Anal. Chem. **2006,** *78,* 2579−2592

Identification and Quantification of Homologous Series of Compound in Complex Mixtures: Autocovariance Study of GC/MS Chromatograms

Maria Chiara Pietrogrande,* Maria Grazia Zampolli, and Francesco Dondi

Department of Chemistry, University of Ferrara, Via L. Borsari, 46, 44100 Ferrara, Italy

The paper describes a method for determining homologous classes of compounds in a multicomponent complex chromatogram obtained under programming elution conditions. The method is based on the computation of the autocovariance function of the experimental chromatogram (EACVF). The EACVF plot, if properly interpreted, can be regarded as a "class chromatogram" i.e., a virtual chromatogram formed by peaks whose positions and heights allow identification and quantification of the different homologous series, even if they are embedded in a random complex chromatogram. Theoretical models were developed to describe complex chromatograms displaying random retention pattern, ordered sequences or a combination of them. On the basis of theoretical autocovariance function, the properties of the chromatogram can be experimentally evaluated, under well-defined conditions: in particular, the two components of the chromatogram, ordered and random, can be identified. Moreover, the total number of single components (SCs) and the separated number of the SCs belonging to the random and ordered components can be determined, when the two components display the same concentration. If the mixture contains several homologous series with common frequency and different phase values, the number and identity of the different homologous series as well as the number of SCs belonging to each of them can be evaluated. Moreover, the power of the EACVF method can be magnified by applying it to the single ion monitoring (SIM) signals to selectively detect specific compound classes in order to identify the different homologous series. By this way, a full "decoding" of the complex multicomponent chromatogram is achieved. The method was validated on synthetic mixtures containing known amount of SCs belonging to homologous series of hydrocarbon, alcohols, ketones, and aromatic compounds in addition to other not structurally related SCs. The method was applied to both the total ion monitoring (TIC) and the SIM signals, to describe step by step the essence of the procedure. Moreover, the systematic use of both SIM and TIC can simplify the decoding procedure of complex chromatograms by singling out only specific compound classes or by confirming the identification of the different homologous series. The method was further applied to a sample containing unknown number of compounds and homologous series (a petroleum benzin, bp 140−160 °C): the results obtained were meaningful in terms of both the identified number of components and identified homologous series.

To date GC analysis is a very rich source of data for chemical analyses, but extracting relevant information from the large, complex data sets is a challenge for information technologies. This is particularly true for hyphenated (GC/MS) and multidimensional (LC−GC, GC−GC, GC×GC) GC techniques, which generate data sets that are 2−3 orders of magnitude larger than for conventional GC.[1,2] The quantity and complexity of GC data makes human analyses of GC signals difficult and time-consuming and motivates the need for computer-assisted signal processing to transform GC data into usable information. Advanced information technologies offer powerful solutions for many of the problems associated with the GC analysis: data handling, processing, analysis, and reporting. In particular, a mathematical approach is very useful to deconvolve incompletely resolved peaks and to interpret the chromatogram, extracting all the analytical information hidden therein, in other words "decoding" the complex chromatogram.[3−10]

Among the many procedures developed, a chemometric approach based on the autocovariance function (ACVF) has proved to be a very powerful tool for interpreting chromatograms of complex mixtures.[10−26] The ACVF method is able to extract

* To whom correspondence should be addressed. E-mail: mpc@unife.it; Fax: 0039 0532 240709.

(1) Christensen, J. H.; Mortensen, J.; Hansen, A. B.; Andersen, O. *J. Chromatogr., A* **2004**, *1062*, 113−123.
(2) Wilson, I. D.; Brinkman, U. A. Th. *J. Chromatogr., A* **2003**, *1000*, 325−356.
(3) Felinger, A. *Adv. Chromatogr.* **1998**, *39*, 201−237.
(4) Felinger, A. In *Data analysis and signal processing in chromatography*; Elsevier: Amsterdam, 1998.
(5) Davis, J. M.; Samuel, *J. High Resolut. Chromatogr.* **2000**, *23*, 235−244.
(6) Davis, J. M.; Pompe, M.; Samuel, C. *Anal. Chem.* **2000**, *72*, 5700−5713.
(7) Shao, X.; Wang, G.; Wang, S.; Su, Q. *Anal. Chem.* **2004**, *76*, 5143−5148.
(8) Wiberg, K.; Andersson, M.; Hagman, A.; Jacobsson, S. P. *J. Chromatogr., A* **2004**, *1029*, 1-X.
(9) Zenkevich, I. G. *Chem. Intell. Lab. Syst.* **2004**, *72*, 233−240.
(10) Dondi, F.; Pietrogrande, M. C.; Felinger, A. *Chromatographia* **1997**, *45*, 435−440.
(11) Felinger, A.; Pasti, L.; Dondi, F. *Anal. Chem.* **1990**, *62*, 1846−1853.
(12) Felinger, A.; Pasti, L.; Dondi, F. *Anal. Chem.* **1990**, *62*, 1854−1860.
(13) Felinger, A.; Pasti, L.; Dondi, F. *Anal. Chem.* **1991**, *63*, 2627−2633.
(14) Felinger, A.; Pasti, L.; Dondi, F. *Anal. Chem.* **1992**, *64*, 2164−2174.
(15) Dondi, F.; Betti, A.; Pasti, L.; Pietrogrande, M. C.; Felinger, A. *Anal. Chem.* **1993**, *65*, 2209−2222.
(16) Pietrogrande, M: C.; Pasti, L.; Dondi, F.; Bollain Rodriguez, M. H.; Carro Diaz, M. A. *J. High Resolut. Chromatogr.* **1994**, *17*, 839−850.

10.1021/ac051491e CCC: $33.50
Published on Web 03/10/2006

information concerning the mixture—number of components, abundance distribution, and recursive chemical structures of compounds—and the separation—separation performance and retention pattern.[11−12] In particular, its power lies in its ability to identify ordered retention patterns, singling them out from the complex chromatogram, which appears crowded with peaks randomly distributed throughout the chromatographic space.[15,19−25] This is the case of the terms of a homologous compounds series that yield an ordered sequence of peaks appearing at repeated distances, i.e., related to addition of a CH_2 group to subsequent terms in the series. The ACVF approach was recently applied to GC/MS chromatograms, in particular to single ion monitoring (SIM) signals [24] and extended from one-dimensional (1D) chromatograms to two-dimensional (2D) separations.[26]

This paper is focused on the separation of homologous series, since they contain very relevant information for characterizing organic compounds in complex mixtures of environmental pollutants in natural, industrial, or environmental samples.[27,28] The recognition of the presence of terms of an homologous series and information on the distribution pattern of the terms, i.e., odd/even predominance, are diagnostic for assessment of natural versus anthropogenic origin of contamination sources. The molecular composition of light hydrocarbons in petroleum products is extensively used in organic geochemistry studies, i.e., identification of sedimentary organic matter or biological precursors.[29−32] In fact, the presence of terms of a homologous series of compounds presenting a nonrandom distribution of the number of carbon atoms can be regarded as a biochemical signature, since they are the product of biochemical reactions due to living organisms.[27,33]

Up to now, the ACVF approach has been applied to random multicomponent mixtures and chromatograms, i.e., chromatograms containing a great number of single components (SCs) whose retention times are randomly distributed over the chromatographic space. In these cases, the mixture complexity, i.e., the number of SCs, can be quantitatively determined and the method was fully validated.[10−21] In the case of ordered multicomponent chromatograms, only the presence of the homologous series was detected by using the ACVF approach, without systematically facing the estimation of the number of SCs belonging to the specific classes.[22,24] This paper discusses how the ACVF can be used to fully extract information on the different homologous series present in a complex sample: detection of the presence of terms of an ordered series, characterization of their specific chemical structure, and quantification of the number of components in the homologous series. In particular, pertinent expressions for determining the above-mentioned attributes are derived and the specific conditions in which such expressions can be applied are described. They are applied to determine the number of homologous series and the number of SCs belonging to each series and other SCs not structurally related to each other. By this way, a method to experimentally evaluate Giddings sample dimensionality parameters[33] is faced.

This procedure is experimentally checked on both synthetic and unknown mixtures. We also investigated how the ACVF approach can be improved when applied to SIM signal in GC/MS. In fact the SIM-ACVF combination seems very promising since SIM detection makes it possible to reduce the signal complexity by selectively monitoring specific chemical classes of compounds.[1,24,34]

THEORY

Autocovariance Function Method. The chemometric approach studies the autocovariance function that can be directly computed (experimental ACVF, EACVF) from the experimental chromatogram acquired in digitized form, using the following expression:[11,12]

$$\mathrm{EACVF}(\Delta t) = \frac{1}{M}\sum_{j=1}^{N-l}(Y_j - \hat{Y})(Y_{j+l} - \hat{Y}) \qquad l = 0,1,2...M-1 \quad (1a)$$

$$\Delta t = l\tau \quad (1b)$$

where Y_j is the digitized chromatogram signal, $\hat{Y}$ its mean value, M the truncation point in the EACVF computation, Δt the EACVF correlation time, and τ the time interval between the subsequent digitized positions. Δt assumes discrete values with l ranging from 0 to $(M - 1)$.

Theoretical expressions (theoretical ACVF, TACVF) have been developed to express ACVF in terms of the hidden separation parameters, i.e., number of SCs, m, SC peak standard deviation, σ, and retention pattern structure.[12] They require theoretical models to describe complex chromatograms: many functions can be developed to describe the infinity of real cases. Here two limit cases of retention patterns are considered, i.e., a Poissonian (P) distribution that describes a completely disordered separation where SC retention positions are uniform randomly distributed over the chromatographic axis and an ordered (O) distribution.[11,14] In the following, the simplest approach assuming chromatographic peaks of Gaussian shape with constant width,

(17) Pietrogrande, M. C.; Dondi, F.; Felinger, A. *J. High-Resolut. Chromatogr.* **1996**, *19*, 327−332.
(18) Pietrogrande, M. C.; Ghedini, D.; Velada, G.;. Dondi, F. *Analyst* **1998**, *123*, 1199−1204.
(19) Felinger, A.; Pietrogrande, M. C. *Anal. Chem.* **2001**, *73*, 618A−626A.
(20) Pietrogrande, M. C.; Coll, P.; Sternberg, R.; Szopa, C.; Navarro-Gonzalez, R.; Vidal-Majar, C.; Dondi, F. *J. Chromatogr., A* **2001**, *939*, 69−77.
(21) Pietrogrande, M. C.; Tellini, I.; Felinger, A.; Dondi, F.; Szopa, C.; Sternberg, R.; Vidal-Madjar, C.; Dondi, F. *J. Sep. Sci.* **2003**, *26*, 569−577.
(22) Pietrogrande, M. C.; Tellini, I.; Pasti, L.; Dondi, F.; Szopa, C.; Sternberg, R.; Vidal-Madjar, C.; Dondi, F. *J. Chromatogr., A* **2003**, *1002*, 179−192.
(23) Pietrogrande, M. C.; Tellini, I.; Szopa, C.; Felinger, A.; Coll, P.; Navarro-Gonzales, R.; Sternberg, R.; Vidal Madjar, C.; Raulin, F. *Planet. Space Sci.* **2003**, *51*, 581−590.
(24) Pietrogrande, M. C.; Zampolli, M. G.; Dondi, F. *Ann. Chim. (Rome)* **2004**, *94*, 721−731.
(25) Pietrogrande, M. C.; Zampolli, M. G.; Dondi, F.; Szopa, C.; Sternberg, R.; Buch, A.; Dondi, F. *J. Chromatogr., A* **2005**, *1071*, 255−261.
(26) Marchetti, N.; Felinger, A.; Pasti, L.; Pietrogrande, M. C.; Dondi, F. *Anal. Chem.* **2004**, *76*, 3055−3068.
(27) Pio, C. A.; Alves, C. A.; Duarte, A. C. *Atmos. Environ.* **2001**, *35*, 1365−1375.
(28) Brodskii, E. S. *J. Anal. Chem.* **2002**, *57*, 585−591.
(29) Hostettler, F. D.; Kvenvolden, K. A. *Environ. Forensics* **2002**, *3*, 293−301.
(30) Huang, H.; Larter, S. R.; Love, G. D. *Org. Geochem.* **2003**, *34*, 1673−1687.
(31) Ahmed, M.; George, S. C. *Org. Geochem.* **2004**, *35*, 137−155.
(32) Greenwood, P. F.; Arouri, K. R.; Logan, G. A.; Summons, R. E. *Org. Geochem.* **2004**, *35*, 331−346.
(33) Giddings, J. C. J. *J. Chromatogr.* **1995**, *703*, 3−15.
(34) Vékey, K. *J. Chromatogr., A* **2001**, *921*, 227−236.

i.e., constant standard deviation σ, will be discussed. This assumption is usually true under optimized programmed temperature conditions.[10,16] Also, the case of nonconstant peak width has been developed.[13,21]

Let us consider a disordered multicomponent chromatogram with m_P SCs (the suffix *P* refers to Poissonian retention pattern. The corresponding TACVF is given by[11]

$$\mathrm{TACVF_P}(\Delta t) = \frac{A_{\mathrm{T,P}}^2(\sigma_{\mathrm{h,P}}^2/a_{\mathrm{h,P}}^2 + 1)}{2\sqrt{\pi} m_P \sigma X} \mathrm{e}^{-[(\Delta t)^2/4\sigma^2]} \qquad (2)$$

where

$$A_{\mathrm{T,P}} = \sqrt{2\pi} m_P a_{\mathrm{h,P}} \sigma \qquad (3)$$

is the total area of the chromatogram, $a_{\mathrm{h,P}}$ and $\sigma_{\mathrm{h,P}}^2$ are, respectively, the mean and the variance of SC peak abundance (i.e., peak maximum height of a SC peak), and X is the total time range of the chromatogram.[11]

An ordered pattern in a multicomponent chromatogram is formed by a sequence of SC peaks: this is the case of the terms of an homologous series separated under optimized temperature programming conditions where the retention time of the nth term is described by

$$t_R(n) = c + bn \qquad n = 0,1,2...n_{\max} \qquad (4)$$

where $n_{\max}$ is the highest value of the SC number of the series, c represents the contribution of a specific functional group to the overall retention, and b is the retention increment between terms of the homologous series, e.g., the CH_2 retention time increment. In the most general experimental separations, where temperature programming conditions are far from the optimal ones (i.e., eq 4 is not strictly true), a proper linearization procedure can be applied to the experimental signal to transform the time axis into a new scale to make retention a homogeneous process yielding constant retention increments between subsequent terms of the homologous series.[21] If the term with $n = 0$ is absent in the considered homologous series, the number of SCs will be equal to $n_{\max}$. The constants c and b are called the phase and frequency indicators of the sequence, respectively, according to Gidding's definition.[33] In this case, the expression of TACVF can be obtained from ref 26, under the assumption that the SC term with $n = 0$ (see eq 4) is absent:

$$\mathrm{TACVF_O}(\Delta t) = \sum_{k=0}^{k = n_{\max} - 1} \frac{A_{\mathrm{T,O}}^2}{2\sigma\sqrt{\pi} X (n_{\max} - k)} \left(\frac{\sigma_{\mathrm{h,O}}^2}{a_{\mathrm{h,O}}^2} + 1 \right) \mathrm{e}^{-[(\Delta t - bk)^2/4\sigma^2]} \qquad (5)$$

$$A_{\mathrm{T,O}} = \sqrt{2\pi} n_{\max} a_{\mathrm{h,O}} \sigma \qquad (6)$$

where $A_{\mathrm{T,O}}$ is the total area of the ordered chromatogram, $a_{\mathrm{h,O}}$ and $\sigma_{\mathrm{h,O}}^2$ are respectively the mean and the variance of SC peak abundance (i.e., peak maximum height of a SC peak) in the O-type chromatogram. Note that this expression does not contain the phase c, and thus, information concerning the sequence phase is lost: the ACVF retains only the recursivity of an ordered structure. According to eq 5, the $\mathrm{TACVF_O}$, and therefore the $\mathrm{EACVF_O}$ plot, shows well-defined Gaussian peaks of standard deviation equal to $(2\sigma)^{1/2}$, located at interdistances bk, corresponding to repeated interdistances between terms of the homologous series (eq 5). These peaks are called *deterministic* since they reflect the order of the sequence; their height decreases on k, but their shape is independent of k. The subscript O identifies that an ordered retention pattern is present in the multicomponent chromatogram.

The developed models (eqs 2 and 5) are independent of the distribution of the SC abundance (abundance model, AM): such a function yields different values of the ratio σ_h^2/a_h^2, called SC peak height dispersion. Here two limit examples of abundance patterns are considered: the constant AM, where peak heights are nearly constant yielding $\sigma_h^2/a_h^2 = 0$, and the exponential AM, where SC abundances are randomly distributed around the mean value a_h value yielding $\sigma_h^2/a_h^2 = 1$.[11−14] Note that the last distribution is also the most probable one among the all possible distributions compatible with a given number of total SCs, since it contains the maximum entropy.[35]

The original complete procedure is based on the fitting of EACVF to TACVF to obtain information on sample complexity—number of components, m—and on the separation system, mean peak width, σ.[11−16] A simplified procedure based on simple computation on EACVF and graphical inspection of the EACVF plot has also been developed to obtain the same information.[17−25] In this paper, this procedure, developed for generally random chromatograms, is extended to ordered chromatograms and to compound chromatograms containing both Poissonian and ordered components.

Decoding Random-Type Multicomponent Chromatograms. In the case of Poissonian retention pattern, the first region in the $\mathrm{EACVF_P}$ plot ($0 < \Delta t \leq 4\sigma$), according to eq 2, is expected to be half of a Gaussian peak of standard deviation equal to $(2\sigma)^{1/2}$, showing a shape averaged on the shape of all the peaks present in the chromatogram. From $\mathrm{EACVF_P}$ half-height peak width, $d_{h/2}$, the mean peak standard deviation can be simply estimated:[17]

$$\sigma = d_{h/2}/1.665 \qquad (7)$$

From the value of $\mathrm{EACVF_P}$ at the origin ($\Delta t = 0$), it is possible to estimate the number of SCs of the chromatogram, m_P, by rearranging eq 2:[17]

$$m_P = \frac{A_{\mathrm{T,P}}^2(\sigma_{\mathrm{h,P}}^2/a_{\mathrm{h,P}}^2 + 1)}{\mathrm{EACVF_P}(0) d_{h/2} 2.129 X} \qquad (8)$$

In this equation, the quantities $A_{\mathrm{T,P}}$ and X can be determined from the experimental chromatogram and $d_{h/2}$ can be determined over the experimental $\mathrm{EACVF_P}$ plot (eq 7). On the contrary, the exact value of the quantity $\sigma_{\mathrm{h,P}}^2/a_{\mathrm{h,P}}^2$ cannot be experimentally determined from the chromatogram due to SC peak overlapping. However, $\sigma_{\mathrm{h,P}}^2/a_{\mathrm{h,P}}^2$ can be approximated by the $\sigma_{\mathrm{M,P}}^2/a_{\mathrm{M,P}}^2$ value,

(35) Dondi, F.; Bassi, A.; Cavazzini, A., Pietrogrande, M. C. *Anal. Chem.* **1998**, *70*, 766−773.

which is the peak maximum dispersion ratio, i.e., the dispersion ratio computed from the observed peak maximums in the chromatogram:

$$\sigma^2_{h,P}/a^2_{h,P} \approx \sigma^2_{M,P}/a^2_{M,P} \tag{9}$$

Consequently m_P can be estimated by using the following equation:

$$m_P = \frac{A^2_{T,P}(\sigma^2_{M,P}/a^2_{M,P} + 1)}{\mathrm{EACVF}_P(0)d_{h/2}2.129X} \tag{10}$$

The quantity m_P is related to the Poisson distribution, and its standard deviation is known to be equal to $m_P^{1/2}$. Consequently, eq 10 yields an estimate $m_{est,P} = m_P \pm m_P^{1/2}$. The approximation error connected to eq 9 has been discussed elsewhere.[14]

Equations 9 and 10 can be used not only in the case of a strictly Poissonian retention pattern but also to other cases of random patterns.[10−17] A useful form of eq 10 can be obtained if the total area $A_{T,P}$ is expressed as

$$A_{T,P} = m_P A_{m,P} \tag{11}$$

where $A_{m,P}$ is the average area of SC peaks in the "random" multicomponent chromatogram. By combining eqs 10 and 11, one obtains

$$m_P = \frac{\mathrm{EACVF}_P(0)2.129d_{h/2}X}{A^2_{m,P}(\sigma^2_{M,P}/a^2_{M,P} + 1)} \tag{12}$$

Equation 12 clearly shows the direct proportionality between m_P and $\mathrm{EACVF}_P(0)$ and m_P under conditions of constancy of $A_{m,P}$ value.

Decoding Ordered Multicomponent Chromatograms. Equations similar to eqs 10 and 12 can be derived even in the case of O multicomponent chromatograms. By using the same assumptions employed in deriving eqs 10 and 12, one obtains

$$n_{max} - k = \frac{A^2_{T,O}(\sigma^2_{M,O}/a^2_{M,O} + 1)}{\mathrm{EACVF}_O(bk)d_{h/2}2.129X} \tag{13a}$$

$$n_{max} - k = \frac{\mathrm{EACVF}_O(bk)2.129d_{h/2}X}{A^2_{m,O}(\sigma^2_{M,O}/a^2_{M,O} + 1)} \tag{13b}$$

where the quantity $A_{n_m,O}$ is the mean area of SC peaks defined by

$$A_{T,O} = n_{max}A_{m,O} \tag{14}$$

and $\sigma^2_{M,O}/a^2_{M,O}$ is the peak maximum dispersion ratio. The different peaks in the EACVF_O plot will be well distinct provided that $b > 4\sigma$ (see eq 5). In this case, the SC number n_{max} can be determined by using the peak at the origin, in analogy with eq 12:

$$n_{max} = \frac{A^2_{T,O}(\sigma^2_{M,O}/a^2_{M,O} + 1)}{\mathrm{EACVF}_O(0)d_{h/2}2.129X} \tag{15a}$$

The first deterministic peak can also be used and in this case one has

$$n_{max} = \frac{A^2_{T,O}(\sigma^2_{M,O}/a^2_{M,O} + 1)}{\mathrm{EACVF}_O(b)d_{h/2}2.129X} + 1 \tag{15b}$$

Consequently, an experimental check of the ordered character of the chromatogram can be obtained by comparing the n_{max} values estimated from eqs 15a and b. Note that the $A_{T,O}$, X, and $d_{1/2}$ values are experimentally accessible parameters.

Decoding Compound Multicomponent Chromatograms with One Homologous Series. If a complex mixture is formed by one homologous series of n_{max} SCs (ordered sequence, eq 4) and of an ensemble of uncorrelated m_P SCs, (random component), the total number of SCs, m_{tot}, and the total area of the compound multicomponent chromatogram will be given by

$$m_{tot} = m_P + n_{max} \tag{16a}$$

and

$$A_{T,tot} = A_{T,P} + A_{T,O} \tag{16b}$$

respectively. A method to estimate m_{tot}, m_P, b, and n_{max} from the EACVF_{tot} plot of the compound chromatogram, under specific conditions, will be described. In analogy with eqs 10 and 12, let us put the following equations (to be proved):

$$m_{tot} = \frac{A^2_{T,tot}(\sigma^2_{M,tot}/a^2_{M,tot} + 1)}{\mathrm{EACVF}_{tot}(0)d_{h/2}2.129X} \tag{17}$$

$$m_{tot} = \frac{\mathrm{EACVF}_{tot}(0)2.129d_{h/2}X}{A^2_{m,tot}(\sigma^2_{M,tot}/a^2_{M,tot} + 1)} \tag{18}$$

where the quantity $A_{m,tot}$ is the mean area of SC peaks defined by

$$A_{T,tot} = m_{tot}A_{m,tot} \tag{19}$$

and $\sigma^2_{M,tot}/a^2_{M,tot}$ is the peak maximum, dispersion ratio in the compound chromatogram. Equation 18 can be obtained by combining eq 13b with $k = 0$, eqs 12 and 16a, and further by assuming that

$$\mathrm{EACVF}_{tot}(0) = \mathrm{EACVF}_P(0) + \mathrm{EACVF}_O(0) \tag{20}$$

$$A_{m,tot} \approx A_{m,P} \approx A_{m,O} \tag{21}$$

$$\sigma^2_{M,tot}/a^2_{M,tot} \approx \sigma^2_{M,P}/a^2_{M,P} \approx \sigma^2_{M,O}/a^2_{M,O} \tag{22}$$

Equations 21 and 22 mean that SCs in both the ordered and the

Poissonian components of the multicomponent chromatogram have equal average peak areas and peak height dispersion ratios, respectively. Equation 20 expresses the rule of the variance additivity for independent variables, remembering that EACVF(0) has the meaning of a variance (see eq 1a). This condition holds true for the present case since the Poissonian part of the compound chromatogram is completely random and thus not correlated with the ordered one. Consequently, m_{tot} can be estimated from $EACVF_{tot}$ by using eq 17, under the conditions described by eqs 21 and 22.

The n_{max} value, and therefore also the m_P value, can be obtained from $EACVF_{tot}$ (see eq 16a), under given conditions. In fact, let us assume the following property of the EACVF (to be discussed in the following):

$$EACVF_{tot}(kb) = EACVF_O(kb) \quad \text{for} \quad k > 0 \qquad (23)$$

Equation 13b together with the conditions expressed by eqs 20−23 becomes:

$$n_{max} - k = \frac{EACVF_{tot}(bk)2.129d_{h/2}X}{A^2_{m,tot}(\sigma^2_{M,tot}/a^2_{M,tot} + 1)} \qquad (24)$$

$$\text{conditions: eqs 20, 21−23;} \quad b \geq 4\sigma$$

Equation 24 means that that n_{max} can be evaluated from $EACVF_{tot}$, even if under strict conditions. The condition $b \geq 4\sigma$ means that the SC peaks belonging to the homologous series are each other sufficiently resolved in the chromatogram (eq 4) and their correlation does not interfere with that inside a SC component peak. In this case, the first and the subsequent deterministic peaks of the $EACVF_O$, i.e., the $EACVF_O(kb)$ peaks for $k > 0$, fall well separated from the origin ($\Delta t = 0$), beyond $\Delta t = 4\sigma$ (see eq 5). Under these conditions, eq 23 also holds true since the $EACVF_P \approx 0$, in the region $\Delta t \geq 4\sigma$, and the $EACVF_{tot}$ plot is only made by the $EACVF_O$ (eq 20, see refs 11−15 for a full discussion).

A simple form of eq 24 can be obtained for $k = 1$, by combining eq 18 and eq 24 for

$$n_{max} = m_{tot}\frac{EACVF_{tot}(b)}{EACVF_{tot}(0)} + 1 \qquad (25)$$

$$\text{conditions: eqs 20, 21−23;} \quad b \geq 4\sigma$$

The conditions expressed by eqs 21 and 22 require a comment since they are not very common in the practice. In particular, the hypothesis expressed by eq 21—i.e., the average SC abundance of the SCs belonging to the total mixture and that to a given homologous series are the same—seems critical: in fact, it is usually the condition that one homologous series can be either predominant or in trace with respect to the majority of the other SCs. Otherwise, eq 22 can be more or less met in practice since it establishes that the degree of randomness on the SC peak height dispersion ratio is similar for either the random component or the ordered component (homologous series). Therefore, the case when eq 21 does not hold true will be developed in the following. By combining eq 17 or 18 with eq 13a or 13b, respectively, for $k = 1$, $b \geq 4\sigma$ and under the conditions described by eqs 20−23, one can obtain

$$n_{max} = m_{tot}\frac{EACVF_{tot}(0)}{EACVF_{tot}(b)} \times \frac{A^2_{T,o}}{A^2_{T,tot}} + 1 \qquad (26a)$$

$$\text{conditions: eqs 20, 21−23;} \quad b \geq 4\sigma$$

and

$$n_{max} = m_{tot}\frac{EACVF_{tot}(b)}{EACVF_{tot}(0)} \times \frac{A^2_{m,tot}}{A^2_{m,O}} + 1 \qquad (26b)$$

$$\text{conditions: eqs 20−23;} \quad b \geq 4\sigma$$

respectively. Consequently, in the general case, when the ratio of the SC area of structured class with respect to that of the totality of the SCs is unknown, the sole EACVF ratio and m_{tot}—obtained from eq 17—cannot yield a quantitative estimate of SC number of the homologous series but only an "apparent" SC number. Equations 26a and b suggest that a useful strategy of study can be the use of selective detectors combined to universal ones. In fact they make it possible to selectively detect specific compound classes, and thus, the contribution of the different classes can be decoded from the total mixture and quantitatively estimated. A point deserves a comment: as a consequence of the factor $A^2_{T,o}/A^2_{T,tot}$ in eqs 26a and b, the EACVF(b) will result increased or decreased in the case of an homologous series present as a predominant or a trace component, respectively. The detection of an homologous series as a minor component will be unfavorable. This comes from two concomitant effects: the $EACVF_{tot}(b)$ is very low because it is inversely proportional to $A^2_{T,o}$ and it can be lost in the random background of the $EACVF_{tot}$ around b, which can be significantly high in the cases of low m_{tot} values.[11−14] Consequently, the topic needs specific experimental investigation.

Decoding Compound Multicomponent Chromatograms with More Than One Homologous Series. The possibility of obtaining the SC number of the ordered component or its relative fraction with respect to the total SC number, m_{tot} under eqs 20−23 conditions (see eqs 24 and 25) can be immediately extended to the case of more than one homologous series, all displaying the same frequency b value. The retention time positions for each homologous series i will be given by

$$t_{R,i}(n_i) = c_i + bn_i \qquad n_i = 0,1,2...n_{max,i} \qquad i = 1,2..i_{max} \qquad (27)$$

where i_{max} is the number of homologous series in the ordered component of the compound multicomponent chromatogram. This case is relevant since it can correspond to a multicomponent mixture made by several linear homologous series (linear hydrocarbons, alcohols, ketones, etc.) all the with common CH_2 increment: the b quantity (frequency) will represent the retention increment due to CH_2 whereas the c_i quantity (phase) will represent the effect of the functional group. The total number of SCs of the Ordered component to be introduced in

eqs 26a and b will be

$$n_{max} = \sum_{i=1}^{i=i_{max}} n_{max,i} \quad (28)$$

Consequently, the deterministic peak at b in the EACVF will give information on the ordered component, using eqs 24−26b. Under favorable conditions, it is also possible to obtain information on the existence of the different homlogous series (i.e., i_{max}), their identity, and the number of components belonging to the different series ($n_{max,i}$), under the eqs 20−23 conditions. In fact, additional peaks lower than the predominant b peak will appear in the EACVF plot (eq 30): they reflect the repetitivities in the multicomponent chromatogram due to the mutual interaction among the different homologous series present in the mixture. In particular, the repetitivities are equal to the differences between the different c_i values, $\pm\Delta c_j$, with $j = 1,...j_{max}$, where j_{max} is equal to all the distinct combinations with which the phase differences can be chosen, i.e., $j_{max} = \binom{i_{max}}{2}$. For example, in the case of two homologous series, there is only one Δc_j, value, i.e., $\Delta c = |c_1 - c_2|$. In the case of three series, there will be three combinations: $\Delta c_1 = |c_1 - c_2|$; $\Delta c_2 = |c_2 - c_3|$; $\Delta c_3 = |c_1 - c_3|$ (with $\Delta c_3 = \Delta c_1 - \Delta c_2$). Consequently, the number of homologous series, i_{max}, with common frequency value (e.g., the above referred frequent case of addition of CH_2 in several different homologous series) can be obtained by the multiplicity of the main b peak in the EACVF plot. In fact, the Δc_j terms sum and subtract to b to give the observed multiplicity (b):

$$\text{multiplicity}(b) = 2\binom{i_{max}}{2} + 1 \quad (29)$$

Also, the value of the phase differences among the different homologous series can be estimated from the EACVF peaks at $\pm\Delta c_j$, under the condition that the Δc_j peaks are each well resolved. Such $\pm\Delta c_j$ values, under well-defined chromatographic conditions, are diagnostic for specific chemical structures of the terms of the homologous series.[23,25]

Moreover, the heights of these cluster peaks due to the phase differences Δc_j give information on the quantitative composition of the mixture: in fact they will be proportional to the number of repeated retention time differences concurring to determine them. This last property derives from expressions similar to eq 23 or 24 (under eqs 20−23 conditions), which hold true also for repetitivities of type $\pm k\Delta c_j$ in analogy with those of type kb. The $\pm k\Delta c_j$ peaks close to the kb peak will be lower than the central kb peak. In the favorable case of two series with equal number of homologous terms, the two twinning peaks will be half with respect to the central peak (e.g., Figure 4a, b in ref 26), allowing us to assign the number of SCs of each homologous series.

The Δc_j peaks will be distinctly detected, and thus, the O component of the compound chromatogram will be fully decoded, provided that their values do not interfere with each other, be greater than 4σ, and the number of the repeated retention time differences be significant. Moreover, the Δc_j pattern complexity around the main b peak in the EACVF will increase (see eq 29) with the increasing number of homologous series, i_{max}, and under such conditions that the possibility of overlapping among the Δc_j peaks will increase. This phenomenon is also associated with mutual overlapping among the SCs belonging to the different series. This effect will eventually suggest the need to increase the dimensionality of the separation system: the structure of the EACVF plot is related to the concepts of sample dimensionality and separation dimensionality introduced by Giddings.[33] The EACVF appears as a powerful tool for determining dimensionality parameters of an unknown mixture.

If the frequency values of two series are different, i.e., if $b_1 \neq b_2$, the resulting chromatogram and, therefore, the EACVF plot, will show an even more complex pattern. This is due to the interactions between the various c_i and b_i terms. The point lies beyond the aim in the present study and it will not be considered in detail here. Nonetheless, the EACVF plot in one- or in two-dimensional separations is, in principle, able to decode this increased complexity.[26]

Table 1. Organic Compounds Present in the Standard Mixture[a]

Hydrocarbons	
2,3-dimethylpentane	*n*-octane
2-methylhexane	1-pentene
3-methylhexane	cyclohexane
isooctane	methylcyclopentane
Alcohols	
1-propanol	1-heptanol
1-butanol	1-octanol
1-pentanol	1-nonanol
1-hexanol	
Ketones	
2-propanone	2-heptanone
2-butanone	2-octanone
2-pentanone	2-nonanone
2-hexanone	
Esters	
isopropyl acetate	propyl acetate
ethyl propionate	butyl acetate
Aromatics	
benzene	propylbenzene
xylene	butylbenzene
ethylbenzene	

[a] Chromatogram in inset in Figure 2b.

EXPERIMENTAL SECTION

GC/MS analyses were performed on standard and unknown mixtures of organics. Standard mixtures contained organic compounds with the number of carbon atoms ranging from 3 to 11: hydrocarbons, alcohols, ketones, esters, and aromatics (the chemical structures of the studied compounds are listed in Table 1). An unknown sample was a petroleum benzin (bp 140−160 °C) delivered from Fluka (Milan, Italy). A Mega Series 5160 gas chromatograph (Fisons Instruments, Milan, Italy) was coupled with a QMD1000 quadrupole mass spectrometer (Fisons Instruments). The column used was a RTX-20 column (L = 10 m, i.d. 0.25 mm, d_f 1 μm) (Restek, Bellafonte, PA). The analyses were performed under temperature-programmed conditions: 30 °C for 2 min and then an increase to 80 °C at 5 °C/min, which has been found to be the optimal temperature programming conditions.[21−23] The carrier gas was helium at a flow rate of 1.2 mL/min. Split conditions (1:400 split ratio) were used for injection (injection

temperature, 220 °C; injected sample, 1 μL of mixture). The mass spectrometer operated in EI mode (positive ion, 70 eV); mass spectra were acquired with repetitive scanning from 20 to 400 m/z in 1 s.

COMPUTATION

All the programs are written in Fortran and run on a personal computer 2-Ghz (512 RAM) Pentium III. The EACVF was numerically calculated from the digitized chromatogram, according to eq 1a. The developed algorithm concerns the calculation of some separation parameters reported in eqs 10, 13, and 24:[12] the total area of the chromatogram, A_T, was computed by numerical integration. The number of peak maximums in the chromatogram p was detected by using an algorithm based on the comparison of five successive points and a threshold level (1% of the highest peak) to filter out the noise. From these values, the average peak maximum abundance, a_M, and its standard deviation, σ_M^2, were computed and introduced in eqs 10, 13, and 24 to substitute the true theoretical value a_h and σ_h^2/a_h^2.[17] The number of SCs, m_{tot}, computed according to eq 18 was reported as $m_{tot} \pm m_{tot}^{1/2}$ because of the Poissonian character of this variable.

DISCUSSION

The theoretical treatment proved that the EACVF of a compound multicomponent chromatogram—i.e., a multicomponent chromatogram formed by both the ordered and the Poissonian components (or only an ordered component made by several homologous series)—shows a series of peaks due to the components of the different homologous series, in addition to the main peak at the origin. By studying the EACVF plot, a complete decoding of both the ordered and the Poissonian component can be achieved under well-defined conditions, expressed by eqs 20−23: the most severe condition is that expressing the equality of the SCs average quantities of the ordered and Poissonian components (eq 21). The peak at the origin allows one to determine the total number of SCs (m_{tot}) (eqs 13a, 15a, and 17); the deterministic peaks at different Δt positions correspond to the SCs belonging to homologous series characterized by this specific retention increment (eqs 13a, 15a, and 25). Cluster peaks around a main deterministic peak at given Δt correspond to number and type of different homologous classes with common frequency (eq 29), and their heights are related to the number of SCs belonging to the different classes (eq 25).

In the practice, we can consider the EACVF plot as a "class chromatogram", since it makes it possible a "separation, identification, and quantification by classes" of the mixture components, under well-defined conditions and strategies. The reliability of the method must be tested in the practice in order to single out its advantages and eventual limits with respect to general conditions of the experimental analysis. In this paper, the EACVF method was checked on three types of mixtures.

Mixtures 1. This first group of mixtures were formed only by ordered components, i.e., a known number of SC terms of homologous series with a known concentration and abundance distribution. Several cases were considered, containing either only one or several homologous series. By this way, the ability of the method was checked in determining the total number of SCs (eqs 13a, 15a, and 15b,) and identifying the different homologous series and the number of SCs belonging to them (eqs 15a and b, 28, and 29).

Mixtures 2. These were standard mixtures made by an ordered component plus some uncorrelated SCs of known abundance quantity and distribution. This type of mixture contained both ordered and Poissonian components displaying SC abundance values always following the assumptions described by eqs 21 and 22, under different conditions. With these mixtures, we checked the ability of the method in determining the total number of SCs (eq 17), those separately belonging to both the ordered and the Poissonian components (eq 17, 25, and 16a), and identifying the different homologous series (eq 29).

Mixture 3. This comprises petroleum benzin (bp 140−160 °C), which is an unknown mixture of hydrocarbon compounds containing 7−10 carbon atoms.[32,37] Likely, this type of mixture contained an unknown but limited number of SCs (hydrocarbons) with both an ordered and a Poissonian component. It is thus an example of a compound multicomponent mixture for which the conditions expressed by eqs 20−23 are not a priori known.

The power of the method was enhanced by coupling it with the MS selective detection: it was applied to GC/MS signals obtained in the SIM detection mode. The SIM chromatograms were investigated since they selectively retain, in a simplified signal, information on class of compounds with specific chemical structures.[24,31,32] This approach was applied to the case of mixtures 2 and mixture 3.

Method Validation on Chromatograms of Standard Mixtures Formed Only by Ordered Components (Mixtures 1). The method was validated by application to chromatograms obtained from standard mixtures containing a known number of components. They were formed by hydrocarbons and oxygenated compounds with the number of carbon atoms ranging from 3 to 11, including compounds belonging to different homologous series (chemical structures in Table 1). They are some of the basic compounds used to characterize petroleum products;[29−32,36] in particular, oxygenated compounds are the chemical signature of atmospheric oxidative processes or microbial degradation and their identification is fundamental for the assessment of contamination sources.[27,28,30] Analyses were performed in optimized temperature-programmed conditions to obtain a linear retention pattern, i.e., a constant retention contribution of CH_2 addition in the terms of a homologous series (b term in eq 4).[33]

The reliability of the simplified approach (eqs 13 and 25) to estimate the fraction of components belonging to homologous series was tested. The mixtures were prepared by mixing proper concentrations of standard compounds so that all the components of the ordered sequences display nearly the same mean peak area values (eq 21) and the same peak maximum dispersion ratio (eq 22). In an unknown mixture, no "a priori" information is available on the SC peak height distribution: it can be "a posteriori" estimated by the value experimentally computed for $\sigma_{M,tot}^2/a_{M,tot}^2$.

As an example of an homologous series, the chromatogram of a series on 1-alcohols (1-propanol to 1-nonanol, mixture 1a) is reported in Figure 1: it looks like an ordered sequence of seven peaks located at a constant interdistance, 2.8 min (i.e., the frequency b of the series, eq 4). This ordered structure is singled

(36) Varotsis, N.; Guieze, P. *J. Pet. Sci. Eng.* **1996**, *15*, 81−89.
(37) McLafferty, F. W.; Turecek, F. *Interpretation of Mass Spectra*; University Science Books: Sausalito, CA, 1993.

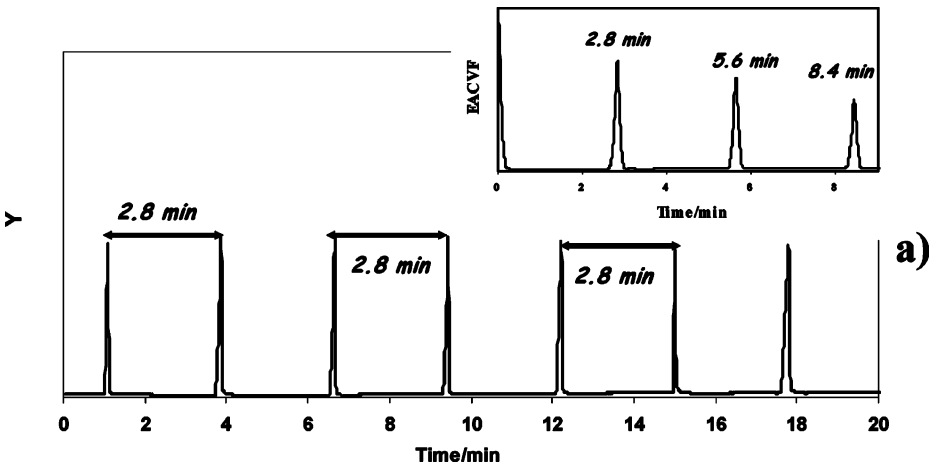

Figure 1. Chromatogram of a series on 1-alcohols (1-propanol to 1-nonanol, mixture 1a). Inset: EACVF plot computed on the chromatogram.

out by the EACVF plot (inset in Figure 1a, solid line) computed on the chromatogram: well-shaped deterministic peaks are present at 2.8 min and multiple values at 5.6 and 8.4 min, due to repeated interdistances (see eq 5). This result means that, under these chromatographic conditions, the EACVF peaks at 2.8, 5.6, and 8.4 min are diagnostic for the presence of a sequence of a homologous series.

To study the effect of combination of different ordered sequences, a standard mixture (mixture 1b) containing three homologous series (j_{max} =3) was analyzed at the same temperature-programmed conditions (TIC chromatogram in Figure 2a): the mixture contains seven 1-alcohols (C_3-C_9), seven ketones (C_3-C_9), and five aromatics (benzene to butylbenzene) (Table 1). This mixture represents the simplest case where SC peak height is described by a constant function, i.e., $\sigma_h^2/a_h^2 = 0$, and the conditions described by eqs 21 and 22 hold true for this sample. The EACVF was computed on the experimental chromatogram and its plot reported in Figure 2b (lower trace). The EACVF plot shows deterministic peaks at 2.8 and 5.6 min, revealing the presence of the homologous series; in fact, under these chromatographic conditions, they correspond to the frequency b common for all the series. From the EACVF value of the deterministic peak at 2.8 min, the number of terms of the homologous series can be estimated by using eq 13a ($k = 1$); the obtained values is $n_{max} = 19$ that exactly corresponds to the real SC number in the sample. The feature of the EACVF plot in the region close to the main peak at 2.8 min retains information on the chemical composition of the mixture: the lower deterministic cluster peaks are due to the combinations of the different phase c_i values for the three homologue classes (eq 29).[22] Specific peaks diagnostic for the chemical classes studied can be identified: $\Delta c_1 = 0.5$ min is the peak interdistance between aromatics and ketones, $\Delta c_2 = 0.6$ min between ketones and alcohols, and $\Delta c_3 = 1.7$ min between alcohols and aromatics (see arrows in Figure 2a and b). It can be seen that the plot is symmetric versus the symmetry axis at 1.4 min, i.e., 2.8/2 min; the symmetric cluster peaks at $b \pm \Delta c_j$ are also evident in the plot, i.e., deterministic peaks at 2.3, 2.2, 2.7, 3.3, 3.4, and 3.9 min, respectively.

Method Validation on Chromatograms of Standard Mixtures Formed by an Ordered Component plus Some Uncorrelated SCs (Mixtures 2). The EACVF approach was applied to samples containing compounds with different uncorrelated structures, in addition to the previous samples formed by terms of homologous series. This is a way to investigate the ability of the EACVF to extract information on ordered sequences even if they are hidden in the signal complexity.

As a first example, a mixture containing 31 organic compounds was analyzed at the same temperature-programmed conditions (mixture 2a, TIC chromatogram in inset in Figure 2a). The mixture was prepared so that the assumptions described by eqs 21 and 22 can be a priori verified. The condition concerning mean peak area (eq 21) was achieved by mixing proper concentrations of standard compounds in the mixture so that nearly the same mean area was obtained from the whole mixture and from the ordered component, i.e., $A_{m,tot} \approx A_{n_{max},O}$. The condition on distribution of SC peak height (eq 22) can be a posteriori verified by computing $\sigma_{M,tot}^2/a_{M,tot}^2$ and $\sigma_{m,O}^2/a_{m,O}^2$ quantities from the chromatograms obtained from the whole mixture and the mixture containing terms of the ordered sequences; the obtained values are 0.08 and 0.03 for $\sigma_{M,tot}^2/a_{M,tot}^2$ and $\sigma_{m,O}^2/a_{m,O}^2$, respectively, which means a constant distribution of SC peak height. The EACVF was computed on the experimental chromatogram (plot in upper trace of Figure 2a). From EACVF at $\Delta t = 0$, the number of components is correctly estimated as $m_{tot} = 31 \pm 6$ (eq 18).

The EACVF plot computed on the complex mixture (31 components, upper trace in Figure 2b) shows a very similar feature compared to that of the ordered mixture (19 components, lower trace in Figure 2b). Therefore, a simple visual inspection of the EACVF plot makes it possible to extract information on the chemical composition of the mixture since it retains information on the ordered components of the complex chromatogram. From the main deterministic peak at 2.8 min, the number of terms of the homologous series can be estimated by using eq 13 ($k = 1$); the obtained values is $n_{max} = 19$, which exactly corresponds to the real number of terms of homologous series in the sample. The lower EACVF peaks at $\Delta c_1 = 0.5$ min, $\Delta c_2 = 0.6$ min, and $\Delta c_3 = 1.7$ min, and their combination with 2.8 min, are diagnostic for the presence of three homologous series (alcohols, aromatics, ketones) since they are due to the combinations of the different phase c_i values of the three sequences (eq 29). Moreover, the height values of these peaks are related to the number of components belonging to each class (eq 26a); from the EACVF at 0.6 min, i.e., the interdistance between alcohols and ketones, it is computed that seven terms of these classes are present in the

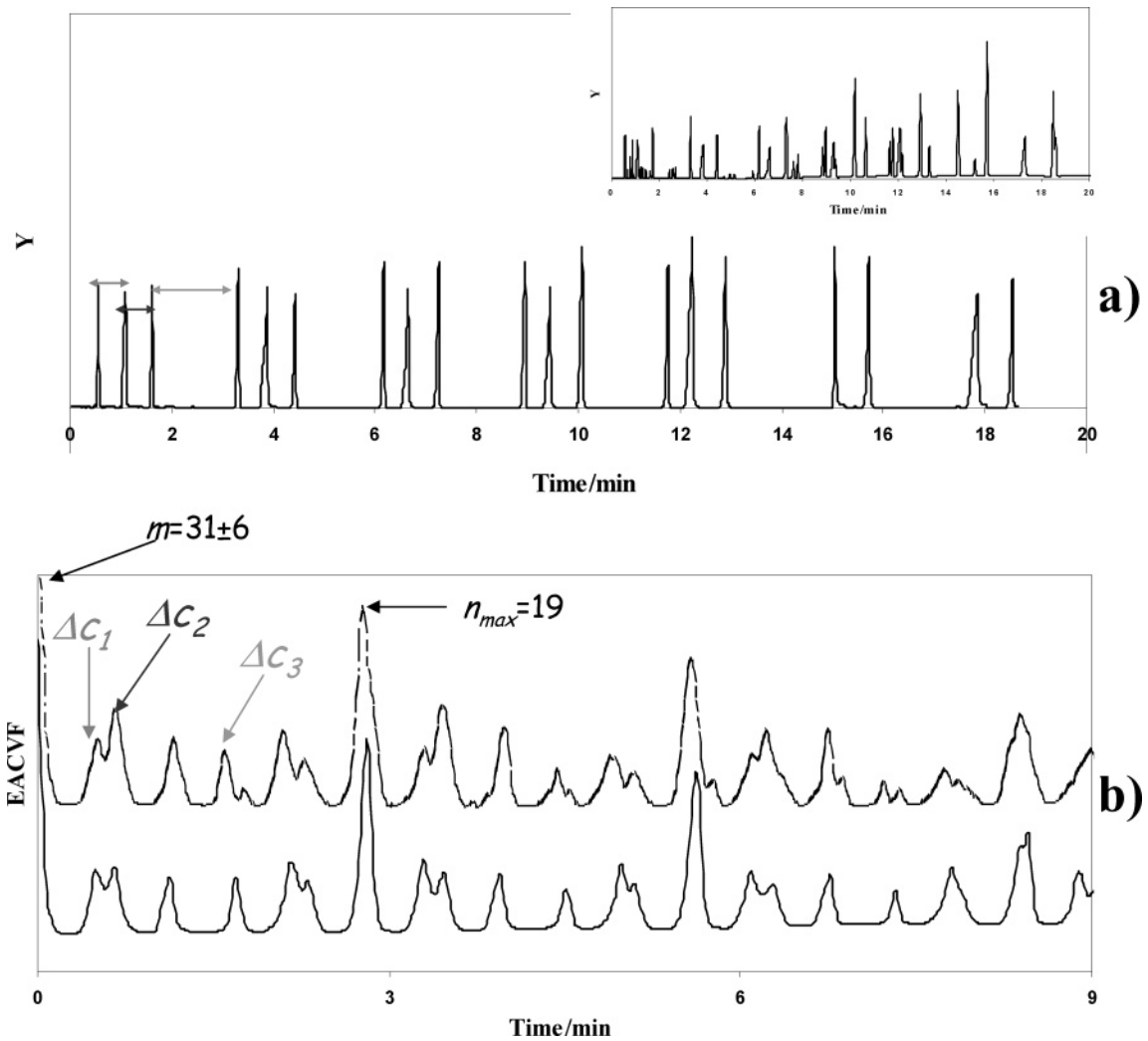

Figure 2. GC/MS chromatograms of complex mixtures with constant AM (mixture 1b). (a) TIC chromatogram of a mixture of 1-alcohols (C_3−C_9), ketones (C_3−C_9), and aromatics (benzene to butylbenzene). Inset: TIC chromatogram of a mixture containing 31 organic compounds, including 1-alcohols (C_3−C_9), ketones (C_3−C_9), and aromatics (benzene to butylbenzene). (b) EACVF plot computed on the TIC chromatograms: lower trace, mixture containing 1-alcohols, ketones, and aromatics; upper trace, mixture containing 31 organic compounds. Constant interdistances between classes are signed by colored arrows: Δc_1, aromatics−ketones in red, Δc_2, ketones−alcohols in blue, and Δc_3, alcohols−aromatics in green.

mixture. The interdistances for the pairs ketones−aromatics and alcohols−aromatics are 0.5 and 1.7 min, respectively; the EACVF computed at these values, EACF (0.5 min) ≈ EACF (1.1 min) correspond to the presence of five aromatic compounds. Indeed, some differences between the EACVF plots of the compound (upper trace in Figure 2b) and the ordered mixture (lower trace in the same figure) can be singled out; they may be due to other correlated molecular structures present in the mixture. Such interferences of different sequences are lost in a random retention pattern of a complex mixture containing a high number of components (higher than 50);[11,35] the studied mixture does not strictly fulfill this condition of complexity (number of components lower than 50).

A check of the obtained results and extraction of additional information on the complex chromatogram can be obtained from the MS signal. In particular, by studying the mass spectra of the sample components, it can be possible to select specific ion fragments to be used in SIM acquisition to selectively detect only terms of a given homologous series.[1,7,24,31−34] As an example, for the studied sample (TIC chromatogram in inset of Figure 2a), it was possible to acquire SIM chromatograms showing only terms of each series (Figure 3a): m/z 31 (CH_2OH^+) is the characteristic fragment for primary alcohols, m/z 43 (CH_3CO^+) for ketones, and m/z 78 + 91 (C_6H_6, $C_6H_5CH_2^+$ fragments) for aromatics.[24] The EACVF plots (Figure 3b) computed on the SIM chromatograms single out the order (deterministic peaks at 2.8 and 5.6 min) and make it possible to estimate the number of terms of each homologous series (n_{max}) from the EACVF value at $\Delta t = 2.8$ min (eq 26a). The number of components is accurately estimated for each series: $n_{max} = 7$ for primary alcohols (m/z 31), $n_{max} = 7$ for ketones (m/z 43), and $n_{max} = 5$ for aromatics (m/z 78 + 91). Moreover, the SIM chromatograms can be used to check the assumption of constant AM of the terms of the homologous series; in fact, the $\sigma^2_{m,O}/a^2_{mO}$values computed on the SIM signal represent only the terms of the ordered sequences. This assumption was verified since the $\sigma^2_{m,O}/a^2_{m,O}$ values obtained for primary alcohols, ketones, and aromatics were 0.03, 0.02, and 0.02, respectively. These results also show a good agreement with the

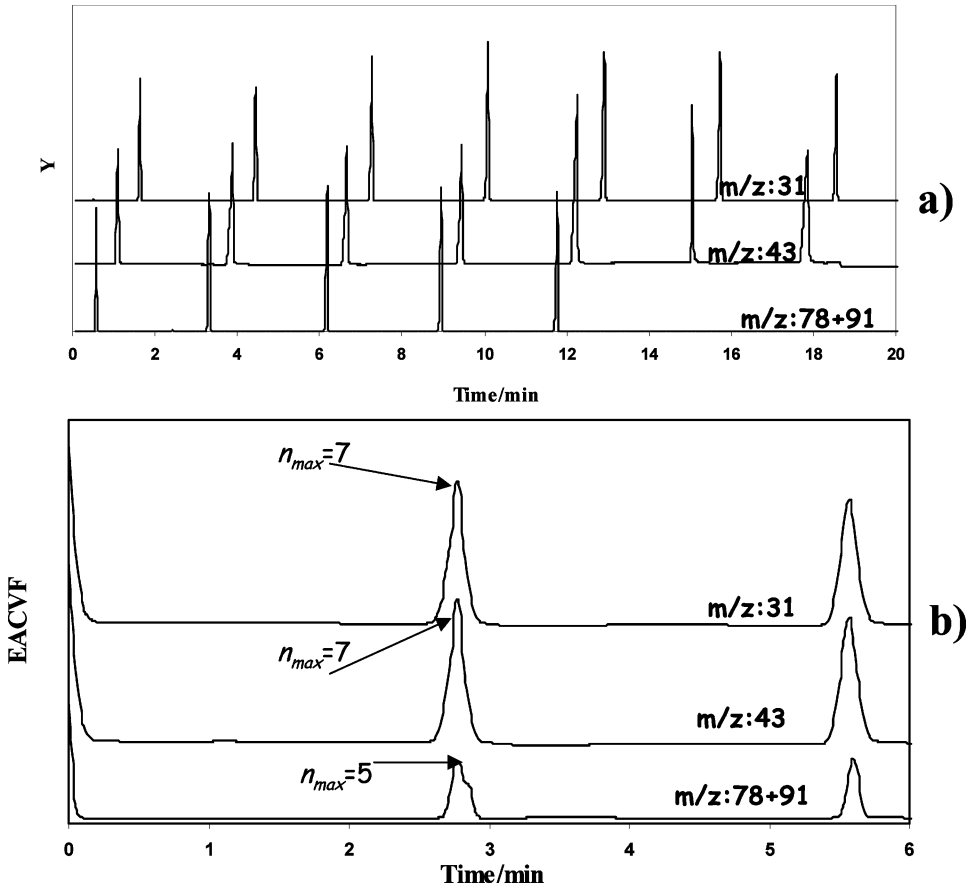

Figure 3. SIM chromatograms of a mixture containing 31 organic compounds, including 1-alcohols, ketones, and aromatics (mixture 2a). (a) SIM chromatograms at $m/z = 78 + 91$ (aromatics), $m/z = 43$ (ketones), and $m/z = 31$ (1-alcohols); (b) EACVF plots computed on the SIM chromatograms.

data separately obtained from the TIC chromatogram of each standard mixtures: $\sigma^2_{m,O}/a^2_{m,O} = 0.03$. These results are based on the assumption that the SIM method displays the same relative sensitivity as the TIC signal to yield a correct estimation of $\sigma^2_{m,O}/a^2_{m,O}$ value.[37] In this specific case, this is only an acceptable approximation. A rigorous investigation of this topic lies beyond the aims of this study.

In the second example studied (mixture 2b), the sample components display an abundance distribution described by an exponential function; the theoretical peak height dispersion ratio, σ^2_h/a^2_h, is 1 and an experimental $\sigma^2_{M,tot}/a^2_{M,tot}$ value ~1 is measured. It is the most frequent case since it has been theoretically demonstrated and experimentally verified that it is the "limit" distribution for the maximum complexity of the mixture.[35] The studied standard mixture contains 39 organic compounds; 7 components are terms of the homologous series of ketones (C_3-C_9), in addition to other uncorrelated compounds. The ketone concentration was chosen to yield exponentially distributed peak heights of all the components, i.e., $\sigma^2_{M,tot}/a^2_{M,tot} \approx 1$, with a mean peak area close to the value obtained from all the SCs, i.e., $A_{m,tot} \approx A_{n_{max},O}$ (eq 21). The TIC chromatogram obtained under temperature-programmed conditions is reported in inset in Figure 4a. The EACVF was computed on the experimental chromatogram and its plot reported in Figure 4a. Its feature simply singles out the presence of an homologous series identified by the presence of the deterministic peaks at 2.8 and 5.6 min. The total number of components can be estimated from EACVF at $\Delta t = 0$ (eq 18) as $m_{tot} = 39 \pm 6$. The fraction of them belonging to the homologous series is computed by using eq 26a; the obtained value is $n_{max} = 7$, which exactly corresponds to the real number of ketones in the sample. This result proves the reliability of the simplified approach based on eq 26a, in this general case of exponential AM, if the condition $A_{m,tot} \approx A_{n_{max},O}$ holds true. The results obtained can be confirmed by studying the simplified signal obtained from the complex mixture by the SIM detection mode. By selecting a proper m/z value (43, fragment CH_3CO^+), only the peaks corresponding to ketones are selectively shown (Figure 4b). On this chromatogram, the assumption of the exponential AM can be verified by computing $\sigma^2_{m,O}/a^2_{m,O} = 1.02$. Moreover, the number of the components of the series, n_{max}, can computed from the EACVF value at $\Delta t = 2.8$ min (eq 26a) (deterministic peak in inset in Figure 4b). The obtained result, $n_{max} = 7$, shows the reliability of the procedure.

Application to Chromatograms of the Unknown (Mixture 3). The applicability of the method was tested on a sample containing an unknown number of organics. A petroleum benzin (bp 140−160 °C) was submitted to GC/MS analysis (TIC chromatogram in Figure 5a), and the EACVF was computed on the experimental chromatogram (Figure 5b, upper trace), assuming a threshold level of 1% of the highest peak. From the EACVF(0), it is possible to estimate the SC number present in the mixture: $m_{tot} = 49 \pm 7$ (eq 17). The EACVF plot (Figure 5b, upper trace) shows a well-defined deterministic peak at $\Delta t = 2.8$ min that can be diagnostic to identify the presence of homologous series

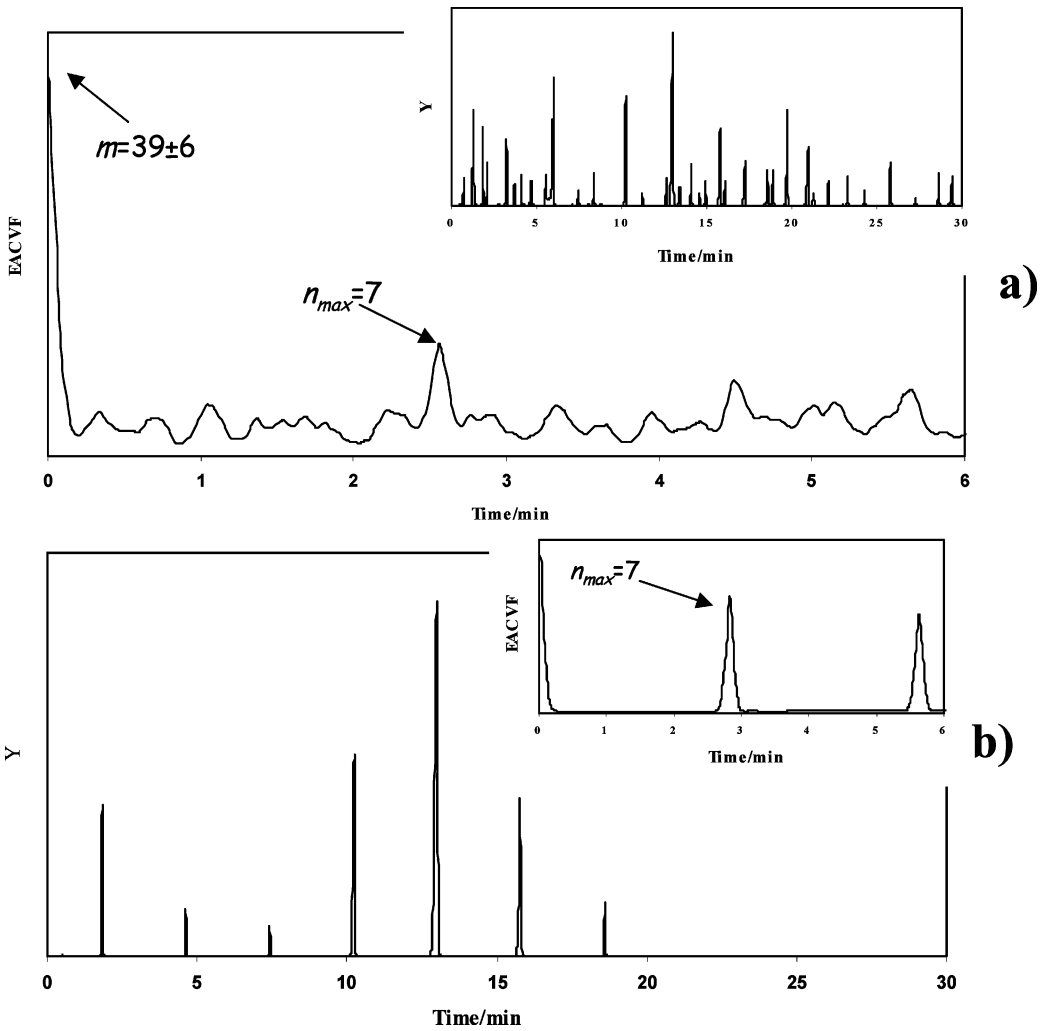

Figure 4. GC/MS chromatogram of a standard mixture (mixture 2b) containing 39 organic compounds with exponential AM, including ketones (C_3-C_9). (a) EACVF plot computed on the TIC chromatogram. Inset: TIC chromatogram; (b) SIM chromatogram at $m/z = 43$ (ketones). Inset: EACVF plot computed on the SIM chromatogram.

displaying a CH_2 group increment, i.e., of an ordered component together with a Poisson component. However, for this unknown mixture, no conditions expressed by eqs 21 and 22 can be a priori assumed. In this case, the estimate of $n_{max} = 18$ for the ordered component, obtained by applying eq 25 to the EACVF at $\Delta t = 2.8$, should be considered only an "apparent" number of terms of the homologous series, since the true value should be obtained from eq 26a, which requires the a priori knowledge of $A_{m,tot}/A_{n_{max}}$.

The petroleum benzin mixture was further investigated under SIM conditions operating at $m/z = 57 + 71 + 85$ (fragments $C_4H_9^+$, $C_5H_{11}^+$, and $C_6H_{13}^+$) (inset in Figure 5a); it mainly corresponds to select SC acyclic alkanes, both straight chain and branched compounds, containing 7−10 carbon atoms.[32,37] Information on the SC peak height distribution for all the acyclic alkanes can be verified a posteriori by computing the quantities $\sigma^2_{M,tot}/a^2_{M,tot}$ on the TIC and SIM chromatograms, respectively, assuming that the TIC and SIM signals display the same relative sensitivities. The obtained values were 1.02 (TIC) and 1.03 (SIM), respectively, i.e.

$$(\sigma^2_{M,tot}/a^2_{M,tot})_{TIC} \cong (\sigma^2_{M,tot}/a^2_{M,tot})_{SIM} \cong 1$$

They indicate that the abundance distributions are described by the same exponential function (expected value of 1) for the two detection conditions. This finding is not surprising since one must remember that the exponential distribution is the most probable one.[15] Consequently, by changing the detection conditions from TIC to SIM, the SC distributions look similar.

The EACVF was computed on the SIM signal (lower trace in Figure 5b); from the EACVF(0) value, the total number of acyclic alkanes $m_{tot} = 27 \pm 5$ is estimated (eq 17), significantly lower than that obtained under TIC detection conditions,[49] as expected. In fact, mixture 3 contains cyclic, aromatic, and other SCs.

By using eq 25, a value $n_{max} = 17$ is obtained from the EACVF at $\Delta t = 2.8$ min calculated on the SIM chromatogram (Figure 5b, lower trace, bold numbers). This result should be considered an "apparent" value, as stated above. However, this value shows an excellent agreement with that obtained under TIC detection ($n_{max} = 18$); this suggests that this may be a true value, since it is obtained under very different conditions of universal and selective detection and from the recursive retention properties (CH_2 increment) singled out by EACVF. This is proved in the following. In fact the "true" n_{max} value is unique and one should obtain unique true n_{max} values (eq 26b) under both TIC and SIM detection

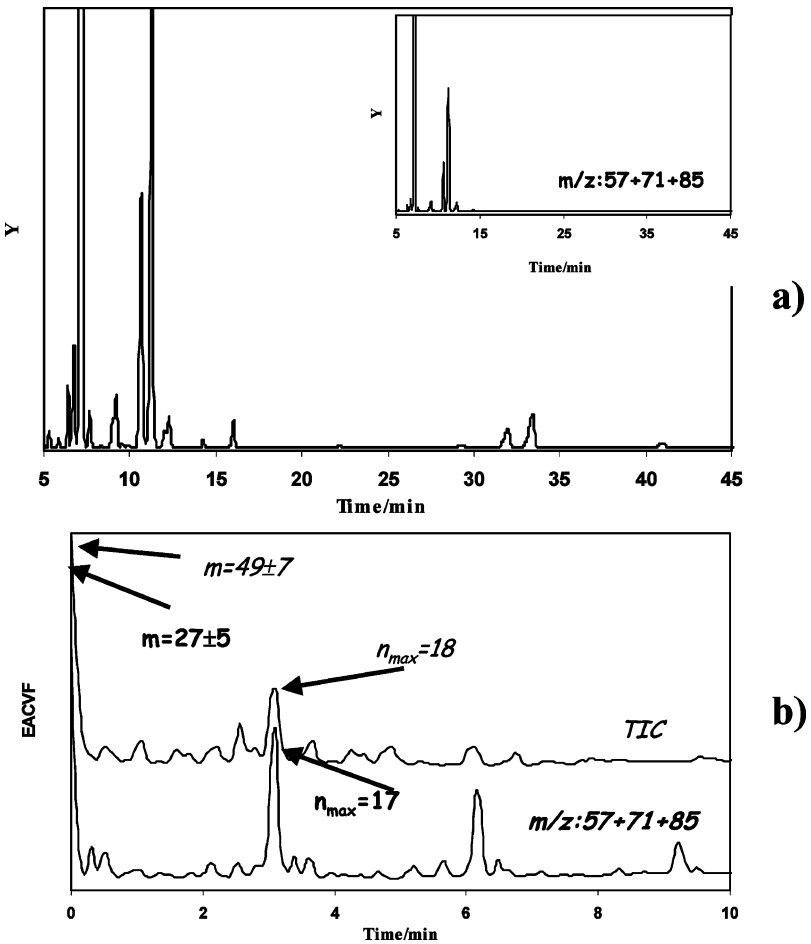

Figure 5. GC/MS chromatograms of a petroleum benzin (bp 140−160 °C, mixture 3). (a) TIC chromatogram. Inset: SIM chromatogram at $m/z = 57 + 71 + 85$ (*n*-alkanes). (b) EACVF plots computed on the SIM chromatogram (lower trace, values in bold) and the TIC chromatogram (upper trace).

conditions. Under these conditions, from eq 26b one should have

$$\left(\frac{\mathrm{EACVF}_{\mathrm{tot}}(b)}{\mathrm{EACVF}_{\mathrm{tot}}(0)} \times \frac{A^2_{\mathrm{m,tot}}}{A^2_{\mathrm{m,O}}}\right)_{\mathrm{TIC}} = \left(\frac{\mathrm{EACVF}_{\mathrm{tot}}(b)}{\mathrm{EACVF}_{\mathrm{tot}}(0)} \times \frac{A^2_{\mathrm{m,tot}}}{A^2_{\mathrm{m,O}}}\right)_{\mathrm{SIM}}$$

Moreover, since nearly equal values of $n_{\max}$ were obtained from eq 25, as above seen, one has

$$\left(\frac{\mathrm{EACVF}_{\mathrm{tot}}(b)}{\mathrm{EACVF}_{\mathrm{tot}}(0)}\right)_{\mathrm{TIC}} \cong \left(\frac{\mathrm{EACVF}_{\mathrm{tot}}(b)}{\mathrm{EACVF}_{\mathrm{tot}}(0)}\right)_{\mathrm{SIM}}$$

and thus, by comparing these two last conditions, one obtains

$$\left(\frac{A_{\mathrm{m,tot}}}{A_{\mathrm{m,O}}}\right)_{\mathrm{TIC}} \cong \left(\frac{A_{\mathrm{m,tot}}}{A_{\mathrm{m,O}}}\right)_{\mathrm{SIM}}$$

Because, as above pointed, the dispersion ratios under TIC and SIM detection were the same, all these findings are likely compatible only when $A_{\mathrm{m,tot}}/A_{\mathrm{m,O}} \approx 1$, since the m_{tot} values were significantly different in the two detection conditions ($m_{\mathrm{tot}} = 49$ and 27, for TIC and SIM, respectively, see above). Consequently, the eq 21 and 22 conditions hold true and the $n_{\max}$ values estimated from eq 25 should be true.

The number of homologous series in mixture 3 can be inferred from the EACVF lower deterministic peaks in the region close to the main peak at 2.8 min, according to eq 29. The symmetry axis at 1.4 min, i.e., 2.8/2 min, is evident in both the plots. Even if the noisy EACVF background can obscure other features, the symmetric peaks at $b \pm \Delta c_{\mathrm{i}}$, due to the combinations of the different phase c_{i} values for the ordered sequences (eq 29) make it possible to identify at least three different homologue classes with uncorrelated chemical structure present in the mixture.

The above-reported findings constitute only an example of the richness of information that can be obtained from the present approach. Some of the above-reported findings call for additional systematic investigation on both SIM chromatograms and EACVF. Likewise, a deeper investigation of the specific effects of signal noise, the increase on column peak capacity, application of proper linearization procedures should be done in order to make sharper the correlation features and single out additional hidden features;[22] this requires a specific study and lies beyond the aim of the present one.

CONCLUSION

The simplified EACVF procedure developed here proves to be a simple data processing method to efficiently handle a multicomponent chromatogram in order to characterize the chemical composition of the complex sample. The method is

particularly powerful in identifying the presence of an ordered sequence of compounds, singling it out from the complexity of the disordered chromatogram; the two components, ordered and disordered, can be separated and quantitatively evaluated; that is, number of compounds of each pattern can be estimated. Such information can be extracted by handling the simple FID signal, without any information on the chemical structure of the components. Moreover, the power of the method is significantly magnified if combined with the SIM detection. In fact, the EACVF essentially singles out structure retention correlation of thermodynamics origin, whereas SIM provides further selectivity to the method related to selected molecular structures.

The information obtained by the method makes it possible to analytically characterize the sample not only in terms of identification and quantification of selected SCs but also of identification and quantification of the specific SC homologous series building up the total mixture. Therefore, the present procedure seems to be not merely a powerful chemometric tool for handling complex chromatograms but also a new approach for a comprehensive characterization of a complex multicomponent mixture. In fact, the EACVF plot can be really considered as a "SC class chromatogram" giving a separation, identification, and quantification by classes, which is additional information, compared to the overall and sometime indistinct sequence of overlapping peaks.

At present, the application of the developed procedure is limited by fulfilling the severe conditions of the SC concentration, in particular, the average abundance of the SCs belonging to the homologous series compared to that of the total mixture. However, the good data obtained from the benzin sample and the possibility to check their availability by handling the SIM signal seem to be a very promising result concerning the applicability of this simple method to unknown real samples. It is clear that the present procedure must be extended to the general, most usual condition, where the concentration of the homologous series is different from that of the majority of the other SCs. For this aim, further theoretical development and application to real cases are under study.

Moreover, the method robustness toward experimental limitations is under study: how the procedure is powerful in overcoming problems related to experimental chromatograms obtained in unfavorable conditions acquisition, such as nonlinear temperature-programming conditions and noisy signals. Another limit of the present procedure may be the high concentration of the sample components—yielding overloading effects with consequent peak shape distortion—and a wide concentration range (over several orders of magnitude) of the detectable components—EACVF may be mostly affected by the predominant components obscuring the least abundant compounds.

At present, the procedure has been tested on a limited number of classes of compounds, but it is obvious that it is general and can be extended to different classes of compounds, if specific fragments for SIM detection are selected, and to more complex mixtures. Finally, the present results obtained from studies of 1D chromatograms can be the starting point for their extension to 2D multicomponent separations, whose 2D-ACVF was proved to exhibit useful properties for decoding highly complex multicomponent mixtures.[26]

ACKNOWLEDGMENT

This work has been supported by the Italian Space Agency (Grant IR/054/02), the Italian University and Scientific Research Ministry (Grants 2003039537_005, 2005032388_004), and by the University of Ferrara, Italy.

GLOSSARY GLOSSARY

*2*D ACVF	two-dimensional autocovariance function
ACVF	autocovariance function
a_h	mean SC peak height
a_M	mean peak maximum
A_m	mean area of the SC peaks in the chromatogram, eq 14
AM	abundance model, i.e., constant and exponential
$A_{m,O}$	mean area of the SC peaks of ordered sequences
A_T	total area of the chromatogram
b	frequency of the homologous series, eqs 4, 27
c	phase of the homologous series, eqs 4, 27
C	constant (abundance model)
$d_{h/2}$	half-height peak width, eq 7
E	exponential (abundance model)
EACVF	experimental autocovariance function, eq 1a
h	SC peak height
i_{max}	maximum number of homologous series in a compound chromatogram, eq 27
j_{max}	maximum number of distinct combinations between phases of homologous series in a compound chromatogram, eq 28
l	discrete interdistance on which EACVF is computed, eq 1b
k	interdistance order between SC positions in an ordered sequence
m	number of SCs
M	maximum Δt interdistance extension, over which EACVF is computed, eq 1a
m_{tot}	number of SCs in compound multicomponent chromatogram, i.e., random + ordered components, eq 16
n_{max}	maximum number of the term in a given homologous series, eq 4
O	ordered
P	Poissonian
SC	single component
TACVF	theoretical autocovariance function, eqs 2,5
t_R	retention time
x	retention time coordinate
X	retention time range of the chromatogram
$\hat{Y}$	mean value of the signal of the digitized chromatogram
Y_j	signal of the digitized chromatogram
Greek Symbols	
Δ	difference

Δc_i	difference between c_i phases of homologous series in a compound chromatogram
Δt	time interdistance between subsequent points in the chromatogram
σ	standard deviation of a chromatographic peak
σ_M^2	variance of peak maximum distribution
σ_h^2	variance of SC peak abundance/height
σ_h^2/a_h^2	SC peak abundance/height dispersion ratio
σ_M^2/a_M^2	peak maximum dispersion ratio
τ	time interval between two subsequent digitized positions, eq 1b

Suffixes (if not specified)

i	index of each homologous series in a compound chromatogram, eq 28
j	index of distinct combination between phases of homologous series in a compound chromatogram
n	index of terms of an homologous series, eq 4
O	ordered retention pattern
P	Poissonian retention pattern
tot	compound multicomponent chromatogram, i.e., Poissonian + ordered components, eq 16a

Received for review August 18, 2005. Accepted December 29, 2005.

AC051491E