Some Problems of Regression Analysis in the Spectrodissymmetry

[a]M. LIŠKA, [b]J. ANTALÍK, and [b]J. VAVRA

[a]Institute of Inorganic Chemistry, Slovak Academy of Sciences, CS-842 36 Bratislava

[b]Department of Physical Chemistry, Faculty of Chemical Technology, Slovak Technical University, CS-812 37 Bratislava

The possibilities of application of spectral dependence of light scattering dissymmetry (Z) in distributional analysis of polydispersions with spherical particles are described. The calculation of the best estimates of distribution function parameters and the estimate of standard deviations were performed numerically using the Simplex method. For theoretical computations of light scattering characteristics the exact Lorenz—Mie theory was applied.

The algorithm of the numerical search of distribution parameters was tested on the experimental values of Z of polystyrene latexes from literature. The best estimates of the median and breadth parameters determined were in very good harmony with the results from literature.

In our previous papers [1—3] the possibility of application of the light scattering dissymmetry (Z) by determination of size or distribution of sizes of latex spherical particles on the basis of comparison of the theoretical and experimental values of Z was pointed out. The aim of this paper is to study the application of optimization methods by distributional analysis of polydispersions on the basis of spectral dependence of Z.

THEORETICAL

In the theory of light scattering the coefficient of dissymmetry Z is defined as the ratio of two light fluxes Φ scattered at two angles of observation Θ

$$Z(\Theta_1, \Theta_2) = \frac{\Phi(\Theta_1)}{\Phi(\Theta_2)} \quad (1)$$

As the light scattering is a function of size and form of the scattering particle, the dissymmetry of light scattering may serve as a sensitive optical quantity which characterizes the scattering system.

In accordance with the exact Lorenz—Mie theory of light scattering the theoretical values of Z for polydisperse system of nonabsorbing spheres are given by

$$Z_n^{calc}(r_M, \sigma, \Theta_1, \Theta_2, m, \lambda) =$$

$$= \frac{\int_0^\infty f(r; r_M, \sigma)\, i_n(r; \Theta_1, m, \lambda)\, dr}{\int_0^\infty f(r; r_M, \sigma)\, i_n(r; \Theta_2, m, \lambda)\, dr} \quad (2)$$

where $f(r; r_M, \sigma)$ is normalized biparametric size distribution function with parameters r_M, σ and $i_n(r; \Theta, m, \lambda)$ are Lorenz—Mie intensity functions, which

are dependent on the scattering angle Θ, refractive index of the dispersed particles relative to that of the dispersion medium m, radius of particle r, and wavelength of the radiation in the system λ.

As m is characteristic of disperse system and the wavelength can be chosen according to available photometer, Z can be evaluated from eqn (2) when the size distribution function is known.

From the practical point of view we are concerned with the inverse problem of determining the particle size distribution from scattering data. As the direct inversion of eqn (2) is not possible, the distribution parameters are searched by numerical optimization process.

The Lorenz—Mie intensity function for nonpolarized radiation i_n is evaluated as an average value of intensity function for vertically and horizontally polarized primary beam

$$i_n(r; \Theta, m, \lambda) = \frac{1}{2}\left[i_1(r; \Theta, m, \lambda) + i_2(r; \Theta, m, \lambda)\right] \quad (3)$$

Hereby for vertical component i_1 and horizontal component i_2 it is valid

$$i_1(r; \Theta, m, \lambda) = \left|\sum_{n=1}^\infty \frac{2n+1}{n(n+1)}(a_n\pi_n + b_n\tau_n)\right|^2 \quad (4)$$

$$i_2(r; \Theta, m, \lambda) = \left|\sum_{n=1}^\infty \frac{2n+1}{n(n+1)}(a_n\tau_n + b_n\pi_n)\right|^2 \quad (5)$$

where π_n and τ_n are angular coefficients expressed in terms of the Legendre polynomials and a_n, b_n are coefficients defined by means of the Bessel functions. The number of the terms in series (4) and (5) can be determined by means of the following em-

pirical equation [4]

$$n_{max} = 1.84 \, \alpha^{0.904} + 4 \qquad (6)$$

where α is the parameter of size ($\alpha = 2\pi r/\lambda$). Details of numerical computations can be found in the literature [5, 6].

From experimental determination the values of dissymmetry $Z_{n,i}^{exp}$ for $\Theta_{1,i}$, $\Theta_{2,i}$, λ_i for $i = 1, 2, ... N$ could be obtained (N is the number of experimental values of Z). If we assume that the values $Z_{n,i}^{exp}$ are determined with a constant variance, the best estimate of r_M and σ can be obtained by minimization of an objective function in the form of nonweighted sum of squares of deviations between the theoretical and experimental values of Z according to the relation

$$U(r_M, \sigma) = \sum_{i=1}^{N} \Big[Z_n^{calc}(r_M, \sigma, \Theta_{1,i}, \Theta_{2,i}, m, \lambda_i) -$$
$$- Z_{n,i}^{exp} \Big]^2 = min \qquad (7)$$

If it can be assumed that the values $Z_{n,i}^{exp}$ are loaded with identical relative error, then the expression (7) can be written in the form

$$U(r_M, \sigma) = \sum_{i=1}^{N} \Big[\big(Z_n^{calc}(r_M, \sigma, \Theta_{1,i}, \Theta_{2,i}, m, \lambda_i) -$$
$$- Z_{n,i}^{exp} \big) / Z_{n,i}^{exp} \Big]^2 = min \qquad (8)$$

where the values $(Z_{n,i}^{exp})^{-2}$ have the meaning of weight coefficients.

Owing to the fact that U is a function only of two variables — r_M and σ ($\Theta_{1,i}$, $\Theta_{2,i}$, m, λ_i are known), minimization can be made with an advantage on the basis of some nongradient method (*e.g.* SIMPLEX method), or a gradient method with numerical computation of the gradient.

The estimate of standard deviations of distribution parameters r_M and σ can be obtained, if the course of the function $U(r_M, \sigma)$ in the vicinity of its minimum U_{min} is substituted with the quadratic form [7]

$$\Delta U(r_M, \sigma) = U(r_M, \sigma) - U_{min} = X^T A X \qquad (9)$$

where **A** is the information matrix of Hess, **X** is the column vector of deviations of r_M and σ from their optimal values r_M^* and σ^*; then X^T (the row vector to vector **X**) can be expressed in the form

$$X^T = \{X_1, X_2\} = \{r_M - r_M^*, \sigma - \sigma^*\} \qquad (10)$$

Hereby $U_{min} = U(r_M^*, \sigma^*)$.

Diagonal elements A_{kk} of matrix **A** can be numerically estimated after the choice of elements of the vector **X** in the form $X_i = \pm \delta_{ik} d_k$, where $i = 1, 2$ and

$d_k > 0$ is the adopted step, with which adequate parameter is changed.

The value of diagonal elements can be then computed as the average of ΔU for positive and negative shift of parameters r_M and σ according to the relations

$$A_{11} = \frac{1}{2} \Big[\Delta U(r_M^* - d_1, \sigma^*) + \Delta U(r_M^* + d_1, \sigma^*) \Big] / d_1^2 \quad (11)$$

$$A_{22} = \frac{1}{2} \Big[\Delta U(r_M^*, \sigma^* - d_2) + \Delta U(r_M^*, \sigma^* + d_2) \Big] / d_2^2 \quad (12)$$

Nondiagonal elements of matrix **A** can be calculated as arithmetic mean of ΔU for combination of shifts of parameters r_M and σ in positive and negative direction in the form

$$A_{21} = A_{12} = \frac{1}{8} \Big[\Delta U(r_M^* + d_1, \sigma^* + d_2) + \Delta U(r_M^* -$$
$$- d_1, \sigma^* - d_2) - \Delta U(r_M^* + d_1, \sigma^* - d_2) -$$
$$- \Delta U(r_M^* - d_1, \sigma^* + d_2) \Big] / (d_1 d_2) \qquad (13)$$

The values of standard deviation and covariance of parameters can be evaluated from elements of the inverse matrix $(A^{-1})_{ij}$ in the form

$$s(r_M) = [(A^{-1})_{11} U_{min}/(N - 2)]^{1/2} \qquad (14)$$
$$s(\sigma) = [(A^{-1})_{22} U_{min}/(N - 2)]^{1/2} \qquad (15)$$
$$cov(r_M, \sigma) = (A^{-1})_{12} U_{min}/(N - 2) \qquad (16)$$

The correlation coefficient R can be determined from the relation

$$R = \frac{cov(r_M, \sigma)}{s(r_M) \, s(\sigma)} \qquad (17)$$

Presented procedure is acceptable from the point of view of applied approximation of U, if relations $d_1 \doteq s(r_M)$ and $d_2 \doteq s(\sigma)$ are fulfilled. Therefore the calculation is repeated cyclically, so that in the next iteration the steps d_1 and d_2 are substituted by the values of $s(r_M)$ and $s(\sigma)$ from the foregoing iteration. By a reasonable choice of starting estimate of d_1 and d_2 the first two iterations are usually sufficient.

RESULTS AND DISCUSSION

Described method of distributional analysis was tested on the basis of experimental values of dissymmetry and mean particle sizes of polystyrene (PS) latexes published by *Maron et al.* [8]. From the cited paper the values of $Z_{n,i}^{exp}(\Theta_{1,i} = 45°, \Theta_{2,i} = 135°, m, \lambda_i)$ for seven PS latexes with mean particle radius 30—107 nm were taken. The value of $m =$

1.200 for these latexes was adopted as constant for the wavelengths used.

The computation of the best estimate for parameters of particle size distribution was made numerically as described according to the original program in Fortran.

Owing to the frequent applicability [1—4], it was assumed that the distribution function of particle sizes could be expressed as a log-normal distribution of negative order (NOLD) in the form

$$f(r; r_M, \sigma) = \frac{1}{(2\pi)^{1/2}\sigma r} \exp\left[-\frac{1}{2\sigma^2} \ln^2 \frac{r}{r_M}\right] \quad (18)$$

where $\ln r_M = \int_{-\infty}^{\infty} \ln r\, f(r)\, dr$ is the mean value of $\ln r$ and r_M is, in this case, both the median and the geometric mean value r_g, and

$$\sigma = \left[\int_{-\infty}^{\infty} (\ln r - \ln r_M)^2\, f(r)\, dr\right]^{1/2} \quad (19)$$

is the standard deviation of $\ln r$ and is termed the geometric mean standard deviation.

It should be noted that the architecture of program facilitates inclusion of other types of distributions by modification of only one subroutine.

Minimization of sums in expressions (7) and (8) was carried out by SIMPLEX method according to Nelder and Mead [9]. The integrals in eqn (2) for determination of Z were solved by the numerical method. The acceleration of computations was achieved by means of disc file of the values $i_{1,2}(r; \Theta, m, \lambda)$ tabulated with constant step $\Delta\alpha = 0.01$.

Selected experimental values of Z at corresponding wavelengths for PS latexes are summarized in Table 1. This table includes also the values of $Z_{n,i}^{calc}$, obtained by minimization of sums in eqn (8), together with values of relative deviations $\Delta_{rel} Z$

$$\Delta_{rel}Z = \frac{Z_{n,i}^{calc} - Z_{n,i}^{exp}}{Z_{n,i}^{exp}} \quad (20)$$

The relative deviations obtained are in very good agreement with experimental error of Z determination.

Table 1. Comparison of Experimental [8] and Theoretical Values of Z

Latex	λ_i/nm	$Z_{n,i}^{exp}(45°/135°)$	$Z_{n,i}^{calc}(45°/135°)$	$\Delta_{rel}Z/\%$
1	271.1	41.70	40.00	− 4.1
	325.3	166.7	166.7	0.0
	409.4	10.00	9.52	− 5.0
2	325.3	1.24	1.24	0.0
	409.4	1.14	1.14	0.0
	433.6	1.13	1.13	0.0
3	271.1	1.62	1.61	− 0.6
	325.3	1.37	1.39	1.5
	409.4	1.23	1.23	0.0
4	271.1	3.56	3.68	3.4
	325.3	2.34	2.35	0.4
	409.4	1.71	1.69	− 1.2
	433.6	1.59	1.59	0.0
	519.3	1.38	1.38	0.0
5	271.1	7.58	7.63	0.7
	325.3	3.57	3.55	− 0.6
	409.4	2.09	2.11	0.9
6	271.1	33.30	33.30	0.0
	325.3	9.01	8.85	− 1.8
	409.4	3.33	3.41	2.4
7	325.3	13.50	13.50	0.0
	409.4	4.37	4.37	0.0

Table 2 summarizes the values of parameters r_M and σ of NOLD for PS latexes under study together with values of standard deviations $s(r_M)$ and $s(\sigma)$ and the values of Student t-statistics.

From comparison of the values of r_M with adequate experimental values $\bar{r}_{exp}$ we can see in all cases only small difference, hereby $r_M > \bar{r}_{exp}$ is always valid in connection with asymmetry of distribution function assumed.

The values of $t(r_M)$ point out that in all cases with nonzero degree of freedom the statistical significance of r_M is more than 97.5 %.

The values of the breadth parameter σ in harmony with paper [8] represent only some percent from the value $\bar{r}_{exp}$. Statistical significance of σ for latex No. 4 is higher than 99 %, but for latexes No. 1, 5, 6 it is 90 %. It is evident that the statistical method applied in the cases with one degree of freedom ($N = 3$) has only orientational character. But practical application of our method of distributional analysis enables to increase the degrees of freedom without problem: the measurements can be performed at

Table 2. The Calculated Best Estimates of Parameters of NOLD for PS Latexes and Their Statistical Characteristics

Latex	$\bar{r}_{exp}$/nm	N	r_M/nm	$s(r_M)$/nm	$t(r_M)$	σ	$s(\sigma)$	$t(\sigma)$
1	107.0	3	106.47	0.59	180.45	0.042	0.006	7.0
2	30.4	3	28.21	0.81	34.83	0.106	0.019	5.6
3	37.4	3	36.04	1.21	29.79	0.077	0.017	4.5
4	59.0	5	54.38	1.25	43.50	0.114	0.017	6.7
5	69.0	3	66.05	0.79	83.61	0.093	0.013	7.1
6	83.5	3	79.57	0.27	294.70	0.119	0.005	23.8
7	93.0	2	83.67	–	–	0.138	–	–

two wavelengths, at three angles or at different polarization of primary beam.

REFERENCES

1. Vavra, J. and Antalík, J., *Chem. Zvesti 37*, 33 (1983).
2. Vavra, J. and Antalík, J., *Chem. Papers 44*, 769 (1990).
3. Vavra, J. and Antalík, J., *Colloid Polym. Sci. 269*, 1060 (1991).
4. Dautzenberg, H., Hartman, M., and Koch, H., *Acta Polym. 30*, 290 (1979).
5. Kerker, M., *The Scattering of Light and Other Electromagnetic Radiation*. Academic Press, New York, 1969.
6. Bohren, C. F. and Huffman, O. R., *Absorption and Scattering of Light by Small Particles*. (Russian translation.) Mir, Moscow, 1986.
7. Bard, Y., *Nonlinear Parameter Estimation*. Academic Press, New York, 1974.
8. Maron, S. H., Pierce, P. E., and Elder, M. E., *J. Macromol. Sci., Phys. B 1*, 29 (1967).
9. Nelder, J. A. and Mead, R., *Comput. J. 7*, 308 (1965).

Isobaric Binary Equilibrium Data for the System Water—Benzothiazole at 101.3 kPa

[a]E. GRACZOVÁ, [a]J. SUROVÝ, [b]J. KOZMA, and [a]J. OVEČKOVÁ

[a]*Department of Chemical and Biochemical Engineering, Faculty of Chemical Technology, Slovak Technical University, CS-812 37 Bratislava*

[b]*Research Institute of Chemical Technology, CS-836 03 Bratislava*

A set of isobaric vapour-liquid equilibrium data for the water—benzothiazole system at 101.325 kPa is reported. These data were calculated on the assumption of ideal behaviour of the vapour phase and real behaviour of the liquid. In this work, activity coefficients were calculated using the van Laar equation the parameters of which were determined from experimental solubilities. On using the solubility data, both temperature and composition of the heteroazeotropic mixture at the pressure of 101.325 kPa were also obtained.

The binary system water(1)—benzothiazole(2) exhibits a limited miscibility; the solubility of components is low (the mass fraction has an order of magnitude 10^{-3}). Data on physical properties of benzothiazole are, as yet, not very plentiful in the literature [1, 2]. A survey of the literature indicates that up to the present practically no data have been reported for binary equilibrium for this system.

A detailed description of the method used for the determination of vapour-liquid (V-L) equilibrium is given in our previous work [3]. Solubilities were determined experimentally by the turbidity method. The saturated vapour pressure of benzothiazole was measured by the static method [4, 5]. The so obtained set of data enabled to calculate the composition of the heteroazeotropic mixture for the water—benzothiazole system and θ—x—y equilibrium data at 101.325 kPa.

EXPERIMENTAL

Material used was research grade 99.9 % benzothiazole supplied by VUCHT (Research Institute of Chemical Technology, Bratislava) with refractive index n(D, 20 °C) = 1.6425 identical with the data point of literature [1], water was bidistilled.

Solubilities of the system water—benzothiazole were determined by the turbidity method. Heterogeneous binary samples of known composition were prepared. These samples in sealed glass ampoules were thoroughly stirred and heated in a water or glycerol bath up to the temperature at which one homogeneous liquid was formed. By cooling, the temperature was found at which heterogeneous phases appeared. The temperature of the water bath was determined with an accuracy of ± 0.1 K. The samples were weighed on an analytical balance with