Nonlinear least squares fitting applied to copolymerization modeling

Alex M. van Herk*, Thomas Dröge²

1 Laboratory of Polymer Chemistry, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands
2 Institut für Physikalische Chemie, Universität Göttingen, Tammanstraße 6, D-37077 Göttingen, Germany

(Received: June 11, 1997; revised manuscript of July 14, 1997)

SUMMARY:
The choice of the statistical method to determine the reactivity ratios in copolymerization modeling is shown to be very important. Problems in literature, as well as possible pitfalls when using available statistical programs that are in itself correct are pointed out. These problems mainly involve (knowledge of) the error structure, as the error structure determines the weighting scheme of the data points in fitting procedures. A simple, robust, statistically correct non-linear least squares (NLLS) method is reintroduced which is based on the visualization of the sum of squares of residuals in the so-called sum of squares space (SSS). The advantages of this method include the fact that the method is easy to understand and can be implemented in simple computer programs, as well as the fact that the method allows important aspects of the error structure to be incorporated. Furthermore, in the SSS the joint confidence interval (JCI) with exact shape can be constructed. The exact shape of the JCI is not always ellipsoidal and following a normal distribution, depending on the linearity of the fitted equations. This can sometimes lead to wrong conclusions.

Introduction
In the field of copolymerization modeling the use of the correct statistical procedures to obtain parameters is an important issue. Even up to this date people are using incorrect methods as it often seems to be a problem to decide which statistical method is correct for the problem at hand. Researchers encounter these problems for example in the determination of reactivity ratios in copolymerization from compositional data¹⁴, and even more complicated, from average propagation rate coefficients¹) and also in the determination of activation parameters⁷,⁸).

Especially in the field of determining reactivity ratios, although the correct methods are outlined in literature³,⁴a) already a long time ago, we can still find many examples in recent literature where major errors are made. Not only is it important to use correct methods of parameter estimation but also the issue of experimental design and model discrimination⁴a,b,c) is of paramount importance for obtaining accurate values for the parameters.

For example, without knowledge of the error structure in the measurements it is not possible to weight the measured data correctly in parameter estimation and also the construction of correct joint confidence intervals is therefore impossible. According to the authors’ opinion there are three main reasons for that:
1) The researchers that perform experiments to obtain reactivity ratios are usually not statisticians and have difficulties in comprehending the statistical papers on this topic.

2) The people that write the papers on the statistical methods usually forget that in real life the people that do the experiments try to do not too many experiments. One example is in obtaining an estimate of the variance of a measurement. In order to obtain a proper estimate of the variance that can be used in weighting the data, statistically one needs at least eight replicate measurements. There are only few, if any, papers in the field of reactivity ratio estimation where this is done. Yet statistical methods are applied in this area where it is assumed that the variance is known.

3) Even if the importance of using the right statistics is realized by the experimentators they are not in the possession of an appropriate computer program or have no possibilities to write such a program, so they resort to a simple linearization method. The Fineman-Ross linearization method is still used in 1997 although already in 1973 and before it was shown to be obsolete. Recently it was pointed out that with the advent of rapid computers a simple and old method of least squares fitting can be applied which has some advantages over the iterative methods commonly applied. This method comprises the calculation of the sum of squares of the residuals over a wide range of the parameters and constructing and visualizing this sum of squares of the residuals space (further abbreviated as SSS). In this SSS the optimum set of parameters is found as a minimum in the sum of squares of the residuals. Although the computational effort is larger than the iterative procedures, the method has the advantage that there is less risk of running into a local minimum and also from the SSS the true joint confidence interval (JCI) with correct shape can be constructed. This method was already applied to copolymerization problems in 1964 by Behnken and in 1965 by Tidwell and Mortimer. Recently this method was applied again to the determination of reactivity ratios from low conversion determinations of copolymer composition as a function of feed composition (determination of $r_1$ and $r_2$ in the ultimate model or restricted penultimate model). Also for the closely related topic of the determination of activation parameters of reactivity ratios and propagation rate constants this method has recently been applied.

One of the drawbacks of this method is that the independent variable is assumed to be errorless. However in most of the above mentioned fitting problems this assumption is justified as will be shown in the examples treated in this paper.

The aim of this paper is to point out the importance of knowing the error structure in relation to the proper weighting of the measurements during the fitting procedure. A new aspect that will be shown is that sometimes the JCIs with exact shape but approximate size are to be preferred over the joint confidence intervals with approximate shape but exact size.
The non-linear least squares method by visualizing the sum of squares space

Finding the optimal parameter values

The method is based on the principle of minimizing the sum of squares of weighted residuals. In contrast to other methods this one does not look for the minimum in the sum of squares of the residuals space by iterative procedures but visualizes the SSS in a user defined region and finds the optimal parameter values.

The sum (over \( n \) data points) of squares (\( ss \)) of residuals as a function of, for example, the two parameters \( o_1 \) and \( o_2 \) is defined as:

\[
ss(o_1, o_2) = \sum_{i=1}^{n} \{w_i[y_i - f(x_i, o_1, o_2)]^2\} \tag{1}
\]

where \( f(x_i, o_1, o_2) \) is the function that relates the independent variable \( x_i \) to the dependent (response) variable \( y_i \) and \( w_i \) is the weighting factor for data point \( i \). In this SSS, constructed by plotting \( ss(o_1, o_2) \) as a function of \( o_1 \) and \( o_2 \), the optimum set of the parameters is found, e.g. \( \delta_1 \) and \( \delta_2 \), at a minimum sum of squares \( ss(\delta_1, \delta_2) \). The principle is represented schematically in Fig. 1.

The advantage of this method is that, when the user defines a proper region where the solutions for the parameters are to be found, always the true minimum is found. There is no chance of getting stuck in a local minimum.

In general the least squares methods require that the errors are independent and normally distributed with a common variance. In order to comply to this restriction in cases that the variance is not equal for all measurements, appropriate weighting of the data needs to be performed, depending on the error structure.

Three weighting schemes can be used to weight the data according to Eq. (1):

1. The best weighting scheme is to use the reciprocal variance; \( w_i = 1/\sigma_i^2 \) where \( \sigma_i \) is the standard deviation in a normal distribution. Instead of using the exact variances, an estimate of the variance obtained from doing replicates can be used; \( w_i = 1/s_i^2 \) where \( s_i \) is the estimated standard deviation in a normal distribution.

The question arises how many replicate measurements are needed to get a reliable estimate of the variance. With a small number of replicates these estimates are likely to be very unreliable. Bardsley et al. showed that at least eight replicates should be obtained in order to use estimated variances in weighting. In general no reliable estimates for the variances in the measurements are available in the copolymerization literature. With a small number of replicates it is also very difficult to make a statement about the type of error distribution (e.g. a normal distribution). Quite often it is not clear what is meant with “the error in the copolymer composition is 5%”, as this could be a standard deviation, the outer boundaries of a normal distribution (about three times the standard deviation) or is it the outer boundary of a uniformly distributed error. With only a few replicates it is most likely to represent the outer boundaries of an unknown error distribution that quite often is treated as the outer boundaries of a uniformly distributed error. When the errors are not too large, one can convert these outer boundaries \( \pm \varepsilon_i \) to an estimated standard deviation \( s_i \) in a normal distribution through:
Fig. 1. Schematical representation of the method of non-linear least squares by visualization of the sum of squares space. The upper part of the scheme shows the sum of squares of residuals space, \( ss(o_1, o_2) \) as a function of the parameters \( o_1 \) and \( o_2 \). At the optimum combination of parameters \( (\hat{o}_1, \hat{o}_2) \), right arrow, there is a minimum in the sum of squares space corresponding to the best fit of the function through the data (lower part of the scheme, drawn line). If another combination of the parameters \( o_1, o_2 \) is taken (left arrow) this does not result in a good fit.
\( \varepsilon_i = \sqrt{3} \, s_i \). An alternative for doing many replicate measurements in order to obtain estimates of the variance is an elaborate error discussion as for example is done in ref.\(^5\). As soon as there is no reliable estimate of the variance one has to resort to another weighting scheme.

(2) The assumption of constant relative error (also called a multiplicative error): This assumes that the variances of the observed responses are proportional to the square of the theoretical responses \( \sigma_i^2 \equiv f(x_i, \delta_1, \delta_2)^2 \). Actually the weights derived from the best fit responses should be used. A variation to this method is to use the weights calculated from the measured responses; \( w_i = 1/y_i^2 \). When the data are very noisy this can lead to biased and/or uneven weighting. This approach leads to (unreasonably) small variance estimates if the measured responses are small and to computational problems if the responses are very close to zero.

(3) The assumption of a constant variance (also called an absolute or additive error): Here it is assumed that the errors are effectively constant over the range of measurements and therefore also independent of the values of the observed responses \( y \)-values), \( w_i = 1 \).

So the purpose of weighting the measurements is to comply to the restriction that the least squares method can only be applied to measurements where the errors have a common variance, are independent and normally distributed.

The choice for a particular weighting type depends on the type of error structure in the data set. Often the exact errors are not known but the absolute or relative error is thought to be constant. If one of these assumptions is made, careful inspections of the table of residuals should be performed in order to see whether this assumption was right. A method to check the residuals in a quantitative way are the run and sign tests\(^10\).

If the error in the independent variable is not negligible, in principle the error in (all) variables should be used as pointed out by Joshi in 1973\(^9\). An option is to transpose the error in \( x_i \) to an error in \( y_i \) by using the derivative of the function to \( x \) and apply the nonlinear least squares method\(^5\). The Error-in-Variables-Model\(^11,12\) is dealing with the errors in all variables in a correct way. This model forms the basis of an often used microcomputer program (RREVM) to calculate reactivity ratios\(^12\).

In this program the entered errors are the basis of a calculation of the variance-covariance matrix which in turn is used to construct the joint confidence intervals, all on the basis that the entered errors can be treated as known errors. It is the authors' impression that the program is most seldom used in a correct way, in the sense that enough replicates are taken in order to obtain a reliable estimate of the error (see discussion above on weighting scheme 1).

### Finding the joint confidence intervals with exact shape

The unbiased (exact shape) joint confidence interval for the parameter estimates is bounded by contours in the sum of squares space corresponding to those of the likelihood function\(^3\):
In the case that only error estimates are available (e.g. the relative/absolute error for all the measurements is approximately equal but not exactly known), the determination of this joint confidence interval (JCI) is performed by a so called F-test\textsuperscript{3,11} where the probability F-values are used at level z (usually 90 or 95% probability):

\[ ss(\hat{\sigma}_1, \hat{\sigma}_2)_z \leq ss(\hat{\sigma}_1, \hat{\sigma}_2) + p \ s^2 F_z(p, n-p) \]  

where \( F_z(p, n-p) \) represents a value from the F-distribution at level z at p and n-p degrees of freedom (p, the number of parameters, equals two in the above mentioned examples) and n data points, \( ss(\hat{\sigma}_1, \hat{\sigma}_2) \) is the sum of squares of residuals at the minimum and \( s^2 \) is an estimate of the true variance \( \sigma^2 \) which can be obtained from the residual sum of squares (if the model is correct): \( s^2 = ss(\hat{\sigma}_1, \hat{\sigma}_2)/(n-p) \), which after substitution in Eq. (2a) leads to:

\[ ss(\hat{\sigma}_1, \hat{\sigma}_2)_z \leq ss(\hat{\sigma}_1, \hat{\sigma}_2) \{ 1 + p/(n-p) F_z(p, n-p) \} \]  

– When the errors in the measurements are exactly known, for example by statistical analysis of (a minimum of eight) replicate measurements, the \( \chi^2 \) distribution can be used to construct the JCI\textsuperscript{11}:

\[ ss(\hat{\sigma}_1, \hat{\sigma}_2)_z \leq ss(\hat{\sigma}_1, \hat{\sigma}_2) + \sigma^2 \chi^2_z(p) \]  

where \( \chi^2_z(p) \) represents a value from the \( \chi^2 \)-distribution at level z at p degrees of freedom and \( \sigma^2 \) corresponds to the true variance of the dependent variable. Usually an estimate of \( \sigma^2 \) is used (\( s^2 \)) which is calculated from the known errors. The latter approach, obviously, in general leads to smaller JCIs because more is known about the actual error in the measurements.

In the case of estimated errors, the joint confidence region is determined by the sum of squares in the minimum (thus determined by the actual calculated residuals and not by the estimated errors) and the F-value for that particular number of points and parameters.

In the case of known errors, the \( \chi^2 \) distribution in combination with the known errors and the minimum sum of squares is used to construct the JCI (Eq. (3)).

The JCIs as calculated by this approach are the unbiased joint confidence intervals with exact shape (but approximate size) and as such can deviate from symmetric ellipses produced by for example the RREVM program because there the (symmetric) ellipses with approximate shape (but exact size) are calculated\textsuperscript{12} according to a method described by, for example, Hautus et al.\textsuperscript{13} The departure from exact ellipsoidal shape depends on the linearity of the function and is even used as a measure of linearity of a functional relationship\textsuperscript{14}. The Arrhenius equation is an example of an equation that can be linearized, therefore the exact shape of the JCIs corresponds to an ellipse (when the logarithm of the pre-exponential factor and the activation energy are taken as the parameters). On the other hand equations that are non-linear will have JCIs that can strongly deviate from ellipses (see last example in this paper).
The estimated standard deviation for the parameters can be calculated from the joint confidence intervals assuming a normal (or Gauss) distribution of the error around the optimum parameters. The 95% probability level, for example, is associated with $\pm 1.96$ times the standard deviation in a normal distribution. This standard deviation is only valid if the JCIs are symmetric around the optimum parameters and follow a normal distribution, in other cases only the JCIs themselves should be used to make statements about the errors in the parameters.

**Commonly made mistakes in determining reactivity ratios**

Many papers$^{1-4,9,11,13,15,16}$ deal with a comparison of the use of different statistical methods to obtain reactivity ratios, like linear regression, non-linear regression and the EVM method. From the literature a few basic problems arise:

- transformation of the original measurements in order to suit e.g. linearized equations often also leads to an unwanted change in the error structure which makes the method of least squares invalid because it is assumed that the errors are independent and normally distributed with a common variance. In some special cases a logarithmic transformation in combination with a multiplicative error structure in the original data leads to a correct error structure again$^7$, if the new error structure is taken as being additive (constant absolute error).
- Appropriate weighting of the measurements is usually overlooked.
- The joint confidence intervals, because the total sum of squares of residuals space is usually not available, are constructed as a symmetrical ellipsoid which is only a good approximation if the function is not too non-linear.
- Usually the error structure in the measurements is not known but treated as if it were known and therefore wrong statistics are used. For example the RREVM program uses the $\chi^2$-distribution (Eq. (3)) and therefore treats the entered errors as known errors!
- The effect of conversion in copolymerization is underestimated as discussed for example by Chee$^{16}$.

The method of visualizing the sum of squares space$^1$ can overcome most of the above mentioned problems, although it is not able to account for the errors in the independent variables. It is however believed by the authors that if the errors in the independent variables are small there is little loss in using the least squares method over the EVM method. The least squares method, in the form as proposed by one of the authors$^7$, offers the benefits that it is easily understood, the algorithm can easily be programmed in computer languages like BASIC or FORTRAN or even in spreadsheet programs like Excel and aspects of error structure and weighting can easily be incorporated.

The method will be applied to some copolymerization problems from literature where a particular emphasis will be put on the size and shape of the joint confidence intervals.
Some examples in copolymerization

Copolymer composition versus feed composition

Until this date one can find many examples in literature where linearized copolymer equations are used to fit the copolymerization data\(^{17a,b}\).

The first example is taken from a recent publication where still the Fineman-Ross linearized copolymer equations were used\(^{17a}\). The publication deals with the determination of copolymerization parameters of methyl methacrylate (MMA) and dodecyl methacrylate (DMA) by means of FTIR spectroscopy. In this publication the Fineman-Ross and Kelen-Tudos linearized copolymerization equations (of the ultimate model) are used to analyze copolymer composition expressed in mole fraction of MMA in the copolymer \((F_{\text{MMA}})\) as a function of the mole fraction of MMA in the feed \((f_{\text{MMA}})\).

Tab. 1. Composition of copolymers of MMA-DMA determined by FTIR spectroscopic measurements in CDC\(_3\)\(^{17a}\)

<table>
<thead>
<tr>
<th>(f_{\text{MMA}})</th>
<th>0,1</th>
<th>0,3</th>
<th>0,4</th>
<th>0,5</th>
<th>0,6</th>
<th>0,7</th>
<th>0,8</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F_{\text{MMA}})</td>
<td>0,059</td>
<td>0,243</td>
<td>0,364</td>
<td>0,486</td>
<td>0,583</td>
<td>0,721</td>
<td>0,824</td>
</tr>
</tbody>
</table>

The reevaluation of the data is performed with the non-linear least squares (NLLS) method applied on the original copolymer equation:

\[
F_1 = \frac{r_1 f_1^2 + f_1 f_2}{r_1 f_1^2 + 2 f_1 f_2 + r_2 f_2^2}
\] (4)

In the publication no indication of the size or structure of the errors is given. It is inferred from the data that the error in the feed composition is very small and therefore the NLLS method can be used instead of the EVM method. The error in the copolymer compositions is not estimated in the publication. It is very important to decide whether there is a constant absolute or a constant relative error in \(F_{\text{MMA}}\). Because it is not clear from the original data what error structure to assume both assumptions will be compared. In Fig. 2 the optimum parameters \(r_1\) and \(r_2\) are plotted together with their joint confidence intervals. The methods of absolute or relative error minimization give different optimum parameters because the data are weighted differently. The program RREVM\(^{12}\) gives the same parameter set as NLLS with relative error minimization (Eq. (1), weighting scheme 2); \(r_1 = 1,71, r_2 = 1,95\). Absolute error minimization (Eq. (1), weighting scheme 3) gives: \(r_1 = 1,43, r_2 = 1,67\).

In the RREVM program errors of 1% in the feed composition and 5% in the copolymer composition were taken with which the approximate JCI was constructed (Fig. 2, curve a). Changing the error in the feed composition from 1% to 0,1% does not result in a noticeable decrease in the size of the JCI as constructed by the RREVM program, indicating that an error of 1% in the feed composition in this case
Fig. 2. Optimum parameters and joint confidence intervals for the reactivity ratios of the copolymerization of methyl methacrylate and dodecyl methacrylate: a) — RREVM program, 1% error in feed composition and 5% error in copolymer composition, b) - - - - NLLS, known errors of 5% in the copolymer composition, using Eq. (3), c) - - - - NLLS, constant relative error, using Eq. (2), d) - - - - NLLS, constant absolute error, using Eq. (2). □: Minimum in SSS according to methods a, b and c. ▲: Minimum in SSS according to method d

does not contribute to the size of the JCI. Performing NLLS with known errors of 5% (treated as outer limits of a uniform error distribution, see above) in the copolymer composition and using Eq. (3), the JCI comes closest to the one given by the RREVM program. The small differences in form and shape result from the fact that the RREVM program uses an approximation for the construction of the JCI[13]. Also the JCI for the NLLS method with relative (weighting scheme 2, curve c) and with absolute error constant (weighting scheme 3, curve d) where constructed using Eq. (2). Both these JCI are larger than the JCI constructed with the RREVM program.

From this comparison we can conclude that the weighting scheme can have a dramatic effect on the obtained optimum parameters and therefore every report of reactivity ratios should be accompanied by a discussion of the error structure in the copolymer composition (see also ref.[5]). Furthermore, the size of the JCI depends on whether one assumes known or estimated errors. The average fitting error for the copolymer composition in the NLLS method amounts to 3%. When using the F-distribution (Eq. (2)) to construct the JCI, this JCI is still larger than the one constructed with a \(\chi^2\)-distribution, although here an error of 5% has been used in the copolymer composition (RREVM program and Eq. (3))!

In Tab. 2 the reactivity ratios obtained with the different methods are compared, together with the estimated standard deviations. It is obvious that the Fineman-Ross and the Kelen-Tüdös methods give wrong answers. All the methods with relative
Tab. 2. Reactivity ratios for the copolymerization of MMA with DMA obtained with different methods

<table>
<thead>
<tr>
<th>Method and weighting scheme (ws) to obtain parameters and JCI</th>
<th>( r_1 )</th>
<th>( r_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fineman-Ross(^{17a})</td>
<td>1,49 ± 0,09(^a)</td>
<td>1,79 ± 0,12</td>
</tr>
<tr>
<td>Kelen-Tüdös(^{17a})</td>
<td>1,58 ± 0,33</td>
<td>1,88 ± 0,32</td>
</tr>
<tr>
<td>NLLS, Eq. (2), ws 3</td>
<td>1,43 ± 0,20</td>
<td>1,67 ± 0,20</td>
</tr>
<tr>
<td>NLLS, Eq. (2), ws 2</td>
<td>1,71 ± 0,20</td>
<td>1,95 ± 0,15</td>
</tr>
<tr>
<td>NLLS, Eq. (3), ws 2</td>
<td>1,71 ± 0,19</td>
<td>1,95 ± 0,12</td>
</tr>
<tr>
<td>RREVM(^{13})</td>
<td>1,71 ± 0,16</td>
<td>1,95 ± 0,10</td>
</tr>
</tbody>
</table>

\(^a\) Estimated standard deviations.

error minimization give the same answers for the reactivity ratios but yield different estimated standard deviations. The RREVM method uses the errors entered by the users as being the outer limits of a known uniform distribution and constructs the JCI with the \( \chi^2 \)-distribution. Because a linearization is used the JCIs are always symmetric ellipses. In this particular example, the RREVM program gives a too small JCI because the errors in the copolymer composition are not accurately known and therefore the F-distribution should be used to construct the JCI.

If the sum of squares space around the minimum corresponds to a normal distribution it is possible to obtain an estimated standard deviation (e.s.d) from the JCIs. For example, the 75%, 90% and 95% probability levels are associated with 1,15 e.s.d., 1,645 e.s.d. and 1,96 e.s.d., respectively. If the JCIs constructed on these three probability levels are symmetric and give the same e.s.d. values within say 5% then the sum of squares space around the minimum conforms to a normal distribution and one can safely obtain an e.s.d from it. For the data tabulated in Tab. 1 the JCIs on three levels of probability, constructed with Eq. (3) using the \( \chi^2 \)-distribution with 5% error (outer limits of a uniform distribution) in the copolymer composition, the following e.s.d.'s were obtained: on probability levels of 75%, 90% and 95%, for \( r_1 \) 0,197, 0,179 and 0,170, respectively, and for \( r_2 \) 0,123, 0,112 and 0,103, respectively. So the sum of squares space around the minimum does not conform to a normal distribution. Therefore in this case the JCIs themselves should be shown as an indication of the errors and it is not correct to give standard deviations in this case. Because the copolymerization equation (Eq. (4)) is not too non-linear, when optimal experimental design is applied as proposed by Tidwell and Mortimer\(^4\), the sum of squares space usually conforms to a normal distribution.

It must be noted that one always has to be careful in using fitting programs in that one is aware of how the error in the measurements is treated in the computer program. Sometimes the entered error is treated as a standard deviation, sometimes as three times the standard deviation and (for example in RREVM) sometimes as the outer boundaries of a uniform distribution. Although in the latter case the estimated entered error is treated as the outer boundary of a uniform distribution the actual error structure must still be normally distributed.
**Average propagation rate coefficient versus feed composition**

The second example concerns the average propagation rate coefficient as a function of the feed composition. In this case we took some unpublished results obtained for the pulsed laser polymerization (PLP) on styrene-MMA\(^\text{18}\). In Tab. 3 the average propagation rate coefficient as a function of the feed composition is shown.

Tab. 3. The average propagation rate coefficient \(\langle k_p \rangle\) as a function of the feed composition \(f_{Sty}\) for styrene-MMA

<table>
<thead>
<tr>
<th>(f_{Sty})</th>
<th>(\langle k_p \rangle/(1 \cdot \text{mol}^{-1} \cdot \text{s}^{-1}))</th>
<th>(f_{Sty})</th>
<th>(\langle k_p \rangle/(1 \cdot \text{mol}^{-1} \cdot \text{s}^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>630.8</td>
<td>0.493</td>
<td>264.1</td>
</tr>
<tr>
<td>0</td>
<td>650.6</td>
<td>0.493</td>
<td>272.5</td>
</tr>
<tr>
<td>0</td>
<td>621.5</td>
<td>0.593</td>
<td>254.6</td>
</tr>
<tr>
<td>0</td>
<td>621.5</td>
<td>0.593</td>
<td>254.6</td>
</tr>
<tr>
<td>0.098</td>
<td>464.3</td>
<td>0.693</td>
<td>237.6</td>
</tr>
<tr>
<td>0.098</td>
<td>464.3</td>
<td>0.693</td>
<td>237.6</td>
</tr>
<tr>
<td>0.194</td>
<td>379.7</td>
<td>0.796</td>
<td>228.0</td>
</tr>
<tr>
<td>0.194</td>
<td>391.7</td>
<td>0.796</td>
<td>228.0</td>
</tr>
<tr>
<td>0.294</td>
<td>342.6</td>
<td>0.897</td>
<td>231.3</td>
</tr>
<tr>
<td>0.294</td>
<td>332.2</td>
<td>0.897</td>
<td>231.3</td>
</tr>
<tr>
<td>0.393</td>
<td>291.5</td>
<td>1</td>
<td>246.2</td>
</tr>
</tbody>
</table>

This example is chosen because for this particular problem this is one of the largest datasets we know. (The dataset contains some exact duplications of \(\langle k_p \rangle\) for a given composition.) Also the system Sty-MMA has been studied most extensively\(^1,2\). The data can be described with the penultimate model according to which:

\[
\langle k_p \rangle = \frac{\bar{r}_1 f_1^2 + 2f_1 f_2 + \bar{r}_2 f_2^2}{\bar{r}_1 f_1/k_{11} + \bar{r}_2 f_2/k_{22}}
\]  

(5)

where

\[
\bar{k}_{11} = \frac{k_{111}(f_1 r_1 + f_2)}{f_1 r_1 + f_2/s_1} \quad \text{and} \quad \bar{k}_{22} = \frac{k_{222}(f_2 r_2 + f_1)}{f_2 r_2 + f_1/s_2}
\]

\[
\bar{r}_1 = \frac{r_1(f_1 r_1 + f_2)}{f_1 r_1 + f_2} \quad \text{and} \quad \bar{r}_2 = \frac{r_2(f_2 r_2 + f_1)}{f_2 r_2 + f_1}
\]

and

\[
 r_1 = \frac{k_{111}}{k_{112}} \quad r_2 = \frac{k_{222}}{k_{221}} \quad r_1' = \frac{k_{211}}{k_{212}} \quad r_2' = \frac{k_{122}}{k_{121}} \quad s_1 = \frac{k_{211}}{k_{111}} \quad s_2 = \frac{k_{122}}{k_{222}}
\]

In the restricted penultimate model\(^9\) \(r_1 = r_1'\) and \(r_2 = r_2'\). If we take styrene as monomer one and methyl methacrylate as monomer two and fix the following parameters: \(k_{111} = 246 \text{ 1} \cdot \text{mol}^{-1} \cdot \text{s}^{-1}\), \(k_{222} = 631 \text{ 1} \cdot \text{mol}^{-1} \cdot \text{s}^{-1}\), \(r_1 = 0.48\) and \(r_2 = \)
Fig. 3. Optimum values for $s_1$ and $s_2$ with the joint confidence interval for styrene and methyl methacrylate at 50°C. a) NLLS using Eq. (2) and constant relative error, b) using the EVM approach with 0.5% standard deviation in the feed composition and 1% standard deviation in the $k_p$-values.

With weighting scheme 3 (absolute error constant) the joint confidence interval even stretches to infinity in both directions for $s_2$. Actually this particular example is quite favorable because there are many measurements and the homopropagation rate coefficients do not differ too much. Even in this case it is almost impossible to obtain an estimate for one of the two $s$-values. Therefore one should be most critical if parameter estimates for both $s$-values are given. The insensitivity for one of the two $s$-values of $k_p$ can be understood if we look at the contribution of the term in Eq. (5) that contains $s_2$ (that is the second term between brackets in the denominator). In this particular example ($k_{111}$ is about 1/3 of $k_{222}$) that term only contributes a few percent to the total value of the denominator (less than 25% above $f_1 = 0.2$). Only the $k_p$ values

0,42\textsuperscript{18}, we can obtain estimates for $s_1$ and $s_2$ with a two parameter fit. As is common in fitting $k_p$ values in this way, the relative error is assumed constant. We minimize the sum of squares according to Eq. (1) with weighting scheme 2 and construct the joint confidence intervals with Eq. (2).

As can be seen in Fig. 3a the parameter $s_1$ is well determined but $s_2$ is not. Whereas the JCI for $s_1$ stretches from 0.29 to 0.32, the JCI for $s_2$ stretches from 1.25 to 11. This phenomenon is quite common in fitting $k_p$ data with the (restricted) penultimate model. The $s$-value referring to the addition of the monomer with the highest homopropagation rate coefficient (in this case monomer two, methyl methacrylate) is poorly determined. With the same set of data and fitting with weighting scheme 3 (absolute error constant) the joint confidence interval even stretches to infinity in both directions for $s_2$. Actually this particular example is quite favorable because there are many measurements and the homopropagation rate coefficients do not differ too much. Even in this case it is almost impossible to obtain an estimate for one of the two $s$-values. Therefore one should be most critical if parameter estimates for both $s$-values are given. The insensitivity for one of the two $s$-values of $k_p$ can be understood if we look at the contribution of the term in Eq. (5) that contains $s_2$ (that is the second term between brackets in the denominator). In this particular example ($k_{111}$ is about 1/3 of $k_{222}$) that term only contributes a few percent to the total value of the denominator (less than 25% above $f_1 = 0.2$). Only the $k_p$ values
with a very low \( f_1 \) contribute considerably to the value of the denominator. So also here experimental design is necessary\(^{4a,b,c} \), where at least some measurements are needed very close to \( f_1 = 0 \) to obtain a reasonable estimate for \( s_2 \). As up to this date no studies in this sense have appeared in literature there are no reliable sets of \( s_1 \) and \( s_2 \) available\(^{20} \) and therefore, at least from experimental data, no inferences about trends in the product \( s_1 \cdot s_2 \) can be made\(^{21} \).

The data were also fitted with the EVM approach\(^{22} \) where known errors were used. These were entered as standard deviations (0.5% standard deviation in the feed composition and 1% standard deviation in \( k_p \)) in the program which were taken as 50% of the average actual fitting error in the NLLS fit in order to be able to make a comparison in the size of the JCl's.

When comparing the JC1 with approximate shape (linearized) joint confidence intervals using the EVM approach\(^{22} \) with the actual JCl's (exact shape, approximate size) there are larger deviations in shape because the used equation for \( k_p \) (Eq. (5)) is less linear than the differential copolymerization equation (Eq. (4)).

The difference in size stems from the fact that in the NLLS method the F-distribution (Eq. (2)) is used where unknown errors are assumed whereas in the EVM approach the \( \chi^2 \)-distribution is used (Eq. (3)) where known errors are used. Furthermore the ellipsoid suggests a much higher accuracy for \( s_2 \) than is justified!

**Conclusion**

The choice of the statistical method to determine the reactivity ratios in copolymerization modeling is shown to be very important. Problems in literature seem to be that proper weighting is not always conducted and that the construction of the joint confidence intervals can be done in different ways. When using available computer programs the treatment of the entered errors in the measurements is not always clear and can lead to errors in the size of the joint confidence intervals. In the case of non-linear functions the use of joint confidence intervals with exact shape is preferred. In the case that the independent variable is errorless, it is shown that the method of visualization of the sum of squares space and constructing the joint confidence intervals in this space has advantages over the other methods applied in copolymerization modeling in the sense that the method is easy to understand and implement in simple computer programs and also that important aspects of the error structure of the data can easily be incorporated. The knowledge of the error structure in the measurements is shown to be very important and the error structure determines the method of weighting the data in NLLS fitting procedures\(^{23} \). The shape of the joint confidence intervals is not always ellipsoidal and following a normal distribution, depending on the linearity of the fitted equations\(^{24} \).

**Acknowledgement:** The authors want to acknowledge Bart Manders and Mark Peeters for making their copolymerization data available as a model system for these calculations and also for useful suggestions and Eric Nies for applying the EVM method on these copolymerization data.
A. M. van Herk, T. Droge

14) H. N. Linssen, Statistica Neerlandica 29, 93 (1975)
18) B. G. Manders, M. F. Peeters, A. M. van Herk, to be published
19) T. Fukuda, Y.-D. Ma, H. Inagaki, Macromolecules 18, 17 (1985)
22) A program described in “L.T.M.E. Hillegers, PhD Dissertation, Eindhoven University, The Netherlands (1986)” was used to apply the EVM approach on Eq. (5)
23) A computer program implementing the principles outlined in this publication is available from the first author
24) A recent paper coming to similar conclusions is: A. Kaim, P. Oracz, Macromol. Theory Simul. 6, 565 (1997)