Eawag
Swiss Federal Institute of Aquatic Science and Technology

eawag aquatic research

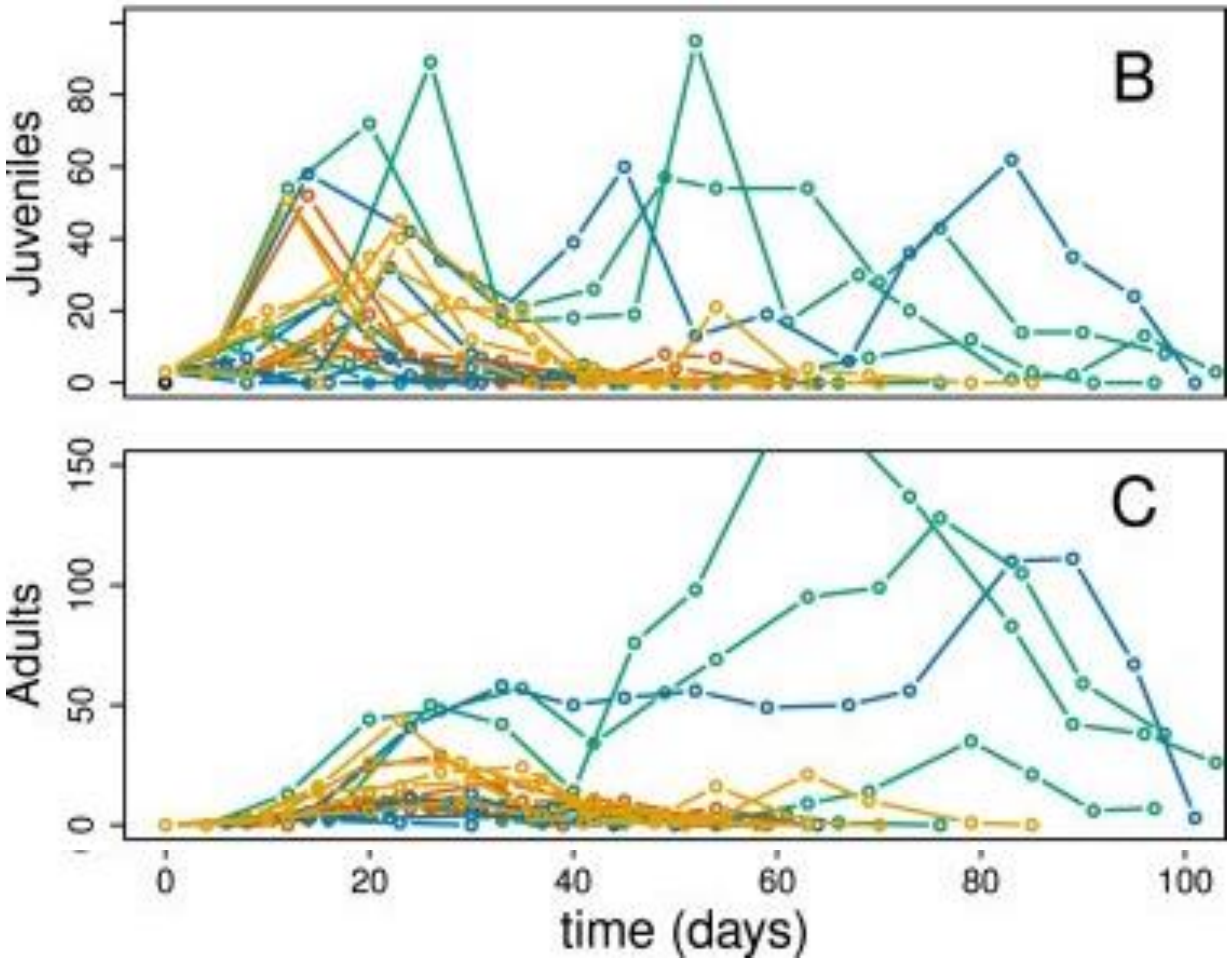
Stochasticity in Models

21.5.2024 – Andreas Scheidegger

Goals of today

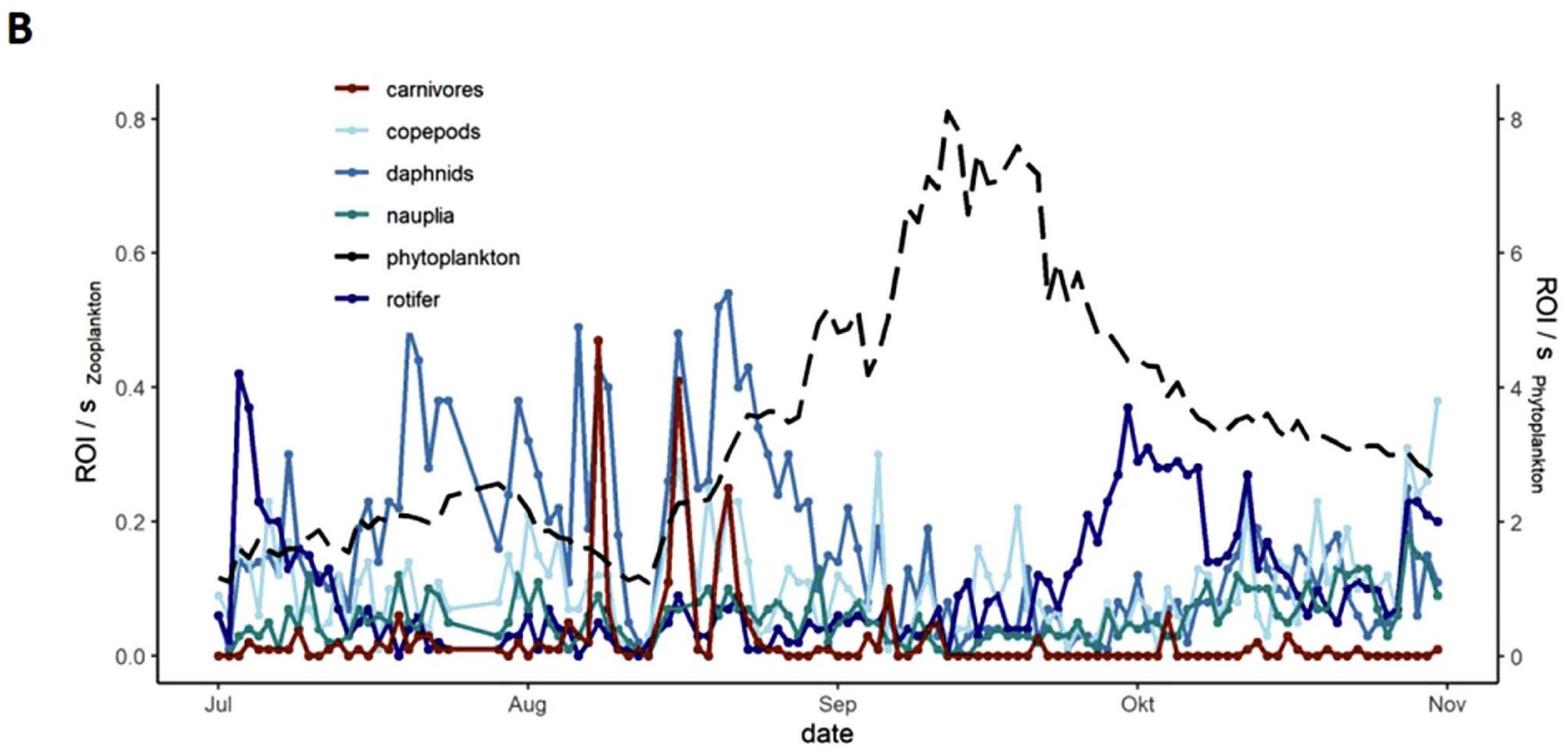
i. You can identify reasons to include stochasticity in a model
ii. You can distinguish error propagation and *stochastic parameters*
iii. You can know the effect of Ornstein-Uhlenbeck process parameters

Does this look like our model ouput? Why not?



Palamara et al. 2023. Investigating the effect of pesticides on Daphnia population dynamics by inferring structure and parameters of a stochastic model.
https://doi.org/10.1016/j.ecolmodel.2022.110076

Does this look like our model ouput? Why not?

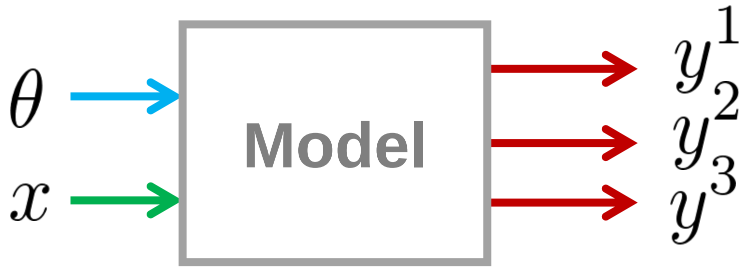


Merz, Ewa, et al. (2021). "Underwater dual-magnification imaging for automated lake plankton monitoring." *Water Research* 203 117524.

Models are wrong: sources of uncertainty

Almost all models suffer from these sources of uncertainty:

1. Input uncertainty
2. Parameter uncertainty
3. Model structure uncertainty
4. Output (measurement) uncertainty
5. (Scenario uncertainty)



Sources of stochasticity

eawag aquatic research

- **Environmental stochasticity**
  - Spatial variability
  - Missing factors
  - Fast fluctuations that are not measured
- **Genetic stochasticity**
  - Changes in the genetic composition of a population even in the absence of selective forces
- **Demographic stochasticity**
  - Individual behaviour influences the population
  - E.g. death, birth, food intake are individual

Almost all models suffer from these sources of uncertainty:

1. Input uncertainty
2. Parameter uncertainty
3. Model structure uncertainty
4. Output (measurement) uncertainty
5. (Scenario uncertainty)

In which category of uncertainty do this stochastic influences belong to?

Sources of stochasticity

eawag aquatic research

- **Environmental stochasticity**
  - Spatial variability
  - Missing factors
  - Fast fluctuations that are not measured
- **Genetic stochasticity**
  - Changes in the genetic composition of a population even in the absence of selective forces
- **Demographic stochasticity**
  - Individual behaviour influences the population
  - E.g. death, birth, food intake are individual

Almost all models suffer from these sources of uncertainty:

1. Input uncertainty
2. Parameter uncertainty
3. Model structure uncertainty
4. Output (measurement) uncertainty
5. (Scenario uncertainty)

Where should we add noise?

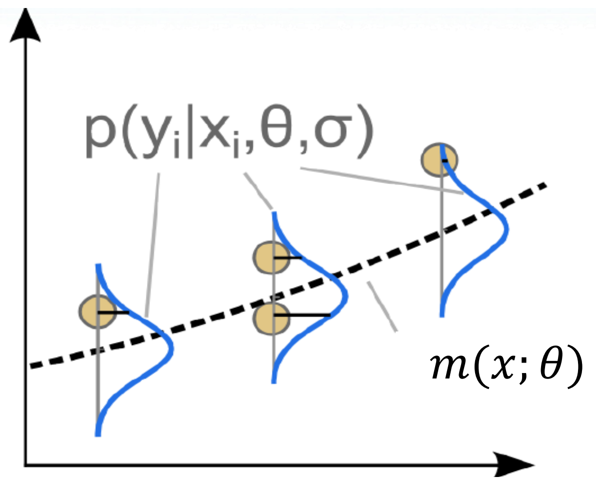
eawag aquatic research

Ordinary Differential Equation:
→ additive noise on solution

$$\frac{dx}{dt} = f(x, t, \theta)$$

ODE solver

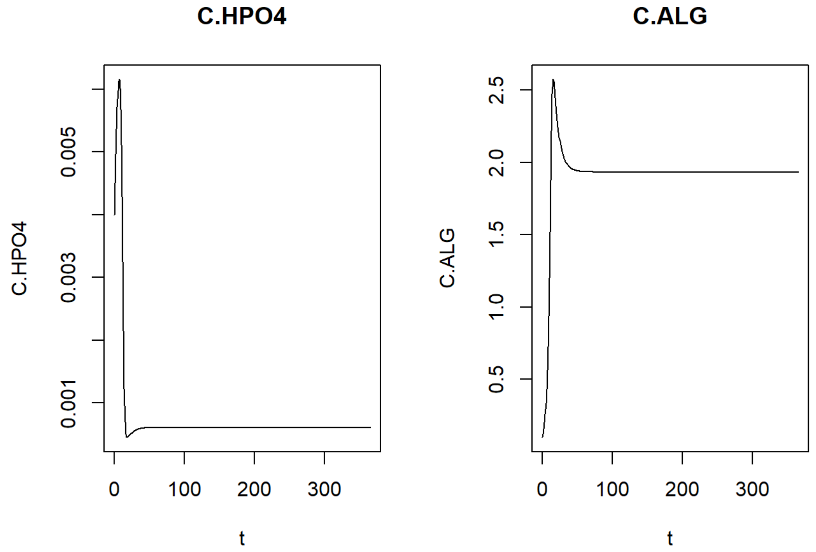
$$X(t) = m(t, \theta)$$



Phytoplankton Model

$$\frac{\mathrm{d}C_{\mathrm{HPO_4^{2-}}}}{\mathrm{d}t} = \frac{Q_{\mathrm{in}}}{V}\left(C_{\mathrm{in,HPO_4^{2-}}} - C_{\mathrm{HPO_4^{2-}}}\right) - \alpha_{\mathrm{P,ALG}} \cdot k_{\mathrm{gro,ALG}} \frac{C_{\mathrm{HPO_4^{2-}}}}{K_{\mathrm{HPO_4^{2-},ALG}} + C_{\mathrm{HPO_4^{2-}}}} C_{\mathrm{ALG}}$$

$$\frac{\mathrm{d}C_{\mathrm{ALG}}}{\mathrm{d}t} = -\frac{Q_{\mathrm{in}}}{V} C_{\mathrm{ALG}} + k_{\mathrm{gro,ALG}} \frac{C_{\mathrm{HPO_4^{2-}}}}{K_{\mathrm{HPO_4^{2-},ALG}} + C_{\mathrm{HPO_4^{2-}}}} C_{\mathrm{ALG}} - k_{\mathrm{death,ALG}}\, C_{\mathrm{ALG}}$$



Process	Substances / Organisms HPO4 $[\mathrm{gP/m^3}]$	ALG $[\mathrm{gDM/m^3}]$	Rate
Growth of algae	$-\alpha_{\mathrm{P,ALG}}$	1	$\rho_{\mathrm{gro,ALG}}$
Death of algae		-1	$\rho_{\mathrm{death,ALG}}$

Demo

Additive Noise

(Ordinary Differential Equation + Gaussian Error)

Where should we add noise?

eawag aquatic research

Ordinary Differential Equation:
→ additive noise on solution

$$\frac{dx}{dt} = f(x, t, \theta)$$

ODE solver

$$X(t) = m(t, \theta) + \epsilon(t)$$

Stochastic Differential Equation
→ internal noise

$$\frac{dx}{dt} = f(x, t, \theta) + \eta(t)$$

SDE solver

$$X(t) = ?$$

Phytoplankton Model as Stochastic Differential Equation

$$\frac{dC_{\mathrm{HPO_4^{2-}}}}{dt} = \frac{Q_{\mathrm{in}}}{V}\left(C_{\mathrm{in,HPO_4^{2-}}} - C_{\mathrm{HPO_4^{2-}}}\right) - \alpha_{P,\mathrm{ALG}}\, k_{\mathrm{gro,ALG}}\, \frac{C_{\mathrm{HPO_4^{2-}}}}{K_{\mathrm{HPO_4^{2-},ALG}} + C_{\mathrm{HPO_4^{2-}}}}\, C_{\mathrm{ALG}} + \eta_1(t)$$

$$\frac{dC_{\mathrm{ALG}}}{dt} = -\frac{Q_{\mathrm{in}}}{V} C_{\mathrm{ALG}} + k_{\mathrm{gro,ALG}}\, \frac{C_{\mathrm{HPO_4^{2-}}}}{K_{\mathrm{HPO_4^{2-},ALG}} + C_{\mathrm{HPO_4^{2-}}}}\, C_{\mathrm{ALG}} - k_{\mathrm{death,ALG}}\, C_{\mathrm{ALG}} + \eta_2(t)$$

Demo

Internal Noise

(Stochastic Differential Equation)

Stochastic Differential Equations and Mass Balance

Assume a very simple model with two reactive substances U_1 and U_2:

$$\frac{\mathrm{d}U_1}{\mathrm{d}t} = -aU_1U_2 + \eta_1$$

$$\frac{\mathrm{d}U_2}{\mathrm{d}t} = +aU_1U_2 + \eta_2$$

Does the total mass, $U_1 + U_2$, remain constant?

→ If we want to keep the mass balance closed, we can only make the *rates* stoahcastic.

eawag aquatic research

Phytoplankton Model with time-depending growth rate

$$\frac{dC_{\mathrm{HPO_4^{2-}}}}{dt} = \frac{Q_{\mathrm{in}}}{V}\left(C_{\mathrm{in,HPO_4^{2-}}} - C_{\mathrm{HPO_4^{2-}}}\right) - \alpha_{P,\mathrm{ALG}}\, k_{\mathrm{gro,ALG}} \frac{C_{\mathrm{HPO_4^{2-}}}}{K_{\mathrm{HPO_4^{2-},ALG}} + C_{\mathrm{HPO_4^{2-}}}} C_{\mathrm{ALG}}$$

$$\frac{dC_{\mathrm{ALG}}}{dt} = -\frac{Q_{\mathrm{in}}}{V} C_{\mathrm{ALG}} + k_{\mathrm{gro,ALG}} \frac{C_{\mathrm{HPO_4^{2-}}}}{K_{\mathrm{HPO_4^{2-},ALG}} + C_{\mathrm{HPO_4^{2-}}}} C_{\mathrm{ALG}} - k_{\mathrm{death,ALG}}\, C_{\mathrm{ALG}}$$

$$\frac{dk_{\mathrm{gro,ALG}}}{dt} = \frac{1}{\tau}\left(\mu - k_{\mathrm{gro,ALG}}\right) + \eta(t,\sigma)$$

Ornstein-Uhlenbeck process

Ornstein-Uhlenbeck process

A random process with autocorrleation and a tendency to revert to a mean value

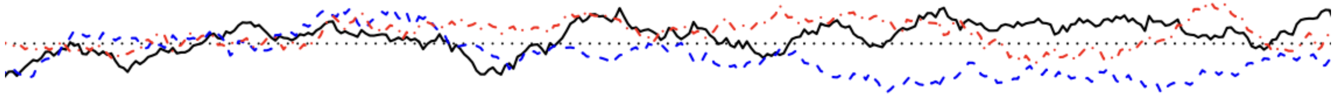
$$\frac{dx}{dt} = \frac{1}{\tau}(\mu - x) + \eta(t)$$

$$dx = \frac{1}{\tau}(\mu - x)dt + \sigma\sqrt{\frac{2}{\tau}}dW$$

μ mean of the process

τ autocorrelation length of the process

X_t is normal distributed at all time t with:



For more details see manuscript, section 9.2.3

Demo

Time-depending growth rate

Summary

Summary

- Some factors contributing to uncertainty can be best modeled as «internal» stochastic processes
  - Genetic stochasticity
  - Demographic stochasticity
  - Environmental stochasticity

- Often we want to make a parameter stochastic over time – this preserved mass balance

- The Ornstein-Uhlenbeck process is a good model for stochastic time-depending parameters

Why use a model with internal stochasticity?

The good

- Conceptually cleaner. We model the uncertainty where we think it originates for,
- The model becomes very expressive.
  This should also causion us: maybe a pattern in the data is just the result of noise!

The bad

- It is hard to get a feeling how the noise will affect the solutions.

The ugly

- Parameter Inference becomes *much* harder

Why use a ”normal” ODE model?

The good

- Easier to reason about
- Can be sufficient to describe the *average* behaviour – often this is of more interest
- Parameter inference is much easier

The bad

- We can become overconfident and accidently calibrate our model to noise in data

The ugly

- The additive noise is most likely inadequate description of the uncertainty. Hence our estimations of the uncertainties tend to be overly optimistic

Appendix

Stochastic Differential Equations

Commonly used notation for stochastic differential equations:

$$dx = f(x,t,\theta)dt + g(x,t,\theta)dW$$

where W denotes a Wiener process.

We interpret this as a differential equation with some noise component:

$$\frac{dx}{dt} = f(x,t,\theta) + \eta(t)$$

(this notation is mostly used in phyics)

Numerical solving a ODE: Euler method

$$\frac{dx}{dt} = f(x, t, \theta)$$

$$x_{t+\Delta_t} = x_t + \Delta_t\, f(x_t, t, \theta)$$

Do not use the Euler methods for real models, it is numerically unstable and inefficient

Solving a SDE: Euler–Maruyama method

$$dx = f(x,t,\theta)dt + g(x,t,\theta)dW$$

$$x_{t+\Delta_t} = x_t + \Delta_t\, f(x_t,t,\theta) + g(x_t,t,\theta)\sqrt{\Delta_t}\,\epsilon_t$$

where $\epsilon_t \sim \mathrm{N}(0,1)$

Also here, you may want to use better methods