

(51) International Patent Classification:  
**G06K 9/46** (2006.01)(21) International Application Number:  
PCT/US2009/061031(22) International Filing Date:  
16 October 2009 (16.10.2009)

(25) Filing Language: English

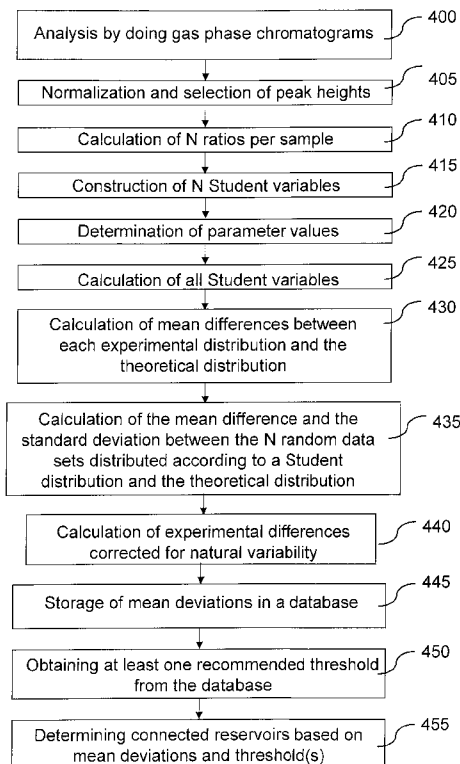
(26) Publication Language: English

(30) Priority Data:  
0805967 27 October 2008 (27.10.2008) FR(71) Applicant (for CA only): **SCHLUMBERGER CANADA LIMITED** [CA/CA]; 525 - 3rd Avenue S.W., 7th Floor, Calgary, Alberta T2P 0G4 (CA).(71) Applicant (for FR only): **TECHSIA S.A.** [FR/FR]; Parc Euromédecine, 340 Rue Louis Pasteur, F-34790 Grabels (FR).(71) Applicant (for all designated States except AG, AO, BA, BB, BW, BZ, CA, CL, CN, CR, CU, DM, DO, FR, GD, GE, GH, GM, GT, HN, HR, IN, IS, KE, KM, KN, KP, LA, LC, LK, LR, LS, LY, MA, ME, MG, MK, MN, MW, MZ, NA, NG, NI, PE, PG, PH, RS, SC, SD, SG, SL, SM, ST, SV, SY, SZ, TZ, UG, US, VC, VN, ZM, ZW): **LOGINED B.V.** [NL/NL]; Gevers Deynootweg 61, S Gravenhage, NL-2586 BJ The Hague (NL).(71) Applicant (for AG, AO, BA, BB, BW, BZ, CL, CN, CR, CU, DM, DO, GD, GE, GH, GM, GT, HN, HR, IN, KE, KM, KN, KP, LA, LC, LK, LR, LS, LY, MA, ME, MG, MK, MN, MW, MZ, NA, NG, NI, PE, PG, PH, RS, SC, SD, SG, SL, SM, ST, SV, SY, SZ, TZ, UG, VC, VN, ZM, ZW only): **PRAD RESEARCH AND DEVELOPMENT LIMITED**; Craigmuir Chambers, P.O. Box 71, Road Town, Tortola (VG).(71) Applicant (for IS only): **SCHLUMBERGER TECHNOLOGY CORPORATION** [US/US]; 5599 San Felipe, Suite 100, Houston, Texas 77056 (US).

[Continued on next page]

(54) Title: PROCESS AND APPARATUS FOR PROCESSING SIGNALS

Figure 4



(57) Abstract: The process of determining similarities of signals resulting from a sample analysis step using the same type of measuring device, characterized by the fact that it comprises a step of determining a value for each parameter of an uncertainty model related to said signals by processing homologous characteristic elements in at least two signals derived from the analysis of similar samples, and a step of determining a measurement of similarity between signals based on each parameter value of the said uncertainty model and of homologous characteristic elements in the different signals whose similarity is being measured. In some embodiments, during the value determination step, a mathematical uncertainty model is optimized so that it corresponds to the theoretically expected differences between the homologous characteristic elements in the various signals from similar samples



(72) **Inventor; and**

(75) **Inventor/Applicant** (*for US only*): **NOUVELLE, Xavier** [FR/FR]; 817 Ancien Chemin du Moulin, F-34270 Les Matelles (FR).

(74) **Agents:** **GALLOWAY, Bryan** et al.; 5599 San Felipe, Suite 100, Houston, Texas 77056 (US).

(81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,

SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— with international search report (Art. 21(3))

## PROCESS AND APPARATUS FOR PROCESSING SIGNALS

5           This invention concerns a process and an apparatus for processing signals. It applies in particular to the processing of signals representative of the measurements made on samples in order to determine similarities between them. More specifically, this invention applies to the signals generated by a chromatograph in order to determine whether two samples of oil come from reservoirs or horizons that are connected.

10           The characterization of reservoir continuity provides information aimed at reducing the key uncertainties in an oil field whose exploitation is being contemplated, and at planning and implementing optimum reservoir development. Such characterization is therefore of great interest to the oil industry.

15           Reservoir characterization and reservoir continuity studies can be achieved in various ways. Some examples of characterization methods aimed at estimating reservoir compartmentalization are PVT (Pressure, Volume, Temperature) measurements, isotope analysis, GCMS (Gas Chromatography - Mass Spectrometry) techniques or FTIR (Fourier Transform Infra Red) spectroscopy, and multidimensional gas chromatography.

20           Among these methods, the use of oil fingerprints obtained by the analysis of gas chromatograms of crude oil is one of the quickest and least expensive.

          The oil fingerprint technique is relatively simple to implement. It consists of comparing various gas chromatograms of samples from different wells in an oil field. The differences between the chromatograms are used as indicators to identify possible barriers between the reservoirs covered by these wells.

25           Since the beginning of the 1980's, this technique, called ROF (Reservoir Oil Fingerprinting), has been widely used to estimate the connections between reservoirs or, what amounts to the same thing, the presence of flow barriers between two reservoirs. The ROF method is based on the comparison of several chromatograms obtained in the same chromatographic conditions. More specifically, it is based on the differences between the peak height ratios of the various samples analyzed.

30           The two peak heights used to calculate each ratio are generally selected so that they are close to one another in order to avoid differences due to phenomena other than reservoir compartmentalization, such as evaporation, gravitational gradients, and the immobilization of heavy components in the chromatograph. In effect, the incoherencies introduced by these mechanisms are particularly significant for components having significant retention time differences.

It is generally accepted that the accuracy of chromatographic analysis is on the order of 1 to 3%. Based on that range, differences in peak height ratios of 5 to 10% or even more cannot be attributed to analytical errors and should represent real differences in crude oil composition. It is therefore customary to use a restricted selection of the most discriminating ratios in order to separate crude oils into meaningful groups. These groups can then be represented on star diagrams that represent the ratios in one plane in order to facilitate comparison.

However, this technique presents serious limitations, particularly with respect to the uncertainties as to peak height ratios: the repeatability deviations of the analysis conditions and the deterioration of the chromatographic column over time play a decisive role that is not taken into account. Furthermore, the uncertainty as to each ratio is highly dependent upon the chromatographic peaks used in this ratio. Problems such as coelutions (simultaneous detection of different components poorly separated by chromatography) or measurement noise result in specificity of the error assigned to each ratio. This problem can lead to erroneous interpretations of the chromatographic data. As proof of this, changing the list of the peak height ratios to be used in a star diagram can lead to very different results.

In addition, since only a restricted number of peak height ratios is considered, the star diagram is only a partial representation of the spatial topology of all of the peak height ratios. Furthermore, slight but numerous differences in composition between samples can have a considerable impact. This compositional variability and the subsequent difficulty of restricting the number of peak height ratios selected can lead to erroneous results.

Other interpretations using GC (Gas Chromatography) peak height ratios in statistical methods are also possible, such as ascending hierarchical classification, main component analysis, or fuzzy logic classification.

However, it is important to note that all these methods are based on relative differences in peak height ratios, the uncertainty of which is undetermined. The results obtained are therefore not expressed in absolute terms and are restricted to the comparison of crude oil samples on a scale that is never the same from one series of comparisons to another.

In other words, the current state of the art does not provide a scale on which differences between samples can be universally measured. The experience acquired on past projects is therefore of little use for future projects. This is all the more annoying because there may be slight differences in composition within a single reservoir or reservoirs that are after all connected (phenomenon of compositional gradients). In the current state, since the amplitude of these phenomena cannot be universally quantified, specialists cannot determine whether differences between samples do or do not signify actual permeability barriers between reservoirs.

This invention is aimed at remedying these disadvantages.

To that end, this invention is first aimed at a process of determining similarities between signals from a sample analysis step using a single type of measurement apparatus, characterized in that it comprises:

5           - a step of determining a value for each parameter of an uncertainty model related to said signals by processing homologous characteristic elements in at least two signals derived from the analysis of similar samples, and

10           - a step of determining a similarity measurement between signals based on each parameter value of the said uncertainty model and of homologous characteristic elements in the signals whose similarity is being measured.

The term "homologous" refers to signals that are characteristic of a specific measurement present in the various analyses, i.e., peaks of a given compound in different signals (in the specific case of chromatography, the peaks of a given compound present in different chromatograms).

15           As a result of these methods, the similarity measurement is directly related to the uncertainty of the measurements. It is observed that the higher the number of peak height ratios associated with the signals, the more precise the determination of the values of the uncertainty model and the more accurate the similarity measurement. Note here that the similar samples do not necessarily come from a single initial sample: the same sample may  
20           merely be analyzed twice, or two samples may be independently taken from the same reservoir and then analyzed.

In some embodiments, during the value determination step, a mathematical uncertainty model is optimized so that it corresponds to the theoretically expected differences between the homologous characteristic elements in the signals from similar samples.

25           In some embodiments, during the value determination step, a mathematical uncertainty model is optimized so that the distribution of the Student variables constructed from the said model corresponds to a theoretical Student distribution with one degree of freedom of the differences between the homologous characteristic elements in the signals from similar samples.

30           In some embodiments, during the parameter value determination step, the parameter values minimize the quadratic distance between a theoretical cumulative distribution function and the cumulative distribution function obtained with the characteristic elements of a pair of signals from similar samples.

35           In some embodiments, during the similarity measurement determination step, the mean difference is determined between the distribution obtained by calculating the differences between the homologous characteristic elements in the signals whose similarity

measurement is being determined and a theoretical distribution based on the said parameter values.

In some embodiments, during the similarity measurement determination stage, the mean difference is determined between the distribution obtained by calculating the differences between the homologous characteristic elements in the signals whose similarity measurement is being determined and a theoretical Student distribution based the said parameter values.

In some embodiments, during the similarity measurement determination stage, the mean difference is determined between the distribution obtained by calculating the difference between the homologous characteristic elements in the signals whose similarity measurement is being determined and a theoretical Gaussian distribution based on the said parameter values.

In some embodiments, during the similarity measurement determination stage, a deviation between the samples is supplied based on the said mean difference.

It is observed that this deviation is independent of the number N of accessible ratios and independent of the sample analysis system, that this deviation is global and absolute (the deviation being independent of the sample analysis system). This deviation makes it possible to create case study databases directly usable for comparisons with any new study project or any new sample analysis apparatus. In effect, since the deviations between the samples are expressed on an absolute scale, the knowledge acquired at one site can be utilized for subsequent studies. For example, for an application to an oil field exploitation project, these databases make it possible in particular to establish composition (or similarity) variation thresholds after which a barrier can be considered to exist between the reservoirs that supplied the crude oil samples.

In some embodiments, during at least one determination step, the homologous characteristic elements are peak heights present in the different signals. It is observed that this implementation method is preferentially suited to the use of a Gaussian distribution.

In some embodiments, during at least one determination step, the homologous characteristic elements are peak height ratios present in the different signals.

It is observed that this embodiment is preferentially suited to the use of a Student distribution.

In some embodiments, during the analysis step, a gas chromatograph is used to supply chromatogram signals.

It is therefore possible to determine the similarities between samples that can be analyzed with such a chromatograph, such as samples of crude oil.

In some embodiments, during at least one determination step, an uncertainty model described by the following equation is used:

$$(\sigma_{Hi}^{Th})^2 = (\alpha \cdot Hi + \beta)^2$$

in which  $\alpha$  and  $\beta$  are parameters to be optimized and  $H_i$  is the normalized height of peak  $i$ .

The inventor has in fact determined that such an uncertainty model does  
5 correspond to numerous sample analyses, such as chromatographic analyses.

Second, this invention is aimed at an apparatus for determining similarities of signals from a sample analysis step using a single type of measurement apparatus, characterized in that it comprises:

- a means of determining a value for each parameter of an uncertainty model  
10 related to the said signals by processing homologous characteristic elements in at least two signals from the analysis of similar samples, and

- a means of determining a similarity measurement between signals based on each parameter value of the said uncertainty model and homologous characteristic elements in the signals whose similarity measurement is being determined.

15 Since the advantages, objectives, and special characteristics of this apparatus are similar to those of the process according to this invention, as briefly explained above, they are not described herein.

Other advantages, objectives, and special characteristics of this invention will become apparent from the following description, which is for explanatory purposes and in no  
20 way limits the attached drawings, in which:

- Figure 1 schematically represents a first implementation method of a device according to this invention,

- Figure 2 schematically represents an oil field with wells,

- Figure 3 schematically represents extracts from gas chromatograms of samples  
25 of crude oil, and

- Figure 4 represents, in the form of a logic diagram, the steps taken in a specific embodiment of the process according to this invention.

As explained above, this invention concerns, in general, a process and an apparatus for processing signals. It applies in particular to the processing of signals  
30 representative of measurements made on samples in order to determine similarities between them, as explained in relation to Figure 1. More specifically, this invention applies to signals generated by a chromatograph in order to determine whether two samples of oil come from reservoirs or horizons that are connected, as explained in relation to Figure 2.

Figure 1 shows a local terminal 105 equipped with a printer 110, a means of  
35 capturing physical quantities 135, and a means of access 115 to a network 120 to which a server 125 is connected. The server 125 is furnished with a database 130.

The local terminal 105 is, for example, a commonly used computer. The local terminal 105 commands the printing of sample similarity tables and the interpretation of such tables. The means of access 115 to the network 120 is, for example, a modem of a known type permitting access to the network 120, for example the internet.

5           The server 125 is of a known type. The database 130 contains at least a list of similarity or deviation measurements between samples already processed using the process according to this invention, along with sample identifiers and possibly means of capturing physical quantities 135 used to analyze the said samples.

10           The terminal 105 contains software which, when run, implements the steps in the process according to this invention. The server 125 contains software which, when run, implements the steps of a similarity or deviation measurement storage and access process.

Alternatively, the terminal 105 does not contain specific software but implements a web browser and a web service contained in the server 125.

15           The terminal 105 and/or the server 125 is adapted to determining similarities in signals from the same means of capture 135, also called "measuring device." To that end, the terminal 105 and/or the server 125 contains:

- a means of determining a value for each parameter of an uncertainty model related to the said signals by processing homologous characteristic elements in at least two signals from the analysis of similar samples, and
- 20           - a means of determining a similarity measurement between signals based on each parameter value of the said uncertainty model and homologous characteristic elements in the signals whose similarity measurement is being determined.

25           These means are, for example, composed of a microprocessor 140 and memories 145 containing an operating system 150 and application software 155 containing instructions to implement the process according to this invention.

Lastly, in a known manner, the local terminal 105 is equipped with a display screen 160 and means of control 165, for example a keyboard and a mouse.

30           The following describes an implementation of this invention for processing signals from a means of capture 135 composed of a gas chromatograph adapted to produce chromatograms of crude oil samples. However, this invention is not limited to this type of means of capture, but on the contrary extends to all types of means of capturing physical quantities, such as sound, mechanical, electrical, or magnetic signals, vibrations, or images. It makes it possible to determine a level of similarities between signals from such means of capture. For example, this invention can be used to recognize or differentiate voices, 35           pronounced words and phrases, objects present in visual or audio scenes, seismic shocks, and electromagnetic waves.



In the case of application to the oil industry, this invention uses a new statistical method for analyzing gas chromatograms, which is based on a coherent quantification of the uncertainty of peak height measurements. The quantification of the uncertainties inherent to each peak height permits a more precise differentiation. Note that it may be used in other fields, such as commingled streams allocation.

Figure 2 shows underground crude oil reservoirs 210, 215, and 220 in an oil field 205. Sample collection wells 225, 230, and 235 permit the collection of samples in each of the reservoirs 210, 215, and 220, respectively.

One of the problems posed for the exploitation of this oil field is the number of wells necessary to reach all the reservoirs. Correlatively, there is an effort to find reservoirs that are connected by an oil flow such as reservoirs 210 and 215 in Figure 2, and reservoirs that are not connected, for example reservoir 220 is not connected to reservoir 210 or to reservoir 215. The processing of gas chromatograms using the process according to this invention makes it possible to estimate a similarity of the chemical compositions of the crude oil samples collected and, accordingly, to provide a probability of connection between reservoirs.

Two chromatogram extracts 305 and 310 illustrated in Figure 3 show that each chromatogram extract contains peaks, 315 to 325, and 330 to 340, respectively. These peaks may be facing each other on the two chromatograms, like peaks 315 and 330. They may also be very close on the same chromatogram, like peaks 335 and 340. If the height of one peak influences the height of the other, there is coelution.

In the remaining of the description, the peak height ratios of each chromatogram are considered that chromatogram's footprint. Note that as many peaks as possible, even all the peaks, are used. These peaks are common to all the chromatograms, i.e., the peaks have the same Kovats index to within one chromatographic peak width.

It should be noted here that Kovats indices are relative retention values calculated so as to be independent, to the first order, of the chromatographic conditions in which the samples were analyzed. The Kovats index of a compound *i* is calculated using the following equation:

$$\text{KovatsID}(i) = 100 \cdot (n_{\text{Ci-1}} + (\log_{10}(t_i) - \log_{10}(t_{\text{Ci-1}})) / (\log_{10}(t_{\text{Ci}}) - \log_{10}(t_{\text{Ci-1}})))$$

where:

KovatsID(*i*) = Kovats index of compound *i*,

$n_{\text{Ci-1}}$  = Number of carbons of the *n*-paraffin located just before compound *i*,

$t_i$  = Retention time of compound *i*,

$t_{\text{Ci-1}}$  = Retention time of the *n*-paraffin located just before compound *i*, and

$t_{\text{Ci}}$  = Retention time of the *n*-paraffin located just after compound *i*.

As can be seen in Figure 4, in a specific embodiment, the first step 400 of the process consists of making a chromatogram for each sample collected, and identifying the chromatograms that involve samples from the same sample collection well.

Then, in a step 405, the peaks of the signals representing the chromatograms are normalized so as to offset the differences in injected volume of the samples in the chromatograph system, as expressed below with respect to Equation (4). In this same step 405, the homologous peaks from all the chromatograms are selected, i.e., those having the same Kovats index. This provides characteristic measurements specific to each sample in the form of the selected peak heights.

In a step 410, N ratios of these characteristic measurements are calculated for each sample.

In a step 415, N Student variables are constructed from the N ratios of each pair of samples and from an uncertainty model whose parameters are to be optimized.

In a step 420, from the pairs of analyses assumed to be similar and from the Student variables, the values of the parameters that optimize the uncertainty model are determined.

In a step 425, all the Student variables for all the other pairs of analyses are calculated using the parameters optimized in step 420.

In a step 430, for each pair of analyses, the mean difference between the distribution of the N Student variables and a theoretical Student distribution with one degree of freedom is calculated.

In a step 435, the mean deviation is established between sets of N random data distributed according to a Student distribution and a theoretical Student distribution. The standard deviation is also calculated. These values represent the expected natural variability for N variables.

In a step 440, the mean deviation determined in step 435 is subtracted from the mean difference determined in step 430 so as to define a zero difference when two analyses are similar. The uncertainty as to the mean difference between two analyses is also given by the standard deviation determined in step 435.

With respect to steps 415 and 420, a mathematical model is constructed in order to estimate the uncertainties as to the peak height ratios between two chromatograms using the pairs of analyses pertaining to the same crude oil sample(s) and using the theoretical distribution of the expected errors (in the case of peak height ratios: Student distribution with one degree of freedom).

The uncertainty model must meet two criteria: it must be sufficiently flexible so that the distribution of the Student variables (constructed from the uncertainty model) is in keeping with the Student distribution with one degree of freedom, but is at the same time

sufficiently general to apply to a series of analyses relatively close to one another in time. Numerous tests were conducted in order to find the best compromise between these two criteria.

With respect to the model distribution curve (Student distribution with one degree of freedom), it can be shown that the variables constructed from Equation (2) below are distributed according to a Student law with one degree of freedom. In this case, this model curve is therefore fixed once and for all. However, the technique used could also be applied to more conventional measurements having a Gaussian type error, for example using peak heights instead of peak height ratios. In this case, since the variables are not in the form of ratios, it is simply necessary to use the normal law instead of the Student law with one degree of freedom.

The number of analyses to be done successively on a single batch, i.e., a single series of analyses with the same means of capture, in this case the same chromatograph, is restricted for two main reasons. First, the chromatographic analysis of crude oil takes time. In addition, chromatographic data are increasingly altered as more analyses are done with the same chromatographic column.

Since the number of measurements of the height of the same peak of the same fluid is very limited, the common uncertainty calculations, such as standard deviation, cannot be done with precision.

For the implementation of the process according to this invention, the numerous peak height ratio measurements in each chromatogram are used. The Student variable  $t$  is used, which is expressed as:

$$t = \frac{(\bar{x} - \mu)}{\sigma_e / \sqrt{n}} \quad (1)$$

in which:

$\bar{x}$  is the mean value of the sample for a descriptive variable  $x$ ,

$\mu$  is the expectation for variable  $x$ ,

$\sigma_e$  is the standard deviation estimator for the population, and

$n$  is the number of independent measurements in the sample.

In our case, the variables are the peak height ratios.

For two chromatograms representing the same crude oil, the expectation  $\mu$  is zero for each variable. In such a configuration, the Student variable  $t$  associated with each peak height ratio can be written:

$$t_{ij} = \frac{\left( \left( \frac{H_i}{H_j} \right)_{Chromato.1} - \left( \frac{H_i}{H_j} \right)_{Chromato.2} \right)}{\sqrt{\left( \sigma_{H_i/H_j} \right)_{Chromato.1}^2 + \left( \sigma_{H_i/H_j} \right)_{Chromato.2}^2}} \quad (2)$$

where  $H_i$  is the normalized height of peak  $i$ .

The distribution of such a variable is known as Student distribution with one degree of freedom. When the number of variables  $t_{ij}$  for each chromatogram is reasonably high, the theoretical distribution of such variables is approached if the denominator in Equation (2) is correctly estimated.

Four error factors can be determined in the measurements in the chromatograms:

- The first concerns the peak height in the measurement, where it can be considered that the lower the peak, the higher the relative error;

- The second refers to the phenomenon of coelution, which has a direct effect on peak height; the greater the coelution, the more a peak is modified by the contribution of neighboring peaks; the ratio between the peak heights and the height of its highest valley gives a good representation of the degree of coelution;

- The third concerns the difference between the retention times of a given compound in successive analyses; in effect, if the retention time of a given compound varies from one analysis to another, the peak height measurements may be estimated erroneously due to the phenomenon of coelution, the positioning of the baseline, or erroneous identifications;

- The last factor involves the height of the baseline: the higher the baseline, the greater the uncertainties on the peak height measurement. All these factors were integrated in conceptual models. Each of these models was tested with different sets of data in order to determine its accuracy and its strength.

In order to optimize the conceptual model, since a pair of chromatograms comes from the same crude oil sample, the cumulative distribution function, or CDF, of the  $t$  variables on each peak height ratio should follow the same trend as the theoretical  $t$ -Student distribution with one degree of freedom.

These considerations were used to optimize the models: their optimum values are the values that minimize the quadratic distance between the theoretical CDF and the experimental CDF obtained with pairs of chromatograms from the same sample of crude oil.

The best uncertainty model is therefore the one that minimizes this criterion, but also the one that has the best potential of being universal. This latter requirement is evaluated by splitting the pairs of chromatograms of the same samples of crude oil into two groups: the first is used to optimize the parameters of the model and the second is used to

validate its universality. Universality means that the validation data (pairs of samples not used for the optimization of the uncertainty model but acknowledged to be similar) must exhibit a similarity comparable to that found for the data used for optimization. If this is not the case, the uncertainty cannot be used, for example due to an excessively rapid change in the measurement system. It is therefore recommended that the same sample be analyzed at the end of the series of samples as at the beginning. Alternatively, the parameter values determined at various stages of the series of samples analyzed are interpolated.

The various simulations lead to the reduction of the number of variables in the uncertainty estimate and the best conceptual model can be described by the following equation:

$$(\sigma_{Hi}^{Th})^2 = (\alpha.H_i + \beta)^2 \quad (3)$$

in which  $\alpha$  and  $\beta$  are parameters to be optimized and  $H_i$  is the normalized height of peak  $i$ .

The peak heights are normalized in order to avoid potential effects of scale, using the following formula:

$$(H_i^{Fluid\ I})_{Norm} = \frac{H_i^{Fluid\ I}}{\left( \frac{1}{N_H} \sum_j^{N_H} H_j^{Fluid\ I} \right)} \quad (4)$$

In effect, as illustrated in Figure 3, peak amplitude may vary considerably from one fluid to another without increasing the relative uncertainty.

The calculation of the propagation of uncertainties can then be expressed as

$$\sigma_{Hi/Hj}^{Th} = \frac{1}{H_j} \sqrt{(\alpha.H_i + \beta)^2 + \left( \frac{H_i}{H_j} \right)^2 (\alpha.H_j + \beta)^2} \quad (5)$$

which permits the transposition of the peak height uncertainties to peak height ratios.

The distributions of the experimental (Equation 2) and theoretical Student variables are then compared in order to obtain an absolute measurement of the differences between the footprints composed of the chromatogram peak height ratios.

It is known that the characterization of the barriers between oil reservoirs is often complicated by the variability of the fluid composition in a given reservoir. Since the conventional method used to characterize compartmentalization is applied on a relative scale, and with a restriction of the peak height ratios, the composition variability due to true compartmentalization or due to thermo-gravitational (or other) effects cannot be quantified.

In order to solve this critical problem, the inventor has calculated an absolute deviation between the samples of crude oil. From one case to another, since the deviations

are expressed on a comparable scale, the exact involvement of flow barriers and internal reservoir variability can be quantified and therefore taken into account. However, since the uncertainties on peak height ratios are not the same from one series of analyses to the next, the inventor has opted to use the Student t variable distribution curves obtained for each pair  
 5 of fluids. Since each analysis is transposed in terms of t-variables, and since these variables are independent of the amplitude of the uncertainties, the mean deviation between the theoretical and experimental CDF values is expressed on a universal scale. Using the natural variability of N random variables (N being the number of peak height ratios used) distributed according to a Student distribution with one degree of freedom, the uncertainty on the  
 10 evaluation of difference can also be estimated.

The simplest way to evaluate the mean deviation between the theoretical and experimental CDF values consists of applying the following formula:

$$D_{Fluide1 / Fluide2} = \frac{1}{N} \sqrt{\sum_{i=1}^n (CDF_{Fluide1 / Fluide2}(t_i) - CDF_{Theo}(t_i))^2} \quad (6)$$

in which:

15  $CDF_{Fluide1 / Fluide2}(t_i)$  is the CDF associated with each experimental variable  $t_i$

$CDF_{Theo}(t_i)$  is the CDF associated with the same theoretical variable  $t_i$ , and

N is the number of t-variables.

However, this method is not sufficient when some experimental CDF values are narrower than the theoretical CDF. In such cases, the calculated deviation between the CDF  
 20 values would be large, while any single-point comparison of the pair of chromatograms would be closer than expected.

This situation appears when the pairs of chromatograms used to establish the uncertainty model are less distant than others that were not used. This may occur because the CDF comparison method is sensitive enough to register the gradual deterioration of the  
 25 chromatograph columns during the cycle of analyses of a set of crude oil samples.

The mean deviation between the experimental and theoretical CDF values is finally calculated using the following formula:

$$D_{Fluide1 / Fluide2} = \frac{1}{n} \sum_{i=1}^n d_i \quad (7)$$

The values of the  $d_i$  depend on the values of the variables  $t_i$ :

30  $si \ t_i < 0 : d_i = CDF_{Exp}(t_i) - CDF_{Theo}(t_i) \quad (8)$

$si \ t_i > 0 : d_i = CDF_{Theo}(t_i) - CDF_{Exp}(t_i) \quad (9)$

This result avoids penalizing experimental distributions that are narrower than expected according to the uncertainty model: such a case would be represented by a negative difference.

5 With respect to the expression of the uncertainty on the various calculations, a generator of random variables distributed according to a Student distribution with one degree of freedom was constructed. It can be used to do a simulated CDF with N random variables to calculate its differences from the theoretical CDF according to Equations (7), (8), and (9).

10 This process can also be repeated hundreds of times in order to estimate its standard deviation. This latter value corresponds to the natural variability of simulations of N variables. If the number of random variables N coincides with the number of independent ratios of the set of real data, it can be used as a reliable estimate of the uncertainty on the difference between experimental and theoretical CDF values.

15 In a step 445, the deviations between the footprints are stored on an absolute scale in a database. Which permits the creation of case study databases directly usable for comparisons with any new study project. In effect, since the deviations between samples are expressed on an absolute scale, the knowledge acquired on a series of analyses can be utilized for subsequent studies.

20 Among other things, these databases make it possible to establish composition variation thresholds beyond which a barrier can be considered to exist between the reservoirs that supplied the oil samples.

In a step 450, recommended deviation thresholds can be recovered from the database beyond which samples are considered as not coming from interconnected reservoirs.

25 In a step 455, on the basis of the recommended thresholds or thresholds specifically determined for the project, samples that come from reservoirs that are probably interconnected can be estimated based on the deviations between samples.

Table 1 below concerns two data sets consisting of several sample analyses. Each data set contains analyses of fluids from different wells but from the same oil field.

30 Some samples were analyzed twice in order to construct the uncertainty model on each data set and in order to validate it.

In Table 1, the order in which the analyses were made has been preserved. This can be important because the quality of the peak height measurements can deteriorate as the analyses progress. For each data set, the uncertainty model was constructed with the first sample.

Table 1

| Data set 1 |        | Data set 2 |        |
|------------|--------|------------|--------|
| Analysis   | Sample | Analysis   | Sample |
| A1         | 1      | A1         | 1      |
| B1         | 2      | B1         | 2      |
| C1         | 3      | C1         | 3      |
| D          | 4      | D1         | 4      |
| E          | 5      | D2         | 4      |
| F          | 6      | C2         | 3      |
| G          | 7      | B2         | 2      |
| C2         | 3      | A2         | 1      |
| B2         | 2      |            |        |
| A2         | 1      |            |        |

The following describes an example of implementation of the process according to this invention. In this example, only sample 1 (analyses A1 and A2) was used to obtain the uncertainty model for each data set. The result is given in Table 2 below:

|    | A1     | A2    | B1     | B2    | C1     | C2    | D     | E     | F      |
|----|--------|-------|--------|-------|--------|-------|-------|-------|--------|
| A2 | -0.002 |       |        |       |        |       |       |       |        |
| B1 | 0.198  | 0.195 |        |       |        |       |       |       |        |
| B2 | 0.196  | 0.194 | -0.023 |       |        |       |       |       |        |
| C1 | 0.182  | 0.181 | 0.156  | 0.157 |        |       |       |       |        |
| C2 | 0.180  | 0.180 | 0.159  | 0.161 | -0.015 |       |       |       |        |
| D  | 0.190  | 0.189 | 0.145  | 0.151 | 0.073  | 0.093 |       |       |        |
| E  | 0.198  | 0.197 | 0.087  | 0.099 | 0.157  | 0.160 | 0.146 |       |        |
| F  | 0.192  | 0.191 | 0.062  | 0.063 | 0.156  | 0.157 | 0.145 | 0.087 |        |
| G  | 0.196  | 0.193 | 0.049  | 0.059 | 0.154  | 0.157 | 0.142 | 0.056 | -0.002 |

Analyses B1 and B2 (sample 2) and analyses C1 and C2 (sample 3) validate the uncertainty model because they are not used to construct it. The analyses of these two pairs are narrower than expected, since their difference with respect to theory is slightly negative (-0.023 and -0.015, respectively). This can be explained by the order in which the analyses were done. Analyses F and G cannot be differentiated either (result of -0.002). This result means that samples F and G come from reservoirs that are connected, or from separated reservoirs in which the crude oil has the same chemical composition (also called "false negatives").

The other analyses show significant differences in composition according to the values given in Table 2 and the 95% confidence interval (equivalent to  $\pm 0.019$ ). Note that in order to determine this value of  $\pm 0.019$  the standard deviation is multiplied by the multiplier



1.96 found in the table of Student fractiles. Such a result suggests the existence of flow barriers between the reservoirs represented by these analyses. However, the fluid composition variability in a single reservoir is actually not very well known.

As a result, the maximum threshold value beyond which the presence of a flow barrier between two reservoirs is certain may be greater than zero. The specific implementation method of the process according to this invention explained herein supplies absolute measurements between two footprints. It would therefore be sufficient to set this threshold a single time for a well-known project.

However, by processing the thresholds used on several past projects, a threshold to apply to a new project can be pre-estimated. As knowledge accumulates, it is also possible to determine the expected composition deviations for representatives of a given reservoir based on various parameters such as the distance between sample collection points, the nature of the reservoir rock, fractures, etc.

On a second data set, the distribution analysis result given in Table 3 below for the 95% confidence interval (equivalent to  $\pm 0.015$ ) suggests that the differences between footprints can be entirely attributed to the uncertainty on the peak height measurements. It can be deduced from this that all the samples from this data set come from the same reservoir or are "false negatives." The existence of flow barriers cannot be deduced from differences in composition.

|    | A1     | A2     | B1     | B2    | C1     | C2     | D1     |
|----|--------|--------|--------|-------|--------|--------|--------|
| A2 | -0.001 |        |        |       |        |        |        |
| B1 | -0.013 | -0.014 |        |       |        |        |        |
| B2 | -0.009 | -0.001 | -0.021 |       |        |        |        |
| C1 | 0.005  | 0.011  | -0.001 | 0.006 |        |        |        |
| C2 | 0.004  | 0.018  | -0.003 | 0.006 | -0.003 |        |        |
| D1 | 0.011  | 0.002  | -0.002 | 0.006 | -0.008 | -0.006 |        |
| D2 | 0.001  | 0.008  | 0.002  | 0.002 | -0.013 | -0.016 | -0.005 |

In conclusion, the process according to this invention provides an estimate of peak height ratio uncertainty on the basis of solid theoretical knowledge, even if the number of measurements of each peak height for a given fluid is very limited.

The lesson provided by this invention involves analyzing the distribution of the differences between two representations (in this case, chromatograms) of the same fluid (for example, the same sample), instead of focusing on the values of the peak height ratios associated with those representations. The uncertainty model constructed on two footprints from the same sample (in this case, crude oil) is applied to the other footprints made with the same equipment (in this case, the same chromatograph column), and the number of analyses made for the same data set must be sufficiently restricted to preserve the generality

of the model. In the implementation method explained in Example 2 above, each sample results in a pair of analyses supplying pairs of footprints, only one of which was used to construct the uncertainty model.

5 In addition, the lesson provided by this invention involves determining a deviation between footprints on an absolute scale. The distribution of t variables determined for a pair of analyses is compared to the theoretical distribution obtained if those analyses had come from the same sample. The mean deviation, compared to the expected natural variability, provides an absolute measurement of the differences between these two footprints.

10 The inventor has determined that these uncertainties between two analyses can be used as a strong basis for estimating the actual deviation between them.

## CLAIMS

1 - Process for determining similarities of signals from a sample analysis step using the same type of measuring device, characterized in that it comprises:

5           - a step of determining a value for each parameter of an uncertainty model related to said signals by processing homologous characteristic elements in at least two signals derived from the analysis of similar samples, and

          - a step of determining a similarity measurement between signals based on each parameter value of the said uncertainty model and of homologous characteristic elements in  
10 the signals whose similarity is being measured.

2 - Process according to Claim 1, characterized by the fact that during the value determination step a mathematical uncertainty model is optimized so that it corresponds to the theoretically expected differences between the homologous characteristic elements in the signals from similar samples.

15 3 - Process according to either of Claims 1 or 2, characterized by the fact that during the value determination step a mathematical uncertainty model is optimized so that the distribution of the Student variables constructed from the said model corresponds to a theoretical Student distribution with one degree of freedom of the differences between the homologous characteristic elements in the signals of similar samples.

20 4 - Process according to either of Claims 2 or 3, characterized by the fact that during the step of determination of the value of each parameter, the parameter values minimize the quadratic distance between a theoretical cumulative distribution function and the cumulative distribution function obtained with the characteristic elements of a pair of signals from similar samples.

25 5 - Process according to Claim 4, characterized by the fact that during the similarity measurement determination step, the mean difference is determined between the distribution obtained by calculating the differences between the homologous characteristic elements in the signals whose similarity is being measured and a theoretical distribution based on the said parameter values.

30 6 - Process according to Claim 5, characterized by the fact that during the similarity measurement determination step, the mean difference is determined between the distribution obtained by calculating the differences between the homologous characteristic elements in the signals whose similarity is being measured and a theoretical Student distribution based on the said parameter values.

35 7 - Process according to Claim 5, characterized by the fact that during the similarity measurement determination step, the mean difference is determined between the distribution obtained by calculating the differences between the homologous characteristic elements in

the signals whose similarity is being measured and a theoretical Gaussian distribution based on the said parameter values.

8 – Process according to any of Claims 5 to 7, characterized by the fact that during the similarity measurement determination step, a deviation between the samples is supplied  
5 based on the said mean difference.

9 – Process according to any of Claims 1 to 8, characterized by the fact that during at least one determination step the homologous characteristic elements are homologous peak heights in the different signals.

10 – Process according to any of Claims 1 to 9, characterized by the fact that during at least  
10 one determination step the homologous characteristic elements are homologous peak height ratios in the different signals.

11 – Process according to any of Claims 1 to 10, characterized by the fact that during the analysis step a gas chromatograph is used to supply chromatogram signals.

12 – Process according to any of Claims 1 to 11, characterized by the fact that during at least  
15 one determination step an uncertainty model described by the following equation is used:

$$\left(\sigma_{Hi}^{Th}\right)^2 = (\alpha.H_i + \beta)^2$$

in which  $\alpha$  and  $\beta$  are parameters to be optimized and  $H_i$  is the normalized height of peak  $i$ .

13 – Apparatus for determining similarities of signals from a sample analysis step using the  
20 same type of measuring device, characterized in that it comprises:

- a means of determining a value for each parameter of an uncertainty model related to the said signals by processing homologous characteristic elements in at least two signals from the analysis of similar samples, and

- a means of determining a measurement of similarity between signals based on  
25 each parameter value of the said uncertainty module and of homologous characteristic elements in the different signals whose similarity is being measured.

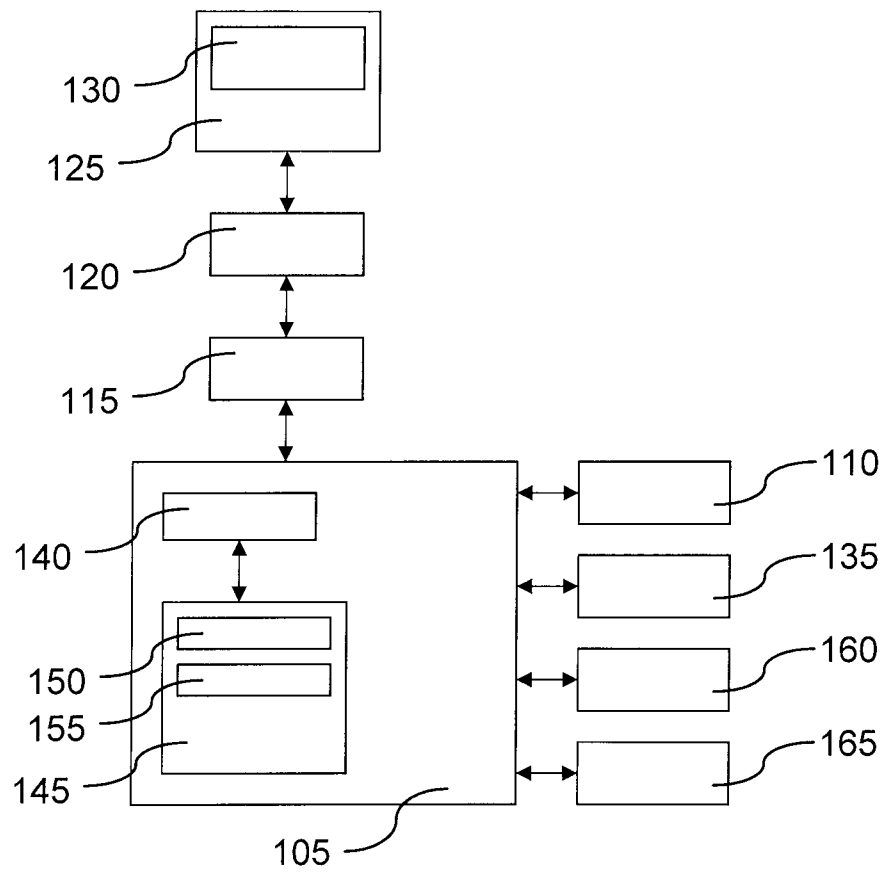


Figure 1

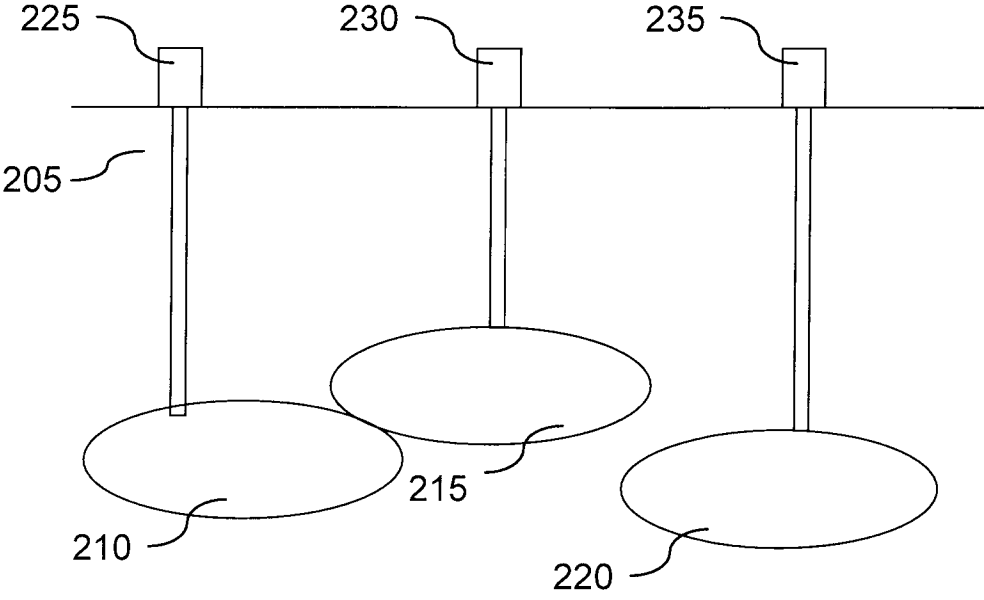


Figure 2

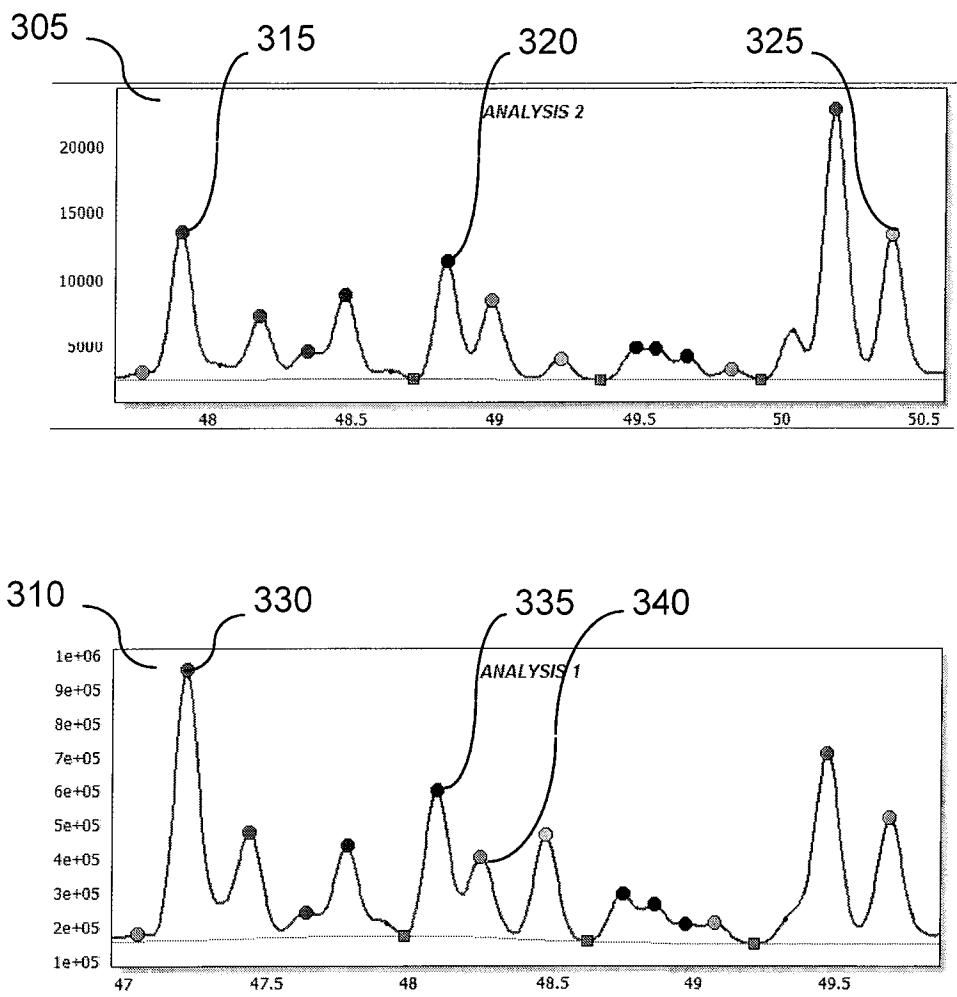
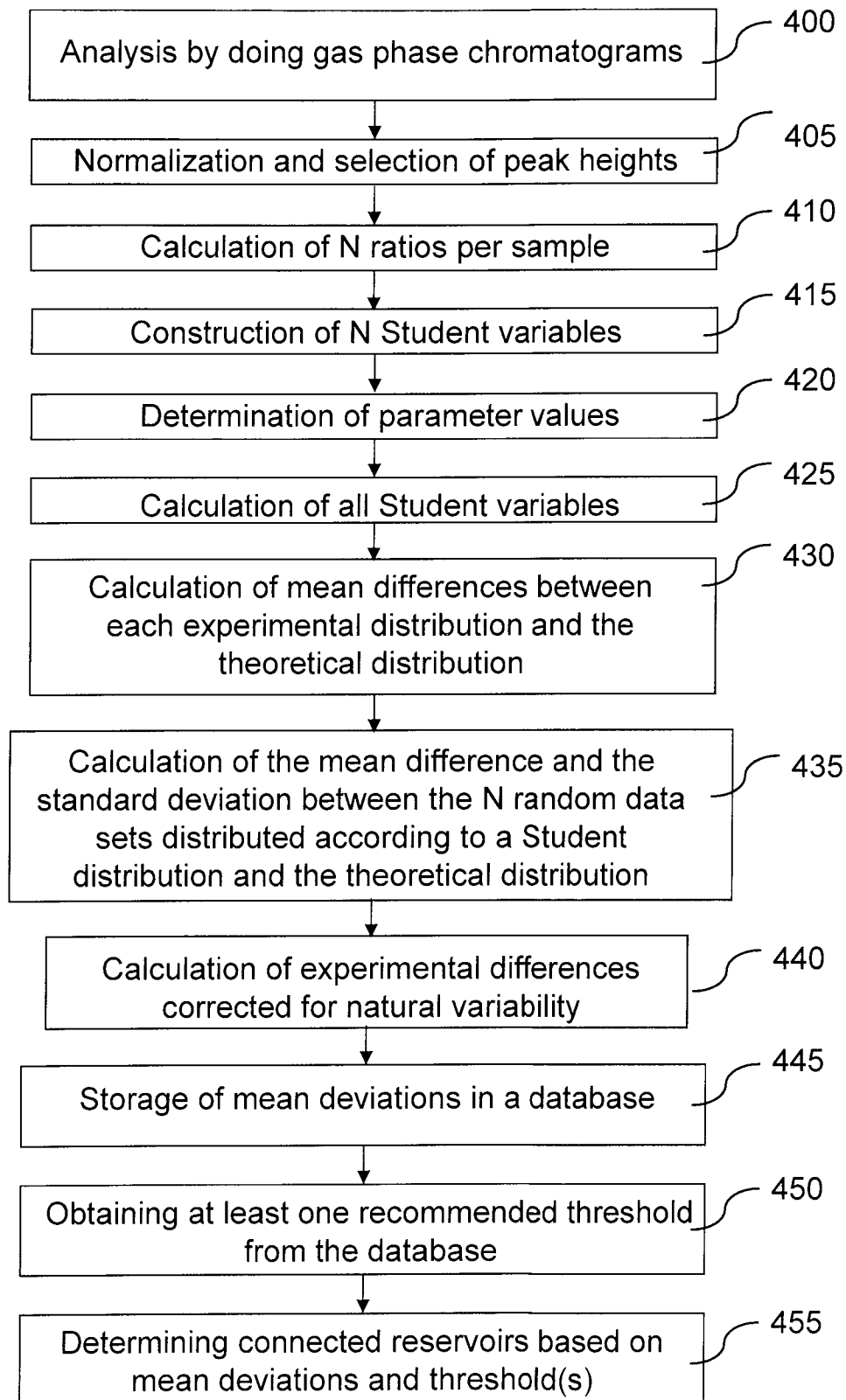


Figure 3

Figure 4





## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2009/061031

A. CLASSIFICATION OF SUBJECT MATTER  
IPC(8) - G06K 9/46 (2009.01)  
USPC - 382/154

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC(8) - 9/00, 9/32, 9/34, 9/36, 9/46 (2009.01)  
USPC - 382/128, 154, 232

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MicroPatent

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|-----------|--------------------------------------------------------------------------------------|-----------------------|
| Y         | US 2008/0031528 A1 (CROMBEZ et al) 07 February 2008 (07.02.2008) entire document     | 1-3, 13               |
| Y         | US 2008/0039706 A1 (CHEFD'HOTEL et al) 14 February 2008 (14.02.2008) entire document | 1-3, 13               |
| Y         | US 4,984,160 A (SAINT FELIX et al) 08 January 1991 (08.01.1991) entire document      | 3                     |
| A         | US 2003/0197972 A1 (MORRIS) 23 October 2003 (23.10.2003) paragraph 0034              | 3                     |
| A         | US 2005/0141757 A1 (AYACHE et al) 30 June 2005 (30.06.2005) entire document          | 1-3, 13               |

☐ Further documents are listed in the continuation of Box C.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

07 December 2009

Date of mailing of the international search report

15 DEC 2009

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, Virginia 22313-1450  
Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300  
PCT OSP: 571-272-7774

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US2009/061031

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 4-12  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.