Reliability in Kinetic Evaluation of Environmental Metabolism Data - Assessment and the Influence of Model Specification

Johannes Ranke (University of Bremen and Eurofins Regulatory AG), René Lehmann (Federal Environmental Agency, Germany)

contact: jranke@uni-bremen.de
JohannesRanke@eurofins.com
Rene.Lehmann@uba.de

In kinetic modeling, rate constants and formation fractions are often fitted directly to observed data applying statistical methods that depend on the assumption of normal distribution of the estimators. As rate constants and formation fractions can only take on positive values and formation fractions are restricted to a sum of 1, this assumption is not satisfied which induces bias to statistical tests and assessments of parameter uncertainty. To solve this problem kinetic models can be reparameterised using log transformation as proposed *e.g.* by Bates and Watts (1988) for rate constants and using isometric log ratio (ILR) transformation (Egozcue *et al*. 2003) to fit formation fractions, making use of the fact that a set of formation fractions can be treated as compositional data.

The basic model

Let $Y = f(Y_0, \theta, t) + \varepsilon$ be the model to be fitted to observed data $y_1, \ldots, y_k$ where Y_0 is the vector of initial values of the observed variables, θ is the parameter vector and ε is a multivariate random variable. For $x \in \mathbb{R}$ let $diag(x)$ be the matrix with entries x on its diagonal and 0 otherwise. In this work, we assume ε_i to be independent and identically normal distributed with expected value $E(\varepsilon) = 0_n$ and covariance matrix $V(\varepsilon) = diag(\sigma^2)$ where σ is the standard deviation of the normal distribution of ε_i.

Fitting rate constants

Kinetic rate constants can only be positive real values. However, when using least squares optimization, parameters are estimated iteratively together with their standard error and statistical significance tests are commonly applied. However, estimating a standard error supposes normal distribution. To solve this problem, kinetic rate constants k are replaced by terms of the form $\log k$‘ as proposed e.g. by Bates and Watts (1988). The estimators of the k‘ values are assumed to better fit the assumption of normal distribution.

Fitting formation fractions

Formation fractions always have values between 0 and 1. Furthermore, in fitting a single degradation experiment they always have to sum up to exactly 1. These two properties lend them to treatment as compositional data. In order to better fulfill the assumptions of least squares fitting, it is proposed here to fit formation fractions using isometric log ratio transformation (ilr) in the form proposed by Filzmoser and Hron (2008):

Let $x = (x_1, \ldots, x_D)$ be a formation fraction with $x_i > 0$ for all $i = 1, \ldots, D$ and $\sum_{i=1}^{D} x_i = 1$.

Then the ILR transform of x is denoted $z = (z_1, \ldots, z_{D-1})'$ and given by $z_i = \sqrt{\frac{i}{i+1}} \log \frac{\sqrt[i]{\prod_{j=1}^{i} x_j}}{x_{i+1}}$ for $i = 1, \ldots, D\text{-}1$.

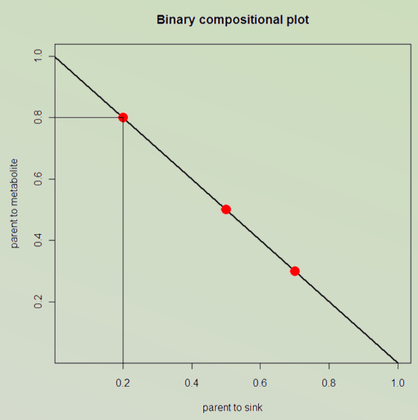
Confidence limits for model parameters

Standard optimisation routines generally provide a standard error for the parameter estimate in their output. The reparameterisation of the model aimed to obtain a symmetrical distribution of estimators that more closely follow a normal distribution. Therefore it is proposed to construct confidence intervals for the transformed parameters based on the assumption that they follow a normal distribution.

For a simple one-to-one relationship between transformed parameters and model parameters, the boundaries of these confidence intervals can simply be backtransformed for the construction of confidence intervals of the model parameters.

In the ILR transformation, parameters are not transformed one by one, but in parameter sets. In order to construct confidence intervals for the transformed parameters, it is suggested to use Monte Carlo Analysis, feeding transformed parameters as random variables with the standard error obtained by the fitting routine. Confidence intervals for the backtransformed parameters can then be obtained as percentiles of the backtransformed random variables.

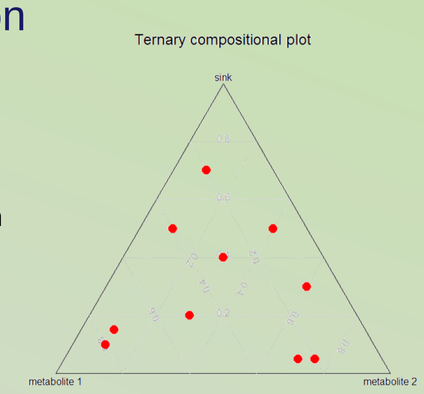
Calculation of confidence intervals is not yet implemented in mkin.



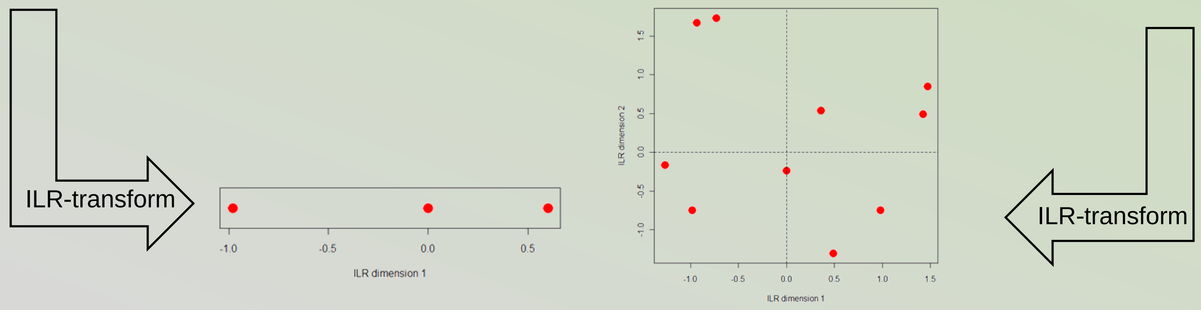
Formation fractions for flows to sink and to metabolite

ILR data representation

Formation fractions always sum up to a constant where the fraction for each pathway can only be positive. Statistical standard procedures can only be applied if the data underlie so called euclidean metric. ILR transformation is used to solve this issue.



Formation fractions for flows to sink and to two metabolites



Output from mkin 0.8.11 as published on CRAN

```
...
Starting values for optimised parameters:
             initial   type lower upper
parent_0       100.0  state     0   Inf
k_parent_sink    0.1 deparm     0   Inf
k_m1_sink        0.1 deparm     0   Inf
k_parent_m1      0.1 deparm     0   Inf

Fixed parameter values:
   value  type
m1     0 state

Optimised parameters:
              Estimate Std. Error t value   Pr(>t)
parent_0      9.960e+01  1.614e+00  61.720  < 2e-16 ***
k_parent_sink 4.792e-02  3.750e-03  12.778 3.05e-15 ***
k_m1_sink     5.261e-03  7.159e-04   7.349 5.76e-09 ***
k_parent_m1   5.078e-02  2.094e-03  24.248  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

differences

Development version mkin 0.9.2 from r-forge

```
...
Starting values for optimised parameters:
             initial   type transformed
parent_0       100.0  state  100.000000
k_parent_sink    0.1 deparm   -2.302585
k_parent_m1      0.1 deparm   -2.302585
k_m1_sink        0.1 deparm   -2.302585

...
Optimised, transformed parameters:
             Estimate Std. Error
parent_0       99.598      1.614
k_parent_sink  -3.038      0.078
k_parent_m1    -2.980      0.041
k_m1_sink      -5.248      0.136

Backtransformed parameters:
                Estimate
parent_0      99.598483042
k_parent_sink  0.047920119
k_parent_m1    0.050777613
k_m1_sink      0.005260653
...
```

Future aspects in research

Relaxing the assumption of i.i.d. error variables ε_i fitting a nonlinear mixed effects model can be used to overcome the problem of inhomogeneous error variances often observed in experiments. Inhomogeneous error variances cause serious bias in statistical tests and computation of confidence intervals if not adjusted. Allowing for different error variances for parent and metabolites has been previously proposed. However, a more general variance model would allow for a more precise description of the uncertainties of the fitting procedure.

References

Filzmoser, P.; Hron, K. (2008): *Outlier Detection for Compositional Data Using Robust Methods*; Mathematical Geosciences; 40; 233-248

Egozcue, J. J.; Pawlowsky-Glahn, V.; Mateu-Figueras, G.; Barceló Vidal, C. (2003): *Isometric Logratio Transformations for Compositional Data Analysis*; Mathematical Geology; 35; 279-300

Bates, D. M.; Watts, D. G. (1988): *Nonlinear regression analysis and its applications.* Wiley-Interscience.

Poster presented at SETAC World Meeting, Berlin, 21 May 2012