THE LIMIT OF DETERMINATION AND THE LIMIT OF NONZERO INFORMATION CONTENT

Karel ECKSCHLAGER

Institute of Inorganic Chemistry, Czechoslovak Academy of Sciences, 250 68 Prague-Řež

The limit of nonzero information content, which is the lowest result of quantitative determination that is accurate and precise enough to have nonzero information content, is suggested to serve as the lowest limit of applicability of an analytical method whose results involve a systematic error. The limit of determination at which the relative precision of results is $\sigma_r = 10\%$ can be used as the lowest limit of applicability of a method whose results are accurate.

Publications dealing with the problem of the limit of determination in quantitative trace analysis[1−7] and the ample references therein document the importance of this quantity as well as the fact that neither its definition nor the way of its determination have been solved uniquely.

Initially, the limit of determination was established analogously as the limit of detection. Kaiser[1,2] defined the latter as the concentration x_N corresponding to the lowest signal discernible from the background or blank. The calibration dependence of the signal (y) on the analyte concentration (x) is a function $y = f_C(x)$, the analytical dependence is a function $x = f_A(y)$. The lowest signal discernible from the background or blank can be defined as

$$y_N = y_0 + 3\sigma(y_0),\qquad(1)$$

where y_0 is the background or blank level and the standard deviation $\sigma(y_0)$ characterizes the background or blank noise. The limit of detection then can be expressed as

$$x_N = f_A(y_N) = f_A(y_0 + 3\sigma(y_0)) = x_0 + 3\sigma(x_0),\qquad(2)$$

where $\sigma^2(x_0)$ is the variance of results of determination of the concentration x_0. This way of determining the limit of detection was later refined and modified; the various approaches to this problem have been compared by Long and Winefordner[8] (see also refs[3,4]).

Part XX in the series Theory of Information as Applied in Analytical Chemistry; Part XIX: Collect. Czech. Chem. Commun. *50*, 1359 (1985).

Analogously as the limit of detection, the limit of determination can be considered as

$$x_B = f_A(y_0 + w\sigma(y)) = x_0 + w\sigma(x) . \tag{3}$$

No unique convention, however, exists as to the value of the coefficient w. Whereas, for instance, Kaiser[2] uses $w = 6$ for the "Reinheitsgarantie" (purity guarantee), Currie[3] recommends $w = 10$ for the "limit of quantitative determination", arguing that at this value the relative standard deviation of determination is $\sigma_r(x) = 10\%$, which is a precision he regards acceptable. Also in more recent, mathematico--statistically founded work[3-8], the limit of determination is defined with respect to the precision of results, their accuracy being not taken into account.

Currie's definition has led to a new concept of the limit of determination, considering it as the lowest result that is still acceptably precise. This concept has been treated in detail by Liteanu and Rica in their monograph[4]; the authors also introduced information entropy as an additional criterion of acceptable precision of results. Beyermann[9] defines the limit of determination generally as the lowest result that is precise and accurate enough to enable the actual analyte content to be satisfactorily estimated. A tool for finding which result affords a satisfactory estimate of the actual analyte content is information theory[10-12].

THEORETICAL

Information content of a result of quantitative analysis, in terms of divergence measure[10-12] for a case where the expected analyte content lies within the interval of $\langle x_1, x_2 \rangle$ and the results of analysis obey the normal distribution $N(\mu, \sigma^2(x))$, is given by the relation

$$I(r; p, p_0) = \ln\left[(x_2 - x_1)/(\sigma(x)\sqrt{(2\pi e)})\right] - (1/2)\left[\delta/\sigma(x)\right]^2 , \tag{4}$$

where the mean error $\delta = |X - \mu|$, X being the true value of analyte content. In practice the analytical procedure is usually optimized for the results to be as precise as possible and the mean error to be statistically insignificant at a chosen significance level $(1 - \alpha)$, i.e., for δ not to exceed the value of $\sigma(x) z(\alpha)$, where $z(\alpha)$ is the critical value of the normal distribution. Information content according to Eq. (4) is negative at $\delta > \sigma(x)\left\{\ln\left[(x_2 - x_1)^2/(2\pi e\sigma^2(x))\right]\right\}^{1/2}$. Although negative information is not defined by classical information theory, it is reasonable, with regard to the semantic and pragmatic meaning of information as obtained by quantitative analysis[12], to interpret $I(r; p, p_0) < 0$ as a case where entirely incorrect results misinform us[11,12]. We usually put

$$I = I(r; p, p_0) \quad \text{for} \quad \delta < \sigma(x)\sqrt{\left\{\ln\left[(x_2 - x_1)^2/(2\pi e\sigma^2(x))\right]\right\}} \tag{5a}$$

and

$$I = 0 \quad \text{for} \quad \delta \geqq \sigma(x) \sqrt{\{\ln\left[(x_2 - x_1)^2/(2\pi e\sigma^2(x))\right]\}}, \tag{5b}$$

the case where $I = I(r; p, p_0)$ being referred to as a result with nonzero information content.

Nonzero information content can be adopted as meeting the requirement of a sufficiently precise and accurate result in Beyermann's definition of the limit of determination[9]. Putting $x_1 = x_0$, i.e. inserting the background or blank level for x_1, and substituting $\sigma(x) z(\alpha)$ for δ, Eq. (4) shows that nonzero information is obtained if

$$x_2/\sigma(x) \geqq \sqrt{(2\pi e)} \exp\left[(1/2) z^2(\alpha)\right] + x_0/\sigma(x). \tag{6}$$

Thus, we can define the limit of nonzero information content $x_1 = x_2$ with x_2 from Eq. (6) as the analyte content (e.g. concentration), as

$$x_1 = x_0 + A(\alpha) \sigma(x), \tag{7}$$

where

$$A(\alpha) = \sqrt{(2\pi e)} \exp\left[(1/2) z^2(\alpha)\right]. \tag{8}$$

Eq. (7) thus is analogous to Eq. (3) for the limit of detection, $A(\alpha)$ being in place of the coefficient w. $A(\alpha)$ values for $z(\alpha)$ corresponding to $0\cdot80 \leqq (1 - \alpha) \leqq 0\cdot95$ are given in Table I.

Similarly, for an analytical method involving a systematic error significant at the $(1 - \alpha)$ level, the maximum relative standard deviation $\sigma_{r,max}(x)$ for the results to have nonzero information content can be determined as

$$\sigma_{r,max}(x) = 100\sigma(x)/x_1 = 100/A(\alpha) = (100/\sqrt{e}) \operatorname{erf}(z(\alpha)), \tag{9}$$

TABLE I

$A(\alpha)$ values for critical $z(\alpha)$ values corresponding to significance of the mean error at $(1 - \alpha)$ levels between $0\cdot80$ and $0\cdot95$

$(1 - \alpha)$	$z(\alpha)$	$A(\alpha)$
0·8000	1·2816	9·40
0·8165	1·3300	10·00
0·8500	1·4400	11·65
0·9000	1·6449	15·99
0·9250	1·7800	20·15
0·9500	1·9600	28·21

where erf(z) is the error function,

$$\mathrm{erf}\,(z) = (1/\sqrt{}(2\pi))\exp\left[-(1/2)\,z^2(\alpha)\right].\tag{10}$$

DISCUSSION

The limit of nonzero information content x_1 according to Eqs (7), (8) can be adopted as a quantification of the limit of determination in the sense of Beyermann's definition, regarding a result sufficiently precise and accurate if it has nonzero information content. However, precision and accuracy of analytical results are properties that are not entirely independent. Information content according to Eq. (4) depends on σ and on $z(\alpha) = (\delta/\sigma)$, i.e., on the level at which δ is statistically significant. Iso-curves can be plotted for various information content values and $(1 - \alpha)$ and $\sigma_r(x)$ pairs. The zero iso-inform is shown in Fig. 1b.

The procedure for finding the limit of determination according to Currie's definition is illustrated in Fig. 1. The relative standard deviation $\sigma_r(x) = 100\sigma(x)/x$ is chosen, e.g. $\sigma_r(x) = 10\%$, and the corresponding x_N is found on the curve of dependence of $\sigma_r(x)$ on analyte content x for the analytical method under consideration. The limit of nonzero information content is determined likewise, only the $\sigma_r(x)$ value is not arbitrarily chosen; instead, it is found for the lowest $(1 - \alpha)$ level at which statistical insignificance of systematic error can be guaranteed, e.g. by calcula-
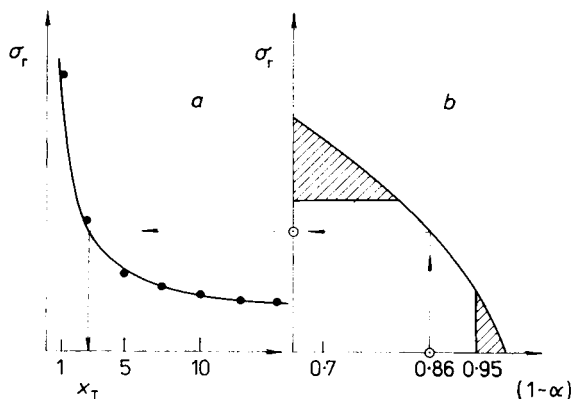

FIG. 1

Establishing the limit of nonzero information content for the determination of nickel. *a* Dependence of σ_r on nickel content; *b* course of $\sigma_{r,\max}$ for results involving a systematic error significant at the $(1 - \alpha)$ level (zero iso-inform). The value corresponding to the maximum relative standard deviation $\sigma_{r,\max}$ is found in Fig. 1b for $(1 - \alpha) = 0.86$, and for this σ_r, the limit of nonzero information content $x_1 = 2.70$ is found in Fig. 1a

tion using Eq. (9) or by using the iso-inform (Fig. 1b). Fig. 1 corresponds to a case where δ/σ is known to be 1·5, i.e. δ significant at the $(1 - \alpha) = 0·86$ level. The corresponding $\sigma_{r,max}(x)$ value is found in Fig. 1b, and the analyte content x_1 corresponding to nonzero information content is determined from the $\sigma_r(x)$ vs x plot for the analytical method in question (Fig. 1b). It will be clear that while the plot in Fig. 1a pertains to a particular analytical procedure, the zero iso-inform in Fig. 1b is generally valid. If the $\sigma_r(x)$ vs x plot has a minimum, which is not infrequent, x_1 is determined as the lower of the x values corresponding to the $\sigma_{r,max}(x)$ value found. The limit of nonzero information content can be determined by using Table I, by a simple calculation using Eqs (8), (9) or graphically. It is, however, necessary to determine the systematic error by analysis of standard samples. In some instances the level of significance $(1 - \alpha)$ need not be determined, calculation of the $A(\alpha)$ value (Eq. (8)) for the experimentally determined $z(\alpha)$ ratio being sufficient.

However, it is not easy to decide — particularly in trace analysis or where large numbers of standard samples are unavailable — whether the mean error $\delta_\pm = (X - \mu)$- which can be both positive or negative, is really "systematic". This is only unambit guous if the error has always the same sign and is sufficiently high, e.g., if it is significant at a level $(1 - \alpha) \geq 0·816$. Therefore, in the case of low mean errors of different signs, we will characterize the limit of determination in trace analysis in terms of Currie's definition, and only if the systematic error is significant at a level $(1 - \alpha) \geq 0·816$, the possible effect of δ on the actual analyte content estimate will be taken into account and the limit of nonzero information content will be given. Basically, then, definition (3) will be used choosing $w = 10$ for $0 \leq z(\alpha) \leq 1·33$ and $w = A(\alpha)$ for $1·33 \leq z(\alpha) \leq 1·96$. A case where δ/σ exceeds 1·96 is inacceptable in practice, giving evidence of a systematic error so high that it should be eliminated by appropriate calibration.

Establishing the Limit of Determination and the Limit of Nonzero Information Content; An Example

The procedure for establishing the limit of determination and the limit of nonzero information content will be illustrated on a case of the photometric determination of copper and nickel in high-purity alkali hydroxides after extraction and reextraction[13]. It is clear from the data in Tables II and III that the results are not accurate; the mean error $\delta_\pm$ is invariably positive, i.e., the results are lower. The dependence of σ_r on the analyte content x can be expressed by hyperbolic regression in the form

$$\sigma_r = a + b/x , \tag{11}$$

where the coefficients can be determined as

$$a = \left[\sum x \sum x\sigma_r - n \sum x^2\sigma_r\right]/\left[(\sum x)^2 - n \sum x^2\right] \tag{12}$$

TABLE II

Results of determination of copper after extraction and reextraction

X	μ	σ	σ_r	$\hat{\sigma}_r$	δ	$z = \delta/\sigma$
1·00	0·87	0·1070	10·70	11·14	+0·13	1·215
2·50	2·29	0·1495	5·98	5·98	+0·21	1·405
5·00	4·77	0·2171	4·34	4·26	+0·23	1·059
7·50	7·13	0·2760	3·68	3·69	+0·37	1·341
10·00	9·65	0·3480	3·48	3·40	+0·35	1·006
12·50	12·09	0·4038	3·23	3·23	+0·41	1·015
15·00	14·54	0·4620	3·08	3·12	+0·46	0·996

Average $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ *1·148 ± 0·159*

$A(\alpha) = \sqrt{(2\pi e)} \exp[(1/2) . 1\cdot148^2] = 7\cdot9876 \ 100 . A(\alpha) = 12\cdot52$
Hyperbolic regression: $\hat{\sigma}_r = 2\cdot5452 + 8\cdot59377/x$
Limit of determination: $x_N = 8\cdot59377/(10 - 2\cdot5452) = 1\cdot15$
Limit of nonzero information content: $x_I = 8\cdot59377/(12\cdot52 - 2\cdot4552) = 0\cdot86 \Rightarrow \sigma_r < 10\%$
The lower limit of applicability of the method for the determination of copper will be represented by the limit of determination.

TABLE III

Results of determination of nickel after extraction and reextraction

X	μ	σ	σ_r	$\hat{\sigma}_r$	δ	$z = \delta/\sigma$
1·00	0·76	0·182	18·20	17·53	+0·24	1·319
2·50	2·20	0·213	8·52	8·41	+0·30	1·408
5·00	4·61	0·261	5·22	5·36	+0·39	1·494
7·50	6·97	0·323	4·31	4·35	+0·53	1·641
10·00	9·48	0·380	3·80	3·84	+0·52	1·368
12·50	11·79	0·439	3·51	3·54	+0·71	1·617
15·00	14·19	0·508	3·39	3·33	+0·81	1·594

Average $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ *1·492 ± 0·120*

$A(\alpha) = \sqrt{(2\pi e)} \exp[(1/2) . 1\cdot492^2] = 12\cdot5783 \ 100 . A(\alpha) = 7\cdot95$
Hyperbolic regression (see Fig. 1a): $\hat{\sigma}_r = 2\cdot3209 + 15\cdot21184/x$
Limit of determination: $x_N = 15\cdot21184/(10 - 2\cdot3209) = 1\cdot98 \Rightarrow I(r; p, p_0) \leqq 0$
Limit of nonzero information content: $x_I = 15\cdot21184/(7\cdot95 - 2\cdot3209) = 2\cdot70$
The lower limit of applicability of the method for the determinatin of nickel will be represented by the limit of nonzero information content.

and

$$b = \left(\sum x \sigma_r - a \sum x\right)/n .\qquad(13)$$

The fit of this regression to the experimental data is good, as documented by the small differences between the calculated $\hat{\sigma}_r$ and observed σ_r values (for nickel, see also Fig. 1a). For the determination of copper and nickel, $\delta/\sigma \approx 1.15$ and 1.49, respectively, which correspond to significance at levels $(1 - \alpha) \approx 0.75$ and 0.86, respectively. For copper, the limit of nonzero information content is $x_I = 0.842$; however, $\sigma_r = 12.52\%$, so that at this limit the results do not have the required precision of $\sigma_r = 10\%$, and therefore, the value of $x_B = 1.127$ at which the results have the relative precision of 10% and nonzero information content will be taken as the lower limit of applicability of the analytical method. For nickel, the limit of nonzero information content $x_I = 2.725$ will be regarded as the lower limit of applicability of the method because at the limit of determination $x_B = 1.989$ the hazard exists that the results may have zero information content. (It is noteworthy that ultimately a precise and accurate method has been arrived at[13] by including AAS as the analytical finish and by modifying the calibration procedure.)

CONCLUSIONS

The limit of detection x_N according to Eq. (2) and the limit of determination x_B according to Eq. (3) using the w value as discussed above delimit three regions of analyte content:

1. Region of $x < x_N$, where the presence of the analyte in question cannot be proved; the only information gained from the absence of the analytical signal is that $0 \leqq \leqq x < x_N$. The information content of this finding according to refs[10,11,14,15] is given as $\ln (x_{max}/x_N)$, $x_{max} \leqq x_N$, where x_{max} is the highest expected analyte content.

2. Region of $x_N \leqq x \leqq x_B$, where analyte can be detected but not determined without the hazard that the information content of the result may be zero even in case that its mean error is statistically insignificant. The information content of analyte detection has been discussed, e.g. in refs[11,16-18].

3. Region of $x > x_B$, where analyte can be determined, the results having always nonzero information content according to Eq. (4) and relative precision $\sigma_r \leqq 10\%$. We can rightly suppose that such results lead to a satisfactory estimate of the actual analyte content. The information content of such results has been dealt with in refs[10-12].

REFERENCES

1. Kaiser H.: Fresenius Z. Anal. Chem. *209*, 1 (1965).
2. Kaiser H.: Fresenius Z. Anal. Chem. *216*, 80 (1966).
3. Currie L. A.: Anal. Chem. *40*, 586 (1968).
4. Liteanu C., Rica I.: *Statistical Theory and Methodology of Trace Analysis.* Horwood, Chichester 1979.
5. Ebel S., Kamm K.: Fresenius Z. Anal. Chem. *316*, 382 (1983).
6. Luthardt M., Than E., Heckendorff H.: Fresenius Z. Anal. Chem. *326*, 331 (1987).
7. Vogelsang J.: Fresenius Z. Anal. Chem. *328*, 213 (1987).
8. Long G. G., Winefordner J. D.: Anal. Chem. *55*, 712A (1983).
9. Beyermann K.: *Organische Spurenanalyse*, p. 15. Thieme, Stuttgart 1982.
10. Eckschlager K., Štěpánek V.: *Information Theory as Applied to Chemical Analysis.* Wiley, New York 1979.
11. Eckschlager K., Štěpánek V.: *Analytical Measurement and Information.* Research Studies Press, Letchworth 1985.
12. Danzer K., Eckschlager K., Wienke D.: Fresenius Z. Anal. Chem. *327*, 312 (1987).
13. Eckschlager K.: Collect. Czech. Chem. Commun. *34*, 1321 (1969).
14. Eckschlager K., Štěpánek V.: Mikrochim. Acta *1978*, I, 107.
15. Eckschlager K., Štěpánek V.: Mikrochim. Acta *1981*, II, 143.
16. Cleij P., Dijkstra A.: Fresenius Z. Anal. Chem. *298*, 97 (1979).
17. Liteanu C., Rica I.: Anal. Chem. *51*, 1986 (1979).
18. Eckschlager K., Král M.: Collect. Czech. Chem. Commun. *49*, 2342 (1984).

Translated by P. Adámek.