure correction and daily normalisation

Pressure correction:

$$x_{ijk}' = \mathbf{P}x_{ijk}^* \quad (\text{A.4})$$

Evaluation of daily mean:

$$x_{ik}' = \frac{\sum_{j=1}^N x_{ijk}'}{N} \quad (\text{A.5})$$

Normalisation of daily mean:

$$x_{ik}' = \mathbf{N}\bar{x}_{ik}' \quad (\text{A.6})$$

Evaluation of 'overall' mean:

$$\bar{x}_i' = \frac{\sum_{k=1}^P x_{ik}'}{P} \quad (\text{A.7})$$

A.3. Procedure (iii)—run-by-run normalisation

Normalisation of each run:

$$x_{ijk} = \mathbf{N}x_{ijk}^* \quad (\text{A.8})$$

Evaluation of daily mean:

$$\bar{x}_{ik} = \frac{\sum_{j=1}^N x_{ijk}}{N} \quad (\text{A.9})$$

Evaluation of 'overall' mean:

$$\bar{x}_i = \frac{\sum_{k=1}^P \bar{x}_{ik}}{P} \quad (\text{A.10})$$

Appendix B

The importance of finding the covariance term correctly [16] can be illustrated by the calculation of a physical property (Z_{mix}) defined by:

$$Z_{\text{mix}} = \sum_i Z_i x_i \quad (\text{B.1})$$

where Z_i denotes the value of the property being measured² for component i . In the general case, the uncertainty is given by:

$$u^2(Z_{\text{mix}}) = \sum_i Z_i^2 u^2(x_i) + \sum_i \sum_{j \neq i} Z_i Z_j u(x_i, x_j) \quad (\text{B.2})$$

If covariance is ignored, the uncertainty of the component-by-difference method is given by:

$$u^2(Z_{\text{mix}}) = \sum_{i=2}^q (Z_i^2 + Z_1^2) u^2(x_i) \quad (\text{B.3})$$

The correct result for the component-by-difference method is:

$$u^2(Z_{\text{mix}}) = \sum_{i=2}^q (Z_i - Z_1)^2 u^2(x_i) \quad (\text{B.4})$$

² In this Appendix, we use Z to indicate a general property, not the compressibility factor of the gas mixture, which is often denoted Z by other authors.

which can be compared to the correct result for normalisation:

$$u^2(Z_{\text{mix}}) = \frac{\sum_{i=1}^q (Z_i - Z_{\text{mix}})^2 u^2(x_i^*)}{\sum_{w=1}^q x_w^{*2}} \quad (\text{B.5})$$

References

- [1] International Standard ISO 6976(E), Natural gas—calculation of calorific values, density, relative density and Wobbe index from composition, 1995.
- [2] P. Ulbig, D. Hoburg, *Thermochim. Acta* 382 (2002) 27.
- [3] C.J. Cowper, in: E.R. Adlard (Ed.), *Chromatography in the Petroleum Industry*, Elsevier, Amsterdam, 1995.
- [4] C.J. Cowper, A.J. DeRose, *Analysis of Gases by Chromatography*, Pergamon, Oxford, 1983.
- [5] B.D. Skrbic, M.J. Zlatkovic, *Chromatographia* 17 (1983) 44.
- [6] ASTM Standard D1945-03, Standard Test Method for Analysis of Natural Gas by Gas Chromatography, 2003.
- [7] ASTM Standard D1946-90, Standard Practice for Analysis of Reformed Gas by Gas Chromatography, 2000.
- [8] International Standard ISO 6974-3, Natural gas—determination of composition with defined uncertainty by gas chromatography—Part 3: determination of hydrogen, helium, oxygen, nitrogen, carbon dioxide and hydrocarbons up to C₈ using two packed columns, 2001.
- [9] T.J. Lechner-Fish, *Am. Lab.* 28 (1996) 37.
- [10] J.S. Stufkens, H.J. Bogaard, *Anal. Chem.* 47 (1975) 383.
- [11] G.C. Rhoderick, *J. Chromatogr. A* 1017 (2003) 131.
- [12] International Standard ISO 6974-5, Natural gas—determination of composition with defined uncertainty by gas chromatography—Part 5: determination of nitrogen, carbon dioxide and C₁ to C₅ and C₆+ hydrocarbons for a laboratory and on-line process application using three columns, 2001.
- [13] A. Stuart, K. Ord, *Kendall's Advanced Theory of Statistics*, sixth ed., Arnold, London, 1994.
- [14] Guide to the Expression of Uncertainty in Measurement, International Organisation for Standardisation, Geneva, 1995.
- [15] International Standard ISO 6974-2, Natural gas—determination of composition with defined uncertainty by gas chromatography—Part 2: measuring system characteristics and statistics for processing of data, 2002.
- [16] W. Hässelbarth, W. Bremser, in: *Correlation in Natural Gas Composition Data*, Proceedings of the International Gas Research Conference, Amsterdam, 2001.