Nonlinear modeling: Case study: river discharge

Introduction to Statistical Modelling

Prof. Joris Vankerschaver

Model overview

Setting

Streeter-Phelps model

Use water pollution as water quality monitoring tool. Describes how dissolved oxygen decreases in a river along a certain distance by degradation of biological oxygen demand.

- Aerobic bacteria gradually remove organic pollution downstream of pollution source
- Reactions
  - Aerobic removal of biochemical oxygen demand
  - Oxygen transfer between atmosphere and water
- Assumption: plug-flow stream
- Simple dynamical model (nonlinear)

Setting

Typical values for rivers

- Biochemical oxygen demand (BOD)
    - Not polluted: BOD < 1mg/l
    - Mildly polluted: 2mg/l $<$ BOD $<$ 8mg/l
- Dissolved oxygen (DO)
    - Maximal saturation: DO $= 12.9$mg/l
    - Typical value in freshwater stream: DO ≈ 9mg/l
    - Threat to aquatic life: DO < 5mg/l

Model

Constant flow rate: location doesn't matter, dynamic model **in time**

$$\frac{d\text{BOD}}{dt} = \text{BOD}_{in} - k_1\text{BOD}$$
$$\frac{d\text{DO}}{dt} = k_2(\text{DO}_{sat} - \text{DO}) - k_1\text{BOD}$$

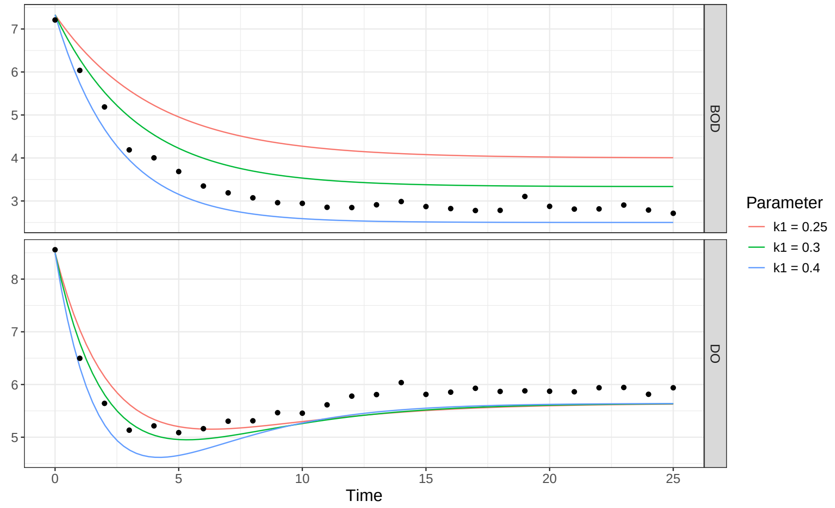
where

- BOD_{in}: BOD flux of waste discharge ($\text{mg} \cdot \text{l}^{-1} \cdot \text{min}^{-1}$)
- DO_{sat}: dissolved oxygen concentration at saturation
- k_1: deoxygenation rate (min^{-1})
- k_2: reaeration rate, rate at which oxygen can be absorbed from the atmosphere (min^{-1})

Model inputs

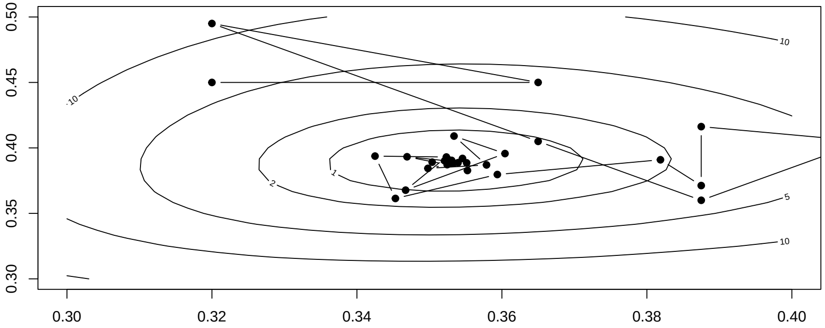
- Initial conditions:
  - $\text{BOD}_{t=0} = 7.33\text{mg/l}$
  - $\text{DO}_{t=0} = 8.5\text{mg/l}$
- Initial model inputs:
  - $\text{BOD}_{in} = 1\text{mg}\cdot\text{l}^{-1}\ \cdot\text{min}^{-1}$
  - $\text{DO}_{sat} = 8.5\text{mg}\cdot\text{l}^{-1}$
  - $k_1 = 0.3\text{min}^{-1}$ (**unknown**)
  - $k_2 = 0.4\text{min}^{-1}$ (**unknown**)

Model trajectories

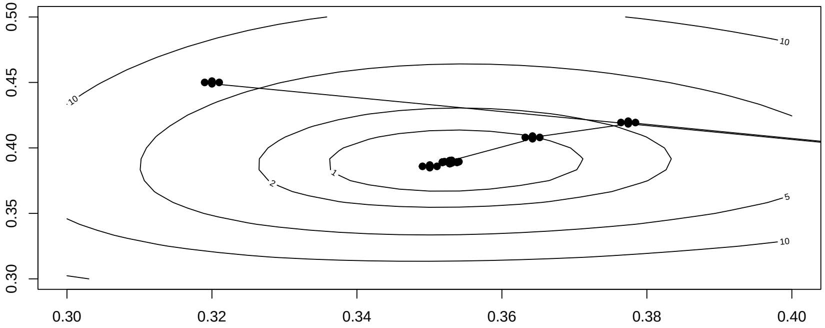


Parameter estimation

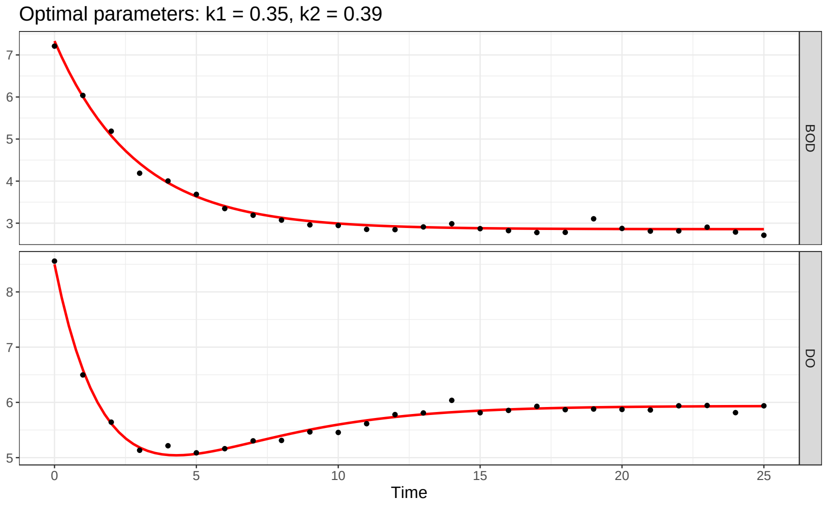
Parameter estimation - simplex method



Parameter estimation - BFGS



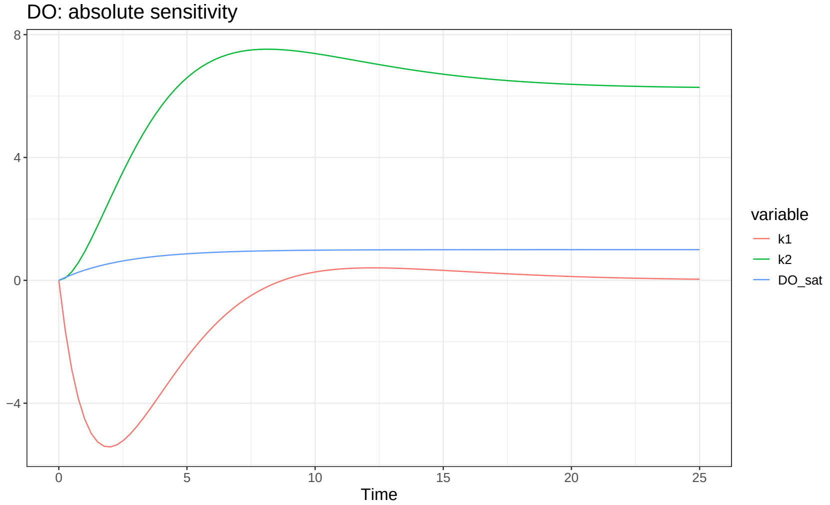
Optimal parameter result



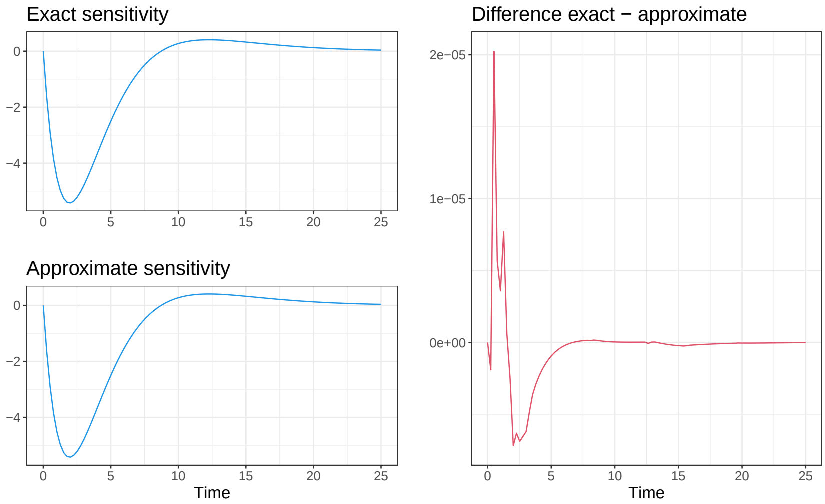
Sensitivity analysis

Absolute sensitivity functions

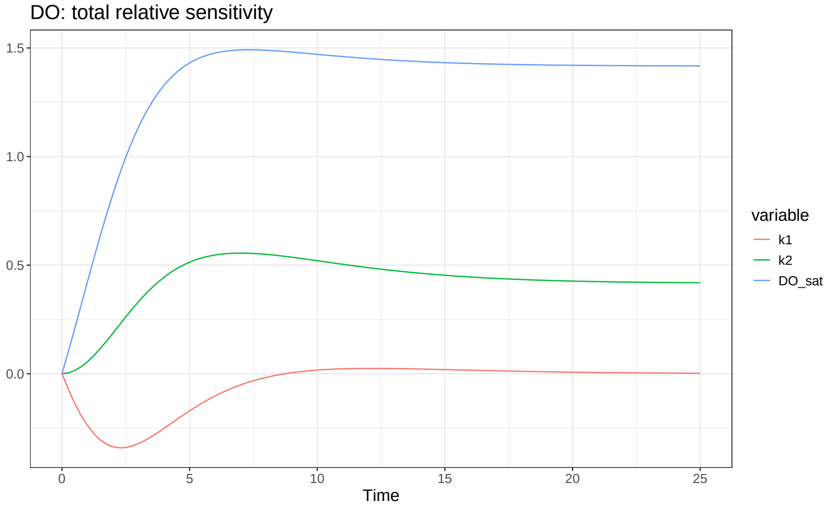
Absolute sensitivity of DO with respect to k_1 and k_2.



Difference exact approximate



Relative sensitivity functions



Conclusions

- Note difference in values with absolute sensitivities
- DO seems more sensitive to k_2 than to k_1
- Extrema at slightly different time points: information concerning correlation between both parameters
    - Maximal sensitivity at same time point: parameters strongly correlated (impact of change in parameters similar)
- Studying sensitivity very valuable: sensitivity used in many techniques for model analysis

Aside: quality of estimation

- DO measurements at different times: measurement error (obtained manually) is 0.05 mg$\cdot$l^{-1}
- Estimate simultaneously k_1 and k_2 (assume all other parameters and initial conditions constant)
- Gives $k_1 = 0.353$min^{-1}, $k_2 = 0.389$

Quality of estimation: FIM

Fisher information matrix:

$$\mathsf{FIM} = \sum_{i=1}^{N} \left(\frac{\partial y}{\partial \theta}(t_i) \right)^T Q_i \left(\frac{\partial y}{\partial \theta}(t_i) \right)$$

- Measurement noise $\sigma_{DO} = 0.05 \mathrm{mg\,l}^{-1}$
- Weight "matrix" in the objective function $Q = \sigma_{DO}^{-2}$.

Gives:

$$\mathsf{FIM} = \begin{bmatrix} 3.91 \cdot 10^4 & -2.96 \cdot 10^4 \\ -2.96 \cdot 10^4 & 4.56 \cdot 10^5 \end{bmatrix}$$

Quality of estimation: confidence intervals

Error covariance matrix:

$$\mathsf{C} = \mathsf{FIM}^{-1} = \begin{bmatrix} 2.69 \cdot 10^{-5} & 1.75 \cdot 10^{-6} \\ 1.75 \cdot 10^{-6} & 2.31 \cdot 10^{-6} \end{bmatrix}$$

95% confidence intervals:

$$k_1 : 0.353 \pm 0.011$$
$$k_2 : 0.389 \pm 0.003$$

Covariance:

$$\mathsf{cor}(k_1, k_2) = 0.22$$

Quantitative analysis

- Calculate δ_{rmsq} for DO and different parameters
- Illustration: different measuring schemes for DO:
  - Scheme 1: $t_k = 0 : 0.1 : 2$
  - Scheme 2: $t_k = 0 : 2 : 20$
  - Scheme 3: $t_k = 0 : 2 : 10$
  - Scheme 4: $t_k = 10 : 2 : 20$

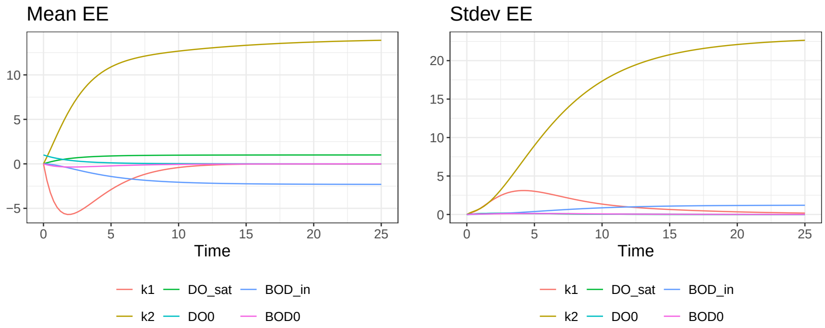
	Scheme 1	Scheme 2	Scheme 3	Scheme 4
k_1	0.24	0.13	0.18	0.18
k_2	0.09	0.44	0.43	0.43
DO_{sat}	0.49	1.32	1.23	1.23
	$k_2 < k_1 < \text{DO}_{sat}$	$k_1 < k_2 < \text{DO}_{sat}$	$k_1 < k_2 < \text{DO}_{sat}$	$k_1 < k_2 < \text{DO}$

GSA: Morris screening: parameter ranges

- Global sensitivity for 6 parameters and initial conditions in designated ranges:

$$k_1 : [0.1; 0.6] \quad k_2 : [0.1; 0.6] \quad \mathsf{DO}_{sat} : [10; 12]$$
$$\mathsf{BOD}_{in} : [0.1; 2] \quad \mathsf{DO}_{t=0} : [6; 10] \quad \mathsf{BOD}_{t=0} : [6; 10]$$

GSA: Morris screening

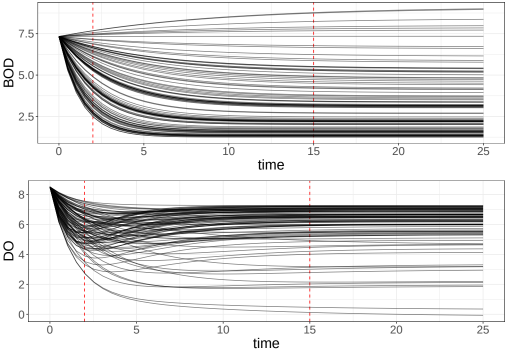


Among the selected parameters:

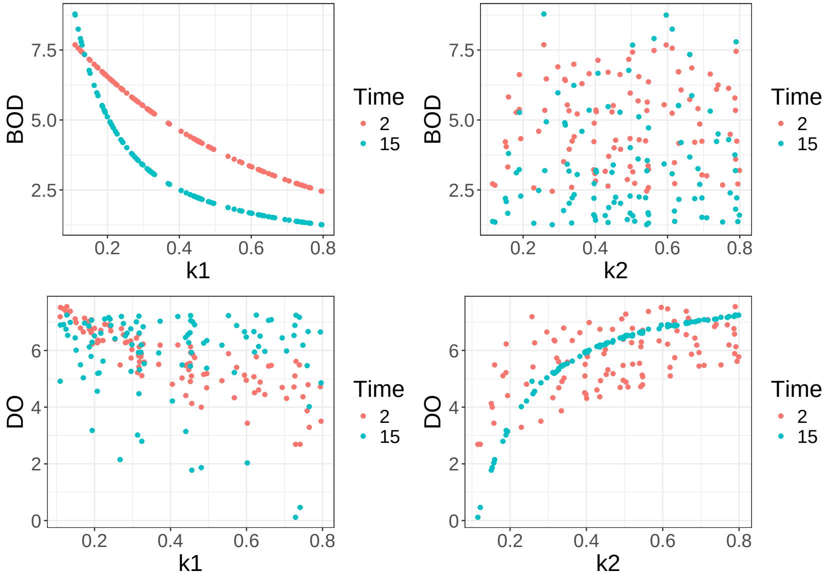
- At time 0, DO is only sensitive to DO_0
- At time 25, DO is sensitive to k_2 and DO_{sat}.

GSA: Monte Carlo

- Monte Carlo: 100 simulations with varying k_1 and k_2
- Parameter ranges: k_1, k_2 uniformly sampled from $[0.1, 0.8]$
- Interested in parameter effect at $t = 2$ and $t = 15$



GSA: Monte Carlo



GSA: Standardized regression coefficients

Previous plots show:

- BOD is sensitive to k_1 at $t = 2, 15$
- DO is sensitive to k_2 at $t = 15$ (and somewhat at $t = 2$)

Regression coefficients for DO at $t = 15$:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -0.342 \\ 6.374 \end{bmatrix}.$$

Standardized regression coefficients (using $\sigma_{k_1} = 0.211$ and $\sigma_{DO} = 1.887$):

$$SRC_{k_1} = -0.038, \quad \text{and} \quad SRC_{k_2} = 0.713.$$