ELSEVIER

Journal of Chromatography A, 686 (1994) 1–10

JOURNAL OF CHROMATOGRAPHY A

Effect of integration parameters on high-performance liquid chromatographic method development and validation

Yves-L. Grize[a,*], Heinz Schmidli[a], Jens Born[b]

[a]*Mathematical Applications Department, Ciba-Geigy Ltd., CH-4002 Basle, Switzerland*

[b]*Research Services Central Analytical Department, Ciba-Geigy Ltd., CH-4002 Basle, Switzerland*

First received 15 April 1994; revised manuscript received 27 July 1994

Abstract

An HPLC method is defined as a specific setting of the physical, chemical and data processing parameters which control the chromatographic analysis. During the development or validation of a method no special attention is usually given to the data processing parameters. However, it turns out that HPLC methods for complex samples can be very sensitive to minor changes in certain numerical integration parameters such as the threshold. A series of statistically designed chromatographic runs for a dyestuff with a very large number of peaks is presented where the interpretation of the data depends crucially on which threshold parameter value is chosen. It is therefore recommended that for development or validation of an HPLC method, especially in the context of quality control of complex substances, greater attention should be paid to the integration parameters.

1. Introduction

High-performance liquid chromatography (HPLC) is a widely used technique in the chemical industries. It has even been claimed that 50–70% of all analytical applications involve HPLC measurements [1].

The setting of the parameters (often called the separation conditions) that control the run of a chromatographic analysis is called a chromatographic method. A great deal of attention has been focused on method development, optimization and more recently method validation. The aim of HPLC method development and optimization is to find the best separation conditions for the reliable measurement of the amounts of individual components in the mixture. One goal of the validation of a given HPLC method is to check its insensitivity to small variations in the separation conditions. Clearly, optimization and validation are linked, as robustness is also a feature that an HPLC method should somehow satisfy to be really optimum.

An important type of application is the quality control and quality assurance of complex chemical substances involving a large number of components, e.g., reactive textile dyestuffs. In these applications, the chromatogram of a manufactured dyestuff is compared with that of a standard (reference) to ensure that the two samples do not differ appreciably. A typical example of the chromatogram of a reactive textile dyestuff is shown in Fig. 1, where two scales are used to show better the complexity of the chromatogram. One of the difficulties in this type of application is the importance that even small

* Corresponding author.

0021-9673/94/$07.00 © 1994 Elsevier Science B.V. All rights reserved
SSDI 0021-9673(94)00711-X

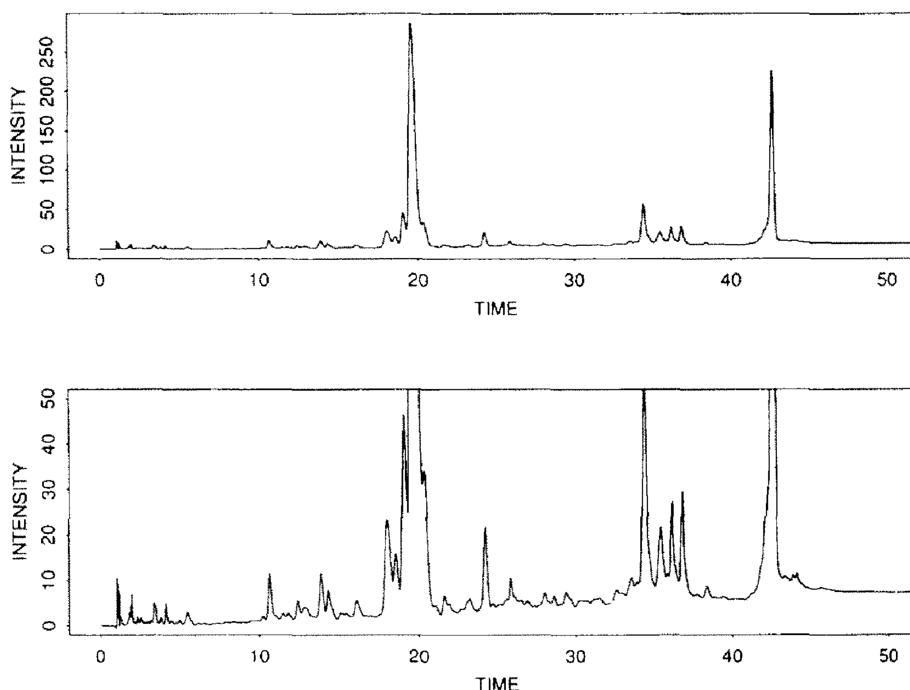

Fig. 1. Typical chromatogram of a reactive textile dyestuff.

unidentified peaks may have on the quality or toxicity of the product. In this specific context, the development of an HPLC method must aim not only at allowing a precise separation and quantification of the main components, but also, and perhaps even more important, at a systematic detection of minor components, sometimes under varying laboratory conditions.

Although the most usual strategy for developing HPLC methods in practice is still by trial and error combined with empirical knowledge and pragmatic rules, the advantages of the use of statistically designed experiments and response surfaces techniques have been recognized and are now used more frequently (e.g., [2–4]). Validation of chromatographic methods, in the past often reduced to tests of repeatability and linearity related to the areas of the most interesting peaks, are now including robustness (or ruggedness) tests. These tests, which must show that a method is insensitive to small changes in the method parameters, are also increasingly based on statistically designed experiments (e.g., [5–7]).

Chromatographic separation depends clearly on a large number of factors, including physical and chemical properties of the sample mixture and of the mobile and stationary phases and data processing parameters. Typically in the development and validation of HPLC methods the physical and chemical parameters are varied while the data processing parameters remain fixed. These data processing parameters can be separated into two groups: hardware parameters and software parameters. The hardware parameters are, e.g., flow control, temperature control, lamp current, photodiode-array detection (DAD)-specific and electronic parameters which influence the detected signals that are stored. The software parameters, on the other hand, are used to interpret and report the results from the stored data. The software parameters are often also called integration and calibration parameters. Usually no or only little attention is paid to the integration parameters. Further, manufacturers usually do not provide any indications on how to choose the settings of these parameters, and in practice a value in the middle of the

permitted range is arbitrarily chosen and kept fixed.

However, during our investigations on the validation of an HPLC method for a quality control application, we found instead that slight modifications of the values of certain integration parameters have surprising effects on the robustness of a method measured by the number of detected peaks. Why the number of detected peaks was chosen as the robustness criterion will be explained in Section 2.1. More specifically, when comparing a series of chromatograms obtained under different separation conditions (which is exactly what one does when using statistically designed experiments), the choice of integration parameter values may change the result of this comparison considerably: a method that first appears to be robust (number of detected peaks stable) may, after changing the integration parameters, lose its robustness (number of detected peaks unstable). In other words, the relative importance of the effects of the separation parameters may change when the integration parameters change. In practice, this means that two experimenters, after having performed exactly the same experiments, but having analysed their results under only slightly different settings of the integration parameters, may come to completely different conclusions in deciding which separation condition is better. This finding provided the main motivation for this paper, where our aim is to warn HPLC method developers about the possible sensitivity of the interpretation of their results to apparently innocuous integration parameter settings. In fact, we believe that data processing parameters must also be considered during HPLC method development and validation. We hope that the developers of HPLC methods will in future give more attention to these influences and provide advice to practitioners on how to adjust the integration parameters.

The paper is organized as follows: in Section 2 we describe the experimental background, explain why we chose the number of peaks as our robustness criterion and describe the different integration parameters; in Section 3 the effect of one integration parameter, the threshold, on the robustness analysis of an HPLC method used for the quality control of a dyestuff is described; and in Section 4 conclusions are drawn.

2. Experimental background

2.1. Robustness criterion

The usual optimization criterion for HPLC method development is the maximization of the minimum separation (or resolution) between adjacent peaks under the hypothesis that the number of components to be detected in the sample is known. However, in many industrial applications, e.g., in reactive textile dyestuffs production, this number of components is usually large and unknown. Of course, a chromatogram with a high resolution and a small number of peaks may be obtained, but this may be due to co-elution of many components. Also, in quality control or quality improvement applications, i.e., those of interest to us, it is a priori possible that certain small peaks will affect the quality of the product. For example, a very small amount of red in a yellow dyestuff has catastrophic consequences. For a typical chromatogram, see again Fig. 1. As it is not known in advance which component will affect quality, the total number of detected components is an important criterion for method selection.

One might at first think that it is enough simply to set the instrument threshold at its highest sensitivity, i.e., the one which corresponds to the chosen signal-to-noise ratio (e.g., 6σ), and then proceed with the experimentation. However, at this value peak detection is no longer reliable and far too many peaks are detected (in our case almost 200; see Table 1). There is therefore an unknown reasonable range of possible threshold values above the detection limit which still allows the detection of many potentially relevant small peaks and nevertheless gives a good resolution. In practice, one chooses an arbitrary value on the sensitive side of the threshold scale and keeps it fixed. The HPLC method is then developed at this threshold. During method development the number of

detected peaks is one response of interest among others such as minimum resolution and retention time. Multi-criteria optimization techniques (see, e.g., [8]) can be used at this stage. Finally, for validation purposes, the number of detected peaks is a very simple summary measure to check the robustness at detecting many small peaks. Of course, other measures could also be used, e.g., which monitor the area percentage of the major peaks. However, these other criteria are much more complex and more error-prone than the number of detected peaks. Especially in the context of validation of a method for complex substances with many small peaks, the number of peaks appears simple and yet sufficiently informative.

There are situations other than the quality control of complex substances where the total number of detected peaks can be used as a robustness criterion. For example, in the analysis of environmental samples, an a priori unknown number of components with very low concentrations need to be detected. Here again one is primarily concerned with maximizing the number of detectable components. Clearly, after this has been achieved, the HPLC method must be further improved by optimizing other criteria such as resolution of peaks known to be important or retention time. Finally, an important validation criterion is the stability of the number of detected peaks under changes in certain parameters.

2.2. *Experimental conditions*

The sample mixture was a commercially available reactive dye, Cibacron Red C-2G (500 mg dissolved in water). The chromatographic method to be tested was a gradient method. The end-points of the mobile phases were (1) a mixture of 10% acetonitrile and water and (2) 10% water and acetonitrile. Methanol was used as additional organic modifier. As ion-pairing agent tetrabutylammonium perchlorate was used and sodium citrate was used as a buffer. The columns were 125 mm × 4 mm I.D. and filled with 5-μm Hypersil ODS.

The chromatographic system was a Hewlett-Packard Model 1090M with three low-pressure pumps, autoinjector and column oven and DAD was applied. One of eight possible channels was used to acquire chromatograms at 254 nm. The spectral resolution chosen was 4 nm and the time resolution for the chromatograms was 0.003–0.004 min. The spectra and the chromatograms were stored on completion of the chromatographic run. Data editing and calculations such as determination and integration of peaks were executed afterwards using the stored data.

The background noise σ was measured using a blank chromatogram and a signal-to-noise ratio of 6 was taken throughout to ensure that no artifactual peaks would occur. This means that only peaks greater than 6σ were ultimately reported even if the chosen threshold value was smaller.

2.3. *Integration parameters*

The evaluation of the acquired raw data is done by mathematical integration. The integration algorithms identify peaks which are characterized by position on the time scale, height, area, width at half-height, symmetry, etc. The integration sensitivity can be adjusted by three integration parameters: threshold, area reject and peak width [9]. These three parameters are briefly described.

Threshold

The threshold is a value which expresses the minimum peak height detected by the integrator. This value lies on a scale from −12 to 25. On that scale, −12 is the most sensitive threshold and 25 the least sensitive threshold. The correspondence between this scale and the DAD milli-absorption units (mAU) scale is given by $\mathrm{mAU} = 2^{(th+2)}$, where *th* is the threshold value. Because the HP 1090M system used a 16-bit A/D converter, only sixteen different threshold values can be used.

Area reject

The decision on whether or not to store a peak is made at the end of the peak. The area reject

sets the area of the smallest expected peak. Thus the integrator ignores any peak which is smaller than this value. The smallest possible area reject value is zero.

Minimum peak width

The minimum peak width sets the value of the narrowest peak to be detected. Thus the integrator ignores any peak whose width is smaller than this value. The unit of peak width is time measured in minutes. The narrowest possible peak is 0.001 min.

A separate study not described here has shown that the threshold is the most influential of these three parameters on the number of detected peaks. Therefore, subsequently we only study the effect of this parameter.

3. Effect of threshold on the robustness analysis of an HPLC method

3.1. Experimental design

Before a newly developed HPLC method is used on a routine basis, e.g., for the control of product quality in different laboratories of a company, its robustness to small changes in the factor levels should be checked. If no significant effect is found then the method is robust and can be used routinely. Otherwise, either the influential factors have to be controlled more precisely, or a new, more robust method has to be developed. Statistically designed experiments can be used to test the robustness of a method by systematically varying the most important factors, for example, according to a fractional factorial experiment. Designed experiments are very helpful as the information needed can be obtained with a minimum number of experiments, which is especially important if there are many factors. Further, such experiments where the factors are varied simultaneously allow the estimation of possible interactions between the factors (synergisms). Recently, a number of such statistically designed experiments for robustness studies of HPLC have been reported (see, e.g., [7]).

For a newly developed method for determining the reactive dye Cibacron Red C-2G, previous experiments had identified four factors that were likely to affect the robustness of the method if deviations from their nominal values would occur. The four factors and their minimum and maximum deviations are as follows: I, the amount of ion pair agent, 0.4–0.6 g/l (nominal 0.5 g/l); M, amount of methanol, 28–32% (v/v) [nominal 30% (v/v)]; P, pH value, 6.2–6.6 (nominal 6.4); and C, the producer of columns, 0 = HP (Hewlett-Packard) or $+1$ = ST (Stagroma).

The factor C summarizes the properties of the columns of the different distributors, such as number of theoretical plates, filling procedures and ageing characteristics. It was included because different laboratories of the company may well use different column types if not otherwise specified.

In order to study the effects of such deviations, the factors were systematically varied according to a full factorial design, hence giving $2^4 = 16$ experiments. In addition, four of these experiments were repeated to provide information on the repeatability of the measurements. For each setting of the factors a chromatographic run was obtained. Then this chromatogram was analysed for each threshold value between -12 and 4 and the reported number of detected peaks was stored. The design and the number of peaks obtained for each factor combination and each different threshold value are given in Table 1.

3.2. Effect of threshold on the number of detected peaks and on repeatability

First the general effect of threshold on the number of detected peaks is examined. Fig. 2 shows this effect for the four different separation conditions that were run twice. We see that in each instance the number of peaks decreases monotonically as the value of the threshold parameter increases. This is, of course, what one would expect from the definition of this parameter. On the left side of the scale the curves are roughly constant. Clearly in that area the thres-

Table 1
Experimental design and results, showing the design (columns C, I, M, P) and for each of the 20 experiments the number of detected peaks obtained for each threshold value varying from −12 to 4

Run no.	C	I	M	P	Threshold																
					−12	−11	−10	−9	−8	7	−6	−5	4	−3	−2	−1	0	1	2	3	4
1	1	0.4	28	6.2	180	179	179	175	167	147	133	119	100	76	53	34	21	8	4	1	1
2	1	0.4	28	6.6	191	190	189	183	165	155	140	117	99	75	54	33	18	7	4	1	1
3	1	0.4	32	6.2	165	164	162	162	156	148	132	118	96	71	52	34	18	9	4	2	1
4	1	0.4	32	6.2	148	148	148	147	141	133	124	108	91	69	50	31	14	6	4	2	1
5	1	0.4	32	6.6	179	178	178	172	165	153	134	118	100	71	54	32	18	8	4	1	1
6	1	0.4	32	6.6	162	160	161	156	149	141	129	110	93	71	52	30	15	7	4	1	1
7	1	0.6	28	6.2	165	165	163	161	155	145	135	114	95	69	46	26	18	9	4	2	2
8	1	0.6	28	6.6	172	171	171	165	150	128	123	104	85	69	44	26	17	9	4	2	2
9	1	0.6	32	6.2	155	153	153	152	143	128	115	101	80	65	38	26	16	7	4	2	2
10	1	0.6	32	6.6	179	178	178	172	155	136	131	110	83	67	43	25	16	9	4	2	2
11	0	0.4	28	6.2	193	191	190	182	169	153	142	132	121	89	59	45	22	8	4	3	1
12	0	0.4	28	6.6	179	179	178	184	171	150	139	128	109	88	60	43	21	9	5	3	1
13	0	0.4	32	6.2	193	192	192	187	169	143	135	121	104	82	62	40	21	9	5	3	1
14	0	0.4	32	6.6	167	167	168	166	146	141	134	121	112	93	66	43	26	12	5	3	1
15	0	0.6	28	6.2	180	180	177	175	166	152	141	126	110	85	60	33	20	9	6	2	2
16	0	0.6	28	6.6	195	192	193	186	177	159	142	129	112	84	57	36	19	9	5	2	2
17	0	0.6	32	6.2	183	183	182	181	168	155	143	123	99	82	56	32	23	10	4	2	2
18	0	0.6	32	6.2	180	180	179	176	165	154	144	121	101	80	52	33	22	10	5	3	2
19	0	0.6	32	6.6	197	196	195	188	176	165	148	133	104	82	54	30	19	10	5	3	2
20	0	0.6	32	6.6	187	186	186	184	172	158	138	124	101	83	55	30	18	11	5	4	3

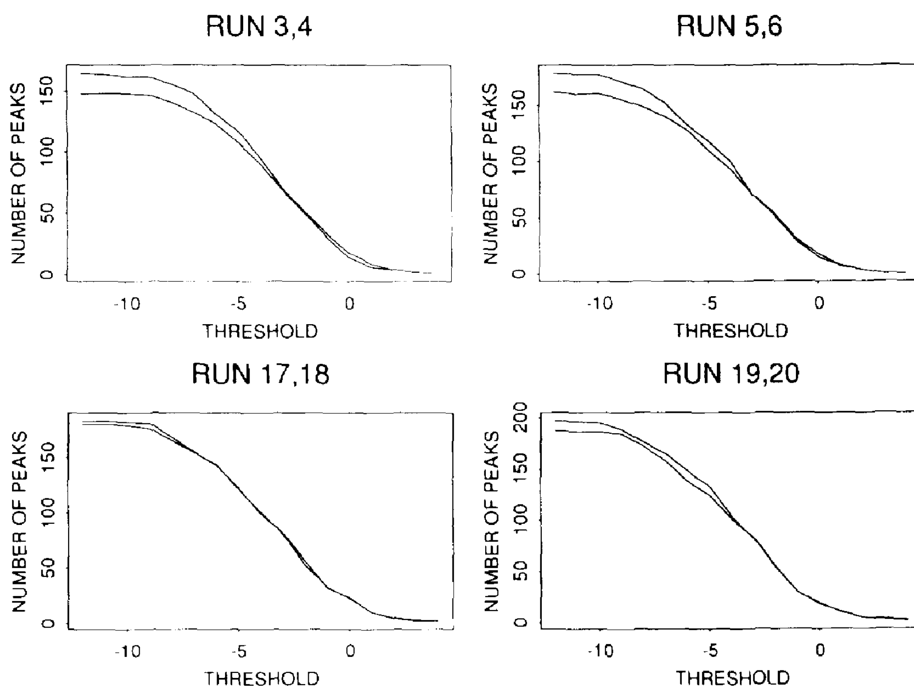

Fig. 2. Repeatability of measurements. The chromatograms corresponding to four different separation conditions (the run numbers correspond to Table 1) were obtained twice and analysed with different values of the threshold parameter. The graphs show how the number of detected peaks changes with the threshold.

hold was below the 6σ detection limit. As can be seen from Table 1, this was the case for all 20 chromatographic runs.

For the dyestuff under consideration the routine threshold value used in the laboratory before our study was undertaken was -2. It is interesting to observe that it is exactly around this value, say in the range from -5 to 0, that the number of peaks varies most. Fig. 2 shows also that in the four repeated cases the size of the threshold effect is clearly distinct from the repeatability error.

That the setting of the threshold parameter affects the number of peaks is, of course, obvious. What is interesting here is the strong non-linearity of this effect and therefore the high sensitivity to threshold values. This fact is, in our opinion, not sufficiently emphasized by the manufacturers of HPLC instruments. However, the full impact of this sensitivity will be shown below, when the number of detected peaks in different chromatograms is used to infer which are the most crucial parameters affecting the quality of the chromatographic run.

Note: when other data handling systems are used (e.g., Spectra-Physics and PE Nelson systems), the aspect of the curves of Fig. 2 may change, but they remain strongly non-linear.

The repeatability between two similar HPLC runs is summarized in Fig. 3, which shows how the relative standard deviation (R.S.D.) changes with the threshold value. For a given threshold, the R.S.D. was calculated as $(S^2/N)^{1/2}$, where S^2 is the average of the sample variance over the four runs which were repeated once (here simply the average squared differences in the number of peaks) and N is the average number of peaks. In the range from -12 to 0, the R.S.D. remains roughly constant at 5%. This is a desirable feature and allows one to compare directly results obtained for different values of the threshold. For positive values the R.S.D. is no longer meaningful because the average number of peaks is almost zero.

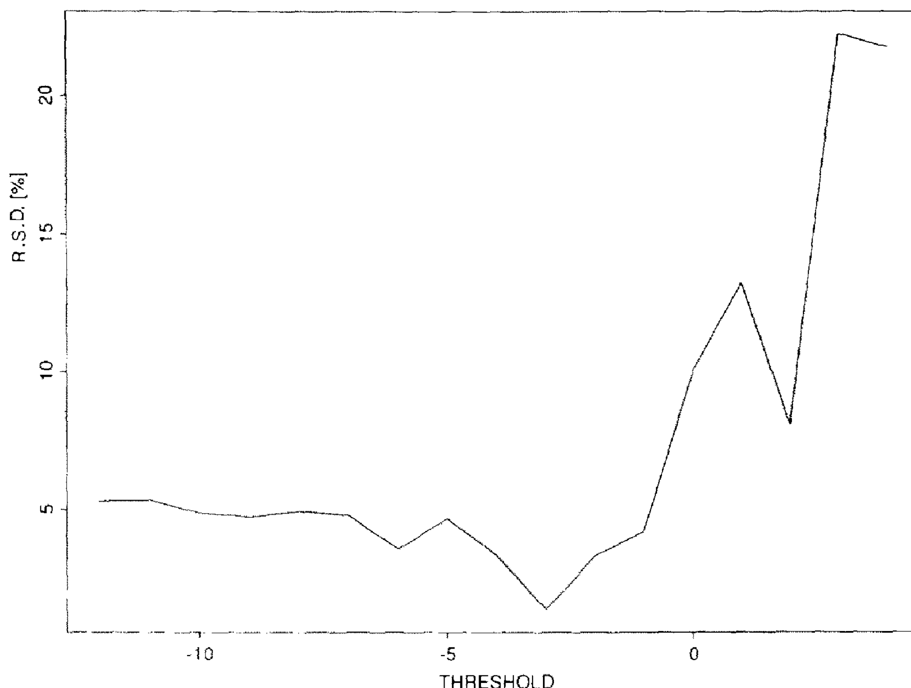

Fig. 3. The relative standard deviation is essentially constant for threshold values in the range −12 to 0.

3.3. *Effect of threshold on the analysis*

A full factorial design allows the estimation of all main effects of the four factors *C*, *P*, *I* and *M* and their eleven interactions *CP*, *CI*, . . . , *CPI*, . . . , *CPIM*. Hence, for a fixed value of the threshold, the 20 experiments can be analysed jointly by multiple regression. Such an analysis gives the effect (model coefficient) of the fifteen model variables *C*, *P*, . . . , *CPIM* on the number of peaks together with their *t*-values. The *t*-value of an effect is the ratio of the model coefficient and its estimated standard deviation. Absolute *t*-values larger than, say, 2 correspond roughly to a significance level below 5% and thus suggest that the effect of the corresponding variable should not be neglected, i.e., that the corresponding variable poses a robustness problem.

To visualize the results, the *t*-values for a range of threshold values are displayed in Fig. 4. This type of graph is called a λ-plot (see [10]) and conveniently summarizes the entire experiments and analyses by showing which are the likely relevant factors for each threshold value. Each connected line corresponds to the *t*-values of one of the fifteen model variables. Those lines that stand out have been labelled by the name of their corresponding model variable. For example, one sees that for a threshold of −5, only the factor *C* has an effect, whereas for the only slightly different threshold of −3 the three factors *C*, *I* and *M* are important and also the interaction between *M* and *P* (and perhaps even other interactions). Hence, by considering only the data corresponding to a threshold of −5, one would think that no robustness problem due to changes in *I*, *P* or *M* could occur. Therefore, it seems sufficient to specify the column type (here HP since this type tends to increase the number of peaks, as can be seen from Fig. 4) to ensure stability in the number of detected peaks. This is not, of course, what would have been deduced if only the data corresponding to a threshold of −3 had been considered.

Further, the graph shows also that the effect of column type (*C*) is consistently present (and always negative) almost throughout the range of threshold values. On the other hand the factor

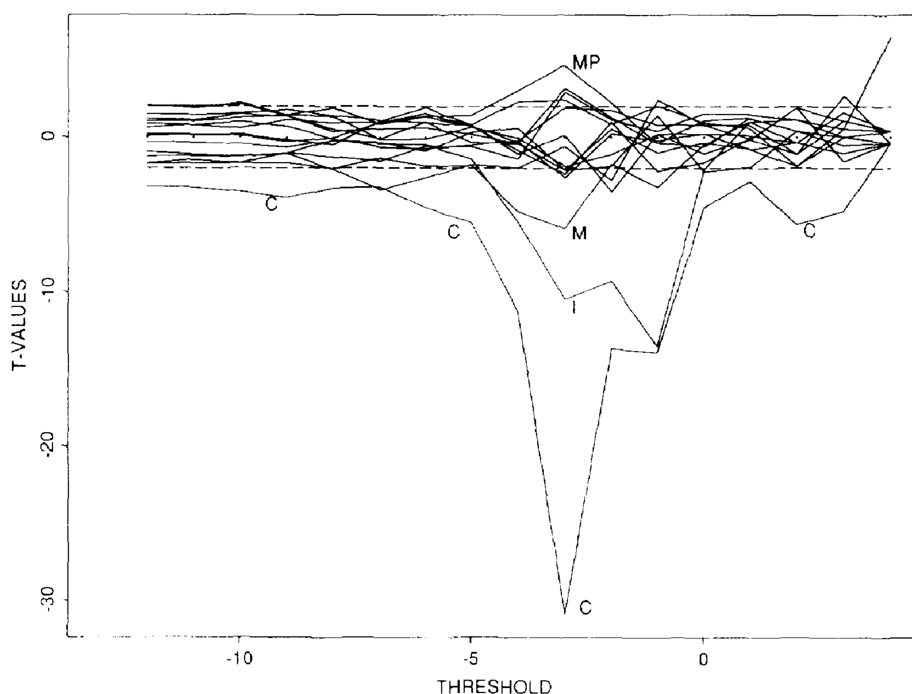

Fig. 4. λ-Plot of the 15 model variables. The plot illustrates how the t-values of each model variable varies with the threshold. Variables whose t-values exceed 2 in absolute value (horizontal dashed lines at 2 and −2) are likely to influence the number of detected peaks.

amount of methanol (M) is only significant for threshold values between −4 and −3. Finally, one also sees that it is for a threshold value of −3 that the greatest number of significant model variables occurs. This shows the variety of information that can be read from a λ-plot.

However, the main conclusion of the analysis is that the effects of column type, amount of ion-pairing agent and amount of methanol depend critically on the value of the threshold parameter. In other words, there is an interaction between threshold and the factors C, I and M. At this stage the experimenter faces the following difficult choice: either a specific threshold value can be chosen once and for all as adequate, perhaps based on some other aspects of the chromatogram, past experience or a priori knowledge of the expected number of peaks, or the HPLC method must be modified, so as to be less sensitive to small modifications of the threshold. Which road to pursue may be situation dependent. In any case, the experimenter has been warned that the threshold value is crucial in analysing the results of this series of experiments.

After having found this unsuspected effect of the threshold, we re-analysed other robustness experiments, not reported here, carried out with the same dyestuff. Our results were similar in that again we found that the analysis of factorial experiments was different depending on the threshold value.

4. Conclusions

In those situations where the number of detected peaks is at least one of the main concerns such as for substances with many important small components, this investigation illustrates a number of issues:

(1) The non-linearity of the number of detected peaks with respect to the value of the

threshold has shown that great care should be taken when setting the value of this parameter.

(2) The value of the threshold should be incorporated in any analysis involving the comparison of chromatograms such as obtained after performing a series of experiments like a factorial design. The analysis may differ greatly with the threshold.

(3) Any method validation study should consider the influence of the threshold. In fact if it does not seem to be possible to specify which threshold is adequate, the method should be robust for a reasonable range of threshold values. This means that the threshold and perhaps further data processing parameters should be incorporated in the experimental design together with the other separation parameters.

(4) During method development the situation is even more difficult. Different settings of the threshold could affect the optimization strategy and lead to very different optimum settings of the factors.

Although the points discussed above were made only in relation to the threshold parameter, similar conclusions could probably be drawn for other data processing parameters. Observe also that these points are not hardware dependent, but are a consequence of the necessary processing of the data.

In our opinion, the issues in this paper illustrate well a general feature of modern scientific measurements in that the complexity of modern laboratory instrumentation often discourages some from trying to understand how their system works. The neglect of apparently innocuous parameters may have dramatic consequences.

Acknowledgement

We thank S. Lahely for her great help with the laboratory work and one of the Editors for his critical but encouraging remarks which helped us to better formulate and justify our approach.

References

[1] F. Erni, presented at the *17th International Symposium on Column Liquid Chromatography, Hamburg, 9–14 May 1993*, Abstract No. O-111.
[2] B. Bourguignon, F. Marcenac, H.R. Keller, P.F. de Aguiar and D.L. Massart, *J. Chromatogr.*, 628 (1993) 171–189.
[3] P. Billot and B. Pitard, *J. Chromatogr.*, 623 (1992) 305–313.
[4] Q.S. Wang, R.Y. Gao, B.W. Wang and D.P. Fan, *Chromatographia*, 38 (1994) 187–190.
[5] M. Mulholland and J. Waterhouse, *J. Chromatogr.*, 395 (1987) 539–551.
[6] M. Mulholland, *Trends Anal. Chem.*, 7 (1988) 383–389.
[7] J.A. Van Leeuwen, L.M.C. Buydens, Vandeginste, G. Kateman, P.J. Schoenmakers and M. Mulholland, *Chemometr. Intell. Lab. Syst.*, 10 (1991) 337–347; and references cited therein.
[8] H.R. Keller, D.L. Massart and J.P. Brabs, *Chemom. Intell. Lab. Syst.*, 11 (1991) 175–189.
[9] Hewlett Packard Analytical Division, *HPLC ChemStation Fact File*, (1987) 3-27–3-47.
[10] G.E.P. Box, *Technometrics*, 30 (1988) 1–17.