

# Test for non-linearity concerning linear calibrated chemical measurements

Lutz Brüggemann · Wolfgang Quapp ·  
Rainer Wennrich

Received: 8 May 2006 / Accepted: 19 August 2006 / Published online: 5 October 2006  
© Springer-Verlag 2006

**Abstract** Linear calibrated chemical measurements are based on the assumption of linearity. Due to influences of matrices at real samples the condition of linearity can be violated. Therefore, a profound examination has to attach much importance on the linearity of calibration. However, different procedures have been applied for this purpose in literature. In order to examine linear calibration for non-linearity, a recently presented procedure is compared with conventional techniques. The associated statistical models and estimations are represented. The applicability of the different procedures is demonstrated in some practical examples, the determination of arsenic in surface water samples taken from different sites in abandoned mining areas in central Germany. Recommendations for using the indicators and tests of non-linearity are given.

**Keywords** Linear calibration · Non-linearity

## Introduction

In the international standard ISO 11843 [1] procedures are presented for the computation of the critical value of the response variable, the critical value of the net state variable, and the minimum detectable value of the net state variable.

The calibration model used is based on the assumption that the calibration function is linear and that the standard deviation is linearly dependent on the net state variable and is given by

$$Y_{ij} = b_0 + b_1 x_i + \varepsilon_{ij}, \quad (1)$$

where  $x_i$ : is the symbol for the net state variable in state  $i$ ;  $\varepsilon_{ij}$ : are random variables which describe the random component of sampling, preparation and measurement error.

It is assumed that the  $\varepsilon_{ij}$  are independent and normally distributed with expectation zero and variance  $V(\varepsilon_{ij}) = \sigma^2(x_i)$ .

The well-known estimates for the weighted linear regression (WLSR) [2] are calculated from the  $N = I * J$  data pairs of the calibration data set  $\{x_i, y_{ij}\}$ , where  $i = 1, \dots, I$  and  $j = 1, \dots, J$ , respectively.

For this study the weights are estimated by  $w_i = 1/\sigma^2(x_i) := 1/s_i^2$ , where  $s_i$  symbolizes the standard deviation of the repeated measurements of response  $y$  at  $x_i$ .

For the regression coefficients  $\hat{b}_1 = \frac{\sum w_i \sum w_i x_i y_{ij} - \sum w_i x_i \sum w_i y_{ij}}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2}$  and  $\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$  are used, where  $\bar{x} = \sum w_i x_i / \sum w_i$ ;  $\bar{y} = \sum w_i y_{ij} / \sum w_i$  and the predicted value of the regression line at position  $i$  is given by  $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$ .

## Test for non-linearity

The correctness of the model and the goodness of fit have to be checked with the help of the residual plot belonging to the residuals

$$e_{ij} = \sqrt{w_i}(y_{ij} - \hat{y}_i). \quad (2)$$

L. Brüggemann (✉) · R. Wennrich  
Department of Analytical Chemistry,  
UFZ Centre for Environmental Research Leipzig-Halle,  
Permoserstrasse 15,  
04318 Leipzig, Germany  
e-mail: lutz.brueggemann@ufz.de

W. Quapp  
Institute of Mathematics, University of Leipzig,  
Augustusplatz 10/11,  
04109 Leipzig, Germany

**Table 1** Main components of the water samples under investigation

	Cospuden recultivated mining lake water Ex. 1	Kulkwitz Recultivated mining lake water Ex. 2	Luppe river water Ex. 3	Pleisse river water Ex. 4	RL 111 mining lake water Ex. 5	Schwel-Vollert mining lake water Ex. 6
Analyte	Concentration (mg/L)					
Na	47	74	117	60	7	19
K	8	16	30	10	3	6
Ca	325	352	140	138	206	225
Mg	78	43	37	35	27	49
Fe	<0.1	<0.1	0.9	5	130	<0.1
NO <sub>3</sub> <sup>-</sup>	3	3	27	63	1	<1
Cl <sup>-</sup>	73	175	170	98	8	610
SO <sub>4</sub> <sup>2-</sup>	1050	910	350	315	1150	82

By visual inspection of the residual plot, deviations of linearity, normality and variance homogeneity, and also that of outliers are recognized.

With the help of the Durbin-Watson test it can be checked for a sequential dependence in which each error (and so residual) is correlated with those before and after it in the sequence. The test focuses specifically on the differences between successive residuals on the base of the Durbin-Watson statistic

$$D = \frac{\sum_{u=2}^N (e_u - e_{u-1})^2}{\sum_{u=2}^N e_u^2} \quad (3)$$

For  $D$  was shown [2], that

- $0 \leq D \leq 4$  always
- If successive residuals are positively serially correlated,  $D$  will be near 0
- If successive residuals are negatively serially correlated,  $D$  will be near 4
- The distribution of  $D$  is symmetric about 2

The residual standard deviation of the regression

$$s_e := \hat{\sigma} = \sqrt{\sum_{i,j} (y_{ij} - \hat{y}_i)^2 / (N - p)} \quad (4)$$

is an estimate of the variance of the measurements (pure experimental error), if the model is correct.

That means, that for  $p=2$  (linear calibration function), a comparatively large value of  $s_e$  can indicate weak non-linearity. Likewise, for  $p=2$  a small value of the coefficient of determination of the regression

$$R^2 = 1 - \frac{\sum_{i,j} (y_{ij} - \hat{y}_i)^2}{\sum_{i,j} (y_{ij} - \bar{y})^2}, \quad (5)$$

which represents the part of the variance that is explained by regression (this quantity is equal to the square of the correla-

**Table 2** Experimental parameters for the determination of arsenic

ICP spectrometer	CIROS, Spectro A.I.
Wavelength	189.042 nm (background corrected)
Power	1.4 kW
Nebulizer	Cross flow
Uptake rate	1.9 mL min <sup>-1</sup>
Ar Gas flow	Aerosol gas (0.9 L min <sup>-1</sup> ) Cooling gas (14 L min <sup>-1</sup> ) Auxilliary gas (0.8 L min <sup>-1</sup> )

tion coefficient), can indicate weak non-linearity. However, a possible misuse of this indicator must be avoided (see for instance [3, 4, 8]).

The suitability of regression model should be proven by a special statistical lack-of-fit-test [5], which is based on analysis of variance (ANOVA). In this way the residual sum of squares of regression is separated into two components: the sum of squares due to lack of fit (LOF) and the pure “error” sum of squares (PE)

$$\sum_{i,j} w_i (y_{ij} - \hat{y}_i)^2 = \sum_i J w_i (\bar{y}_i - \hat{y}_i)^2 + \sum_{i,j} w_i (y_{i,j} - \bar{y}_i)^2. \quad (6)$$

The following  $F$ -test is based on the means of the deviation squares concerned,  $MS_{LOF} = \sum_i J w_i (\bar{y}_i - \hat{y}_i)^2 / (I - 2)$  and  $MS_{PE} = \sum_{i,j} w_i (y_{i,j} - \bar{y}_i)^2 / (N - I)$ . The hypothesis  $H_0$  (regression model is suitable) is examined by means of the test value  $F_{\text{test}} = MS_{LOF} / MS_{PE}$ , which is compared with the corresponding value of the  $F$ -distribution with  $I - 2$  and  $N - I$  degrees of freedom at the significance level  $\alpha$ . For  $F_{\text{test}} > F_{\alpha; I-2; N-I}$   $H_0$  is rejected, i.e. a non-linearity referring to the calibration model (1) is indicated.

Further possibilities for testing for non-linearity results from application of the polynomial calibration model of

**Table 3** Calibration data sets for the calibration experiments for measurement of the arsenic content of six water (concentrations and responses are denoted by  $x$  and  $y$ , respectively). *Ex* example

Calibration level (standard addition) (mg L <sup>-1</sup> )	Cospuden recultivated mining lake water Ex. 1	Kulkwitz recultivated Mining lake water Ex. 2	Luppe river water Ex. 3	Pleisse river water Ex. 4	RL 111 mining lake water Ex. 5	Schwel-Vollert mining lake water Ex. 6
$x$	$y$	$y$	$y$	$y$	$y$	$y$
0.0	-43	-89	-119	-36	-14	21
0.0	-59	31	-36	31	-44	21
0.0	-92	10	-80	-70	-48	-63
0.0	-31	-17	-18	-17	-63	9
0.5	1,436	1,340	1,329	1,283	1,313	1,222
0.5	1,386	1,352	1,280	1,211	1,243	1,226
0.5	1,376	928	1,376	1,406	1,270	1,367
0.5	1,396	909	1,278	1,307	1,307	1,299
2.0	5,527	5,152	5,029	5,212	4,804	4,773
2.0	5,429	5,493	5,177	5,040	4,891	4,785
2.0	5,179	5,520	4,959	5,072	4,831	5,203
2.0	5,377	5,370	4,933	5,124	4,896	5,120
5.0	13,554	12,720	12,749	12,911	12,350	12,661
5.0	13,848	13,206	12,985	12,708	12,406	12,808
5.0	13,940	13,133	13,315	13,466	12,735	13,063
5.0	13,185	13,566	13,389	13,148	12,663	13,144
10.0	26,080	25,560	25,460	25,649	24,833	24,575
10.0	27,439	25,258	25,292	24,970	25,180	25,000
10.0	27,500	25,960	26,109	25,981	25,101	24,648
10.0	25,768	26,394	25,579	25,340	25,296	24,660

degree  $p-1 \geq 2$

$$Y_{ij} = b_0 + \sum_{l=1}^{p-1} b_l x_l^l + \varepsilon_{ij}. \tag{7}$$

By means of Mandel’s fitting test [6], it can be determined as to whether the residual variances, resulting from the linear and the quadratic calibration function, significantly differ. The hypothesis  $H_0$  (no significant difference between the residual variances) is examined by means of the test value

$$F_{\text{test}} = ((N - 2)s_{e,2}^2 - (N - 3)s_{e,3}^2) / s_{e,3}^2. \tag{8}$$

This is compared with the corresponding value of the  $F$ -distribution with 1 and  $N-3$  degrees of freedom at the significance level  $\alpha$ . For  $F_{\text{test}} > F_{\alpha;1;N-3}$   $H_0$  is rejected. Therefore, the calibration problem is fitted better by a quadratic function than by the linear one. Thus, an indication of non-linearity of the calibration function is obtained.

Recently, another non-linearity test based on the model (7), created by Mark and Workman [7], was suggested. To avoid correlations between the various powers of  $X$ , a transformation of the original variable  $X$  is accomplished in the following manner

$$x_1 := x \quad \text{and} \quad x_l := (x - z_l)^l \tag{9}$$

resulting in the multiple linear regression model

$$Y_{ij} = b_0 + \sum_{l=1}^{p-1} b_l x_l + \varepsilon_{ij}. \tag{10}$$

Evaluation of non-linearity of the calibration is reduced on testing the statistical significance of each of the coefficients.

The condition

$$\text{Cov}(x, x_l) = 0 \tag{11}$$

gives equations solvable for  $z_l$ .

To improve the representation, in the following the abbreviation

$$H_l = \sum x^l (x - \bar{x}) \tag{12}$$

is used.

For  $l = 2$  one obtains

$$\sum (x - \bar{x})(x - z_2)^2 = 0 \tag{13}$$

$$\sum x^2 (x - \bar{x}) - 2z_2 \sum x(x - \bar{x}) = 0 \tag{14}$$

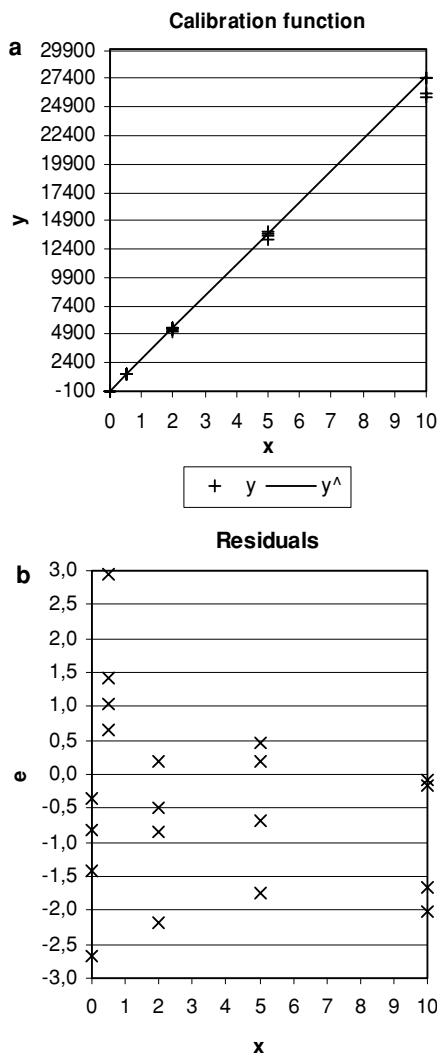
$$z_2 = H_2 / 2H_1. \tag{15}$$

**Table 4** Results for variance homogeneity tests refer to the calibration experiments of the examples represented in Table 3. *Ex* example

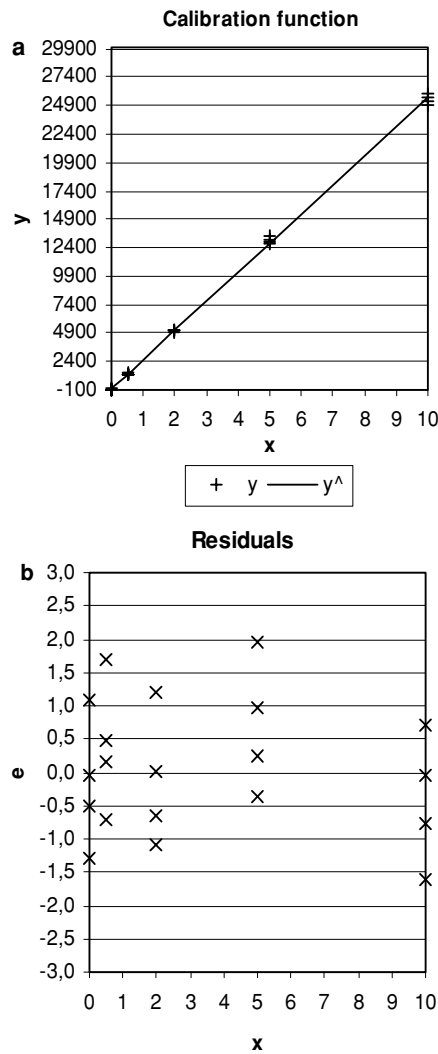
Test quantity	Cospuden recultivated mining lake water Ex. 1	Kulkwitz recultivated mining lake water Ex. 2	Luppe river water Ex. 3	Pleisse river water Ex. 4	RL 111 mining lake water Ex. 5	Schwel-Vollert mining lake water Ex. 6	Critical value of the corresponding distribution
<i>F</i> -test							
$S^2_{f \text{ min}}$	691.667	2,738.250	2,059.583	1,776.667	421.583	1,632.000	
$S^2_{f \text{ max}}$	813,023.250	24,3201.333	124,528.667	186,387.333	38,693.667	50,082.000	
$F_{\text{test}} = S^2_{f \text{ max}}/S^2_{f \text{ min}}$	1,175.45	88.82	60.46	104.91	91.78	30.69	$F_{J-1;J-1;1-\alpha/2} = 15.4$
	Significant	Significant	Significant	Significant	Significant	Significant	
Bartlett-test							
$X^2_{\text{test}} = (1/c)^*$	114.2	100.8	92.5	96.7	82.4	86.8	$X^2_{J-1;1-\alpha/2} = 11.1$
$\sum [(J - 1) \ln(S^2_w / S^2_f)]$	Significant	Significant	Significant	Significant	Significant	Significant	
	VH violated (strong)					VH violated (weak)	

**Table 5** Results for non-linearity indicators and tests refer to the calibration experiments of example 1 represented in Table 3. *non-lin* non-linearity; *Ex* example

	Cospuden recultivated mining lake water Ex. 1	Kulkwitz recultivated mining lake water Ex. 2	Luppe river water Ex. 3	Pleisse river water Ex. 4	RL 111 mining lake water Ex. 5	Schwel-Vollert mining lake water Ex. 6
Non-linearity indicators and tests	Results					
Coefficient of determination	0.997995	0.998811	0.999108	0.999417	0.999492	0.999522
Residual standard deviation	1.461	1.128	1.289	1.017	1.794	1.427
Visual inspection of residuals	Some indication of non-lin	Little indication non-lin	Little indication non-lin	No indication non-lin	Little indication non-lin	Little indication non-lin
Test for serial correlation of residuals (Durbin-Watson)	Test is inconclusive	No non-lin	No non-lin	No non-lin	No non-lin	No non-lin
<i>F</i> -test: Lack of fit	Non-lin	No non-lin	Non-lin	No non-lin	Non-lin	Non-lin
Mandel's fitting test	Non-lin	Non-lin	No non-lin	No non-lin	No non-lin	Non-lin
Test for non-linearity (Mark/Workman) (With $p = 4$ : Significance of $b_2$ or $b_3$ )	Non-lin	Non-lin	No non-lin	No non-lin	No non-lin	Non-lin



**Fig. 1** a Regression line of the WLS-regression to the calibration of example 1. b Residuals of the WLS-regression to the calibration of example 1



**Fig. 2** a Regression line of the WLS-regression to the calibration of example 4. b Residuals of the WLS-regression to the calibration of example 4

$l = 3$  yields

$$\sum (x - \bar{x})(x - z_2)^3 = 0 \tag{16}$$

$$\sum x^3(x - \bar{x}) - 3z_3 \sum x^2(x - \bar{x}) + 3z_2^2 \sum x(x - \bar{x}) = 0 \tag{17}$$

$$z_3 = H_2/2H_1 \pm \sqrt{H_2^2 - (4/3)H_1H_3/2H_1}. \tag{18}$$

However, it can be proved that for the radicand

$$H_2^2 - (4/3)H_1H_3 \leq 0 \tag{19}$$

is valid, i.e., for  $z_3$  only the real solution

$$z_3 = H_2/2H_1 \tag{20}$$

can be achieved and the correlation between  $x$  and  $(x - z_3)^3$  has a minimum value, however Eq. (11) is not completely fulfilled. Equations (17), (18) and (20) differ slightly from the results of Mark/Workman [7], because the equation [A14] given in the authors' paper [7] is not completely correct.

In the following the hypothesis  $H_0$  ( $b_l$  not significant) is examined by means of the test value

$$t_{l,\text{test}} = b_l/s_{bl}, \tag{21}$$

which is compared with the corresponding value of the  $t$ -distribution with  $N-p$  degrees of freedom at the signifi-

**Table 6** Results for non-linearity indicators and tests refer to the calibration experiment of Example 1 (taken from the corresponding MS Excel spreadsheet)

	$N$ (measurements)		20	20	20
	$I$ (calibration levels)		5		
	$P$ (dimension of the model)	1	2	3	4
			$y = b_0 + b_1x$		
Visual inspection	$b_l(l = 1, \dots, p - 1)$	- 21.647	2760.718		
	$e = \sqrt{w}(y - \hat{y})$		Some indication for non-linearity		
Test for serial correlation of residuals (Durbin/Watson)	$D_{\text{test}} = \sum (e_u - e_{u-1})^2 / \sum e_u^2$		1.063 (test inconclusive)		
	$D_{\text{Low},N,p,1\%}$		0.95		
	$D_{\text{Up},N,p,1\%}$		1.15		
			$y = b_0 + b_1x + b_2x^2$		
ANOVA	$b_l(l = 1, \dots, p - 1)$	- 40.232	2843.589	- 20.009	
	$SS_{\text{Tot}} = \sum w(y - y_M)^2$		19170.281	19170.281	
	$SS_{\text{Reg}} = \sum w(y_M - \hat{y})^2$		19131.851	19143.860	
	$SS_{\text{Res}} = \sum w(y - \hat{y})^2$		38.431	26.421	
	$SS_{\text{PE}} = \sum w(y - y_{iM})^2$		15.000		
	$SS_{\text{LOF}} = \sum w(y_{iM} - \hat{y})^2$		23.431		
$F$ -test: lack of fit	$F_{\text{test}} = [SS_{\text{LOF}}/(I - p)]/[SS_{\text{PE}}/(N - I)]$		7.810		
	$p$ -level; $df1 = I - p$ , $df2 = N - I$		0.002262 (significant)		
Coeff. of determination	$R^2 = 1 - SS_{\text{Res}}/SS_{\text{Tot}}$		0.998997	0.998622	
Correlation coefficient	$R = \sqrt{R^2}$		0.999656	0.999311	
Residual variance	$S_e^2 = SS_{\text{Res}}/(N - p)$		2.135	1.554	
Residual SD	$S_e = \sqrt{S_e^2}$		1.461	1.247	
Mandel's fitting test	$F_{\text{test}} = [(N - 2)S_{e,2}^2 - (N - 3)S_{e,3}^2]/S_{e,3}^2$		7.727252		
	$p$ -level; $df1 = 1$ , $df2 = N - 3$		0.012840 (significant)		
Test for non-linearity (Mark/Workman)				$y = b_0 + b_1x + b_2(x - z_2)^2$	
	$b_l(l = 1, \dots, p - 1)$	461.097	2643.279	- 20.009	
	$s_{bl}$	174.200	49.030	7.198	
	$t_{1,\text{test}} = b_l/s_{bl}$		53.911	- 2.780	
	$p$ -level; $df = N - p$		0.000000 (significant)	0.012840 (significant)	
				$y = b_0 + b_1x + b_2(x - z_2)^2 + b_3(x - z_3)^3$	
	$b_l(l = 1, \dots, p - 1)$	482.437	2591.878	- 2.277	3.779
	$s_{bl}$	166.414	56.079	12.717	2.283
	$t_{1,\text{test}} = b_l/s_{bl}$		46.219	- 0.179	1.655
	$p$ -level; $df = N - p$		0.000000 (significant)	0.860138 (not significant)	0.117314 (not sign.)

cance level  $\alpha$ . For  $t_{\text{test}} > t_{1-\alpha/2;N-1}$   $H_0$  is rejected. Thus, an indicator for non-linearity of the calibration function is obtained.

The partial  $F$ -statistic with 1 and  $\nu$  degrees of freedom is exactly equal to the square of the  $t$ -statistic with  $\nu$  degrees of freedom obtained from Eq. (21) (see reference [2])

$$F_{1,df,1-\alpha} = \{t_{df,1-\alpha/2}\}^2; \tag{22}$$

therefore, the tail probability to the test value  $t_{2,\text{test}}$  of the quadratic regression coefficient is identical with the corresponding probability to Eq. (8).

The aim of the present work is both to investigate the dependence of matrix influence on the linearity of calibration and to compare the novel non-linearity test [7] with the conventional non-linearity tests represented above by means of practical examples.

## Experimental

An environmentally important task is the determination of arsenic in surface water samples, e.g. in areas influenced by mining activities. Depending on the sampling site, the effects of different matrices with different strengths can occur. These can affect the calibration. For this study, water samples were collected from several lakes in abandoned lignite mining areas and two rivers, which pass through a mining region in Germany. For the determination of arsenic ICP atomic emission spectrometry, a relatively less sensitive analytical technique for this task, was chosen. For the determination of arsenic, an ICP atomic emission spectrometer CIROS (Spectro A.I.) with pneumatic nebulization (cross-flow) was used. The samples are distinctly different concerning their matrix composition (Table 1). The experimental parameters are given in Table 2.

For each sample a calibration was done based on standard addition (0, 0.5, 2, 5, 10 mg L<sup>-1</sup> of arsenic) in four parallel runs. The designed calibration experiments for the comparison of the non-linearity tests represented above are compiled in Table 3. A preceding examination of variance homogeneity with all examples resulted in this condition being violated (see Table 4). Therefore weighted linear regression models for the calibration were used.

## Results and discussion

The results concerning all calibration experiments are compiled in Table 5. Despite significant differences in the sample composition concerning the matrix elements, the sensitivity of the calibration (not discussed here) was not an influence in the calibration range. As shown by the data for the different natural water samples, we found different results in testing the linearity of the calibrations despite the facts that the calibrations are based on a standard addition procedure and the use of a less sensitive ICP atomic emission spectrometry technique for the determination of arsenic. This indicates that the linearity of the calibration is influenced by the varying matrices.

One can see that all tests in the case of strong non-linearity (example 1, see Figs. 1a,b, example 6) indicate “non-linearity”, although the calibration was only done in a calibration range of 1.5 orders of magnitude in concentration. That means that the calibrations of examples 1 and 6

have to be examined. In the case of strong linearity (example 4, see Figs. 2a,b), all tests indicate “no non-linearity”. More difficult to judge are the cases characterized by a weak non-linearity (examples 2, 3, and 5). In these cases, the result of the Mandels fitting test is preferred, because this test gives a special result, i.e., that a quadratic calibration function is more suitable as a linear function.

For the non-linearity test [7] and the model with  $p - 1 = 2$ , it is shown that the Mandels fitting test yields the same tail probability 0.012840 (see Table 6), demonstrating the validity of Eq. (22). The test for serial correlation of residuals is proved as being a little sensitive in the treated cases of linear calibration. This test is ineffective in our case of calibration data, which are typically chemical measurements ( $I = 5$  calibration levels,  $J = 4$  replicates). Also, the residual standard deviation of the regression and the coefficient of determination can not be used in the treated cases as an indicator for non-linearity.

## Conclusion

From the results shown one must conclude that an appropriate statistic examination of the linearity is necessary, working with samples characterized by a strong matrix load. The residual plot absolutely has to be checked. Moreover, the lack-of-fit test (based on replications) and above all the Mandels fitting test are proven as being efficient tools for the detection of non-linearity.

## References

1. ISO 11843-2 (2005) Capability of detection: methodology in the linear calibration case. Draft, ISO Geneva
2. Draper NR, Smith H (1998) Applied regression analysis. 3rd edn. Wiley, New York
3. Van Loco J, Elskens M, Croux C, Beernaert H (2002) Accred Qual Assur 7:281–285
4. Huber W (2004) Accred Qual Assur 9:726
5. Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J (1997) Handbook of chemometrics and quality metrics. part A, Elsevier, Amsterdam, pp 171–230
6. Mandel J (1964) The statistical analysis of experimental data. Wiley, New York
7. Mark H, Workman J (2005) Chemometr Spectrosc 20(9):26–35
8. Davies AMC, Fearn T (2006) Spectrosc Eur 18(2): 31–32