

# Dependence of computed copolymer reactivity ratios on the calculation method. II. Effects of experimental design and error structure

**Citation for published version (APA):**

Hautus, F. L. M., Linssen, H. N., & German, A. L. (1984). Dependence of computed copolymer reactivity ratios on the calculation method. II. Effects of experimental design and error structure. *Journal of Polymer Science, Polymer Chemistry Edition*, 22(12), 3661-3671. <https://doi.org/10.1002/pol.1984.170221204>

**DOI:**

[10.1002/pol.1984.170221204](https://doi.org/10.1002/pol.1984.170221204)

**Document status and date:**

Published: 01/01/1984

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Dependence of Computed Copolymer Reactivity Ratios on the Calculation Method. II. Effects of Experimental Design and Error Structure

F. L. M. HAUTUS, H. N. LINSSEN,\* and A. L. GERMAN, *Laboratory of Polymer Chemistry, Eindhoven, University of Technology, Eindhoven, The Netherlands*

## Synopsis

Two calculation methods for estimating reactivity ratios, one method based on the differential Alfrey–Mayo equation and one based on the integrated form of this model, are compared with respect to precision and bias. Both methods are characterized by the use of information about the monomer feed composition only and are assumed to be valid up to high conversion. As only monomer feed composition has to be analyzed, several sampling designs are feasible. Two extreme designs can be distinguished. One consists of repetitive sampling of the initial and final monomer feed mixture, whereas the other consists of sequential sampling during the course of the reaction. The influence of both designs on the calculated  $r$ -values is investigated by means of simulation. In the present paper the second calculation method, based on the integrated form, is solved by a nonlinear least squares method considering errors in both variables. This method required additional information about the error structure of the data. As this information is mostly of approximate nature, the influence of misspecification of this error structure on the calculated  $r$ -values is also examined.

## INTRODUCTION

In 1944 the simple (differential) copolymer equation was derived by Alfrey and Goldfinger<sup>1</sup> and Mayo and Lewis.<sup>2</sup> This equation describes the ratio of the instantaneous rates of consumption of the monomers as function of the instantaneous monomer feed ratio and the two reactivity ratios (the  $r$ -values). By means of GLC techniques the instantaneous monomer feed ratio can be estimated directly.<sup>3</sup> There are, however, no techniques available to directly estimate the ratio of the instantaneous monomer consumption. A number of methods to compute reactivity ratios are developed, based on the differential equation and on data consisting of the initial monomer feed ratio and the resulting copolymer composition.

Three factors, two of experimental and one of mathematical nature, will introduce systematic errors in the calculated reactivity ratios:

- (1) The ratio of the instantaneous monomer consumption is approximated by the copolymer composition.
- (2) Most copolymerization reactions inevitably show a drift in the molar feed ratio, so an approximated monomer feed ratio must be chosen.

\* Department of Mathematics, Eindhoven University of Technology, Eindhoven, The Netherlands.

- (3) The reactivity ratios are computed, using simple least squares methods, thereby neglecting the error in the so-called independent variable.

For many copolymer reactions the error in the calculated  $r$ -values, caused by the approximations due to factors 1 and 2 increases with increasing conversion.

In the present paper two calculation methods, which are assumed to be valid up to high conversion, are considered. One method is based on the differential copolymer equation, the second on the integrated form of this equation. Both methods obtain all required information from the monomer feed composition, so the error due to factor 1 does not exist in this case. The advance of using a computer allows the use of the integrated copolymer equation such that the systematic error due to factor (2) can be eliminated. Patino-Leal et al.<sup>4</sup> and Van der Meer et al.<sup>5</sup> gave examples of significant systematic errors due to factor 3. They also described a calculation method to cope adequately with the error structure of the observations.<sup>5</sup> As this method is intricate and difficult to implement, relative to the method based on the differential equation, it seems worthwhile to determine those cases where it yields significantly better results. To examine both methods properly, experimental data should be simulated because only in that case error structure and true  $r$ -values are known exactly. This is done in the following way.

First, a number of representative  $r$ -values are chosen. For each couple of  $r$ -values true values of the monomer feed ratio and the corresponding conversion are computed with the aid of the integrated copolymer equation. Then observations are simulated by disturbing the generated ideal GLC peak areas by random error. Finally, estimated  $r$ -values are recalculated for each method. In order to obtain a sample of calculated  $r$ -values simulation is repeated several times for each method. These samples are compared according to several criteria.

Gas-liquid chromatographic techniques allow for the sampling during the course of copolymerization reactions. A number of experimental designs are now conceivable. The two extreme cases are:

- (a) the "repeated sampling" design: samples are taken repetitively but preceding and succeeding the period of reaction, in each case  $n/2$  times, where  $n$  is the total number of "observations" for each experiment; and
- (b) the "sequential sampling" design: samples are taken sequentially  $n$  times during the period of reaction, at approximately equal time intervals.

The effect of the choice of experimental design on the calculated  $r$ -values has been neglected in literature so far and is investigated in the first part of the present paper.

The second calculation method, the integrated copolymer equation, is solved with a nonlinear least squares method which takes into account errors both in conversion and in monomer feed ratio. So this method requires additional information on the observational error structure. As this information will mostly be of an approximate nature, the sensitivity of the calculated  $r$ -values with respect to misspecification of the error structure is investigated also by means of simulation.

## SUMMARY OF THE CALCULATION PROCEDURES USED

For the calculation of  $r$ -values Alfrey and Mayo derived the simple differential copolymer equation

$$\frac{d[M_1]}{d[M_2]} = \frac{r_1 q + 1}{r_2/q + 1} \quad (1)$$

where  $d[M_1]/d[M_2]$  is the ratio of the instantaneous rates of consumption of the monomers by chain propagation and  $q = [M_1]/[M_2]$  is the molar ratio of the instantaneous monomer feed concentrations.

Watts et al.<sup>6</sup> transformed this equation into

$$\frac{d \ln M_1}{d \ln M_2} = \frac{r_1 q + 1}{r_2 + q} \quad (2)$$

For many copolymerization reactions  $d \ln M_1/d \ln M_2$  appears to be rather constant within the kinetic experiment during the reaction even up to high conversion. This phenomenon was used by defining  $\beta$  as

$$\frac{d \ln M_1}{d \ln M_2} = \beta \approx \text{constant}$$

During the course of the reaction the monomer feed ratio ( $q$ ) is drifting so an average feed ratio is introduced by

$$q_A = (q_0 + q_F)/2 \quad (3)$$

where  $q_0$  and  $q_F$  are the initial and final monomer feed ratios, respectively. The following equation is now assumed to be valid

$$\beta = \frac{r_1 q_A + 1}{r_2 + q_A} \quad (4)$$

$\beta$  can now be estimated for each experiment by simple linear regression.<sup>6</sup> The set  $\beta$ 's and  $q_A$ 's obtained in this way, can be used to calculate the  $r$ -values employing a nonlinear least squares method based on eq. (4). This procedure will be referred to as the WLS (Watts, Linssen, and Schrijver) method.

The second method presently investigated is based on the integrated form of the Alfrey-Mayo equation. Integration yields

$$z_2 = \left(\frac{q}{q_0}\right)^{-1-x_2} \left(\frac{x_2 q - x_1}{x_2 q_0 - x_1}\right)^{1+x_1+x_2} \quad (5)$$

where  $x_1 = 1/(r_1 - 1)$ ,  $x_2 = 1/(r_2 - 1)$ ,  $z_2 = M_2/M_{20}$ , and  $M_{20}$  and  $M_2$  are the initial and instantaneous concentration of monomer 2.

Van der Meer et al.<sup>5</sup> used a nonlinear least squares algorithm considering errors in both variables, developed by Linssen,<sup>7</sup> for the estimation of  $r_1$  and

$r_2$ . This method of estimating  $r$ -values will be referred to as the VLG (Van der Meer, Linssen, and German) method.

### SIMULATION AND COMPARISON

Simulation is based on the presumption that reaction mixtures containing two monomers and a solvent are analysed quantitatively by means of GLC. These primary experimental data allow all necessary variables to be obtained (monomer feed ratio, conversion  $M_1$ ,  $M_2$  etc.). The solvent peak is used as internal standard and is thus to be considered as one of the primary experimental data.

The integrated form of the Alfrey-Mayo eq. (5) is based on the conversion of one of the monomers. For simulating purposes, this form is not recommended. When the conversion of one of the monomers is selected, the conversion of the other monomer may attain extremely high or low values especially if the  $r$ -values are very dissimilar. Therefore, a measure of conversion based on the total number of moles reacted is introduced here. This "total conversion" is defined as

$$\text{Conversion}_t = \frac{M_{10} - M_1 + M_{20} - M_2}{M_{10} + M_{20}} = 1 - z_t \quad (6)$$

where  $M_{10}$  and  $M_{20}$  are initial number of moles of  $M_1$  and  $M_2$ , respectively; and  $M_1$  and  $M_2$  are instantaneous number of moles of  $M_1$  and  $M_2$ , respectively.

The equation used to simulate the "experiments" is based on this "total conversion" and can be derived by combining eqs. (5) and (6)

$$z_t = \frac{(q+1)}{(q_0+1)} \cdot \left(\frac{q}{q_0}\right)^{-1-x_2} \left(\frac{x_2 q - x_1}{x_2 q_0 - x_1}\right)^{1+x_1+x_2} \quad (7)$$

For selected couples of  $r$ -values, 10 "experiments" with initial monomer feed ratio ( $q_0$ ) varying from 5 in steps of 0.5 down to 0.5, are chosen. Within each set of "experiments" the 'total conversion' is selected (20 or 40%) and feed ratios are computed by means of eq. (7). Then, for each "experiment" ideal GLC peak areas are generated (realistic values are chosen for the areas of solvent- and initial- $M_2$  peak).

In case of repeated sampling design only initial and final 'observations' are simulated and these ideal peak areas are now randomly disturbed by a normal error with mean zero and standard deviation of  $1/\sqrt{15}\%$ . This standard deviation is equal to the standard deviation of the average of 15 random disturbances each with standard deviation of 1%.

In case of sequential design, the "total conversion" is given 30 different values, starting at 0 and increasing by a fixed amount till a maximum value of 20 and 40% respectively, and the corresponding monomer feed ratios are computed. From these values ideal peak areas are calculated. 'Observations' are simulated by disturbing these ideal peak areas using normal error with zero mean and standard deviation 1%. From these disturbed areas monomer feed ratio, degree of conversion, beta, etc, are computed, and these data are

used to estimate the  $r$ -values using the two calculation methods considered in this paper.

In this way 50 estimates are obtained for both methods and each design. Note that the difference between the repeated- and sequential sampling design for calculation method WLS is only caused by the estimation of  $\beta$ . This  $\beta$  is determined for each experiment either by the initial and final "observation" with error of  $1/\sqrt{15}\%$ , (WLSR) or by 30 "observations" each with error 1% spread over the periode of reaction (WLSS). The average monomer feed ratio is in both cases defined by eq. (3).

There are a number of ways to compare the performance of both methods. Two useful quantitative measures are the "mean distance"

$$MD = \sum_{p=1}^{50} \sqrt{(r_{i1} - R_1)^2 + (r_{i2} - R_2)^2} / 50, \quad (8)$$

and the "mean relative distance"

$$MRD = \sum_{p=1}^{50} \sqrt{(1 - r_{i1}/R_1)^2 + (1 - r_{i2}/R_2)^2} / 50, \quad (9)$$

where  $(r_{i1}, r_{i2})$  represent the estimated  $r$ -values, and  $R_1$  and  $R_2$  the corresponding true alues.

In Table I the MRD's are displayed for all cases and both designs considered here. Displaying MD's would give similar results.

The results of the simulation can also visually be represented by "approximately smallest 50% frequency regions." These are (approximately) the smallest regions that contain 50% of the pairs of estimated  $r$ -values. The regions are ellipsoids in the  $r_1$ - $r_2$ -plane given by

TABLE I  
Mean Relative Distance (MRD) for VLG and WLS and Repeated and Sequential Sampling Designs

$R_1$	$R_2$	Total conversion (%)	VLG		WLS	
			Repeated	Sequential	Repeated	Sequential
0.20	10	20	0.079	0.149	0.112	0.518
		40	0.024	0.045	0.086	0.169
0.35	0.20	20	0.055	0.098	0.062	0.145
		40	0.023	0.043	0.029	0.061
0.45	0.02	20	0.419	0.906	0.531	1.320
		40	0.165	0.337	0.422	0.371
0.60	0.50	20	0.034	0.064	0.033	0.059
		40	0.014	0.028	0.014	0.027
0.90	0.30	20	0.043	0.082	0.043	0.093
		40	0.017	0.035	0.018	0.038
2.0	0.10	20	0.158	0.229	0.171	0.270
		40	0.056	0.078	0.128	0.098
4.0	0.70	20	0.055	0.101	0.066	0.211
		40	0.023	0.040	0.030	0.060
10	0.30	20	0.161	0.370	0.271	0.864
		40	0.050	0.104	0.156	0.259

$$(r_1 - \bar{r}_1, r_2 - \bar{r}_2) \mathbf{S}^{-1} \begin{pmatrix} r_1 - \bar{r}_1 \\ r_2 - \bar{r}_2 \end{pmatrix} = \chi_2^2(0.50)$$

where  $\bar{r}_1$  and  $\bar{r}_2$  denote the average value of the reactivity ratios,  $\mathbf{S}$  denotes the estimated covariance matrix, and  $\chi_2^2(0.50)$  the 50% point of the chi-square distribution on 2 degrees-of-freedom.

In Figure 1 all 50 estimated pairs of  $r$ -values for both methods and both designs are indicated as given in the legend, also the 50% regions are displayed. In Figures 2 and 3 these 50% regions are displayed for all cases and both methods.

In the sequel the following terminology will be used. When the area of its 50% region is small (large) the method under consideration will be said to have a large (small) "precision." When the distance between the center of the 50% region and the pair of true values of  $r_1$  and  $r_2$  is small (large), the method will be said to have a small (large) "bias." Small precision or high bias will give rise to a large mean distance MD and large mean relative distance MRD.

Figures 2 and 3 and Table I lead to the following conclusions:

- (1) Increasing conversion causes smaller mean relative distance for all cases.
- (2) Repeated sampling design is superior to the sequential sampling design with respect to the MRD. (For WLS there are two exceptions  $r_1 = 0.45$  and  $r_1 = 2$ , both 40% conversion.)
- (3) For the WLS method sequential sampling design may introduce

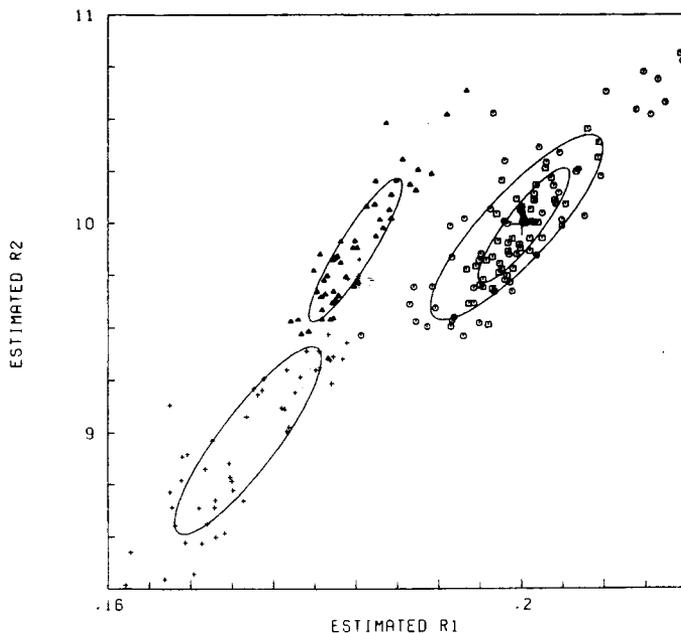


Fig. 1. Fifty simulated estimates for  $r_1$  (0.20) and  $r_2$  (10.0) and associated approximately smallest 50% frequency regions, for both methods and both designs: ( $\square$ ) VLGR, ( $\circ$ ) VLGS, ( $\triangle$ ) WLSR, (+) WLSS. Total conversion, 40%.

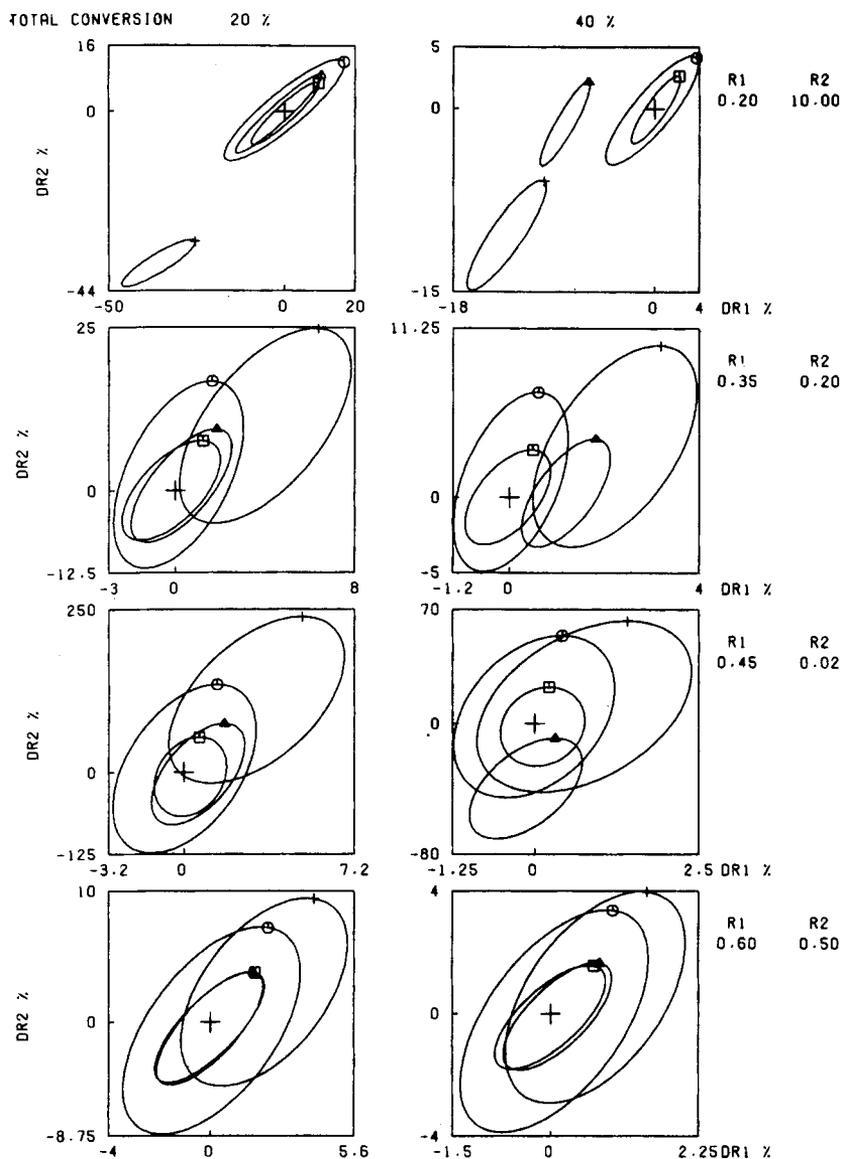


Fig. 2. 50% Ellipsoids for methods (□) VLGR, (○) VLGS, (△) WLSR, and (+) WLSS.

bias. This bias however becomes less if conversion increases (compare WLSS 20—40% conversion). On the other hand, increasing conversion also means increasing bias (compare WLSR 20—40% conversion). Besides, both causes of bias seem to be negatively influenced by dissimilar  $r$ -values ( $r_1 = 0.45, r_2 = 0.02$  have the largest MRD).

- (4) In case of VLG the sampling method introduces no bias. This means that the increase of the MRD is entirely due to loss of precision. Note that WLSS may be 'precisely wrong' whereas WLSR is 'roughly right' ( $r_1 = 10, r_2 = 0.3, 20\%$  conversion).

The repeated sampling design is superior to the sequential sampling design. This is caused by the fact that in case of repeated sampling, the 15 observations at both extremes of the observational range carry all the information, whereas in the sequential sampling case, the information is spread over the whole observational range.

For VLG this difference causes only a loss in precision and gives no rise to bias. For WLS however  $\beta$  and  $\delta = 1/\beta$  is estimated, using simple linear regression for each experiment, by interchanging the role of the dependent and independent variables.

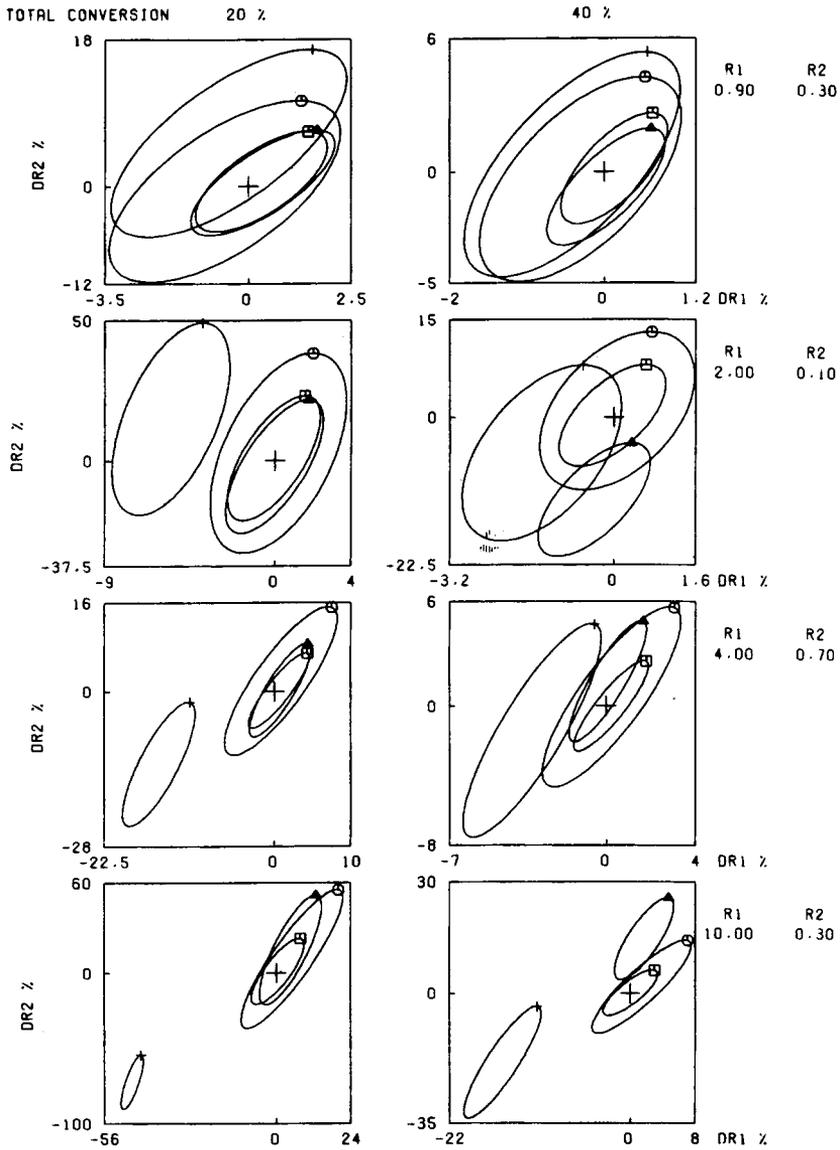


Fig. 3. 50% Ellipsoids for methods (□) VLGR, (○) VLGS, (△) WLSR, and (+) WLSS.

The error in the dependent variable must be substantially larger than that in the independent variable, in order for the linear regression method to yield reliable results, and so, either the estimate for  $\beta$  or the estimate for  $1/\beta$  may be unreliable.

An advantage of the sequential sampling design is that it allows for model-testing: deviations from the Alfrey-Mayo model can be detected. As long as there is some doubt about the validity of the Alfrey-Mayo model, a purely repeated sampling design is not recommended.

### SENSITIVITY OF VLG WITH RESPECT TO MISSPECIFICATION OF ERRORSTRUCTURE

The gas-liquid chromatographic technique for the analysis of a copolymerization reaction mixture yields direct information on the instantaneous monomer feed composition in terms of three peak areas:  $A_s$  = the observed area of solvent,  $A_1$  = observed area of monomer 1, and  $A_2$  = observed area of monomer 2. The observational error in the peak areas are assumed to be independent with standard deviations:

$$\sigma(A_k) = \lambda_k a_k \tau \quad k = s,1,2 \quad (10)$$

where  $a_k$  denotes the "true" value of the observed area and  $\tau$  is a common, possibly unknown, scale factor. The meaning of  $\lambda_k$  will be explained as follows. Denoting

$$\Delta A_k = A_k/a_k - 1$$

then

$$\sigma(\Delta A_k) = \lambda_k \tau$$

This means that on an average the relative errors in  $A_1$  and  $A_2$  differ a known factor  $\lambda_1/\lambda_s$ , respectively  $\lambda_2/\lambda_s$  from the relative error in  $A_s$ . The effects these errors have on the monomer feed ratio and the degree of conversion can be computed by means of the law of propagation of errors<sup>8</sup> and is described in Van der Meer et al.<sup>5</sup>

It is essential for the application of VLG that the ratios  $p_1 = \lambda_1/\lambda_s$  and  $p_2 = \lambda_2/\lambda_s$  are known. When this knowledge is only of an approximate nature, it is important for the experimentalist to know how sensitive VLG is with respect to misspecification of the errorstructure. One way to quantify this sensitivity is to repeatedly estimate the reactivity ratios each time using different, but plausible, values of  $p_1$  and  $p_2$ . If the resulting estimates do not differ significantly the experimentalist can rest assured. If they do, then additional information concerning  $p_1$  and  $p_2$  is needed, as it is well known in statistical literature<sup>9</sup> that  $p_1$  and  $p_2$  are not estimable from the copolymerization data alone.

During the present investigation a sensitivity study as described above is carried out as follows: It is assumed that  $p = p_1 = p_2$ . Data are simulated with  $p = 1$  and analyzed according to five specifications:

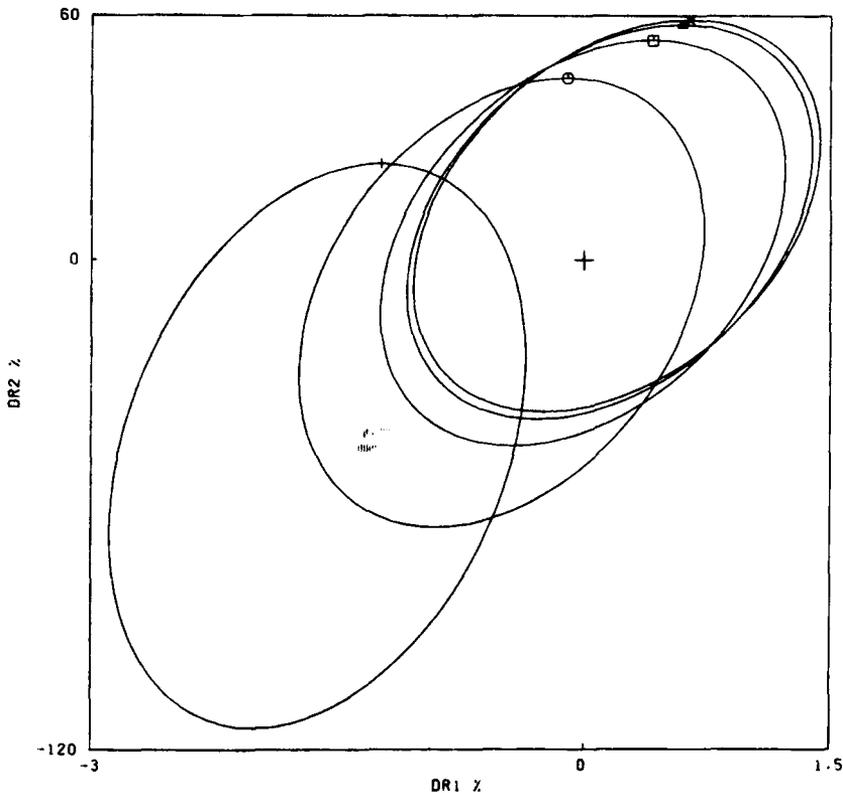


Fig. 4. Approximately smallest 50% frequency regions for  $r_1$  (0.45) and  $r_2$  (0.02), misspecified error structure and calculation method VLGS: ( $\square$ )  $p = 1$ , ( $\circ$ )  $p = 1/2$ , ( $\triangle$ )  $p = 2$ , ( $+$ )  $p = 1/4$ , ( $\times$ )  $p = 4$ . Total conversion, 40%.

$p = 1$  (the right one),  $1/2$ ,  $2$ ,  $1/4$ , and  $4$

Both sampling designs are considered. The results for two representative cases are displayed in Figures 4 and 5 for the sequential sampling design. These two figures and the calculated results lead to the following conclusions:

- (1) For the repeated sampling designs the estimates were found to be hardly sensitive to the values of  $p$ .
- (2) For the sequential sampling designs the bias increases for misspecification errors. Especially for  $p = 0.25$  (the errors in solvent peaks are specified to be four times as large as they actually are), the bias is significant.

The somewhat surprising conclusion (1) makes it plausible that the biases of the WLSR results as displayed in Figures 2 and 3 are due to linearization alone, at least for observational error magnitude as used in this paper. It is assumed that there is no bias in the observations themselves. Systematic observation error is expected to result in seriously biased estimated, as noted by Petrak and Pitts.<sup>10</sup>

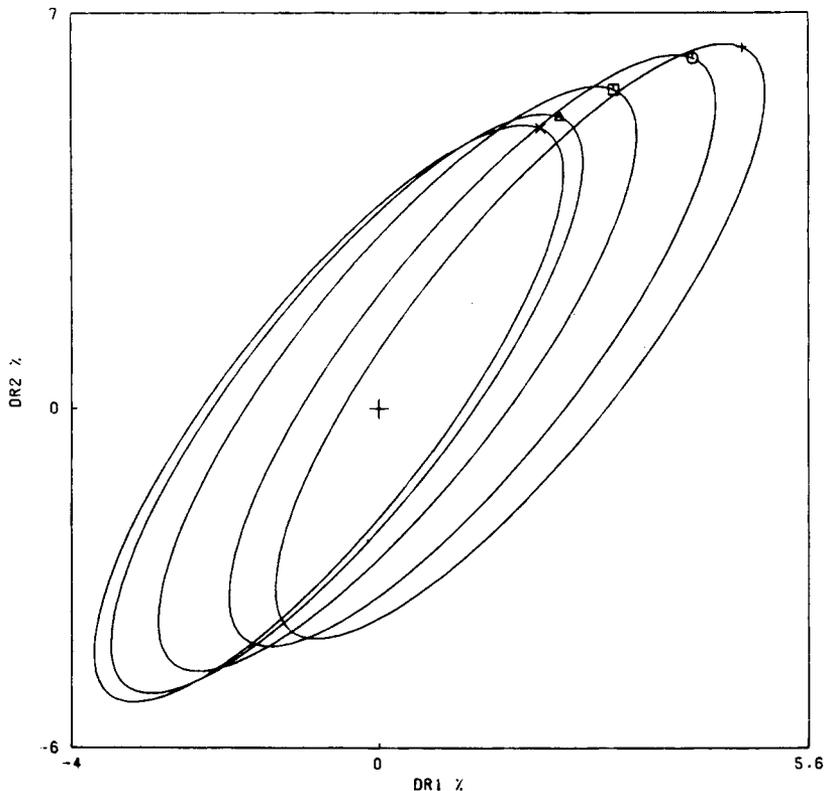


Fig. 5. Approximately smallest 50% frequency regions for  $r_1$  (4.00) and  $r_2$  (0.70), misspecified error structure and calculation method VLGS:  $p = 1$ , (○)  $p = 1/2$ , (△)  $p = 1/4$ , (+)  $p = 1/4$ , (X)  $p = 4$ . Total conversion, 40%.

The authors thank Mr. H. Willemsen for the development of the software required to represent the data graphically.

### References

1. T. Alfrey Jr. and G. Goldfinger, *J. Chem. Phys.*, **12**, 205 (1944).
2. F. R. Mayo and F. M. Lewis, *J. Am. Chem. Soc.*, **66**, 1594 (1944).
3. A. L. German and D. Heikens, *J. Polym. Sci. Part A-1*, **9**, 2225 (1971).
4. H. Patino-Leal, P. M. Reilly, and H. F. O'Driscoll, *J. Polym. Sci. Polym. Lett. Ed.*, **18**, 219 (1980).
5. R. Van der Meer, H. N. Linssen, and A. L. German, *J. Polym. Sci. Polym. Chem. Ed. V* **16**, 2915 (1978).
6. D. G. Watts, H. N. Linssen, and Y. Schrijver, *J. Polym. Sci. Polym. Chem. Ed. V* **18**, 1285 (1980).
7. H. N. Linssen, "Nonlinear Regression with Nuisance Parameters: An Algorithm to Estimate the Parameters," *Proceedings 1st European Meeting of Statisticians*, Grenoble, 1976, North Holland, Amsterdam, 1977.
8. J. Mandel, *The Statistical Analysis of Experimental Data*, Interscience, New York, 1964, p. 74.
9. M. E. Solari, *J. R. Stat. Soc. B* **31**, 372-375, (1969).
10. K. L. Petrak and E. Pitts *Polymer*, **24**, 729-732 (1983).

Received March 5, 1984

Accepted May 7, 1984