SDE analysis of growth and energy intake for pigs

Helle Sørensen, Mathematical Sciences (NAT)
Anders Strathe, Animal Nutrition (LIFE)

KU Statistics Network meeting 2009
Jørlunde, January 22.

Outline

This is an application from my LIFE-time.

In short:

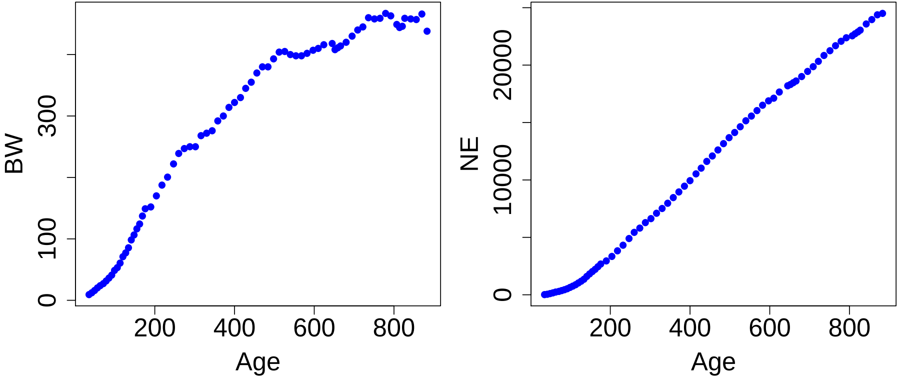
- What did we do already and what are the results?
- What is there to be skeptical about?

More detailed:

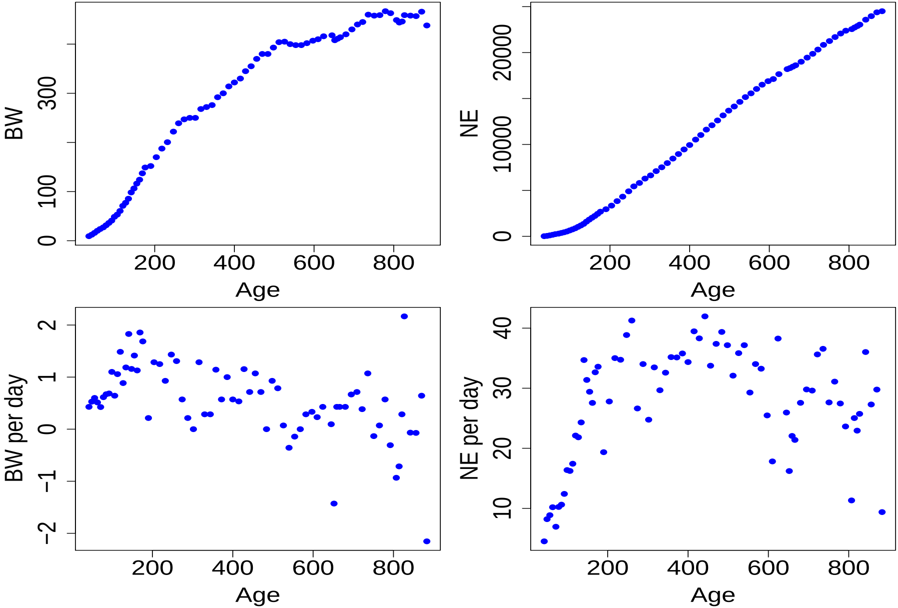
- The data and the questions
- The model: deterministic (ODE) and stochastic (SDE) version
- Estimation and results
- Investigation in a simpler model:
  - transformation
  - properties of the estimation procedure
  - random effects

Data from pig no. 6205: cumulated values

- Data from weaning to (near) maturity
- Bodyweight (*BW*) and net energy intake (*NE*) measured simultaneously weekly/every second week
- Data series to maturity for 13 pigs (shorter for another 175 pigs)



Data from pig no. 6205: increments per day



Questions of interest and concern

Nutrition

- Interplay between BW and NE — simultaneous model for (BW, NE)
- Estimation of maintenance requirement
- Prediction of body energy (BE) — only measurable at slaughter
- Estimation of population parameters and comparison of groups

Statistics

- SDE model as a way to incorporate 'dynamic' or 'intrinsic' noise?
- Variance homogeneity, transformation, auto-correlation
- Estimation methods and their properties, in particular their ability to distinguish between dynamic noise and measurement noise
- Random effects

Unusual data. SDE-models new to the animal science literature.

Outline

This is an application from my LIFE-time.

In short:

- What did we do already and what are the results?
- What is there to be skeptical about?

More detailed:

- The data and the questions
- The model: deterministic (ODE) and stochastic (SDE) version
- Estimation and results
- Investigation in a simpler model:
  - transformation
  - properties of the estimation procedure
  - random effects

The deterministic model

A two-dimensional ODE-model for (BW, NE).

Assumptions:

- NE rate quadratic as a function of metabolic weight ($BW^{0.75}$):

$$\frac{dNE}{dt} = \theta_1 \cdot BW^{0.75} + \theta_2 \cdot BW^{2 \cdot 0.75}$$

- Intake goes to maintenance and growth:

$$\frac{dBE}{dt} = \frac{dNE}{dt} - \theta_4 \cdot BW^{0.75}$$

- Relation between BW and BE (unobservable)

$$BW = \theta_5 \cdot BE^{\theta_6}$$

The deterministic model (cont'd)

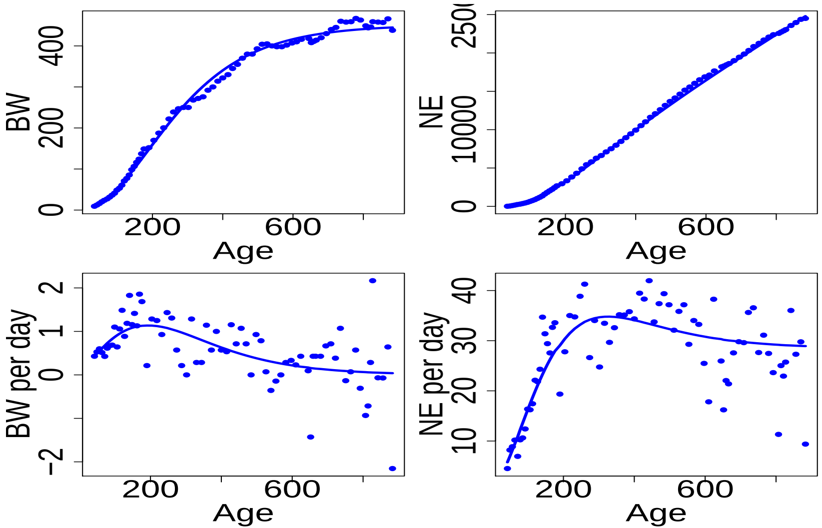
The assumptions lead to a 2-dimensional ODE for (BW, NE):

$$\frac{dBW}{dt} = f_{BW}(BW, NE, \theta_1, \theta_2, \theta_4, \theta_5, \theta_6)$$
$$\frac{dNE}{dt} = f_{NE}(BW, NE, \theta_1, \theta_2)$$

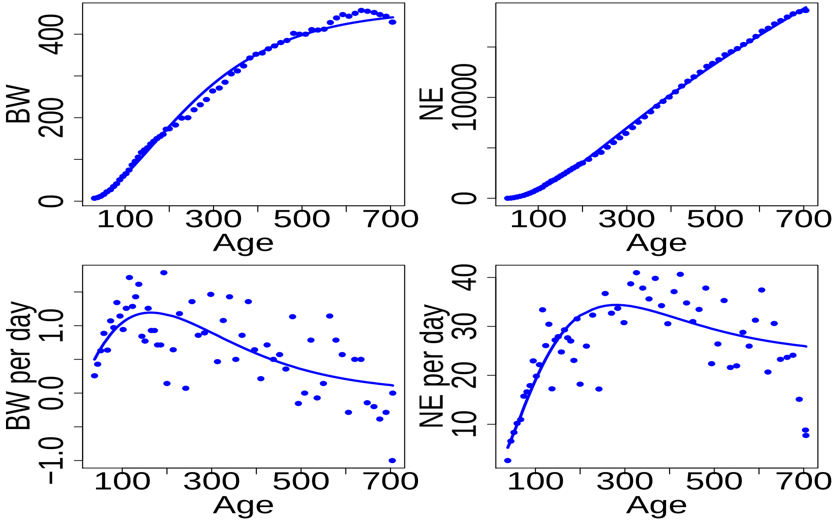
Is the model at all able to "catch" the gereral trends in the data?

- More or less, yes
- Auto-correlation (of course) in BW and NE
- The peak around 150–200 days for $\Delta BW/\Delta t$ perhaps not 'sharp' enough

Data from pig no. 6205



Data from pig no. 5365



Statistical models

Simple model for $y_i = (BW, NE)$:

$$y_i = \text{ODE solution} + e_i$$

with auto-correlated residuals.

Alternative: State space model with system equation defined by a SDE.

- SDE-model implicitly defines auto-correlation
- SDE-model defines models for increments over time-intervals of any length in a consistent way
- State space formulation: dynamic noise as well as measurement noise.
- Dynamic noise may originate from circumstances in the system that we cannot (or will not) include in the system model.
- The system equation is closely linked to scaled differences:

$$\frac{\Delta y}{\sqrt{\Delta t}} = \frac{\Delta(\text{ODE solution})}{\sqrt{\Delta t}}$$

SDE-based state space model

System (or state) equation: Two-dimensional SDE,

$$\begin{pmatrix} dBW \\ dNE \end{pmatrix} = \begin{pmatrix} f_{BW}(BW, NE, \theta) \\ f_{NE}(BW, NE, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{12} & \sigma_2 \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}$$

where

- f_{BW} and f_{NE} come from the ODE model, depend on $\theta_1, \ldots, \theta_6$
- σ_1, σ_2 and σ_{12} are parameters
- w_1, w_2 are independent Wiener processes

Measurement equation: The k'th observation pair, at time t_k:

$$\begin{pmatrix} y_{1,k} \\ y_{2,k} \end{pmatrix} = \begin{pmatrix} BW_{t_k} \\ NE_{t_k} \end{pmatrix} + \begin{pmatrix} e_{1,k} \\ e_{2,k} \end{pmatrix}$$

where $(e_{1,k}, e_{2,k})'$ are independent, two-dimensional Gaussian.

Comments on the model

What does the system equation mean?

For Δt "small",

$$\begin{pmatrix} \Delta BW \\ \Delta NE \end{pmatrix} \approx \begin{pmatrix} f_{BW}(BW, NE, \theta) \\ f_{NE}(BW, NE, \theta) \end{pmatrix} \Delta t + \sqrt{\Delta t} \begin{pmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{12} & \sigma_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

where u_1, u_2 are iid. standard normal.

Questions:

- What happens when Δt is not so small?
  True cond. distribution of $(\Delta BW_{t_k}, \Delta NE_{t_k})$ given $(BW_{t_{k-1}}, NE_{t_{k-1}})$?
- (BW, NE) observed with measurement noise (so observations are not Markov). Conditional distribution given the whole past?

Variance inhomogeneity and transformation

Not surprisingly, variance inhomogeneity is an issue.

One option: Diffusion coefficients and measurement noise variance defined as functions of states (BW, NE). Not handled by the software.

Another option: transformation.

Usually: 'SDE + transformation = Ito's formula'

Again, this is not handled by the software. And who says the original scale is the right one for constant variance?

Our approach: 'find scale where additive noise is reasonable'

- Transform-both-sides of the ODE-model so that parameter interpretations are maintained
- Diffusion term and measurement errors added as before

Outline

This is an application from my LIFE-time.

In short:

- What did we do already and what are the results?
- What is there to be skeptical about?

More detailed:

- The data and the questions
- The model: deterministic (ODE) and stochastic (SDE) version
- Estimation and results
- Investigation in a simpler model:
  - transformation
  - properties of the estimation procedure
  - random effects

Estimation

Parameters: the ODE-parameters (θ's), diffusion parameters (σ's), measurement noise variances (s's).

Estimation approach from Kristensen and Madsen, 2003 (DTU):

The conditional distribution of y_k given the past is approximated by the Gaussian distribution $\rightarrow$ likelihood function is a product of Gaussian densities.

The mean and variance are computed by the extended Kalman filter:

- prediction equations: predict state at time t_k from $y_1, \ldots, y_{k-1}$
- updating equations: 'predict' state at time t_k from $y_1, \ldots, y_{k-1}, y_k$

Implemented in the CTSM software (Kristensen and Madsen) and the R-package PSM (Mortensen and Klim).

Results from the analysis — and problems?

The square root transform turned out to work 'best'.

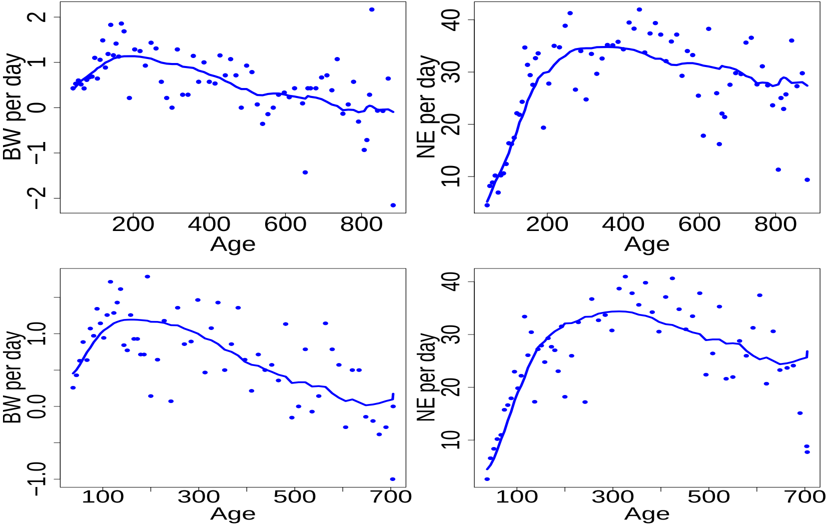
The good story:

- Reasonable fits and biologically meaningful parameter estimates
- Reasonable prediction of body energy — only measurable at slaugther.

The problems:

- Have used $\sigma_{12} = 0$ corr. to independent SDE noise for *BW* and *NE*. Not realistic — might affect SE's of the parameters of interest.
- Measurement noise variances estimated to zero for almost all animals. 'True' or is it too hard to distuinguish it from SDE noise?
- Properties of the estimation procedure: Bias? Coverage of confidence intervals? SDE noise versus measurement noise?
- The pigs analysed one by one. Would be natural to analyse all data simultaneously with random effects $\leftrightarrow$ population parameters.

Fitted drift for pig no. 6205 and 5365

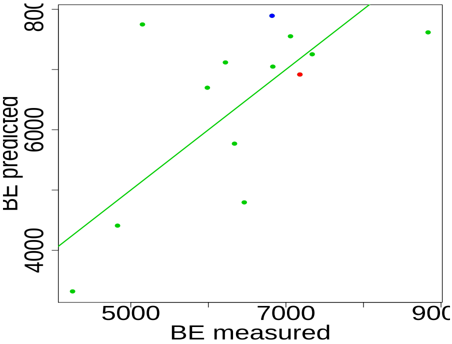


Prediction of body energy (BE)

The body energy (BE) was measured at slaughter.

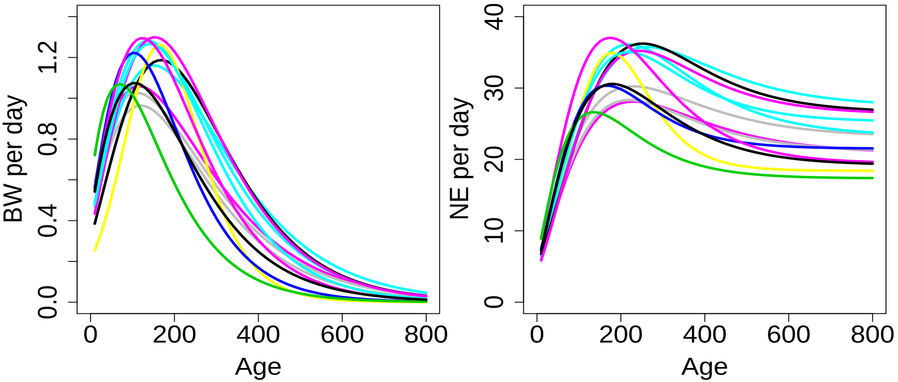
Predicted value,

$$\widehat{BE} = \left(\frac{BW}{\hat{\theta}_5}\right)^{1/\hat{\theta}_6}$$



Estimated ODE solutions for all pigs

Estimates correspond to the following increments of the ODE solutions:



- Natural to incorporate a random pig effect on some of the parameters
- Phase variation ↔ time-warping (functional data analysis)

Outline

This is an application from my LIFE-time.

In short:

- What did we do already and what are the results?
- What is there to be skeptical about?

More detailed:

- The data and the questions
- The model: deterministic (ODE) and stochastic (SDE) version
- Estimation and results
- Investigation in a simpler model:
  - transformation
  - properties of the estimation procedure
  - random effects

Properties of the estimation procedure

Investigations in a simpler set-up in order to speed up computations:

- Only weight observations (BW)
- Simpler ODE-model: Gompertz,

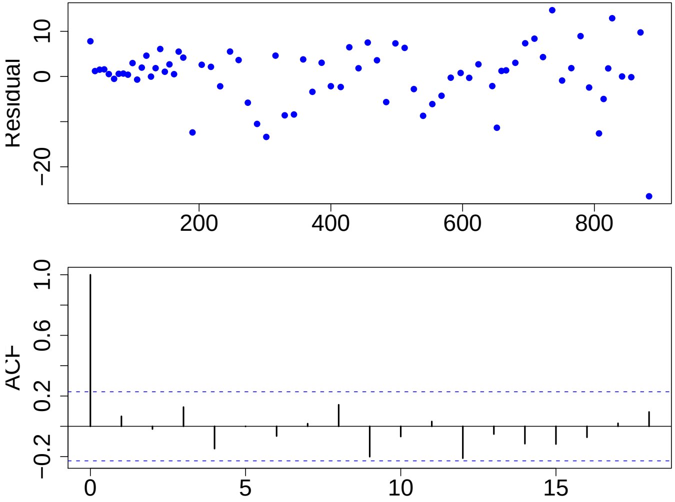
$$\frac{dBW}{dt} = \alpha \cdot BW \cdot \log(K/BW)$$

  where K is the 'asymptotic' and maximal weight, α determines the decrease in growth rate.

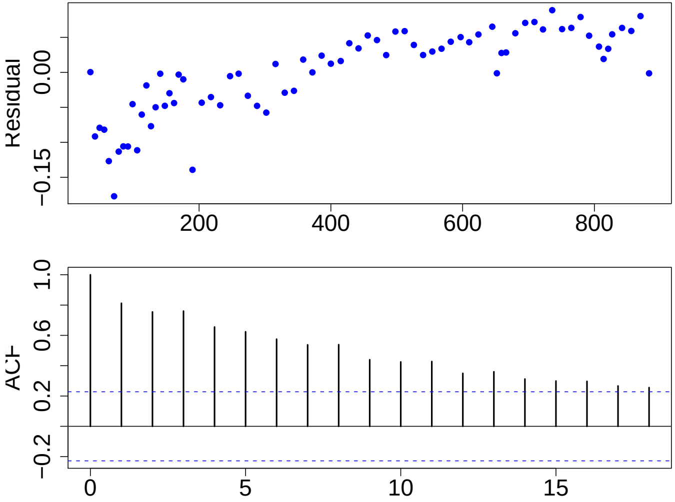
Statistical model: squareroot transformation of ODE, diffusion term, and measurement noise:

$$\begin{aligned} \frac{d\sqrt{BW}}{dt} &= \frac{1}{2}\alpha\sqrt{BW}\log\left(\frac{K}{\sqrt{BW}^2}\right)dt + \sigma\, dw_t \\ y_{1,k} &= BW_{t_k} + e_{1,k} \end{aligned}$$

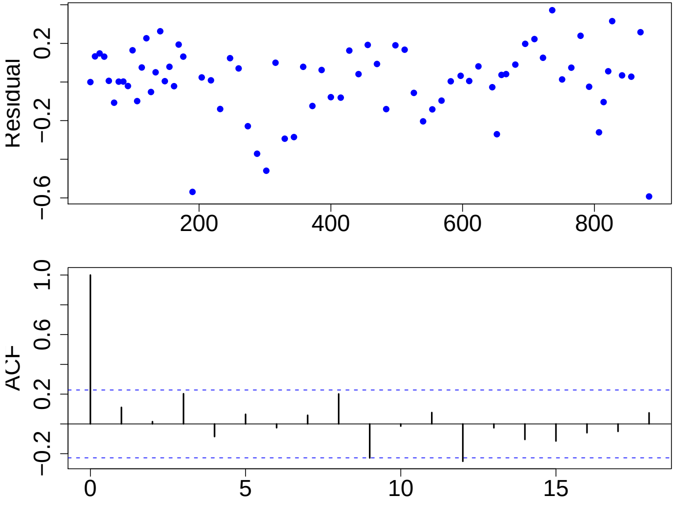
Original



Log



Squareroot



Simulations

Observation times equal to those of pig 6205.

Structural parameters b and K equal to values estimated for pig 6205:

$$K = 437, \qquad b = 0.00772$$

Three different regimes for σ (SDE noise) and s (measurement noise)

A: $\sigma \approx 1000 \cdot s$

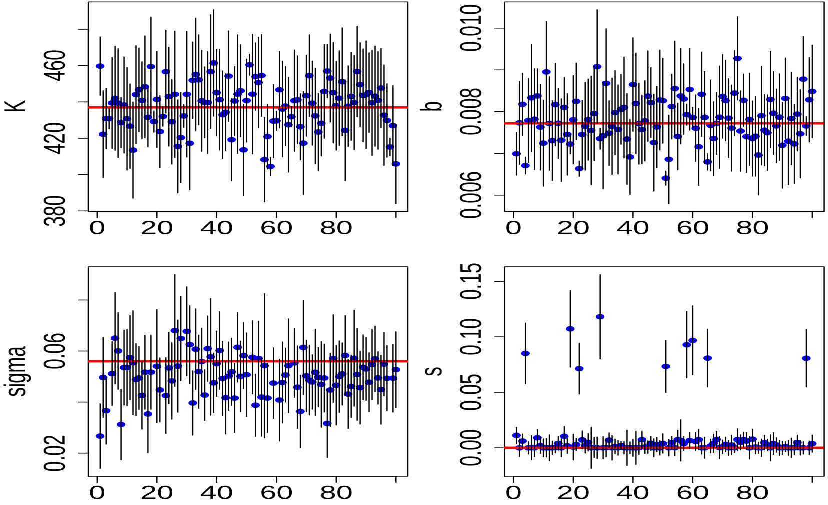
B: $s \approx 1000 \cdot \sigma$

C: $\sigma \approx s$

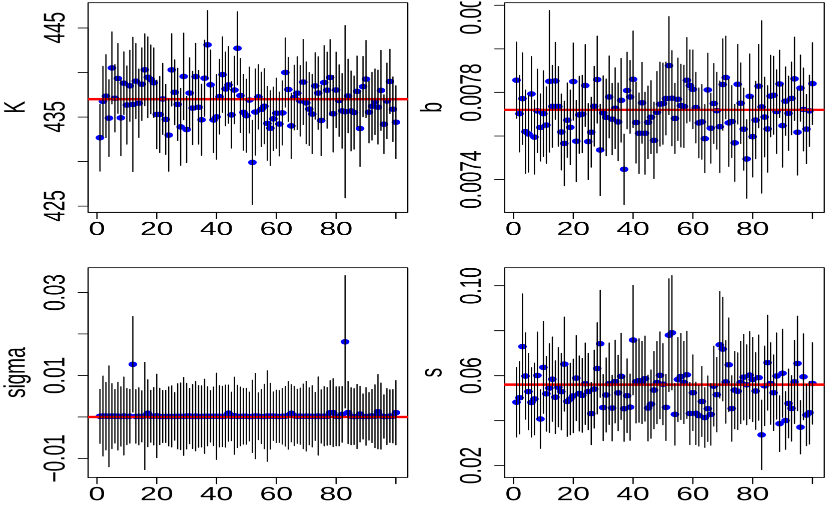
100 simulated paths.

For each path: computation of estimate and 95% confidence intervals (based on the estimated Fisher information).

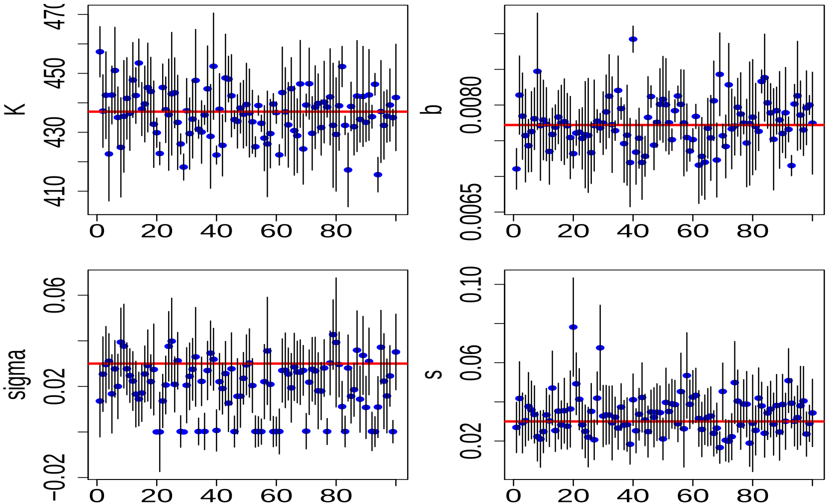
Case A: $\sigma = 0.0560$, $s = 0.00005$



Case B: $\sigma = 0.00005$, $s = 0.0560$



Case C: $\sigma = 0.03$, $s = 0.03$



Conclusions

Conclusions on simulations

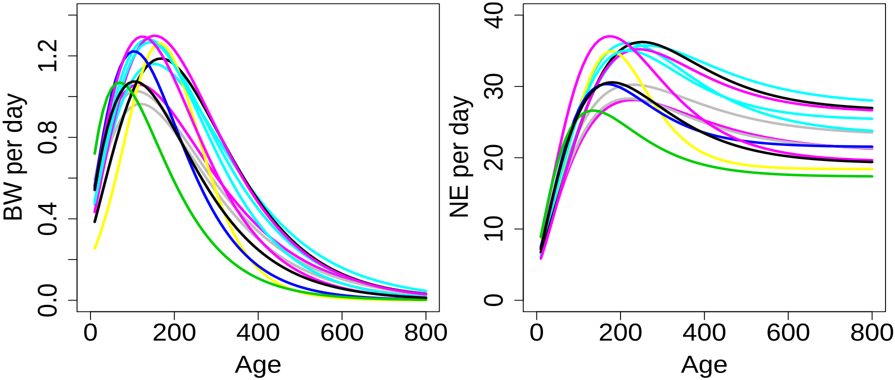
- There *are* problems with the estimation of σ and s
- In particular the procedure sometimes finds the wrong regime (A: 9/100 and C: 18/100).
- ... but the K and b estimates are (almost) unaffected by this: no bias
- Coverage of K and b confidence intervals seems to be okay, except in case C where it is only about 0.8.

Comments on random effects

- Methods for SDE-models with random effects and measurement noise are implemented in the R-package PSM.
- ... at least in principle: I had serious problems with estimation of the between-animal variation of parameters
- Alternative approach: time-shifts/time-warping

Estimated ODE solutions for all pigs

Estimates correspond to the following increments of the ODE solutions:



- Natural to incorporate random effects
- Phase variation $\leftrightarrow$ time-warping (functional data analysis)