Linked Michaelis-Menten type models for methanogenesis data

"A stable isotope titration method for measurement of the contribution of acetate and carbon dioxide reduction to methane production"

By N.D. Gray, J.N.S. Matthews and I.M. Head

University of Newcastle upon Tyne

Statistical details of the analysis.

The rate of production of $^{13}CH_4$ (in μmol per cc of slurry per hour) increases with the amount of ^{13}C substrate present. The form of this dependence is proposed to follow Michaelis-Menten kinetics, i.e.

$$^{13}\mathrm{CH}_4 = \frac{B[\mathrm{S}_{\mathrm{labelled}}]}{K + [\mathrm{S}_{\mathrm{labelled}}]}$$

The rate of production of unlabelled methane, $^{12}CH_4$ is also measured and *a priori* we would expect the total production of methane to be unaffected by the concentration of labelled substrate. Consequently, we would expect the production of $^{12}CH_4$ to be governed by an equation of the form:

$$^{12}\mathrm{CH}_4 = A - \frac{B[\mathrm{S}_{\mathrm{labelled}}]}{K + [\mathrm{S}_{\mathrm{labelled}}]}$$

where A is the total production of methane.

Estimates for the parameters governing the Michaelis-Menten kinetics, B and K and of the total production of methane, A can be obtained by fitting a suitable statistical model to these data. Three models will be fitted. The details of these will be presented and the rationale for each explained.

Model I: common error, constant total methane production.

The model to be fitted to each observation is as follows.

$^{13}CH_4$ production: $$y_{ij}^{13} = \frac{Bs_i}{K + s_i} + e_{ij}$$

$^{12}CH_4$ production: $$y_{ij}^{12} = A - \frac{Bs_i}{K + s_i} + e_{ij}$$

Here s_i denotes the ith substrate concentration employed in the experiment (i=1,2,3,4, corresponding to the four concentrations used) and $y_{i1}^{12}, y_{i2}^{12}, y_{i3}^{12}$ (respectively $y_{i1}^{13}, y_{i2}^{13}, y_{i3}^{13}$) are the three rates of production of $^{12}CH_4$ (respectively $^{13}CH_4$) observed at this concentration. The e_{ij} are the residuals or error term. In this model, with

common error, the error terms are all assumed to follow independently a Normal distribution with zero mean and common standard deviation s.

The model will be fitted in a single procedure: i.e. we will not fit the $^{12}CH_4$ and $^{13}CH^4$ observations separately. This is because some of the parameters of the model (most importantly *B* and *K*) influence both sets of observations.

Model II: separate error, constant production.

The model to be fitted to each observation is more or less as before, i.e.

$^{13}CH_4$ production: $$y_{ij}^{13} = \frac{Bs_i}{K + s_i} + e_{ij}^{13}$$

$^{12}CH_4$ production: $$y_{ij}^{12} = A - \frac{Bs_i}{K + s_i} + e_{ij}^{12}$$

The difference from Model I is that the residual terms for $^{13}CH_4$ observations have standard deviation s_{13}, which is not necessarily equal to s_{12}, the value for the $^{12}CH_4$ observations. The reason for this is two-fold. First, the quantities are measured using different techniques, which do not necessarily have the same characteristics. Second, looking at the variation within the replicate measurements at each substrate concentration, the $^{12}CH_4$ values seem to be measured with less precision than are the $^{13}CH_4$ values.

Model III: separate error, non-constant production.

The model to be fitted to each observation is now

$^{13}CH_4$ production: $$y_{ij}^{13} = \frac{Bs_i}{K + s_i} + e_{ij}^{13}$$

$^{12}CH_4$ production: $$y_{ij}^{12} = A + Cs_i - \frac{Bs_i}{K + s_i} + e_{ij}^{12}$$

The difference from Model II is that the total production of methane is not now assumed to be unaffected by the concentration of labelled substrate. This is model provides a simple check on an important assumption underlying model II. The hypothesis of constant total methane production can be assessed by testing the null hypothesis that C=0.

Details of fitting etc.

Each of the models was fitted by maximum likelihood (ML) and the standard errors of the parameter estimates obtained from the expected information matrix. This is standard statistical theory. The following informal remarks are intended for those for whom this is unfamiliar methodology.

The likelihood is a mathematical function of the data and, more importantly, of the unknown parameters. So, e.g., in Model II the likelihood can be thought of as a function of the unknowns, namely A, B, K s_{12} and s_{13}, which we can write succinctly as $f(A, B, K, s_{12}, s_{13})$. This function is related to the probability (assuming the model) of observing the data that have been seen. The method of ML says that we should estimate the unknown parameters by the values which make this probability maximal.

So, the method now reduces to a mathematical problem, namely to find the values of A, B, K s_{12} and s_{13} which maximises $f(A, B, K, s_{12}, s_{13})$. In this instance we cannot find all these values by analytical means. However, if we were to fix B and K at arbitrary values, we can find the values of the other parameters analytically, albeit in terms of the arbitrary B and K. This function, call it $g(B, K)$, is known as the profile likelihood and we maximise it using numerical methods. Doing as much as you can analytically means you only have to maximise a function of two variables and this is somewhat more stable. The profile log-likelihoods are given in Appendix I.

The other aspect of ML is that the value of the likelihood (actually the log of this quantity) can under some circumstances, be used to assess the relative fit of different models. This method is used to compare models I and II and models II and III.

A further feature of the ML method is that it can be used to derive standard errors of the estimates of the parameters. This is done using the expected information matrix, which for model II is given in Appendix II.

The proportion of total methane production due to acetoclastic methanogenesis can be estimated as B/A. The standard error of this ratio can be approximated using the delta-method as:

$$\sqrt{\frac{s_B^2}{A^2} + \frac{B^2 s_A^2}{A^4} - 2\frac{B s_{AB}}{A^3}}$$

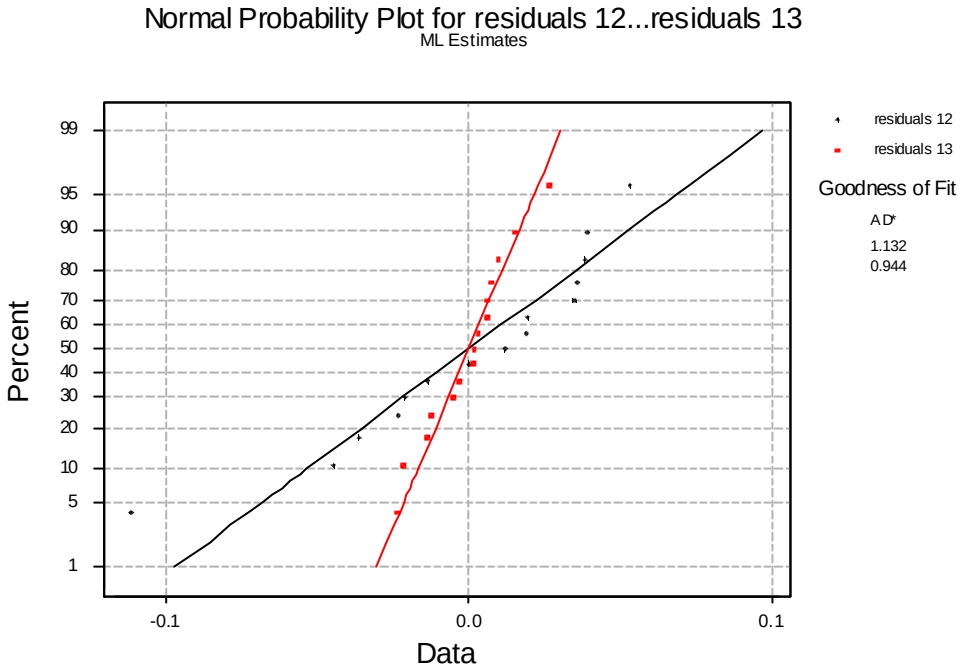
where s_A^2, s_b^2 are estimates of the variances of the estimates A and B, respectively and s_{AB} estimates the covariance between the estimates of A and B.

How well does the model fit?

This is a very broad question which encompasses many issues, including those concerned with the background theory. Here we consider only a narrow statistical assessment of the models.

An important aspect of the analysis is that the likelihoods used for the various models assume that the residuals follow a Normal distribution. For a given data point the residual is the difference between the observed rate of methane production and that predicted according to the model under consideration (based on the ML estimates of the parameters for that model). This can be assessed by looking at a Normal probability plot of the residuals. For models II and III, which assume different variances for the ^{13}C curve and ^{12}C curve, separate plots are needed for the two sets of

observations. As an illustration, the plots from the data analysed in the report are shown below. The slopes of the lines differ because the SD of the residuals are indeed quite different. The lines are reasonably straight, which indicates that the assumptions of Normality are tenable.



Appendix I:profile likelihoods

Model I.

For this model the log-likelihood is:

$$\ell(A,B,K,s) = -\tfrac{1}{2}(n_{12}+n_{13})\log s^2 - \frac{1}{2s^2}\{\sum_{i\hat{I}C^{12}}(y_i - A + Bf(s_i;K))^2 + \sum_{i\hat{I}C^{13}}(y_i - Bf(s_i;K))^2\}$$

where n_j is the number of observations on the ^{j}C curve (j=12 or 13) and $f(s;K)=s/(s+K)$.

Maximising over s gives

$$\ell(A,B,K,\hat{s}) = -\tfrac{1}{2}(n_{12}+n_{13})\log(\sum_{i\hat{I}C^{12}}(y_i - A + Bf(s_i;K))^2 + \sum_{i\hat{I}C^{13}}(y_i - Bf(s_i;K))^2)$$, up to

irrelevant constants. Maximising over A gives $\hat{A} = \sum_{i\hat{I}^{12}C}(y_i + Bf(s_i;K))/n_{12}$. So

substituting this leaves the profile log-likelihood as $\ell(\hat{A},B,K,\hat{s})$, which can then be maximised numerically over B and K.

Model II

For this model the log-likelihood is:

$$\ell(A,B,K,s_{12},s_{13}) = -\tfrac{1}{2}n_{12}\log s_{12}^2 - \tfrac{1}{2}n_{13}\log s_{13}^2 - \frac{1}{2s_{12}^2}\sum_{i\in C^{12}}(y_i - A + Bf(s_i;K))^2$$
$$- \frac{1}{2s_{13}^2}\sum_{i\in C^{13}}(y_i - Bf(s_i;K))^2$$

and maximising with respect to the residual SDs gives:

$$\ell(A,B,K,\hat{s}_{12},\hat{s}_{13}) = -\tfrac{1}{2}n_{12}\log \sum_{i\in C^{12}}(y_i - A + Bf(s_i;K))^2 - \tfrac{1}{2}n_{13}\log \sum_{i\in C^{13}}(y_i - Bf(s_i;K))^2$$

again omitting irrelevant constants. As for Model I, the estimator of A is given by $\hat{A} = \sum_{i\in^{12}C}(y_i + Bf(s_i;K))/n_{12}$, so substituting this in the above gives a function in two unknowns, B and K, which can be maximised numerically.

Model III

The calculations here are the same as above, only allowing the slightly more complicated form for the production of $^{12}CH_4$. Analytical maximisation with respect to the residual SDs and also with respect to A and C, again leaving a numerical maximisation over just two unknowns.

Appendix II: the expected information matrix.

This is the expected information matrix for Model II and is recorded here for reference.

For Model II $E(-\partial^2 \ell / \partial q \partial q^T)$

$$\begin{matrix} \frac{n_{12}}{s_{12}^2} & -\frac{1}{s_{12}^2}\sum_{C^{12}} f_i & \frac{B}{s_{12}^2 K}\sum_{C^{12}} f_i(1-f_i) \\ -\frac{1}{s_{12}^2}\sum_{C^{12}} f_i & \frac{1}{s_{12}^2}\sum_{C^{12}} f_i^2 + \frac{1}{s_{13}^2}\sum_{C^{13}} f_i^2 & -\frac{B}{K}\left[\frac{1}{s_{12}^2}\sum_{C^{12}} f_i^2(1-f_i) + \frac{1}{s_{13}^2}\sum_{C^{13}} f_i^2(1-f_i)\right] \\ \frac{B}{s_{12}^2 K}\sum_{C^{12}} f_i(1-f_i) & -\frac{B}{K}\left[\frac{1}{s_{12}^2}\sum_{C^{12}} f_i^2(1-f_i) + \frac{1}{s_{13}^2}\sum_{C^{13}} f_i^2(1-f_i)\right] & \frac{B^2}{K^2}\left[\frac{1}{s_{12}^2}\sum_{C^{12}} f_i^2(1-f_i)^2 + \frac{1}{s_{13}^2}\sum_{C^{13}} f_i^2(1-f_i)^2\right] \end{matrix} \begin{matrix} \\ A \\ B \\ K \end{matrix}$$

Here $f_i = f(s_i; K)$, where $f(s; K) = s/(s + K)$ and we have observed that $\partial f / \partial K = -f(1-f)/K$.