Exercise 11

Growth rates decay as development progresses, eventually leading to growth termination. Many mechanisms have been proposed , but the mechanism underlying this decay has remained elusive [1, 2, 3]. The area growth of the *Drosophila* wing disc has been measured [4]. This offers the opportunity to test a range of candidate models [5]. As part of this problem sheet, you will learn about model selection by example of models of growth control in the *Drosophila* wing disc.

1. **Parameter Fitting and Model Comparison** Growth can be modelled according to a simple ODE,

$$\frac{dT}{dt} = k(t) \cdot T \tag{1}$$

where T is the area at time t and k is the so called growth rate, or area growth rate. If there was no slow down in growth, growth would be exponential and the growth rate would be a constant,

$$k(t) = k_0 \qquad \text{(Model 1, CST)}. \tag{2}$$

We want to compare this model to models with decaying growth rates. Here, we will consider the following alternative functional forms: An exponential decaying growth rate

$$k(t) = k_0 \cdot \exp^{-\delta \cdot t} \qquad \text{(Model 2, EXP)}, \tag{3}$$

a decline inversely proportional to the ara

$$k(t) = k_0 \cdot \frac{T(0)}{T(t)} \qquad \text{(Model 3, PA)}, \tag{4}$$

a decline according to a power law

$$k(t) = k_0 \cdot \frac{1}{t^\delta} \qquad \text{(Model 4, POW)}, \tag{5}$$

an exponential decaying growth rate with an additional constant term

$$k(t) = k_0 \cdot \exp^{-\delta \cdot t} + k_1 \qquad \text{(Model 5, EC)}, \tag{6}$$

and a logistic growth law

$$k(t) = k_0 \cdot (T_{set} - T(t)) \qquad \text{(Model 6, LOG)}. \tag{7}$$

In this exercise sheet, you are supposed to use model selection to identify the best fitting model.

Part 1: Model Fitting In the first step, we seek to find the most likely parameter set, X, for each model based on the given data, D, and additional information, I, that we may have. Using Bayes' theorem, this can be achieved by minimising the likelihood function. Assuming that the measurement errors are independent and distributed according to a normal distribution, the logarithm of the likelihood function, L, depends on the χ^2 function, i.e.

$$L = \ln\left[\text{prob}(D|X, I)\right] = \text{constant} - \frac{1}{2}\chi^2; \qquad \chi^2 = \sum \frac{(d_i - y_i)^2}{\sigma_i^2}, \tag{8}$$

where d_i is the (mean) data point at time point i, y_i the simulated value and σ_i the standard deviation.

a) Download the data set for fitting (*Data_Fit.xlsx*), implement the ODE (Eq. 1) and the different models from above (Eq. 2 - 7) and fit them to the data (use e.q. *lsqnonlin* or *particleswarm*). Determine the parameter set that either minimizes χ^2 for each model. Watch out in which way you have to provide the deviation to the opimizing function you are using (*lsqnonlin* differs from *particleswarm*!).

b) Plot the models vs. the data. Which model can be excluded for both of the fittings?

Part 2: Model Selection Two commonly used model selection criteria are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The two are very closely related but there are some slight differences as you will see. The AIC and BIC are defined as

$$\begin{aligned} AIC &= -2\ln(\hat{L}) + 2 \cdot k \\ BIC &= -2\ln(\hat{L}) + k \cdot \ln(n). \end{aligned} \tag{9}$$

Here, $\hat{L}$ is the maximized value of the likelihood function of the model, k the number of parameters being estimated, and n is the number of data points (=sample size). For both AIC and BIC, the model with the lowest values (also negative) is best.

c) How are $\hat{L}$ and χ^2 related? max $\hat{L}$ ~ min χ^2

d) Calculate the AIC and BIC for all the models. What is the ranking of the models according to the AIC and BIC values. Are there differences? Why?

e) Plot the complete data (*Data_complete.xlsx*) and compare it to the simulations with the optimized parameter sets that you obtained before with the restricted data range. Which models best describe the later/earlier time points? Would you draw a different conclusion now?

As we assume that the measurement errors are independent and distributed according to a normal distribution, AIC and BIC can be calculated also from the residual sum of squares (RSS),

$$RSS = \sum (d_i - y_i)^2. \tag{10}$$

data magnitude problem!

Knowledge of the data variance is thus not required. This is helpful in case of the sparse dataset by Nienhaus and co-workers as the sample size for the very early and late points is too small to determine the variance for those time points. The equations for the AIC and BIC then read

$$\begin{aligned} AIC &= 2 \cdot k + n \cdot \ln(RSS) \\ BIC &= n \cdot \ln\left(\frac{RSS}{n}\right) + k \cdot \ln(n). \end{aligned} \tag{11}$$

f) Repeat the fitting procedure based on residual sum of squares (RSS) for the complete data set. Compare the fits that you obtained previously based on χ^2 for the central data points. Why do the fits differ? Which of the two fits (the one based on RSS or χ^2) would you prefer? How does the RSS-based fit change when you only use the central data points? Does this help? What does this mean for the use of the RSS-based model selection method? Discuss the different approaches in light of your observations!

References

[1] J. Vollmer, F. Casares, and D. Iber. Growth and size control during development. *Open Biol*, 7(11), 2017.

[2] J. Vollmer, P. Fried, M. Sanchez-Aragon, C. S. Lopes, F. Casares, and D. Iber. A quantitative analysis of growth control in the drosophila eye disc. *Development*, 143(9):1482–90, 2016.

[3] J. Vollmer, P. Fried, D. Aguilar-Hidalgo, M. Sanchez-Aragon, A. Iannini, F. Casares, and D. Iber. Growth control in the drosophila eye disc by the cytokine unpaired. *Development*, 144(5):837–843, 2017.

[4] Ulrike Nienhaus, Tinri Aegerter-Wilmsen, and Christof M. Aegerter. In-vivo imaging of the Drosophila wing imaginal disc over time: novel insights on growth and boundary formation. *PloS one*, 7(10):e47594, 2012.

[5] J. Vollmer and D. Iber. An unbiased analysis of candidate mechanisms for the regulation of drosophila wing disc growth. *Sci Rep*, 6:39228, 2016.