

# RTS, a computer program for the experimental set-up and interpretation of ruggedness tests

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

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

Received 8 September 1997; received in revised form 9 January 1998; accepted 17 January 1998

---

## Abstract

A computer program is described for the experimental set-up and interpretation of ruggedness tests. The implemented strategy was based on a number of case studies and contains both recommended designs and minimal designs. The minimal designs reduce the number of experiments, but they cannot be statistically interpreted based on the interaction or dummy factor effects. The use of randomization tests as an alternative statistical interpretation method for the significance of the effects was examined. Some of the minimal designs are expandable to designs with characteristics similar to those of the recommended designs. The program is designed to facilitate the selection of the designs and the interpretation of the results and to prevent or detect problems such as drifting of responses. © 1998 Elsevier Science B.V. All rights reserved.

*Keywords:* Ruggedness test strategy; Method validation; Experimental design; Software

---

## 1. Introduction

The ruggedness of an analytical method can generally be described as the ability to reproduce this method under different circumstances without the occurrence of unexpected differences in the obtained results [1]. A ruggedness test is a part of method validation, which is becoming increasingly important, particularly in the pharmaceutical industry. However a ruggedness test is often not performed, due to the complexity and the many approaches possible. Therefore, after examining the literature [1] and performing a number

of case studies [2–5], a strategy was selected to determine the ruggedness of an analytical method [6]. This strategy was implemented in a MATLAB computer program, called ruggedness test strategy (RTS), which is described in this article. The strategy presented here follows the definition for ruggedness test most frequently used in the chemical literature: ‘A ruggedness test is an intralaboratory experimental study in which the influence of small changes in the operating or environmental conditions, called factors, on measured or calculated responses are evaluated. The changes introduced, reflect the changes that can occur when a method is transferred between different laboratories, different experimentors, different devices, etc.’ [7,8]. According to this definition,

---

\* Corresponding author. Fax: +32 2 4774735; e-mail: [fabi@vub.vub.ac.be](mailto:fabi@vub.vub.ac.be)

the factors one examines are mainly factors that are described in the operating procedure of the method, such as flow rate, pH of mobile phases, etc.

**2. Hardware and software**

RTS was developed on a computer with an Intel 75 MHz Pentium processor and 8 MB RAM, running Microsoft Windows 3.1 and Mathworks MATLAB 4.0.

MATLAB is a technical programming environment for high-performance numeric computation and visualisation.

The minimum configuration for RTS is a computer with a 486 processor with mathematical co-processor and 8 MB RAM, running Microsoft Windows 3.1x and MATLAB 4.x.

**3. Description of the program**

The program is designed to guide the user step by step through the strategy for the experimental set-up and the interpretation of a ruggedness test, as can be seen in the flow chart of the program (Fig. 1). The graphical interface, the on-line help menus and default choices make the program user-friendly, so that it can easily be run by users having little experience with MATLAB or with ruggedness testing. Structured programming and source code make the program easy to update.

RTS consists of two main parts: (i) the selection and creation of the experimental set-up and (ii) the treatment of the results of the experiments in which the experimental data are entered and the calculations and the statistical interpretation are performed.

The program enables experimental data to be manually entered, saved and loaded from disk. Each choice to be made or result is represented on a separate screen. On each screen there is (i) a help-menu, containing a contents description, a ‘Frequently Asked Questions’ file and a context-specific explanation, (ii) a ‘Go on’ and ‘Back’ button allowing fast navigating through the program, (iii) a ‘Quit’ button allowing to leave on

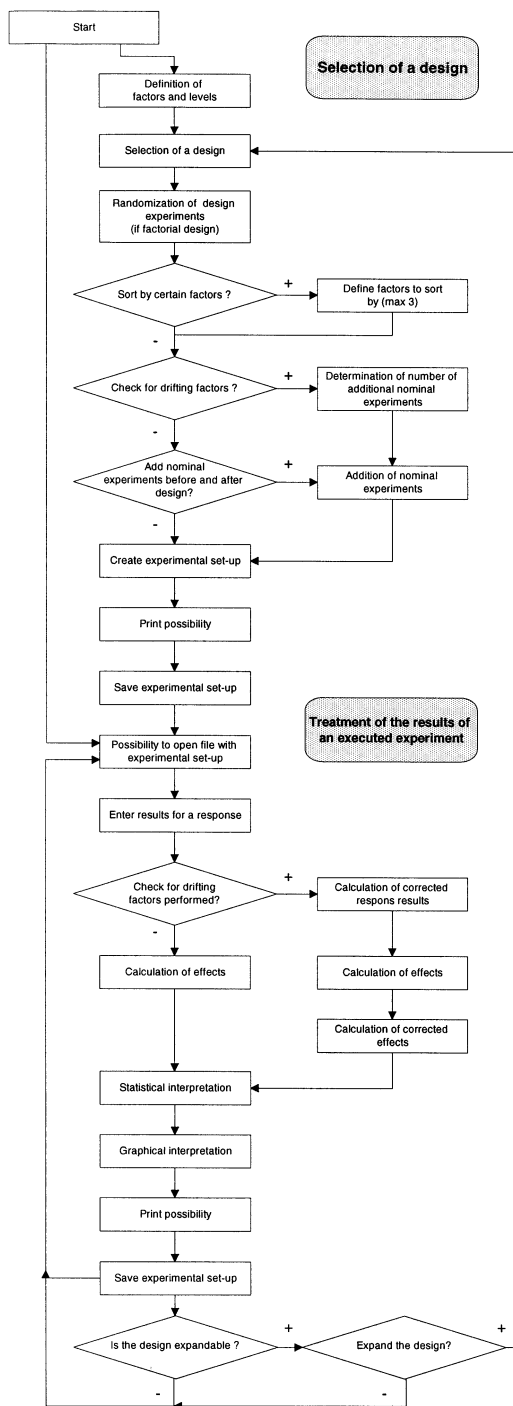


Fig. 1. Flowchart of RTS.

Factor	Description of factors	Levels	
		-1	1
1	Flow rate	0.9 ml/min	1.1 ml/min
2	Column temperature	20°C	30°C
3	Injection volume	50 microliter	70 microliter
4	Detection wavelength	212 nm	216 nm
5	pH of mobile phases	4.8	5.2

Fig. 2. Definition of factors and levels.

each moment the program with or without saving all data (see Fig. 2).

For the choices to be performed by the analyst, user-friendly radio-buttons, sliders and pop-up menus are used with the recommended choices as default values. Question sections are pruned if they are not required. The results are represented numerical and visual. A report can be generated containing the complete set-up and all the results. This report can be used in other applications or imported in other documents.

The program can be downloaded at <http://fab.vub.ac.be/~fab/rts.html>

### 3.1. Creation of the experimental set-up

#### 3.1.1. Selection and definition of the factors and levels to test

With RTS one can investigate 2–20 factors, each at two or three levels. After defining the number of factors and levels, the user can describe

the factors and their levels. In Fig. 2 an example is shown for testing five factors of an HPLC method at two levels. The factors can often easily be selected from the operating procedure for the method. The levels should represent the maximum difference of the values of the factors that can be expected when the method is transferred between different laboratories, different devices, different experimentors, etc. If the factors are tested at only two levels, then they are generally tested at a low and a high extreme level, that is a level smaller and larger respectively than the nominal one (specified in the operating procedure). If the factors are tested at three levels, the nominal and two extremes are usually selected.

A decision support system for the selection of factors and their levels was programmed in Asymetrix Toolbook v1.5 (see Fig. 3). This external module can be called from RTS if desired. Factors (mostly related to chromatography methods) can be selected from a list (see Table 1). For

Proposed factors	Selected factors	Level -1	Level 0	Level 1
<b>Sample preparation</b>	Flow rate	0.9 ml/min	1 ml/min	1.1 ml/min
Sample weight	Column temperature	20	25	30
Shake time	Injection volume	50 microliter	60 microliter	70 microliter
Sonication time	Detector wavelength	212 nm	214 nm	216 nm
Heating temperature	pH of mobile phase	4.8	5.0	5.2
Wash volume				
Extraction volume				
Centrifugation time				
Centrifugation speed				
Concentration internal standard				
Volume derivatisation agent				
Concentration derivatisation agent				
Derivatisation time				
Derivatisation temperature				
Volume of pH adjusting solution				
Concentration of pH adjusting solution				
pH of solution				
Extraction time				
Dilution factor				
<b>Chromatography</b>				
pH of mobile phase				
pH of buffer				
Ionic strength of buffer				
Buffer concentration				

Fig. 3. Decision support system for the selection of factors and their levels.

each factor recommended levels are available. New factors can be added to the list by the analyst at run time. For certain factors, such as mixture compounds, buffers and columns, additional advice is given as described in [9].

### 3.1.2. Selection of the experimental design

RTS proposes a limited number of experimental designs as a function of the number of investigated factors (see Fig. 4). Factorial [10] and Plackett–Burman [11] designs are included in the program (see Table 2). The user can choose between recommended and minimal designs.

The recommended designs were applied in the original strategy [6]. These designs were chosen so that the number of experiments is the lowest possible, but still allow statistical interpretation. Namely, in the factorial designs, the two-factor interactions are not confounded with the main effects (design resolution is at least IV) and at

least three two-factor interactions can be calculated. The Plackett–Burman designs were chosen so that at least three dummy factors are included.

For examining five to eight factors with the recommended designs, there is also the choice between a fractional factorial and a Plackett–Burman design. Only the fractional factorial designs allow to estimate the main effects without being confounded with the two-factor interactions, but for the Plackett–Burman design on the other hand less experiments are needed [6].

The designs of the recommended strategy proved to work in a number of case studies [2–5], but not everyone is willing to perform more experiments than what is strictly necessary and not everybody wants a statistical interpretation of the effects. Often it is sufficient to have an idea of the magnitude of the effects. Therefore minimal designs were added to the strategy [6]. They were chosen so that the number of experiments is the

Table 1  
List of proposed factors and their levels in external module

Factors	Distance between levels
<b>Sample preparation</b>	
Sample weight	1%
Shake time	20%
Sonication time	20%
Heating temperature	5°C
Wash volume	20%
Extraction volume	20%
Centrifugation time	1 min
Centrifugation speed	500 rpm
Concentration internal standard	20%
Volume derivatisation agent	10%
Concentration derivatisation agent	10%
Derivatisation time	20%
Derivatisation temperature	5°C
Volume of pH adjusting solution	10%
Concentration of pH adjusting solution	10%
pH of solution	0.2
Extraction time	10%
Dilution factor	
<b>Chromatography</b>	
pH of mobile phase	0.2
pH of buffer	0.2
Ionic strength of buffer	5%
Buffer concentration	1%
Column temperature	5°C
Injection volume	10 µl
Injection concentration	1%
Flow rate	0.1 ml min <sup>-1</sup>
Concentration tailing suppressor	1%
Slope gradient	5%
Volume organic modifier	5%
Volume aqueous phase	5%
<b>Detector</b>	
wavelength	5 nm
Time constant	0.1 s
Bandwidth	2 nm
Detector cell temperature	5°C
<b>Non procedure related</b>	
Different analysts	
Different instruments	
Different days	1 day
Different laboratories	
Different lot of reagents	
Column manufacturer	
Lifetime column	
Column batch	
Different type of pipette	

lowest possible. This usually means that the statistical interpretation applied for the recommended designs is no longer possible. Some of the minimal designs are expandable to designs similar to the recommended designs [6]. After the evaluation of the results obtained with the minimal designs, one can decide whether or not to expand the design. Expansion is done by performing additional experiments. For fractional factorial designs for instance, another well-defined fraction of the full factorial is performed. The combination of this new design and the original minimal design gives a new one (expanded design) with resolution IV, similar to the recommended designs. For five factors, the minimal design is a quarter-fraction factorial design with eight experiments. This can be expanded to a half-fraction factorial design with 16 experiments, similar to the recommended design for five factors, by performing eight additional experiments.

If the factors are tested at three levels, the designs of Table 2 are reflected. This means that the designs are performed twice: once with the low extreme level and the nominal level, once with the high extreme level and the nominal level. However, such reflected designs are not recommended, since they require additional experiments and most of the time do not result in additional information.

As can be seen in Fig. 4, for each proposed design it is mentioned how many experiments are needed and how the statistical and graphical interpretation will be performed.

### 3.1.3. Sequence of experiments

The design experiments are normally performed in a random sequence. However the user can choose to sort (block) the experiments by one to three factors. This means that first all the experiments with the blocked factor at one level are performed, and thereafter all the experiments with the blocked factor at the other level(s). Within the blocks the experiments are still randomized. This sorting can be useful for practical reasons, but if drift occurs, the estimated effect of the blocked factor(s) will be affected most by the drift [5].

A possible check for drift is also included in RTS. This is done by performing replicated nomi-

Table 2  
Recommended, minimal and expanded designs included in RTS

Number of factors	Recommended design	Number of experiments	Minimal design	Number of experiments	Expanded design	Number of experiments
2	Full factorial for 3 factors: $2^3$	8	Full factorial for 2 factors: $2^2$	4	Full factorial for 3 factors: $2^3$	8
3	Full factorial for 3 factors: $2^3$	8	Half fraction factorial for 3 factors: $2^{3-1}$	4	Full factorial for 3 factors: $2^3$	8
4	Half-fraction factorial for 4 factors: $2^{4-1}$	8	Half-fraction factorial for 4 factors: $2^{4-1}$	8	—	—
5	Half-fraction factorial for 5 factors: $2^{5-1}$	16	Quarter-fraction factorial for 5 factors: $2^{5-2}$	8	Half-fraction factorial for 5 factors: $2^{5-1}$	16
6	Plackett–Burman design for 11 factors	12	—	—	—	—
6	Quarter-fraction factorial: $2^{6-2}$	16	Eighth-fraction factorial: $2^{6-3}$	8	Quarter-fraction factorial: $2^{6-2}$	16
7	Plackett–Burman design for 11 factors	12	—	—	—	—
7	Eighth-fraction factorial: $2^{7-3}$	16	Sixteenth-fraction factorial: $2^{7-4}$	8	Eighth-fraction factorial: $2^{7-3}$	16
8	Plackett–Burman design for 11 factors	12	—	—	—	—
8	Sixteenth-fraction factorial: $2^{8-4}$	16	—	—	—	—
9–11	Plackett–Burman design for 11 factors	12	Plackett–Burman design for 11 factors	12	—	—
9–11	Plackett–Burman design for 15 factors	16	Plackett–Burman design for 11 factors	12	—	—
12	Plackett–Burman design for 15 factors	16	Plackett–Burman design for 15 factors	16	—	—
13–15	Plackett–Burman design for 19 factors	20	Plackett–Burman design for 15 factors	16	—	—
13–15	Plackett–Burman design for 19 factors	20	Plackett–Burman design for 15 factors	16	—	—
16	Plackett–Burman design for 19 factors	20	Plackett–Burman design for 19 factors	20	—	—
17–19	Plackett–Burman design for 23 factors	24	Plackett–Burman design for 19 factors	20	—	—
17–19	Plackett–Burman design for 23 factors	24	Plackett–Burman design for 19 factors	20	—	—
20	Plackett–Burman design for 23 factors	24	Plackett–Burman design for 23 factors	24	—	—

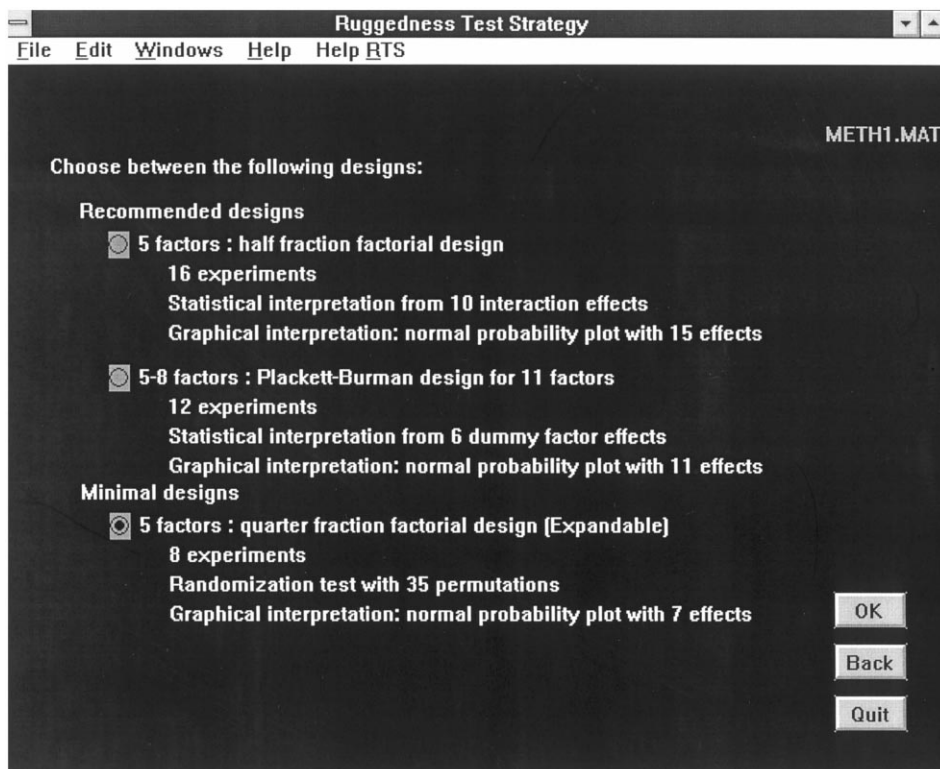


Fig. 4. Selection of design.

nal experiments at regular times between the design experiments. The user can choose the number of experiments at nominal level from a proposed list. These nominal experiments are used to check visually the drifting of the nominal response as a function of time and to perform a correction for the occasionally occurring drift.

If the user chose not to check for drift, the possibility is provided to add a nominal experiment before and after the design experiments. These experiments allow to check if the responses from the experiments at nominal level are as expected at the beginning and the end of the experiments, to obtain an idea of occasionally occurring drift and to normalize the effects [6].

RTS then presents the matrix of the non-randomized or sorted design, the matrix of the experiments ranked in the order in which they must be performed and a full description of the levels of each factor for each experiment, ready to be used by the analyst.

In the following, an example of the output is shown for testing five HPLC factors (shown in Fig. 2) with a minimal design consisting of eight experiments. The experiments are sorted by the factor 'pH of mobile phases' (factor E) and every four design experiments a nominal experiment is added.

#### *Matrix of the design*

Five factors: quarter fraction factorial design (Expandable)

Exp	A	B	C	D	E	BC + DE	BE + CD
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1

6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Matrix of ranked experiments

Exp	A	B	C	D	E
0	0	0	0	0	0
2	1	-1	-1	-1	-1
7	-1	1	1	-1	-1
4	1	1	-1	1	-1
5	-1	-1	1	1	-1
0	0	0	0	0	0
1	-1	-1	-1	1	1
6	1	-1	1	-1	1
3	-1	1	-1	-1	1
8	1	1	1	1	1
0	0	0	0	0	0

*Experimental set-up*

## 1: Experiment 0 (experiment at nominal level)

Flow rate = 1.0 ml min<sup>-1</sup>  
 Column temperature = 25°C  
 Injection volume = 60 µl  
 Detection wavelength = 214 nm  
 pH of mobile phases = 5.0

## 2: Experiment 2

Flow rate = 1.1 ml min<sup>-1</sup>  
 Column temperature = 20°C  
 Injection volume = 50 µl  
 Detection wavelength = 212 nm  
 pH of mobile phases = 4.8  
 ...[Truncated]...

## 11: Experiment 0 (experiment at nominal level)

Flow rate = 1.0 ml min<sup>-1</sup>  
 Column temperature = 25°C  
 Injection volume = 60 µl  
 Detection wavelength = 214 nm  
 pH of mobile phases = 5.0

## 3.2. Treatment of the experimental results

## 3.2.1. Calculation and presentation of effects

After performing the experiments, the experi-

mental set-up can be loaded again into the second part of RTS for the treatment of the results and the obtained data are entered manually into the system. This part of the program can, if necessary, be used more than once for a given experimental set-up, each time for a different response. Resolutions between peaks, contents of the main substances, peak heights, etc. are examples of responses which could be examined for chromatographic methods.

If one checked for drift, the corrected response results are calculated [6] and presented, and the nominal results are plotted as a function of time (see Fig. 5). The percentual change of the response between the beginning and the end of the experiments is calculated. Based on this information, the analyst can decide if drifting indeed occurs and whether he uses the effects calculated from the measured or corrected responses.

Effects are calculated from the measured results and if available also from the corrected results [6]. If the factors are tested at three levels, two effects are calculated: one for the low level interval and one for the high level interval. The effects are presented in tables (see Fig. 6) and in bar plots (see Fig. 7).

## 3.2.2. Statistical interpretation of the results

To identify statistically significant effects from the **recommended designs**, a *t*-test is performed. The absolute value of the effect of a factor *X* ( $|E_X|$ ) is considered to be significant if the *t*-test value (*t*) is larger than a critical value (*t*<sub>critical</sub>), which is the tabulated *t*-value with an appropriate number of degrees of freedom

$$t = \frac{|E_X|}{(SE)_e} \Leftrightarrow t_{\text{critical}} \quad (1)$$

or, in other words, if the effect of the factor is larger than a critical effect value ( $E_{\text{critical}}$ )

$$|E_X| \Leftrightarrow E_{\text{critical}} = t_{\text{critical}} \cdot (SE)_e \quad (2)$$

The standard error (SE)<sub>e</sub> is estimated from effects considered negligible, namely two-factor



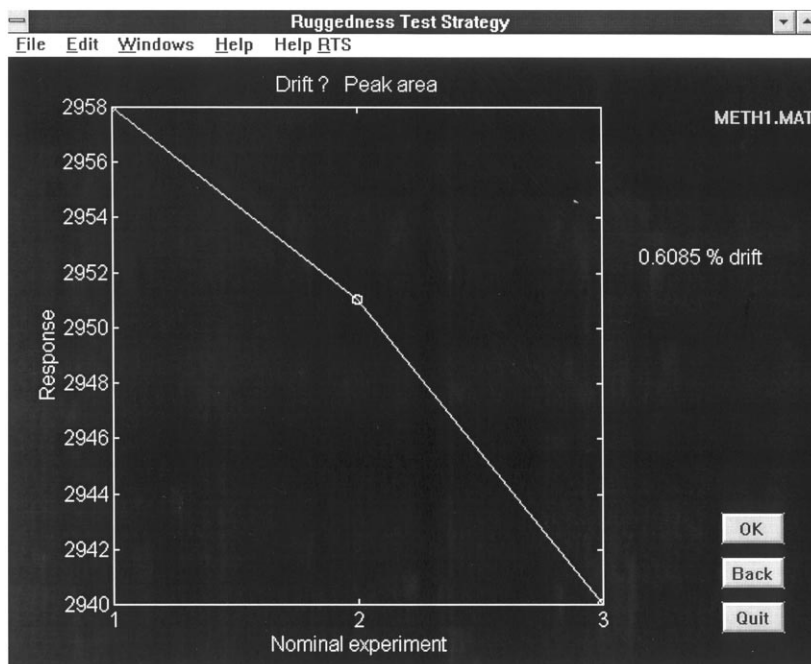


Fig. 5. Nominal results plotted in function of time.

interaction effects for factorial designs or dummy factor effects for Plackett–Burman designs.

The critical effect is calculated for a significance level  $\alpha = 0.05$  and  $\alpha = 0.01$ . Significance of an estimated effect at  $\alpha = 0.05$  is indicated with \* and at  $\alpha = 0.01$  with \*\*. These critical effects are also indicated on the bar plots of the effects.

The **minimal designs** do not always have enough interactions or dummy factor effects for a statistical interpretation. Other possibilities [2] exist, such as estimating  $(SE)_e$  from (i) the variance from R replicate measurements at nominal level, (ii) the variance from duplicated experiments in the design, (iii) the variance of the experiments of the design or (iv) one could perform an independent  $t$ -test with hypothesis  $H_0: \bar{Y}(+) = \bar{Y}(-)$  where  $\bar{Y}(+)$  and  $\bar{Y}(-)$  are the mean of the responses where factor  $X$  is respectively at the high level and low level. However all those methods seem to overestimate or underestimate the experimental error in most cases [2]. Methods (i) and (ii) require additional experiments, which must be avoided in a minimal design. Method (iv) is based on the hypothesis of normal distribution

and homoscedasticity, which will not be the case when there is an effect.

Therefore randomization tests [12,13] were examined as statistical interpretation method. In a randomization test the significance is not determined from statistical tables, but from the distribution of a test statistic generated by randomly assigning the experimental data to the different conditions (methods, treatments, levels of factors,...). They have the advantage that they are truly distribution-free: no assumptions of normality or homoscedasticity are required. The hypothesis test here is not based on whether a certain  $t$ -value is exceeded, but on ranking. In our strategy, a test statistic ( $t$ ), equivalent to that for an independent  $t$ -test,

$$t = |\bar{Y}(+) - \bar{Y}(-)| \quad (3)$$

is calculated for each factor  $X$ . Then the data (i.e. the measured responses) are permuted repeatedly and the test statistic is calculated for each of the resulting data permutations. These data permutations, including those representing the examined factor, constitute the reference set for determining

Factors (interactions)	Effect	Normalized effect	Probability	Significance
Flow rate	-0.85	-0.02874	1	-
Column temperature	10.45	0.3533	0.6	-
Injection volume	11.55	0.3905	0.5429	-
Detection wavelength	-4.35	-0.1471	0.7714	-
pH of mobile phases	-86.25	-2.916	0.02857	**
BC+DE	2.85	0.09635	0.8571	-
BE+CD	4.45	0.1504	0.7429	-

METH1.MAT  
Peak area

Effects from corrected results

Print  
OK  
Back  
Quit

Fig. 6. Tables of effects.

significance. A simulated example is shown in Table 3: the effects of four factors are examined at two levels in eight experiments. Permutation 1 (first row) is considered as the original data and represents the effect of factor A: the four first experiments, factor A was at level (–) and the four last experiments factor A was at level (+). The reference set exists of every possible partition of the eight responses that assigns four to level (–) and four to level (+). Some of the permutations can represent another factor, if one look e.g. at Table 4 how factor B is represented (– – + + – – + +), one finds that this equals permutation 10. The proportion of data permutations in the reference set that have test statistic values larger than or equal to the value for the examined factor is the *P*-value (significance or probability value). The effects are considered significant when the *P*-value is smaller than 0.01 ( $\alpha = 0.01$ ) or 0.05 ( $\alpha = 0.05$ ). Significance of an estimated effect at  $\alpha = 0.05$  is indicated with \* and

at  $\alpha = 0.01$  with \*\* (see Fig. 6). If there are only 35 data permutations (eight experiments) an effect is indicated with \* if the *P*-value is 2/35 (second largest test statistic) and with \*\* if the *P*-value is 1/35 (largest test statistic). In our example, the factor B has as *P*-value 1/35 (lowest probability) and is therefore considered as significant on the level  $\alpha = 0.01$ .

The disadvantage of randomization tests is the long computer time needed for in some cases large number of permutations. This time is minimized by using the simplified equivalent test statistic (Eq. (3)) and by using random data permutation. The simplified test statistic can be used because it results in the same ranking order as the test statistic of the *t*-test [12]. When using random data permutation, the test statistic values are only calculated for a certain number of permutations, randomly selected from all the possible ones. In RTS 2000 random data permutations are calculated or all the systematic data permutations if

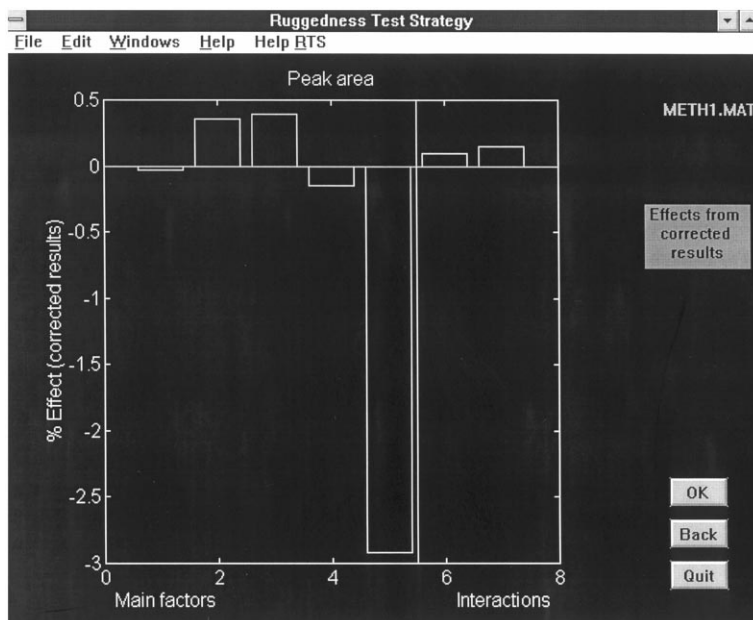


Fig. 7. Bar plot of effects.

there are less than 2000. It has been shown that the power of a randomization test with as few as 1000 random data permutations is almost equal to that of the randomization test with systematic data permutation [14].

However some problems are to be expected when there is more than one significant effect. A strong significant effect will probably cause an underestimation of the significance of the other effects. This is due to the fact that the test statistics for certain data permutations may be confounded by other effects of the design. This is solved by performing a correction for each significant effect. If a strong significant effect is found, the responses are corrected by adding half of the absolute value of the effect to the responses belonging to the level which gives the lowest responses, and by subtracting half of the absolute value of the effect from the responses belonging to the level which gives the highest responses (i.e. elimination of highest effect from data set). The randomization test is then computed with the corrected responses. This is repeated until no new significant effects are found. An example of this procedure on the data of Table 3 is shown in

Table 4. Four factors are tested at two levels in eight experiments. With the *t*-test based on the interaction effects, three significant effects are found (A, B and C). The randomization test only indicates one effect (B) as significant. The significance of the effects of factor A and C is underestimated due to the confounding of the effect of factor B in other data permutations. After correction of the responses for factor B, two significant effects (B and A) are found. After correction for factor A, three significant effects (A, B and C) are found. Correction for factor C doesn't result in more significant effects. This means that the same effects are indicated as significant with the *t*-test based on interaction or dummy effects as with the randomization tests with correction.

The **expanded designs** are statistically interpreted in the same way as the recommended designs.

### 3.2.3. Graphical interpretation of the effects

The normal distribution of the effects is verified graphically by plotting the effects in a normal probability plot (see Fig. 8) [15,16]. Non-significant effects are normally distributed around zero

Table 3  
35 data permutations for 8 experiments (2 levels)

Permutation	Level (-)				Level (+)				Test statistic	Ranking
1 (factor A)	823	876	1021	1078	928	979	1119	1177	101.25	9
2	823	876	1021	928	1078	979	1119	1177	176.25	2
3	823	876	1021	979	1078	928	1119	1177	150.75	3
4	823	876	1021	1119	1078	928	979	1177	80.75	13
5	823	876	1021	1177	1078	928	979	1119	51.75	20
6	823	876	1078	928	1021	979	1119	1177	147.75	4
7	823	876	1078	979	1021	928	1119	1177	122.75	7
8	823	876	1078	1119	1021	928	979	1177	52.25	19
9	823	876	1078	1177	1021	928	979	1119	23.55	30
10 (factor B)	823	876	928	979	1021	1078	1119	1177	197.25	1
11	823	876	928	1119	1021	1078	979	1177	127.25	5
12	823	876	928	1177	1021	1078	979	1119	98.25	11
13	823	876	979	1119	1021	1078	928	1177	101.75	8
14	823	876	979	1177	1021	1078	928	1119	72.75	16
15	823	876	1119	1177	1021	1078	928	979	2.75	32
16	823	1021	1078	928	876	979	1119	1177	75.25	15
17	823	1021	1078	979	876	928	1119	1177	49.75	21
18	823	1021	1078	1119	876	928	979	1177	20.25	31
19	823	1021	1078	1177	876	928	979	1119	49.25	22
20	823	1021	928	979	876	1078	1119	1177	124.75	6
21 (factor C)	823	1021	928	1119	876	1078	979	1177	54.75	18
22	823	1021	928	1177	876	1078	979	1119	25.75	28
23	823	1021	979	1119	876	1078	928	1177	29.25	25
24	823	1021	979	1177	876	1078	928	1119	0.25	35
25	823	1021	1119	1177	876	1078	928	979	69.75	17
26	823	1078	928	979	876	1021	1119	1177	96.25	13
27	823	1078	928	1119	876	1021	979	1177	26.25	27
28	823	1078	928	1177	876	1021	979	1119	2.75	33
29 (factor D)	823	1078	979	1119	876	1021	928	1177	0.75	34
30	823	1078	979	1177	876	1021	928	1119	28.25	26
31	823	1078	1119	1177	876	1021	928	979	98.25	10
32	823	928	979	1119	876	1021	1078	1177	75.75	14
33	823	928	979	1177	876	1021	1078	1119	46.75	24
34	823	928	1119	1177	876	1021	1078	979	23.25	29
35	823	979	1119	1177	876	1021	1078	928	48.75	23

and in a normal probability plot they tend to fall on a straight line, while significant effects deviate from this line and are located at the sides of the figure. This line preferably should fit to the non-significant experiments. Drawing the least squares line through all effects is not a good solution, since significant effects are in fact outliers to the straight line formed by the non-significant ones and they will attract the line into their direction. This makes it, in some cases, more difficult to judge whether a point deviates from the non-significance line or not. Therefore a more robust line

is obtained by calculating the least squares line through the smallest (in absolute values) 80% of the effects.

#### 4. Case study

As an example we consider one of the experiments described in reference [5]. The effect of four factors on the HPLC assay for tetracycline (TC) and its impurities was examined. For four factors, the RTS program proposes (see Table 2) as rec-

Table 4  
Comparison of *t*-test and randomization tests with and without correction

	Factor A	Factor B	Factor C	Factor D	AB and CD	AC and BD	AD and BC	Response
Experiment 1	-	-	-	-	+	+	+	823
Experiment 2	-	-	+	+	+	-	-	876
Experiment 3	-	+	-	+	-	+	-	1021
Experiment 4	-	+	+	-	-	-	+	1078
Experiment 5	+	-	-	+	-	-	+	928
Experiment 6	+	-	+	-	-	+	-	979
Experiment 7	+	+	-	-	+	-	-	1119
Experiment 8	+	+	+	+	+	+	+	1177
Effect	101.3	197.3	54.75	0.75	2.75	-0.25	-2.75	
Significance ( <i>t</i> -test/interaction effects)	**	**	**	-	-	-	-	
Significance (randomization test)	-	**	-	-	-	-	-	
Significance (randomization test+correction for B)	**	(**)	-	-	-	-	-	
Significance (randomization test+correction for C)	(**)	(**)	**	-	-	-	-	
Significance (randomization test+correction for A)	(**)	(**)	(**)	-	-	-	-	

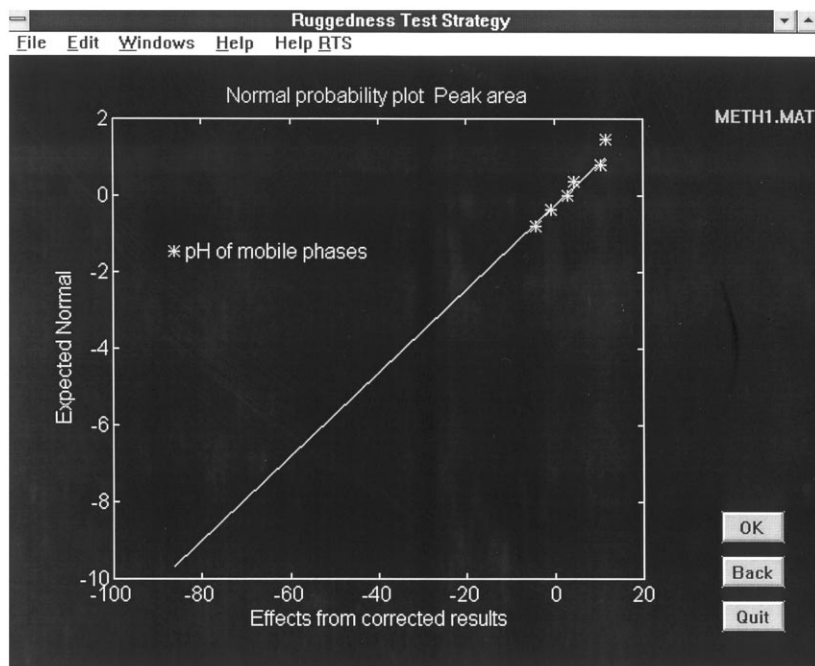


Fig. 8. Normal probability plot of effects.

ommended design a half-fraction factorial design of eight experiments with a statistical interpretation based on three interaction effects. As minimal design, the same half-fraction factorial design of eight experiments is proposed, but when this option of the program is chosen, a randomization test with 35 permutations will be used as statistical interpretation method. For examining four factors, the minimal design has no benefits (the number of experiments could not be reduced here), but in this case study it is interesting for comparisons. Originally the recommended design option was selected. The report file is shown below for the response 'relative retention of tetracycline (to epitetracycline)':

RTS—RUGGEDNESS TEST STRATEGY  
VERSION 1.1.11 (Dec 1997)  
Report RRET\_TC.MAT  
22-Dec-97

Description of tested factors and their levels

Factor	Low level	High level
--------	-----------	------------

NH <sub>4</sub> ox/ NH <sub>4</sub> phosphate	0.19 M/0.38 M	0.21 M/0.42 M
Vol. DMF in mob. phase	260 ml	280 ml
pH of mobile phase	7.45	7.85
Flow of mobile phase	0.9 ml min <sup>-1</sup>	1.1 ml min <sup>-1</sup>

Matrix of the design.

Four factors: Half-fraction factorial design

Exp	A	B	C	D	AB +CD	AC +BD	AD +BC
1	-1	-1	-1	-1	1	1	1
2	1	-1	-1	1	-1	-1	1
3	-1	1	-1	1	-1	1	-1
4	1	1	-1	-1	1	-1	-1
5	-1	-1	1	1	1	-1	-1

6	1	-1	1	-1	-1	1	-1
7	-1	1	1	-1	-1	-1	1
8	1	1	1	1	1	1	1

No check for drift was performed  
Effects from measured results

Matrix of ranked experiments

Exp	A	B	C	D
1	-1	-1	-1	-1
2	1	-1	-1	1
3	-1	1	-1	1
4	1	1	-1	-1
5	-1	-1	1	1
6	1	-1	1	-1
7	-1	1	1	-1
8	1	1	1	1

Factor	Effect	Significance
NH4ox/ NH4phosphate	-0.00375	—
ml DMF in mob phase	-0.1113	*
pH of mobile phase	0.3488	**
Flow (ml min <sup>-1</sup> )	0.02925	—
AB+CD	0.02325	—
AC+BD	-0.00375	—
AD+BC	-0.03725	—
Critical value 5%	0.08096	
Critical value 1%	0.1486	

Experimental set-up

1: Experiment 1

NH4ox/NH4 phosphate = 0.19 M/0.38 M  
Vol. DMF in mob. phase = 260 ml  
pH of mobile phase = 7.45  
Flow = 0.9 ml min<sup>-1</sup>  
...[Truncated]...

8: Experiment 8

NH4ox/NH4 phosphate = 0.21 M/0.42 M  
Vol. DMF in mob. phase = 280 ml  
pH of mobile phase = 7.85  
Flow = 1.1 ml min<sup>-1</sup>

Response: Rel Retention TC

Exp	Response
1	1.5890
2	1.5950
3	1.5210
4	1.5150
5	2.0080
6	1.9480
7	1.8070
8	1.8520

In order to compare the *t*-test based on the interaction effects with the randomization test, the computations were redone for the case in which the minimal design was selected. This gave the following results:

Factor	Effect	Probability	Significance
NH4ox/ NH4phosphate	-0.00375	0.9714	—
ml DMF in mob phase	-0.1113	0.02857	**
pH of mobile phase	0.3488	0.02857	**
Flow (ml min <sup>-1</sup> )	0.02925	0.7714	—
AB+CD	0.02325	0.8571	—
AC+BD	-0.00375	0.9714	—
AD+BC	-0.03725	0.7143	—

With the *t*-test based on interaction effects, the effect of factor 'Volume DMF in mobile phase' was found to be significant at the 5% significance level and the effect of factor 'pH of mobile phase' at the 1% level. The randomization test indicates both effects as significant at the 1% level (explanation, see higher). However, the difference between both results is not large if one considers that in the first computation the absolute value of the effect of factor 'Volume DMF in mobile phase' (0.1113) was close to the critical effect at the 1% level (0.1486). It has to be remarked that two different effects can have the same probability value, due to the repeated randomization test (see higher).

## 5. Conclusion

A MATLAB computer program is developed for the experimental set-up and interpretation of ruggedness tests for mainly procedure related factors.

The program is designed to facilitate the selection of the designs and the interpretation of the results and to prevent or detect problems such as drifting. The graphical interface, on-line help menus, default values, etc. make RTS very user friendly, so that only knowledge about the analytical method is needed to define possibly critical factors and to decide which measurement responses need to be studied. Thanks to the decision support system for the selection of factors and their levels, it is less likely that the user will forget some important factors or choose too extreme values for the extreme levels.

The strategy and program have been successfully applied in academical study and industrial practice.

## Acknowledgements

Parts of this work were funded by the Research Contract Nr. 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 thank C. Hartmann of Novartis for the scientific discussions and for her advice about the RTS-program.

## References

- [1] Y. Vander Heyden, D.L. Massart, Review of the use of robustness and ruggedness in analytical chemistry, in: M.M.W.B. Hendriks, J. De Boer, A.K. Smilde (Eds.), *Robustness of Analytical Chemical Methods and Pharmaceutical Technological Products, Data Handling in Science and Technology 19*, Elsevier, Amsterdam, 1994, pp. 79–147.
- [2] 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.
- [3] Y. Vander Heyden, C. Hartmann, D.L. Massart, L. Michel, P. Kiechle, F. Erni, Ruggedness tests on a HPLC assay: comparison of tests at two and three levels by using two-level Plackett–Burman designs, *Anal. Chim. Acta* 316 (1995) 15–26.
- [4] 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 weight polymers, *J. Chromatogr. A* 756 (1996) 89–106.
- [5] Y. Vander Heyden, D.L. Ann Bourgeois, Massart, Influence of the sequence of experiments in a ruggedness test when drift occurs, *Anal. Chim. Acta* 347 (1997) 369–384.
- [6] 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.
- [7] 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, *Chemometr. Intell. Lab. Syst.* 10 (1991) 337–347.
- [8] G.T. Wernimont, in: W. Spendley (Ed.), *Use of Statistics to Develop and Evaluate Analytical Methods*, Association of Official Analytical Chemists, Arlington, VA, pp. 78–82.
- [9] Y. Vander Heyden, F. Questier, D.L. Massart, Ruggedness testing of chromatographic methods: selection of factors and levels, accepted for publication, *J. Pharm. Biomed. Anal.* 18 (1998) 43–56.
- [10] E. Morgan, *Chemometrics, Experimental design, Analytical Chemistry by Open Learning*, J. Wiley and Sons, Chichester, 1991.



- [11] R.L. Plackett, J.P. Burman, The design of optimum multifactorial experiments, *Biometrika* 33 (1946) 305–325.
- [12] E.S. Edgington, in: D.B. Owen (Ed.), *Randomization Tests*, 2nd edn., Marcel Dekker, New York, 1987.
- [13] P. Vankeerberghen, C. Vandenbosch, J. Smeyers-Verbeke, D.L. Massart, Some robust statistical procedures applied to the analysis of chemical data, *Chemometr. Intell. Lab. Syst.* 12 (1991) 3–13.
- [14] E.S. Edgington, Approximate randomization tests, *J. Psychol.* 72 (1969) 143–149.
- [15] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: a Textbook*, Elsevier, Amsterdam, 1988, p. 488.
- [16] A. Holtsberg, A rudimentary statistics toolbox (qqnorm.m), Lund Institute of Technology, Department of Mathematical Statistics, <http://www.maths.lth.se/matstat/andersh/stibox.tar>