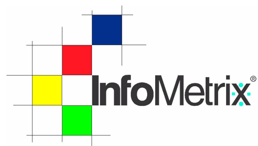

Aligning Chromatograms

Brian G. Rohrback and L. Scott Ramos
Infometrix, Inc., Bothell, WA

Abstract

A common problem in chromatography is retention time instability. There are many physical mechanisms that are employed to mitigate retention time shift: careful selection and replacement of columns, sample pretreatment, improvements in flow control, or temperature programming precision. If more precise control of the instrument parameters is not practical, there are software tools for aligning data that can be applied depending on the method: these tools make use of internal standards, external standards and target chromatograms. This poster outlines how each can be used to adjust automatically for retention time shift in both upstream and downstream petroleum applications.

Introduction

A well-known limitation of chromatography is non-reproducibility of retention times. For some applications, this issue may not be of great importance, but for most analysts it is a significant concern. With complex chromatograms, the peak assigning algorithm built in to most chromatographic software may fail when retention times of key peaks have shifted out of the expected time range. Instrument companies have tried to circumvent the problem with techniques such as retention time locking, which uses dynamic pressure control to try to match retention in the current run to that of a defined target run. However, the technique does not produce exactly matching times and cannot be applied to historical data. In addition, any form of chromatography that uses a liquid mobile phase (*e.g.,* liquid chromatography, thin layer chromatography, capillary electrophoresis), cannot be easily controlled in the pressure/flow regime.

An example of a sample requiring alignment is drawn from a refining application in which chromatographic drift is frequently encountered. The following figure shows two chromatograms of similar material where many peaks have shifted significantly in the second sample with respect to those in the first.

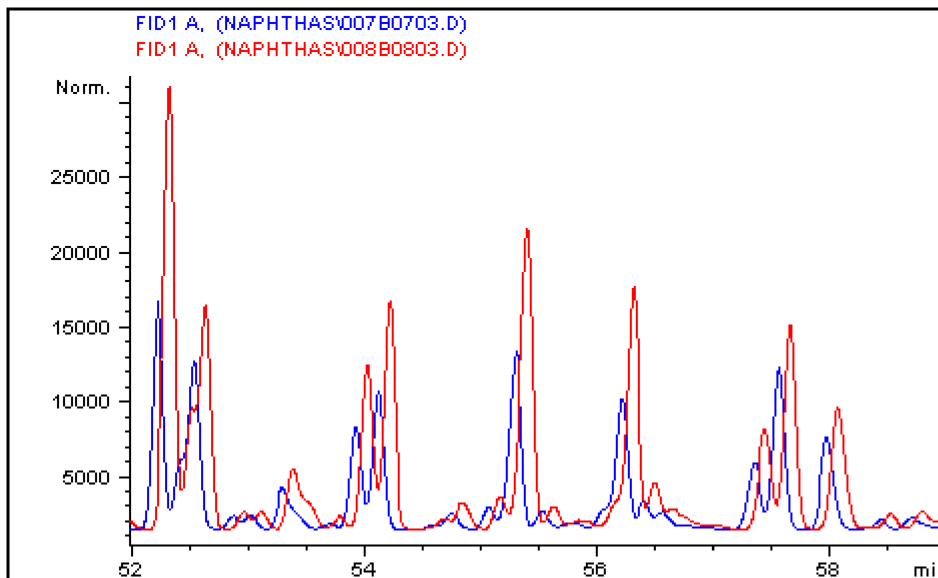

Figure 1. Two chromatographic traces of naphthas showing retention time variability

Following alignment of the sample to the target (the blue trace above is the target), the misalignment has been almost completely eliminated, as seen in the next figure.

© 2003, Infometrix, Inc.

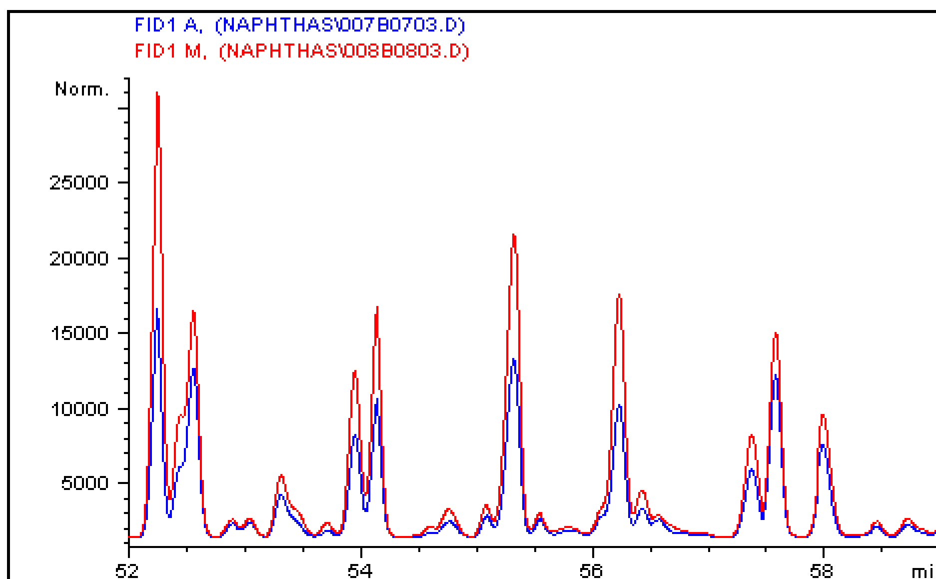

Figure 2. Adjusting the red chromatogram using LineUp corrects for retention time shift

There are two traditional and one new software methods for doing this alignment. All have the advantage of being amenable to automation (*i.e.,* no user intervention is required during routine processing). A disadvantage is that these approaches can disguise systematic drift and in turn hide larger problems.

Internal Peak Indexing. Incorporated into every chromatographic software system is the ability to adjust peak identifications based on one or more internal standards. Although most of these systems do not have the ability to adjust the chromatographic trace itself to reflect "standard time", this adjustment is available in some third party software (such as Pirouette®). In most cases, internal standards are added to the sample prior to injection and are chosen such that they bracket the peaks that are of most interest. Typically, the more internal standards employed the better the alignment.

Advantages	Disadvantages
Internal standards are subject to precisely the same chromatographic conditions as all other peaks.	The chromatographic software MUST find the retention markers.
Processing is very fast (seconds).	If there is too much drift, you need to recalibrate with an external standard.

External Peak Indexing. External standards are also in use to align chromatograms. In this case, a standard is run either before the analysis, or just after, to provide a check on peak assignments. Although not commonly practiced, the shift in the external standard peaks can be monitored to shift the analyte chromatogram in a similar way.

Advantages	Disadvantages
Ensures a consistent suite of marker peaks.	Most chromatographic software does not support the automated re-indexing of peaks.
Processing is very fast (seconds).	The standard must be run frequently and is limited by the run-to-run variability of retention time.

External Pattern Matching. A new software-based approach, using multivariate alignment algorithms, can minimize the deviations in chromatograms that are not in alignment. In this case, the algorithm aligns a sample to one or more defined target chromatograms by maximizing the correlation of one chromatogram to another*. This means that the sample and the target must be reasonably similar in

order for this procedure to work.

Advantages	Disadvantages
Requires no a priori method development, no internal or external peak standards.	Requires that you have a similar chromatogram to act as a gold standard.
Alignment failures can be flagged automatically.	Processing time from a few seconds to a minute; extreme chromatographic shifts increase processing time.

EXPERIMENTAL

Samples in this study were gathered from several sources: upstream and downstream; on-line and in-lab chromatography systems. HPLC results are from an Agilent ChemStation, and the GC data were originally processed from Agilent ChemStation or from Scientific Software's EZChrom *Elite*™. Data processing was performed using Pirouette® (for internal and external peak indexing) or with LineUp™ software (for external pattern matching), both from Infometrix, Inc. Comparisons of results were performed using Matlab® (The MathWorks, Inc.).

RESULTS AND DISCUSSION

The most critical issue with chromatographic data when used for any type of routine analysis is to insure that chromatograms from different injections appear as if there has been little or no shift in retention from run to run. Because non-reproducibility in retention is common in real world chromatographs, some form of processing is required to minimize the effects of this lack of reproducibility.

The Easy Case. Some sources of chromatography data have a sufficient number of readily identified marker peaks such that retention indices (*e.g.,* Kovats index) can be used. In the first example, crude oils from a mixed-source reservoir are to be processed to quantify changes in production composition over time. Chromatograms in this type of study are often collected over a period of years, which makes reliance on the chromatography hardware to keep things in alignment practically impossible. Fortunately, because the oils are not biodegraded, we have a series of markers that can be used as alignment standards.

Each of these data forms requires special handling to achieve alignment. Thus retention indices, for either the internal standard or the external standard approach, require simultaneous or companion analysis of sets of marker compounds. Kovats indices require the simultaneous analysis of a series of n-alkanes, but other indices can be derived which serve the same purpose. In either case, the absolute retention times are modulated to a scale in which, ideally, the same compound eluted from different samples will have identical (modulated) time or time index. Few chromatography systems include the ability to compute retention indices (Kovats or otherwise), but all will compute relative retention times based on a calibration or peak ID table.

Alignment in this well-controlled environment is shown below using the n-alkanes as internal standards.

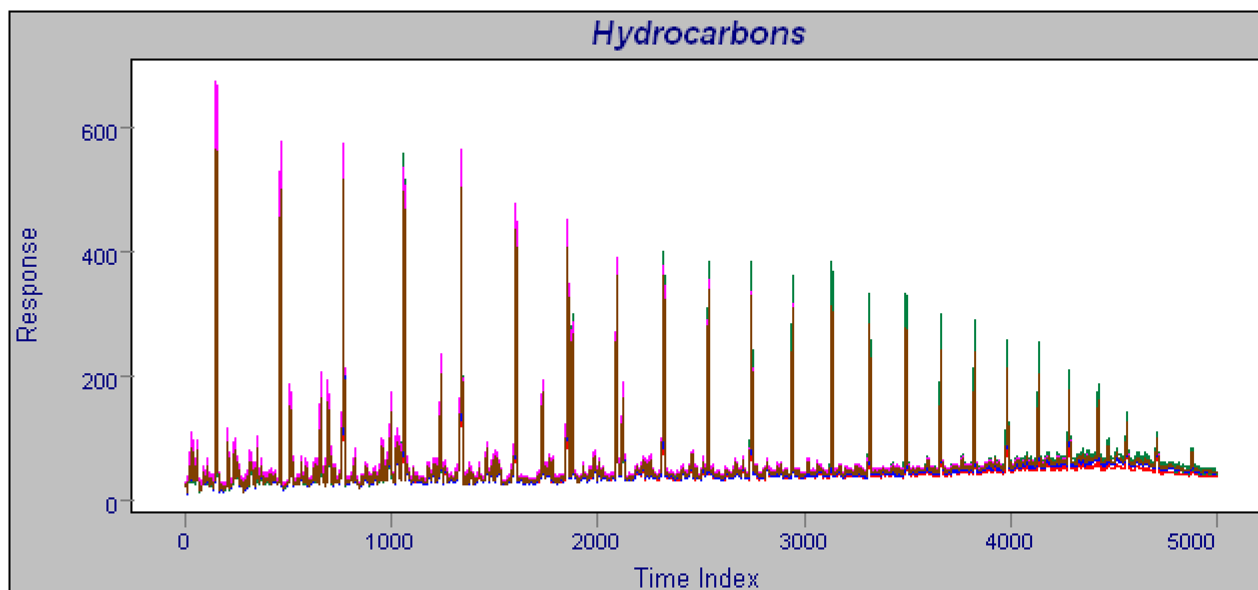

Figure 3. Correcting retention time shift using the Kovats (internal peak indexing) approach; full chromatogram.

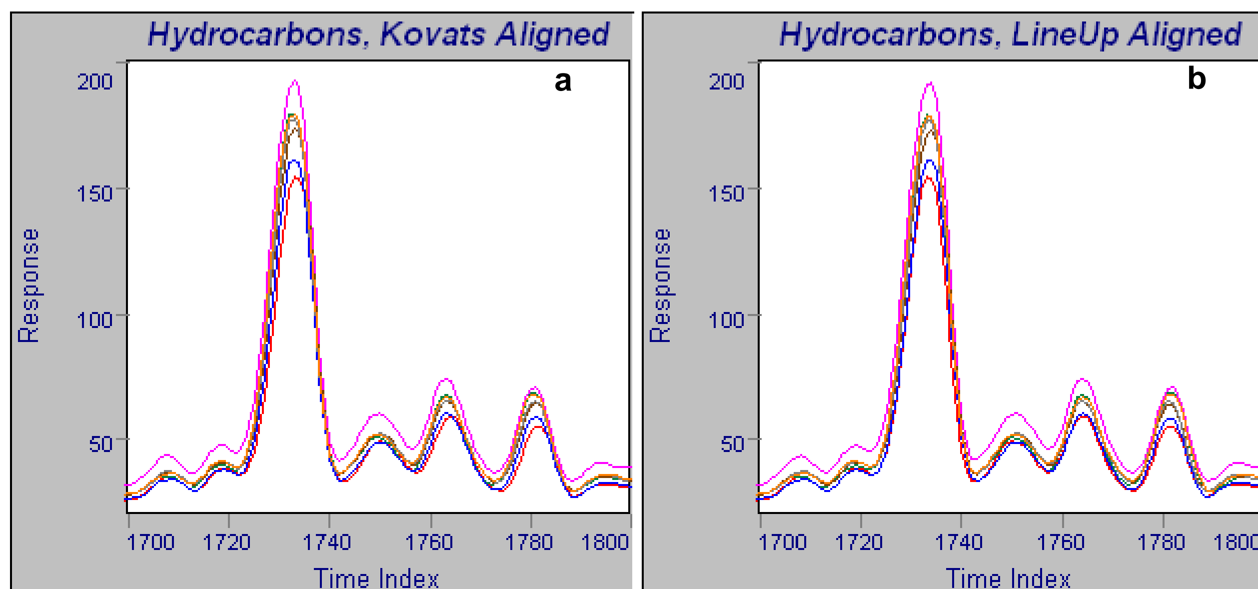

Figure 4. (a) Kovats approach, zoomed region between C_{16} and C_{17}. (b) Correcting retention time shift using LineUp (external pattern matching); zoomed region between C_{16} and C_{17}.

A similar result (Figure 4b) is obtained without requiring the identification of the alkanes by using the external pattern matching algorithm.

More Difficult. We often do not have the luxury of having retention markers spread evenly throughout the chromatogram. When possible, we can use standards in regions without much peak activity to form the basis of alignment. One example of HPLC data is shown below.

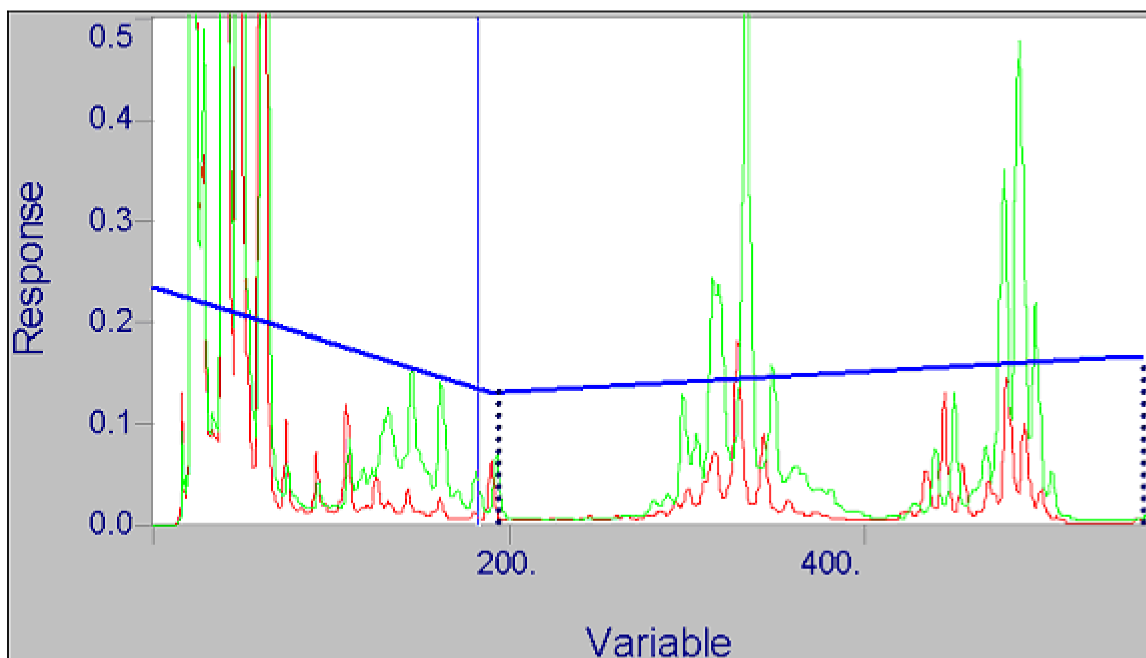

Figure 5. Setting up internal standards to bracket the region of interest; dotted lines point to positions of two standards

In this example a series of fifteen chromatograms were collected. These data were bracketed by QC samples such that they were processed in three sets over a two-day period. Due to the design, we can compare the approaches of external peak markers, internal peak markers and external pattern matching for their accuracy and precision of alignment.

Overlaying the raw chromatograms, we see that the retention time variability, even for these successive runs, is quite high (Figure 6a). There appears to be a uniform drift as well as a random error component. The resulting alignment (Figure 6b) in all cases reflects the near elimination of retention time variation.

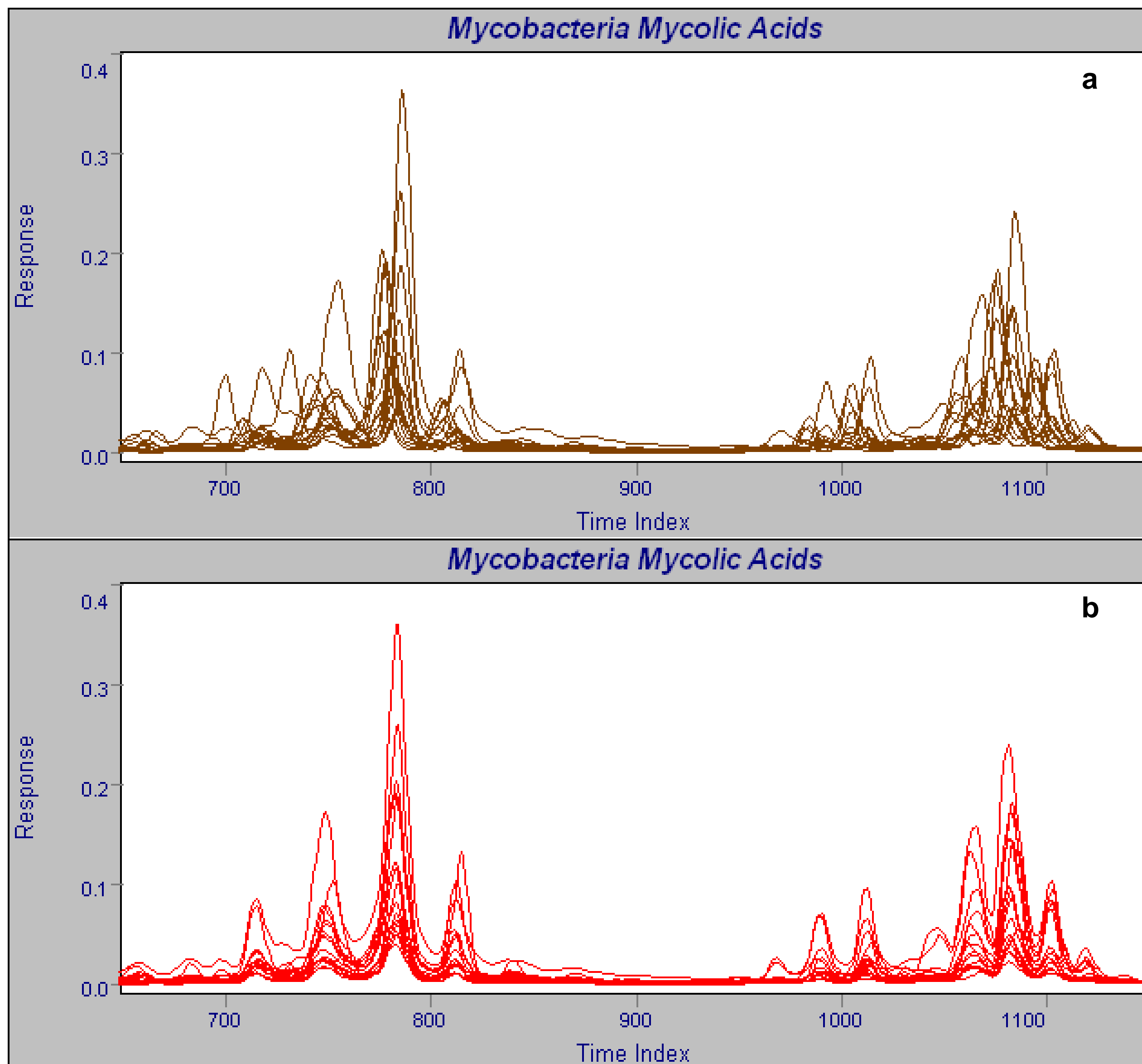

Figure 6. Results of 15 HPLC runs collected over a two-day period, not including QC samples; (a) the raw data; (b) the data aligned by LineUp (external pattern alignment)

In order to assess the quality of alignment, we computed a similarity value for each sample, compared to the QC standard, both before and after processing. The values are summarized in the following plot.

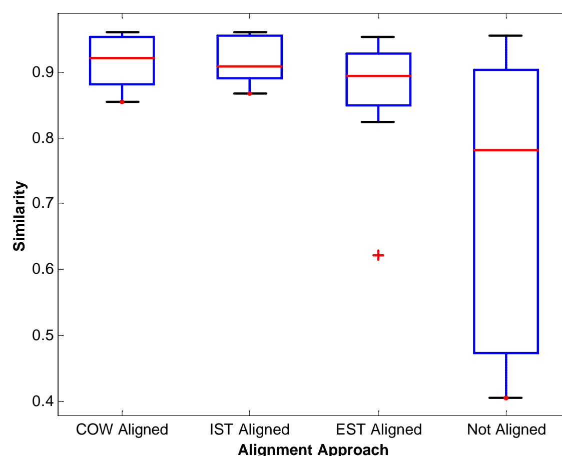

Figure 7. Box and whisker plots for the HPLC data with the range, first quartile, median and third quartile marked

There is clear improvement in the chromatograms after alignment by any of the three techniques (see Figure 6b). The raw data show significant variability that makes evaluation of the trace difficult and raises the risk that some peaks will be misassigned. The box and whisker plot for the external standard (EST) approach shows the range, median and quartile information for 14 of the samples, with the 15th plotted as a lone outlier. This sample was most time-distant from the running of the external QC standard. The internal standard and the external pattern match show very consistent results.

The HPLC data shown here mirror the results of the crude oil data first shown. When internal standards are available, the alignment of chromatographic traces is reliable. There are concerns about using external *peaks* as the basis for standardization, but external *patterns* give results comparable to a well-controlled internal standard run.

The On-Line World. The ability to use internal standards is simpler in an analytical laboratory, but there are significant drawbacks to adding standards to mixtures in a process environment. The retention problem is present in on-line settings due to typically less control over the environmental conditions, flow and column parameters. In the example below of low molecular weight hydrocarbon analysis (C_1 to C_9), there is enough variability to affect the accuracy of the analysis.

In this case, due to the lack of reliable retention time markers, only the external pattern match approach is viable.

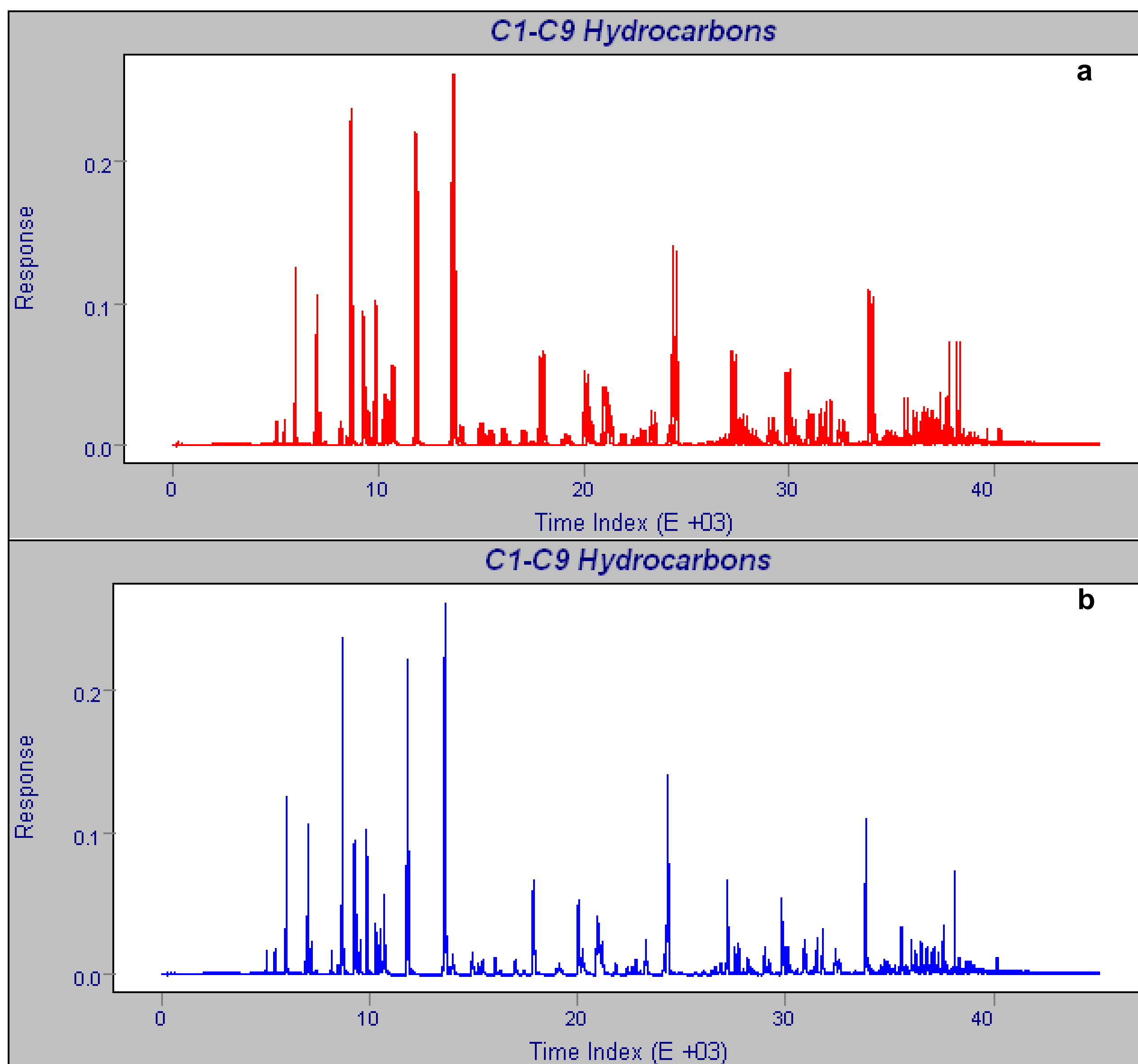

Figure 8. An overlay of the 5 original traces (a) show unstable retention times where a similar overlay of the 5 aligned traces (b) show them to be in sharp focus.

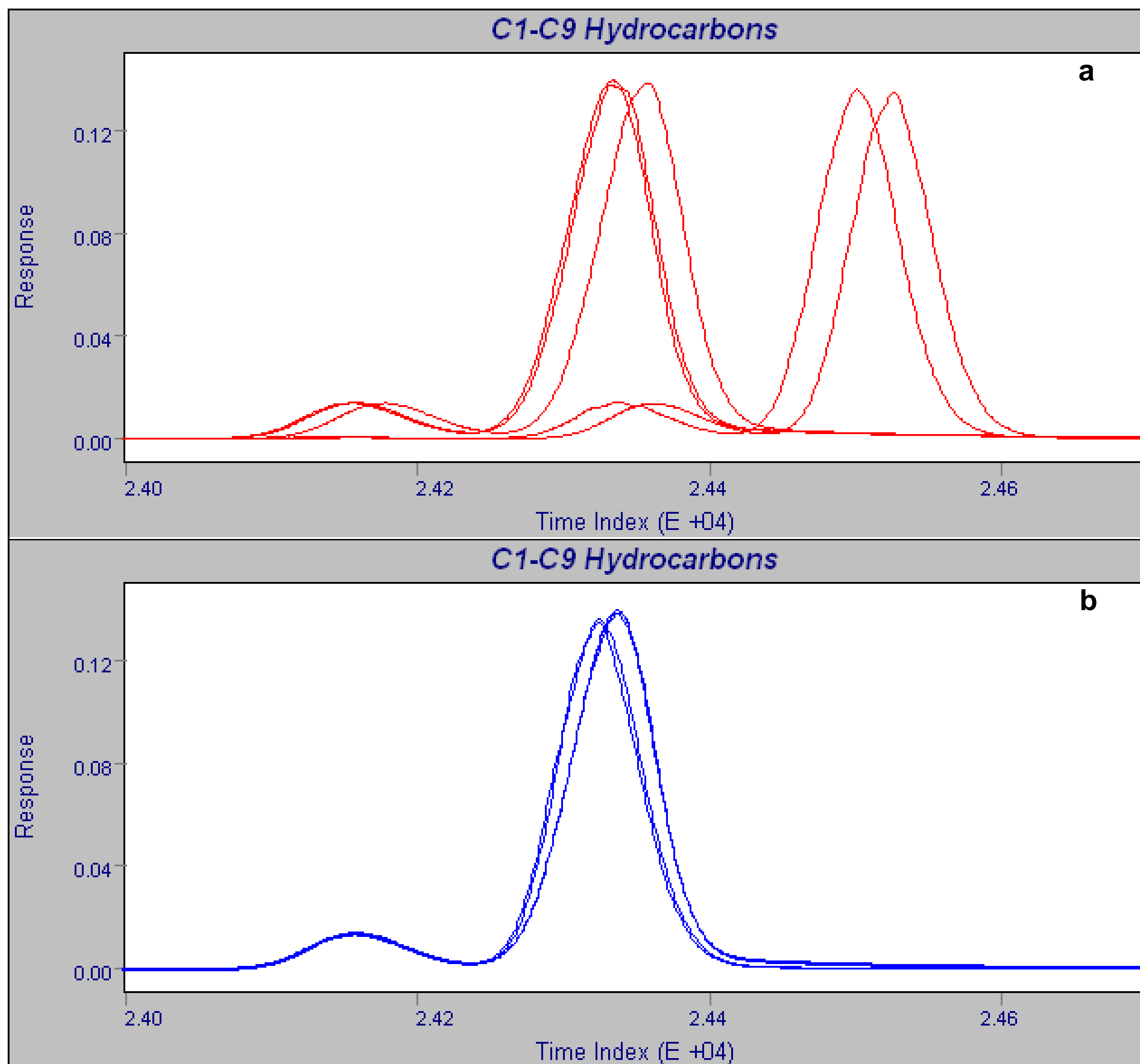

Figure 9. Zooming into the middle of the chromatogram, the retention time correction is highlighted; (a) unaligned and (b) aligned profiles

The tightening of the chromatographic distribution is well-illustrated by performing a principal component analysis (PCA) of the chromatograms, as shown in the following figure. In this view, each chromatogram is represented by a single point. The distance between these points is a function of the similarity of the traces: the closer the points, the more similar the chromatograms.

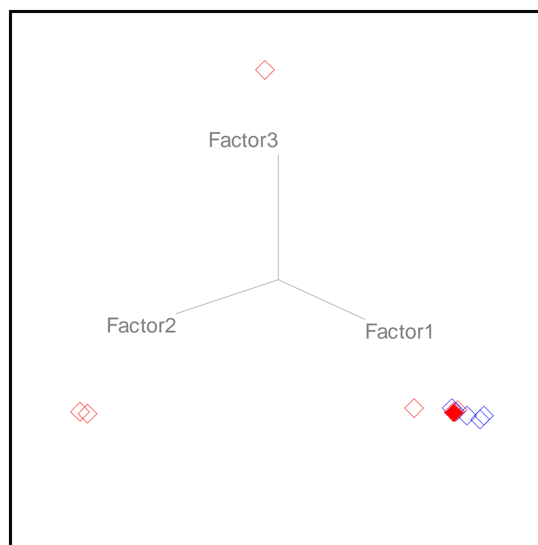

Figure 10. PCA Scores show scatter in the original (red) and a tight cluster after alignment (blue). Alignment standard is the filled point.

The scores of the pre-alignment chromatograms are widely spread, while the aligned traces are clustered tightly around the sample chosen to be the alignment standard.

CONCLUSIONS

For any alignment method to succeed, some basis for the alignment has to be present. In the case of internal peak indexing, standards of marker compounds need to be analyzed with the samples. Ideally, these markers would either be natively present, spiked into the sample matrix or at least co-injected with the samples.

In the case of external peak markers, it is clearly possible to analyze a mixture of marker compounds in a separate analysis from the sample. But we then are forced to assume that these times are similar enough to the times that would have been found had the markers been a part of the sample. This external standard approach makes a potentially risky assumption that the retention times do not shift significantly from run to run, within a day's analysis window.

The use of an external pattern standard does not require either coinjection or even the identification of individual peaks in the chromatogram prior to alignment. This capability allows alignment to be employed in an automated way prior to the integrator getting its hands on the data. It also requires no user intervention, allowing on-line applications to generate improved data.

REFERENCES

*The external pattern matching alignment is based on Correlation Optimized Warping, an algorithm continuing its development at the Technical University of Denmark by the CPB research group at BioCentrum and image analysis group at IMM under the direction of Associate Professor Jørn Smedsgaard.

Kováts, E., *Adv. Chromatogr.*, **1**: 229 (1965).

LineUp Version 2.0 *Software Manual*, Infometrix, Inc., p 21 (2003).

Nielsen, N-P. V.; J. M. Carstensen; and J. Smedsgaard, *Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping*, J. Chromatogr. A, **805**:17-35 (1998).

Pravdova, V.; B. Walczak; and D.L. Massart, *A comparison of two algorithms for warping of analytical signals*, Anal. Chim. Acta, **456**:77-92 (2002).

Wang, C.P. and T.L. Isenhour, *Time-warping algorithm applied to chromatographic peak matching gas chromatography/fourier transform infrared/mass spectrometry*, Anal. Chem., **59**:649-654 (1987).