STATISTICAL TREATMENT OF MATERIAL BALANCE OF SEPARATION OPERATIONS

František MADRON

Research Institute of Inorganic Chemistry,
400 60 Ústí nad Labem

Received April 10th, 1978

A method of complex statistical treatment of material balances of separation operations without chemical reactions has been proposed. The method which is suitable for checking of correctness of material balance of separation operations in the laboratory and industrial size is simple and easy for programming on digital computer. The computational procedure is illustrated on an example of material balance of an absorption unit.

In experimental determination of material balances of separation operations a case can occur where a greater number of quantities has been measured than is necessary for formulation of a material balance. In such cases it is mostly not possible, due to experimental errors, to complete exactly the material balance. The redundant information included in the experimental measured data can be used for verification of correctness of the performed material balance (determination of gross and systematic errors of meassurement) and for increasing the accuracy of experimental values.

In this study which is based on the earlier publications on statistic treatment of the material balance of chemical reactors[1-3] a simple method of statistic evaluation of the material balance of separation operations without chemical reactions is proposed.

THEORETICAL

Let us consider the separation unit which is demonstrated in Fig. 1. A mixture of compounds which should be separated in the unit (stream *1*) is in contact with the stream *4*, mass transfer takes place and two streams *2* and *3* originate whose composition differs from those of streams *1* and *4*. By this simple diagram can be demonstrated *e.g.* absorption and extraction and if the flow rate of stream *4* equals to zero, the simple continuous distillation as well.

Let us limit ourselves, for the sake of simplicity, to cases when stream *4* does not include separated compounds and let us assume that the equipment is operated under steady state. We denote the molar flow rate of the i-th component in the k-th stream as n_{ik} (positive in the case of outlet stream and negative at the inlet stream). The function of separation unit can be then characterized by the separation degree s_i

$$s_i = -n_{i2}/n_{i1}\ ; \quad i = 1, 2, \ldots, I\ . \qquad (1)$$

If the quantities n_{ik} were measured for 1, 2, ..., I and $k = 1, 2, 3$ the experimental separation degrees can be determined from two relations – on basis of streams *1* and *2* from relation

$$s^*_{i2} = -n^*_{i2}/n^*_{i1}\ ; \qquad (2)$$

and of streams *1* and *3* from relation

$$s^*_{i3} = 1 + \left(n^*_{i3}/n^*_{i1}\right)\ ; \quad i = 1, 2, \ldots, I\ , \qquad (3)$$

where the asteriscs denote the quantities determined by measurements.

Though the values of separation degrees computed from Eqs (*2*) and (*3*) should be equal (there obviously holds $- n_{i1} = n_{i2} + n_{i3}$) due to experimental and other errors these values will probably differ. From the magnitude of deviations of these quantities it is possible to judge whether in the material balance there act only small random errors of measurements, whose range is approximately known or whether there appear gross errors of measurements, losses of compounds, errors in formulation of material balance equations *etc.* Further, the procedure for statistical analysis of material balance of separation unit illustrated in Fig. 1 is proposed. In absence of

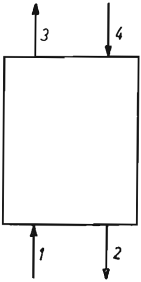

FIG. 1
Separation Unit

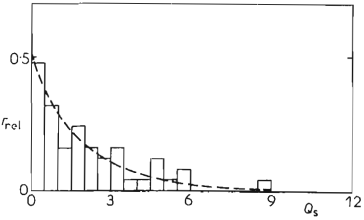

FIG. 2
Histogram of Quantities Q_s Out of 50 Simulated Experiments
----- Probability density of $\chi^2(2)$ distribution.

rough errors the statistically smoothed values are obtained of separation degrees or of individual quantities n_{ik}.

Stochastic Model

The simplified stochastic model of the separation unit can be written in the form of equation

$$s_{ik}^* = s_i + e_{ik}\,; \quad i = 1, \ldots, I\,; \quad k = 2, 3\,, \tag{4}$$

where s_{ik}^* are the quantities defined by Eqs (*2*) and (*3*), s_i are the actual (unknown) degrees of separation and e_{ik} are errors corresponding to quantities s_{ik}^*.

If we form the column vector

$$s^* = (s_{12}^*, s_{13}^*, s_{22}^*, s_{23}^*, \ldots, s_{i2}^*, s_{i3}^*, \ldots, s_{I2}^*, s_{I3}^*)^{\mathrm{T}}$$

with the number of elements $2I$ and a similarly formed column vector $\mathbf{e}_s$ formed by errors e_{ik} we can write Eq. (*4*) in the matrix form

$$\mathbf{s}^* = \boldsymbol{A} \,.\, s + e_s\,, \tag{5}$$

where $\mathbf{s}$ is the column vector of actual values of absorption degrees (number of elements I) and $\boldsymbol{A}$ the matrix of the dimension $(2I \times I)$ whose elements are given by the relations

$$a_{ij} = 1 \quad \text{for} \quad i = 2j \quad \text{or} \quad i = 2j - 1 \tag{6}$$

$$a_{ij} = 0 \quad \text{in other case} \tag{7}$$

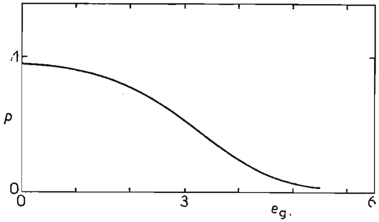

FIG. 3

Error of II-nd Kind in Dependence on the Magnitude of Positive Gross Error (e_g, σ) of Quantity x_{11}^*

Let us assume that the errors e_{ik} are random variable with normal distribution with the zero mean value and covariance matrix $\mathbf{W} = E\{\mathbf{e}_s . \mathbf{e}_s^T\}$.

The statistical estimate of degrees of separation on the principle of maximum likelihood is then given by the relation[4]

$$\hat{\mathbf{s}} = (\mathbf{A}^T . \mathbf{W}^{-1} . \mathbf{A})^{-1} . \mathbf{A}^T . \mathbf{W}^{-1} . \mathbf{s}^* \tag{8}$$

at the assumption that matrix $\mathbf{W}$ is regular.

The estimate of errors e_{ik} are the residuals $\hat{e}_{ik} = s_{ik}^* - s_i$. The measure of adequacy of the mathematical model (*5*) for description of separation operations is the quadratic form

$$Q_s = \hat{\mathbf{e}}_s^T . \mathbf{W}^{-1} . \hat{\mathbf{e}}_s , \tag{9}$$

which is a random variable with χ^2 distribution[4] with I degrees of freedom $\chi^2(I)$. In the case when model (*5*) is adequate, the quantity Q_s will occur with the probability $1 - \alpha$ in the interval $\langle 0; \chi^2_{1-\alpha}(I)\rangle$, where $\chi^2_{1-\alpha}(I)$ is the critical value of $\chi^2(I)$ distribution for the level of significance $1 - \alpha$. In the opposite case it is possible to assume that the model (*5*) is not correct (there appear gross errors *etc.*).

Statistical evaluation of the material balance based on computation of the separation degrees according to Eqs (*2*) and (*3*) is simple but it does not enable statistical smoothing of values $\hat{n}_{ik}$. In the following part an alternative method is given which is directly based on meassured values n_{ik}^* which does not have this disadvantage.

If we denote the separated components as d_i, the separation can be schematically written as

$$(d_i)_1 \rightarrow (d_i)_2 \quad i = 1, 2, \ldots, I \tag{10}$$

$$(d_i)_1 \rightarrow (d_i)_3 \tag{11}$$

where terms (*10*) and (*11*) are representing the transfer of the *i*-th component from the stream *1* to the stream *2* or *3*. The system of Eqs (*10*) and (*11*) is analogical to the system of chemical equations and, consequently, n_{ik} can be expressed by use of quantities r_{ik} which are an analogy of extents of chemical reactions[5]

$$\begin{aligned} n_{i1} &= -r_{i1} - r_{i2} \\ n_{i2} &= r_{i1} \qquad\qquad i = 1, 2, \ldots, I \\ n_{i3} &= r_{i2} , \end{aligned} \tag{12}$$

where the quantities n_{i1} are negative since the reactants of fictive chemical reactions (*10*) and (*11*) are concerned. If again the column vectors are formed $\mathbf{n} = (n_{11}, n_{12}, n_{13}, \ldots, n_{I1}, n_{I2}, n_{I3})^T$ and $r = (r_{11}, r_{12}, \ldots, r_{I1}, r_{I2})^T$ it is possible to write the system of Eqs (*12*) in the matrix form

$$\boldsymbol{n} = \boldsymbol{B} \,.\, \boldsymbol{r} \,, \tag{13}$$

where the elements of the matrix **B** of the dimension ($3I \times 2I$) are given by relations (*14*)

$$\begin{aligned} &b_{ij} = -1 \,, \quad \text{for} \quad 2i = 3j - 1 \quad \text{or} \quad 2i = 3j - 4 \\ &b_{ij} = 1 \quad \text{for} \quad 2i = 3j + 1 \quad \text{or} \quad 2i = 3j \\ &b_{ij} = 0 \text{ in the other cases.} \end{aligned} \tag{14}$$

After introduction of measured quantities and of their errors, the mathematica model of separation (Eq. *13*) can be written in the form

$$\boldsymbol{n}^* = \boldsymbol{B} \,.\, \boldsymbol{r} + \boldsymbol{e}_n \,, \tag{15}$$

where $\mathbf{e}_n$ is the column vector $(e_{11}, e_{12}, e_{13}, \ldots, e_{I1}, e_{I2}, e_{I3})^T$ of errors of measured quantities $\boldsymbol{n}^*$. Statistical treatment of the material balance is then identical to the previous case (Eqs (*8*) and (*9*)) and is given by the equation

$${}^{\wedge}\boldsymbol{r} = (\boldsymbol{B}^T \,.\, \boldsymbol{F}^{-1} \,.\, \boldsymbol{B})^{-1} \,.\, \boldsymbol{B}^T \,.\, \boldsymbol{F}^{-1} \,.\, \boldsymbol{n}^* \tag{16}$$

and after substitution into Eq. (*13*) the relation (*17*) is obtained

$${}^{\wedge}\mathbf{n} = \boldsymbol{B} \,.\, {}^{\wedge}\mathbf{r} \tag{17}$$

giving the smoothened values of quantities ${}^{\wedge}n_{ik}$. The hypothesis that the model (*15*) is adequate for description of the separation operation is tested by use of the quadratic form (*18*) which again has χ^2 distribution with I degrees of freedom.

$$Q_n = {}^{\wedge}\mathbf{e}_n^T \,.\, \boldsymbol{F}^{-1} \,.\, {}^{\wedge}\mathbf{e}_n \tag{18}$$

Example: Let us consider absorption of stack gases from industrial process. The compounds A and B in a mixture with inert compounds (stream No *1*) are introduced into the absorption column where they are contacted with the absorption liquid (stream No *4*). Two streams are leaving the absorber — the liquid one (2) and stack gases (stream 3).

Measured are the following rates of streams N_k and concentrations

$$x_{ik}\colon N_1^*, N_4^*, x_{11}^*, x_{21}^*, x_{12}^*, x_{22}^*, x_{13}^*, x_{23}^* \,,$$

where the compounds are numbered by A—1 and B—2.

The following equations for calculation of the quantity n^*_{ik} from the directly measured values N^*_k and x^*_{ik} are formed:

$$n^*_{ik} = N^*_k x^*_{ik} \quad i = 1, 2\,; \quad k = 1, 2, 3\,, \tag{19}$$

where the quantities N^*_2 and N^*_3 are not measured directly but are calculated from the next balance relations

$$N^*_3 = -N^*_1(1 - x^*_{11} - x^*_{21})/(1 - x^*_{13} - x^*_{23}) \tag{20}$$

$$N^*_2 = -N^*_4/(1 - x^*_{12} - x^*_{22})\,. \tag{21}$$

In Table I are given the directly measured quantities and their relative standard deviations (variation coefficients). It is possible to assume on basis of the character of measurement, that the directly measured quantities have a normal distribution and that they are statistically independent. The quantities n_{ik} and s_{ik} were calculated from relations (*19*)—(*21*) and (*2*) and (*3*) (Table II).

The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ had in this case the form

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{22}$$

The covariance matrices $\boldsymbol{W}$ and $\boldsymbol{F}$ were calculated after linearisation of Eqs (*2*), (*3*) and (*19*) by the method described *e.g.* in literature[6,7]. This method is based on expansion of Eqs (*2, 3, 19*) into the Taylor's series according to the directly measured quantities with the 1st order terms respected and with the following calculation of covariances of linear forms of directly measured random quantities. Details on this method can be found in the cited literature. We give here for illustration the covariance matrix of degrees of separation $\boldsymbol{W}$, which is

$$\boldsymbol{W} = \begin{bmatrix} 5{\cdot}92 \,.\, 10^{-4} & -8{\cdot}97 \,.\, 10^{-5} & 5{\cdot}29 \,.\, 10^{-4} & -8{\cdot}98 \,.\, 10^{-7} \\ -8{\cdot}97 \,.\, 10^{-5} & 9{\cdot}60 \,.\, 10^{-5} & -1{\cdot}58 \,.\, 10^{-6} & 8{\cdot}66 \,.\, 10^{-7} \\ 5{\cdot}29 \,.\, 10^{-4} & -1{\cdot}58 \,.\, 10^{-6} & 1{\cdot}50 \,.\, 10^{-3} & -8{\cdot}51 \,.\, 10^{-5} \\ -8{\cdot}98 \,.\, 10^{-7} & 8{\cdot}66 \,.\, 10^{-7} & -8{\cdot}52 \,.\, 10^{-5} & 2{\cdot}06 \,.\, 10^{-5} \end{bmatrix}$$

The statistically smoothed quantities s_i and n_{ik} calculated from relations (*8*) and (*17*) are given in Table III. The values of quadratic forms Q_s and Q_n calculated from relations (*9*) and (*18*) were 0·42 and 0·40. As the critical value of χ^2 (2) distribution for the level of significance 95% is 5·991 it is possible to assume that the material balance of the absorber is well formulated.

By linearisation of relations in the calculation of covariance matrices we make certain inaccuracy. For determination of this effect 50 experiments were simulated on the digital computer. To the values of quantities x_{ik} and N_k which satisfied exactly the material balance were added the random quantities with the normal distribution and zero mean value and standard deviations given in Table I. With so simulated experimental values the quantities Q_s were calculated. In Fig. 2 is given the histogram of these values together with theoretical χ^2 (2) distribution.

TABLE I

Measured Values x_{ik}^* (mole fraction), N_{ik}^* (mol . s^{-1}) and their Relative Standard Deviations γ

	x_{11}^*	x_{12}^*	x_{13}^*	x_{21}^*	x_{22}^*	x_{23}^*	N_1^*	N_4^*
	0·0316	0·0762	0·011	0·00566	0·0188	0·00062	—80·6	—19·4
γ	0·02	0·01	0·02	0·03	0·01	0·03	0·03	0·005

TABLE II

Values n_{ik}^* (mol . s^{-1}) and s_{ik}^* (—)

n_{11}^*	n_{12}^*	n_{13}^*	n_{21}^*	n_{22}^*	n_{23}^*	s_{12}^*	s_{13}^*	s_{22}^*	s_{23}^*
—2·55	1·64	0·863	—0·456	0·403	0·049	0·642	0·661	0·884	0·893

TABLE III

Statistically Smoothed Quantities $\hat{}n_i$ (mol . s^{-1}) and $\hat{}s_i$ (—)

$\hat{}n_{11}$	$\hat{}n_{12}$	$\hat{}n_{13}$	$\hat{}n_{21}$	$\hat{}n_{22}$	$\hat{}n_{23}$	$\hat{}s_1$	$\hat{}s_2$
—2·49	1·64	0·856	—0·452	0·404	0·048	0·657	0·893

Let us consider the errors which can be made at judging the correctness of the material balance. The probability that we reject the hypothesis on the correctness of formulation of the material balance, if this is correct, (probability of the Ist kind error) is obviously given by the level of significance and in this case it is 5%. The probability of error of IInd kind *i.e.* acceptance of the hypothesis in the case when this is wrong depends on the magnitude of the gross or systematic error which we have made at balancing. This problem cannot be solved in general. In Fig. 3 is given for illustration the probability of the IInd kind error p in dependence on the magnitude of positive gross error in determination of concentration x_{11} (the magnitude of gross error is given as the multiple of the standard deviation from Table I).

DISCUSSION

The proposed method of statistical treatment of the material balance of separation operations is simple and easy for programming on the computer. Beside the usual cases of the checking of correctness of formulation of the material balance in the laboratory or industrial scale it can be applied also in data equisition by the on-line computer.

In this study are given two computation procedures. The first one is based on determination of the degrees of separation based on two outlet streams, the second one is based on introduction of the fictive chemical reaction concept. The advantage of the first method lies in smaller dimension of the matrices used in calculations. Though the results in both used methods for the here studied case do not differ considerably, in general case it is necessary – if possible – to prefer the second method. This is because in the calculation of the degrees of separation according to relations (*2*) and (*3*) the part of information included in quantities n^*_{ik} is lost. This can result in the increase in the probability of the IInd kind error.

An important part of the method forms determination of covariance matrices of the measured quantities. In this study the covariances were calculated from standard deviations of the directly measured quantities. This was based on the assumption that the errors of the directly measured quantities are the randomly uncorrelated variables with the normal distribution and the zero mean value. These assumptions can be often accepted in practice. From Fig. 2 results that the linearisation used in the calculation of covariances has no significient effect on distribution of quantities Q at the assumption that standard deviations of the measured quantities do not exceed several percent of the measured value.

The proposed method of evaluation of the material balance is based on the assumption that the measured quantities do include surplus information concerning the material balance. In the majority of cases it is not possible to measure all quantities appearing in the balance (all concentrations and streams). A certain number of these quantities can be calculated from relations of the type of Eqs (*20*) and (*21*). The use of these additional relations results in increase of the absolute values of non-diagonal terms of covariance matrices $\boldsymbol{W}$ and $\boldsymbol{F}$ and in increasing correlation between the smoothened quantities $^\wedge s_i$ and $^\wedge n_{ik}$. The use of additional balance relations results also in the increase of the probability of the IInd kind error. It can also happen that for some compound the balance is no more overdetermined and statistical treatment of the balance for this component is obviously senseless.

LIST OF SYMBOLS

$\boldsymbol{A}$	matrix defined in Eq. (*5*)
$\boldsymbol{B}$	matrix defined in Eq. (*13*)
$\mathbf{e}_n$	vector of errors of quantities $\boldsymbol{n}$
$\mathbf{e}_s$	vector of errors of quantities $\mathbf{s}$
$E(\,)$	mean value of random variable
$\boldsymbol{F}$	covariance matrix of vector $\mathbf{e}_n$
I	number of separated compounds
$\boldsymbol{n}$	vector of flow rate of compounds through individual streams
N_k	flow rate of the k-th stream
Q_n	quadratic form defined in Eq. (*18*)

Q_s	quadratic form defined in Eq. (*9*)
$\mathbf{r}$	vector of extents of fictive chemical reactions
x	concentration (mole fraction)
$\mathbf{W}$	covariance matrix of vector $\mathbf{e}_s$
γ	relative standard deviation
σ	standard deviation

Subscripts

i	i-th compound
k	k-th stream

Superscripts

-1	inversion matrix
T	transposed matrix
*	measured quantity
$\wedge$	smoothed quantity

REFERENCES

1. Václavek V.: This Journal *34*, 2662 (1969).
2. Madron F., Vaněček V.: This Journal *42*, 1805 (1977).
3. Madron F., Veverka V., Vaněček V.: AIChE J. *23*, 482 (1977).
4. Himmelblau D. M.: *Process Analysis by Statistical Methods.* Wiley, New York 1973.
5. Aris R.: *Introduction to the Analysis of Chemical Reactors.* Prentice—Hall, New York 1965.
6. Box M. J.: Technometrics *12*, 219 (1970).
7. Hahn G. J., Shapiro S. S.: *Statistical Models in Engineering.* Wiley, New York 1967.

Translated by M. Rylek.