

DMFit manual

Version 3.5 By József Baranyi

Institute of Food Research, Norwich Research Park
Norwich NR4 7UA,
United Kingdom.

DISCLAIMER

DMFit is freely distributable shareware package, assuming that you do not modify, reverse-engineer it, and you do not build it in other application.

You are also licensed to duplicate DMFit to others, provided that you do not charge for the copy in any way. Under no circumstances may you sell or rent the program or distribute it for any fee, without the explicit written permission of the authors.

NO WARRANTY OF ANY KIND, FOR ANY USE, BY ANYONE, AT ANY TIME, FOR ANY REASON.

This product is provided on an "as-is" basis, the entire risk as to the results and performance of the software is assumed by the recipient. Neither the author, nor the Institute of Food Research, warrant, guarantee or make any representations regarding the use of, or the results obtained with, the software, in terms of correctness, accuracy, reliability, currentness, legality or otherwise. In no event shall the author or the Institute of Food Research or anyone else who has been involved in the creation, production or delivery of this product be held liable for any direct, indirect, consequential or incidental damages (including but not limited to damages for loss of profits, interruption of business, loss of information and the like) arising out of the use or inability to use this product.

If you use DMFit for studies, reports, publication, etc., acknowledge this program and the model of Baranyi and Roberts (1994) as listed in the Reference.

CONTENT

INTRODUCTION

DATA PREPARATION

VARIABLES

INPUT PARAMETERS

OUTPUT PARAMETERS

RUNNING DMFIT ON YOUR CURVES

EXAMPLE

SECONDARY MODELLING

REFERENCE

INTRODUCTION

DMFit is a Microsoft Excel Add-in, running under Windows, to fit curves where a linear phase is preceded and followed by a stationary phase. The main difference between this model and other sigmoid curves like Gompertz, Logistic, etc. is that the mid-phase is really very close to linear, unlike those classical sigmoid curves which have a pronounced curvature there.

DMFit is part of the system used in-house at the Institute of Food Research to model the time-variation of the logarithm of cell concentrations of bacterial batch cultures. (DM: Dynamic Modelling).

This document assumes that the user is familiar with the basics of Excel, as well as with the Windows operating system. The terminology and notation used here are compatible with the papers listed in the reference therefore they are not explained extensively.

DATA PREPARATION

Each measured curve must have a name. An Excel sheet must be prepared, listing all these names in its first column, possibly together with the environmental conditions under which the respective curves were measured. These conditions should be recorded in the second, third, etc. columns, parallel to their curves. The recorded conditions do not play any role in the fitting itself, but can be useful for your purposes. The prepared index sheet will be called INDEX0 by DMFit. The first line of the sheet must be a heading, as shown below.

When using DMFit, follow the next demo:

Copy the enclosed files onto your hard disk. Open DMFit3_5.XLAM. A new menupoint, DMFit will appear in the ribbon. right hand corner of the menubar. If using it first time, just type some data to be fitted in two columns of an empty sheet, then rename your sheet as you like (e.g. TRY1). Highlight your data, together with a heading first row, and choose the menupoint DMFit – Primary modelling – Fit curve defined by selection

DMFit will create a new book, with an Index sheet called Index0 and a data sheet where your curve is placed. The index sheet is divided into two parts: input and output columns. These parts are separated by a thick column border. You can read the respective column (field) names in the heading. The first column contains the names of those curves you want to fit.

When preparing more curves to be fitted, create an index sheet (as Index0 in the enclosed demo INDEX.XLSX file) and put your curves in a datasheet, as in the 'Logc' sheet in that demo. Make sure that the curve names appear in the first column and that their first character is alphabetic (not a number). The actual curves must be in the respective second and third columns as shown in the 'Logc' sheet. Note that the name of the datasheet (logc in our example) must be identical to the heading of the first column of the index sheet.

The values in the datasheet of observations are numbers (except the heading), showing either increasing or decreasing tendency. Only those values are considered which are larger than the values 'Lower' and 'Upper' set in the Generic control parameters menupoint.

A third column, beside the time and observation columns, is optional; here, you can give weights for the data points. In this case, the heading, too, and all the rows should contain 3 items (x variable, y variable, weight).

The most frequent reason for program error is the inappropriate data format! The list of points in a curve should contain numbers only!

VARIABLES

DMFit fits the sigmoid function described by Baranyi and Roberts (1994). It has 4 main parameters:

`rate, lag, y0, yEnd`

and two curvature parameters:

`mCurv, nCurv.`

Estimates of the four main parameters will be in the output columns of the index sheet.

The primary parameter, called 'rate', is the potential maximum rate of the model (μ_{\max} in the original paper), which can be negative, too, for decay. It is only a potential rate because, theoretically, this rate is never exactly reached by the fitted curve, due to the applied limiting functions which has a slight effect even in the linear phase. However, the difference between the potential and the real maximum rate (denoted as 'rate(num)' by the program) is negligible if the mCurve curvature parameter is sufficiently large.

'lag' denotes the lag parameter as described by Baranyi and Roberts (1994).

The parameter `y0` is the initial point of the sigmoid curve.

The parameter `yEnd` is the upper asymptote of the sigmoid curve (or lower asymptote if the curve is decreasing).

Curvature parameters: `nCurv` and `mCurv`, (at the beginning and end of the linear phase, resp.) with values between 0 and 20. These parameters will not be optimised by the program, but possibly omitted, one or both of them, if parsimony requires so.

The definition of `mCurv` is: `mCurv=m*abs(yend-y0)`, where `m` is the Richard curvature (see Baranyi and Roberts, 2000). The default value for `mCurv` is 10.

For the lower curvature parameter

a/ definition: $nCurv = \nu / \mu$ (see reference) for `modelCode=2` (DModel)
(default value: $nCurv=1$);

b/ definition: $nCurv = h_0 = \nu * lag$ (see reference) for `modelCode=3` (scalefree DModel)
(default value: $nCurv=h_0$).

The above two options can be set in the DMFit menupoint 'Generic control parameters'. The option 3 results in a scale-independent model, which is useful for situations other than bacterial viable counts.

Among the possible models, also "Gompertz" and "Biphasic" are listed for comparison purposes. The Gompertz function is analysed extensively by Zwietering et al (1991) and this program uses the parameterisation he introduced. The biphasic function consists of two joint linear segments. In that case, the slope of the first segment plays the role of μ_{max} the location of the joint is the "lag" while the slope of the second segment will be the parameter "yend".

If you don't want DMFit to investigate, whether a curvature parameter can be zero, make its initial value bold, in the input columns (before the thick line); this font generally does not allow an initial value to be changed.

No transition phase is equivalent to setting the respective curvature parameter to zero. For example, the setting $nCurv=0$ and $mCurv=0$ generates linear fit.

$nCurv \cdot 20$, $mCurv \cdot 9999$ are symbolic values; they denote 'infinity' to indicate that the curvatures are real breakpoints (intercepting straight lines). In fact, values bigger or equal to 20 result the same fitting as the biphasic model.

INPUT PARAMETERS

As has been mentioned, in the input part of the index sheet, you can give initial values for the parameters (optional). This helps the non-linear curve fitting algorithm to find, in dubious cases, the best fit in the vicinity of your given initial estimates.

A possible index sheet for input should look like this:

logc	temp	pH	NaCl		mCurv	nCurv	rate	lag	y0	yEnd	gear
B301_100	15	7.3	0.5								
B301_101	10	6.9	2.5								
B301_102	10	7.2	2.5								
B301_103	10	6.9	4.5								
B301_104	15	7	4.5								
B301_105	10	7.2	4.5								
B301_106	15	7.3	4.5								
B301_42	10	6	4.6		10	1					
B301_67	10	5.1	0.5								

For this Index sheet, the observed curves could be recorded in the same workbook, in a sheet called as the first column heading (in our case it is “logc”).

The last column has a so-called gear control parameter. Its value can be from 0 to 5. Its effect is similar to its everyday meaning: gear=5 (default) means fast progress in the fitting procedure but possibly missing a hidden best fit; gear=1 is slow but more safe. The option gear=0 either generates a very rough fit or plots that particular curve which is induced by user given initial estimates.

For the sake of example, consider the input table above. In the eighth row, the user prescribed mCurv=10 for the after-linear-phase curvature parameter of the curve B301_42. The **bold** font indicates that zero value for the respective curvature would not be allowed, even if, according to the F-test, built in the program, the data indicated so.

OUTPUT PARAMETERS

On the right hand side of the column gear, DMFit calculates the parameters of the best fit complying with the input controls, as well as some other statistical/computational indicators:

```
errCode,  mCurv,  nCurv,  rate(pot),  se(rate),  lag,
se(lag),
rate(num), r1=rate*lag,  se(r1),  yEnd,  se(yEnd),
se(fit),  R^2_stat
```

errcode: Error code:

0 - OK

- 1 - discontinuous problem: model of intercepting straight lines, or gear=0.
- 2 - some parameters have high standard error (rel. error is greater than 200%)
- 3 - ill conditioned problem
- 4 - running into singularities during the calculations
- 5 - error in input data in the index sheet or error in the raw data of the curves
- 6 - the curve to be fitted is not found in the actual directory

`mCurv, nCurv, rate... se(yEnd)` : Final estimates of the model parameters and their standard errors. The parameter `rate(num)` is the potential maximum rate which, at small curvature parameters, can differ significantly from `rate(num)`, the maximum rate calculated numerically.

`se(fit)` : Standard error of fitting (estimated standard deviation of the observed independent values, assumed to be constant).

`R^2_stat`: Adjusted R-square statistics of the fitting.

RUNNING DMFIT ON YOUR CURVES

If you want to fit more curves, prepare them as described above. With the index sheet as active sheet, set the cursor to the name of the curve to be fitted and select the first menupoint of DMFit. This fits only one curve. If you select the second menupoints, that fits all the curves from the actual row and below, up to an empty cell in the first column in the index sheet. The above two menupoints create plots about each fitting and place them onto separate sheets, as you will notice. If you want to see these plots collected on one sheet, choose the menupoint 'Insert/Control Plotsheet' under DMFit. Then the graphs will be collected as small plots on a single sheet. These menupoints plot the past fittings, retrospectively, together, so you can compare the effect of the different settings of the input part of the index sheet.

The menupoint 'Parameters only' does not provide plots, only the parameters in the output part of the index sheet.

SOME TECHNICAL LIMITS:

- max number of points in a curve to be fitted = 500
- max number of plots: 120

DEMO

This example, with weights, magnifies the significance of the first point, while practically omits the 5th point:

TIME , LOGC, WEIGHTS
0.000 , 3.39, 10
5.960 , 3.98, 1
8.080 , 4.41, 1
10.170 , 4.96, 1
19.750 , 7.45, 0.01
21.250 , 7.79, 1
32.830 , 8.76, 1
49.750 , 8.78, 1

A DEMO

1. Copy and open the files DMFIT.XLAM and INDEX.XLSX in a working directory. The “Logc” sheet of the INDEX.XLSX workbook are log10 cell concentrations of batch bacterial cultures as in Gibson et al (1988). You should see the menupoint DMFIT in the menubar.

2. Use the 'Insert plotsheet' from the “Utilities” menupoint of DMFit.

Primary modelling

In the Index sheet, position the cursor on the first curve and select the menupoint DMFIT - FIT ALL THE CURVES FROM ACTIVE ROW. You will get plots in the created “Plots” sheet, as well as growth parameters in the output part (right hand side of the column gear) of the index sheet.

Secondary modelling

You can model the variation of the estimated growth parameters with temperature (temp) and pH, for example with the Regression or the Solver facility of Excel (or, in fact, with any multivariate regression program) and you can get the sheet similar to the “Modelsheet” included in the enclosed Index workbook. However, you can use the “Polynomial (linear regression)” of the “Secondary modelling” facility of DMFit, too. In the latter case, you produce a sheet as the one called `PolyModel'. In these cases we modelled how the fitted rate values in the column `rate(pot)' depended on temperature and pH.

Nonlinear regression is for advanced users only!

For more details, see the next part of this manual.

Validation

Copy the predicted values, generated by the secondary model, in the `rate' column of the input part of the index sheet (left hand side of the column “gear”). These will be now fixed parameters since we will set the “gear” value for 0.

Repeat the steps 7 and 8 for the other growth parameters (not for the curvature parameters and for the initial value y0). For the sake of simplicity, we took the parameters “rl=rate*lag” constant, =1 , respectively), and this relation was used

to give initial values for the “lag” parameter.

Set the control “gear” to zero and repeat the step 3. Now you can compare the environment-modelled, the fitted curves and the original data.

After completing all the steps above, you should get a sheet like “Fits&Preds” in the workbook INDEX.XLS.

SECONDARY MODELLING

DMFit also contains tools for linear and non-linear regression tools to model how some estimated parameters depend on factors possibly recorded in the second, third, etc columns of the index sheet.

When the curves are fitted, highlight the recorded values for the factors, together with the heading in the first row. The response (specific rate, lag time) must be in the last column of the selection. (You can achieve the selection of non-adjacent columns by pressing the “Ctrl” button while clicking on a distant cell). Empty or non-numerical cells (e.g. letters) are not allowed and a warning message is displayed in this case. To perform a regression analysis, choose the menupoint DMFit – Secondary modelling.

A. Linear regression

Linear Regression Input Box

Response(Y):	The last column of the selection.
Y Transform:	This list box gives several options for transformations of the response variable Y.

By clicking the > button, terms can be selected for the regression model. No more than 3 regressors can be selected. Terms will be transferred into the current model in order of their appearance in the select term window.

Next: Opens the Linear Regression Dialog Box

Linear Regression Dialog Box

Input area: contains the terms selected for the model: the response Y and the regressors. The current quadratic model appears below the Input area. Terms can be eliminated from the current model by ticking the checkboxes “Omitted =>”.

Go: Starts the linear regression using the model terms of the current model

Back: Takes the user back to the Linear Regression Input Box. This gives the opportunity to modify the selection of the regressors.

Output area:

Summary: After performing a regression, this window contains important parameters for model adequacy checking.

R²: Coefficient of Determination. R² is the portion of total variability which is explained by the model. It generally increases when more parameters are fitted. This can lead to an over-parameterised model with high R², but poor prediction quality (see Baranyi et al, 1996)

RMSE: Root-mean-square error (standard deviation of the residuals)

Estimated parameters, Standard error, t-values: Shows the terms used in the model and their regression coefficients with standard errors and t statistics.

Save results in a Sheet:

Save the main results of the regression in an excel sheet.

B. Nonlinear regression

Linear and Nonlinear Regression Dialog Box

Select “Nonlinear regression” in the Secondary modelling menu point. Three nonlinear models, available in literature on the growth rate of the micro-organisms as a function of temperature, pH and water activity, are implemented in DMFit:

- a) Ratkowsky-type model (Ratkowsky *et al.*, 1982, Adams *et al.*, 1991, Wijtzes *et al.*, 1993)

$$\mu_{max} = b(T - T_{min})^2 (pH - pH_{min})(aw - aw_{min}) \quad (1)$$

where T_{min} , pH_{min} and aw_{min} are respectively the minimum temperature, pH and aw for growth.

- b) Cardinal Model (Rosso *et al.*, 1995, 1998)

$$\mu_{max} = \mu_{opt} \gamma(pH) \rho(T) \delta(aw) \quad (2)$$

with

$$\rho(T) = \begin{cases} T < T_{min}, & 0 \\ T_{min} < T < T_{max}, & \frac{(T-T_{min})^2(T-T_{max})}{(T_{opt}-T_{min})[(T_{opt}-T_{min})(T-T_{opt})-(T_{opt}-T_{max})(T_{opt}+T_{min}-2T)]} \\ T > T_{max}, & 0 \end{cases}$$

$$\gamma(pH) = \begin{cases} pH < pH_{min}, & 0 \\ pH_{min} < pH < pH_{max}, & \frac{(pH-pH_{min})(pH-pH_{max})}{(pH-pH_{min})(pH-pH_{max})-(pH-pH_{opt})^2} \\ pH > pH_{max}, & 0 \end{cases}$$

$$\delta(aw) = \begin{cases} aw < aw_{min}, & 0 \\ aw_{min} < aw < aw_{max}, & \frac{(aw-aw_{min})^2(aw-aw_{max})}{(aw_{opt}-aw_{min})[(aw_{opt}-aw_{min})(aw-aw_{opt})-(aw_{opt}-aw_{max})(aw_{opt}+aw_{min}-2aw)]} \\ aw > aw_{max}, & 0 \end{cases}$$

where T_{opt} , pH_{opt} , aw_{opt} are respectively the optimum temperature, pH and water activity for growth, T_{max} , pH_{max} are the maximum temperature and pH allowing the bacterial growth.

c) Gamma Model (Zwietering *et al.*, 1992)

$$\mu_{max} = \mu_{opt} \gamma(T) \gamma(pH) \gamma(aw) \quad (3)$$

$$\text{with } \gamma(T) = \left(\frac{T-T_{min}}{T_{opt}-T_{min}} \right)^2 \quad \gamma(pH) = \frac{pH-pH_{min}}{pH_{opt}-pH_{min}} \quad \gamma(aw) = \frac{aw-aw_{min}}{1-aw_{min}}$$

Theoretical domain of applicability of the models

Model	Temperature	pH	a _w
Ratkowsky type model	Suboptimal range	Suboptimal range	Entire range allowing growth
Cardinal Model	Entire range allowing growth	Entire range allowing growth	Entire range allowing growth
Gamma Model	Suboptimal range	Suboptimal range	Entire range allowing growth

Partial versions of the equations (1), (2) and (3) can be used in the case where only one or two environmental factors are to be considered. For example, if the user selects the Ratkowsky-type model and pH and temperature as the regressors, the form of the selected model is:

$$\mu_{max} = b(T - T_{min})^2 (pH - pH_{min}) \quad (4)$$

Nonlinear Regression Model Dialog Box

Response(Y):	contains the headline of the last column of the selection.
Y Transform:	This list box gives several options for transformations of the response variable Y.
Model:	This list box allows the selection of the models implemented in DMFit.

By clicking the > and < button, terms can be selected or unselected for the modelling. The models implemented in DMFit can describe only the effects of temperature, pH and water activity. If selected, the temperature headline must appear in the temperature box, the pH headline in the pH box and the water activity headline in the water activity box.

Next: Opens the Non Linear Regression Dialog

Non Linear Regression Dialog Box

Go: Starts the nonlinear regression.

Back: Takes the user back to the Linear Regression Input Box. This gives the opportunity to modify the selection of the model and/or the regressors.

Save results in a Sheet:
Save the main results of the regression in an Excel sheet.

<u>Output area:</u> If a convergence is not obtained, a warning message is displayed. The user can restart the regression, starting with different initial values.	
Summary:	After performing a regression, this window contains parameters for model adequacy checking.
R ² stat:	Coefficient of Determination. R ² is the portion of total variability which is explained by the model. It generally increases when more parameters are fitted. This can lead to an over-parameterised model with high R ² , but poor prediction quality (see Baranyi et al, 1996).

SE of fit:	Standard error of fit
Estimated parameters	
Standard error:	It shows the terms used in the model and their regression coefficients with standard errors.

MCP

A useful utility called MCP is also added to the “Secondary modelling” menu. You have to select the values of those environmental factors at which the observations were made (temp and pH columns in our index sheet). Then run MCP, which will produce the Minimum Convex Polyhedron “shrink-wrapping” the selected points, as in the sheet “MCP” in Indexx.xls. For more details see Baranyi et al, 1996.

REFERENCE

- Adams, M.R., Little, C.L., Easter, M.C. 1991. Modelling the effect of pH, acidulant and temperature on the growth rate of *Yersinia enterocolitica*. J. Appl. Bacteriol., 71, 65-71.
- Baranyi, J. & Roberts, T.A. (1994). A dynamic approach to predicting bacterial growth in food. Int. J. Food Microbiol. 23, 277-294.
- Baranyi J., Ross T., Roberts T.A. and McMeekin T. (1996). The effects of overparameterisation on the performance of empirical models used in Predictive Microbiology. Food Microbiol. 13. 83-91.
- Baranyi, J. and Roberts, T.A. (2000). Principles and application of predictive modelling of the effects of preservation factors on micro-organisms. In: The microbiological safety and quality of foods. Ed. Lund B., Baird-Parker and Gould, G. Aspen Publishers Inc.
- Gibson, A.M., Bratchell, N. and Roberts, T.A. (1988). Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride and storage temperature. Int. J. Food Microbiol. 6, 155-178
- Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., (1990). Numerical Recipes. Cambridge University Press. Cambridge.
- Ratkowsky, D.A., Olley, J., McMeekin, T.A., Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. J. Bacteriol., 149, 1-5.
- Rosso, L., Lobry, J.R., Bajard, S., Flandrois, J.P. 1995. Convenient model to describe the combined effects of temperature and pH on microbial growth. Appl. Environ. Microbiol., 61, 2, 610-616.
- Rosso, L. 1998. Predictive microbiology: Validation of the models in the industrial context. COST 914 workshop. Validation of predictive models in wide range of European foods, 10-11 december, Cordoba Spain.
- Wijtzes, T., McClure, P.J., Zwietering, M.H., Roberts, T.A. 1993. Modelling bacterial growth of *Listeria monocytogenes* as a function of water activity, pH and temperature. Int. J. Food Microbiol., 18, 139-149.
- Zwietering, M.H., Jongenburger, I., Rombouts, F.M., and van't Riet, K. (1990). Modelling of the bacterial growth curve. *Appl. Environ. Microbiol.* 56. 1875-1881.

Using the spreadsheet CBPRED.XLSM

CBPRED.XLSM is a macro-enabled Microsoft Excel spreadsheet.

In the first part of the sheet, above Row 34, six growth curves are plotted and, in the cells B27:F32, estimations can be seen for their maximum specific growth rates, minimum doubling times and the times required to reach an increase of 10^n cfu/ml in concentration (taking account of lag, too), where the user provides the value of n in the cell A24. For example, 1000-fold increase: $n=3$. Also the user should input the values of the controlling factors in the cells G27:J32 and the codes of the organisms in the cells K27:K32, referring to the cells A11:A22. If an inputted code in K27:K32 is zero or empty then no prediction is given and the curve will not be represented on the chart.

The first two environmental factors are the temperature and pH; the other two are ef_3 and ef_4 . These are identified following the name of the respective organisms, in B11:B22. Generally, the third factor is water activity (a_w). Because, to adjust a_w , salt has been used most commonly at IFR, the salt content (in percentage) is given as third controlling factor for the first two organisms. From that value, the value of a_w is calculated in P27:P32. If something else is used to adjust a_w then set the required a_w level by trial and error, changing I27:I32.

In the cells L27:L32 the inoculum sizes are to be typed as log₁₀ cfu/ml. In the range M27:M32, the user should type the other initial value, α_0 , for each organism. α_0 characterizes the physiological state of the microbial population at the inoculation. This is a rescaled form of a similar quantities denoted by h_0 in Baranyi and Roberts (1994). α_0 is that proportion of the potential specific growth rate at the inoculation which is utilized by the cells, therefore its value can be between 0 and 1 only.

Assuming that the cells were at a similar physiological state at inoculation in the laboratory experiments which provided the data for the model, an average value for α_0 is suggested in the cells Q27:Q32.

The time scale can be changed in the range H2:H22.

In each cell of the range O27:O32, the average multiplicative error of the estimated maximum specific growth rate can be found. This error, in this model, is the same as that of the respective lag time. It is the standard error of the quadratic response surface fitted to the LN-transformed values of the measured specific growth rates. It can be considered as a percentage, what proportion of the predicted specific growth rate (doubling time, lag time) is its standard error.

The fields N27:N32 indicate whether the values of the controlling factors are outside (cell content: N/A) or inside (cell content: 1) the region where the experiments were carried out. The limits, in the order 'lower limit for the first, upper limit for the first, lower limit for the second, ... etc', are fixed in the block of coefficients, left from the column U, under the respective heading 'lim'.

In the second part of the sheet, from Row 34, a matrix of similar predictions can be obtained about a particular organism, the code of which should be typed in D35. The lower and upper bounds of the first three environmental factors, together with the step-values, should be typed in the G36:I38 cells to create the matrix under Row 41.

The value of the fourth environmental factor is fixed in K36.

This matrix of predicted values can be obtained by using the macro button.

You can distinguish four types of cells in the spreadsheet:

1. Standard: containing estimated values, headings or other information or empty cells.
2. Yellow: containing results of mathematical operations not useful for the user (from the column P).
3. Light grey: input area filled by the user.
4. Blocked: area with dashed border for model coefficients, for the user's own, similarly created, model.