ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
18 (1998) 43–56

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

Ruggedness testing of chromatographic methods: selection of factors and levels[1]

Y. Vander Heyden, F. Questier, L. Massart *

ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium

Received 24 October 1997; received in revised form 22 December 1997; accepted 3 January 1998

Abstract

The first step in a ruggedness test is the selection of factors to be examined and their levels. In this paper, both topics are discussed, thereby completing a strategy described earlier. It is demonstrated, by means of some examples, that depending on the formulation (definition) of a factor, information that is physically more or less meaningful is extracted from the experimental design results. Among others, the inclusion of the compounds of a buffer and of the components of a mixture in a screening design were examined. A general guideline to select the levels of the factors in a ruggedness test was proposed. Some special cases, i.e. asymmetric intervals around the nominal level, were also discussed.

Keywords: Ruggedness testing; Method validation; Experimental design; Factor selection; Level selection

1. Introduction

Ruggedness testing usually consists of an intralaboratory experimental study in which the influence of small but deliberate changes in the operating parameters or environmental conditions, called factors, on responses of the method are evaluated. These deliberate changes reflect those that can occur when a method is transferred between different laboratories, experimentators, instruments, etc. [1–4]. A ruggedness test is performed as a part of the validation of an analytical method. In a ruggedness test, an experimental design approach is mostly performed. The first steps in such a procedure are: (a) the selection of the operational or environmental factors to be investigated; and (b) the selection of levels for the factors. In this text, we will have a closer look at the selection and the definition of factors and levels, thereby completing a strategy described earlier [5].

The same set of factors can be entered in the experimental design in different ways and this can lead to information that is physically more or less meaningful depending on the formulation (definition) of the factors. As an example, the formulation of two types of factors will be discussed, namely those derived from the components con-

* Corresponding author. Tel.: +32 2 4774737; fax: +32 2 4774735; e-mail: fabi@vub.vub.ac.be

[1] Presented at the 7th Meeting on Recent Developments in Pharmaceutical Analysis, Island of Elba, Italy, September 16–20, 1997.

0731-7085/98/$ - see front matter
PII S0731-7085(98)00174-5

stituting a mixture and those being the compounds of a buffer. Problems occurring when different chromatographic columns are examined in a design are discussed as a third type of factor. The problems discussed in this text also occur when screening designs are applied in method optimisation [6–8].

After defining the factors, their levels have to be chosen. A general rule for defining factor levels is proposed. Some special cases, such as the definition of asymmetrical intervals around the nominal level, also are examined.

This text is meant as an aid for analytical chemists, pointing out some difficulties when defining the factors and levels in a ruggedness test. The methodology is illustrated with a ruggedness test designed for the high performance liquid chromatographic (HPLC) assay of a drug.

2. Discussion

The HPLC assay is extracted from a monography and had the following characteristics. The mobile phase consisted of acetonitrile/phospate buffer, 250:750 V/V. The phosphate buffer was prepared by dissolving 6.8 g potassium dihydrogen phosphate (MW = 136.1 g mol^{-1}) and 1.8 ml phosphoric acid (1.7 g ml^{-1}, 85% m/m, MW = 98.8 g mol^{-1}) in 1000 ml of water. A liquid chromatograph equipped with an isocratic solvent delivery system, a variable wavelength detector, an automatic injection system and a data acquisition system were used. The column was Merck RP Select B, Lichrospher 60 A, 5 μ, 125 × 3.9 mm ID. The flow rate of the mobile phase was 1 ml min^{-1}; the injection volume of the injected solutions was 10 μl and the detector wavelength was 220 nm. Analyses were performed at room temperature (20–30°C).

The following factors were originally selected from the operating procedure to be tested in a ruggedness test (A) concentration KH_2PO_4 in the aqueous phase, (B) volume H_3PO_4 in the aqueous phase, (C) volume acetonitrile in the mobile phase, (D) volume of aqueous phase in the mobile phase, (E) flow of mobile phase, (F) temperature of the column, and (G) detection wavelength. Each factor was examined at two levels (the extreme levels), which were respectively chosen smaller and larger than the operating conditions (nominal levels) (Table 1). Those factors were examined in a saturated fractional factorial design for seven factors (Table 2).

2.1. The selection and formulation of factors

2.1.1. Factors representing the components of a mixture

Factors C and D are the organic and aqueous parts of the mobile phase. They constitute a mixture and a property of mixtures is that they are subject to the constraint

$$\sum_{i=1}^{p} x_i = 1 \quad i = 1, 2, \ldots, p \tag{1}$$

where x_i is the fraction of the ith component, and p the total number of components in the mixture. Not all mixture components can be controlled independently and any component is determined when the fractions of the other $p-1$ ones are defined [9,10]. In our example, the fraction of the aqueous phase is determined once the fraction of acetonitrile is defined and vice versa. However, the volumes acetonitrile and aqueous phase in the mobile phase were entered in the design as two distinct factors (factors C and D in Tables 1 and 2). This is not appropriate for the reasons explained below.

Table 1
The factors examined and their levels

Factor	Levels		
	(−)	(+)	(0)
A: KH_2PO_4 (g l^{-1})	6.7	6.9	6.8
B: H_3PO_4 (ml l^{-1})	1.7	1.9	1.8
C: Acetonitrile (ml)	240	260	250
D: Aqueous phase (ml)	740	760	750
E: Flow (ml min^{-1})	0.9	1.1	1.0
F: Column temperature (°C)	20	35	20–30
G: Wavelength (nm)	219	221	220

The extreme levels are represented by (−) and (+) and the nominal ones by (0).

Table 2
Saturated fractional factorial design for seven factors (a) and experimental set-up derived from the theoretical design (b)

Experiment number	A	B	C	D	E	F	G
(a) Experimental design							
1	+	+	+	+	+	+	+
2	−	+	+	−	−	−	+
3	+	−	+	−	−	+	−
4	−	−	+	+	+	−	−
5	+	+	−	−	+	−	−
6	−	+	−	+	−	+	−
7	+	−	−	+	−	−	+
8	−	−	−	−	+	+	+
(b) Experimental set-up							
1	6.9	1.9	260	760	1.1	35	221
2	6.7	1.9	260	740	0.9	20	221
3	6.9	1.7	260	740	0.9	35	219
4	6.7	1.7	260	760	1.1	20	219
5	6.9	1.9	240	740	1.1	20	219
6	6.7	1.9	240	760	0.9	35	219
7	6.9	1.7	240	760	0.9	20	221
8	6.7	1.7	240	740	1.1	35	221

Generators of the design: D = ABC; E = AB; F = AC; G = BC.
Factors A–G are specified in the text.

In a ruggedness test, the influence of the factors on a response is usually determined by calculating their effects [1,5,11,12]. This is possible because the experimental designs applied have some specific properties. They have the property of orthogonality, which allows to calculate the effects of the factors independently [13]. The fractional factorial design of Table 2 is a balanced two-level design, i.e. in the $N/2$ design experiments a factor F_i is at one level all other factors are $N/4$ times at levels $(-)$ and $(+)$ [13,14]. In the calculation of the effects, the influences of the other design factors cancel. Another property of those designs is that the factors can be varied independently of each other, i.e. the variation of each variable should be uncorrelated to the variation of the others [15]. Factors C and D however are not independent. From Tables 3 and 4, representing an identical set-up as Table 2, it can be observed that factors C and D in reality are at four levels instead of at two. The consequences are that the resulting design is: (1) not a balanced two-level design any more which means that the calculated effects are corrupted, not only for factors C and D but for all factors because the effects of factors C and D do not cancel in the calculation of the others (main effects are confounded with each other); and (2) not orthogonal.

For the above example, one should have entered only the amount of acetonitrile or of the aqueous phase as a factor while the other one is only used as an adjusting component. The effect (coefficient) calculated for the factor is then caused by the change of the component examined or from an opposite change in the adjusting component (Tables 3 and 4, columns C and D). Their effects cannot be separated. The level of an amount of solvent is preferably expressed as a fraction instead of as a volume since a volume provides no information about its contribution to the mixture while a fraction does.

In HPLC analysis, the mobile phase can contain, besides the aqueous phase, more than one organic modifier, yielding mixtures of three or four components. The situation which occurs in ruggedness testing is that both mixture variables and process variables (e.g. flow, temperature, wavelength) need to be combined in an experimental set-up.

To choose an experimental set-up in this context, two approaches are described in the literature [10,16]. The first one deals directly with the p dependent mixture components x_1, x_2, ..., x_p while the second uses $p-1$ mathematically independent variables, called mixture-related variables. The latter approach allows to study both the mixture-related variables and the process variables as normal factorial variables in a screening design [16].

In the first approach, the process variables are examined in a screening design while the mixture components are studied in a mixture design. The experimental set-up to study all variables requires the execution of a mixture design at each of the screening design experimental conditions, or of a screening design at each of the mixture design experimental conditions. This is represented in Fig. 1 for examining two process variables in a factorial design and three mixture components in a set-up with three mixtures situated around the nominal conditions. In general, this approach requires $q \times N$ experiments where q is the number of mixture design experiments and N is the number of screening design experiments.

Let us consider a mobile phase consisting of methanol (MeOH), acetonitrile (ACN) and water with nominal fractions 0.10, 0.30 and 0.60, respectively, that is examined in combination with five process variables. The minimal number of experiments for such a set-up is 24, i.e. $q=3$, which is the smallest number of mixtures that can be defined around the nominal one, and $N=8$, the smallest screening design to examine five factors (e.g. a quarter-fraction factorial for five factors or a Plackett-Burman design for seven factors, containing two dummy variables). For the second approach, the five process variables and two ($p-1$) mixture related variables will be examined in one screening design. The minimal number of experiments to evaluate the seven factors is eight. This can, for instance, be carried out in a sixteenth-fraction factorial for seven factors or a Plackett-Burman design for seven factors.

In Table 5, a representation is made for examining the two process variables and two mixture-related variables of the three component mixture of Fig. 1 in one screening design. Eight experiments are needed whilst twelve were necessary in Fig. 1. Fig. 2 gives a graphical representation of the set-up of Table 5.

The second approach is clearly more economical in experiments than the first one. In ruggedness testing, one wants to perform the lowest number of experiments possible [5], therefore, the second approach should be preferred, while the first one is more suitable for modelling purposes in method optimisation.

A second decision to be taken, after the choice of the experimental set-up, is the selection of the experimental domain around the nominal conditions, more specifically the choice of the levels for the mixture variables and/or the number of mixture design experiments. Let us consider the first experimental set-up approach. Suppose the extreme levels for three mixture variables have been determined (e.g. according to Section 2.2). A hexagonal area around the nominal composition is then created by these levels. This is represented

Table 3
Fractions of acetonitrile (C) and aqueous phase (D) in the different design experiments

Experiment number	A	B	C	D	E	F	G
1	6.9	1.9	0.255 (+)	0.745 (+)	1.1	35	221
2	6.7	1.9	0.260 (+)	0.740 (−)	0.9	20	221
3	6.9	1.7	0.260 (+)	0.740 (−)	0.9	35	219
4	6.7	1.7	0.255 (+)	0.745 (+)	1.1	20	219
5	6.9	1.9	0.245 (−)	0.755 (−)	1.1	20	219
6	6.7	1.9	0.240 (−)	0.760 (+)	0.9	35	219
7	6.9	1.7	0.240 (−)	0.760 (+)	0.9	20	221
8	6.7	1.7	0.245 (−)	0.755 (−)	1.1	35	221

Between brackets the levels from the theoretical design are indicated.

Table 4
Variables matrix of Table 3

Experiment number	A	B	C	D	E	F	G
1	1	1	0.5	−0.5	1	1	1
2	−1	1	1	−1	−1	−1	1
3	1	−1	1	−1	−1	1	−1
4	−1	−1	0.5	−0.5	1	−1	−1
5	1	1	−0.5	0.5	1	−1	−1
6	−1	1	−1	1	−1	1	−1
7	1	−1	−1	1	−1	−1	1
8	−1	−1	−0.5	0.5	1	1	1

in Fig. 3 but with largely exaggerated intervals ($k = 20$, see Section 2.2) between the levels for visual purposes. Selection of the six corner points as mixture experiments would give for the first approach $6 \times N$ experiments which is not feasible and a more economic set-up is recommended. The smallest number of mixture experiments which create a domain around the nominal conditions is three (a triangle) which for instance could be chosen on the borders of the hexagon. Any triangle chosen will examine at the most only one solvent in an interval symmetrically around the nominal level while the others are studied in an asymmetric interval. However there is no reason why this would happen in reality, i.e. why a fraction would deviate more from the nominal level in one direction than in the other (Section 2.2) and moreover the total number of experiments still remain relatively high.

On the other hand, defining $p - 1 = 2$ mixture-related variables in a screening design (second approach) leads to experimental domains for the mixture components as given in Fig. 4. A practical example of the possible level combinations for examining MeOH and ACN as two factors in a two-level factorial design is shown in Table 6. For a given set of extreme levels, the hexagonal domain (Fig. 3), is completely enclosed by the quadrangular areas of Fig. 4. When moreover those two variables are chosen that are examined in the narrowest intervals, the created area comes closest to the hexagonal one, as can be seen in Fig. 4(a,b). Another advantage of this approach compared to the triangular domain is that all variables, also the adjusting component, are varied over intervals situated symmetrically around their nominal levels.

Finally after the selection of the experimental set-up and the mixture component levels, the influence of the factors on the responses has to be estimated. For the mixture-related variables the usual calculation of effects from a screening design can be used. However, for the definition of the effect of a mixture component some considerations can be made. In mixture design terminology the definition of an effect is different from the one applied in screening designs. For mixtures, the effect of component i on a response is defined as the change in the value of the response resulting from a change in the proportion of component i while holding constant the relative proportions of the other components [10]. In screening designs, an effect is defined as the average change in response observed for a change of that factor from one level to the other, measured on each of the two levels of the other design factors, and this definition is illustrated in [13].

Calculating the effect of a mixture component according to the first definition only allows one to enter one mixture component at the time in a screening design. If the effect of, e.g. MeOH is considered, then the ratio between ACN and water should remain constant (1/2) and is used as adjusting mixture. In reality however, deviations from the nominal conditions of MeOH, ACN and water will probably occur and the ratio among components does not remain constant. Therefore, the requirement that the relative proportions of some components remains constant, was dropped. The effect of the mixture components is better determined according to the screening design ter-

minology (Fig. 4(a,b)). This allows to estimate the effects of two components. It leads to the knowledge whether or not a small change in a certain component affects the response (e.g. selectivity). If one component has a significant effect, this means in practice that the mixture composition as a whole is important. Since it is not possible to control only one of the components of a mixture, the composition of the mixture as a whole should be more strictly controlled.

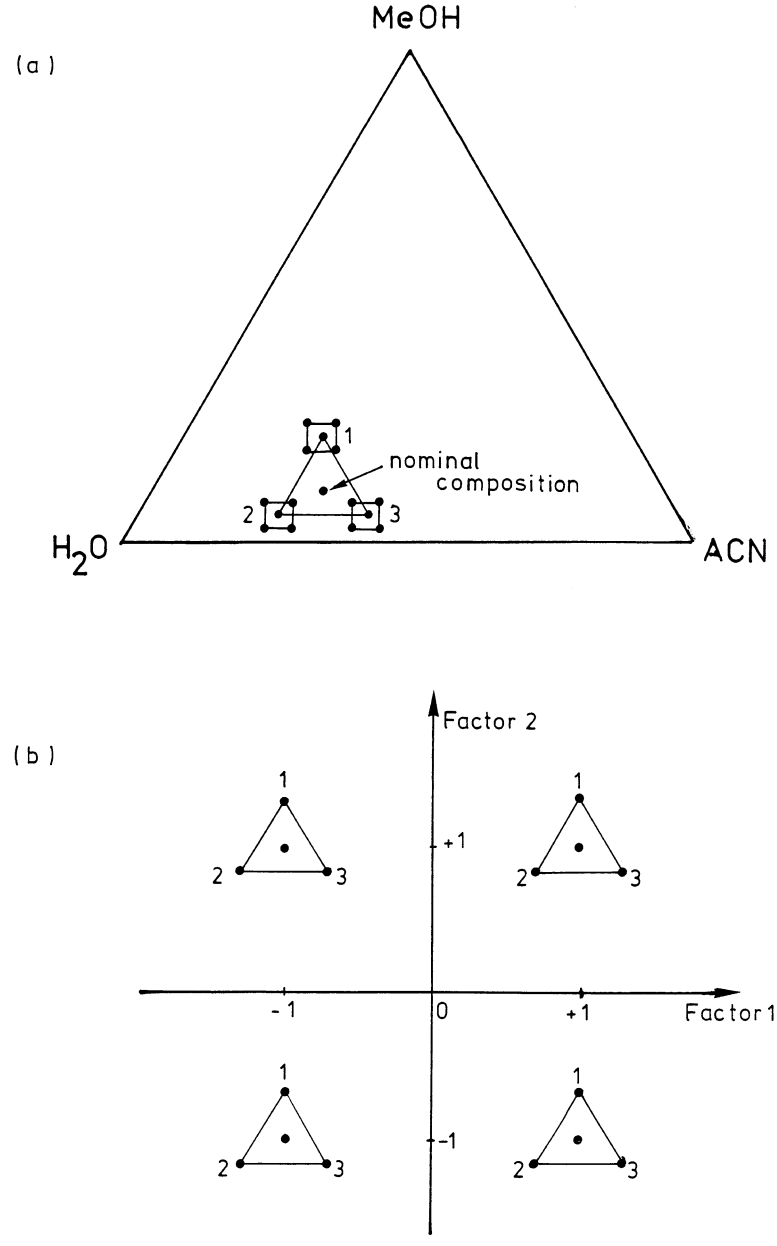

Fig. 1. The combined design approach to examine two process variables in a factorial design and three mixture components in a three experiments mixture set-up: (a) The 2^2 designs at each of the three points of the mixture design; and (b) the mixture designs at each of the four points of the screening design.

Table 5
Half-fraction factorial design for four factors (generator D = ABC)

Experiment number	A	B	C	D
1	+	+	+	+
2	−	+	+	−
3	+	−	+	−
4	−	−	+	+
5	+	+	−	−
6	−	+	−	+
7	+	−	−	+
8	−	−	−	−

Factors A and B are process variables, C and D are mixture variables.

2.1.2. Acidic and basic compounds of a buffer

In the example, factors A and B, the basic and acidic compound concentrations of the buffer KH_2PO_4/H_3PO_4 were entered as two factors. It is indeed possible to examine these factors in this way, because the two concentrations were described in the operating conditions and the design properties are not destroyed. In a ruggedness test, the first concern is to determine the ruggedness of a method and to a smaller extent to explain (the lack of) it. However, if one wants to maximise the information extracted from a ruggedness test it may be interesting to formulate the factors in such a way that the effects have a physical meaning. In the example this is not the case for the effect of factor B (see below).

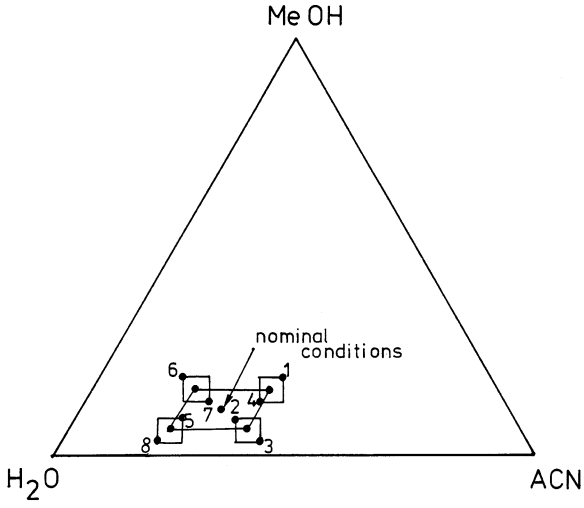

Fig. 2. Graphical representation of the set-up of Table 5 where an approach with mixture-related variables studied in a screening design is shown. 1, 2, ..., 8, are the experiment numbers.

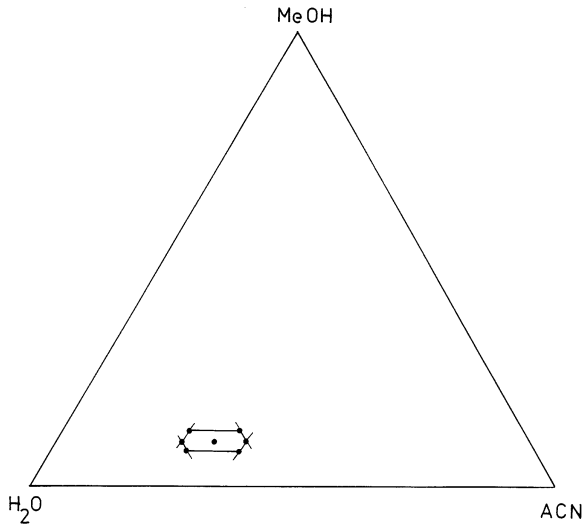

Fig. 3. Hexagonal area created by the extreme levels 0.080 and 0.120 for MeOH, 0.240 and 0.360 for ACN, 0.540 and 0.660 for H_2O. The levels were determined applying the rules of Section 2.2 on the above mentioned mobile phase which was prepared with graduated cylinders of 100 ml (uncertainty internal volume 0.375 ml), 500 ml (uncertainty internal volume 1.88 ml) and 1000 ml (uncertainty internal volume 5 ml) and a constant $k = 20$ was used.

The proton (H^+) activity (concentration) in the mobile phase affects the HPLC responses. The pH of the mobile phase influences the dissociation of some analytes and thus their retention behaviour. The ionic strength (μ) on the other hand can also affect retention behaviour and peak shapes. An increase in ionic strength can for instance reduce tailing [17].

In the example, the effect observed for factor A (Concentration KH_2PO_4 in the aqueous phase) is a measure for the ionic strength, because factor B (volume H_3PO_4 in the aqueous phase) is not contributing. However, for buffer systems composed of two salts, e.g. KH_2PO_4 and K_2HPO_4, both affecting the ionic strength, this is not the case. The effect calculated for factor B on the other hand has little physical meaning. It does not contribute to the ionic strength and is not directly representing the pH since this depends on the

ratio between the KH_2PO_4 and H_3PO_4 concentrations. The above is demonstrated in Table 7 where the pH ($pK_a = 2.12$) and the ionic strength were calculated. One observes that the effect of factor A (Tables 3 and 4) measures that of μ. For the pH, four levels occur due to the original levels of factors A and B. Deduction of the effect of the pH from this design is not possible because it is not a balanced two-level design any more.

In some cases, the pH or μ can be replaced by two other factors representing them. For the buffer applied, these factors can be defined by a and b and represented as

$$\frac{a*[KH_2PO_4]}{a*b*[H_3PO_4]}$$

where a and b are coefficients to be multiplied with the nominal concentrations of the buffer. It has to be mentioned that both factors a and b cannot be entered in a same screening design

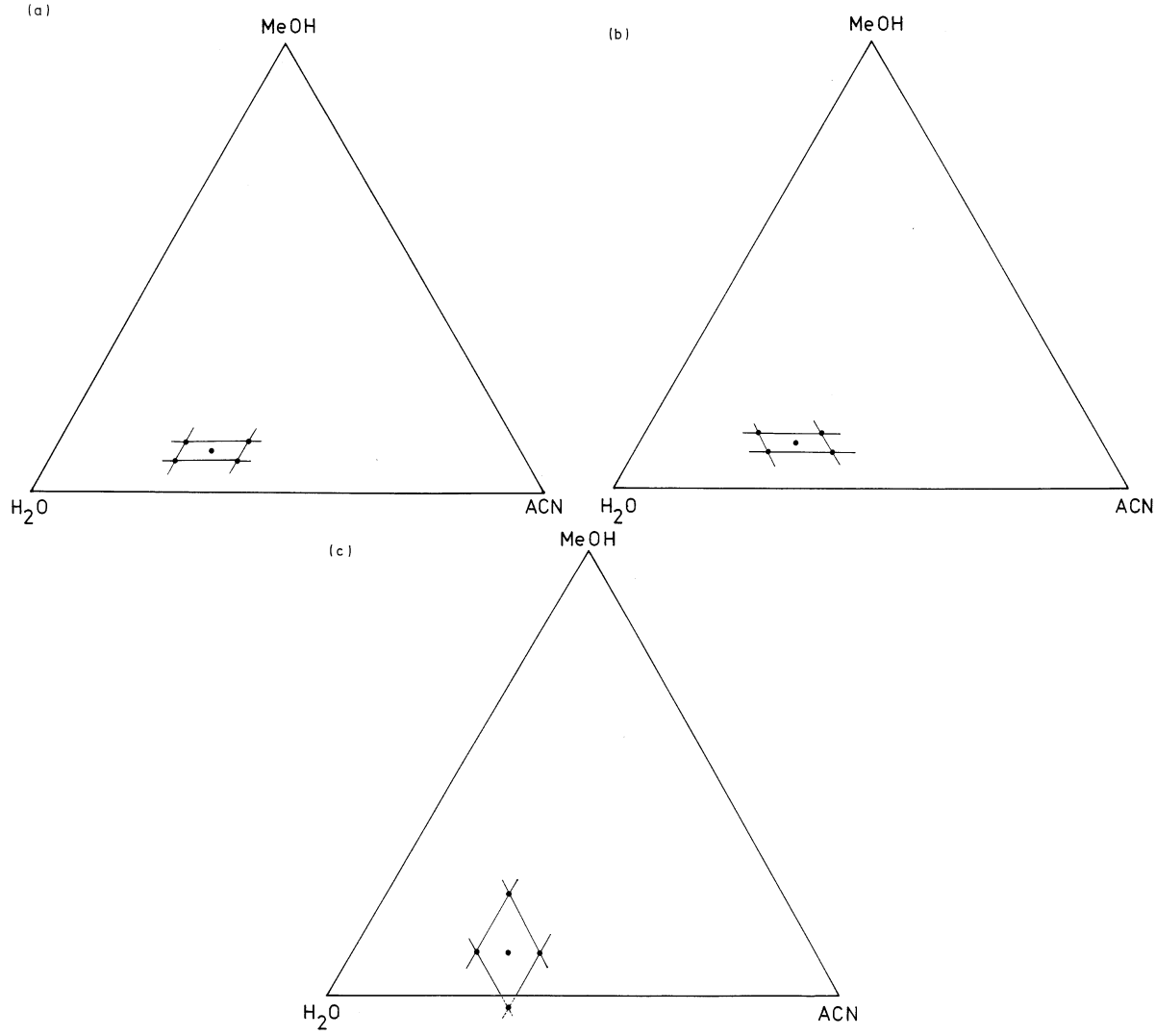

Fig. 4. Areas examined when factors (a) MeOH and ACN, (b) MeOH and H_2O, (c) H_2O and ACN, are examined in a screening design according to the second approach.

Table 6
Possible level combinations for the fractions MeOH (levels: 0.08 and 0.12) and ACN (levels: 0.25 and 0.35) in a two-level factorial design

Theoretical levels		Practical values (fractions)		Adjusting component
MeOH	ACN	MeOH	ACN	H_2O
−	−	0.08	0.25	0.67
+	−	0.12	0.25	0.63
−	+	0.08	0.35	0.57
+	+	0.12	0.35	0.53

unless creating the problems discussed above. Coefficient *a* is multiplied with both compounds and *b* only with the one not contributing to the ionic strength. Both coefficients *a* and *b* have a value of one as a nominal level. For the extreme levels, smaller and larger values are chosen. The coefficient *a* represents the ionic strength, because the ratio between both compounds remains constant but the amount of ions is changing, and coefficient *b* represent the pH, since it changes the ratio between the compounds while the ion concentrations remain constant.

If both buffer compounds contribute to the ionic strength, then the use of the coefficients *a* and *b* has other consequences. Consider for instance the buffer composed with KH_2PO_4 and K_2HPO_4. The coefficient *a* in $a*[KH_2PO_4]/a*[KH_2PO_4]$ represents again the ionic strength. However, the coefficient *b* in $[K_2HPO_4/b*[KH_2PO_4]$, is not representing only the pH. Coefficient *b* affects both the pH and ionic strength and its use is not interesting for this kind of buffer.

2.1.3. Impossible factor and level combinations in screening designs

One should be vigilant not to require impossible factor combinations. An example is the combination of the factors 'manufacturer of column material' and 'batch of material' in one two-level design. Selecting two levels for the manufacturer of material would give manufacturers *I* and *J*. Selecting two levels for the batch of material is not possible since one cannot define batches common to both manufacturers *I* and *J*. A design with four levels for the latter would be created, which is not balanced any more. A possibility to examine this kind of factors would be the use of nested designs [18,19]. Their use in ruggedness testing has been examined [20], personal communication] but a detailed discussion here is outside the scope of this manuscript.

2.2. Defining the factor levels: general guideline

The factor levels are usually defined symmetrically around the nominal level (Table 1). Other examples of levels examined in different case studies can for instance be found in [5,12,21–25]. The interval chosen between them represents the (somewhat exaggerated) limits between which the factors are expected to vary when a method is transferred. In most case studies, the levels are defined by the analyst according to his personal opinion of what is feasible or not.

In some references [1,26], the extreme factor levels are defined as a percentage of the nominal

Table 7
pH and ionic strength calculated for the design experiments of Table 2

	Calculated values			Scaled values	
Exp.	pH	μ		pH	μ
1	2.38	0.0507		−0.43	1
2	2.36	0.0492		−1	−1
3	2.43	0.0507		1	1
4	2.41	0.0492	Scaled	0.43	−1
5	2.38	0.0507	⇒	−0.43	1
6	2.36	0.0492		−1	−1
7	2.43	0.0507		1	1
8	2.41	0.0492		0.43	−1

value, i.e. as 'nominal level $\pm x\%$'. However, this is not an appropriate way. Consider two methods using mobile phases with pH 2.0 and 10.0, respectively. If the extreme levels are defined with a deviation of 10% then they would be 1.8 and 2.2 for the first, and 9.0 and 11.0 for the second method. Usually there is no reason why the ruggedness of the two methods for the influence of the pH should be examined in different intervals.

For quantitative factors, which vary on a continuous scale, we propose to define the extreme levels based on the precision or the uncertainty [27] with which a factor can be set and reset. For instance, the uncertainty in the factor 'pH of a solution' will depend on the uncertainty of the pH meter result and on the uncertainty related to the calibration of the pH meter. Suppose one knows, from a systematic determination of the uncertainties [27], that the pH varies with a confidence level of 95% in the interval $\text{pH} \pm 0.02$ where 0.02 is called the expanded uncertainty (cfr. further). It will be discussed below how to make an estimation of the uncertainty. Due to the uncertainty in the pH, one can expect the nominal pH (pH_{nom}) to vary between the levels $\text{pH}_{\text{nom}} \pm 0.02$. To select the extreme levels in a ruggedness experiment, this interval is enlarged to represent possible variations between instruments or laboratories. This is carried out by multiplying the expanded uncertainty with a coefficient k which gives as extreme levels $\text{pH} \pm k*0.02$. The value for k is in principle chosen arbitrarily but we propose to consider in a first instance the values two, five and ten for it. The value $k = 5$ is used as default value and the values $k = 2$ and 10 are left as alternatives to be used when the analyst prefers larger or smaller intervals for certain factors.

To quantify the (expanded) uncertainty in analytical measurements and, in our case, factor levels, detailed Eurachem guidelines exist [27]. Consider the determination of the uncertainty in the concentration of a solution. Suppose a reagent solution with a nominal concentration of 100 mg l^{-1} is defined in the operating procedure and is prepared in a 100 ml volumetric flask. The concentration C is determined as $C = m/V$ where m is the mass weighed and V the volume. To determine the uncertainty in C, those in m and V are estimated first. The uncertainty on m depends on: (1) the variability of weights determined by difference (final weight is determined by difference); and (2) the uncertainty associated with the calibration of the balance. In (1), Eurachem [27] defines a standard deviation of 0.07 mg for weights up to 50 g and in (2), it is reported that it is within ± 0.1 mg of the displayed value with 95% confidence. The latter quantity needs to be divided by 1.96 (derived from a normal distribution) to give the uncertainty of the component as a standard deviation, $0.1/1.96 = 0.052$ mg. Both components are then combined to give the uncertainty in m, $u_m = \sqrt{0.07^2 + 0.052^2} = 0.087$ or $\text{RSD}_m = 0.0087$.

A critical note should be made here. The above is the procedure proposed by Eurachem to define the uncertainty in the mass, but probably it does not cover all the errors made. For instance, in the mass of substance that is finally brought into the volumetric flask, one of the most important errors is probably the error in the transfer of the weighed material to the volumetric flask. However, this error is not so easily quantified, whilst on the other hand it is not necessarily occurring, e.g. when weighing immediately in the volumetric flask. The introduction of k into the definition of the factor levels should, to a certain degree, take into account the possible occurrence of the errors that are difficult to quantify. Another example of such an uncertainty is the difference in the temperature of a column oven and the real temperature within the analytical column.

The volume V is subject to three main sources of uncertainty: (1) the uncertainty in the stated internal volume of the flask; (2) the variation of filling the flask to the mark; and (3) the flask and solution temperatures differing from the calibrated temperature. The first is indicated by the manufacturer as a $\pm$ figure. For a 100 ml volumetric flask, it is 0.08 ml. Since this figure has no confidence level, a rectangular distribution is assumed and the standard deviation is $0.08/\sqrt{3} = 0.046$ [27]. The uncertainty due to variations in filling can be estimated from a repeatability experiment. A standard deviation for this variation is for instance 0.012 ml (adopted from [27]). Thirdly,

the uncertainty due to temperature difference from the flask calibration temperature can be calculated from an estimate of the temperature range and the coefficients of volume expansion. Only the volume expansion of the liquid is considered since it is considerably larger than that of the flask. Taking a coefficient of volume expansion for water of $2. \times 10^{-4}$°C and a possible temperature variation of ± 3 K gives a 95% confidence interval for a volume V of $\pm(V \times 3 \times 2.1 \times 10^{-4})$. For 100 ml, this gives a 95% confidence interval of 100 ± 0.063 ml and a standard deviation of 0.032 ml (division by 1.96). The uncertainty u_v in the volume V due to the three components is then $u_v = \sqrt{0.046^2 + 0.012^2 + 0.032^2} = 0.057$ ml ($RSD_V = 0.00057$).

The uncertainty u_C in the concentration C is given by $u_C/C = \sqrt{RSD_m^2 + RSD_V^2} = \sqrt{0.0087^2 + 0.00057^2} = 0.0087$. In order to combine the uncertainties associated with each component of a multiplicative expression, the RSD must be used. The standard uncertainty u_C in concentration C is then 0.0087×100 mg $l^{-1} = 0.87$ mg l^{-1}. The expanded uncertainty, which provides an interval within which the value of the factor is believed to lie with a given level is confidence, is ± 1.7 mg l^{-1} for a confidence level of 95%.

From the above example, it is seen that the determination of the extreme levels for the factors according to the Eurachem guidelines is useful but on the other hand rather cumbersome. However, the exercise is of interest, since it will indicate which sources of error are the most important. From a quality point of view, this is useful. However in a ruggedness test one is more interested to have an idea which order of magnitude to choose for the extreme levels, i.e. for a concentration of 100 mg l^{-1} should one examine in the interval 98–102 mg l^{-1}, 90–110 mg l^{-1} or 80–120 mg l^{-1}, to have a realistic situation.

A simpler alternative to the Eurachem guideline is the following. For each measured response, one so-called absolute uncertainty is defined. For instance: (1) for a mass, consider the last number given by the balance or a value specified by the manufacturer, as uncertain, e.g. 0.1 mg for an analytical balance; (2) for a volume, take the uncertainty in the internal volume of the volumetric recipient, specified by the manufacturer, e.g. 0.08 ml for a 100 ml volumetric flask; (3) for a pH value, use the last digit of the display or a value specified by the manufacturer of the pH meter. When a response, e.g. a concentration, is calculated from a number of measured components, the following rules are applied: (1) the absolute uncertainty for a sum or a difference is the sum of the absolute uncertainties in the terms; and (2) the relative uncertainty (i.e. absolute uncertainty/response value) for a product or a quotient is the sum of the relative uncertainties in the terms.

For the above example, we define the absolute uncertainty in the mass as 2×0.1 mg (mass obtained from a difference of two measurements) and in the volume as 0.08 ml. This gives relative uncertainties of 0.02 and 0.0008, and for the concentration of 0.0208. The absolute uncertainty in the concentration is then 2.1 mg l^{-1} which is similar in order of magnitude to the expanded uncertainty determined with the Eurachem guidelines. In both cases, the extreme levels to be examined in a design would be ~ 90 and 110 mg l^{-1} ($k = 5$).

A similar reasoning for solutions is valid for mixtures. Consider a mobile phase 30:70 V/V MeOH/H_2O prepared using graduated cylinders of 500 ml for MeOH (uncertainty internal volume 1.88 ml) and of 1000 ml for water (uncertainty internal volume 5 ml). The fraction of methanol is calculated as $f_{MeOH} = V_{MeOH}/V_{MeOH} + V_{H_2O}$ (volumes considered additive for ease of calculation). The three in Eurachem specified sources of uncertainty (see higher) in V_{MeOH}, expressed as standard deviations, were defined to be 1.085, 0.2 and 0.096, respectively and for V_{H_2O} as 2.887, 0.6 and 0.225. These values give an expanded uncertainty (95% confidence level) of 0.0029 in f_{MeOH}. If the mixture is expressed by f_{H_2O}, the expanded uncertainty is found to be 0.0074.

When applying the alternative guideline, only the uncertainties in the internal volumes are taken into account. The absolute uncertainty in f_{MeOH} is 0.0039 and the one in f_{H_2O} 0.0098. Again both approaches lead to values of a similar order of magnitude.

Depending on the condition, if the fraction MeOH or the one of H_2O is considered, different

uncertainties are found. However, in Section 2.1.1, it was seen that only one fraction should be entered in a screening design. Since the levels of that fraction describe the deviations occurring in the total mobile phase we would recommend to select the extreme levels based on the largest uncertainty found in the fractions of the different components, here in the one of H_2O. The extreme levels are then defined, both for MeOH and H_2O, as $f \pm k$ 0.0074 or $f \pm k$ 0.0098 according to the two guidelines respectively, where a value of $k = 2$ seems large enough here.

Similar reasoning also leads to the definition of the uncertainties in more complex mixtures (Section 2.1.1).

For quantitative factors, the interval between the extreme levels is generally situated symmetrically around the nominal one since there is no reason why a deviation from the nominal level is more probable in one direction than in the other. However, for some factors, the selection of an asymmetrical interval represents more reality. Suppose that a column temperature of 35°C is prescribed. Subsequently, it is not illogical to define room temperature (20°C) as low level since it is possible that the method in some cases will be executed at this temperature, for instance because one does not have a column oven. To determine the high extreme level, the uncertainty in the column oven temperature multiplied by coefficient k could still be used, giving for instance 40°C.

It could be argued that examining the temperature in an interval between 20 and 40°C is not a small perturbation, as the definitions for ruggedness require. However, the idea of ruggedness testing is that the factors are examined for changes that in practice can occur when a method is transferred. In fact, ruggedness tests were originally introduced to simulate these possible perturbations in order to avoid problems of method transfer, for instance, when determining the method reproducibility in an interlaboratory study. That the factors, in general, are examined in small intervals is a consequence of the fact that a transfer usually only is expected to cause changes in the factor conditions within a small interval. It however should not be an a priori requirement. Therefore if one can exclude that after transfer of the method from the laboratory where it is developed to those where it will be used that a method prescribed at 35°C will be executed at room temperature then the above given temperature interval could be replaced by a symmetric one, as for instance, 30–40°C. In the other case, examining the temperature in the interval room temperature − 40°C could be worthwhile considering.

Another example where an asymmetrical interval reflects the variations occurring in reality more, is for the factor 'reaction time', e.g. for a precolumn derivatisation reaction. It is not uncommon that the applied reaction time is sometimes longer than the one prescribed because one thinks that this will not affect the result (the longer, the better…) while too short a reaction time is less probable. Therefore, when the nominal reaction time is 30 min, an interval for the extreme levels between 25 and 45 min may be more adequate than a symmetric one.

3. Conclusions

Proper formulation of factors and selection of their levels, knowledge about experimental design and about the chemical and physical properties of the examined factors are needed to obtain maximal information in a ruggedness test. The following topics should be considered or checked prior to the performance of experiments: (1) can the factors be changed independently from each other and be set and reset reproducibly between experiments; (2) are they formulated appropriately and in an economic way (e.g. mixture factors) (3) are the designs balanced; and (4) are impossible factor or level combinations absent? One should also consider whether one wants merely an estimate of ruggedness, the identification of factors specified in the procedure that affect it or whether one wants to obtain physically interpretable effects also.

The factor levels to be tested in a ruggedness test can be defined based on the uncertainty with which the nominal one can be set and reset. For some factors, an examination of the extreme levels in an interval situated asymmetricaly around the

nominal one can represent better reality than a symmetric one.

Acknowledgements

Parts of this work were funded by the Research Contract NO/03/003 of the Belgian government (The Prime Minister Services, Federal Office for Scientific, Technical and Cultural Affairs, Standardisation Programme) and by the European Community Standards, Measurements and Testing research programme. The authors are grateful to V. Reynders and A. De Schrijver for the technical aid and to B. Boulanger and P. Chiap for discussions on the subject.

References

[1] J.A. Van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, G. Kateman, P.J. Schoenmakers, M. Mulholland, RES, an expert system for the set-up and interpretation of a ruggedness test in HPLC method validation. Part 1: The ruggedness test in HPLC method validation, Chemom. Intell. Lab. Syst. 10 (1991) 337–347.

[2] G.T. Wernimont, Use of Statistics to develop and evaluate Analytical Methods, in: W. Spendley (Eds.), Association of Official Analytical Chemists, Arlington, VA, pp. 78–82.

[3] International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use, Draft Consensus Text, Validation on Analytical Procedures, Released for Consultation, 26 October 1993, pp. 1–7.

[4] H. Fabre, Robustness testing in liquid chromatography and capillary electrophoresis, J. Pharm. Biomed. Anal. 14 (1996) 1125–1132.

[5] Y. Vander Heyden, F. Questier, D.L. Massart, A ruggedness test strategy for procedure related factors: experimental set-up and interpretation, J. Pharm. Biomed. Anal. 17 (1998) 153–168.

[6] S. Boonkerd, M.R. Detaevernier, Y. Vander Heyden, J. Vindevogel, Y. Michotte, Determination of the enantiomeric purity of dexfenfluramine by capillary electrophoresis: use of a Plackett-Burman design for the optimization of the separation, J. Chromatogr. A 736 (1996) 281–289.

[7] K. Jones, Optimization of experimental data, International Laboratory, November 1986, pp. 32–45.

[8] K. Jones, Process scale high-performance liquid chromatography. Part I: An optimisation procedure to maximise column efficiency, Chromatographia 25 (1988) 437–442.

[9] L.B. Hare, Graphical Display of the Results of Mixture Experiments in: R.D. Snee, L.B. Hare, J.R. Trout (Eds.), Experiments in Industry: Design, Analysis and Interpretation of Results, ASQC (1985) 99–109.

[10] J.A. Cornell, Experiments with Mixtures, 2nd ed., Wiley, New York, 1990, pp. 5, 250–256.

[11] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: a textbook, Elsevier, Amsterdam, 1988, pp. 5,101–106, 187.

[12] M. Mulholland, J. Waterhouse, Investigation of the limitations of saturated fractional factorial experimental designs with confounding effects for an HPLC ruggedness test, Chromatographia 25 (1988) 769–774.

[13] E. Morgan, Chemometrics: experimental design, Analytical Chemistry by Open Learning, Wiley, New York, 1991, pp. 54, 81–188.

[14] Y. Vander Heyden, M.S. Khots, D.L. Massart, Three level screening designs for the optimisation and ruggedness testing of analytical procedures, Anal. Chim. Acta 276 (1993) 189–195.

[15] R. Carlson, Design and Optimisation in Organic Synthesis, Data Handling in Science and Technology, vol. 8, Elsevier, Amsterdam, 1992.

[16] D.A. Doornbos, A.K. Smilde, J.H. de Boer, C.A.A. Duineveld, Experimental design, response surface methodology and multicriteria decision making in the development of drug dosage forms, in: E.J. Karjalainen (Ed.), Scientific Computing and Automation (Europe), Elsevier, Amsterdam, 1990, pp. 85–95.

[17] S. Lindsay, High Performance Liquid Chromatography, Analytical Chemistry by Open Learning, Wiley, New York, 1987.

[18] R.R. Sokal, F.J. Rohlf, Biometry, the principles and practices of statistics in biological research, 2nd ed., W.H. Freeman, San Francisco, CA, 1981, pp. 271–320.

[19] International Organisation for Standardisation (ISO), Accuracy (trueness and precision) of measurement methods and results—Part 3: Intermediate measures of the precision of a standard measurement method; International Standard ISO 5725-3:1994(E), First edition.

[20] Y. Vander Heyden, K. De Braekeleer, Y. Zhu, J. Hoogmartens, J. De Beer, D.L. Massart, Nested designs (Nested ANOVA) in ruggedness testing, J. Pharm. Belg. (abstract from poster the Fifth International Symposium on Drug Analysis) 4 (1995) 408.

[21] Y. Vander Heyden, C. Hartmann, D.L. Massart, P. Nuyten, A.M. Hollands, P. Schoenmakers, Ruggedness testing of a size exclusion chromatographic assay for low molecular mass polymers, J. Chromatogr. A 756 (1996) 89–106.

[22] Y. Vander Heyden, A. Bourgeois, D.L. Massart, Influence of the sequence of experiments in a ruggedness test when drift occurs, Anal. Chim. Acta 347 (1997) 369–384.

[23] Y. Vander Heyden, C. Hartmann, D.L. Massart, L. Michel, P. Kiechle, F. Erni, Ruggedness tests on an HPLC assay: comparison of tests at two and three levels by using two-level Plackett-Burman designs, Anal. Chim. Acta 316 (1995) 15–26.

[24] Y. Vander Heyden, D.L. Massart, Y. Zhu, J. Hoogmartens, J. De Beer, Ruggedness tests on the HPLC assay of the United States Pharmacopeia XXIII for tetracycline hydrochloride: comparison of different columns in an interlaboratory approach, J. Pharm. Biomed. Anal. 14 (1996) 1313–1326.

[25] Y. Vander Heyden, K. Luypaert, C. Hartmann, D.L. Massart, J. Hoogmartens, J. De Beer, Ruggedness tests on the HPLC assay of the United States Pharmacopeia XXII for tetracycline hydrochloride. A comparison of experimental designs and statistical interpretations, Anal. Chim. Acta 312 (1995) 245–262.

[26] L.C. Rodríguez, A.M. García Campaña, J.M. Bosque Sendra, R.M. Blanc García, A proposal for a complete ruggedness test of experimental optima in analytical methods, Abstract book of the CAC96 International Conference, Tarragona, Spain, p. 154.

[27] Eurachem, A focus for Analytical Chemistry in Europe, Quantifying Uncertainty in Analytical Measurement, version 5, Eurachem Workshop Draft, September 1994.