Estimation of the Information Content in Ultraviolet−Visible Molecular Absorption Spectrometry

Victor David

Department of Analytical Chemistry, Faculty of Chemistry, University of Bucharest, Sos. Panduri, No. 90, 76.235, Section 5, Bucharest, Romania

The information content obtained from UV−vis molecular absorption spectrometry can be estimated as being dependent on the spectral domain size and the standard deviation of the spectral measurements. The information content can be expressed with Shannon's general formula, taking into account the uncertainty that is solved using an UV−vis measurement. It is shown that this information content as a function of the dependence on the standard deviation reaches a maximum value that could be used to optimize the sensitivity of a spectral measurement process.

1. INTRODUCTION

The foundations of the information theory are due to Shannon.[1] The first applications of probability theory and, later, of information theory in analytical chemistry are due to Kaiser and were related to the spectral analytical methods,[2,3] and after that, Soltzberg and co-workers used this theory to evaluate the pattern classifiers for chemical pattern recognition.[4,5] The application of this theory in analytical chemistry is mainly a consequence of the similarity between an analytical process and a data transmission process from a source to a receptor.[6] Eckschlager's works are among the most useful contributions in this domain.[7]

Even if the problem of the estimation of information content has been widely discussed in the literature, in only a few cases has it been applied to certain analytical processes. Thus, it has been established that this theory can be successfully applied to estimate the information content obtained from the gas chromatography−mass spectrometry technique (GC−MS), taking into account the experimental similarity indexes to estimate the values for the a posteriori probability assigned to each compound.[8] On the other hand, some other information quantities (gain, profitability) enable us to compare the relative suitability and effectiveness of various analytical methods and procedures for different applications.[9]

UV−Vis molecular absorption spectrometry has only a few applications in qualitative analysis, but nevertheless, it is still widely used in quantitative analysis, mainly as a detection method in high-performance liquid chromatography. For this reason, it could be desirable to establish the information content that can be obtained from such a determination. In this paper, the problem of estimating the information content of an UV−vis spectral measurement will be discussed, giving a new point of view on this application. This contribution to the topic can be used to optimize, from the information point of view, the outcome of a spectral measurement, to obtain maximum information from UV−vis absorption spectrometry.

2. THEORETICAL CONSIDERATIONS

The real value of an UV−vis absorbance, denoted by A_r, measured at a certain wavelength λ is situated within the interval given below, with a probability (p_A) given by Gauss' law:

$$\bar{A} \pm n\sigma_A \tag{1}$$

where $\bar{A}$ is the main value for m repeated measurements, having a standard deviation σ_A, and n is a parameter which gives this interval in units of standard deviation and it is correlated with p_A.

This interval has a magnitude denoted by ΔA and is

$$\Delta A = 2n\sigma_A \tag{2}$$

The probability mentioned above can be computed from Gauss' law:

$$p_A = \frac{1}{\sigma\sqrt{2\pi}} \int_{\bar{A}-n\sigma_A}^{\bar{A}+n\sigma_A} \exp\left\{-\frac{(A-\bar{A})^2}{2\sigma_A{}^2}\right\} dA \tag{3}$$

After performing the variable change

$$Z = \frac{A - \bar{A}}{\sigma_A} \tag{4}$$

then p_A becomes

$$p_A = 2\Phi(n) \tag{5}$$

where $\Phi(n)$ is Laplace's function, which is a known parameter.

An UV−vis spectral domain of absorbances measured at a fixed wavelength, $[A_{min}, A_{max}]$, can be divided in a number of interval s, using as a unit the value ΔA described above

$$s = \frac{A_{max} - A_{min}}{2n\sigma_A} \tag{6}$$

Table 1. Values of Information Content Given by Formula 8 for $A_{\min} = 0.0$, $A_{\max} = 1.0$, and Different Values of the Parameters n, σ, and $\phi(n)$

		ΔH for						
n	$\Phi(n)$	$\sigma = 0.0001$	$\sigma = 0.0005$	$\sigma = 0.0010$	$\sigma = 0.0015$	$\sigma = 0.0025$	$\sigma = 0.0050$	$\sigma = 0.0100$
0.25	0.0987	3.282	2.824	2.627	2.511	2.366	2.168	1.971
0.50	0.1915	5.619	4.730	4.347	4.123	3.841	3.458	3.075
0.75	0.2734	7.422	6.152	5.606	5.286	4.883	4.336	3.789
1.0	0.3413	8.763	7.179	6.496	6.097	5.594	4.911	4.288
1.25	0.3944	9.708	7.877	7.088	6.627	6.045	5.257	4.468
1.50	0.4322	10.318	8.307	7.440	6.933	6.295	5.428	4.562^a
1.75	0.4599	10.671	8.535	7.615	7.077	6.399	5.479^a	4.560
2.0	0.4772	10.837	8.621^a	7.666^a	7.108^a	6.405^a	5.451	4.496
2.25	0.4878	10.881^a	8.616	7.640	7.070	6.351	5.375	4.400
2.50	0.4938	10.847	8.554	7.567	6.989	6.261	5.274	4.286
2.75	0.4970	10.772	8.464	7.470	6.888	6.156	5.162	4.168
3.00	0.4987	10.678	8.363	7.365	6.782	6.047	5.049	4.052
4.00	0.5000	10.288	7.966	6.966	6.381	5.644	4.644	3.644

^{a} Maximum of ΔH in relationship to the parameter n.

The probability as A_r to be situated in one of these intervals is given by

$$p_s = \frac{1}{s} = \frac{2n\sigma_A}{A_{\max} - A_{\min}} \quad (7)$$

Known information concepts, such as a priori and a posteriori probabilities, together with their corresponding information entropies, found in a general expression for information content ΔH, can be avoided by using the uncertainty that is solved by an analytical process, and in our case, it is given by the expression of p_A above.

The information content (ΔH_λ) obtained from an UV-vis determination at a constant value of wavelength λ becomes

$$\Delta H_\lambda = -\sum_{i=1}^{s} p_i \log_2 p_i = 2\Phi(n) \log_2 \frac{A_{\max} - A_{\min}}{4n\sigma_A\Phi(n)} \quad (8)$$

Expression 8 depends on three variables: n, σ_A, and the difference $A_{\max} - A_{\min}$. This is a logarithmic function with respect to the variables σ_A and $A_{\max} - A_{\min}$ and a convex function with respect to n, such that an optimum value for n can be estimated in order to obtain a maximum value for ΔH.

If the measurements are made at several wavelengths λ_k ($k = 1, 2, ..., m$), then the information amount (ΔH_m) will be the sum of m single-wavelength measurements:

$$\Delta H_m = \sum_{k=1}^{m} \Delta H(\lambda_k) \quad (9)$$

Using the same procedure described above for computing the information content (ΔH) and the statement that the probability of occurring one event between two independent events is a product between their probabilities, then the information content (ΔHUV-vis) obtained from a spectral measurement at m wavelengths, lying between $\lambda_{\max}$ and $\lambda_{\min}$, is given by

$$\Delta H_{\text{UV-vis}} = 4\Phi(n)\Phi(m) \log_2 \frac{(A_{\max} - A_{\min})(\lambda_{\max} - \lambda_{\min})}{16nm\sigma_\lambda\sigma_A\Phi(n)\Phi(m)} \quad (10)$$

In this formula, only σ_A and σ_λ can depend on λ, while the other parameters are constant over a spectral domain used in UV-vis spectrometry. In this study, we will take into account a homoscedastic standard deviation for the absorbance or wavelength measurements; i.e., they do not depend on the parameter λ.

3. APPLICATIONS AND RESULTS

The information content obtained from an UV-vis measurement, at $\lambda = $ constant, can be easily computed by means of formula 8; the values of $A_{\min}$ and $A_{\max}$ are usually known, and even if in many analytical experiments they are situated within the interval $[-4, +4]$, we will consider in the following calculations only the interval which is obeyed to Lambert-Beer's law, i.e., between 0 and 1. The values of information content are given in Table 1.

As can be observed from Table 1, the information content obtained from the spectral experiment has a maximum value for a constant value of the standard deviation. In this way, we can optimize the analytical results obtained from an UV-vis measurement, such as to maximize information.

One of the most important parameters of an analytical process is the sensitivity of the measurements, denoted by S_i, which is defined as a change in the response of the device output to a change in the concentration of the measured component in the input.[10] The optimum value of the sensitivity (S_{opt}) corresponds to a maximum value of information content and, according to the previous theoretical model, is given by $S_{\text{opt}} = n\sigma_A$, such that $\Delta H = $ maximum. For $S < n\sigma_A$, a large number of spectral intervals s would be obtained but having a low probability of representing the real value, while for $S > n\sigma_A$ a small number of spectral intervals would be obtained but having a high probability of representing the real value. Of course, between these two extremes, an optimum value could be obtained, and for that, we need the maximum of information content depending on σ_A. On the other hand, according to the values of information given in Table 1, the maximum of ΔH will shift to the smaller values of parameter n, with an increase of the standard deviation value.

As has been already shown[11] for a general quantitative analysis, the sensitivity of an UV-vis spectral measurement can be optimized, such as the analytical results, to contain a maximum information content. This is illustrated in Table

Table 2. Sets of Absorbance Values (*A*) Containing Maximum Information, Dependent on the Parameter σ from Table 1

ΔH=max						
$n = 2.25$	$n = 2$	$n = 2$	$n = 2$	$n = 2$	$n = 1.75$	$n = 1.5$
0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0002	0.00010	0.0020	0.0030	0.0050	0.0100	0.0150
0.0004	0.00020	0.0040	0.0060	0.0100	0.0200	0.0300
0.0006	0.00030	0.0060	0.0090	0.0150	0.0300	0.0450

2, for UV−vis absorbances, where their values that provide a maximum information content are given for several values of the standard deviation.

One final remark is given by the assumption that a standard deviation can be always estimated a priori to a certain measurement. The information content obtained in UV−vis absorption spectrometry does not depend on the nature of the sample being analyzed or on the concentrations of the analytes, the only condition being that this concentration be situated within the interval where Lambert−Beer's law is fulfilled.

4. CONCLUSIONS

The information content obtained from an UV−vis spectral measurement can be computed by using the general concepts from information theory. In this study, we propose a mathematical method for computing the information content resulting from UV−vis absorption spectrometry. The most important parameter in this estimation is the standard deviation of the measurements, which must be known a priori to a certain measurement. On the other hand, the maximum information content is influenced by the value of the standard deviation, a property that could be used in choosing the best analytical results in UV−vis absorption spectrometry.

REFERENCES AND NOTES

(1) Shannon, C. E. *Bell Syst. Technol. J.* **1948**, *27*, 623.
(2) Kaiser, H. *Anal. Chem.* **1970**, *42* (24A), no. 2.
(3) Kaiser, H. *Anal. Chem.* **1970**, *42* (26A), no. 4.
(4) Soltzberg, L. J.; Wilkins, C. L.; Kaberline, S. L.; Lam, T. F.; Brunner, T. R. *J. Am. Chem. Soc.* **1976**, *98*, 7139.
(5) Soltzberg, L. J.; Wilkins, C. L.; Kaberline, S. L.; Lam, T. F.; Brunner, T. R. *J. Am. Chem. Soc.* **1976**, *98*, 7144.
(6) David, V. *Rev. Chim. (Bucharest)* **1989**, *40*, 428.
(7) Eckschlager, K.; Stepanek, V. *Information Theory as Applied to Chemical Analysis*; John Wiley & Sons: New York, 1979.
(8) David, V.; Medvedovici, A.; Tache, F. *Ann. Chim. (Roma)* **1998**, *88*, 577.
(9) Obrusnik, I.; Eckschlager, K. *Anal. Chem.* **1990**, *62,* 565.
(10) Liteanu, C.; Rica, I. *Statistical Theory and Methodology of Trace Analysis*; Ellis Horwood: Chichester, 1980.
(11) David, V. *Rev. Roum. Chim.* **1987**, *32*, 539.

CI9801320