Copolymerization Reactivity Ratio Inference: Determining Confidence Contours in Parameter Space via a Bayesian Hierarchical Approach

Robert Reischke*

Dr. Robert Reischke
Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
Email Address: reischke@astro.ruhr-uni-bochum.de

Keywords: reactivity ratios, confidence ellipses, confidence contours, parameter space, copolymerization kinetics, parameter estimation, polymer reaction engineering

Confidence contours in parameter space are a helpful tool to compare and classify determined estimators. For more intricate parameter estimations of non-linear nature or complex error structures, the procedure of determining confidence contours is a statistically complex task. For polymer chemists, such particular cases are encountered in determination of reactivity ratios in copolymerization. Hereby, determination of reactivity ratios in copolymerization requires non-linear parameter estimation. Additionally, data may possess (possibly correlated) errors in both dependent and independent variables. A common approach for such non-linear estimations is the error-in-variables model yielding statistically unbiased estimators. Regarding reactivity ratios, to date published procedures neglect the non-Gaussian structure of the error estimates which is a consequence of the non-linearity of the model. In this publication, this issue is addressed by employing a Bayesian hierarchical model, which correctly propagates the errors of all variables. We detail the statistical procedure in chemist friendly language to encourage confident usage of our tool. Our approach is based on a Python program requiring minimal installation effort. A detailed manual of the code is included in the appendix of this work, in an effort to make this procedure available to all interested polymer chemists.

1 Introduction

Determination of confidence ellipses in parameter space is a helpful tool to compare and classify the determined estimators. By visualization of the confidence intervals in terms of the estimators, different evaluation procedures (models) may be compared and judged according to their overlap in 2D-parameter space. Overall, the complexity of determining confidence ellipses in parameter space is a function of the respective model (M) and the error structure of the data. Hereby, model denotes the type of function to be used in the fitting procedure (e.g. linear, polynomial, non-linear). The error structure of the data must also be accounted for, as data points may need to be weighted differently.

The procedure of determining confidence ellipses for non-linear estimations is a statistically complex task. For polymer chemists, such models are relevant in determining copolymerization reactivity ratios. Here, the non-linear Mayo-Lewis Terminal-model (ML) [1] is most often employed. It is given by

\[
F_2(f_2; r_1, r_2) = \frac{r_2 f_2^2 + f_2(1 - f_2)}{r_2 f_1^2 + 2 f_2(1 - f_2) + r_1(1 - f_2)^2} \equiv M(f_2; r_1, r_2),
\]

where \( F_i \) and \( f_i \) are the mole fractions of the \( i \)-th component in the bulk and the (co-)polymer respectively. The quantities \( r_i \) are the reactivity ratios of each propagating chain end. They are defined as the ratio \( r_1 \equiv k_{11}/k_{12} \) and \( r_2 \equiv k_{22}/k_{21} \). Here, \( k_{ij} \) are the propagation rate constants of the respective radical and monomer combination. Since equation (1) is a non-linear equation (in the parameters \( r_1 \) and \( r_2 \)), parameter estimation techniques via linear regression are prone to fail, yielding biased results. Furthermore, the data used to estimate the reactivity ratios, pairs \( (f_2, F_2(f_2)) \), have errors both in the dependent, \( F_2 \), and independent \( f_2 \) variable. In addition, the errors may be correlated depending on the setup of the polymerization experiment. Using the correct statistical technique is therefore of paramount importance to obtain unbiased estimates of reactivity ratios. Including errors in the independent variable and incorporating the non-linearity of the problem can be done elegantly with the error-in-variables-method (EVM) [2, 3]. In 2018, Scott et al. [4] used this method to provide an accessible EVM framework by providing a ready-to-use computational package.
While the EVM model provides unbiased parameter estimates, the errors calculated in [4] are still assumed to be entirely described by the covariance of the parameters in question. While this is asymptotically true, weakly constraining data (e.g. few data points, or data collected in a region of the data space which is not sensitive to the parameters) might render this assumption false. In particular, the non-linearity of the problem always induces non-Gaussianities in the Bayesian posterior. Therefore, the presentation of the errors as symmetric 1σ uncertainties does not show the whole picture. The same is true when plotting confidence intervals as ellipses from the parameter covariance. This is especially important for small reactivity ratios which have an $r_i$ value close to 0. Due to the constraints that $r_i \geq 0$, the probability is shifted to higher values. The described approach does not take this into account and therefore, underestimates the error in higher value areas.

Following this situation in current literature, this publication aims to shed light on advantages and disadvantages of the published methods as well as introducing a new approach. An open source code will be provided with this work in order to allow the access to a user-friendly Python Script to enable polymer chemists to review their own as well as published reactivity ratio data with a state of the art approach.

We will study the particular case of equation (1), but the framework can be used for any model, provided the user gives a prescription of the model in terms of a Python function depending on its parameters, provides the data and its covariance.

This publication is structured as follows.

Firstly, an overview of the published approaches in determination of confidence contour determination provided in section 2. This includes a comparison of our newly proposed approach with the available methods.

Secondly, section 3 contains the detailed mathematical foundation of the proposed concept for determination of confidence contours for reactivity ratios. As this section is quite technical, a chemist friendly translation is provided in appendix A.

Thirdly, a case study and proof of concept is conducted in section 4. Hereby, published confidence contours of copolymerization data-sets are recreated and enhanced further via our tool. Deviations are discussed in regard to the assumed error structure (gaussian vs. non-gaussian). Fifthly, section 5 contains more information on the general usage of the code. A user-manual for the code is provided in appendix B.
2 Classification of Approach

In literature, several methods of estimating confidence intervals for reactivity ratios have been published during the last decades [5, 4]. These methods are based on different mathematical approaches and strongly differ in applicability, complexity of the calculation procedure as well as availability. A classification of all approaches may be done by comparing their validity for different error structures and models. Firstly, the error structure of \( f_2 \) and \( F_2 \) is crucial. Here, it is necessary to distinguish between three different error structures: (i) an error on the dependent variable \( (F_2 \pm \Delta F_2) \), (ii) errors on both variables \( (F_2 \pm \Delta F_2 \& f_2 \pm \Delta f_2) \) and (iii) correlated errors on both variables \( (F_2 \pm \Delta F_2(f_2) \& f_2 \pm \Delta f_2(F_2)) \).

Secondly, approaches may be grouped according to their validity in regard to linear and non-linear equations. To give an overview regarding the consideration of error structures and equation types, the two published approaches of Van Herk et al.[5], Scott et al.[4] and the newly presented approach of this work are compared in table 1.

Table 1: Overview and comparison of published approaches in determination of confidence contours for reactivity ratios in parameter space. Presented are the considered error structures (see text) as well as the validity for different types of equations. Hereby, ✓ and × denote whether a certain is considered.

<table>
<thead>
<tr>
<th>considered aspects</th>
<th>Van Herk et al.[5]</th>
<th>Scott et al.[4]</th>
<th>present work</th>
</tr>
</thead>
<tbody>
<tr>
<td>errors in x</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>errors in y</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>correlated errors</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>applicable on linear equations</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>applicable on non-linear equations</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>

The method presented by Van Herk et al. may be used to infer reactivity ratios for data sources with Gaussian errors in the \( F_2 \)-value, while errors in the \( f_2 \) are neglected. Hereby, van Herk relies on calculation and investigation of the sum-of-squares. Although the applicable error structure is limited to y-errors \( (F_2 \pm \Delta F_2) \), the approach is highly valued among polymer chemists due to its comprehensibility and feasible self-implementation with basic programming skills.

The approach presented by Scott et al. is more advanced with consideration in errors in dependent and independent variable. It additionally allows for correlation in errors. In comparison to van Herk, the underlying mathematics are much more intricate and are not comprehensible for standard-chemists who do not possess an in depth knowledge of statistics.

Although the approach of Scott et al. is a more advanced approach in determination of confidence contours in reactivity ratios, it does possess certain disadvantages. Hereby, a Taylor expansion is employed in error determination which assumes a Gaussian likelihood. By calculating the second derivative with respect to the parameters of the logarithmic likelihood function, the Gaussian errors for the parameters may be determined. For linear functions, this approach perfectly depicts reality as all higher derivatives are zero. However, this procedure leads to a misrepresentation of errors for non-linear functions since the Taylor expansion does not end with the second derivative. The consideration of higher derivatives is not necessarily helpful for solving the problem as a normalization for the function’s integral is no longer possible.

Coming from this complication, another approach is employed in this work which is related to the approach of van Herk as it relies on screening contour lines. However, this screening is conducted for the likelihood density function (in opposition of the sum-of-squares approach by van Herk). Via the normalized likelihood density function, a contour for a certain probability may be numerically calculated by varying the integration limits. By this, non-Gaussianities may be properly represented by the shape of the confidence intervals. The pathway of the approach presented in this work is schematically shown in figure 1 (see ).

This is especially of importance, if the errors are asymmetrically distributed. This may be the case if values (and errors) are correlated for both variables (e.g. if \( F_2 \) is calculated from \( f_2 \) and not measured separately via spectroscopy). With asymmetrical errors, the standard \( \chi^2 \) formalism can not be as easily applied for the error propagation. In order to fill this gap and to consider such cases, the presented approach accounts for both errors in \( f_2 \) and \( F_2 \), while also allowing for a correlation between the two and being applicable on non-linear equations.
The tool presented here is capable to estimate the reactivity ratios and their respective errors to high fidelity for all three error-structures as long as the model can be linearized in the order of magnitude of the error bars. Given this, the likelihood of the data can be determined to be a Gaussian with a modified covariance matrix (Figure 1, step 1). Using Bayes theorem, the likelihood of the data \( f_{2,i} \) and \( F_{2,j} \) can be connected to the probability of the parameters \( r_1 \) and \( r_2 \). For more information on Bayes theorem, marginal and conditional probabilities we refer to appendix A.3. The reactivity ratios are estimated using a maximum likelihood estimator. While the distribution of data was determined by the Gaussian likelihood, the probability distribution of the parameters \( r_1 \) and \( r_2 \) can be non-Gaussian. Therefore, confidence intervals are not necessarily ellipses and should consequently be referred to as confidence contours in parameter space.

### 3 Inference of Reactivity Ratios

In this section, we will describe the inference of the reactivity ratios and how the likelihood (denoted \( \Phi \) by Scott et al. [4]) marginalized over the independent variable can be written in a closed form. The expressions here follow closely those presented in [6]. We will be quite specific to the problem at hand given in equation (1), but everything is applicable to any other model without loss of generality. This section is very technical and we provide some chemist friendly illustrative examples and a more detailed explanation in appendix A.

#### 3.1 Data Structure

Considering observed data, consisting of \( n \) measurements \( f_{2,i} \) and \( F_{2,j} \) with \( i = 1, ..., n \), these \( n \) measurements are bundled into the vectors \( f_2 = (f_{2,1}, ..., f_{2,n})^T \) and \( F_2 = (F_{2,1}, ..., F_{2,n})^T \). Furthermore, assuming both \( f_2 \) and \( F_2 \) being Gaussian distributed around the (un-observable) true data values, which we label \( f_2^t \) and \( F_2^t \). Providing the full covariance matrix of the stacked data vector fully describes the distribution:

\[
C = \begin{pmatrix}
C_{f_2 f_2} & C_{f_2 F_2} \\
C_{F_2 f_2} & C_{F_2 F_2}
\end{pmatrix},
\]

where the \( C_{ij} \) are the individual covariance matrices defined as

\[
C_{f_2 f_2} = E[(f_2 - f_2^t) \otimes (f_2 - f_2^t)]
\]

\[
= \begin{pmatrix}
(f_2 - f_2^t)_1(f_2 - f_2^t)_1 & \cdots & (f_2 - f_2^t)_1(f_2 - f_2^t)_n \\
\vdots & \ddots & \vdots \\
(f_2 - f_2^t)_n(f_2 - f_2^t)_1 & \cdots & (f_2 - f_2^t)_n(f_2 - f_2^t)_n
\end{pmatrix}
\]
and similarly for the others. Here, \( E[ \cdot ] \) denotes the expectation value and \( \otimes \) the tensor product, that is, in this case, multiplying the \( i \)-th with the \( j \)-th component of the vector \( (f_2^i - f_2^j) \) to obtain a matrix. We know that the true model satisfies \( F_2^i = \mathcal{M}(f_2^i, r_1, r_2) \) subject to equation (1). Finally, a parameter vector \( \theta = (r_1, r_2)^T \) is defined for less clutter, in general this would be a \( d \)-dimensional vector. For the case of reactivity ratio estimation, \( d = 2 \) as we aim for determine two estimators \((r_1, r_2)\).

### 3.2 Bayesian Hierarchical Model

In order to describe the degree of belief of the values of the parameters \( \theta \) being close to the true once the posterior distribution, \( p(\theta|f_2, F_2) \), is required. By Bayes theorem, it is given by

\[
p(\theta|f_2, F_2) = \frac{L(f_2, F_2|\theta)p(\theta)}{p(f_2, F_2)} ,
\]

with the likelihood \( L(f_2, F_2|\theta) \), the prior \( p(\theta) \) and the evidence \( p(f_2, F_2) \). The evidence can be absorbed in the normalization of the posterior since it does not depend on the model. Note that the quantities in equation (4) are probability densities since the random variables are continuous, see appendix A for more details. If not mentioned otherwise we will assume a flat prior with support in \( r_1, r_2 \in (0, \infty) \).

The crucial part is to find the proper likelihood. In the presence of unknown independent variables, it is given by (up to a normalization)

\[
L(f_2, F_2|\theta) \propto \int \int p(f_2, F_2, f'_2, F'_2|\theta) \, df'_2 \, dF'_2 ,
\]

that is, we marginalize over the latent variables (the true values related via the model). It is this likelihood which is maximized in [4] by iterations through the outer and inner loop, that is finding best fit value (the maximum likelihood estimator (MLE)) for given independent variables and then find the MLE for the independent variable. This procedure is repeated until converged. One could set up a full hierarchical model by promoting the independent true variables to parameters and sampling from the full joint distribution function with Markov chain Monte Carlo (MCMC) methods. We would like to stress that this would be the correct way to proceed and that the MLE found in [4] will be the same (the posterior distribution would still be non-Gaussian though). However, one can simplify the problem dramatically with a few assumptions which are very well satisfied in the inference of reactivity ratios.

Let us go back to equation (5): the integrand is the full joint distribution of the data and the true values given the model. Recall that each integration measure is \( n \)-dimensional: \( df'_2 = df'_{2,1} \ldots df'_{2,n} \). This expression can be rewritten in terms of conditional probabilities using Bayes theorem:

\[
L(f_2, F_2|\theta) \propto \int \int p(f_2, F_2|f'_2, F'_2, \theta)p(F'_2|f'_2, \theta)p(f'_2|\theta) \, df'_2 \, dF'_2 .
\]

Since the model, given by equation (1), must be satisfied for the true values we can write

\[
p(F'_2|f'_2, \theta) = \delta_D(F'_2 - \mathcal{M}(f'_2, \theta)) ,
\]

with the Dirac delta function \( \delta_D \) defined in its distributive form:

\[
f(x) = \int f(y)\delta_D(x - y) \, dy ,
\]

for any function \( f \). This makes the \( F'_2 \) integration trivial, leading to:

\[
L(f_2, F_2|\theta) \propto \int p(f_2, F_2|f'_2, \mathcal{M}(f'_2, \theta), \theta) \, df'_2 ,
\]

where the prior for \( f'_2 \) was assumed to be flat (this could be generalized, but this is not considered in the work here). Apart from the last assumption, everything has been completely general. In order to solve the remaining integral, we need to make a two crucial assumptions:
3.3 Confidence Intervals on Non-Gaussian Parameter Spaces

(i) The likelihood, \( p(f_2, F_2|f_2', M(f_2, \theta), \theta) \) is a Gaussian.

(ii) The model can be linearized in the vicinity (within the errors) of the observed values \( f_2 \):

\[
F_2 = M(f_2, \theta) \approx M(f_2, \theta) + G(f_2)(f_2' - f_2),
\]

where the components of \( G \) are given by

\[
(G(f_2))_{ij} = \frac{\partial(F_2)}{\partial(f_2')}_{f_2'=f_2}.
\]

The remaining integration is then merely a standard Gaussian integral. After a lengthy calculation, one finds that the new likelihood is given by a Gaussian with a modified covariance matrix:

\[
L(f_2, F_2|\theta) \propto \frac{1}{\sqrt{\det R}} \exp \left[ -\frac{1}{2} (F_2 - F_2')^T R^{-1} (F_2 - F_2') \right],
\]

where the new covariance is given by

\[
R = C_{F_2F_2} - C_{F_2f_2}^T G^T - GC_{f_2F_2} + GC_{f_2f_2} G^T,
\]

exactly propagating the errors into the new covariance by virtue of Gaussian error propagation. The MLE is now inferred by maximizing equation (12) by any standard technique, removing the necessity of the inner loop in [4]. Note that the covariance now depends on the model parameters via \( G \).

3.3 Confidence Intervals on Non-Gaussian Parameter Spaces

Estimating the MLE, \( \hat{\theta} \) from equation (12) is a piece of cake. Estimating the error is, however, a slightly different story. As mentioned in the introduction, the parameter space is inherently non-linear due to the model and therefore necessarily features non-Gaussianities. Errors on parameters are commonly quoted via the covariance between those parameters which itself can be obtained by the inverse of the Hessian (Fisher matrix, \( F \)) of the log-likelihood:

\[
(C^{-1})_{\alpha\beta} = E \left[ \frac{\partial^2 \log L}{\partial \theta^\alpha \partial \theta^\beta} \right] \equiv (F)_{\alpha\beta},
\]

where \( \alpha, \beta = r_1, r_2 \) label the parameters. This is, however, only a lower bound on the error by virtue of the Cramér-Rao bound which states that errors are bounded from below by the Fisher information matrix. The equality holds for a Gaussian posterior (or an affine parameter space) and it converges to the equality as \( 1/n \), where \( n \) is the number of (independent) measurements taken. Thus, if there are only few data points available, non-Gaussian features in the posterior become present. Furthermore, the limitation \( r_{1,2} > 0 \) will skew the posterior towards larger parameter values.

If we normalize \( p(\theta|f_2, F_2) \) a 100q%-confidence intervals are then given by the contour \( C \) of \( p(\theta|f_2, F_2) \) satisfying:

\[
q = \int_C p(\theta|f_2, F_2) \, d\theta.
\]

An alternative way to compute the confidence contours is MCMC which generates directly samples from the posterior \( p(\theta|f_2, F_2) \). This method makes marginalization much easier and scales well with the dimensionality of the parameter space. For parameter spaces with low dimensionality, as in the example studied here, we will directly integrate over the posterior via equation (15).

4 Case Studies

4.1 Comparison with previous work

In this part we will compare our code to the results obtained by other methods to provide a proof of concept. Hereby, we employ published copolymerization data[7, 8], for which confidence contours have been determined
by van Herk et al.[5] and Scott et al.[4] respectively. In total, three different analysis are carried out whose data pairs are labeled A, B and C in table 3. Hereby, case A is an individual copolymerization experiment, while C and B refer to the same experiment. B and C merely denote a switch in the reference monomer (monomer 1 or monomer 2), in order to prove the validity of our approach in comparison to Scott et al. The errors used for the analysis are constant relative errors, in accordance with the previously published works. In the dependent variable \((F_i)\), we assume \(\sigma_{y,(A,B,C)}/y = 0.05\). For the error in the independent variable \((f_i)\), we neglect the error for the comparison to van Herk et al., but consider an error of \(\sigma_{x,(A,B)}/x = 0.01\) and for comparison to contours derived by Scott. This is due to the different considerations of the approaches, as van Herk does not account for errors in independent variables (see table 1).

Table 2: Data used in section 4 to infer reactivity ratios and the respective errors. Hereby, A denotes an individual copolymerization data set. B and C refer to the same data set, but the respective reference monomer is switched. To analyze the data we assume constant relative errors of 5% on the dependent variable \((F_i)\). For the independent variable \((f_i)\) 0% is assumed for the analysis to be compared to van Herk et al.[5] and 1% for the analysis and comparison of Scott et al.[4] This difference in error structure in the independent variable is reasoned in the respective model assumptions (see table 1).

<table>
<thead>
<tr>
<th>reference observable case</th>
<th>[7]</th>
<th>[8]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f_2)</td>
<td>(F_2)</td>
<td>(f_1)</td>
</tr>
<tr>
<td>A</td>
<td>0.1</td>
<td>0.059</td>
</tr>
<tr>
<td>A</td>
<td>0.3</td>
<td>0.243</td>
</tr>
<tr>
<td>A</td>
<td>0.4</td>
<td>0.364</td>
</tr>
<tr>
<td>B</td>
<td>0.5</td>
<td>0.486</td>
</tr>
<tr>
<td>B</td>
<td>0.5</td>
<td>0.585</td>
</tr>
<tr>
<td>B</td>
<td>0.7</td>
<td>0.721</td>
</tr>
<tr>
<td>B</td>
<td>0.8</td>
<td>0.824</td>
</tr>
</tbody>
</table>

The results of the van Herk et al. and Scott et al. approach for the respective reactivity ratios of A, B and C as published by the authors are listed in table 3.

Table 3: Published reactivity ratios from data in table 3 by van Herk et al. and Scott et al. via their respective approach, the specified errors are included.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(r_1)</td>
<td>1.71 ± 0.19</td>
<td>1.19</td>
</tr>
<tr>
<td>(r_1)</td>
<td>1.95 ± 0.12</td>
<td>1.87</td>
</tr>
</tbody>
</table>

In table 4 the results of the analysis via the approach presented in this work are summarized. First, it is noteworthy that our results agree with those found by both van Herk[5] and Scott[4]. Case A serves merely as a consistency check, since the method used in [5] agrees with the one presented here if the errors of the dependent variable vanish. In particular, they also take into account the full shape of the posterior distribution. The agreement of the results for the cases B and C, where EVM was used, is an excellent demonstration that our code works as it should. In figure 2 the best fit curves together with the data are shown. Since there is no correlation between the different data points it is clear by eye that the curves are a good fit to the data.

Regarding the determined values of reactivity ratios, everything up to this point is exactly the same as in [4]. In order to highlight the differences between our and the previously published approach one has to look at the confidence regions in the two dimensional parameter space. Table 4 lists the one-dimensional marginal errors 1-\(\sigma\) errors of the posterior distribution. This is in general a misleading statement and what it actually corresponds to is that the probability of finding the true value of the parameter in question around the MLE is 68.3%. For a one-dimensional Gaussian, this is exactly the 1-\(\sigma\) around the mean (which equals the MLE). If the underlying distribution is, however, non-Gaussian this interval changes, as does the two-dimensional contour. The results presented in table 4 therefore correspond to presenting results in the following way:

\[
   r_i = r_i(MLE) \pm \sigma_{\text{Gauss}}(r_i) \quad \text{and} \quad r_i = r_i(MLE) \pm \sigma_{\text{non-Gauss}}(r_i),
\]

for the Gaussian and non-Gaussian case respectively. Inspecting table 4 already reveals some discrepancies
between the Gaussian and the actual error. In particular the correct error intervals are always smaller to the left of the peak of the posterior and larger to the right, showing a skewness in its distribution. This trend can be seen very well in figure 3, where the confidence level \( q = 68.3 \), 95.4, 99.7% contours (compare with equation (15), corresponding to 1-\( \sigma \), 2-\( \sigma \) and 3-\( \sigma \) for a Gaussian respectively) are plotted for the exact (blue) and the Gaussian posterior (red). It should be noted that case A is again only for consistency, since in [5] already investigated the full shape of the posterior but was not able to include errors in the independent variable in the analysis. Another visible trend is that the differences between the Gaussian and exact contours becomes larger for larger values of \( q \). The reason for this is that the exact distribution has a much more pronounced tail to large values of \( r_1 \) than the Gaussian one. In case B one can see the largest differences and the corresponding error estimates differ by a factor of more than 2, even more so the Gaussian 99.7% contour lies within the 95.4% contour of the exact distribution.

![Figure 2: Best fit curves for the data (red) in table 3 (blue line) as determined via the approach of this work. The corresponding best fit values (reactivity ratios) are summarized in table 4. From left to right the figures correspond to case A, B and C.](image)

<table>
<thead>
<tr>
<th>case</th>
<th>( r_1 ) (MLE)</th>
<th>( r_2 ) (MLE)</th>
<th>( \sigma_{\text{Gauss}}(r_1) )</th>
<th>( \sigma_-(r_1) )</th>
<th>( \sigma_+(r_1) )</th>
<th>( \sigma_{\text{Gauss}}(r_2) )</th>
<th>( \sigma_-(r_2) )</th>
<th>( \sigma_+(r_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.71</td>
<td>1.95</td>
<td>0.19</td>
<td>0.18</td>
<td>0.23</td>
<td>0.12</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>B</td>
<td>1.72</td>
<td>1.01</td>
<td>0.46</td>
<td>0.24</td>
<td>0.97</td>
<td>0.23</td>
<td>0.12</td>
<td>0.47</td>
</tr>
<tr>
<td>C</td>
<td>1.19</td>
<td>1.88</td>
<td>0.15</td>
<td>0.14</td>
<td>0.16</td>
<td>0.064</td>
<td>0.061</td>
<td>0.067</td>
</tr>
</tbody>
</table>

Table 4: Reactivity ratios obtained for the three cases listed in table 3 via the proposed approach of this work. As specified in the text, two sets of errors are provided for each reactivity ratio. Firstly, for the Gaussian case only a single limit is provided due to the symmetry of the distribution \( \sigma_{\text{Gauss}}(r_i) \). Secondly, the errors are given as the 68.3% contours assuming a Gaussian posterior distribution and taking its exact shape into account. Hereby, a lower and an upper error, \( \sigma_- \) and \( \sigma_+ \), are provided respectively for each reactivity ratio. Any \( \sigma \) in this table corresponds to the boundary of the one-dimensional interval which contains 68.3% probability of the one-dimensional marginal posterior distribution. For more information on this we refer to appendix A.

It was already noted in [4] that there is an asymmetry in cases B and C. As mentioned they do contain the same data just with a different identification of the comonomer \( M_1 \), in particular \( f_1 = 1 - f_2 \) and the same for \( F_2 \). While the results are consistent within the statistical error bars (note that \( r_1 \) has to switch place with \( r_2 \) as well) they should be equal since the same data was used. The reason for the remaining difference is the assumption of constant relative error, yielding larger errors for case B and therefore weaker constraints on the parameters. In case of constant absolute errors, both would yield the same result.

Lastly it should be noted that we do not intend to question the validity of the model, equation (1), when fitting the data used here. This question was already extensively addressed in [4] finding that the difference for the data analysed is well within the statistical error. Our procedure is, however, completely free to use the cumulative form of the copolymer composition model (at the expense of some computational time).

### 4.2 Validity of the Procedure

In order to integrate out the latent variables, i.e. the true values of the independent and dependent variable, two assumptions were made whose validity has to be discussed.
The first assumption is that the likelihood \( p(f_2, F_2|f_2^i, \mathcal{M}(f_2^i, \theta), \theta) \) is a multivariate Gaussian. This is an assumption made for simplicity and the lack of knowledge of the actual distribution. For the distribution \( p(f_2, F_2|\theta) \) a Gaussian distribution is assumed in all analysis, since they are all based on \( \chi^2 \). Assuming that this also holds for the latent variables should thus not be of any large concern. Furthermore, the latent variables are integrated out. For the second assumption one can check its validity for the present situation by inspecting figure 2. The assumptions requires that the model can be linearised around the true value within the errors. It is evident from figure 2 that this requirement is very well satisfied. For the model at hand only relative errors of well above 20\% would cause the assumption to break down.

If one wants to drop either assumption (or both) the only proper way out would be to design a Bayesian hierarchical forward model by sampling directly from the distribution \( f_2^i \) and \( F_2 \). One could another step back and sample directly on the observables of the experiment with some assumed error distribution. This would captures the non-linearity at each stage but we would argue that the gain should be marginal as discussed above.

## 5 Code description

The code is written in Python and only requires standard packages such as numpy, scipy and matplotlib. Also a \LaTeX{} installation is required for the plots. The installation and usages is described in more detail in appendix B.

As input the code uses a configuration file were all important parameters can be specified. The main source file then just requires the user to provide the model and its derivative with respect to the independent variable as a Python function which depends on the parameters and the data. Once everything is specified the code internally maximizes equation (12) with respect to the parameters to find the MLE. At the MLE the Fisher matrix is calculated to define the Gaussian errors (step 1 in figure 1). Subsequently a root search is carried out on the function \( f(C) = 1 - q(C) \), where \( q(C) \) is given by equation (15), to find the confidence contours corresponding to \( q = 0.683, 0.954, 0.997 \). Function \( q(C) \) contains the Bayes theorem depicted in step 2 in figure 1, while the root search corresponds to step 3. The results are then used to produce figures such as figure 2 and figure 3. Additionally output files containing the data to produce these figures are generated, given the user the opportunity to self-plot the results if desired.

To compute the corresponding one-dimensional errors one needs to marginalize over the other parameter. For the Gaussian case this is trivial since the marginalization can be carried out analytically:

\[
\sigma_{i, \text{Gauss}} = \sqrt{(F^{-1})_{ii}} ,
\]

for the \( i \)-th parameter. In the general case we have to calculate the marginal distribution, \( p(\theta_i) \), of the \( i \)-th parameter

\[
p(\theta_i) = \int p(\theta)d\theta_1 \ldots d\theta_{i-1}d\theta_{i+1} \ldots d\theta_n ,
\]
where we denote the joint posterior distribution as $p(\theta)$, i.e. equation (12) as a function of the parameters. In figure 4 we show the marginal distribution for case B with the Gaussian curve in red and the exact approximation in red.

One can define the cumulative distribution function (CDF):

$$P(\theta_i) = \int_{-\infty}^{\theta_i} d\theta'_i p(\theta'_i)$$

The 1-$\sigma$ interval are then defined via the following condition

$$\sigma_- = \hat{\theta}_i - P^{-1}(0.159) \quad \text{and} \quad \sigma_+ = P^{-1}(0.841) - \hat{\theta}_i,$$

where $P^{-1}$ is the inverse CDF and $\hat{\theta}_i$ is again the MLE of the $i$-th parameter. The results are all saved in a text files.

6 Conclusion

In this work we have proposed a method to incorporate the error in the independent variable which does not rely on the EVM but instead marginalizes over the latent variables analytically. This procedure allows for normal minimization of the likelihood which assumes the standard form but with a modified covariance matrix for which all the machinery is readily available. The advantages of this procedure are the following

(i) Due to the analytic marginalization the computation is computationally not expensive. Compared to the EVM the inner loop of the iterative scheme (see [4]) is not necessary. Even though the method makes a few assumptions we have shown that they are very well satisfied for all practical purposes.

(ii) We are able to calculate non-Gaussian errors on the parameter thus taking into account the full shape of the posterior distribution. This is a feature which was studied for example in [5] (but without including errors in the independent variable) but is ignored in the literature otherwise. This can lead to biased error estimates depending on the data (see appendix A for more details).

We have implemented a Python program which uses a configuration file and which is easy to use and, considering the advantages listed above, provides the statistically most sound analysis method for reactivity ratios. With this, polymer chemists interested in classification and sound evaluation of measured or published reactivity ratio data are enabled to employ a state of the art in approach. Based on this, reliable and fundamental reactivity ratios may implemented in (co-)polymerization models to work towards depiction and better understanding of structure-property relationships in (co-)polymers. The usage of the program is described in detail in appendix B

Figure 4: One-dimensional marginal distribution for $r_1$ and $r_2$ for case B. The blue line shows the exact distribution while the red one shows the Gaussian approximation.
and it can be used on easily on any operating system with a working Python3 installation. While the examples here were carried out for the instantaneous model we also have implemented the cumulative model. The program therefore provides an alternative to the one presented in [4] while also computing the error structure correctly.

Acknowledgements

RR is supported by the European Research Council (Grant No. 770935). The author thanks Alex van Herk and Simon Harrison for valuable discussions and comments on the manuscript.

References


A Some additional statistical foundations

In this appendix we intend to provide some more details on the statistical tools used and to give a few illustrative examples for the reader who is not so familiar with statistical inference.

A.1 Relevant Statistics Vocabulary

**statistical inference**: process in statistic data analysis. It is the overarching statistical concept of this publication. In statistical inference, samples may be tested (investigated) to check whether they fulfill a certain hypothesis. For this, the sample is firstly assigned a model (here: non-linear mayo lewis), to represent the data. Secondly, a preposition (claim) may be derived via the model. In our case, the preposition is the confidence interval for the estimators (here: reactivity ratios).

**maximum likelihood estimation**: maximum likelihood estimation is another important statistical concept as applied in this work. It aims to estimate the parameters (here: reactivity ratios) of an assumed probability distribution based on observed data (measured $f_2$ and $F_2$). Hereby, the maximum likelihood estimator (MLE) is yielded, which corresponds to a set of parameters (here: $r_{1,\text{best}}$ and $r_{2,\text{best}}$) which best describes the underlying observables in terms of the assumed model (here: non-linear mayo lewis).
sample and parameter space: data space corresponds to the measured or observed variables (here: comonomer content in bulk $f_i$ and polymer $F_i$). In order to judge and compare the values obtained by regression of statistical inference, a transfer to parameter space may be helpful. Hereby, the dimensions of parameter space correspond to the estimators or fitting parameters (here: reactivity ratios). By determination of confidence contours in parameter space, models may be compared.

Bayes theorem: theory which links conditional probabilities of events. In simple terms, it returns the probability of event $A$ occurring given that event $B$ has already occurred ($p(A|B)$). This probability is equal to the probability of the switched just described case ($B$ occurs given that $A$ already occurred $p(B|A)$), scaled by the ratio of probabilities of $A$ and $B$ themselves (equation (21)).

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$ (21)

discrete and continuous variables: the definition of discrete and continuous variables in terms of statistics is related to the probability density function. For a continuous variable, the cumulative probability density function is steady. For a discrete variable, the cumulative probability density function is unsteady, e.g. due to sudden offsets. Measured or observed data (here: comonomer content in bulk $f_i$ and polymer $F_i$) is commonly assumed to be continuous.

probability density function (PDF): function of a continuous random variable in sample space. The function returns the probability or a relative likelihood of a variable possessing a certain value. In combination with a flat prior, the absolute likelihood of a variable taking on any specific value is zero. This is due to the allowance of the variable to take on any value (no values are preferred).

flat or uninformative prior: a prior contains information regarding a certain variable that we definitely know of (boundary condition). If a prior is called flat, this means that within the boundary conditions, no specific values are preferred. In our case, the prior does contain the information that the reactivity ratios cannot take on negative values. However, the reactivity ratios are allowed to possess all values above zero with the same possibility during the inference process.

informative prior/posterior: an informative prior may be understood as an updated flat prior (which did previously not contain any helpful information). This is achieved by appliance of the Bayes- theorem and the likelihood estimation to yield the posterior.

A.2 Finding the maximum likelihood

The first step of the parameter inference is to find the parameters maximizing the posterior. This corresponds to finding the best combination of estimators ($\hat{\theta}$: $r_1$ and $r_2$ in our specific case) which best describes the data set. This approach is related to well-known regression of data which is often used in chemistry. Here, the sum of squares are minimized to find the best set of fitting parameters. In the case of the likelihood distribution, we are looking for the maximum to find the best estimators ($\hat{\theta}$: $r_{1,\text{best}}$ and $r_{2,\text{best}}$). If we choose an uninformative or flat prior one can just maximize the likelihood since the evidence (the measured data $f_2$ and $F_2$), $p(D)$, does not depend on the parameters. The corresponding estimator, $\hat{\theta}$, is therefore called the maximum likelihood estimator (MLE) which are those parameters, $\hat{\theta}$ satisfying

$$\frac{\partial p(D|\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}} = 0 \ ,$$ (22)

given the data, this are $d$ equations for $d$ parameters.

The likelihood $p(D|\theta)$ is often denoted $L(D|\theta)$ and is in our case given by equation (12). In practice the logarithmic likelihood $L := -2 \ln L$ is often minimized since it is numerically much more stable. What is done technically is to run an optimizer on $L(\theta)$ to find $\hat{\theta}$, e.g. a Nelder-Mead algorithm.

The outlined procedure is in principle straightforward. However, the elephant in the room is the likelihood, how do we know how it looks like? In principle the only way to find out is to create many realizations of the same data set to obtain samples of $p(D|\theta)$. This is often unpractical and certain assumptions are made. One of the most common one is the assumption of the likelihood being a multivariate Gaussian making the life of statisticians much easier. It corresponds to least-square fitting since maximizing the log-likelihood of a Gaussian
corresponds to the sum of squares. Another motivation for the Gaussian distribution is the central-limit-theorem stating that any sum of independent and identically distributed random variables with finite variance tend to a Gaussian distribution.

### A.3 Bayes theorem, marginal and conditional probabilities

We assume that we have acquired some data \( D \), this could be a set of pairs \( \{ f_2, F_2 \} \), and a model \( M \), e.g. the Mayo-Lewis equation which depends on parameters \( \theta \). In our case, \( \theta \) symbolizes the reactivity ratios \( r_1 \) and \( r_2 \). At the heart of parameter inference stands Bayes theorem. It relates the known objects: the distribution of the data given the parameters, \( p(D|\theta) \) and the evidence \( p(D) \) to the desired quantity: the posterior \( p(\theta|D) \). The Bayes theorem relates the two quantities by successive application of the definition of conditional probabilities:

\[
p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{23}
\]

The posterior is now interpreted as a probability distribution in the parameters and it therefore describes our belief or confidence in the parameters after the data has been observed/measured. On the other hand stands the prior \( p(\theta) \) which contains any prior information about the parameters. For us, \( p(\theta) \) for example includes that reactivity ratios must be greater than zero.

If the data and the parameters can only assume discrete (and finite) values, the \( p(x) \) in equation (23) are probabilities. However, in this work the variables are continuous. Therefore, the are probability density functions (PDFs). A PDF, \( p(x) \) is assigned a probability \( dP = p(x)dx \) of finding \( x \) in the interval \( [x,x+dx] \) and one requires:

\[
1 = \int p(x)dx ,
\]

where the integral is carried out over the support of \( p(x) \). The remainder of the paper focuses on the continuous case. If the random variable \( x \) is multidimensional it is denoted in bold: \( \mathbf{x} \). This is relevant for us, as we aim to obtain confidence ellipses in 2D-space for combinations of \( r_1 \) and \( r_2 \). Consequently, the highest dimension we are concerned with in this publication is two. In this case we write the joint probability density function \( p(x) = p(x_1,...,x_d) \) for a \( d \)-dimensional vector. In this case, the joint PDF is normalized over the \( d \)-dimensional volume:

\[
1 = \int \cdots \int _{\text{d integrals}} p(x) \ dx_1 \cdots dx_d \equiv \int p(x) \ dx \tag{25}
\]

for simplicity of the notation, the \( d \) integrals are usually not written explicitly. For any joint PDF one can define the marginal PDF where one or more variables have been integrated out. For example \( p_{\text{marg}}(x_i) \) is called the marginal PDF of \( x_i \), where all other parameters have been integrated out over:

\[
p_{\text{marg}}(x_i) = \int \cdots \int _{d-1 \text{ integrals}} p(x)dx_1 \cdots dx_{i-1}dx_{i+1} \cdots dx_d . \tag{26}
\]

In contrast, the conditionalized PDF is defined such that all other parameters are kept at a certain values:

\[
p_{\text{cond}}(x_i) = p(x_{0,0},...,x_{i-1,0},x_i,x_{i+1,0},...,x_{d,0}) . \tag{27}
\]

The conditionalized PDF can therefore be thought of as a through the joint PDF.

Coming from this general description of the Bayes theorem, the approach is easily transferred to the specific case of reactivity ratio estimation.

In terms of estimating reactivity ratios and their confidence intervals, equation (23) is transformed to equation (28) (which is also shown in the technical section 3.2). Hereby, the data \( D \) is defined by the comonomer contents \( (f_2, F_2) \) and the estimator \( \theta \) corresponds to the set of reactivity ratios. The variables are denoted bold as they consist of at least two-dimensional vectors.

\[
p(\theta|f_2, F_2) = \frac{L(f_2, F_2|\theta)p(\theta)}{p(f_2, F_2)} , \tag{28}
\]
A.4 Estimating the uncertainties of the MLE

Once the MLE, \( \hat{\theta} \), \( r_{1,\text{best}} \), and \( r_{2,\text{best}} \), has been found, one needs to specify the uncertainties on the estimator. As stated before, for linear parameters, this amounts to Gaussian error propagation and it suffices to provide the covariance matrix of the parameters which can be turned into \( n - \sigma \) contours. If the model is non-linear the problem becomes more involved. It is instructive to recall the meaning of the confidence interval or the coverage probability: A confidence interval \( I \subseteq \mathbb{R}^d \) of the posterior distribution \( p(\theta|D) \) gives the probability

\[
P = \int_I p(\theta|D) \, d\theta, \tag{29}
\]

that the true value lies in this region of parameter space. This of course provides that the model is correct and that there are no systematic effects biasing the MLE. Phrased differently, if you repeat the same experiment as before the value will fall into the confidence interval with the probability \( P \). In practice the conditional error is basically never used and one either quotes one- or two-dimensional marginal errors. Thus one integrates out all but one (two) parameters as in equation (26) and calculates \( I \) on the one(two)-dimensional subspace. The next two sections deal with confidence intervals in the Gaussian and the general case.

Table 5: \( \chi^2 \) for joint parameter contours for one and two parameters [9].

<table>
<thead>
<tr>
<th>contour</th>
<th>( P )</th>
<th>( d=1 )</th>
<th>( d=2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1( \sigma )</td>
<td>68.3%</td>
<td>1.00</td>
<td>2.30</td>
</tr>
<tr>
<td>2( \sigma )</td>
<td>95.4%</td>
<td>4.00</td>
<td>6.17</td>
</tr>
<tr>
<td>3( \sigma )</td>
<td>99.7%</td>
<td>9.00</td>
<td>11.8</td>
</tr>
</tbody>
</table>

A.5 Confidence intervals for a Gaussian posterior distribution

Assume that the posterior distribution is well approximated by a multivariate Gaussian distribution. In this case, the logarithmic likelihood assumes a \( \chi^2 \) form:

\[
\chi^2(\theta) = (\theta - \hat{\theta})^T C^{-1}_{\theta} (\theta - \hat{\theta}), \tag{30}
\]

where \( C^{-1}_{\theta} \) is the inverse parameter covariance matrix of the aforementioned one(two)-dimensional subspace and \( \theta \) are the corresponding parameters in this subspace. As an example: if three parameters are given and one is interested in the first and last parameter, the second row and column have to be removed from the parameter covariance. A few probabilities enclosed in the interval and their corresponding \( \chi^2 \) values are listed in table 5 from which it is clear that the one-dimensional case corresponds exactly to the \( \sigma \)-contours, while ellipses in two dimensions need to be stretched due to the volume factor. It remains to specify the covariance matrix which is given by the inverse of the Fisher matrix as defined in equation (14). Finally, equation (30) shows that two-dimensional confidence intervals in the Gaussian case are always ellipses or circles.

A.6 Confidence intervals in the general case

If the posterior distribution it not Gaussian, one cannot write down equation (30) as it is. An instructive way to approach non-Gaussianities is to look at the Taylor expansion of the expected logarithmic likelihood, \( L \) at \( \hat{\theta} \):

\[
E[L(\theta)] = E[L|_{\theta=\hat{\theta}}] + \sum_{\alpha, \beta=1}^{d} E \left[ \frac{\partial^2 \log L}{\partial \theta^\alpha \partial \theta^\beta} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})_\alpha (\theta - \hat{\theta})_\beta + \ldots \tag{31}
\]

The expectation value is applied to make the expression independent of the data. Note that the first term in the Taylor expansion vanishes due to equation (22) and that the zeroth order term is an irrelevant constant which can be absorbed in the normalization of the posterior distribution. In case \( L \) assumes the form of equation (30) as it is for a Gaussian distribution, the series expansion would stop after the second order. If, however, the underlying distribution is non-Gaussian higher order terms can become important. In fact, the Gaussian distribution is
the only distribution where only the second order term is non-vanishing, all other distributions have an infinite amount of other terms in their respective Taylor expansions. Though these additional contributions might be small.

To find the one(two)-dimensional confidence intervals in the general case there are two possibilities (i) directly integrating the posterior or (ii) sampling from the posterior distribution with MCMC (Markov chain Monte Carlo). Since the problem at hand is low dimensional \( d = 2 \) method (i) will be used. The details are described in section 5.

B Usage of the code

B.1 File Structure

Once downloaded and unpacked, the data structure of the tool is visible: it is set up in three main folders. Hereby, the folder data contains the user-specific data to be processed via the tool. For the main example of this publication, this corresponds to defining \( f_2 \) and \( F_2 \) data points. Obviously, the folder output stores the results generated by the tool after its appliance onto the user-defined input data. Here, the user finds automatically generated figures as well as the results in .txt files. Lastly, the third folder source contains the code of the tool as well as the input.config file in which a user is able to specify important aspects such as error structures or confidence levels.

B.2 Installation

The tool is easy to install, you require a running Python3 installation. The package requirements are automatically handled by the included file setup.py which can be used with pip install ., to install all dependencies of the code. Otherwise one can always manually install the required packages which are numpy, scipy and matplotlib.

B.3 Three-Step Usage

Once the installation is done, the code is used in three steps:
1. A set of data has to be provided by the user in the input folder.
2. Define parameters in input.config in the source folder.
3. Execute code by running python main.py.

B.4 Set-up of the Configuration File

The input.config is crucial in appliance of the tool, as it allows a user to define the type of analysis to be conducted. Within the file, rows starting with a hash (#) are comments providing explanation of the needed input. In the latter file one needs to specify the path where the data is stored which is used for the fit, particularly the code requires the data pairs \((X, Y)\) in a file with 2 columns and \(N\) rows, where \(N\) is the number of data points. Additionally the errors on the two variables need to be specified, this can be done by providing files containing the components of the \(N \times N\) covariance matrix for \(C_{XX}, C_{YY}\) and \(C_{XY}\) respectively. Alternatively one can just specify constant absolute or relative errors. The generated output may defined within the configuration file in terms of the desired filename as well as colors or axis labels on the generated plots.

B.5 Output Files

After execution, the code produces plots with the 2D contours using the Fisher approximation as well as the full shape of the posterior. Similarly the 1D marginal posterior distributions are plotted and the best fit curve with the data.

The data to produce the plots are also stored in files, so that a user is enabled to plot the results themselves. In particular the full 2D-posterior distribution is stored on a three parameter grid in a text file. The grid may be
plotted as a 2D-contour or a 3D-plot (e.g. in OriginLab). The respective confidence contours may be extracted by defining the respective confidence levels provided in the second text output file.