# Tutorial Fitting Parameters for ODEs

Essential Skills for Bioinformatics and Biocomplexity

**Topics covered:**

- Fitting ODE-models to data (using `grind.R`).
- What model parameters can(not) be estimated?
- Distinguishing between models and overfitting.

**Preparatory reading:**
Ram, Y. *et al.*, Predicting microbial growth in a mixed culture from growth curve data, PNAS 116(29): 14698–15707 (2019).

## 1 Getting started with Grind

In the paper by **?** ] you studied earlier this week the individual growth rate of each strain is described by extending the logistic growth equation $\mathrm{d}N/\mathrm{d}t = rN(1 - N/K)$ into

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{q_0}{q_0 + \mathrm{e}^{-mt}}\, rN\left[1 - \left(\frac{N}{K}\right)^v\right] ,$$

where $v$ is a parameter defining the curvature of the density dependent function $[1 - (N/K)^v]$ and the leading term describes the adjustment of the bacteria to the new environment of the experiment.

To study this growth model, we will work with an R script called `grind.R` which is a wrapper around the R-packages `deSolve`, `FME` and `rootSolve` developed by Karline Soetaert and colleagues [**? ? ? ?** ]. These packages allow one to solve differential equations, find their steady state, and perform nonlinear parameter estimation. In the first exercise, we talk you through the process of using Grind, simply using fixed parameter values. In later exercises we turn to the fitting of parameters.

**Question 1**

**a**. To analyse the model open the file `StartingGrindAndRam.Rmd`. It is in the subfolder of the GitHub repo you cloned on Tuesday. Parameter values are filled in in the vector `p`, and initial conditions (initial state) are put in the vector `s`. After defining the model, parameter values and initial conditions you see the command `run()` which integrates the model numerically and provides a time plot
Familiarize yourselves with the script, and execute it chunk by chunk.

**b**. Change the value of $K$ and study what happens

**c**. Change the value of $v$, while keeping it at least 1 and study what happens.

**d**. Modify `StartingGrindAndRam.Rmd` to write a function for the delay phase and plot it. What does this function do?

**e**. What happens if you change $m$? What happens if you change $q0$? Study this also in the full model.

A full manual of Grind is available in the form of a tutorial (tbb.bio.uu.nl/rdb/grindR/tutorial.pdf).

# 2 Parameter fitting in Grind

**Question 2** In the previous section we played around with the effects of the different parameters. Now we are going to try and fit ourselves the values of these parameters to the data obtained in the Ram et al. paper.

**a**. The data is in the folders Fig3, Fig4 and Fig5, in the form of '.csv'-files. Open the R script `BasicParameterFittingRamPaper.Rmd`. Execute the first two code chunks to load grind and plot those data of Figs 3, 4 and 5 that correspond to experiment A. Note that the strains are called 'R' and 'G', for red and green, respectively. Why are there a series of points per single time point in the plots for Fig3 and Fig4 but not Fig5?

**b**. Now we are going to fit the 6 parameters of the growth model to the data in figure 3A for the red strain. Code is provided in the next two code chunks defining the model, which parameters to fit, any relevant upper or lower bounds on parameter values and the fitting itself. Compare your parameter estimates with those in Table S2. Note that in the Rmd script we transform the experimental data by taking their log. Why do we do that, did the authors do that as well, and does it make a difference?

**c**. Now run the next line of code which will give you statistics for the fit you just performed. First consider the first part. What would the estimated value for the variable N mean? Given that for all estimated parameter values there is a standard deviation what does this imply?

**d**. Now consider the second part of the statistics. What would it mean if parameters were strongly (anti)correlated, could you think of examples? What would this suggest for a model if this happens?

**e**. Now we move to figure 3B. In our script lag parameters, $q_0$ and $m$, are not ignored when fitting the data, while this is done by Ram et al. based on the fact that in the B experiment conditions are such that no lag phase is expected. What values do you obtain for these parameters and does this make sense?

**Question 3** We have redone the fitting of Fig3A as done in the paper, and done it for Fig3B already slightly differently. Now we are going to consider if things could have been done still better.

**a**. The authors fitted the data in Fig 3 A1 and 3 B1 separately, while they are data from the same strain. It would thus make sense to fit these data sets simultaneously. Which parameters do you think should be similar between the two experiments, and which parameters do you think should be different between the experiments?

**b**. Perform the simultaneous fitting to the datasets of Fig3A and Fig 3B using the supplied code chunks. Make sure you understand what is happening in the code. Why would this approach of fitting to two data sets simultaneously be better?

**c**. Is this reflected in the fit statistics? What should we ideally have done?

**Question 4**. After estimating six parameter from the growth curves for strains growing in isolation in Fig. 3, two different strains were grown together and their total optical density (OD) was measured (see Fig. 4). The Fig. 4 data was then fitted to the summed population size, $N_1 + N_2$, of a 2-dimensional version of the same model

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = \frac{q_{0,1}}{q_{0,1} + \mathrm{e}^{-m_1 t}}\, r_1 N_1 \left[1 - \left(\frac{N_1}{K_1}\right)^{v_1} - c_2\, \frac{N_2^{v_2}}{K_1^{v_1}}\right] ,$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = \frac{q_{0,2}}{q_{0,2} + \mathrm{e}^{-m_2 t}}\, r_2 N_2 \left[1 - \left(\frac{N_2}{K_2}\right)^{v_2} - c_1\, \frac{N_1^{v_1}}{K_2^{v_2}}\right] .$$

The two $c_i$ parameters provide the relative competitive strength of the other species (when $c_i = 1$ the intraspecific competition is equal to the interspecific competition). Fitting this model with 14 free parameters to the single sigmoid curve of data points representing the total OD (see Fig. 4) would clearly be infeasible. Instead, only $c_1$ and $c_2$ were estimated, and the 12 other parameters were copied from the fitting of the Fig. 3 data for both red and green strains.

**a**. Define the 2D model in the Rmd file, make sure to give it a different name from the original 1D model.

**b**. Run the code chunks provided for defining the parameter vector and for determining the initial conditions. Which parameters should be fitted?, define these in the code.

**c**. Next do the actual fitting. Look at both the curves and parameter fits and fit statistics obtained. How do the obtained c values relate to the curves? What do they mean biologically? How thrustworthy are these findings?

**d**. Given the total of 14 parameters one may wonder whether the authors are not overfitting their data, and whether a smaller number of parameters could suffice. As an example of reducing parameter numbers, redo the above fitting assuming a single competition parameter c. Hint: To do this redefine the model by renaming the c1 and c2 parameters into c (also do this in the p-vector !). How do your fitting results change and what c value do you obtain?

September 28, 2021, Kirsten ten Tusscher, Dieter Stoker, Rob J. de Boer, Utrecht University