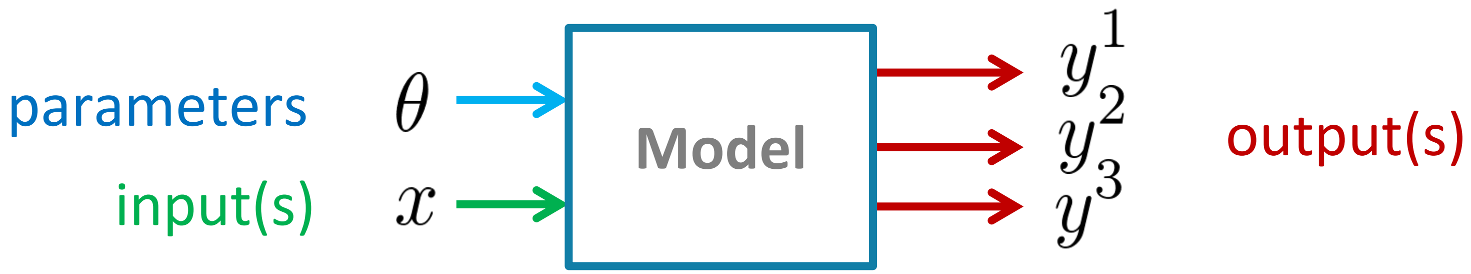
Exercise 6

Modelling Aquatic Ecosystems FS25

Today’s agenda

- Work on task 1-2, discuss the questions and results
- Break
- Work on task 3-4, discuss the questions and results
- Open questions

Task 1 - Uncertainty analysis



I know the *uncertainty of the inputs or parameters.*

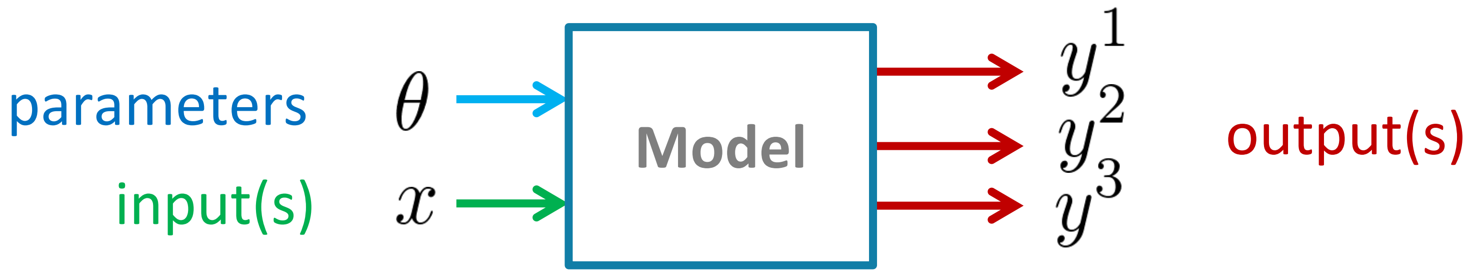
What is the resulting uncertainty of the *outputs*?

How can I compute it? E.g., **Monte Carlo Error propagation**

Task 1 - Questions

- How would you decide on the standard deviation for the different parameters?
- How large has `N` to be to get stable results?
- (optional) Compute the mean and sd of the outputs at `t = 365`
- (optional) Make a histogram of the model outputs at `t = 365`

Task 2-3-4 – Parameter estimation



What *parameters make the outputs* most similar to the observations?

Where do we have information about parameter values?

I. Laboratory experiments
II. Scientific literature
III. Calibration with observational data

What are the different model calibration techniques?

i. Manual calibration
ii. Minimizing a loss function
iii. Maximum Likelihood estimation – **a special loss function**
iv. Bayesian inference – **combine field data with other information**

Task 2 – Likelihood function

A likelihood function $p(\mathrm{data}|\theta)$ answers the following question:

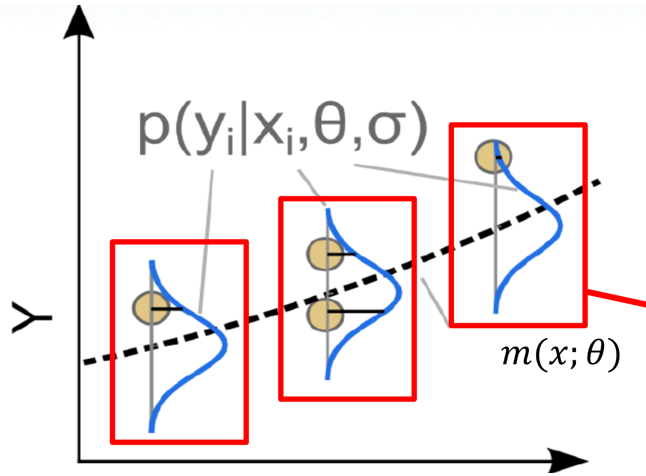
> "Given a *stochastic* model that generates random data. If the parameters are set to θ, what is probability (density) that the randomly generated data equal the observed?"

Task 2 – Likelihood function

deterministic model

$$\hat{Y}_i = m(x;\theta) + \epsilon_i$$

noise



Likelihood for all observations

$$p(\boldsymbol{y}|\boldsymbol{x},\theta,\sigma) = \prod_i p_i(y_i|x_i,\theta,\sigma)$$

Parameters, model, observational data

```r
# Formulation of a likelihood function for the lake plankton model
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

loglikeli <- function(par, system, obs, verbose=FALSE){
  # negative parameter values lead to a likelihood of zero or a log likelihood of
  # minus infinity:
  if ( any(par<=0) ) return(-Inf)

  # set the parameters equal to the current values given as the first function argument
  # (keep the other parameters):
  system@param[names(par)] <- par

  # set the start time to zero and the other output times to those with observations:
  system@t.out <- c(0,as.numeric(rownames(obs)))

  # calculate the deterministic results of our model:
  res <- calcres(system)

  # calculate the log likelihood using independent, normal distributions
  # with extracting the standard deviations from the parameter vector
  ll <-
    sum(c(dnorm(x=obs[,"C.HPO4"], mean=res[-1,"C.HPO4"], sd=par["sd.obs.HPO4"], log=TRUE),
          dnorm(x=obs[,"C.ALG"] , mean=res[-1,"C.ALG"] , sd=par["sd.obs.ALG"], log=TRUE)))

  # print parameters and likelihood if the verbose mode was selected:
  if ( verbose ) { print(par); cat("loglikeli =",ll,"\n") }

  # return the log likelihood value
  return(ll)
}
```

Time to work on the exercise

Task 3 - Questions

- What happens if you choose different initial values `par.ini`?

- How do you interpret `sd.obs.HPO4` and `sd.obs.ALG`?

Task 4 - Questions

- What can you get out of the parameter histograms?
- Experiment with different prior distributions.
- Try different initial values. What happens, if the initial values are very far off?
- What happens if the sample size is too small?
- Are some parameters correlated?

Open questions?