Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 992 (2003) 11–27
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Advances in methodology for the validation of methods according to the International Organization for Standardization Application to the determination of benzoic and sorbic acids in soft drinks by high-performance liquid chromatography

Inmaculada García[a], M. Cruz Ortiz[a,*], Luis Sarabia[b], Carmen Vilches[c], Elisa Gredilla[c]

[a]*Department of Chemistry, Faculty of Sciences, University of Burgos, Pza. Misael Bañuelos s/n 09001, Burgos, Spain*
[b]*Department of Mathematics and Computation, Faculty of Sciences, University of Burgos, Pza. Misael Bañuelos s/n 09001, Burgos, Spain*
[c]*Servicio Territorial de Sanidad y Bienestar Social de Burgos, Junta de Castilla y León, Paseo Sierra de Atapuerca s/n 09071, Burgos, Spain*

Received 19 September 2002; received in revised form 27 January 2003; accepted 27 January 2003

Abstract

Robust chemometric techniques such as least median of squares regression, H15 Huber estimator and Lenth's method are fundamental tools in the validation of analytical methods since they contribute the strategies needed to estimate efficiently parameters such as robustness, linear range, selectivity, accuracy (trueness and precision) and the capability of detection. In addition, the capability of discrimination defined as a generalisation of the capability of detection for any nominal concentration is evaluated. The new strategy proposed is applied to the validation of a chromatographic method for use in systematic analysis.

Keywords: Validation; Accuracy; Soft drinks; Food analysis; Regression analysis; Chemometrics; Robust testing; Benzoic acid; Sorbic acid; Organic acids

1. Introduction

The quality of food products, the sanitary state and the commercial value of many products are established on the basis of chemical measurements. Thus, maintaining or improving the reliability of these determinations is a crucial aspect on which a great deal of effort is spent. Not only is the importance of chemical measurements clear, but also the need to guarantee their quality to evaluate, as far as is possible, the economic and social consequences which may result from mistaken analytical measurements. Legislation covering food additives, which is becoming stricter, obliges the analyst to develop new analytical methods and as a result to validate them, in order to identify the potential sources of error which may affect them. This is the objective of the method validation and must be done in accordance with the quality requirements established by the

*Corresponding author. Fax: +34-947-258-831.
E-mail address: mcortiz@ubu.es (M.C. Ortiz).

0021-9673/03/$ – see front matter
doi:10.1016/S0021-9673(03)00246-2

International Organization (ISO) [1], which are the accepted guidelines for the homologation and accreditation of chemical laboratories and analytical methods. Specifically, in this article we have used ISO 5725 [2] to study the accuracy (trueness [3] and precision [4,5]) of an analytical method. Also, in Section 5 [6] of this standard, the need is recognised to incorporate a robust methodology which as far as possible avoids the effect of anomalous data. To evaluate the detection limit understood as the capability of detection one can consult ISO 11843 [7]. In the second part of this standard the concept of capability of detection is applied to chemical analysis [8].

Nowadays several guides can be found whose purpose is to discuss the topics related to the validation method according to ISO. As an example EURACHEM has published a guideline [9]. In its Work Programme for the period 2000–2005, it includes some tasks such us ''facilitate the acceptance of traceability concepts and systems within the analytical chemistry community particularly in support of ISO/IEC 17025'' as well as ''establish links between the new ISO/IEC 17025 and regulators and users''.

The validation method has been done by the examination of the following parameters: robustness, linear range/linearity, sensitivity, selectivity, accuracy, trueness, precision, capability of detection and capability of discrimination.

The objective of this work is to establish a methodology to guarantee the validity of a chromatographic method together with the interpretation of the results obtained at each stage of the experimental validation process. To do this we have used a wide range of statistical techniques which guarantee the solidity of the conclusions reached. The statistical methodology usually advocated for the evaluation process, is based on the normality hypothesis but the inevitable presence of anomalous data makes automatic application of the formulas inadvisable. ISO itself recognises that the measurement process must be formulated statistically and that it must be validated with statistical guarantees. For this reason, in this paper we have resorted to the use of non-parametric statistical tools which do not require any particular distribution and are resistant to outlier data. They are based on the median which, in general, provides a better idea of the central trend than the mean of the data itself and since the presence of anomalous data is unavoidable, the use of robust statistical techniques is a key to achieve a reliable analysis. Specifically, the least median of squares regression (LMS), the H15 Huber estimator and Lenth's method have been successfully used for the simultaneous quantitative analysis of benzoic and sorbic acids in soft drinks by high-performance liquid chromatography (HPLC) with diode array detection (DAD) in UV–Vis.

The use of benzoic acid (E210), sorbic acid (E200) and their corresponding salts as preservatives in foods nowadays is very widespread. Despite the fact that fizzy drinks allow the growth of a low number of microorganisms, due to the action of the carbon dioxide itself, preservatives are required to prevent alterations during long periods of storage at room temperature. Although both acids are highly efficient [10] against yeasts, moulds and to a lesser extent against bacteria, sorbic acid has the advantage of being active in less acid media (it can be so until pH values of 6) and in addition has no taste, but has the inconvenience of disappearing in part when the product is boiled.

The permitted quantities of the two compounds in soft drinks has fallen since 1981. Currently, the European Union legislation (Directive 95/2/CE) establishes the limit of 150 mg l^{-1} of benzoic acid (E210) and 300 mg l^{-1} of sorbic acid (E200) if they are found separately and 150 and 250 mg l^{-1}, respectively, if they are found in combination. These values confirm the tendency to substitute benzoic acid and its salts for other preservatives which are equally efficient but do not give taste to foods and are even less toxic.

The Association of Official Analytical Chemists (AOAC) [11] proposes a gas chromatography method for the simultaneous analysis of benzoic and sorbic acids. According to the procedure, both acids must be extracted with ether and then they are transformed into trimethylsilyl (TMS) esters through a pre-column derivatization technique. For the quantitative determination, phenylacetic and caproic acids are used as internal standards for benzoic and sorbic acids, respectively. This process clearly requires excessive analysis time (approximately 45 min per sample) and is too complex. HPLC nowadays is the

preferred analytical technique for the determination of preservatives in food [12–14]; they can also be analysed through capillary electrophoresis [15]. Along these lines, the method proposed in this paper simplifies considerably the analysis, reducing its cost and time (10–12 min).

2. Experimental

2.1. Chemicals

Benzoic acid (>99%) was supplied by Merck (Hohenbrunn, Germany), whereas sorbic acid (>99%) was obtained from Scharlau Chemie (Barcelona, Spain); both of them were used without further purification. Acetonitrile, acetic acid glacial and ammonium acetate, employed for the preparations of the mobile phase were purchased from Panreac (Barcelona, Spain) and they were of HPLC grade. The water used for preparing the buffer and the standard solutions was obtained by the Milli-Q Gradient A10 water purification system of Millipore (Bedford, MA, USA).

2.2. Apparatus

The chromatographic analysis was carried out in a Waters high-performance liquid chromatograph (Bedford, MA, USA) equipped with a Waters 510 Pump and a Waters 717 Injector. The analytical column which operates at room temperature was a Nova-Pak C_{18}, 150×3.9 mm from Waters, and the analysis involving diode array detection was performed in a Waters 996 UV–Vis absorbance detector. The Millenium 32 (version 3.05.01, 1998, Waters) software was used to control the system.

2.3. Chromatographic conditions

The mobile phase contains a mixture of acetonitrile–acetate buffer adjusted to pH 4.4 (40:60, v/v). The buffer was prepared by dissolving 3.84 g of ammonium acetate in 1 l of water and adjusting the pH to 4.4 with acetic acid. Before use, the effluent was filtered through a 0.45-μm membrane filter and degassed in an ultrasonic bath. The chromatographic separation was achieved with isocratic elution at a flow-rate of 0.8 ml min^{-1} and 20 μl of sample were injected into the chromatographic system. The area of both peaks was registered at 254 nm. After each injection, the injector was washed with a mixture of methanol–water (10:90, v/v). The chromatograph requires approximately 40 min to reach the equilibrium.

2.4. Standards and sample solutions

Stock solutions of benzoic acid and sorbic acid at a concentration of 400 mg l^{-1} were separately prepared in water slightly basified with 0.5 ml of 1 *M* NaOH. The stock solution of sorbic acid was then diluted in water to obtain a solution with a final concentration of 20 mg l^{-1}. The working solutions used to build the calibration curves were prepared daily by diluting in water the stock solution of benzoic acid (400 mg l^{-1}) and the diluted solution of sorbic acid (20 mg l^{-1}) at several concentration levels. The calibration curves were built by measuring the peak area of 10 working solutions whose concentration is comprised between 5 and 140 mg l^{-1} of benzoic acid (standards were prepared every 15 mg l^{-1}) and between 0.2 and 3 mg l^{-1} of sorbic acid (standards are not equally spaced). Three instrumental replicates were carried out at the extremes of the calibration curve and two in the rest. The measurements corresponding to the same calibration curve were recorded the same day in increasing order of concentration.

The soft drink samples were adjusted to pH 3 by addition of 4 *M* HCl and then homogenised, filtered through a 0.45-μm filter, and degassed. If the concentration of the preservative in the beverage is higher than the largest one used to build the calibration curve, the drink sample is diluted in water. The presence of benzoic and sorbic acids can be verified by both the retention time and spectrum.

3. Results and discussion

Fig. 1 illustrates the chromatogram of a standard, which contains 140 mg l^{-1} of benzoic acid and 3 mg l^{-1} of sorbic acid; it was obtained with the experimental conditions indicated in Section 2.3. The elution order is (1) benzoic acid (retention time, 1.8

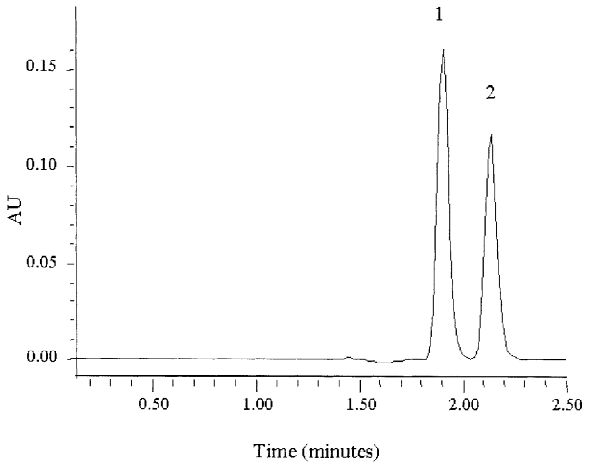

Fig. 1. Characteristic chromatogram of a standard which contains a mixture of 140 mg l^{-1} of benzoic acid (1) and 3 mg l^{-1} of sorbic acid (2). On the ordinate axis is the absorbance under the experimental conditions described in Section 2.3.

min; retention factor, 0.4) and (2) sorbic acid (retention time, 2.1 min; retention factor, 0.6). Values found in the separation factor, 1.5, and the resolution of the column, 1.9, indicate that the analytical method proposed in this work completely separates the analytes. The retention parameters have been estimated in terms of times according to the definitions provided by the IUPAC ([16], pp. 9–30, 9–31 and 9–36).

3.1. Robustness

To state that an analytical procedure is robust means [16] that ''the precision and trueness (accuracy) of the method are insensitive to minor changes in environmental and procedural variables, laboratories, personnel . . . ''. Otherwise, imputed performance characteristics would depend upon a number of uncontrolled factors and would have limited utility; from this it is apparent that robustness is an essential stage in the validation of a procedure. Despite the fact that IUPAC uses the term robustness, in the bibliography two slightly different concepts can be found: ruggedness [17] which evaluates the behaviour of the method compared with changes in the external experimental conditions (such as different laboratories, analysts, equipment, etc.) and robustness [18,19] which studies the effect of the experimental factors intrinsic to the method such as temperature, changes in the composition of the mobile phase, the reaction time, etc. Because the purpose of this paper is the internal validation, that is, the assurance of the accuracy of the method when small but inevitable changes in the experimental conditions occur, the robustness has been analysed instead of the ruggedness.

The influence of each experimental factor on the response has been evaluated using the Plackett–Burman design [20,21] which allows to estimate in an unbiased way the effect of $P-1$ factors with only P experiments (where P is a multiple of four); therefore, we are dealing with designs which allow for rapid a priori control of robustness [22].

The strategy followed to analyse the robustness is: (i) identify those factors which can influence the response, (ii) for each of these factors define the nominal and extreme levels to be accounted for in routine work, (iii) arrange the experimental plan according to the experimental design methodology, (iv) perform the experiments in random order and evaluate each factor effect.

Table 1 shows the seven factors examined together with their levels; the sign (+) corresponds to the nominal level which is the value of the factor normally given in the procedure, and (−) is the extreme level (the maximum separation admitted from the nominal level). Both values should be close so that the model which relates the signal to be analysed with the factors will be linear [22]. This way of describing the robustness was first described by Youden [23] and has been taken up by the AOAC.

Table 1
Experimental factors together with the nominal (+) and extreme (−) levels selected for the Plackett–Burman design

	Factor (units)	Level	
		+	−
b_1	Mobile phase: acetonitrile (%)	40	35
b_2	pH	4.4	4.0
b_3	Concentration of ammonium acetate (g l^{-1})	3.8	3.6
b_4	Mobile phase flow (ml min^{-1})	0.8	0.7
b_5	Injection volume (μl)	20	25
b_6	Wavelength (nm)	254	260
b_7	Stabilisation time (min)	40	30

Table 2
Plackett–Burmann design (experiments 1–8) with three replicates at the nominal level (last three rows) and experimental value of the response (mean concentration predicted from three replicates). (+) Stands for the nominal level and (−) for the extreme level

Order	Factors							Response (mg l^{-1})	
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	Benzoic	Sorbic
Design									
1st	+	+	+	+	+	+	+	88.31	160.35
5th	+	+	−	+	−	−	−	90.83	162.13
3rd	+	−	+	−	+	−	−	76.32	165.87
2nd	+	−	−	−	−	+	+	73.11	154.78
8th	−	+	+	−	−	+	−	86.40	162.70
4th	−	+	−	−	+	−	+	90.70	161.85
7th	−	−	+	+	−	−	+	90.67	161.70
6th	−	−	−	+	+	+	−	86.96	164.54
Replicates									
	+	+	+	+	+	+	+	88.21	164.50
	+	+	+	+	+	+	+	88.87	165.10
	+	+	+	+	+	+	+	87.62	159.70

See Table 1 for the codification of the factors.

To establish the robustness of the method, 11 experiments, each of them involving one calibration curve, were carried out (eight experiments of the Plackett–Burman design itself and three replicates at the nominal level). The conditions and the order in which experiments were performed are reported in Table 2. In each of the 11 experiments, the response to be analysed (the two last columns in Table 2) was the predicted mean concentration of a soft drink sample (three replicates were taken into account to calculate the mean concentration).

The weights of the factors and their P value were estimated by least-squares with NEMROD-W program [24] obtaining the results given in Table 3. To decide whether the change between the nominal and the extreme level of a factor affects the determination of the concentration, the hypothesis test for the significance of the coefficients [25] was evaluated. In this test the null hypothesis is ''the coefficient is zero'', and the alternative hypothesis is ''the coefficient is different from zero''. Since a significance level α, of 0.05 has been chosen, those coefficients whose P value is below 0.05 will be significantly different from zero and must be controlled in the analytical procedure. According to the results presented in Table 3, one would admit the existence of four factors, percentage of acetonitrile (b_1), pH (b_2), the mobile phase flow (b_4) and the wavelength (b_6), influential in the benzoic acid signal, while the method is more robust for the sorbic acid.

However, one must take into account that the

Table 3
Estimated effects for each factor

Factor	Benzoic acid		Sorbic acid	
	Coefficient	P value	Coefficient	P value
b_1	−3.33	<0.01*	−0.70	0.51
b_2	3.59	<0.01*	0.28	0.78
b_3	−0.05	0.80	1.17	0.30
b_4	3.82	<0.01*	0.70	0.51
b_5	0.20	0.33	1.67	0.17
b_6	−1.68	<0.01*	−0.89	0.41
b_7	0.23	0.28	−1.81	0.15

The asterisk (*) indicates that the factor is significant at a significance level of 0.05. See Table 1 for the codification of the factors.

experimental variance to apply the test was determined with only three values (experiments 9, 10 and 11, replicates of the nominal level), so it is likely that the variance will be underestimated and therefore significant factors could be admitted when they are not so. The opposite could happen to the sorbic acid model; an overestimation of the experimental variance as a result of anomalous data in experiments 9, 10 and 11, would lead to consider non-significant effects when in fact they do affect the signal. It is clear that with only three values one cannot guarantee either the absence of underestimation or of anomalous data. That is why the results will be compared with those obtained using other methods (Lenth's method and normal probability plot) which do not need the variance of the replicates to estimate the effects.

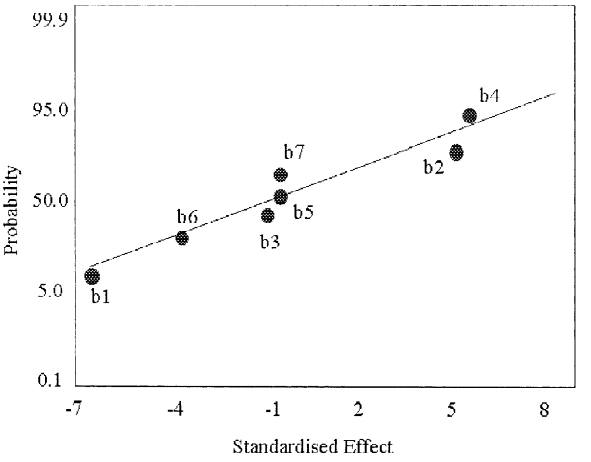

Fig. 2. Normal probability plot of the coefficients of the benzoic acid. See Table 1 for the codification of the factors.

3.1.1. Lenth's method

The first alternative to evaluate the significance of the coefficients from unreplicated designs is Lenth's method [26]. Instead of estimating the experimental variance using experiments 9, 10 and 11, a pseudo-variance is determined by multiplying the median of the coefficients in absolute value by 1.5. Then the significance limit is set at 2.5 times the pseudo-variance. If any coefficient exceeds this significance limit, it is eliminated and the process is repeated with the remaining ones. The advantage of Lenth's method is the use of the median which is a robust estimator and therefore will contribute results independently of the variance estimated by the replicates. This methodology has already been used successfully in the field of electrochemistry [27].

Thus, the significance limit for the benzoic acid is 6.44, and that for sorbic acid is 3.59; because no coefficient (Table 3) is greater than the corresponding limit value it must be admitted that none of them is significant and therefore the procedure is robust.

3.1.2. Normal probability plot

Another option for the analysis of the significance of the coefficients is the normal probability plot [22]. This is based on the fact that if the variation in the data is due solely to a random variation and the changes in level of the factors have no effect on the response, then the coefficients will have a normal distribution such that in a plot on normal probabilistic paper they will be seen to be aligned. On the other hand, the coefficients of the significant factors will not be due to chance, but rather to the change in the level of the factor, such that they will move away from the straight line on which fall the factors that are no significant. Figs. 2 and 3 display the normal probability plots of the coefficients for benzoic and sorbic acids. In the benzoic acid case, the effects of pH (b_2), the mobile phase flow (b_4), acetonitrile (b_1) and the wavelength (b_6), which were initially considered significant, are aligned with the rest, which

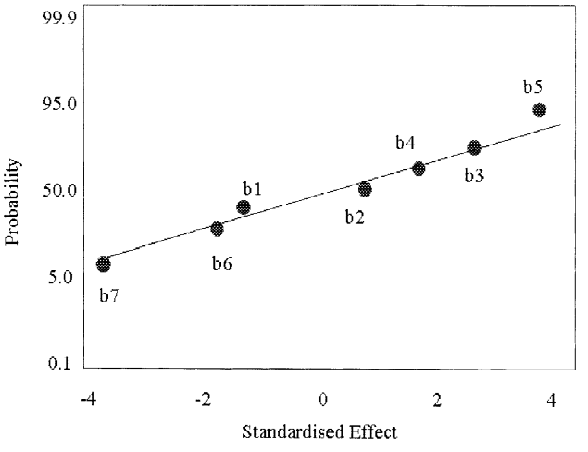

Fig. 3. Normal probability plot of the coefficients of the sorbic acid. See Table 1 for the codification of the factors.

ratifies the conclusion extracted from Lenth's method, no factor is significant so the method is robust. Likewise in the plot corresponding to the sorbic acid (Fig. 3) it can be seen that all the coefficients are aligned and so are not significant.

In this way, through the application of diverse tools unrelated with the classic estimation of the experimental variance, one can conclude that, in the levels investigated, there is no factor which is influential in the quantitative determination of both analytes and therefore that the procedure is robust. This conclusion assures the accuracy of the method against small variations in the experimental conditions which usually exist.

3.2. The observed signal: linear range and calibration function

Calibration is one of the most important steps in chemical analysis. The calibration curve does not need to be linear in the range of application for the method to be efficient, but it is simpler to construct, evaluate and control [25].

The linear model which relates analyte concentration (x) with the signal provided by the analytical instrument (y) is:

$$y = A + Bx + \varepsilon \tag{1}$$

where A is the intercept, B the slope, that is, the parameters of the true but unknown model and ε is the random error which models the uncertainty in the experimental determination (its mean is assumed to be zero). The least-squares (LS) method, which minimises the sum of the squares residuals, Eq. (2), was used for calculating the estimation of the true but unknown parameters, a and b. Supposing that the errors are with normal distribution, with constant variance, and that they are independent of each other and of the concentration level, the estimation by least squares provides the most likely parameters and the estimations are the most precise of the accurate ones.

$$\min_{a,b} \sum_{i=1}^{n} (y_i - (a + bx_i))^2 \tag{2}$$

Despite the excellent features of the least-squares method, it has the inconvenience of being very sensitive to the presence of outliers. Consequently it is not a useful tool for determining the linear range of a calibration curve because any curvature in the response behaves in the same way as outliers so the LS criterion will tend to mask the curvature. Outliers cause lack of normality in the probability distributions which produces the loss of the statistical properties of the LS regression. Besides their presence affects not only the evaluation of the slope and the intercept, as can be deduced from Eq. (2), but also the precision the estimations are determined with. As the presence of outliers highly increases the residual standard deviation, s_{yx} (Eq. (3)), and the confidence intervals depend on the correct assessment of s_{yx}, the subsequent estimations of the slope (Eq. (4)) and the intercept (Eq. (5)) will have large confidence intervals. As a result the parameters will be less precise.

$$s_{yx}^2 = \frac{\sum_{i=1}^{n} (y_i - (a + bx_i))^2}{n - 2} \tag{3}$$

$$s_b^2 = \frac{s_{yx}^2}{\sum_i (x_i - \bar{x})^2} \tag{4}$$

$$s_a^2 = s_{yx}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2} \right) \tag{5}$$

For this reason, a robust regression technique, the least median of squares (LMS) regression, was employed for the determination of those points which are outside the linearity. LMS regression minimises the Eq. (6) which corresponds to the median of the squared residues and has the theoretical property of exact fit, supporting 50% of the anomalous data both on the abscissa axis (leverage data) and on the ordinate axis (outlier data). The theoretical aspects of this regression can be consulted in Ref. [28].

$$\min_{a,b} (\text{median}_i \{(y_i - (a + bx_i))^2\}) \tag{6}$$

The properties of the LMS regression have already been used to determine the linear range of a calibration [29] because, not being affected by outliers, it suffices to examine the residuals to decide which are big and correspond to outliers, that is, outside the linearity.

However the LMS regression has an incon-

venience: it does not make any hypothesis as to the probability of distribution of the error ε of the model (Eq. (1)). Therefore the confidence interval of the parameters cannot be statistically evaluated.

In this paper the subsequent protocol for validating the linearity has been followed: (i) acquisition of the experimental data (concentration, peak area), (ii) perform the LMS regression, (iii) evaluation of the standardised residuals (SR) and rejection of those data whose SR exceeds 2.5 (thus the linear range can be stated), (iv) execution of the LS regression with the aligned experimental points, (v) evaluation of the parameters of the LS regression (slope, intercept and residual standard deviation), (vi) validation of the regression through the corresponding hypothesis tests. The whole process is named reweighted least squares (RLS). This methodology [28] has been implemented in the PROGRESS program (available free from http://win-www.uia.ac.be/u/statis).

To evaluate the linear range the area of the chromatographic peak of 10 standards were measured. The concentrations of the standards varies between 5 and 220 mg l^{-1} of benzoic acid and 0.2 and 30 mg l^{-1} of sorbic acid. Table 4 shows the analytical signal (area of the chromatographic peak, in arbitrary units) obtained for each standard and their standardised residual (SR) with respect to the LMS regression. Data marked with an asterisk (*), are those considered to be non aligned because their standardised residual, in absolute value, is greater than 2.5 so the linear range is 5–145 mg l^{-1} for benzoic acid and 0.2–4 mg l^{-1} for sorbic acid.

After establishing the linear range, we evaluated the parameters (and their confidence intervals) of the calibration model which relates the peak area with the concentration. A second calibration was carried out for each analyte, now with the standards, in concentrations included within the linear range already estimated. Applying the RLS methodology, one finds that the model of benzoic acid has a slope (sensitivity) of 8096 ± 35 au l mg^{-1}, an intercept of 2798 ± 2706 au and the residual standard deviation s_{yx} is 3632 au. That corresponding to sorbic acid has a slope (sensitivity) of $318\,833 \pm 888$ au l mg^{-1}, an intercept of 1216 ± 1226 au and the residual standard deviation is 1314 au. In both cases the confidence intervals were calculated setting the significance level α at 0.05. Of these values, it is noteworthy that the procedure has more sensitivity to the sorbic acid than to the benzoic acid, approximately 40 times higher, due to the area of the spectrum where the measurements are collected. This difference in sensitivities will have consequences in the evaluation of other figures of merit such as the capability of detection.

Once the parameters were estimated, models must be validated to guarantee that they are linear. Thus, it is necessary to verify the assumptions related not only to the function but also to the residuals in order to assure that the model selected is the correct one.

The functional part was validated by means of the following tests[1] [25]: (i) the *lack-of-fit test* (H_0 ''the bias is zero'', H_1 ''the bias is positive''), (ii) the *significance test* (H_0 ''the regression cannot explain the experimental variation'', H_1 ''the regression does explain the experimental variation''). As for the residuals, they must be *normally distributed* (Kolmogorov–Smirnov test and χ^2 test, H_0 ''the data are normally distributed'', H_1 ''they are not so''), *independent* (Durbin–Watson test, H_0 ''the residuals are independent'', H_1 ''they are dependent'') and *homoscedastic* (Bartlett's test and Cochran's test, H_0 ''the variances are not significantly different'', H_1 ''at least one of the variances is different''). The hypothesis tests were performed with STATGRAPHICS [30] obtaining the *P* values shown in Table 5. According to the results it can be concluded that fixing α at 0.05, both models explain the experimental variability observed and they do not have lack of fit. The residuals are independent, homoscedastic and normally distributed because there is not evidence to reject the null hypothesis. Consequently the model will be adequate to describe the linear relationship between the area and the concentration.

3.3. Analysis of the selectivity/specificity

Although the experimental procedure is based on HPLC where the composition of the mobile phase, the column, etc., can be modified so that the presence of other components does not influence the results, the selectivity is an essential condition to be

[1] H_0 refers to the null hypothesis and H_1 to the alternative hypothesis.

Table 4
Analytical signals (au), concentration (mg l^{-1}) of benzoic and sorbic acids of the standards and standardised residual (SR) with respect to the LMS regression to estimate the linear range

Benzoic acid			Sorbic acid		
Conc. (mg l^{-1})	Peak area (au)	SR	Conc. (mg l^{-1})	Peak area (au)	SR
4.99	45 565	−0.45	0.20	67 383	−0.62
4.99	47 231	−0.08	0.20	69 748	−0.07
4.99	50 858	0.73	0.20	70 154	0.02
19.96	165 158	0.01	0.80	263 962	0.06
19.96	175 533	2.31	0.80	268 102	1.01
19.96	168 207	0.69	0.80	264 904	0.27
44.91	363 597	0.58	1.39	453 627	−0.86
44.91	364 086	0.69	1.39	459 345	0.45
44.91	363 984	0.66	1.39	460 883	0.81
69.86	556 509	−0.08	1.99	652 010	0.22
69.86	561 494	1.03	1.99	647 987	−0.70
69.86	554 700	−0.48	1.99	652 334	0.30
94.81	749 940	−0.62	2.58	840 327	−1.01
94.81	754 805	0.46	2.58	845 060	0.08
94.81	759 861	1.58	2.58	846 563	0.43
119.76	956 416	1.73	3.18	1 036 280	−0.48
119.76	949 410	0.18	3.18	1 018 701	−4.53*
119.76	955 174	1.45	3.18	1 029 936	−1.94
144.71	1 142 797	−0.37	3.97	1 305 475	2.17
144.71	1 148 970	0.99	3.97	1 291 692	−1.01
144.71	1 139 848	−1.03	3.97	1 305 141	2.09
144.71	1 197 546	11.74*	3.97	1 296 164	0.02
169.66	1 336 938	−0.76	15.89	4 985 151	−42.06*
169.66	1 396 939	12.51*	15.89	4 887 694	−64.51*
169.66	1 382 401	9.30*	15.89	4 924 043	−56.14*
169.66	1 394 538	11.98*	15.89	4 908 935	−59.62*
195.61	1 595 697	11.41*	23.84	7 121 237	−144.60*
195.61	1 583 363	8.69*	23.84	7 017 279	−168.55*
195.61	1 578 803	7.68*	23.84	7 107 658	−147.72*
195.61	1 584 788	9.00*	23.84	7 131 721	−142.18*
219.56	1 788 926	12.56*	27.81	8 093 410	−217.95*
219.56	1 781 389	10.90*	27.81	8 047 031	−228.63*
219.56	1 764 314	7.12*	27.81	8 094 770	−217.63*
219.56	1 790 455	12.90*	27.81	8 020 131	−234.83*

With an asterisk (*) the signals whose standardised residual in absolute value is greater than 2.5.

verified with univariate or zero order signals. This parameter is defined in the EURACHEM [9] guide as ‘‘the ability of a method to determine accurately the analyte of interest in the presence of other components in a sample matrix under the stated conditions of the test’’.

An efficient way to study the selectivity/specificity is by means of the standard addition method to compare the slopes of the standard addition line and the aqueous calibration line. If there are not matrix interferences, both lines will have the same slope.

To perform the standard addition method, to eight aliquots of a soft drink sample which contains possible interferences, different volumes of standards containing benzoic and sorbic acids were added such that the final concentration added is between 2 and 16 mg l^{-1} of benzoic acid (increasing amounts of 2 mg l^{-1}) and 0.2 and 1.6 mg l^{-1} of sorbic acid (increasing amounts of 0.2 mg l^{-1}). The calibration in aqueous medium was built with standards in concentrations within the linear range established previously. The parameters of both regressions were

Table 5
P values of the hypothesis tests which have been carried out to validate the linearity of the RLS models of benzoic and sorbic acids

Test	*P* value	
	Benzoic acid	Sorbic acid
Function		
Lack of fit	0.24	0.45
Significance	0.00	0.00
Residuals		
Cochran's	0.21	0.53
Bartlett's	0.64	0.54
Durbin–Watson[a]	1.55	1.98
χ^2	0.31	0.52
Kolmogorov–Smirnov	>0.10	>0.10

[a] Critical value 1.3.

determined following the robust methodology (RLS models) described in Section 3.2 to eliminate the outliers, and are shown in Table 6, where $s_{yx}{}^2$ is the residual variance (Eq. (3)), n the number of data used to build the calibration curve, b the sensitivity of the model and $s_b{}^2$ is the standard deviation of the slope (Eq. 4).

The sensitivities of both regressions were compared performing the hypothesis test for the comparison of the slopes of two regression lines (null hypothesis ''the slopes are equal'', alternative hypothesis ''the slopes are different'') [25]. The statistic of the hypothesis test is calculated through the Eq. (7):

Table 6
Parameters of the RLS models in aqueous and in real media

Media	Parameter	Benzoic acid	Sorbic acid
Water			
	$s_{yx}{}^2$	7.59×10^5	1.23×10^6
	n	13	20
	b	8165.51	312 734.10
	$s_b{}^2$	35.64	9.86×10^4
Matrix			
	$s_{yx}{}^2$	7.71×10^5	6.83×10^6
	n	25	25
	b	8216.39	313 781.30
	$s_b{}^2$	910.23	8.24×10^5

s_{yx}^2, residual variance; n, number of data to built the calibration curve; b, slope; $s_b{}^2$ variance associated with the slope.

$$t_{\text{cal}} = \frac{|b_1 - b_2|}{\left(s_{b1}^2 + s_{b2}^2\right)^{1/2}} \tag{7}$$

If the residual variances of both calibration curves are equal (comparison can be performed by means of an F-test) $s_{yx1}^2 = s_{yx2}^2$, the statistic t_{cal} is calculated through Eq. (8) and then compared with the tabulated t value with $n_1 + n_2 - 4$ degrees of freedom at the chosen significance level.

$$t_{\text{cal}} = \frac{|b_1 - b_2|}{\left[\dfrac{(n_1-2)s_{yx1}^2 + (n_2-2)s_{yx2}^2}{n_1+n_2-4}\left(\dfrac{1}{\sum(x_{i1}-\bar{x}_1)^2} + \dfrac{1}{\sum(x_{i2}-\bar{x}_2)^2}\right)\right]^{\frac{1}{2}}} \tag{8}$$

On the other hand, if the residual variances of the calibrations in both media are not equal $s_{yx1}^2 \neq s_{yx2}^2$ then the statistic obtained from Eq. (7), t_{cal}, is compared with a Student's t-distribution: $t' = \left(t_1 s_{b1}^2 + t_2 s_{b2}^2\right)/\left(s_{b1}^2 + s_{b2}^2\right)$. t_1 and t_2 are the theoretical t values at the chosen level of significance with $n_1 - 2$ and $n_2 - 2$ degrees of freedom, respectively.

The statistics calculated, 1.66 (Eq. (8)) and 1.09 (Eq. (7)) for benzoic and sorbic acids, respectively, are smaller than the tabulated ones (2.03 and 2.07, respectively) when the significance level is set to 0.05. It can be concluded that the sensitivities of the aqueous calibration line and the standard addition line are significantly equal which indicates that there is no matrix effect. In other words the method is selective and the calibration can be carried out in aqueous media.

In this section another basic application of the LMS regression has been given. Outliers increases the residual standard deviation s_{yx}, so the subsequent estimations of the parameters would have higher confidence intervals than they should. As a result, the hypothesis tests would tend to be more conservative such that the null hypothesis (the slopes are equal) tends to be accepted when it should not be. Consequently the model would be affirmed to be selective when it is not and there are matrix interferences.

3.4. The two components of the accuracy according to the ISO standard 5725

One requirement to replace one analytical method

with another is that the new one must be accurate. According to ISO 5725-1 [2] accuracy is ''the closeness of agreement between a test result and the accepted reference value''. The term accuracy, when applied to a set of test results, involves a combination of random components (precision) and a systematic error or bias component (trueness). ISO 5725 uses these two terms, trueness and precision, to describe the accuracy of a measurement method. ''Trueness refers to the closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value'' [3]. Precision [4,5] refers to the closeness of agreement between test results. The measurement of precision is usually computed as a standard deviation of the test results, less precision is reflected by a larger standard deviation, whereas trueness has been expressed in terms of bias.

3.4.1. Analysis of the trueness/recovery

Although the method has already been concluded to be selective, the lack of trueness may occur not only when there are analytical interferences, but also when the analytical procedure is very laborious, with various stages, which results in a reduction in recovery. The analysis of the trueness enables to establish the recovery rate, to later apply a correction factor to the measurements obtained.

The standard addition method described in Section 3.3 has been used to compare the difference between the results obtained in the sample with addition and the sample with no addition (''sample + addition'' − ''sample''), with the theoretical value of the addition which is considered as a reference value. If there were no bias, adjusting by least-squares (Eq. (1)), the concentration recovered against the concentration added, one would obtain a regression whose slope is 1 and whose intercept is 0. The deviation of the latter of zero is a measure of the constant error while the deviation of the slope of one is indicative of a proportional systematic error. The random error can be estimated from the standard deviation of the regression s_{yx}. Again, the LMS regression has been applied to avoid the influence of outliers on the estimations.

For the analysis of the trueness/recovery we made use of the experimental data described in Section 3.3. Thus, in the case of benzoic acid, the parameters of the regression were, slope 1.001 ± 0.011, intercept -0.104 ± 0.123 mg l^{-1} and residual standard deviation 0.112 mg l^{-1}; for the sorbic acid the slope was 1.002 ± 0.008, the intercept 0.002 ± 0.009 mg l^{-1} and the residual standard deviation 0.009 mg l^{-1}. The confidence intervals were calculated setting the significance level at 0.05. In both cases it follows that the slope and the intercept are significantly equal to 1 and 0, respectively, then the proportional and constant errors are less than the experimental error itself. In this way, it must be admitted that the method is unbiased and therefore it will not be necessary to apply a correction factor to the results obtained. Finally, the trueness has also been established in terms of recovery, reaching an average recovery for benzoic acid of 101.9% with a standard deviation of 1.3%. For the sorbic acid the average recovery was 100% and the standard deviation 1.3%. In both cases these values were calculated from the recovery obtained at seven different concentration levels.

3.4.2. Analysis of the precision

The second component of the accuracy is the precision which, in the ISO standard 5725 [2], has been defined as ''the closeness of agreement between independent test results obtained under stipulated conditions''. Two different terms [4] are used to define the precision, repeatability and reproducibility are two extremes which describe the minimum and the maximum variability found in the results. Repeatability refers to the precision under repeatability conditions (conditions where independent test results are obtained with the same method, on identical test items, in the same laboratory, by the same operator, using the same equipment, within short intervals of time). Reproducibility is the precision under reproducibility conditions (conditions where test results are obtained with the same method, on identical test items, in different laboratories, with different operators, using different equipment). The evaluation of the precision is based on the analysis of variance (ANOVA) which is a suitable technique for estimating the repeatability and the reproducibility permitting one to take relevant decisions in almost any experimental problem. Since in this paper a single factor (change of day) has been investigated, we will

use the term intermediate precision [5] instead of reproducibility.

To evaluate the precision, the concentration of benzoic and sorbic acids was calculated during two different days and carrying out five replicates each day. The concentrations determined are listed in Table 7. In the terminology of ANOVA there is one factor, different days, with two values ($P=2$) and five replicates ($n=5$) in each of them. In this way it is possible to evaluate the repeatability as the variance of the measurements on a single day (within-day variance),

$$s_0^2 = MS_{\text{error}} = s_r^2 \quad (9)$$

Likewise, one can determine the intermediate precision s_R (Eq. 10) from the variance of the repeatability, s_r^2, and the variance due to change in the factor, $s_L^2 = (MS_{\text{factor}} - MS_{\text{error}})/n$, that is, due to the variability shown by the measurements when changing day.

$$s_R = \sqrt{s_L^2 + s_r^2} \quad (10)$$

The estimations of repeatability, s_r, Eq. (9) and intermediate precision, s_R, Eq. (10) are shown in the second and third columns of Table 8 for each analyte. It was found that the precision values corresponding to the sorbic acid are worse, almost five times higher than those of the benzoic acid. Hence it is essential to verify the absence of anomalous data. ISO proposes firstly the Cochran's test (Section 3.2), which refers to the between-day variability. Then the Grubbs' test (null hypothesis ''the value is outlier'', alternative hypothesis ''this is not the case''), which is basically a test of the within-day variability, to discover possible individual anomalous data. The statistics obtained by applying both tests are given in Table 7 where C is the statistic of the Cochran's test, G_1 indicates that the Grubbs' test has been done with the smallest data, G_5 with the largest, $G_{1,2}$ with the smallest two and $G_{4,5}$ with the largest two. If the test statistic is greater than its 5% critical value and less than or equal to its 1% critical value, the item tested is called a straggler and is indicated by a single asterisk (*). If the statistic is greater than its 1% critical value, the item is called a statistical outlier and is indicated by a double asterisk (**). For testing two outlying observations, outliers and stragglers give rise to values which are smaller than the tabulated 1 and 5% critical values, respectively. Only the outliers are eliminated from the calculation of the precision.

The critical value of the Cochran's test for a data set with two levels and five replicates per level is $C_{2,5}=0.906$ (with the significance level set at 5%) and 0.959 at 1%. As can be seen in Table 7, only in the case of the sorbic acid is it necessary to reject the null hypothesis of the Cochran's test (the variances are equal) and conclude that the variance is significantly different from the rest at this significance level. This inequality may be due to an outlier which excessively raises the variance of the level. For this reason the Grubbs' test which detects the existence of very different data was performed. The critical value of the statistic with five replicates per level at

Table 7
Concentration (mg l^{-1}) of benzoic and sorbic acids determined in a soft drink sample during two different days and performing five replicates per day

Replicates	Benzoic acid		Sorbic acid	
	Day 1	Day 2	Day 1	Day 2
r_1	106.49	105.41	164.70	164.59
r_2	105.55	106.75	165.23	164.54
r_3	104.28	106.66	165.78	164.35
r_4	105.75	107.06	164.58	164.63
r_5	106.21	106.08	176.12**	164.49
C	0.556	–	0.976	–
G_5	0.977	1.022	1.781	1.013
G_1	1.612	1.503	0.544	1.565
G_{45}	0.436	0.458	0.002	0.411
G_{12}	0.096	0.052	0.760	0.086

*Straggler according to ISO norm; **an outlier.
C, the statistic calculated in the Cochran's test; G_5 and G_1, the statistics of the Grubbs' test applied to one data; and G_{45} and G_{12}, the statistics of the Grubbs' test applied to two data.

Table 8
Repeatability, s_r, and intermediate precision, s_R, in mg l^{-1}, calculated according to the ISO standard 5725 and to the H15 estimator

Analyte	With all data		ISO 5725		H15	
	s_r	s_R	s_r	s_R	s_r	s_R
Benzoic acid	0.76	0.86	0.76	0.86	0.75	0.85
Sorbic acid	3.51	3.70	0.37	0.49	0.49	0.72

5% is $G_{0.05,5}=1.715$ and at 1% is $G_{0.01,5}=1.764$. In the case of applying the test for the detection of two outliers, the critical value at 5% is $G_{0.05,5}=0.009$ and at 1% $G_{0.01,5}=0.0018$. These values verify the existence of an outlier in the data set of sorbic acid (displayed by a double asterisk in Table 7). By eliminating this and recalculating the precision one obtains the repeatability and intermediate precision shown in the fourth and fifth columns of Table 8.

However, eliminating the data is not advisable when the aim is to evaluate the variability of the analytical procedure since the real variance reachable in practice is underestimated. An alternative suggested in the ISO standard 5725-5, consists of using robust estimators. The H15 Huber-estimator ($c=1.5$ and ''Proposal 2 Scale'' [31]) is recommended by the Analytical Methods Committee [32] and accepted in the Harmonised Protocol [33]. The technical aspects can be consulted in Refs. [31,34]. It is characterised by being the prototype of estimator with monotone influence function which limits the influence of anomalous data ''moving'' them towards the position of the majority, but maintaining the maximum influence for them. This is done by transforming the original data through the function

$$\Psi_{m\cdot s\cdot c}(x) = \max[m - cs, \min(m + cs, x)] \quad (11)$$

where m and s are the centralization and dispersion parameters, respectively.

The estimator is asymptotically optimum for high quality data not very different from a Student's t-distribution with three degrees of freedom. In addition it gives fairly good protection against high concentrations of data with abnormally large errors. However, the Huber estimator does not reject the outliers but rather maintains their maximum influence although limited.

The robust procedure obtained by adapting the H15 estimator to the problem of the evaluation of the precision consists of two stages and was done exactly as proposed in Ref. [35].

The results were achieved using the INTERLAB program [36] which gives not only the mean, the standard deviation and the precision of a measurements set but also the robust centralization and dispersion parameters calculated by means of the H15 Huber-estimator. The repeatability and intermediate precision estimated in this manner are reported in Table 8. Since one cannot guarantee the absence of anomalous data in any experimental result, this robust procedure avoids the underestimation of the precision of an analytical method. It constitutes a good alternative to the classical or parametric study of data, especially because it is very simple to apply as compared with the scheme of data elimination using the Cochran's test and Grubbs' test. The authors [35] have already applied this procedure for the determination of the precision of a method for the analysis of benzaldehyde by differential pulse polarography (DPP) at various levels of concentration.

3.5. Analysis of the capability of detection

A fundamental aspect which needs to be examined in the validation of any analytical method is its limit of detection, which indicates if an analyte is present or not in the sample. ISO 11843-1 [8] has named this figure of merit the capability of detection or minimum detectable net concentration, x_d, and it has been defined as ''the true net concentration of the analyte in the material to be analysed which leads, with probability $(1-\beta)$, to the correct conclusion that the concentration in the analysed material is different from that in the blank material for a given probability of false-positive, α''. This definition highlights the need to establish not only the probability of false-positive, α (to affirm that the analyte is present when it is not) or probability of type I error, but also β, probability of false-negative (to affirm that the analyte is not present when it is) or probability of type II error. Hence the importance of the evaluation of the probabilities of false-positive and false-negative is made completely clear. Ref. [37] deals in detail with this question. The capability of detection (Eq. (12)) is determined by posing the following unilateral hypothesis test with relation to the presence or absence of analyte in the sample with evaluation of the probabilities of false-positive and false-negative:

Null hypothesis: The analyte is not present

Alternative hypothesis: The analyte is present

$$x_d = \Delta(\alpha,\beta) w_0 \frac{s_{yx}}{b} \quad (12)$$

where $\Delta(\alpha,\beta)$ is the parameter of noncentrality and

w_0 the variance of the estimated concentration at a concentration x of analyte. The procedure is definitively accepted by ISO 11843-2 [8] and IUPAC [16].

From Eq. (12) it follows that to determine the capability of detection it is necessary to take into account the calibration curve which transforms the detection signal into concentration. In an initial stage a model was built for each analyte with eight standards ($I=8$) with concentrations between 0.05 and 0.55 mg l^{-1} of benzoic acid and 0.004–0.035 mg l^{-1} of sorbic acid. The parameters of the models, following the robust RLS methodology, are: slope 7655 ± 244 au l mg^{-1}, intercept 232 ± 70 au and residual standard deviation 79 au in the case of benzoic acid while those corresponding to the sorbic acid model are: slope $307\,884\pm5590$ au l mg^{-1}, intercept 222 ± 87 au and residual standard deviation 88 au. It can be observed that the calibrations are less sensible than those found in Section 3.2 and the residual standard deviation has diminished as a result of working at lower concentration levels [38]. Given that the capability of detection has to be estimated from the parameters of the calibration slope, they must be optimally evaluated. The presence of outliers causes these estimations to be incorrect, and it is normal that at such low concentration levels the procedures are less sensible and anomalous responses appear. To avoid these effects the LMS regression is used.

Once the linear relationship signal-concentration was established, the capability of detection was estimated using the DETARCHI program [39] (available free from the authors), which determines the characteristic curves of detection as a function of the number of replicates for a given probability of false-positive, α. Thus, performing one replicate, it is possible to detect 0.05 mg l^{-1} of benzoic acid and 0.004 mg l^{-1} of sorbic acid with a probability of false-positive (α) equal to 0.05 and a probability of false-negative (β) less than 0.05. The capability of detection achieved for the sorbic acid is below that for the benzoic acid as a result of the greater sensitivity of the model to the first (Section 3.2) and because the residual deviations s_{yx}, remain close to one another. From the results it has to be accepted that the procedure is adequate for detecting the presence of the analytes in the concentration levels permitted (150 and 300 mg l^{-1} of benzoic and sorbic acids, respectively).

3.6. Analysis of the capability of discrimination

Once it has been verified that the model is appropriate for detecting the existence of the corresponding analytes, the following question arises: it is habitual to find up to 130 mg l^{-1} of benzoic acid and 220 mg l^{-1} of sorbic acid in soft drinks, will the model be capable of discriminating the same quantity of analyte at these concentration levels, which are much higher than the capability of detection? To answer this question one needs to analyse a new figure of merit, the capability of discrimination, defined by Sanz et al. [40] as an extension of the concept of the capability of detection for any nominal concentration, x_0. Given a nominal concentration, to know the behaviour of an analytical procedure in samples with similar concentration, the capability of discrimination or minimum discriminable concentration, d, is defined as ''the smallest concentration of analyte that can be distinguished from the nominal concentration with a probability fixed at $1-\beta$ and for a given probability of false non compliance, α''. The definition generalises that of the limit of detection in the sense that the capability of detection is applied when the nominal concentration is zero and the hypothesis test is unilateral. The capability of discrimination on the other hand is applied to any nominal concentration and thus the statistical test will be bilateral.

The need to study the capability of discrimination lies in the fact that in analytical procedures the variance increases with concentration [38]. Therefore, given a technique with a well-established capability of detection, it may not be possible to discriminate the same quantity when it is used in samples with a much higher concentration. The methodology proposed is an adaptation of the Clayton method [8,41] to the bilateral case.

Let x_0 be the nominal concentration for which one will study the minimum discriminable concentration or capability of discrimination. This is determined by testing the hypothesis of the statistical test and as a result one can establish the compliance of the analytical procedure if $X=x_0$ and the non-com-

pliance when $X \neq x_0$. The evaluation is done by means of the following Neyman–Pearson test:

Null hypothesis: the true concentration of the sample is x_0, $X = x_0$;

Alternative hypothesis: the concentration of the sample is not x_0, $X \neq x_0$.

As in the capability of detection the probability of type I error, α = prob{false non-compliance} and the probability of type II error, β = prob{false compliance} are specified. Once the hypothesis test has been established one can determine the capability of discrimination, $d = |x - x_0|$, of an analytical procedure at a nominal concentration x_0 from Eq. (13).

$$d = |x - x_0| = \Delta_x w_{x_0} \frac{s_{yx}}{b} \quad (13)$$

Fig. 4 displays the curves of false compliance, β, for a nominal concentration, x_0, of 30 mg l^{-1} of benzoic acid, as a function of the capability of discrimination, $d = |x - x_0| = |x - 30|$, and the number of replicates, K. The probability of false non-compliance was set at 0.05. These curves are equivalent to the characteristic curves of the capability of detection such that for fixed β, the capability of discrimination d, improves as the number of replicates increases.

Table 9
Capability of discrimination (mg l^{-1}), $d = |x - x_0|$, as a function of the nominal concentration (x_0), of the number of replicates (K), and of the standard residual deviation, s_{yx} (mg l^{-1})

Analyte	x_0 (mg l^{-1})	Replicates			s_{yx}
		$K=1$	$K=3$	$K=5$	
Benzoic acid	30	0.44	0.28	0.24	0.10
	42	0.57	0.34	0.28	0.14
	47	0.62	0.37	0.30	0.16
	70	1.21	0.76	0.64	0.28
Sorbic acid	0.8	0.023	0.014	0.012	0.006
	1.0	0.020	0.013	0.011	0.004
	1.6	0.031	0.019	0.015	0.008
	1.7	0.024	0.015	0.013	0.005

Table 9 provides the values of the capability of discrimination, $d = |x - x_0|$, in mg l^{-1}, estimated at various levels of concentration of benzoic and sorbic acids. The minimum discriminable concentration for the benzoic acid is found between 0.44 and 1.21 mg l^{-1} when the nominal concentration increases from 30 to 70 mg l^{-1} and for the sorbic acid between

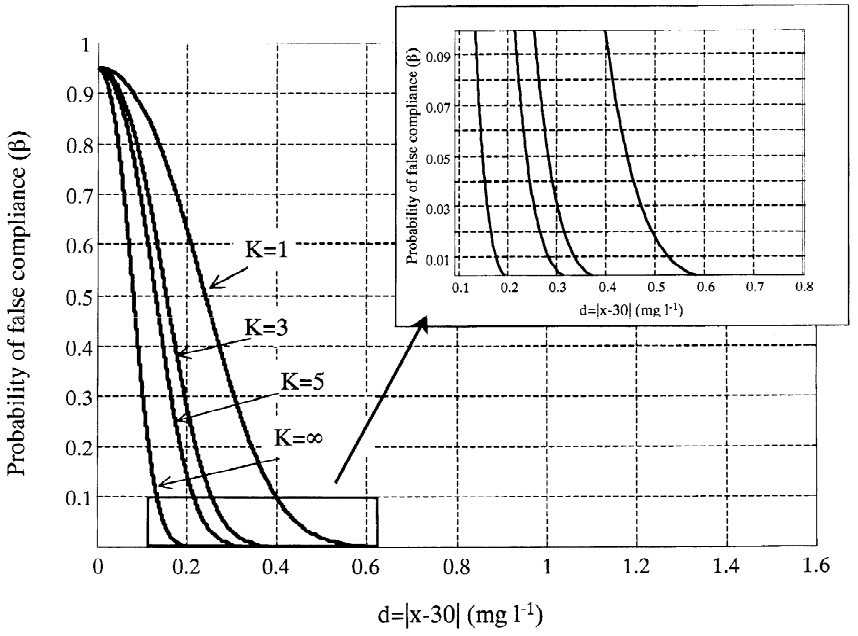

Fig. 4. Capability of discrimination curves for benzoic acid (x_0 = 30 mg l^{-1}) as a function of the number of replicates, K. The probability of false non-compliance α, was fixed at 0.05.

0.020 and 0.031 mg l^{-1} when its nominal concentration goes from 0.8 to 1.7 mg l^{-1}. These values are at least six times greater than the minimum detectable net concentration estimated in Section 3.5 (0.05 and 0.004 mg l^{-1} of benzoic and sorbic acid, respectively). This is due to the increase in the variance of the analytical procedure, last column in Table 9, when the concentration increases.

From the joint study of Table 9 and Fig. 4 it is apparent that the capability of discrimination, just as the limit of detection, improves when the number of replicates K increases because they are inversely proportional. Therefore a way of improving the capability of discrimination is to give as a measure the mean of K experimental measurements. What is more, for the benzoic acid, maintaining α, β and K constant, when the nominal concentration x_0 increases, the capability of discrimination worsens, that is, the minimum discriminable concentration increases. This is due to the significant increase in the residual standard deviation s_{yx} in samples with greater concentration. With regard to the data for the sorbic acid, there is no clear tendency in the capability of discrimination because the variances do not increase significantly with the concentration rise investigated.

4. Conclusions

A method of HPLC has been developed and validated for the joint quantitative determination of benzoic and sorbic acids in soft drinks, reducing by a factor of four the analysis time with respect to the AOAC method. The use of robust statistical tools (LMS regression, H15 estimator and Lenth's method) during the validation process (all the parameters validated are summed up in Table 10) has proved to be efficient in the determination of the robustness, the linear range, the selectivity, trueness/recovery, the precision (repeatability and intermediate precision), the capability of detection and the capability of discrimination. The capability of detection has been established setting the values of the probability of false-positive at 0.05 and false-negative less than 0.05. In addition, the need to extend the concept of capability of detection for any nominal concentration has been emphasised, since in many procedures of chemical analysis it is necessary to measure samples with concentrations far above the limit of detection and one cannot discriminate the same quantity of analyte. The figure of merit which considers this is the capability of discrimination which quantifies the probability of false compliance and the probability of false non-compliance.

Acknowledgements

This work has been partially supported by the Ministerio de Ciencia y Tecnología (DGI, Project BQU2000-0863) and the Consejería de Educación y Cultura de la Junta de Castilla y León (Project BU15/01). I.G. also wishes to thank the Ministerio de Educación Cultura y Deporte for the FPU grant.

Table 10
Summary of the parameters validated for the determination of benzoic and sorbic acids under the experimental conditions described in Section 2

Figure of merit	Benzoic acid	Sorbic acid
Robustness	Yes	Yes
Linear range (mg l^{-1})	5–145	0.2–4
Selectivity/Specificity	Yes	Yes
Recovery (trueness) (%)	101.9	100
Repeatability (mg l^{-1})	0.75	0.49
Intermediate precision (mg l^{-1})	0.85	0.72
Capability of detection (mg l^{-1}) ($\alpha=0.05$, $\beta<0.05$, $K=1$)	0.05	0.004
Capability of discrimination (mg l^{-1}) ($\alpha=0.05$, $\beta=0.05$, $K=1$)	0.44 ($x_0=30$ mg l^{-1})	0.023 ($x_0=0.8$ mg l^{-1})

α is the probability of type I error, β the probability of type II error, K the number of replicates and x_0 the nominal concentration.

References

[1] International Standard ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories, International Organization for Standardization, Geneva, 1999.
[2] International Standard ISO 5725-1, Accuracy (trueness and precision) of measurement methods and results—General Principles and definitions, International Organization for Standardization, Geneva, 1994.
[3] International Standard ISO 5725-4, Accuracy (trueness and precision) of measurement methods and results—Basic methods for the determination of the trueness of a standard measurement method, International Organization for Standardization, Geneva, 1994.
[4] International Standard ISO 5725-2. Accuracy (trueness and precision) of measurement methods and results—Basic methods for the determination of the repeatability and reproducibility of a standard measurement method, International Organization for Standardization, Geneva, 1994.
[5] International Standard ISO 5725-3, Accuracy (trueness and precision) of measurement methods and results—Intermediate measures of the precision of a standard measurement method, International Organization for Standardization, Geneva, 1994.
[6] International Standard ISO 5725-5, Accuracy (trueness and precision) of measurement methods and results—Alternative methods for the determination of the precision of a standard measurement method, International Organization for Standardization, Geneva, 1994.
[7] International Standard ISO 11843-1, Capability of detection—Terms and definitions. International Organization for Standardization, Geneva, Switzerland, 1997.
[8] International Standard ISO 11843-2. Capability of detection—Methodology in the linear calibration case, International Organization for Standardization, Geneva, 2000.
[9] Guide, The Fitness for Purpose of Analytical Methods, A Laboratory Guide to Method Validation and Related Topics, EURACHEM, Teddington, UK, 1998.
[10] A.H. Varman, J.P. Sutherland, in: Beverages—Technology, Chemistry and Microbiology, Chapman and Hall, London, UK, 1994.
[11] Method 983.16, Benzoic Acid and Sorbic Acid in Food, Association of Official Analytical Chemists (AOAC), Gaithersburg, MD, 1995.
[12] L.V. Bui, C. Cooper, J. Assoc. Off. Anal. Chem. 70 (1987) 892.
[13] Q.C. Chen, J. Wang, J. Chromatogr. A 937 (2001) 57.
[14] S.A.V. Tfouni, M.C.F. Toledo, Food Control 13 (2002) 117.
[15] A. Castineira, R.M. Pena, C. Herreo, J. High. Resolut. Chrom. 23 (2000) 647.
[16] J. Inczédy, T. Lengyed, A.M. Ure, A. Gelencsér, A. Hulanicki, in: International Union of Pure and Applied Chemistry (IUPAC), Compendium of Analytical Nomenclature, Blackwell, Oxford, 1998.
[17] H. Günzler, in: Accreditation and Quality Assurance in Analytical Chemistry, Springer, Berlin, 1996.
[18] M.B. Sanz, L.A. Sarabia, A. Herrero, M.C. Ortiz, Talanta 56 (2002) 1039.
[19] L. Cuadros, R. Blanc, A.M. García, J.M. Bosque, Chemom. Intell. Lab. Syst. 41 (1998) 57.
[20] R.L. Plackett, J.P. Burman, Biometrika 33 (1946) 305.
[21] D.C. Montgomery, in: Design and Analysis of Experiments, Wiley, New York, 1997.
[22] G.A. Lewis, D. Mathieu, R. Phan-Tan-Luu, in: Pharmaceutical Experimental Design, Marcel Dekker, New York, 1999.
[23] W.Y. Youden, in: Statistical Techniques for Collaborative Tests, Association of Official Analytical Chemists (AOAC), Washington, 1972.
[24] D. Mathieu, J. Nony, R. Phan-Tan-Luu, L. Prai, NEMROD-W, version 99.1, Marseille, 2000.
[25] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, in: Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997.
[26] R.V. Lenth, Technometrics 31 (1989) 469.
[27] M.B. Sanz, L.A. Sarabia, A. Herrero, M.C. Ortiz, Talanta 56 (2002) 1039.
[28] P.J. Rousseeuw, A.M. Leroy, in: Robust Regression and Outlier Detection, Wiley, New York, 1987.
[29] M.C. Ortiz, J. Arcos, J.V. Juarros, J. López Palacios, L. Sarabia, Anal. Chem. 65 (1993) 678.
[30] STATGRAPHICS Plus for Windows 4.0. Statistical Graphics Corp.
[31] P.J. Huber, in: Robust Statistics, Wiley, New York, 1981.
[32] Analytical Methods Committee, Analyst 114 (1989) 1699.
[33] M. Thompson, R. Wood, J. AOAC Int. 76 (1993) 926.
[34] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, in: Robust Statistics: The Approach Based on Influence Functions, Wiley, Zurich, 1985.
[35] M.B. Sanz, M.C. Ortiz, A. Herrero, L.A. Sarabia, Quim. Anal. 18 (1999) 117.
[36] L.A. Sarabia, M.C. Ortiz, INTERLAB, Burgos, 1995 (available free from the authors).
[37] M.C. Ortiz, L.A. Sarabia, in: R. Cela (Ed.), Avances en Quimiometría Práctica. Universidad de Santiago de Compostela: Servicio de Publicaciones e Intercambio Científico. Santiago de Compostela, 1994, p. 189.
[38] W. Horwitz, Anal. Chim. 67A (1982) 54.
[39] M.C. Ortiz, L.A. Sarabia, Trends Anal. Chem. 13 (1994) 1.
[40] M.B. Sanz, L.A. Sarabia, A. Herrero, M.C. Ortiz, Anal. Chim. Acta 446 (2001) 295.
[41] C.A. Clayton, J.W. Hines, P.D. Elkins, Anal. Chem. 59 (1987) 2506.