

Model Uncertainties and Parameter Estimation

6. May 2026 — Andreas Scheidegger

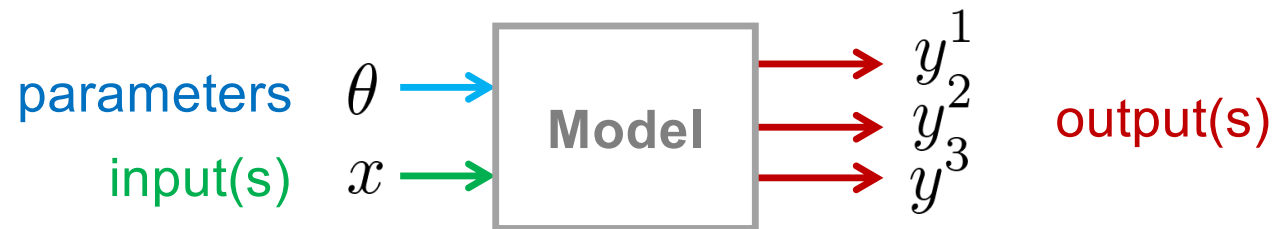
Goals of today

- i. You can identify different *sources* of model *uncertainties*.
- ii. You know how *uncertainties* can be *quantified* mathematically.
- iii. You can assess different ways to obtain good *parameter values*.

Model uncertainties

What is a model?

A computer program:



A mathematical function*:

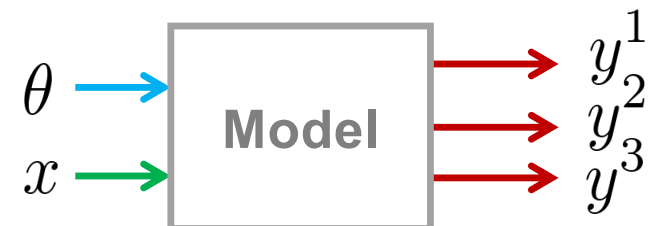
$$\mathbf{y} = m(\mathbf{x}; \boldsymbol{\theta})$$

*if the model output is stochastic, we often use capital letters: $\mathbf{Y} = M(\mathbf{x}; \boldsymbol{\theta})$

Models are wrong: sources of uncertainty

Almost all models suffer from these sources of uncertainty:

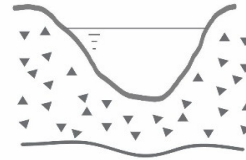
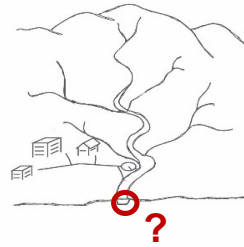
1. Input uncertainty
2. Parameter uncertainty
3. Model structure uncertainty
4. Output (measurement) uncertainty
5. (Scenario uncertainty)



Sources of Model Uncertainty

Input → Process → Response → Observation

Reality



$y_{observed}$

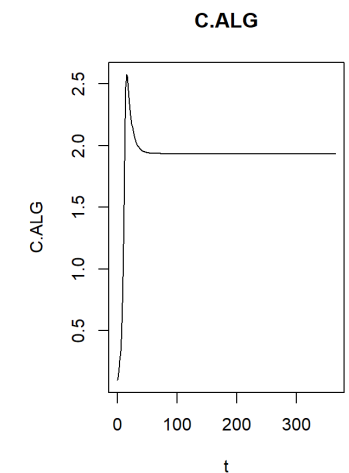
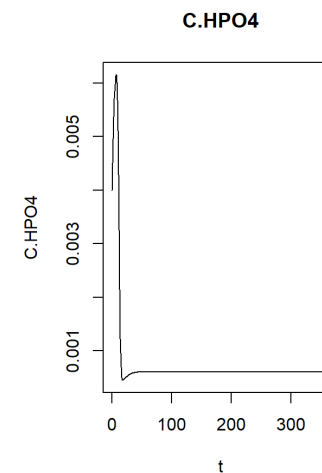
Sources of uncertainty: Phytoplankton Model

Think about examples for each source of uncertainty:

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

Which source do you judge the most relevant?

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

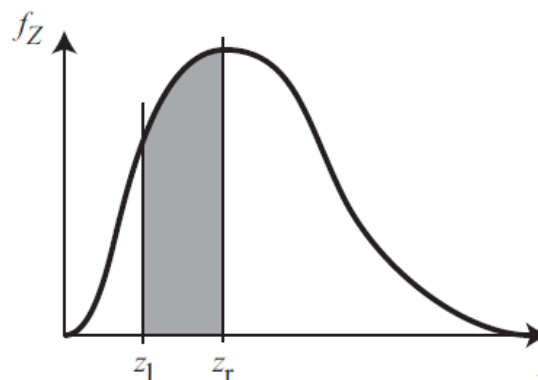


Descriptions of uncertainties

How can we describe uncertainties?

1. Guess tomorrow's air temperature in Zurich.
2. How (un)certain are you about your guess?

→ *Probability distributions* are a natural mathematical framework to describe uncertainties.



Probabilities, are they all the same?

- i. What is the probability to throw a five with a dice?
 - ii. What is the probability to win the Swiss lottery?
 - iii. What is the probability that it snows tomorrow on Uetliberg?
 - iv. What is the probability that it has snowed yesterday on Uetliberg?
- repeatable events*
- unique events*

Probability interpretations*

These are *interpretations*, the mathematical rules (probability calculus) are always the same!

Frequentist interpretation:

- Probability describes the frequency of a random event
- Only applicable if an event can be replicated, e.g. lab experiments, dice, coin flips

Bayesian or subjective interpretation:

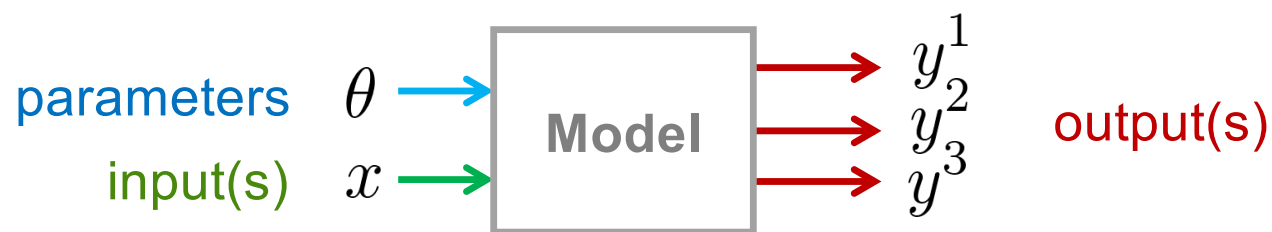
- Probability describes the “knowledge” or “belief” about an event
- Is *subjective* because everybody has different information
- When representing the current state of scientific knowledge, we use “intersubjective” probabilities; an assessment that is shared and agreed by all.

*This is a deep, philosophical topic, see e.g. <https://plato.stanford.edu/entries/probability-interpret/>

Uncertainty Analysis

- Error propagation
- Sensitivity Analysis

Error Propagation



I know the *uncertainty of the **inputs** or **parameters***.
What is the resulting uncertainty of the ***outputs***?

Analytical Error Propagation

Given the standard deviation σ_A and σ_B , what is σ_f ?

Function (model)

Variance of output

$$f = aA \pm bB$$

$$\sigma_f^2 = a^2 \sigma_A^2 + b^2 \sigma_B^2$$

$$f = AB$$

$$\sigma_f^2 \approx (AB)^2 \left[\left(\frac{\sigma_A}{A} \right)^2 + \left(\frac{\sigma_B}{B} \right)^2 \right]$$

For full table see:

https://en.wikipedia.org/wiki/Propagation_of_uncertainty

Based on linear *approximation!*

Monte Carlo Error Propagation

```

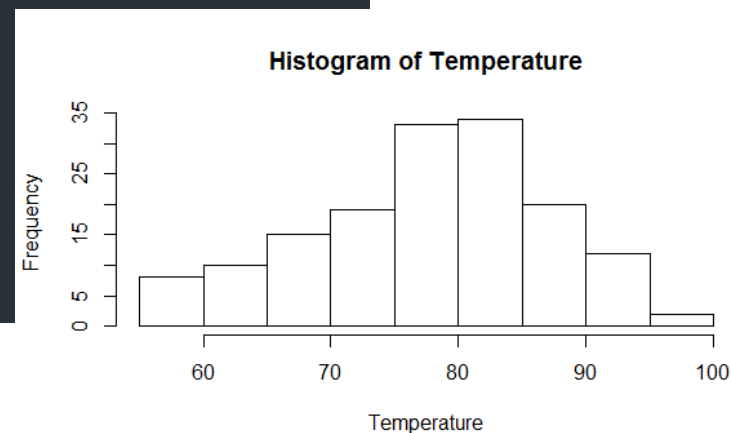
y = empty vector

for i in 1:N
  # sample inputs from distributions
  x ~ rnorm(10, 2)      # Normal( $\mu=10$ ,  $\sigma=2$ )
   $\theta$  ~ runif(0, 2)   # Uniform(0, 2)

  # run model
  y[i] = model(x,  $\theta$ )
end

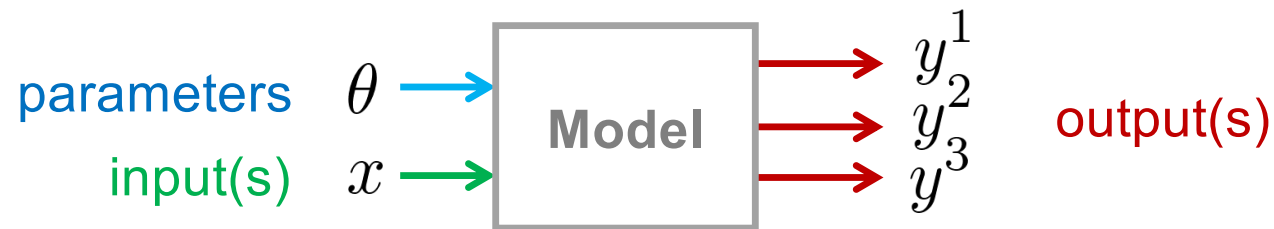
sd(y)
mean(y)
histogram(y)

```



Exact if N goes to infinity

Sensitivity Analysis



How important are the different
parameters and/or model **inputs** for the model **output**?



Factors

(SA does not distinguish between inputs and parameters)

Parameter estimation

Where do we get the numbers from?

How do we get the parameters values?

Parameters phytoplankton model

```
# definition of model parameters:

param  <- list(k.gro.ALG   = 0.5,      # 1/d
               k.death.ALG = 0.1,      # 1/d
               K.HPO4      = 0.002,    # gP/m3
               alpha.P.ALG = 0.003,    # gP/gDM
               A           = 5e+006,   # m2
               h.epi       = 5,        # m
               Q.in        = 5,        # m3/s
               C.HPO4.in   = 0.04,     # gP/m3
               C.HPO4.ini  = 0.004,    # gP/m3
               C.ALG.ini   = 0.1)      # gDM/m3
```

Information on parameter values

I. Laboratory Experiments

- Parameter needs a physical interpretation
- Lab conditions are not always transferable to real system



II. Scientific literature

- Measured under different conditions, varying methodologies, ...
- Only for “famous” parameters possible

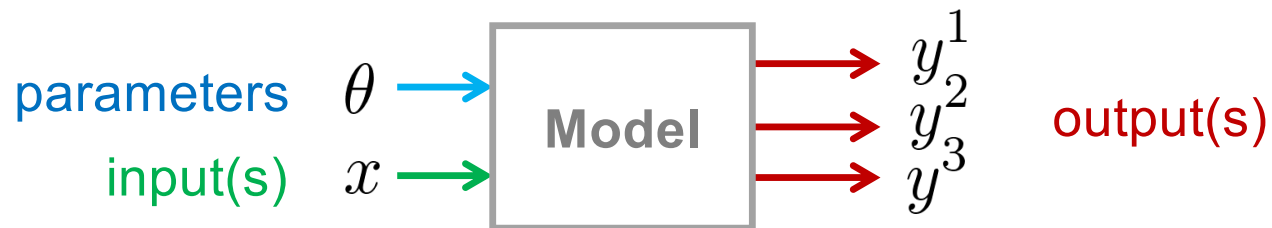


III. Calibration of model output to field observations

- Result depends on the model
- May not be transferable



Model Calibration



Which *parameters* result in *outputs* that best match the observations?

Demo

Manual Calibration

eawag
aquatic research **000**

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

Calibration with loss functions

$$\hat{\theta} = \overbrace{\arg \min_{\theta}}^{\text{take } \theta \text{ at minimum}} L \left(\overbrace{M(x, \theta)}^{\text{model output}}, \overbrace{Y_{\text{obs}}}^{\text{observations}} \right)$$

loss function

- Any loss function that appears sensible can be used
- Parameter uncertainty cannot be quantified!

Examples of loss functions

$$\Phi_{\text{SSE}}[\boldsymbol{\theta}] = \sum_{t=1}^{N_t} (q_t^{\text{obs}} - q_t^{\text{sim}}[\boldsymbol{\theta}])^2$$

$$\Phi_{\text{RMSE}}[\boldsymbol{\theta}] = \sqrt{\frac{1}{N_t} \Phi_{\text{SSE}}[\boldsymbol{\theta}]}$$

$$\Phi_{\text{NS}}[\boldsymbol{\theta}] = 1 - \left(\frac{\Phi_{\text{RMSE}}[\boldsymbol{\theta}]}{\sigma_{\text{obs}}} \right)^2$$

$$\Phi_{\text{SAE}}[\boldsymbol{\theta}] = \sum_{t=1}^{N_t} |q_t^{\text{obs}} - q_t^{\text{sim}}[\boldsymbol{\theta}]|$$

Maximum likelihood estimation

$$\hat{\theta} = \arg \max_{\theta} \underbrace{p_M (Y_{\text{obs}} \mid x, \theta)}_{\text{Likelihood function}}$$

Likelihood function

- The likelihood function is derived from *probabilistic assumptions* about the data generating process.
- Parameter uncertainty can be approximated (see e.g. https://www.sherrytowers.com/mle_introduction.pdf, section 5)

Likelihood function

A likelihood function $p(\text{data} \mid \theta)$ 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 data?”

Note, most of our models are deterministic (i.e. non-random).
The most common approach is then to assume an additional random error on the output.

Likelihood function: example of a deterministic model + noise

deterministic model

$$Y_i = \overbrace{m(x; \theta)}^{\text{deterministic model}} + \underbrace{\epsilon_i}_{\text{noise}}$$

$$\epsilon_i \sim N(0, \sigma^2)$$

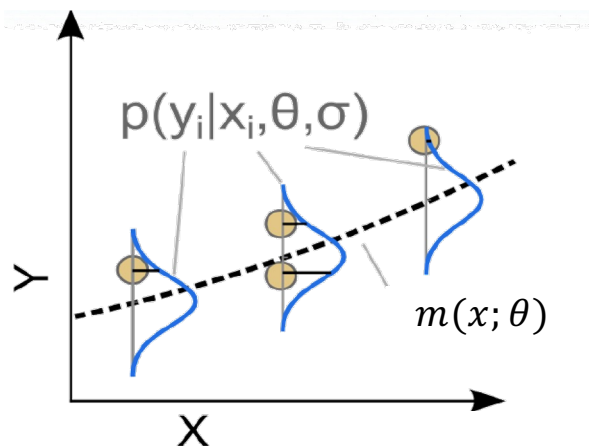
Likelihood for a single observations

$$p_i(y_i | x_i, \theta, \sigma) = \text{pdf of } N(m(x_i; \theta), \sigma^2)$$

i.e. the pdf of a normal distribution with mean = $m(x; \theta)$ and standard deviation σ .

Likelihood for all observations

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



Demo

Likelihood function

Parameter identifiability

$$V = h_{epi} * A$$

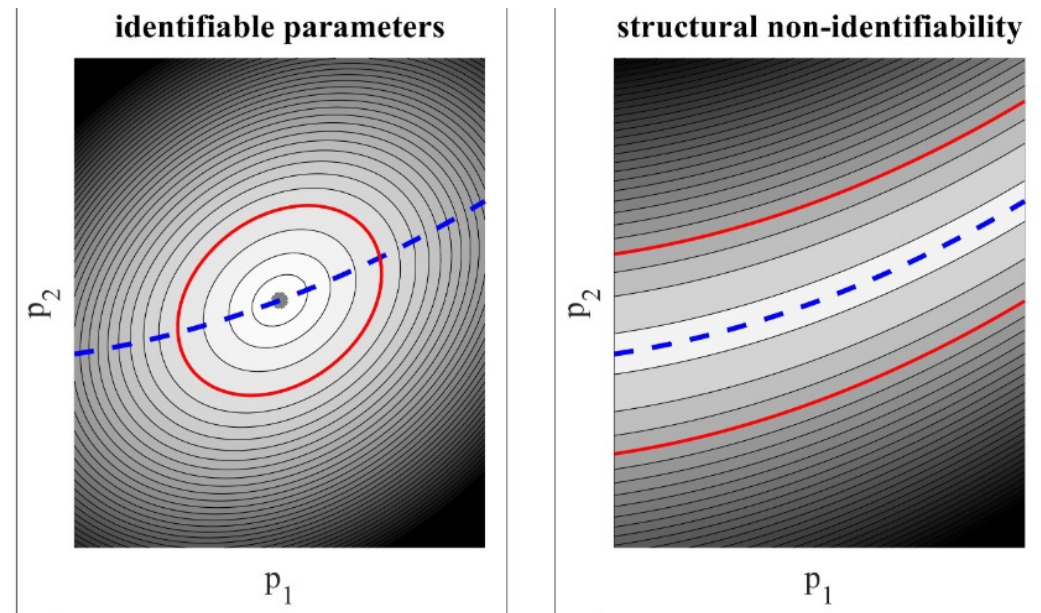
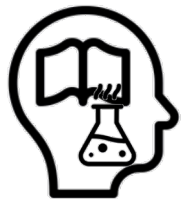


Image from: Wieland, F. G., Hauber, A. L., Rosenblatt, M., Tönsing, C., & Timmer, J. (2021). On structural and practical identifiability. *Current Opinion in Systems Biology*, 25, 60-69.

Bayesian inference: combine all information



- literature
- lab measurements
- experience

+



- model assumptions
- field data



- “updated” knowledge

$$p(\theta)$$

prior distribution

×

$$p(Y_{obs}|\theta)$$




likelihood function
(the same as for MLE!)

∝

$$p(\theta|Y_{obs})$$

posterior distribution

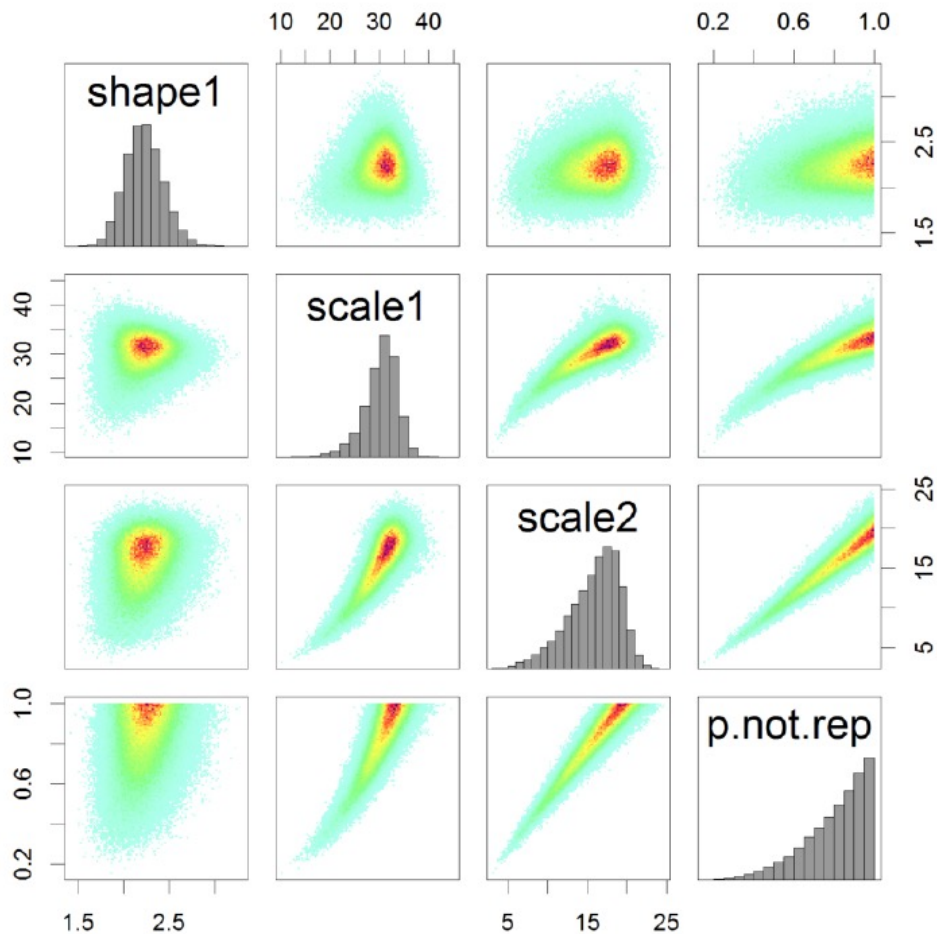
Bayesian inference

- A principled approach to combine:
 - i. Prior information  
 - ii. Data 
- Based on subjective probability interpretation
- More intuitive interpretation of uncertainty intervals
- Computationally demanding (you need to run the model >1000 times)
- Typically used with Monte Carlo Markov Chain (MCMC) sampling

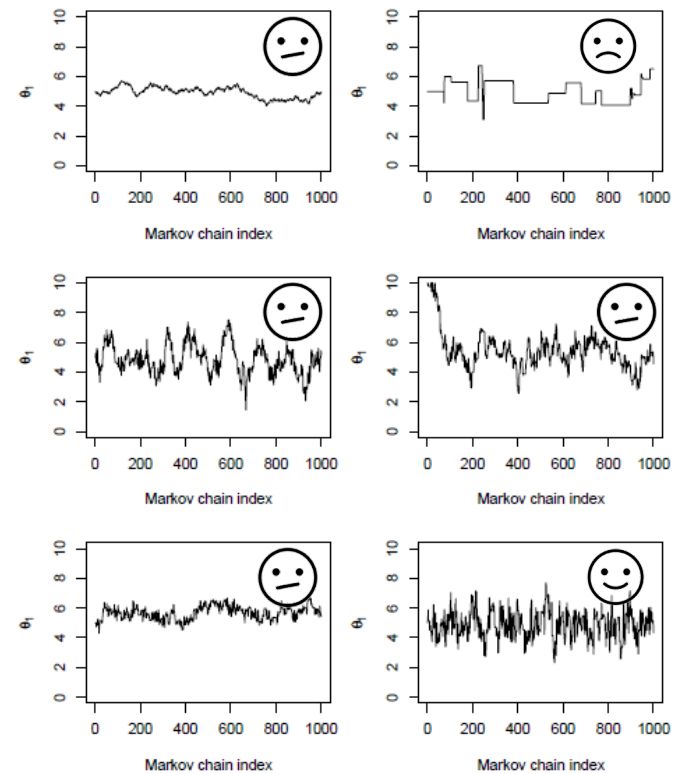
This is often difficult!

$$p(\theta|Y_{obs}) = \frac{p(Y_{obs}|\theta)p(\theta)}{p(Y_{obs})} = \frac{p(Y_{obs}|\theta)p(\theta)}{\int p(Y_{obs}|\theta')p(\theta')d\theta'}$$

Monte Carlo Markov Chain sampling



One- and two-dimensional marginal of the posterior distribution



*Diagnostic of MCMC chains.
Tuning is important!*

Multivariate Random Variables

Joint probability: $p(x, y)$

Marginal probability: $p(x) = \int p(x, y) dy$

Conditional probability: $p(x | y) = \frac{p(x, y)}{p(y)}$

Combining above definitions gives Bayes Theorem:

$$p(x | y) = \frac{p(x, y)}{p(y)} = \frac{p(x) p(y | x)}{p(y)} = \frac{p(x) p(y | x)}{\int p(x') p(y | x') dx'}$$

Demo

Bayesian Regression



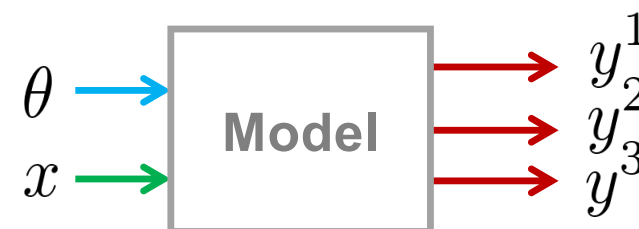
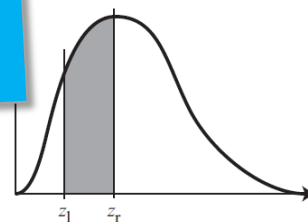
Summary

Model uncertainties

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

Stop worrying and use
Monte Carlo error propagation!

Uncertainties can be described by
probability distributions



Parameter values

Information sources:



i) literature



ii) laboratory experiments



iii) field measurements

Model calibration:



Prior



Bayesian inference

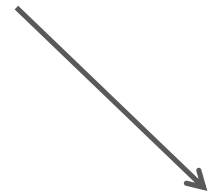


maximum likelihood estimation

minimize loss function



manual calibration



eawag
aquatic research **000**

Bayesian Inference

$$\underbrace{p(\theta \mid Y_{\text{obs}})}_{\text{Posterior distribution}} \propto \underbrace{p_M(Y_{\text{obs}} \mid x, \theta)}_{\text{Likelihood function}} \underbrace{p(\theta)}_{\text{Prior distribution}}$$

- The **likelihood** function is the same as for MLE
- **Prior** describes our “knowledge” about the parameters *before* calibration
- **Posterior** describes our “knowledge” about the parameters *after* calibration

- Bayesian inferences is based on subjective probability interpretation
- Parameter uncertainty is completely described