



KATHOLIEKE  
UNIVERSITEIT  
LEUVEN

# DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0205

ON THE MISUSE OF THE CORRELATION  
COEFFICIENT TO ASSESS LINEARITY OF  
CALIBRATION CURVES

by

J. VAN LOCO  
M. ELSKENS  
C. CROUX  
H. BEERNAERT

D/2002/2376/05

# On the misuse of the correlation coefficient to assess linearity of calibration curves

JORIS VAN LOCO

MARC T. ELSKENS

CHRISTOPHE CROUX

HEDWIG BEERNAERT

*Van loco J. (✉), Elskens M.T. and Beernaert H.*

*Scientific Institute of Public Health, J. Wytsmanstraat 14, B-1050 Brussels, Belgium*

E-mail: [Joris.vanloco@iph.fgov.be](mailto:Joris.vanloco@iph.fgov.be)

Tel.: +32-2-6425356

Fax.: +32-2-6425327

*Croux C.*

*Department of Applied Economics, K.U. Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.*

**Abstract:** The correlation coefficient is commonly used to evaluate the degree of linear association between two variables. However, it can be shown that a correlation coefficient very close to one might also be obtained for a clearly curved relationship. Other statistical tests, like the Lack-of-fit and Mandel's fitting test appear therefore more suitable for the validation of the linear calibration model.

A number of cadmium calibration curves from atomic absorption spectroscopy were assessed for their linearity. All the investigated calibration curves were characterised by a high correlation coefficient ( $r > 0.997$ ) and low quality coefficient ( $QC < 5\%$ ), but the straight-line model was systematically rejected at the 95% confidence level on the basis of the Lack-of-fit and Mandel's fitting tests. Furthermore, significantly different results were achieved between a linear (LRM) and a quadratic regression (QRM) model in forecasting values for mid-scale calibration standards. The results obtained with the QRM did not differ significantly from the theoretically expected value, while those obtained with the LRM were systematically biased. It was concluded that a straight-line model with a high correlation coefficient, but with a lack of fit, yields significantly less accurate results than its curvilinear alternative.

*Keywords: linearity - goodness of fit - correlation coefficient - lack of fit - calibration*

GF-AAS: Graphite Furnace Atomic Absorption Spectrometry

LOF: Lack of fit

LRM: Linear Regression Model

QRM: Quadratic Regression Model

$r$ : Correlation coefficient

## **Introduction**

The linear range of most analytical instruments is known to be limited. Therefore, during method validation the linearity of the calibration curve should be assessed and the working range of the calibration curve be determined. [1,3]. The correlation coefficient  $r$  is commonly used for this purpose, and curves with  $r \geq 0,995$  are usually considered to be linear. Nevertheless, several investigators focussed on the fact that  $r$  might not be a useful indicator of linearity [2,3], and other statistical tests or quality parameters have been suggested to ascertain the goodness of fit of the calibration curve [3, 4, 5].

On the contrary, a calibration curve with  $r \geq 0,995$  can be considered nearly linear. Furthermore, from an inference point of view, linear regression models are easy to implement, compared to curvilinear or non-linear regressions models [6]. Therefore, a straight-line calibration curve should always be preferred over curvilinear or non-linear calibration models if equivalent results can be gained. A prerequisite, however, is that one should be able to assess this equivalence. In other words, is there any evidence for a systematic difference between the results of the two models at a given confidence level?

In this paper, cadmium calibration curves from atomic absorption spectroscopy were tested for their linearity. Alternative curvilinear regressions were proposed when linearity was rejected. Predictions made on the basis of the fitted curve for both linear and curvilinear models were compared.

## Assessing the linearity of calibration curves

Graphite Furnace Atomic Absorption Spectroscopy (GFAAS) is known to have a limited linear calibration range. In order to assess the linearity of the calibration process, several calibration lines for cadmium were constructed over a period of four months. These calibrations were performed using standard solutions prepared from the corresponding high purity metal Baker Cd Atomic Absorption Standard of 1000  $\mu\text{g/ml}$  (NIST traceable). The GF-AAS was programmed to produce a calibration curve with the following concentrations: 0 - 0,8 - 1,6 - 2,4 - 3,2 - 4,0  $\text{ng/ml}$ . The solutions were injected in duplicates.

The linearity of the calibration process was investigated by means of the Lack-of-fit test [3], Mandel's Fitting Test Value [5], the Quality Coefficient [3,4] and the correlation coefficient [2,3]. The results are summarised in Table 1.

<b>Table 1</b>
----------------

The results in Table 1 shows that for the Lack-of-Fit test the linear regression model (LRM) must systematically be rejected at the 95% confidence level ( $F_{\text{crit},95\%} = 5,41$ ), and for the Mandel's Fitting test even rejected at the 99% confidence level ( $F_{\text{crit},99\%} = 10,56$ ). Thus, despite that the correlation  $r$  and quality coefficients  $QC$  are greater than 0,997 and lower than 5% respectively, the linearity of the calibration lines were rejected on the basis of the before mentioned  $F$ -tests. This corroborates the statements of the Analytical Methods Committee [2] that  $r$  should be used with care when evaluating the linearity of calibration lines. Moreover, questions arise regarding the significance of  $QC$ , for which an upper limit of 5% was proposed in assessing the suitability of a calibration process [4]. Here, even with a  $QC$ -value less than 3%, the LRM is rejected at the 95% confidence level (Table 1).

Alternatively, the residual plots give useful information to validate the chosen regression model. The residual plot can be used to check whether the underlying assumptions, like normality of the residuals and homoscedasticity, are met as for evaluating the goodness of fit of the regression model. [3] Figure 1a shows a

residual plot for a LRM. The U-shaped residual plot indicates that a curvilinear regression model should be preferred over a LRM.

## Figure 1

Several authors [3,6,8,9,10,11] recommended alternative calibration functions when linearity of the calibration curve has to be rejected. To correct the non-linearity, a quadratic curvilinear function ( $f(x) = a + bx + cx^2$ ) was chosen. The lack-of-fit tests for the quadratic regression model (QRM) are summarised in table 1. The test for lack-of-fit reveals that this QRM adequately fits the calibration data at 99% (highly significant) confidence level and at the 95% (significant) confidence level in all cases except one. In determining whether the order of the polynomial regression model is appropriate, the significance of the second order coefficient is estimated. The p-value on the second order coefficient, shown in table 1, is systematically smaller than 1%. Consequently, a lower order model should not be considered. In addition, residual plots (figure 1b) were constructed for this QRM. The residuals were randomly scattered within a horizontal band around the centre line. Therefore, the QRM was chosen as the reference model. Note that an increase of the variance can be seen at higher concentrations.

## Predictions made on the basis of the fitted curve for linear (LRM) and quadratic (QRM) models

To gauge the agreement/disagreement between predicted concentrations calculated from the LRM and QRM, a mid-scale calibration standard (2 ng/ml) was systematically injected in duplicates. The instrument signal corresponded to a point close to the centroid of the data cloud, where the confidence limits for the regression line of LRM was the narrowest.

To compare the outcome of both regression models, the predicted concentration of the mid-scale standard was expressed both as a recovery rate and as a relative deviation. Hence, the following formulas were used:

$$\text{Recovery (\%)} = \frac{\text{determined concentration}}{\text{nominal concentration}} \times 100\%$$

$$\text{Relative deviation (\%)} = |100 - \text{recovery(\%)}|$$

To investigate possible time effects several calibration lines were produced in a period of almost four months. The mid-scale standard was determined twice in the beginning and end of an analysis. The recovery rate and relative deviation of the results are summarised in Table 2.

## Table 2

If both curves yielded equivalent results and were not biased, the recovery rate should be around 100%. Figure 2 shows the recovery rates for the mid-scale standard calculated with the LRM and QRM. It clearly appears that the median from the LRM differs from the theoretical value of 100%. In general, the recovery rates are overestimated when calculated with the LRM. A systematic error of about 4% was found.

This result is supported by the one sample Wilcoxon Signed Ranks Test [12]. The null hypothesis (median = 100%) is rejected for the results determined with the LRM ( $p \leq 0,0004$ ), but not for those derived from the QRM ( $p \leq 0,1788$ ). The one sample student *t*-test gave similar outcomes (LRM:  $p \leq 8,3 \cdot 10^{-8}$ , QRM:  $p \leq 0,177$ ).

## Figure 2

Furthermore to gauge whether the results calculated with the QRM were more accurate than those obtained from the LRM the relative deviation of the data were compared. The non-parametric paired "sign" test was chosen for this evaluation because the relative deviations were not symmetrically distributed. The null-hypothesis stating equivalent results from both models must be rejected ( $n = 14$ ,  $n+ = 2$ ,  $p < 0,006$ ). Therefore, our results indicate a systematic bias on the forecast concentration when applying the LRM. The relative deviation for the QRM is significantly smaller than the one obtained from the LRM. On the contrary, there

is no evidence for a difference about the spread or dispersion of the results between both models, which has been confirmed with the Levene test for homogeneity of variances. ( $F_{\text{Levene}} = 2,70$ ;  $p \leq 0,11$ ) [13].

## Conclusion

As claimed by several investigators [2,3], this paper corroborates the fact that the correlation coefficient is not a useful indicator of linearity in calibration model, even for  $r$ -values  $> 0,997$ . In addition, the present results raise the question about the relevance of the quality coefficient QC in assessing the process calibration. Other statistical tests like the Lack of fit and the Mandel's fitting tests seems more appropriate for evaluating the linearity of the calibration curve during method validation. Preferably, the Lack of fit and the Mandel's fitting tests should be used in conjunction with an evaluation of the residual plot.

Furthermore, it is shown that a straight-line model with  $r > 0,997$  and  $QC < 5\%$  but with lack of fit, yielded forecast values for a mid-scale calibration standard that significantly differ from the nominal ones. In general, the recovery tests were overestimated, while the precision on the result was comparable in both LR and QR models. The bias can be considered as significant since the repeatability of the measurements is usually less than 2%. Furthermore, the situation would even be worse whether the comparison had been carried out with either high or low range calibration standards.

In conclusion, the results in this paper indicate that the correlation coefficient is not suitable for assessing the linearity of calibration curves. Statistical tests like the Lack of fit and the Mandel's fitting should be systematically applied during full method validation. It was shown that in this application cadmium concentrations calculated with the LRM were constantly overestimated by about 4%.

## References

- [1] European Commission (1990) 90/515/EEC: Commission Decision of 26 September 1990 laying down the reference methods for detecting residues of heavy metals and arsenic, Official Journal L286: 33-39
- [2] Analytical Methods Committee (1988) Analyst 113: 1469-1471
- [3] Massart DL, Vandeginste BGM, Buydens LMC., De Jong S, Lewi PJ, Smeyers-Verbeke J (1997) Handbook of Chemometrics and qualimetrics: part A, Elsevier Amsterdam
- [4] Vankeerberghen P, Smeyers-Verbeke J (1992) Chemometrics and intelligent Laboratory Systems, 15: 195-202
- [5] Mandel J (1964) The Statistical Analysis of Experimental Data, Interscience Publ. J. Wiley & Sons New York
- [6] Ratkowsky DA (1990) Handbook of Nonlinear Regression Models, Marcel Dekker Inc. New York
- [7] International Organization for Standardization (2000) ISO 11843-2: 2000, Capability of detection - part 2: Methodology in the linear calibration case, ISO Geneva
- [8] International Organization for Standardization (2001) ISO 8466-2: 2001, Water quality - Calibration and evaluation of analytical methods and estimation of performance characteristics - Part 2: Calibration strategy for non-linear second-order calibration functions, ISO Geneva
- [9] MacTaggart DL, Farwell SO (1992) Journal of AOAC International, 75: 594-608
- [10] Wang X, Smeyers-Verbeke J, Massart DL (1992) Analisis, 20: 209-215
- [11] Funk W, Dammann V, Donnevert G (1995) Quality Assurance in Analytical Chemistry, VCH Weinheim
- [12] Sheskin DJ (2000) Handbook of Parametric and Nonparametric Statistical Procedures, Second Edition Chapman & Hall / CRC Press
- [13] Levene H (1960). In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, I. Olkin et al. eds., Stanford University Press, pp. 278-292.

## Tables

Table 1: Value of the Lack-of-Fit (LOF) test, the Mandel's Fitting Test, the Quality Coefficient and the Correlation Coefficient for several linear calibration lines of Cd. For the quadratic regression model, the value of the Lack-of-Fit test, its associated P-value, and the P-value for testing significance of the second order coefficient for the quadratic regression model are represented

Linear regression model				Quadratic regression model		
LOF	Mandel's Test value	QC (%)	r	LOF	P-value on LOF	P-value on second order coefficient
11,08	51,46	3,93	0,9982	0,63	0,6227	0,0000
19,42	56,84	4,23	0,9978	1,58	0,2888	0,0000
7,13	26,29	3,67	0,9985	0,94	0,4792	0,0006
6,99	37,73	3,79	0,9984	0,18	0,9053	0,0002
11,43	58,21	4,03	0,9981	0,31	0,8189	0,0000
29,91	53,02	3,53	0,9986	4,08	0,0676	0,0000
49,80	71,07	3,76	0,9984	5,69	0,0345	0,0000
23,77	73,86	3,19	0,9989	1,66	0,2733	0,0000
31,95	63,37	3,24	0,9988	3,55	0,0875	0,0000
7,49	33,50	2,92	0,9991	0,54	0,6731	0,0003
9,99	55,19	3,95	0,9983	0,15	0,9268	0,0000
10,71	28,65	4,70	0,9975	1,89	0,2321	0,0005
25,21	79,60	3,34	0,9987	1,62	0,2815	0,0000
13,16	35,74	3,37	0,9987	1,93	0,2256	0,0002

Table 2: Recovery and relative deviation results for the linear (LRM) and quadratic (QRM) calibration curves for cadmium.

<b>Cadmium</b>				
<b>Date of analysis</b>	<b>LRM</b>		<b>QRM</b>	
	<b>Recovery</b>	<b>Relative deviation</b>	<b>Recovery</b>	<b>Relative deviation</b>
1/02/01	104,7675	4,767465	100,7667	0,7667
	103,1329	3,132905	99,13229	0,86771
8/02/01	103,6216	3,621582	99,78177	0,21823
	108,2779	8,277901	104,4544	4,4544
1/03/01	104,6719	4,671858	101,1446	1,1446
	106,2291	6,229144	102,7065	2,7065
15/03/01	105,042	5,042017	101,4986	1,4986
	106,7227	6,722689	103,1852	3,1852
2/04/01	104,1185	4,118483	99,81364	0,18636
	102,4143	2,414283	98,11269	1,88731
5/04/01	102,3539	2,35393	97,44148	2,55852
	105,8848	5,884826	100,9665	0,9665
9/05/01	106,3785	6,37847	103,3727	3,3727
	106,3785	6,37847	103,3727	3,3727
21/05/01	101,3516	1,351589	97,63152	2,36848
	102,826	2,826049	99,10214	0,89786
<b>Mean</b>	<b>104,6357</b>	<b>4,635729</b>	<b>100,7802</b>	<b>1,903273</b>
<b>Standard error</b>	<b>1,927598</b>	<b>1,927598</b>	<b>2,202419</b>	<b>1,27907</b>
<b>Median</b>	<b>104,7197</b>	<b>4,719661</b>	<b>100,8666</b>	<b>1,692955</b>
<b>1st Quartile</b>	<b>103,0562</b>	<b>2,979477</b>	<b>99,12475</b>	<b>0,882785</b>
<b>3rd Quartile</b>	<b>106,2665</b>	<b>6,303807</b>	<b>102,8262</b>	<b>2,94585</b>

# Figures

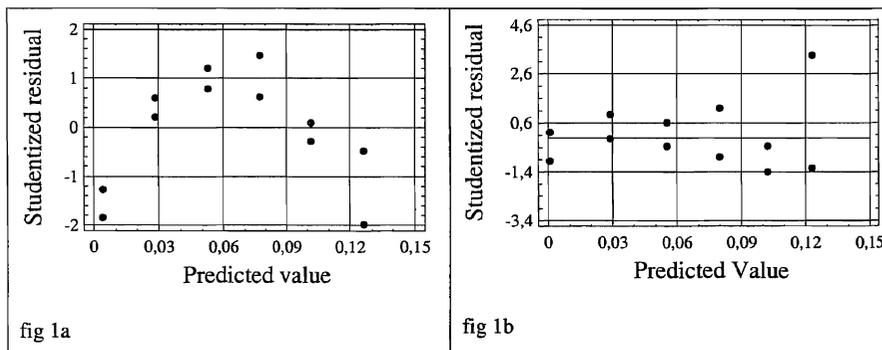


Figure 1: Plot of the residuals of the linear regression model (a) and the quadratic regression model (b) versus the predicted values

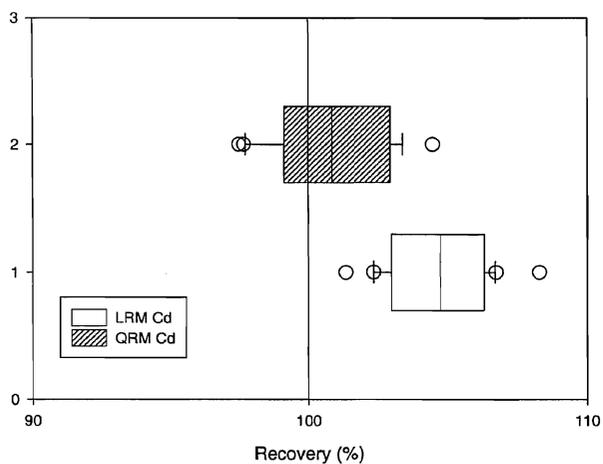


Figure 2: Recovery results for the mid-scale calculated with a second order calibration curve (QRM) and a linear calibration curve (LRM).

