



Lecture 11: Growth Control & Model Selection

Prof Dagmar Iber, PhD DPhil

MSc Computational Biology 2019/20

Contents

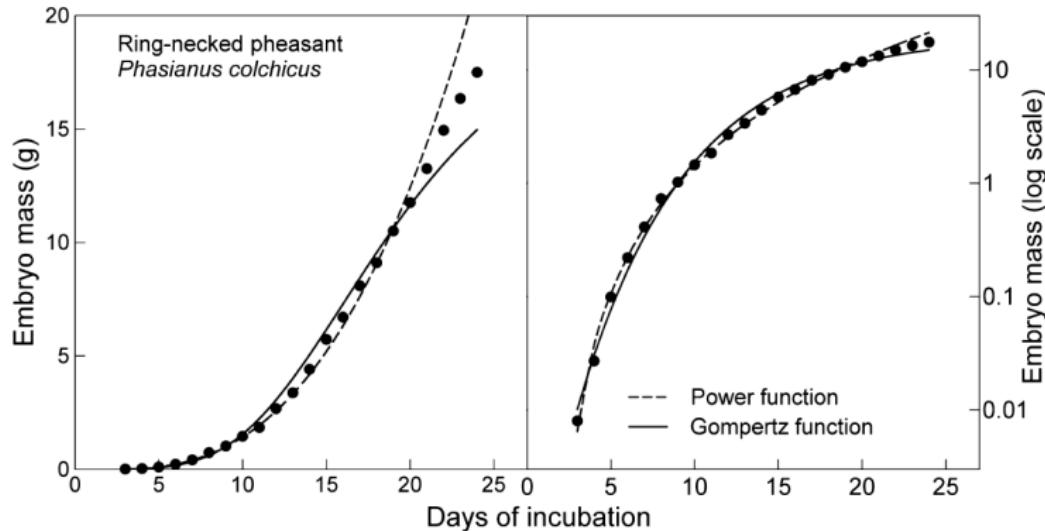
- 1 Developmental Growth Rates decline
- 2 Coordination of Body Growth and Organ Growth
- 3 Models of Organ Growth Control
 - Growth Control by Cell Differentiation?
 - Growth Control by Dilution or constant removal of Growth Factor?
- 4 Data-based Model Selection
 - Parameter Estimation
 - Optimization
 - Model Selection

Human development



Developmental Growth Rates decline

Developmental Growth Curves saturate



[Ricklefs, R. E. (2010). Embryo growth rates in birds and mammals. *Functional Ecology* 24(3): 588-596.]

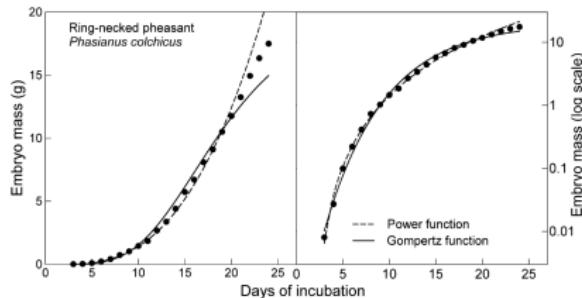
Functional Description of Dev Growth Rates

Powerlaw Growth

$$M = at^b$$

Gompertz Law

$$M = A \cdot \exp(-b \cdot \exp(-kt))$$



Parameter	Description
M	mass at time t
$A = M(t \rightarrow \infty)$	asymptotic mass
$b = \ln(A/M_0)$	final fold-change in tissue size
k	decline in growth rate
t	time since onset of development

Gompertz Law

$$M = A \cdot \exp(-b \cdot \exp(-kt))$$

is the solution of

$$\begin{aligned}\frac{dM}{dt} &= \gamma M \\ \frac{d\gamma}{dt} &= -k\gamma.\end{aligned}$$

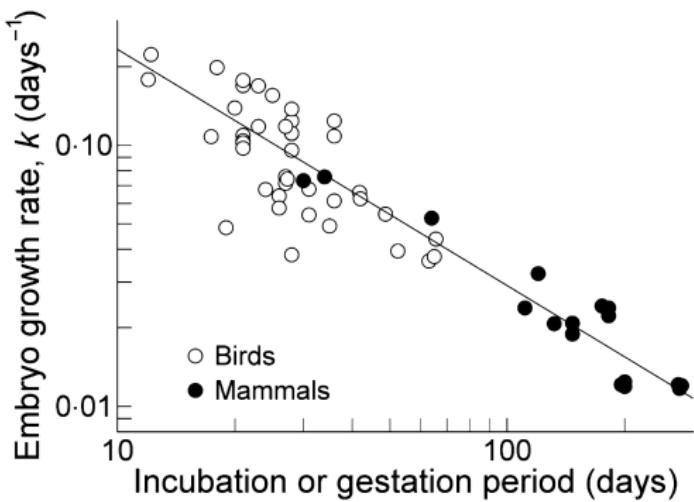
M mass; $A = M(t \rightarrow \infty)$ asymptotic mass; $b = \ln(A/M_0)$ final fold-change in tissue size; k decline in growth rate; t time since onset of development.

The decline in the Growth rate is inversely related to gestation time

Gompertz Law

$$M = A \cdot \exp(-b \cdot \exp(-kt))$$

with M mass; A asymptotic mass; $b = \ln(A/M_0)$; k decline in growth rate; t times since onset of development.



Ricklefs, R. E. (2010). Embryo growth rates in birds and mammals. Functional Ecology 24(3): 588-596.

Continuously Declining Growth Rates in Regenerating Limb

Iten, L. E. and S. V. Bryant (1973). Forelimb Regeneration from Different Levels of Amputation in Newt, *Notophthalmus-Viridescens* - Length, Rate, and Stages. Wilhelm Roux Archiv Fur Entwicklungsmechanik Der Organismen 173(4): 263-282.

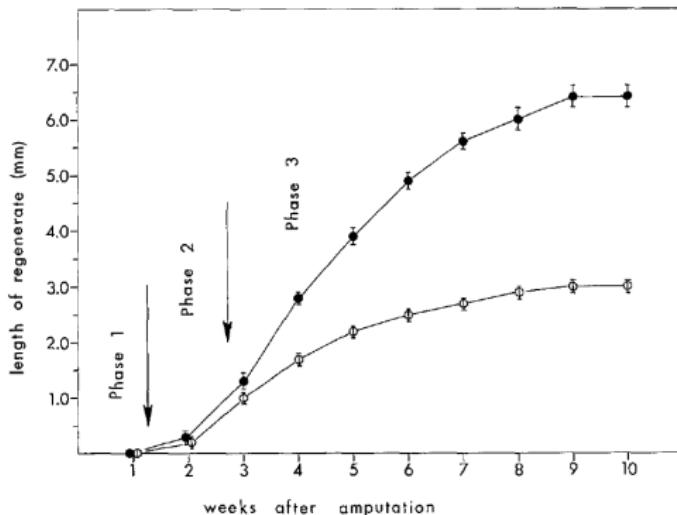


Fig. 13. Graph comparing the mean length of regenerates after amputation at different levels along the limb. Bars represent 2 standard errors of the mean. Closed circles represent regenerates from the proximal level and open circles represent regenerates from the distal level. The three major phases of regeneration are indicated on the graph

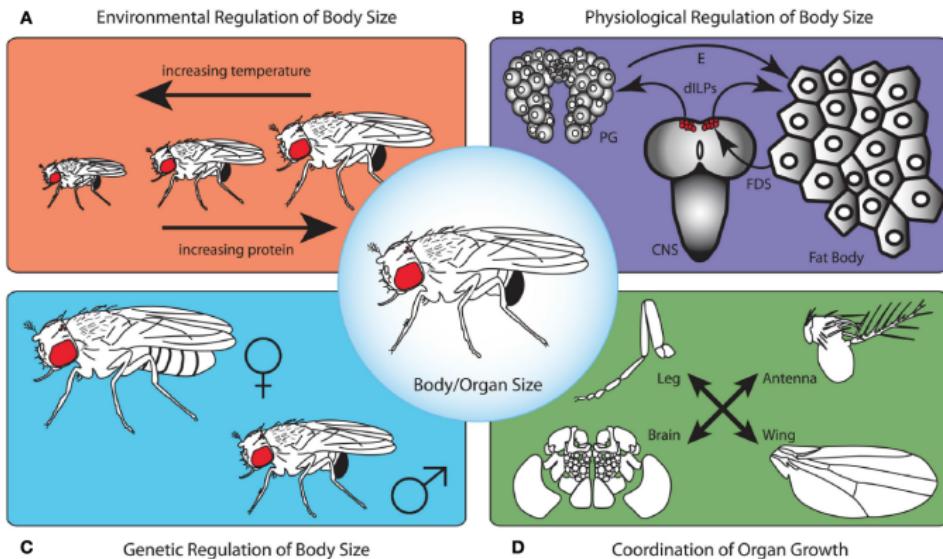
Conclusion

- Developmental Growth Rates decline continuously in a wide range of developmental systems (no counter-examples so far - not even solid cancers or stem cell based systems == organoids)
- The rate of decline is related to gestation time

What is the responsible mechanism?!

Coordination of Body Growth and Organ Growth

Perturbations can affect Size



Mirth, C. K. & Shingleton, A. W. Integrating body and organ size in *Drosophila*: recent advances and outstanding problems. *Front Endocrinol (Lausanne)* 3, 49, doi:10.3389/fendo.2012.00049 (2012).

Allometric Relationships

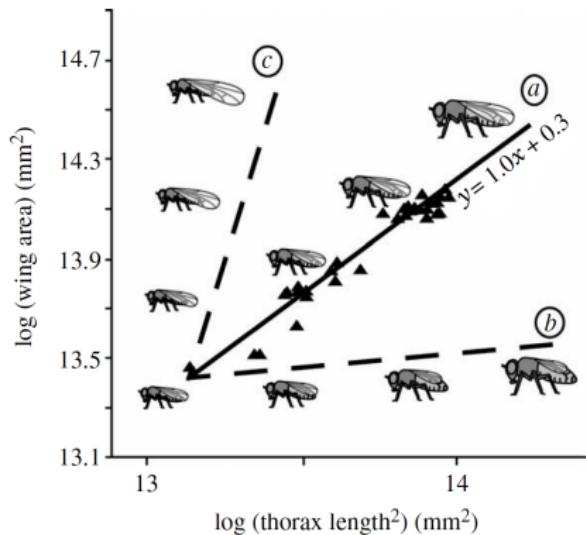
Allometric Relationships

The relationship between the size of one trait and the size of another trait or the body as a whole, are traditionally modelled using the allometric equation

$$y = bx^a \Leftrightarrow$$

$$\log(y) = \log(b) + \alpha \log(x)$$

where x and y are the size of two given traits, respectively.



Shingleton, A. W., et al. (2008). "Developmental model of static allometry in holometabolous insects." Proc Biol Sci 275(1645): 1875-1885.

Coordination of Body and Organ Growth

Allometric Relationships

$$y = bx^a \Leftrightarrow$$

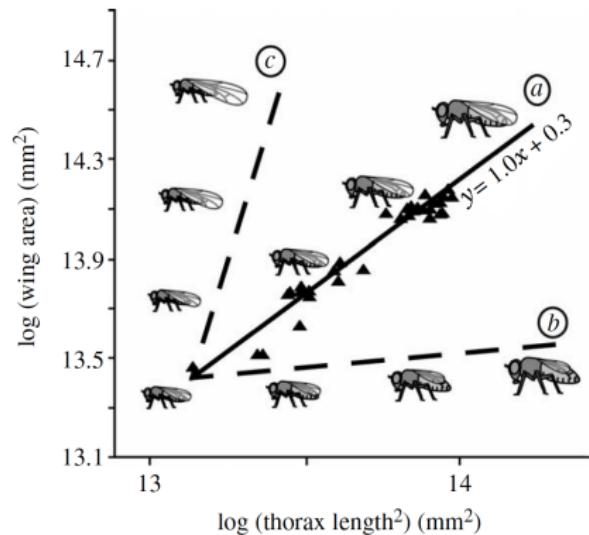
$$\log(y) = \log(b) + \alpha \log(x)$$

$\alpha = 1$: Isometry (case a)

$\alpha < 1$: Hypometry (case b)

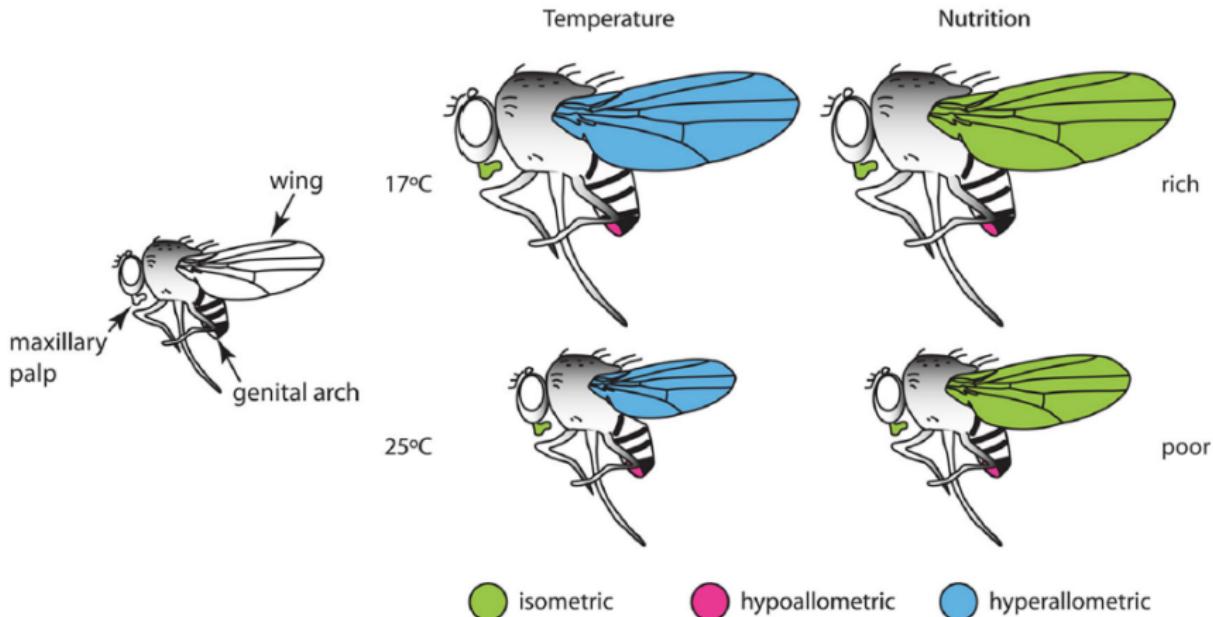
$\alpha > 1$: Hypermetry (case c)

Not all static allometries are linear:
also sigmoidal and discontinuous
cases known.



Shingleton, A. W., et al. (2008). "Developmental model of static allometry in holometabolous insects." Proc Biol Sci 275(1645): 1875-1885.

Allometry in different Appendages



Mirth, C. K. & Shingleton, A. W. Integrating body and organ size in *Drosophila*: recent advances and outstanding problems. *Front Endocrinol (Lausanne)* 3, 49, doi:10.3389/fendo.2012.00049 (2012).

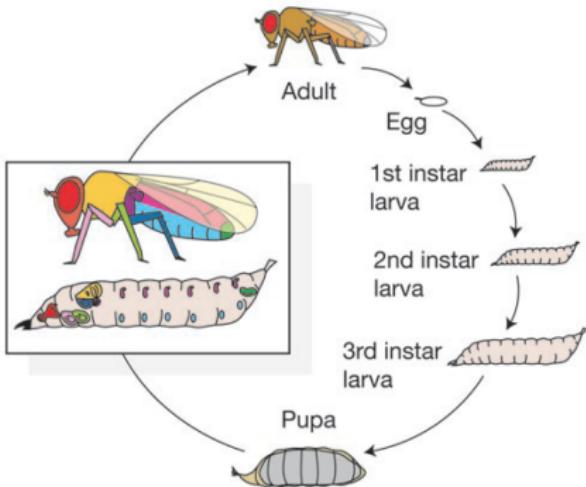
Growth Control in *Drosophila*

In holometabolous insects like *Drosophila*, growth is restricted to the embryonic and larval stages.

The developing insect moults through a series of larval instars before it stops feeding, pupates, undergoes metamorphosis, and finally ecloses from the pupal case as a fully formed adult.

Adult size is limited by the size of the larvae when it stops feeding, and hence stops growing.

Shingleton, A. W., et al. (2007). *Bioessays* 29(6): 536-548.



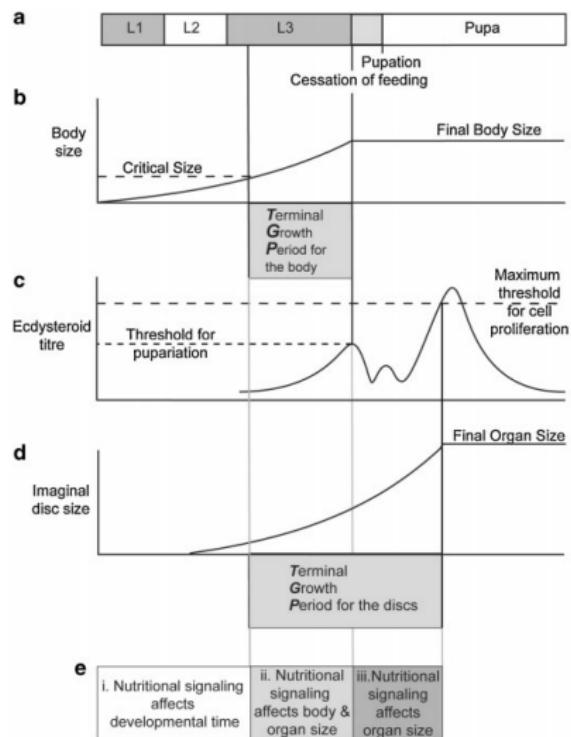
Shingleton, A. W. *Organogenesis* 6, 76-87 (2010).

Coordination of Body and Organ Growth in *Drosophila*

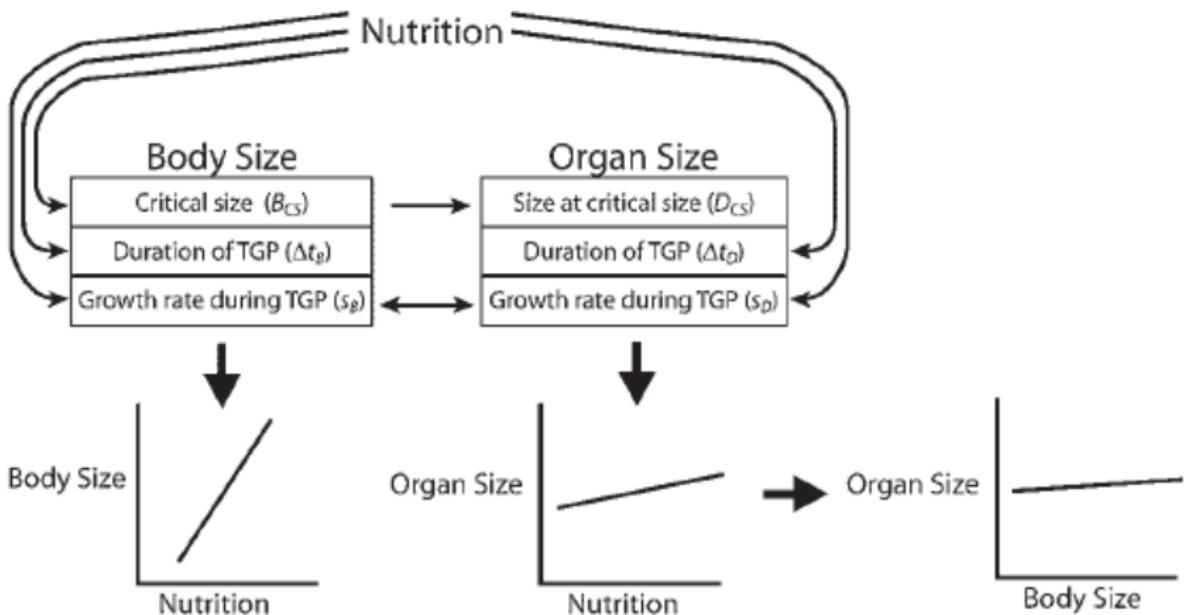
The adult organs develop as imaginal discs within the growing larva.

Although the cells that will become the discs are specified during embryogenesis, the point in development when discs initiate growth varies among discs, with most discs starting growth in the first or second instar.

Shingleton, A. W., et al. (2007). *Bioessays* 29(6): 536-548..

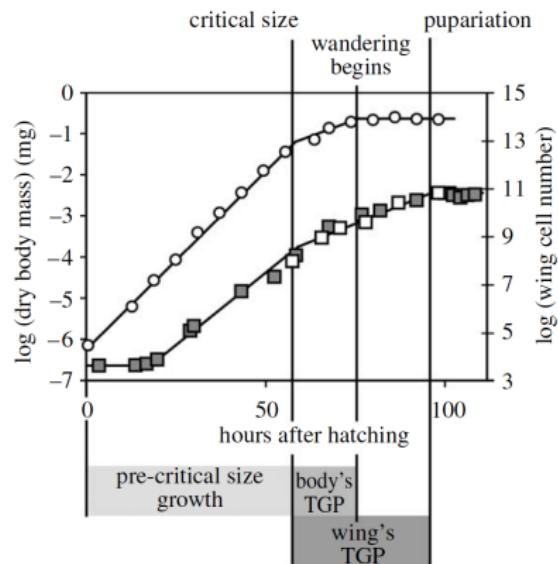


Coordination of Body and Organ Growth



Singleton, A. W., Frankino, W. A., Flatt, T., Nijhout, H. F. & Emlen, D. J. Size and shape: the developmental regulation of static allometry in insects. *Bioessays* 29, 536-548, doi:10.1002/bies.20584 (2007).

Coordination of Body and Organ Growth in *Drosophila*



Shingleton, A. W., et al. (2008). Proc Biol Sci 275(1645): 1875-1885.

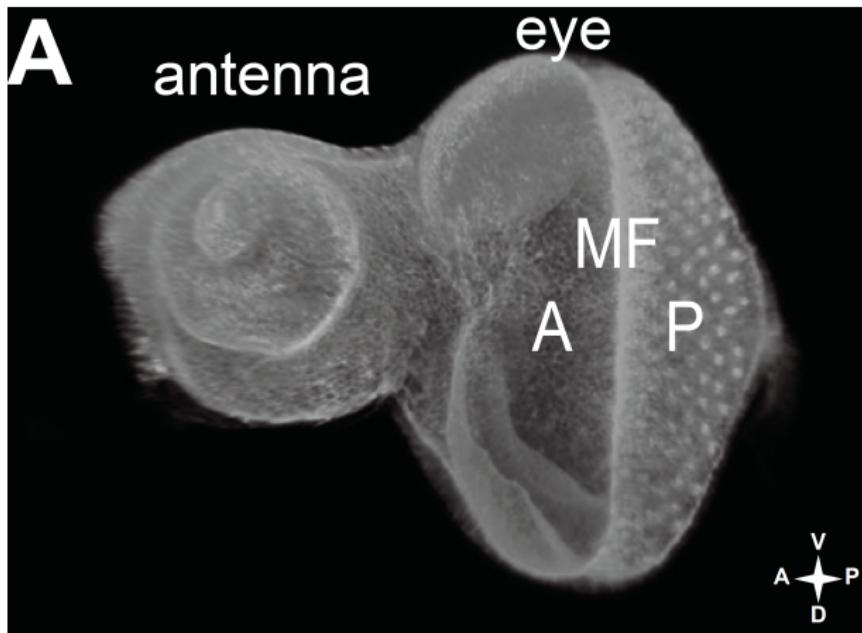
Models of Organ Growth Control

Growth Control by Cell Differentiation?

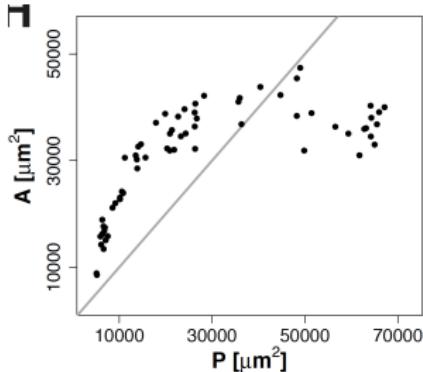
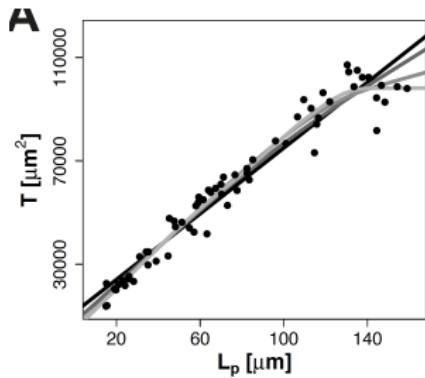
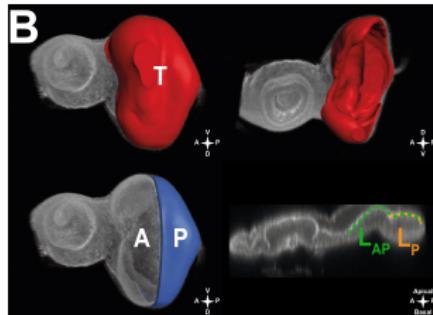
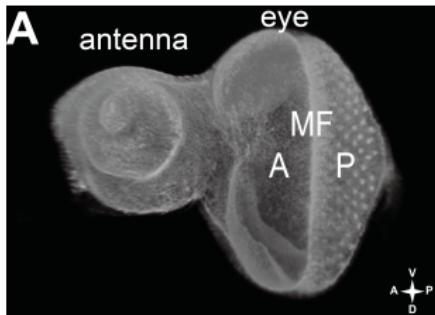
Eye Disc Development

Proliferation occurs mainly on the anterior (A) side. Cells differentiate posterior to the morphogenetic furrow (MF) to become ommatidia.

Growth termination: is this the result of the moving differentiation front (MF) or is growth slowing down?



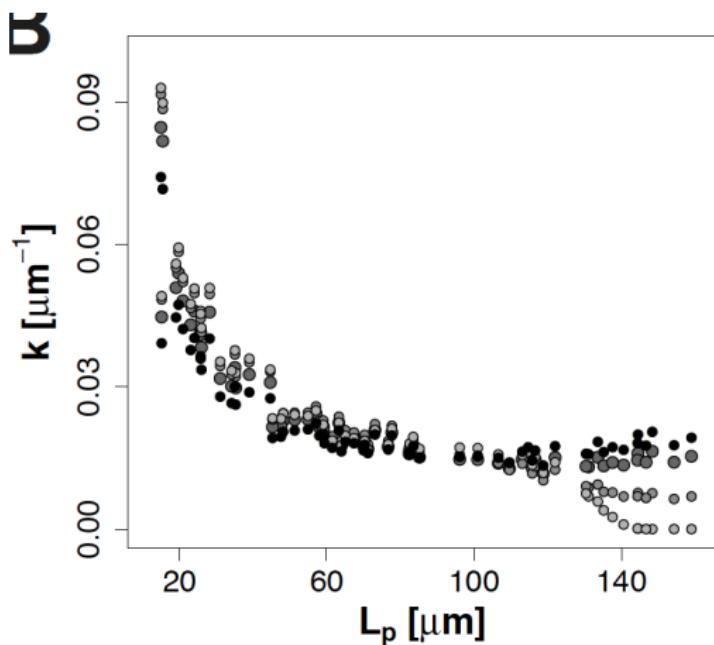
Growth Measurements



Cell Differentiation does not cause Growth Termination

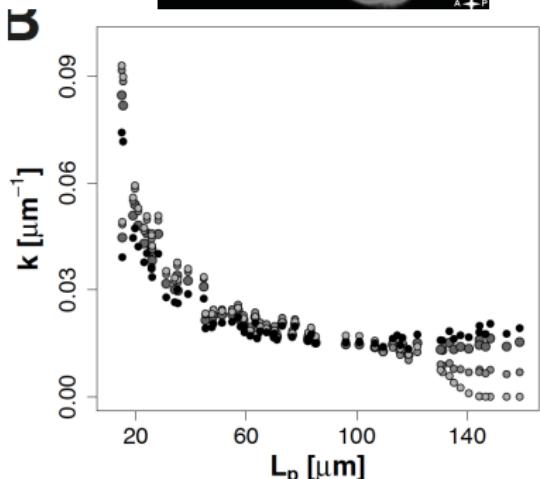
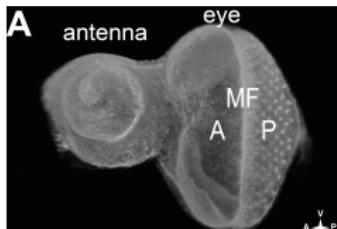
Growth termination: is this the result of the moving differentiation front (MF) or is growth slowing down?

As the **growth rate declines continuously on the anterior side**, cell differentiation is not responsible for growth termination.



Models for Continuously Declining Growth Rates

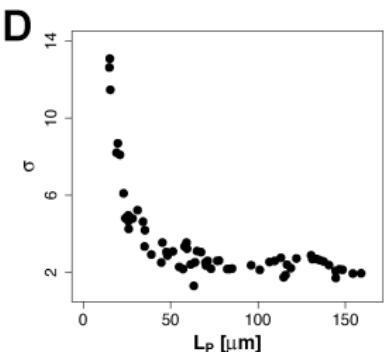
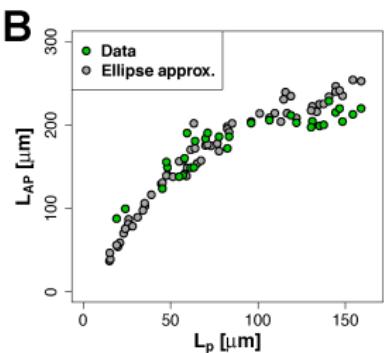
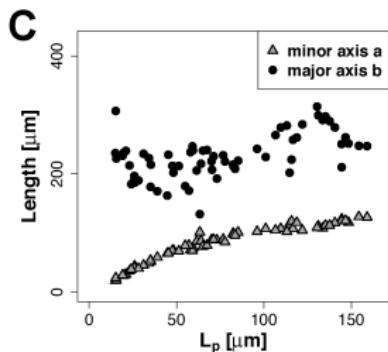
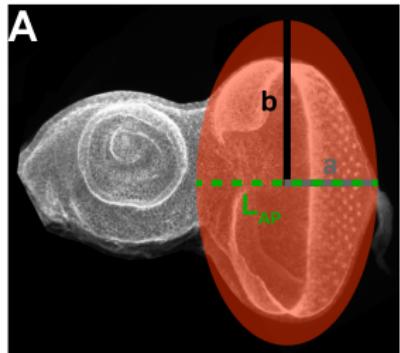
Alternative Models with a continuously declining growth rate for Eye Disc Development



$$\frac{dT}{dL_p} = k \cdot A$$

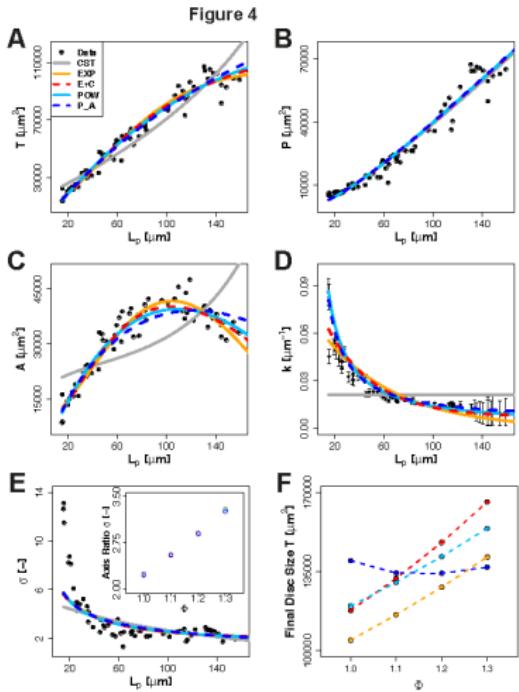
- (CST) $k(L_p) = k_0$
- (POW) $k(L_p) = k_0 \left(\frac{L_p(0)}{L_p} \right)^\delta$
- (P-A) $k(L_p) = k_0 \frac{T(0)}{T}$
- (EXP) $k(L_p) = k_0 \exp(-\delta L_p)$
- (E+C) $k(L_p) = k_0 \exp(-\delta L_p) + k_1$

The Eye Disc can be approximated by an Ellipse



Given T , L_p , L_{AP} from the simulation, we can determine A , P , and $\sigma = \frac{b}{a}$ using the ellipse equation.

Data-based Evaluation of Models



$$\frac{dT}{dL_p} = k \cdot A$$

- (CST) $k(L_p) = k_0$
- (POW) $k(L_p) = k_0 \left(\frac{L_p(0)}{L_p} \right)^\delta$
- (P-A) $k(L_p) = k_0 \frac{T(0)}{T}$
- (EXP) $k(L_p) = k_0 \exp(-\delta L_p)$
- (E+C) $k(L_p) = k_0 \exp(-\delta L_p) + k_1$

Mechanism?

Constant model: Does not fit the data anyway

Power-law model: Fits the data but no biological mechanism (?)

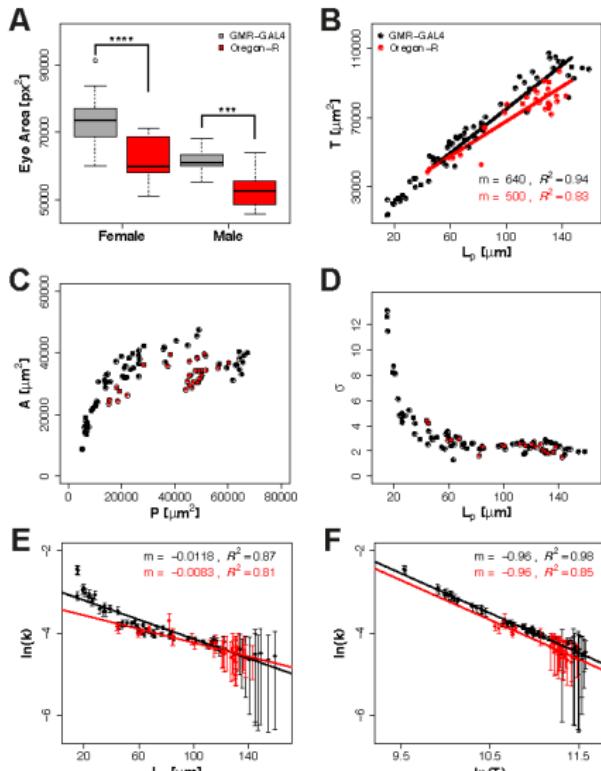
Exponential model: Constant removal of a growth factor (without dilution)

Area-dependent model: Dilution of a growth factor (without degradation)

Growth Control by Dilution or constant removal of Growth Factor?

Test Mechanisms with Smaller Eye Discs

Figure 7



Exponential decline

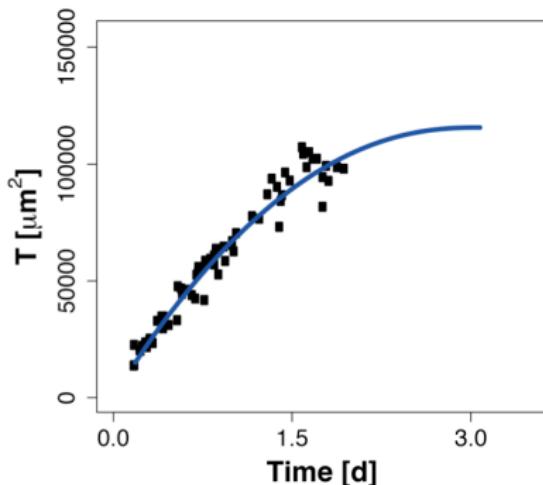
$$k(L_p) = k_0 \exp(-\delta L_p)$$

Area-dependent decline
fits the data best.

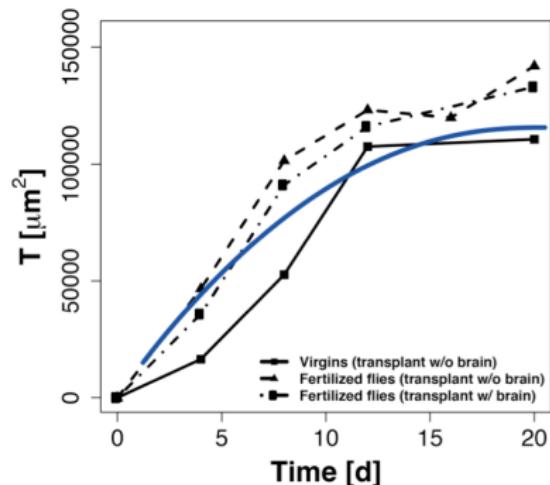
$$k(L_p) = k_0 \frac{T(0)}{T}$$

SIZE PRESERVATION AT LOWER DEVELOPMENTAL SPEEDS

LARVAE

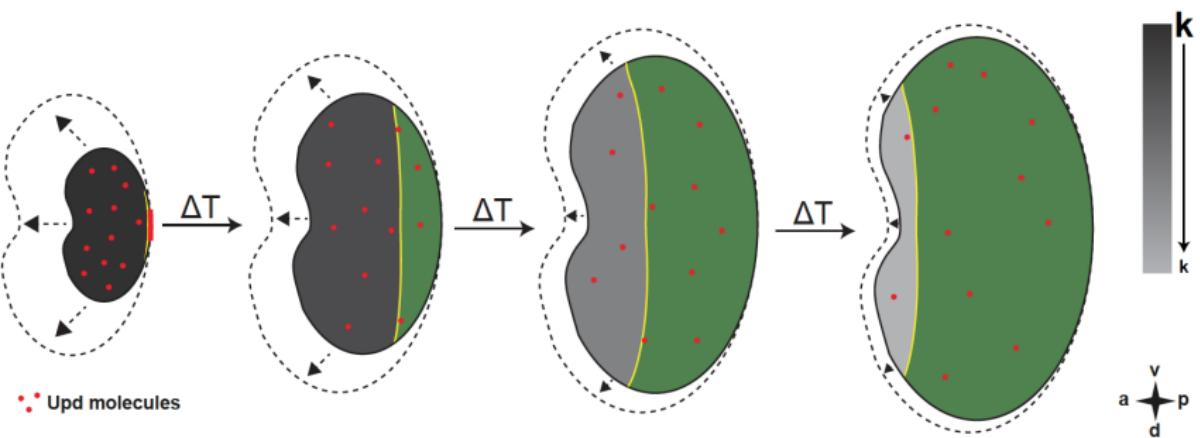


GRAFTED EYE DISCS



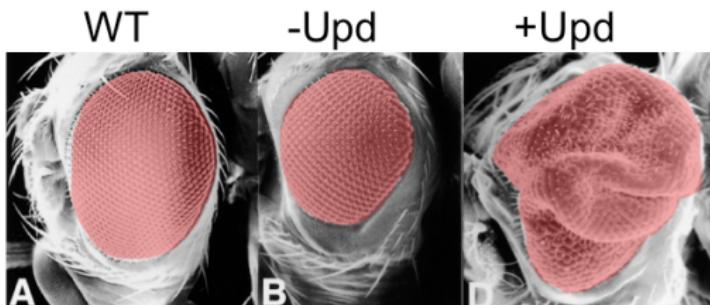
AREA-DEPENDENT Growth Rate naturally preserves final size at lower developmental speeds.

Candidate Mechanism: Dilution of Upd



Candidate Upd

- Production restricted to initial stages of eye development (before MF starts)
- Fast Diffusion & Long Half Life
- Changes in Upd expression affect Eye Size



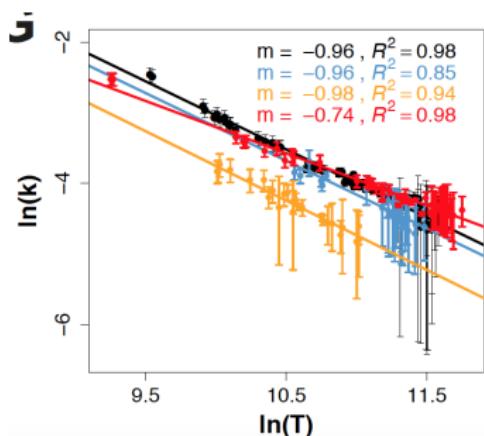
Bach et al., 2003

Control strains:

Gmr $>$, optix $>$,

Mutant strains:

Gmr $>$ Upd, optix $>$ dome

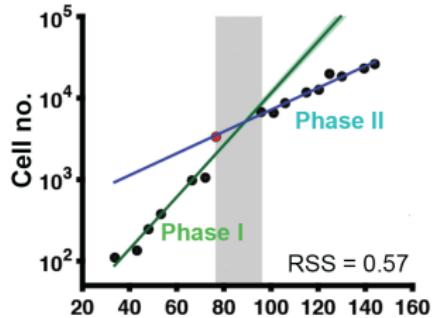
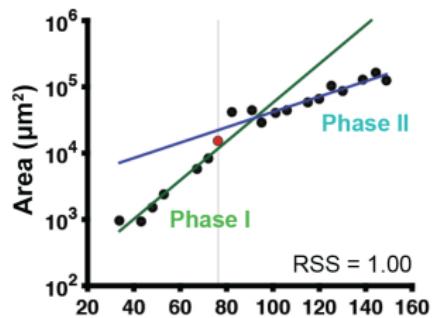
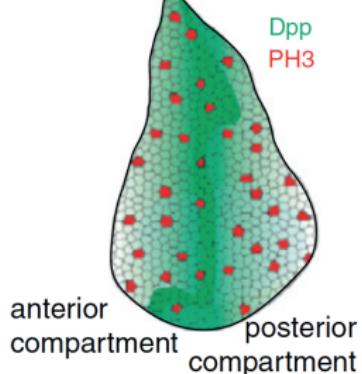
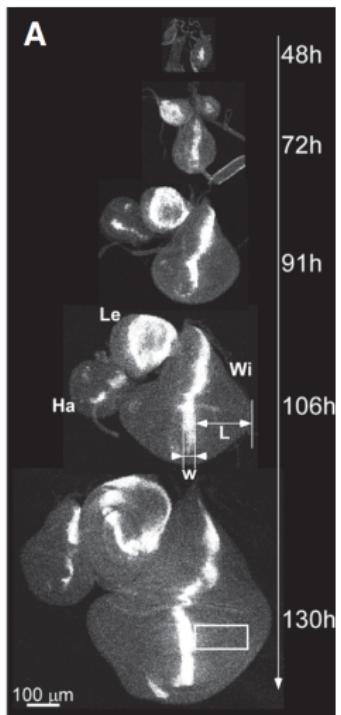


Conclusion on Growth Control in the *Drosophila* eye disc

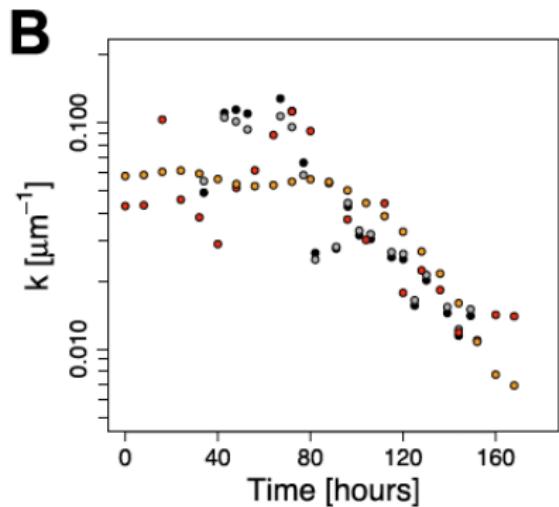
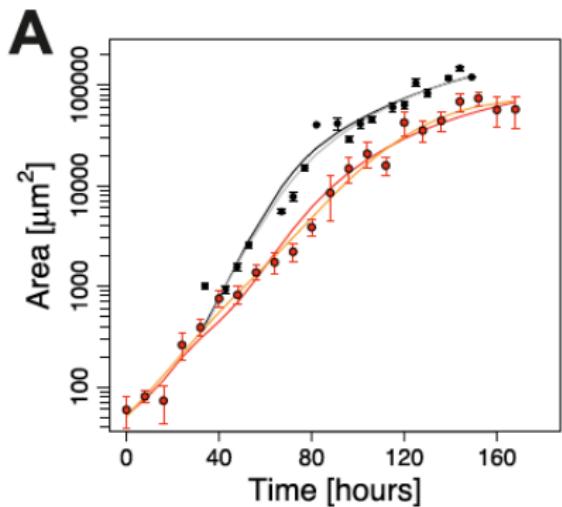
- Observed Decline in Growth Rate in the *Drosophila* eye disc can be explained by Dilution Mechanism
- Dilution of Cytokine Upd good candidate

Is growth control by dilution a likely general mechanism for growth termination during development?

The *Drosophila* Wing Disc as Model System

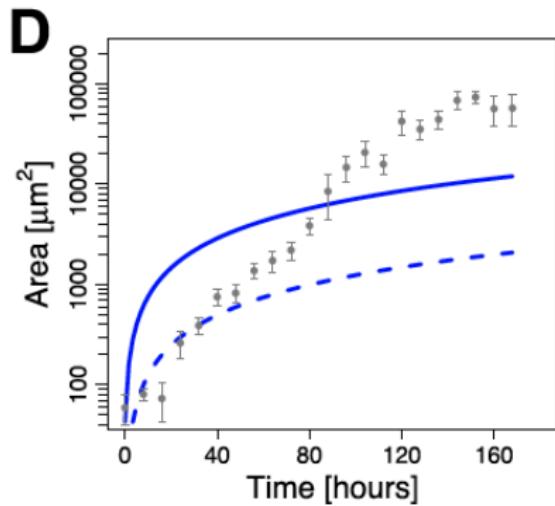
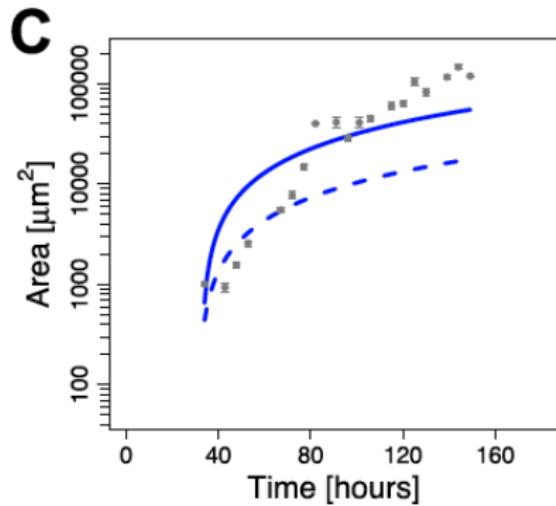


Growth Behaviour of *Drosophila* wing disc

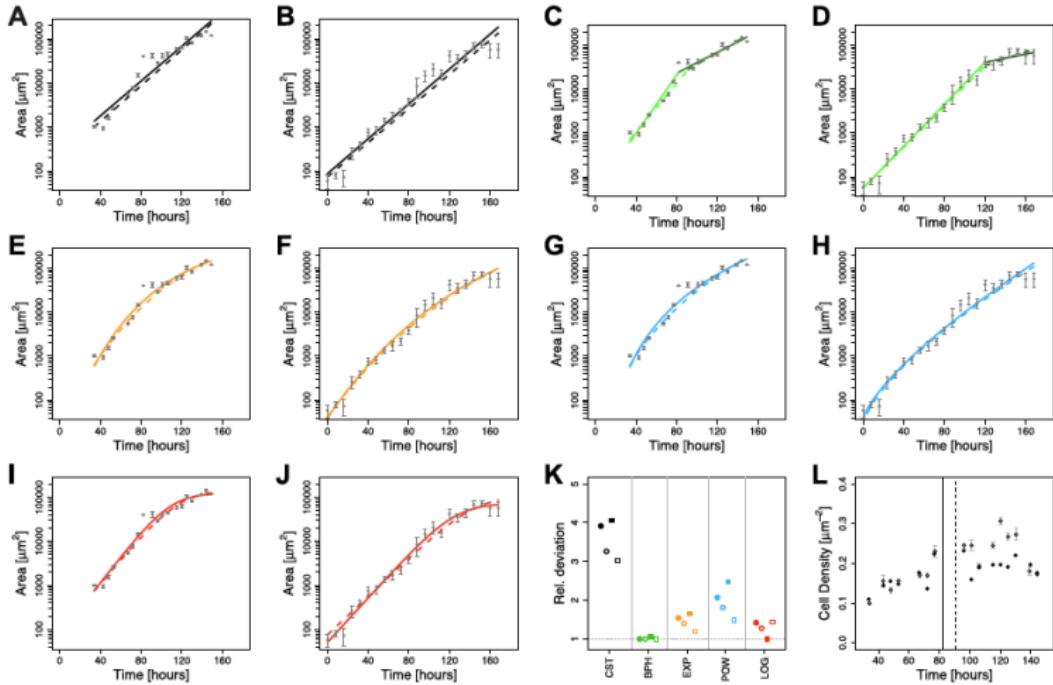


Vollmer, J. & D. Iber (2016), Sci Rep 6: 39228.

Dilution cannot explain Growth Behaviour of *Drosophila* wing disc



Comparison of alternative Growth Models



Vollmer, J. & D. Iber (2016), Sci Rep 6: 39228.

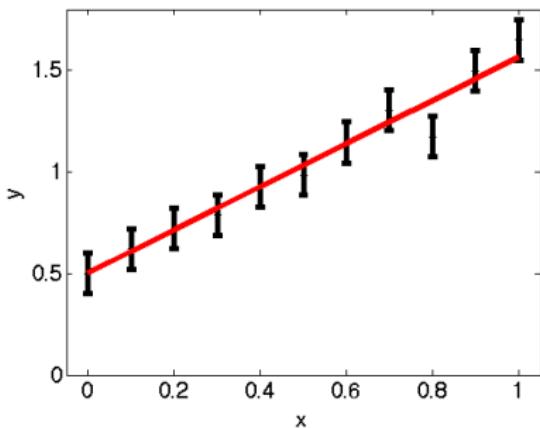
Problem Sheet: Model Selection

- 1** Maximal Likelihood Estimate
- 2** Optimization Algorithms
- 3** Model Selection

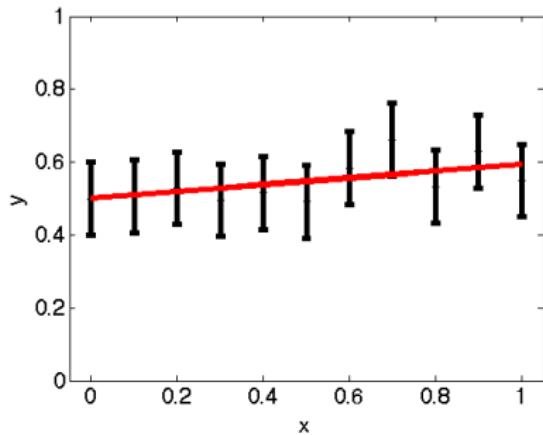
Data-based Model Selection

The problem: How to fit your data?

A fit of the function $y = ax + b$ (red) to the data (black).



A fit of the function $y = ax + b$ (red) to the data (black). Is this overfitting? Should we set $a = 0$?



Task: Fit a model (ODE, PDE, ...) to data - estimate parameters

Questions:

- What algorithms to use?
- How good is the fit?
- Is there a better model?
- Sensitivity?

Answer: Use Bayesian inference

Frequentist: Probability is a frequency (of a random variable)

Bayesian: Probability is a measure of uncertainty

Anything can be assigned a probability ⇒ **Models and parameters have probabilities**

Basic Rules of Probability

Basic Rules of Probability

1. The Complement Rule:

$$P(X|I) + P(\text{not } X|I) = 1 \quad (1)$$

2. The Multiplication Rule:

$$\begin{aligned} P(X, Y|I) &= P(X|Y, I) \times P(Y|I) \\ P(X, Y|I) &= P(Y|X, I) \times P(X|I) \end{aligned} \quad (2)$$

- $P(X)$ is read as 'the probability of X ', where X is an event we are interested in.
- $P(X|I)$ is read as 'the probability of X given I occurs'.
- $P(\text{not } X)$ is read as 'the probability of not X ,' or 'the probability that X does not occur'.

Bayes' Theorem

$$P(X, Y|I) = P(X|Y, I) \times P(Y|I)$$

$$P(X, Y|I) = P(Y|X, I) \times P(X|I)$$

Bayes' Theorem

$$P(X|Y, I) = \frac{P(Y|X, I) \times P(X|I)}{P(Y|I)} \quad (3)$$

- $P(X|Y, I)$ posterior probability
- $P(Y|X, I)$ likelihood
- $P(X|I)$ prior
- $P(Y|I)$ evidence

Bayes' Theorem for Parameter Estimation

- θ : vector with the parameter set that we want to estimate
- D : vector with the set of all measurements
- I : all relevant background information

According to Bayes' Theorem

$$P(\theta|D, I) = \frac{P(D|\theta, I) \times P(\theta|I)}{P(D|I)} \quad (4)$$

Bayesian inference reduces statistics to one equation.

Bayes' Theorem for Parameter Estimation

According to Bayes' Theorem

$$P(\theta|D, I) = \frac{P(D|\theta, I) \times P(\theta|I)}{P(D|I)}$$

- $P(\theta|D, I)$: **posterior probability density function (pdf)** that we want to determine.
- $P(D|\theta, I)$: **likelihood function**
- $P(\theta|I)$: **prior probability density function (pdf)** that reflects our knowledge about the system
- $P(D|I)$: **evidence**, i.e. the likelihood of the data based on our knowledge. Here one could incorporate knowledge about the quality of different experimental techniques or experimental groups.

Simplifying Bayes' Theorem

According to Bayes' Theorem

$$P(\theta|D, I) = \frac{P(D|\theta, I) \times P(\theta|I)}{P(D|I)}$$

We can simplify this equation by noting that

- 1 $P(D|I) = \int P(D|\theta, I)P(\theta|I)d\theta$ does not involve the parameters directly.
- 2 use uniform prior $P(\theta|I) = \text{const}$, i.e. all parameter values are equally probable at the start of the estimation procedure.

In real systems we typically have information that makes certain parameter sets more likely than others and we will discuss later how to incorporate such knowledge.

Simplifying Bayes' Theorem for Parameter Estimation

According to Bayes' Theorem

$$P(\theta|D, I) = \frac{P(D|\theta, I) \times P(\theta|I)}{P(D|I)}$$

We can then write

$$\begin{aligned} P(\theta|D, I) &\propto P(D|\theta, I) \\ \text{posterior pdf} &\propto \text{likelihood function.} \end{aligned} \tag{5}$$

The posterior probability density function (pdf) $P(\theta|D, I)$ is thus proportional to the likelihood function $P(D|\theta, I)$.

Maximum likelihood estimate

$$P(\theta|D, I) \propto P(D|\theta, I) \quad (6)$$

posterior pdf \propto likelihood function.

Maximum likelihood estimate

Our best estimate θ_0 , given by the maximum of the posterior, is equivalent to the solution that yields the greatest value for the probability of the observed data.

Assume independent measurements

Independent Measurements

If we further assume that measurements are independent we have

$$P(D|\theta, I) = \prod_{k=1}^N P(D_k|\theta, I) \quad (7)$$

According to the product rule

$$P(D_k, D_I|\theta, I) = P(D_k|D_I, \theta, I) \times P(D_I|\theta, I) \quad (8)$$

such that for independent measurements

$$P(D_k|D_I, \theta, I) = P(D_k|\theta, I). \quad (9)$$

Likelihood Function

To illustrate the approach we now consider idealized conditions, where

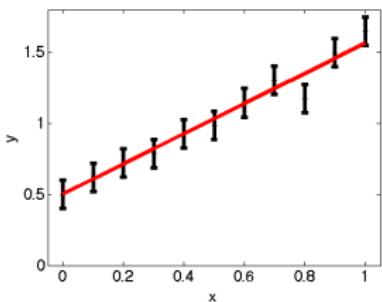
- the noise that is associated with the experimental measurements, D_k , can reasonably be represented by a Gaussian process such that

$$P(D_k|\theta, I) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(F_k - D_k)^2}{2\sigma_k^2}\right). \quad (10)$$

- our background information I includes a knowledge of the expected error bars $\{\sigma_k\}$,
- an adequate relationship f between the parameters θ and the ideal noiseless data F is given,

$$F_k = f(\theta, k). \quad (11)$$

Example: straight line



A fit of the function
 $y = ax + b$ (red) to the
data (black).

$$P(D_k | \theta, I) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp \left(-\frac{(F_k - D_k)^2}{2\sigma_k^2} \right)$$

In this case we have

- $D_k = y_k$ - mean of the data points at a given x_k
- σ_k^2 - variance of the data points at a given x_k
- $F_k = f(\theta, k)$ - model predictions at x_k
- $f(\theta, k) = \theta_1 x_k + \theta_2$ - functional relationship with parameters θ_i

Likelihood Function

$$P(D_k|\theta, I) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(F_k - D_k)^2}{2\sigma_k^2}\right)$$

For independent measurements we have $P(D|\theta, I) = \prod_{k=1}^N P(D_k|\theta, I)$

$$P(D|\theta, I) \propto \exp\left(-\frac{\chi^2}{2}\right) \quad \text{with} \quad \chi^2 = \sum_k R_k^2 = \sum_k \left(\frac{F_k - D_k}{\sigma_k}\right)^2.$$

Residuals

$R_k = \frac{F_k - D_k}{\sigma_k}$ are referred to as residuals.

Least-squares estimate

We now take the logarithm of the posterior pdf and obtain (for independent measurements and uniform prior $P(\theta|I)$)

Log-Likelihood

$$L = \ln(P(D|\theta, I)) = \text{const} - \frac{\chi^2}{2}. \quad (12)$$

Least-squares estimate

Since the maximum of the posterior will occur when χ^2 is smallest, the corresponding optimal solution θ_0 is called least-squares estimate.

Optimization

Optimization Algorithms

1 Local Methods

- Gradient-based Methods
 - Newton-Raphson Iterative Algorithm
 - Levenberg-Marquardt
- Direct, derivative-free Methods
 - Simplex Methods
 - Nelder-Mead Method
 - Conjugate Gradient Method

2 Global Methods

- Simulated Annealing
- Evolutionary Algorithms
- Swarm Particle Methods

Local Optimization Methods: Gradient-based Methods

Iterative Linearization

We can expand the likelihood function around its maximum L_0 as

$$L = L(\theta_0) + \underbrace{\sum_i \nabla_i L(\theta_i - \theta_{i0})}_{=0} + \underbrace{\frac{1}{2} \sum_{i,j} (\nabla \nabla L)_{ij}(\theta_i - \theta_{i0})(\theta_j - \theta_{j0})}_{\frac{1}{2} Q}$$

We will now start at some arbitrary point in parameter space θ_1 where the first derivative is no longer zero, i.e.

$$L = L(\theta_1) + \underbrace{\sum_i \nabla L(\theta - \theta_1)}_{\neq 0} + \underbrace{\frac{1}{2} \sum_{i,j} (\nabla \nabla L)_{ij}(\theta - \theta_1)(\theta - \theta_1)}_{\frac{1}{2} Q}$$

Iterative Linearization

We now develop a Taylor series for ∇L

$$\nabla L = \nabla L(\theta_1) + \nabla \nabla L(\theta_1)(\theta - \theta_1) + h.o.t.$$

At the optimum $\nabla L = 0$. We ignore the higher order terms and rearrange to find for the optimal parameter θ_0

$$\theta_0 \approx \theta_1 - [\nabla \nabla L(\theta_1)]^{-1} \nabla L(\theta_1)$$

The relationship will be exact if $\theta = \theta_0$, or if ∇L is truly linear.

Newton-Raphson Iterative Algorithm

Given a parameter (vector) θ and a likelihood function L :

Newton-Raphson Iterative Algorithm

Start with good estimate θ_1 ; evaluate the gradient vector ∇L at θ_1 .

While $\nabla L > \epsilon$:

- 1 evaluate the second derivative matrix $\nabla\nabla L$ at $\theta = \theta_1$
- 2 calculate an improved estimate $\theta_2 = \theta_1 - [\nabla\nabla L(\theta_1)]^{-1} \nabla L(\theta_1)$
- 3 evaluate the gradient vector ∇L at θ_2

Comments

- If ∇L is linear only one iteration required, i.e. $\theta_2 = \theta_0$ independent of θ_1
- The algorithm will rapidly converge as long as θ_1 is reasonably close to θ_0

Example 1: Fitting a straight line

Suppose we have N data $\{Y_k\}$, with associated error-bars $\{\sigma_k\}$ at positions $\{x_k\}$ and we wish to estimate the best estimate of the slope of a straight line given the intercept. For a straight line the k th ideal datum is given by

$$y_k = mx_k + c \quad (13)$$

where m is the slope and c is the intercept. Substituting $F_k = y_k$, and $D_k = Y_k$ we obtain

$$\chi^2 = \sum_k R_k^2 = \sum_k \left(\frac{F_k - D_k}{\sigma_k} \right)^2 = \sum_k \left(\frac{mx_k + c - Y_k}{\sigma_k} \right)^2. \quad (14)$$

Example 1: Fitting one parameter

In the first instance we assume that c is known and that we need to determine only a best estimate for m .

Given an initial estimate for $m = m_1$ we iterate according to

$$m_{n+1} = m_n - [\nabla \nabla L]^{-1} \nabla L(m_N). \quad (15)$$

where m_N is our estimate after $N - 1$ iterations.

Example 1: Fitting one parameter

$$\chi^2 = \sum_k R_k^2 = \sum_k \left(\frac{F_k - D_k}{\sigma_k} \right)^2 = \sum_k \left(\frac{mx_k + c - Y_k}{\sigma_k} \right)^2 \quad (16)$$

By differentiating χ^2 in Eq. 16 with respect to m we obtain

$$\nabla L(m_N) = -\frac{1}{2} \nabla \chi^2 = -\sum_k \frac{(m_N x_k + c - Y_k) x_k}{\sigma_k^2} \quad (17)$$

$$\nabla \nabla L = -\frac{1}{2} \nabla \nabla \chi^2 = -\sum_k \frac{x_k^2}{\sigma_k^2} \quad (18)$$

$\nabla \nabla L < 0$ so that we indeed obtain the maximal likelihood.

Example 1: Fitting two parameter of a straight line

In the next step we want to estimate both m and c . We use vector-matrix formulation and write

$$\begin{pmatrix} m_{N+1} \\ c_{N+1} \end{pmatrix} = \begin{pmatrix} m_N \\ c_N \end{pmatrix} - [\nabla \nabla L_N]^{-1} \nabla L_N. \quad (19)$$

where

$$\begin{aligned} \nabla L_N &= - \begin{pmatrix} \sum_k \frac{(m_N x_k + c_N - Y_k) x_k}{\sigma_k^2} \\ \sum_k \frac{m_N x_k + c_N - Y_k}{\sigma_k^2} \end{pmatrix}, \\ \nabla \nabla L &= - \begin{bmatrix} \sum_k \frac{x_k^2}{\sigma_k^2} & \sum_k \frac{x_k}{\sigma_k^2} \\ \sum_k \frac{x_k}{\sigma_k^2} & \sum_k \frac{1}{\sigma_k^2} \end{bmatrix}. \end{aligned} \quad (20)$$

Example: Fitting parameters in an ODE model

We now consider a simple ODE of the form

$$\frac{dy}{dt} = f(y) = -py \quad y(0) = 1 \quad (21)$$

where we have time-dependent data Y_k with variance σ_k at the time points t_k .

We seek to estimate the decay rate p .

As before we write

$$L = \text{const} - \frac{\chi^2}{2} \quad \text{with} \quad \chi^2 = \sum_k R_k^2 = \sum_k \left(\frac{y(t_k) - Y_k}{\sigma_k} \right)^2. \quad (22)$$

Example: Fitting parameters in an ODE model

We now seek to maximise the likelihood,

$$L = \text{const} - \frac{\chi^2}{2} \quad (23)$$

which is equivalent to minimise

$$\chi^2 = \sum_k R_k^2 = \sum_k \left(\frac{y(t_k) - Y_k}{\sigma_k} \right)^2. \quad (24)$$

As discussed in the previous lecture, we can use the **Newton-Raphson algorithm**: given an initial estimate for $p = p_1$ we iterate according to

$$p_{N+1} = p_N - [\nabla \nabla L]^{-1} \nabla L, \quad (25)$$

where the RHS is evaluated at p_N , the estimate after $N - 1$ iterations.

Example: Fitting parameters in an ODE model

By differentiating

$$\chi^2 = \sum_k R_k^2 = \sum_k \left(\frac{y(t_k) - Y_k}{\sigma_k} \right)^2. \quad (26)$$

with respect to p we obtain

$$\nabla L(p_N) = -\frac{1}{2} \nabla \chi^2 = -\sum_k \frac{(y(t_k) - Y_k) \frac{\partial y(t_k)}{\partial p}}{\sigma_k^2} \quad (27)$$

$$\nabla \nabla L = -\frac{1}{2} \nabla \nabla \chi^2 = -\sum_k \frac{\left(\frac{\partial y(t_k)}{\partial p} \right)^2}{\sigma_k^2}. \quad (28)$$

$\nabla \nabla L < 0$ so that we indeed obtain the maximal likelihood.

Example: Fitting parameters in an ODE model

$$\begin{aligned}\nabla L(p_N) &= -\frac{1}{2} \nabla \chi^2 = -\sum_k \frac{(y(t_k) - Y_k) \frac{\partial y(t_k)}{\partial p}}{\sigma_k^2} \\ \nabla \nabla L &= -\frac{1}{2} \nabla \nabla \chi^2 = -\sum_k \frac{\left(\frac{\partial y(t_k)}{\partial p}\right)^2}{\sigma_k^2}.\end{aligned}$$

So how can we determine the sensitivities $S_k = \frac{\partial y(t_k)}{\partial p}$? We notice that

$$\frac{d \frac{dy}{dp}}{dt} = \frac{d \frac{dy}{dt}}{dp} = \frac{df(y)}{dp} = \frac{\partial f(y)}{\partial y} \frac{dy}{dp} + \frac{\partial f(y)}{\partial p} = J \frac{dy}{dp} + \frac{\partial f(y)}{\partial p} \quad (29)$$

Example: Fitting parameters in an ODE model

$$\nabla L(p_N) = -\frac{1}{2} \nabla \chi^2 = -\sum_k \frac{(y(t_k) - Y_k) \frac{\partial y(t_k)}{\partial p}}{\sigma_k^2}$$

$$\nabla \nabla L = -\frac{1}{2} \nabla \nabla \chi^2 = -\sum_k \frac{\left(\frac{\partial y(t_k)}{\partial p}\right)^2}{\sigma_k^2}.$$

$$\frac{d\left(\frac{dy}{dp}\right)}{dt} = J \frac{dy}{dp} + \frac{\partial f(y)}{\partial p}$$

The Jacobian J and $\frac{\partial f(y)}{\partial p}$ can be calculated ahead of time and the entire equation can then be integrated alongside the integration of the model ODE.

Example: Fitting parameters in an ODE model

$$\frac{dy}{dt} = f(y) = -py; \quad y(0) = 1$$

$$\frac{dS}{dt} = JS + \frac{\partial f(y)}{\partial p} = -pS - y; \quad S(0) = 0$$

with Jacobian $J = \frac{\partial f(y)}{\partial y} = -p$, $\frac{\partial f(y)}{\partial p} = -y$, and sensitivity $S = \frac{\partial y}{\partial p}$.

OPTIMIZATION: $p_{N+1} = p_N - [\nabla \nabla L]^{-1} \nabla L$

$$\nabla L(p_N) = -\frac{1}{2} \nabla \chi^2 = -\sum_k \frac{(y(t_k) - Y_k) \frac{\partial y(t_k)}{\partial p}}{\sigma_k^2}$$

$$\nabla \nabla L = -\frac{1}{2} \nabla \nabla \chi^2 = -\sum_k \frac{\left(\frac{\partial y(t_k)}{\partial p}\right)^2}{\sigma_k^2}.$$

Example: Fitting 2 parameters in an ODE model

We now estimate two parameter values for a simple ODE of the form

$$\frac{dy}{dt} = f(y) = \rho - dy; \quad y(0) = 10. \quad (30)$$

The parameter vector is then given by $\vec{p} = [\rho, d]^T$.

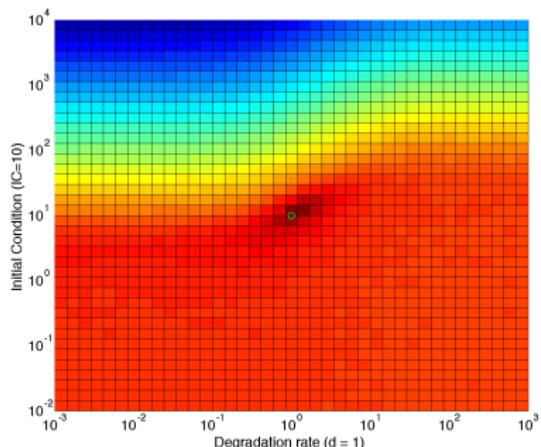
$$\text{OPTIMIZATION:} \quad \vec{p}_{N+1} = \vec{p}_N - [\nabla \nabla L]^{-1} \nabla L$$

$$\begin{aligned} \nabla L &= -\frac{1}{2} \nabla \chi^2 = -\sum_k \frac{(y(t_k) - Y_k)}{\sigma_k^2} \begin{pmatrix} \frac{\partial y(t_k)}{\partial \rho} \\ \frac{\partial y(t_k)}{\partial d} \end{pmatrix} \\ \nabla \nabla L &= -\begin{pmatrix} \sum_k \frac{1}{\sigma_k^2} \left(\frac{\partial y(t_k)}{\partial \rho} \right)^2 & \sum_k \frac{1}{\sigma_k^2} \frac{\partial y(t_k)}{\partial \rho} \frac{\partial y(t_k)}{\partial d} \\ \sum_k \frac{1}{\sigma_k^2} \frac{\partial y(t_k)}{\partial \rho} \frac{\partial y(t_k)}{\partial d} & \sum_k \frac{1}{\sigma_k^2} \left(\frac{\partial y(t_k)}{\partial d} \right)^2 \end{pmatrix}. \end{aligned}$$

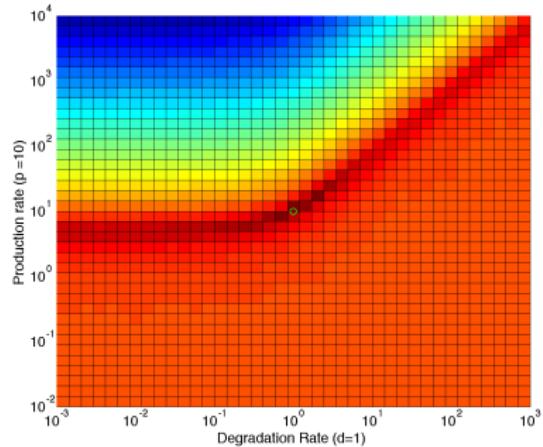
Example: Likelihood Function dependent on parameter values

$$\frac{dy}{dt} = f(y) = \rho - dy; \quad y(0) = 10.$$

Optimizing the IC and d.



Optimizing the correlated production and decay rates.



Levenberg-Marquardt

The stability of the iterative procedure can typically be improved by slowing down a little.

Slowing down could be achieved by multiplying the matrix-vector product $[\nabla \nabla L(X_1)]^{-1} \nabla L(X_1)$ by a fractional constant.

Better results are typically achieved by adding a small (negative) number c to all the diagonal elements of $\nabla \nabla L$:

$$X_{N+1} = X_N - [\nabla \nabla L(X_N) + cI]^{-1} \nabla L(X_N)$$

where I is the identity matrix.

Levenberg-Marquardt

The reason that this additional term stabilizes can be understood by considering the eigenvalue (λ_j) / eigenvector (e_j) pairs. We have

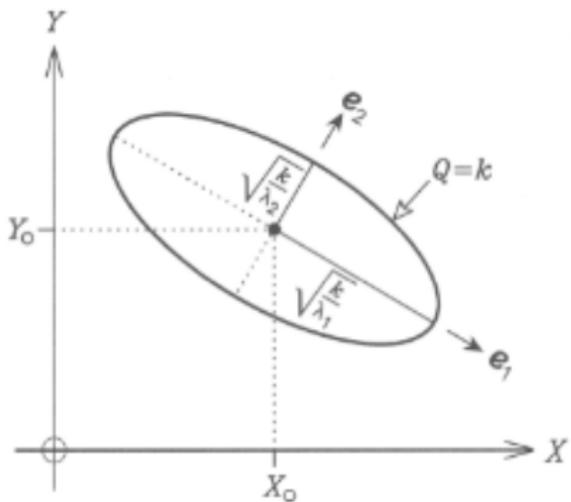
$$[\nabla \nabla L] e_j = \lambda_j e_j$$

The expanded matrix has the same eigenvectors, but different eigenvalues, i.e.

$$[\nabla \nabla L + cI] e_j = (\lambda_j + c) e_j.$$

Levenberg-Marquardt

$$X_{N+1} = X_N - [\nabla \nabla L(X_N) + cI]^{-1} \nabla L(X_N)$$



$$[\nabla \nabla L + cI]e_j = (\lambda_j + c)e_j.$$

For small eigenvalues λ_j (small determinant) as is the case when the ellipse is highly elongated in one direction, c reduces the influence of that direction and increases the determinant, thereby reducing the pre-factor of $\nabla L(X_N)$ as required.

Model Selection

Model Selection

Given some data, what is the best model?

Ideal Approach: Fit the data with the competing models and compare the posterior probability of the models.

In the exercise, you will instead use the Bayesian information criterion (BIC) and the Akaike information criterion (AIC).

Information Criteria

Information Criteria:

$$BIC = -2 \ln \hat{L} + k \ln n \quad AIC = -2 \ln \hat{L} + 2k$$

- \hat{L} : maximal likelihood.
 - k : number of parameters
 - n : number of data points
- AIC and BIC penalize the log of the likelihood with the number of parameters k and the number n of fitted data points.
- Because the first term grows linearly with the number n of fitted data points, while the second term in the BIC is proportional to $\ln n$, the penalty for having too many parameters is diminished as the data set gets larger.

Information Criteria

Information Criteria:

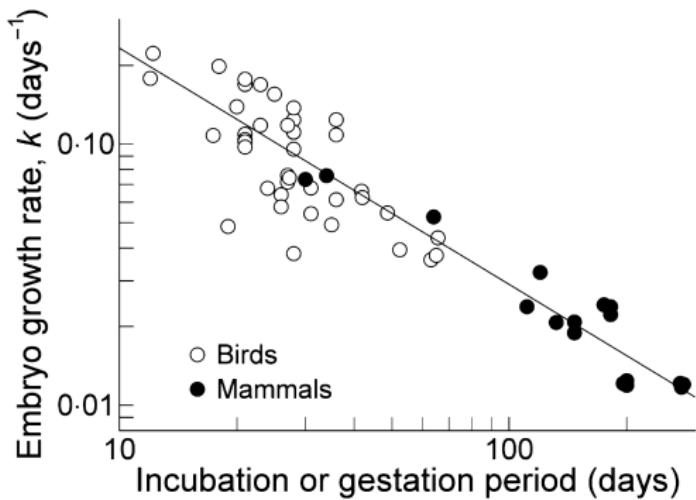
$$BIC = -2 \ln \hat{L} + k \ln n \quad AIC = -2 \ln \hat{L} + 2k$$

Caveats:

- These approximations are generally valid only if the likelihood function is near normal and peaked
- These approximation are only valid for sample size n much larger than the number k of parameters in the model.
- The BIC cannot handle complex collections of models.

How to explain growth termination?

What other mechanisms could you think of to explain growth termination?



Ricklefs, R. E. (2010). Embryo growth rates in birds and mammals. *Functional Ecology* 24(3): 588-596.

Thanks!!

Thanks for your attention!

Slides for this talk will be available at:

<http://www.bsse.ethz.ch/cobi/education>