02427 lecture 09: Advanced Time Series Analysis

Grey-box models

Jan Kloppenborg Møller, Henrik Madsen

jkmo@dtu.dk

Advanced time series analysis – 2026

Program

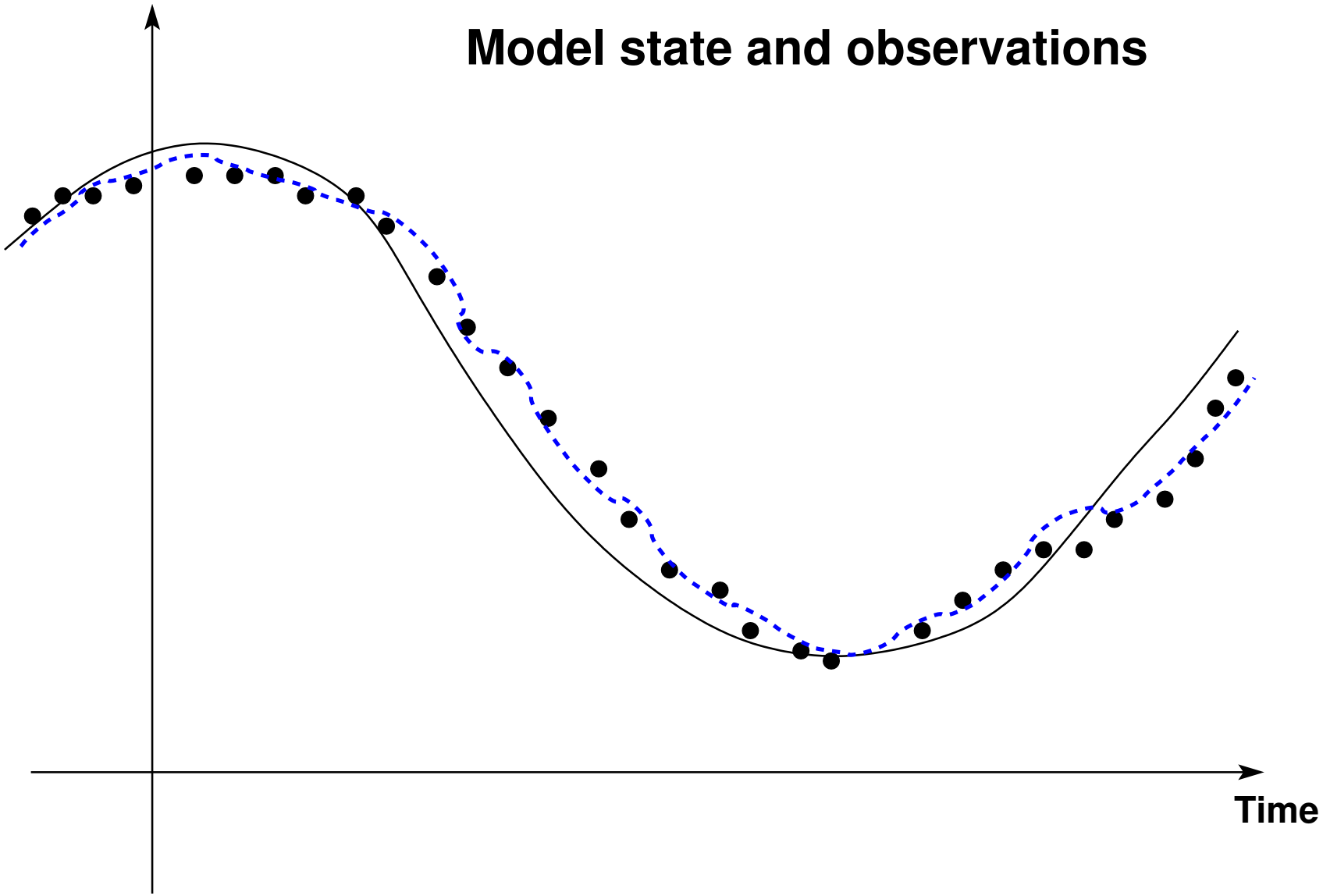
1. Introduction
2. SDE recap
3. Grey-box models
4. Estimation – The maximum likelihood principle
5. State Dependent diffusion
6. A case study

Program

1. Introduction
2. SDE recap
3. Grey-box models
4. Estimation – The maximum likelihood principle
5. State Dependent diffusion
6. A case study

Why consider SDEs ?

- SDE's can combine physical knowledge and data in a proper statistical description
- Physical systems are often influence by random disturbances
- Auto-correlated residuals are often seen. This calls for SDE's instead of ODE's
- Provides a systematic framework for pinpointing model deficiencies
- Basis for improved state estimation and control
- Provides a flexible description of the forecast uncertainty
- Well suited for generation of forecast scenarios (eg. for stochastic programming and decision making under uncertainty)


Model state and observations
Time

- White box

$$\frac{dx}{dt} = f(x_t, u_t, t) \tag{1}$$

- Black box, time-series models (e.g. ARMAX)
- Grey box (Stochastic Differential Equations)

$$\frac{dx}{dt} = f(x_t, u_t, t) + "noise" \tag{2}$$

We will

- Give a proper description of "noise".
- Discuss the properties of an SDE.
- Discuss how to solve an SDE?

Program

1. Introduction
2. SDE recap
   - Integration with respect to Brownian Motion
   - Solution of Stochactic Differential Equations
   - Itô's Lemma
3. Grey-box models
4. Estimation – The maximum likelihood principle
5. State Dependent diffusion
6. A case study

Construction of Brownian Motion

Consider the discrete time process

$$X_{t+\Delta t} - X_t = b(t, X_t)\Delta t + \sigma(t, X_t)\Delta B_t \tag{3}$$

with (the Ito-interpretation)

$$\Delta B_t = B_{t+\Delta t} - B_t \tag{4}$$

and assume that

- $B_0 = 0$ (a.s.)
- $\Delta B_t \sim N(0, \Delta t)$
- with $t_1 < t_2 < t_3 < t_4$ then $Cov[B_{t_2} - B_{t_1}, B_{t_4} - B_{t_3}] = 0$ (independent increments)

Integration with respect to Brownian Motion

Objective: To define the "best" interpretation of integration with respect to Brownian motion.

Define the, discrete time, sum

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta B_j.$$

$\Delta t_j \to 0$ gives the Itô ($B_t = B_{t_j}$ for $t \in [t_j, t_{j+1}]$) integral interpretation

$$X_t = X_0 + \int_0^t b(s, X_s)ds +'' \int_0^t \sigma(s, X_s)dB_s''.$$

We use the notation notation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$$

Example/exercise

Consider the SDE

$$dX_t = bdt + \sigma dB_t$$

with $X_0 = x_0$ (independent of B_t), what is

- The solution to the SDE?
- What is the distribution of X_t?
- What is the joint distribution of $[X_s, X_t]^T$ for $t > s$?

The Itô formula

Let X_t be a random process which satisfy the Itô SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t \tag{5}$$

and let $g(t, X_t) \in C^2([0, \infty) \times \mathbb{R})$. Then $Y_t = g(t, X_t)$ is also an Itô process and

$$dY_t = g_t(t, X_t)dt + g_x(t, X_t)dX_t + \frac{1}{2}g_{xx}(t, X_t)(dX_t)^2 \tag{6}$$

where $(dX_t)^2$ can be computed according to the rules

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0 \quad , dB_t \cdot dB_t = dt \tag{7}$$

Itô Calculus

Let X_t be a random process which satisfy the Itô SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t \tag{8}$$

and let $g(t, X_t) \in C^2([0, \infty) \times \mathbb{R})$. Then $Y_t = g(t, X_t)$ is also an Itô process described by

$$dY = \left(\frac{\partial g}{\partial t} + b\frac{\partial g}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 g}{\partial x^2} \right) dt + \sigma \frac{\partial g}{\partial x} dB \tag{9}$$

Note that the dependency on t and X_t has been left out.

Example/exercise

Consider the SDE

$$dX_t = \theta(\mu - X_t)dt + \sigma dB_t$$

Use Itô's to find the solution to the SDE, given in terms of the distribution of X_t given X_0, use $g(t,x) = e^{\theta t}x$.

Assume you observe X_t at times $[\Delta t, 2\Delta t, ..., N\Delta t,]$, write the observation process as an AR(1) process (including an intercept), and relate the AR parameters and the SDE parameters.

The Fokker-Planck equation

Given the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t \tag{10}$$

the pdf of X_t can be found as the solution to

$$p_t(x, t|x_0) = -\frac{\partial}{\partial x}(b(t, x) \cdot p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma(t, x)^2 \cdot p)$$

You may verify the solution from the previous example.

The Fokker-Planck equation

- Solutions to Fokker-Planck give the full distribution directly
- For simple SDE's with complicated boundaries Fokker-Planck offer the opportunity of simple recursive estimation
- For complicated SDEs Fokker-Planck is difficult to solve (also numerically)

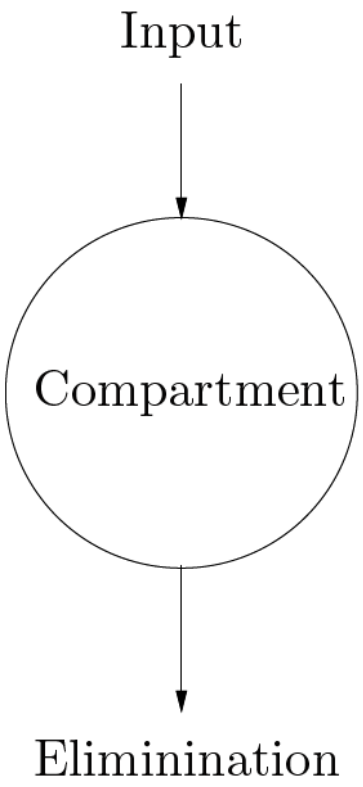
Program

1. Introduction
2. SDE recap
3. Grey-box models
   - Grey Box Modelling
   - Model Identification, Estimation, Selection and Validation
4. Estimation – The maximum likelihood principle
5. State Dependent diffusion
6. A case study

Introduction

- Various methods of advanced modelling are needed for a number of complex physical, chemical and biological systems.
- For a model to describe the future evolution of the system, it must
  1. capture the inherently non-linear behavior of the system.
  2. provide means to accommodate noise due to approximation and measurement errors.
- This calls for methods that are capable of **bridging the gap between physical and statistical modelling**.
- The answer is to consider **grey-box modelling**.

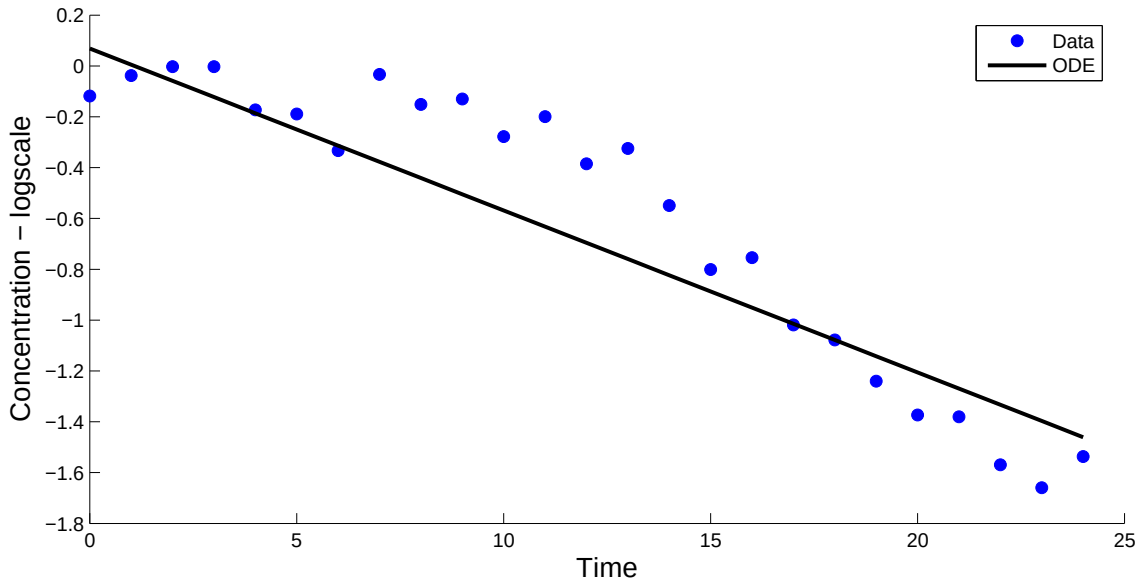
Problem Scenario



Ordinary differential equation

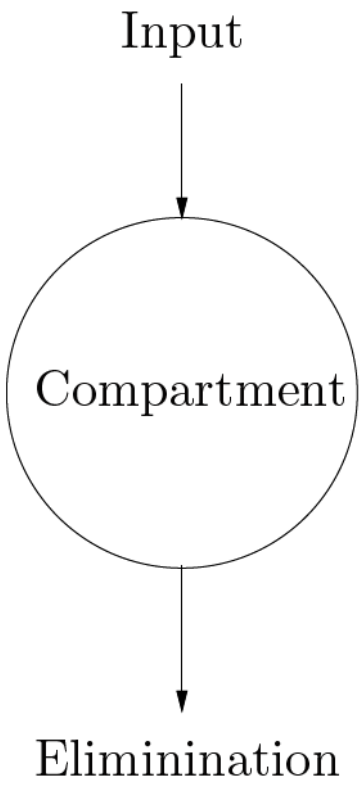
$$\mathrm{d}X = -KX\ \mathrm{d}t$$
$$Y = X + \epsilon$$

ODE



- Autocorrelated residuals!!

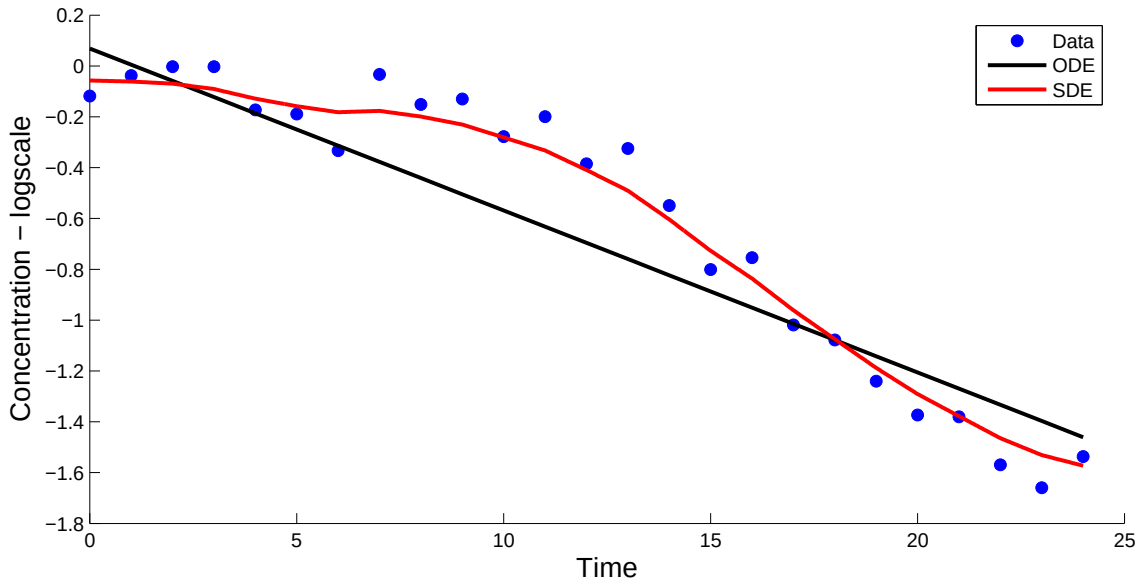
Problem Scenario



Stochastic differential equation

$$\mathrm{d}X = -KX\ \mathrm{d}t + \sigma\ \mathrm{d}w$$
$$Y = X + e$$

ODE vs SDE



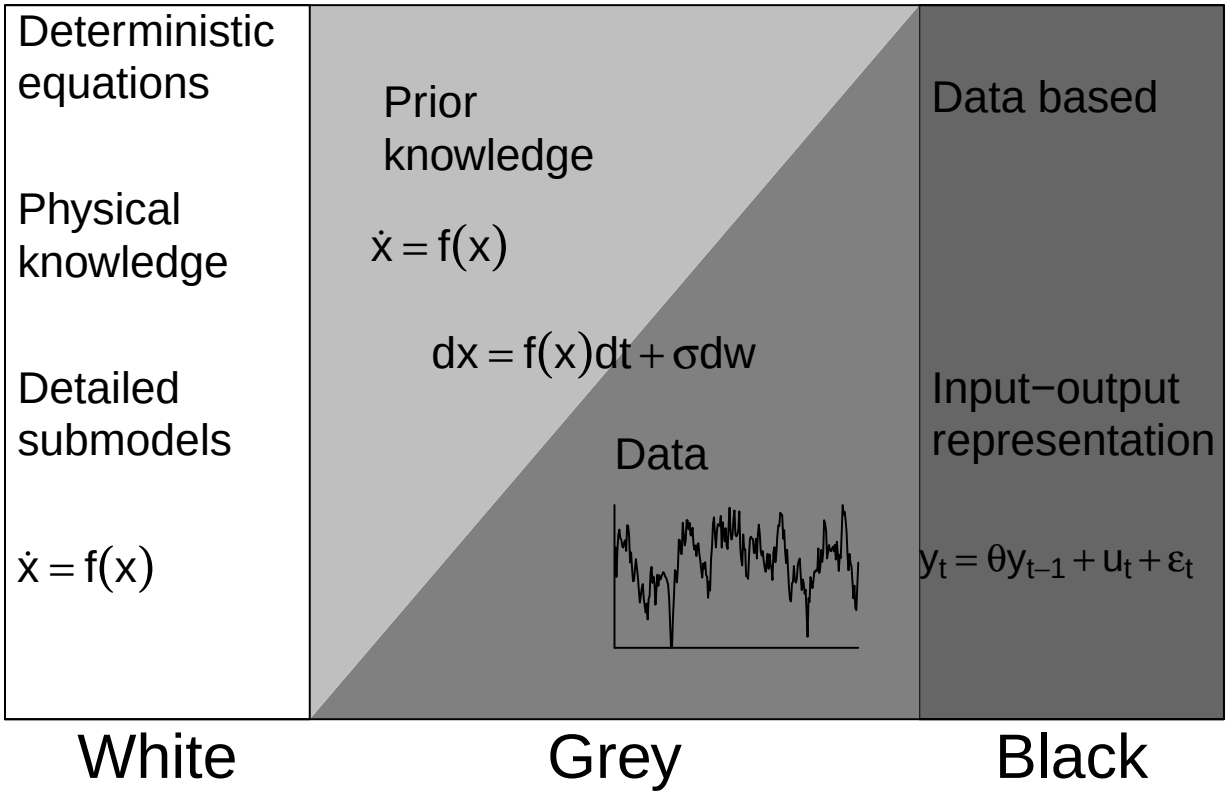
- Uncorrelated residuals
- System noise
- Measurement noise

Example/Exercise

Simulate the system on the previous slide, and discuss what is meant by the solution indicated in the slide.

The grey box modelling concept

- Combines prior physical knowledge with information in data.
- The model is not completely described by physical equations, but equations and the parameters are physically interpretable.



Why use grey box modelling?

- Prior physical knowledge can be used.
- Non-linear and non-stationary models are easily formulated.
- Missing data are easily accommodated.
- It is possible to estimate environmental variables that are not measured.
- Available physical knowledge and statistical modelling tools is combined to estimate the parameters of a rather complex dynamic system.
- The parameters contain information from the data that can be directly interpreted by the scientists.
- Fewer parameters $\rightarrow$ more power in the statistical tests.
- The physical expert and the statistician can collaborate in the model formulation.

Stochastic Differential Equations (SDE's)

- Ordinary Differential Equations (ODE's) provide deterministic description of a system:

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{U}_t, t)dt \quad t \geq 0.$$

  where $\boldsymbol{f}$ is a known function of the time t and the state $\boldsymbol{X}$ and input $\boldsymbol{U}$.
- To describe the deviation between the ODE and the true variation of the state an additive noise term is introduced.
- Physical arguments for including the noise part:
  1. Modelling approximations.
  2. Unrecognized inputs.
  3. Measurements of the input are noise corrupted.

The continuous-discrete time non-linear stochastic state space model

$$\mathrm{d}\boldsymbol{X}_t = f(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{\theta})\,\mathrm{d}\boldsymbol{B}_t, \quad \boldsymbol{X}_{t_0} = \boldsymbol{X}_0$$

Notation

$\boldsymbol{X}_t \in \mathbb{R}^n$	State vector
$\boldsymbol{U}_t \in \mathbb{R}^r$	Known input vector
f	Drift term
G	diffusion term
$\boldsymbol{B}_t$	Brownian motion of dimension, d, with incremental covariance $t\boldsymbol{I}$
$\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$	Unknown parameter vector

Technical University of Denmark DTU

The observation equation

The observations are in discrete time, functions of state, input, and parameters, and are subject to noise:

$$\boldsymbol{Y}_{t_i} = h(\boldsymbol{X}_{t_i}, \boldsymbol{U}_{t_i}, \boldsymbol{\theta}) + \boldsymbol{e}_{t_i}$$

Notation

$\boldsymbol{Y}_{t_i} \in \mathbb{R}^m$	Observation vector
h	Observation function
$\boldsymbol{e}_{t_i} \in \mathbb{R}^m$	Gaussian white noise with covariance $\boldsymbol{\Sigma}_{t_i}$

Observations are available at the time points t_i : $t_1 < \ldots < t_i < \ldots < t_N$

$\boldsymbol{X}_0, \boldsymbol{B}_t, \boldsymbol{e}_{t_i}$ assumed independent for all $(t, t_i), t \neq t_i$

Advantages of HMM with the dynamics described by SDE's

- Provides a decomposition of the total error into process error and measurement error.
- Facilitates use of statistical tools for model validation.
- Provides a systematic framework for pinpointing model deficiencies – will be demonstrated later on.
- Covariances of system error and measurement error are estimated.
- SDE based estimation gives more accurate and reliable parameter estimates than ODE based estimation.
- SDEs give more correct (more accurate and realistic) predictions and simulations.

Methods for Identification, Estimation and Model Validation

- **Model Identification:** See the next slides and above.
- **Parameter Estimation:**
  - Maximum Likelihood Methods
- **Model testing/selections:**
  - Test for significiant parameters (typically t-tests)
  - Test for model reductions (typically likelihood ratio tests)
  - Alternatively: Information criteria
- **Model Validation:**
  - Test whether the estimated model describes the data.
  - Autocorrelation functions – or Lag Dependent Functions.
  - Other classical methods ...

Identification of functional relations

Use **non-parametric methods (kernels, smoothing splines, etc.)** to estimate the **conditional mean** and the **conditional variance**.

- The conditional mean enters the drift term.
- The conditional variance enters the diffusion term.

Identification of Model Structure

- The **diffusion term** gives information for pinpointing model deficiencies.
- Assume that we based on 'large' values of relevant diffusion term(s) suspect $r \in \boldsymbol{\theta}$ to be a function of the states, input or time.
- Then consider the **extended state space model**:

$$\begin{aligned} \mathrm{d}\boldsymbol{X}_t &= f(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{\theta})\,\mathrm{d}\boldsymbol{B}_t, \quad \boldsymbol{X}_{t_0} = \boldsymbol{X}_0 \\ \mathrm{d}r_t &= \mathrm{d}B_t^* \\ \boldsymbol{Y}_{t_i} &= h(\boldsymbol{X}_{t_i}, \boldsymbol{U}_{t_i}, \boldsymbol{\theta}) + \boldsymbol{e}_{t_i} \end{aligned}$$

which corresponds to a **random walk** description of r_t.

Identification of Model Structure

- Do we observe a significiant reduction of the relevant diffusion term(s)?
- In that case calculate the smoothed state estimate $\widehat{r}_{t|N}$ (calculated e.g., by the software tool ctsmTMB.
- Plot $\widehat{r}_{t|N}$ versus the states, inputs and time.
- Identify a possible functional relationship.
- Build that functional relationship into the stochastic state space model.
- Estimate the model parameters and evaluate the improvement – using e.g. likelihood ratio tests.

Program

1. Introduction
2. SDE recap
3. Grey-box models
4. Estimation – The maximum likelihood principle
   - Model Validation
   - Software – CTSM
5. State Dependent diffusion
6. A case study

The likelihood function for the observations

Introduce the information available at time, t_k:

$$\mathcal{Y}_k = (Y_{t_k}, \ldots, Y_{t_0})^T$$

Then the likelihood function of the parameters using all observations until time t_N is

$$\begin{aligned} L_N(\boldsymbol{\theta}) &= p(\mathcal{Y}_N; \boldsymbol{\theta}) \\ &= \left(\prod_{k=1}^{N} p(Y_k | \mathcal{Y}_{k-1}; \boldsymbol{\theta}) \right) p(Y_0; \boldsymbol{\theta}) \end{aligned}$$

Note that we need to condition on all previous observations.
The parameters are found by maximizing the log-likelihood

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log L_N(\boldsymbol{\theta}).$$

Parameter unceartainty

Standard errors are based on the

Definition (Observed information)

The matrix

$$\boldsymbol{j}(\boldsymbol{\theta};\boldsymbol{y}) = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} l(\boldsymbol{\theta};\boldsymbol{y})$$

with the elements

$$\boldsymbol{j}(\boldsymbol{\theta};\boldsymbol{y})_{ij} = -\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\boldsymbol{\theta};\boldsymbol{y})$$

is called the *observed information* corresponding to the observation $\boldsymbol{y}$, evaluated in $\widehat{\boldsymbol{\theta}}$.

The observed information is thus equal to the Hessian (with opposite sign) of the log-likelihood function evaluated at $\boldsymbol{\theta}$. The Hessian matrix is simply (with opposite sign) the *curvature* of the log-likelihood function.

Distribution of the ML estimator

- The standard error of $\widehat{\theta}_i$ is given by

$$\widehat{\sigma}_{\hat{\theta}_i} = \sqrt{\text{Var}_{ii}[\widehat{\theta}]}$$

  where $\text{Var}_{ii}[\widehat{\theta}]$ is the i'th diagonal term of $\boldsymbol{j}^{-1}(\widehat{\boldsymbol{\theta}})$
- Hence we have that an estimate of the dispersion (variance-covariance matrix) of the estimator is

$$D[\widehat{\boldsymbol{\theta}}] = \boldsymbol{j}^{-1}(\widehat{\boldsymbol{\theta}})$$

- An estimate of the uncertainty of the individual parameter estimates is obtained by decomposing the dispersion matrix as follows:

$$D[\widehat{\boldsymbol{\theta}}] = \widehat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\theta}}} \boldsymbol{R} \widehat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\theta}}}$$

  into $\widehat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\theta}}}$, which is a diagonal matrix of the standard deviations of the individual parameter estimates, and $\boldsymbol{R}$, which is the corresponding correlation matrix. The value R_{ij} is thus the estimated correlation between $\widehat{\theta}_i$ and $\widehat{\theta}_j$.

Model validation

The purpose of the model validation is to test if the assumption of the model are fulfilled. If this is the case then the sequence of errors

$$\boldsymbol{e}_k = \boldsymbol{Y}_k - \boldsymbol{Y}_{k|k-1}$$

are independent, but not necessarily indentically distributed, their standardized versions are usually used

$$\tilde{\boldsymbol{e}}_k^i = \frac{\boldsymbol{Y}_k^i - \boldsymbol{Y}_{k|k-1}^i}{\sqrt{\boldsymbol{\Sigma}_{k|k-1}^{yy,ii}}}$$

where superscript i refer to the i'th element, and the condition mean and variance can be extracted from the (Extended) Kalman filter. The sequences $\tilde{e}_k^i$ are iid and in the setup we consider here they are also assumed normal.

Model validation

The sequence(s) $\tilde{e}_k^i$ can be examined using the usual model validation tools (qq-normal plots, ACF plots, preriodiogram, ect.).

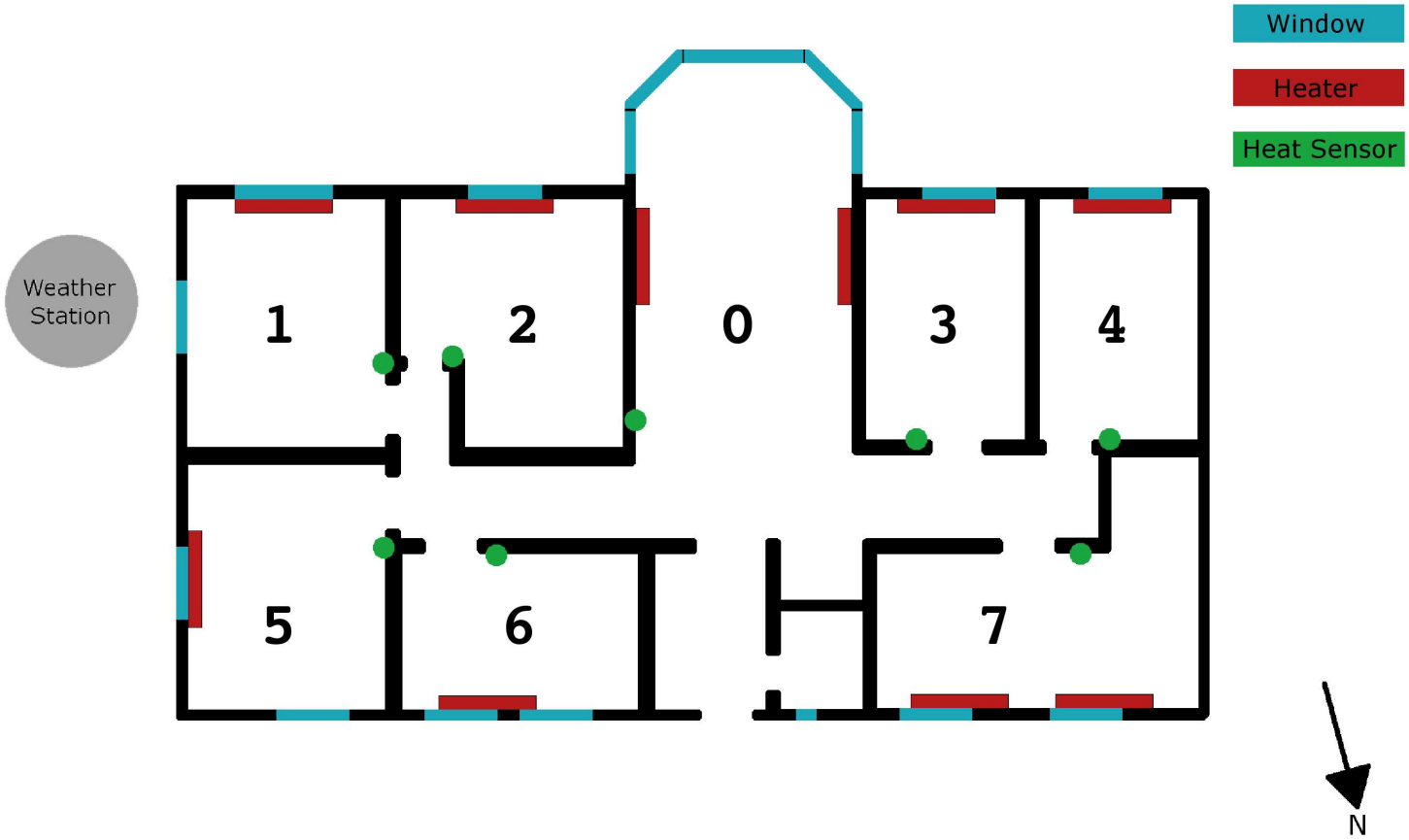
Model validation is usually an interative procedure where model validation is followed by suggetions for model extentions, until a reasonable model is obtained.

Required purpose: Always check you model against the required porpuse (e.g. 5 hours prediction, simulation).

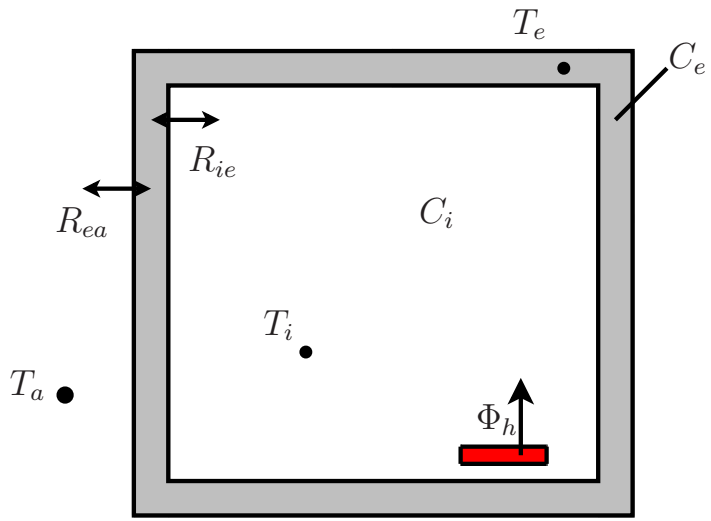
Example: One compartment model

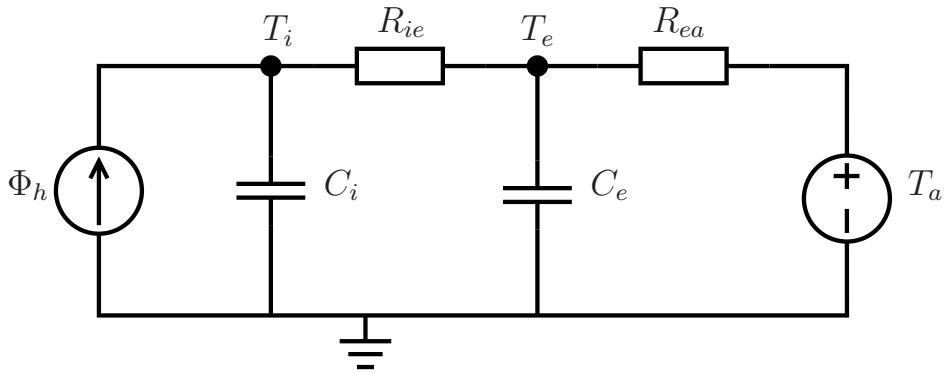
Estimate parameters in the one-compartment model.

Flexhouse Example (Peder Bacher):



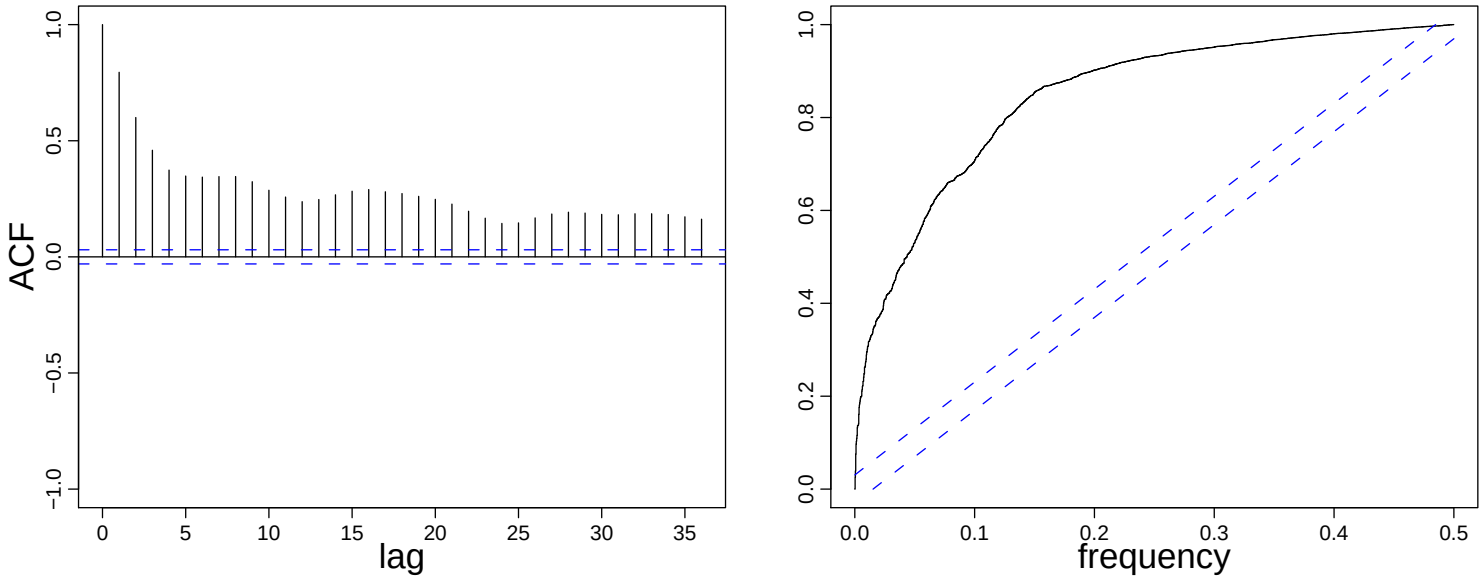
RC-diagram ofte used for illustrating linear models





Model Validation for Model A

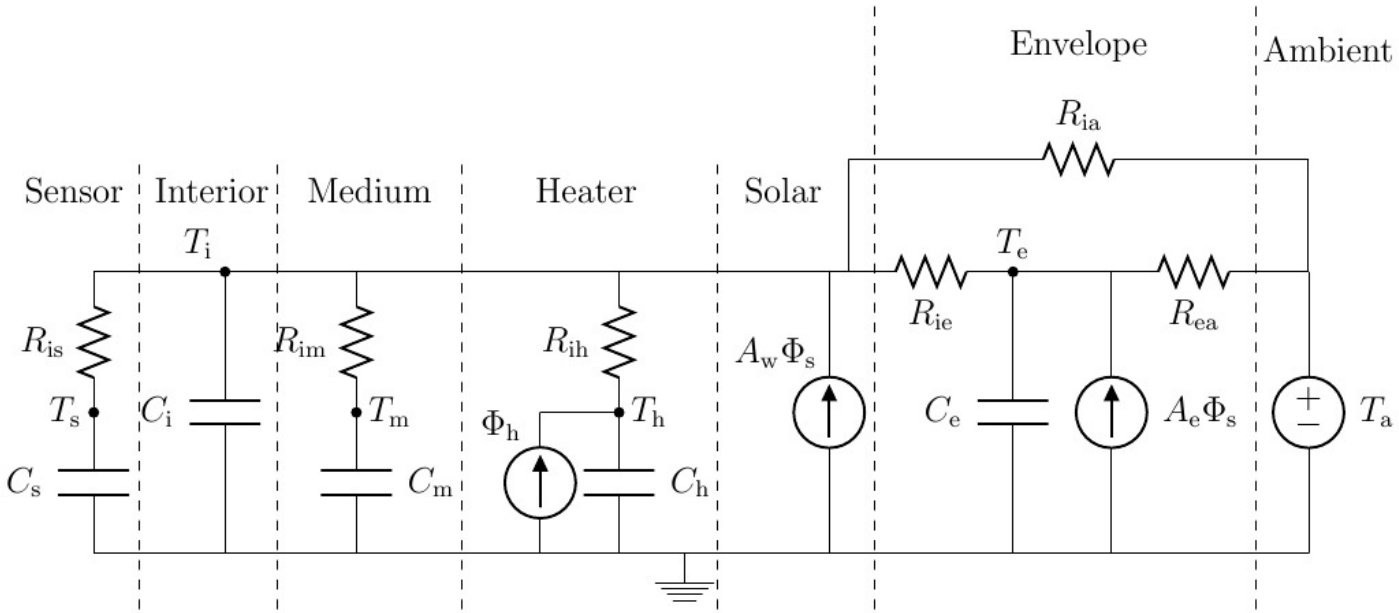
Autocorrelation function and Periodogram for the residuals.



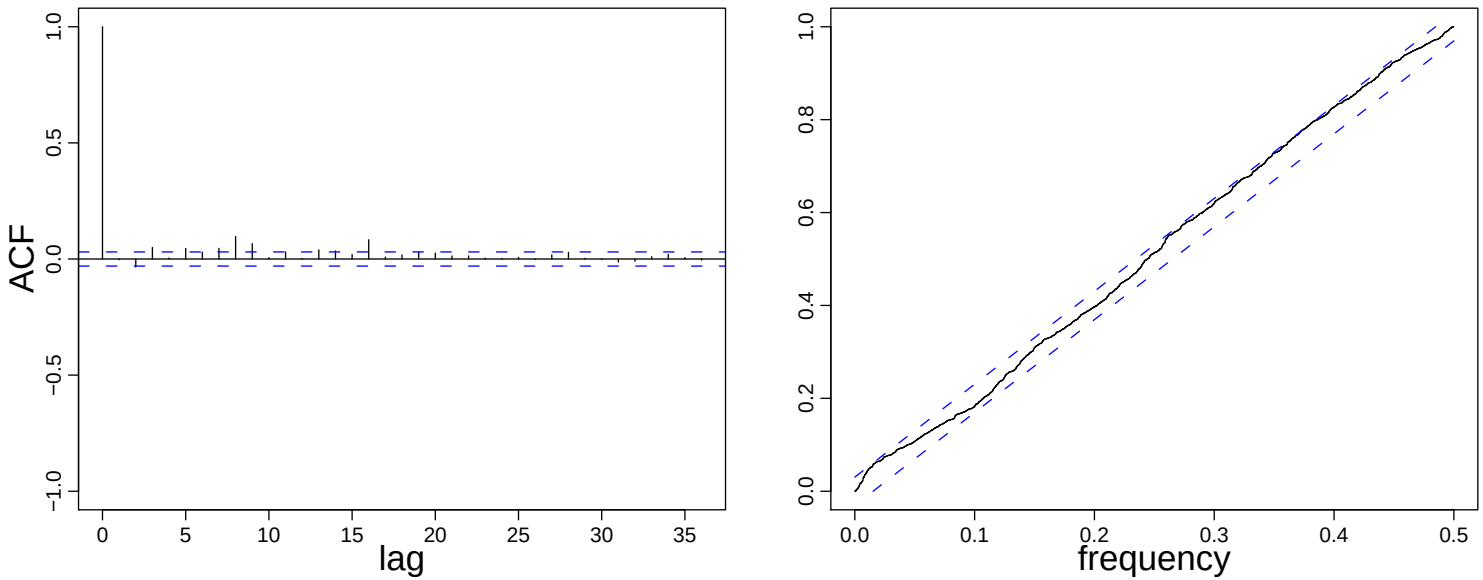
Model is seen not to be adequate.

Model E

After some steps: (Notice that eg. the electric heating system is included)



Model Validation for Model E.



It is concluded that the model is adequate.

Continuous Time Stochastic Modelling (CTSM)

- The parameter estimation here has been performed by using the software CTSM.
- The software has been developed at DTU Compute.
- The new version **ctsmTMB** - see the slides on ctsmTMB.
- The software returns the uncertainty of the parameter estimates as well.

The estimation procedure (CTSM)

CTSM is based on

- The Extended Kalman Filter
- Approximate likelihood estimation (Distribution of $Y_t|\mathcal{Y}_{t-1}$)

and provides eg.

- Likelihood testing for nested models
- Calculations of smoothen state $E[\boldsymbol{X}_t|\mathcal{Y}_T]$
- Calculations of k-step predictions $E[\boldsymbol{X}_t|\mathcal{Y}_{t-k}]$.
- Calculations of noise free simulations $E[\boldsymbol{X}_t|\mathcal{Y}_{t_0}]$

ctsmTMB can use different filters/estimation of the states.

Summary

By applying the grey box modelling approach

- physical/prior knowledge and information in data are combined, ie. we have bridged the gap between physical and statistical modelling.
- many statistical, mathematical and physical methods for model validation and structure modification become available.
- parameter estimates have physical significance - seldom the case for black box models.
- attain more accurate predictions (better long term predictions) and simulations of the system – and more realistic prediction intervals.

Program

1. Introduction
2. SDE recap
3. Grey-box models
4. Estimation – The maximum likelihood principle
5. State Dependent diffusion
   - The diffusion term
   - The Lamperti Transformation
   - Time independent diffusion
   - Multivariate Diffusion

The general setting

The Continous-Discrete Time Stochastic State Space model:

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{B}_t \tag{11}$$

$$\boldsymbol{Y}_k = \boldsymbol{g}(\boldsymbol{X}_{t_k}, t_k, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, \boldsymbol{e}_k), \tag{12}$$

Notice:

- We will use the Itô process interpretation
- In the diffusion term, the diffusion matrix $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times m}$ depends on the states $(\boldsymbol{X}_t)$.

Estimation

- For time series data the likelihood function is a product of one-step transition probabilities, i.e. the product of the probability density functions (pdf's) $p(Y_k|\mathcal{Y}_{k-1})$.
- The conditional densities can in principle be found using the Fokker-Planck equation for **prediction**, and using Bayes rule for **updating**.
- This is however only possible for very simple system equations.

Estimation - cont.

- Since the F-P approach is untracktable we shall use a different approach.
- We will approximate the transition probabilities by Gaussian pdf's.
- However, the filter equations for calculating the conditional mean and covariances take complicated forms if the diffusion term is dependent on the states.
- Our suggestion: Use a transformation of the state space such that the diffusion is independent of the state.
- The transformation (ψ) should form an equivalent relation between the input $\boldsymbol{u}_t$ and the output $\boldsymbol{Y}_k$ and the transformed system equation should depend on the same parameters ($\boldsymbol{\theta}$)

Conceptual diagram

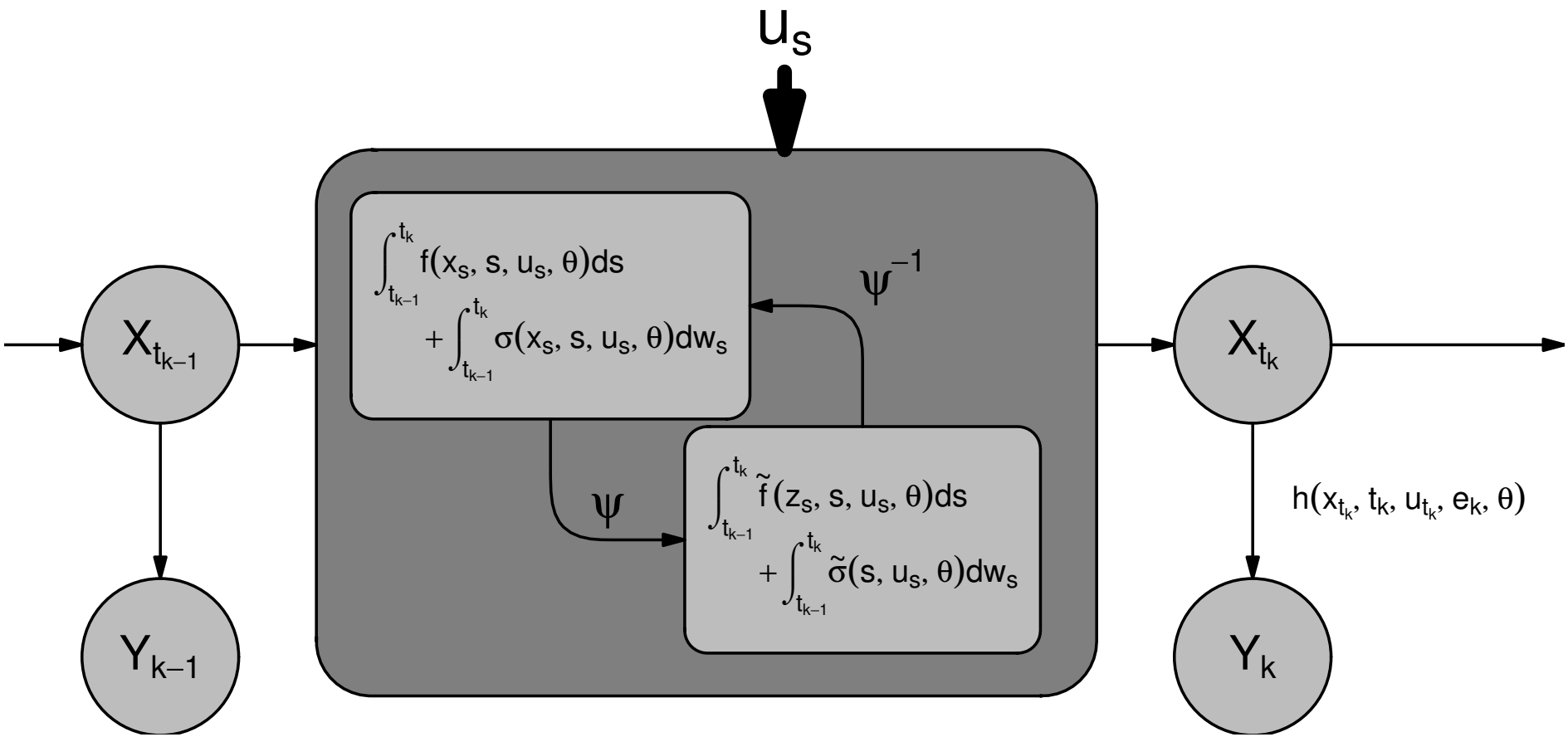


Figure: Lamperti Transformation - Conceptual Diagram

Problem setting

Given a class of SDEs defined by

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{B}_t, \tag{13}$$

where $\boldsymbol{\sigma} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B}_t \in \mathbb{R}^n$ is the standard Brownian motion.

Find transformations $\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t, t)$ or $\tilde{\boldsymbol{Z}}_t = \tilde{\boldsymbol{\psi}}(\boldsymbol{X}_t, t)$ such that

$$d\boldsymbol{Z}_t = \widetilde{\boldsymbol{f}}(\boldsymbol{Z}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \widetilde{\boldsymbol{\sigma}}(t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{B}_t \tag{14}$$

where $\widetilde{\boldsymbol{\sigma}}(\cdot)$ is independent of $\boldsymbol{Z}_t$.
Notice that the parameters of (14) are the same as before.

One-dimensional diffusion – Itô's lemma

Let X_t be an Itô process given by

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dB_t. \tag{15}$$

It is illustrative to express Itô's formula in terms of dw_t rather than dX_t. Notice that we will sometimes write f for $f(X_t, t)$, σ for $\sigma(X_t, t)$ and ψ for $\psi(X_t, t)$.
We obtain

$$dZ_t = \left(\psi_t + \psi_x \cdot f + \frac{1}{2}\psi_{xx} \cdot \sigma^2\right) dt + \psi_x \cdot \sigma dB_t.$$

Lamperti Transformation

Theorem

(Lamperti transform): *Let X_t be an Itô process as in (15), and define*

$$\psi(X_t, t) = Z_t = \int \frac{1}{\sigma(x,t)} dx \Bigg|_{x=X_t}, \tag{16}$$

we assume that ψ is one to one from the state space of X_t onto $\mathbb{R}$ for every $t \in [0, \infty)$. Then Z_t has unit diffusion and is governed by the SDE

$$dZ_t = \left(\psi_t(\psi^{-1}(Z_t, t), t) + \frac{f(\psi^{-1}(Z_t, t), t)}{\sigma(\psi^{-1}(Z_t, t), t)} - \frac{1}{2}\sigma_x(\psi^{-1}(Z_t, t), t) \right) dt + dB_t. \tag{17}$$

The proof is an application of Itô's lemma.

Example: Geometric Brownian motion

Let X_t be an Itô process (SDE) given by

$$dX_t = aX_t dt + \sigma X_t dB_t; \quad X_0 = 1,$$

where σ and a are real constants. Find the Lamperti transformed process and write the solution.

Example: Mean reverting with state dependent noise

Consider

$$dX_t = (b + aX_t)dt + \sigma X_t dB_t. \tag{18}$$

Using (16) we get the same transformation as in the previous example, and Z_t is governed by

$$\begin{aligned} dZ_t &= \left(\frac{b + aX_t}{\sigma X_t} - \frac{1}{2}\sigma \right) dt + dB_t \\ &= \left(\frac{b}{\sigma} e^{-\sigma Z_t} + \frac{a}{\sigma} - \frac{1}{2}\sigma \right) dt + dB_t, \end{aligned}$$

this SDE does not have an explicit solution, but parameter estimation is available through e.g. CTSM and numerical solutions (i.e. the distribution) can be found through simulations.

Multivariate Diffusion

We consider the process

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t)d\boldsymbol{B}_t, \tag{19}$$

with a transformation $\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t, t) = [\psi_1(\boldsymbol{X}_t, t), ..., \psi_n(\boldsymbol{X}_t, t)]$.
In the version of Itô's lemma given above $\boldsymbol{\psi} \in \mathbb{R}^n$. In the general version of Itô's lemma this not a requirement, but we have restricted the attention to equal dimensions of $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$. The derivations below do however easily generalise.

Multivariate Lamperti Transformation

Theorem (Multivariate Lamperti Transform)

Let X_t be an Itô diffusion given by

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t)\boldsymbol{R}(t)d\boldsymbol{B}_t, \tag{20}$$

where $\boldsymbol{R}(t) \in \mathbb{R}^{n\times n}$ is any matrix function of t, and $\boldsymbol{\sigma}(\boldsymbol{X}_t, t) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with diagonal elements $\sigma_{i,i}(\boldsymbol{X}_t, t)$ given by

$$\sigma_{i,i}(\boldsymbol{X}_t, t) = \sigma_i(X_{i,t}, t). \tag{21}$$

Then the transformation

$$Z_{i,t} = \psi_i(X_{i,t}, t) = \int \frac{1}{\sigma_i(x,t)}dx\Big|_{x=X_{i,t}}, \tag{22}$$

will result in an Itô process with the i'th element given by

$$dZ_{i,t} = \left(\frac{\partial}{\partial t}\psi_i(x,t)\Big|_{x=\psi^{-1}(Z_{i,t},t)} + \frac{f_i(\boldsymbol{\psi}^{-1}(\boldsymbol{Z}_t, t), t)}{\sigma_i(\psi_i^{-1}(Z_{i,t}, t), t)} - \frac{1}{2}\frac{\partial}{\partial x}\sigma_i(\psi_i^{-1}(Z_{i,t}, t), t)\right) dt + \sum^{n} r_{ij}(t)dB_{j,t}, \tag{23}$$

Program

1. Introduction
2. SDE recap
3. Grey-box models
4. Estimation – The maximum likelihood principle
5. State Dependent diffusion
6. A case study

Case study: Aim and methods

Aim: To develop a multivariate SDE-model suitable for simulation of phytoplankton

Methods:

- Parameter estimation using CTSM
- Identification using random walk hidden states (visualisations, gam and linear models)
- Considerations on stationarity
- Comparing data and simulations

Skive fjord: A shallow (maximum depth 7 meters) estuary in the northern part of Denmark

- Total volume $V = 113 \cdot 10^6 m^3$, Total surface area $36 \cdot 10^6 m^2$

Inputs:

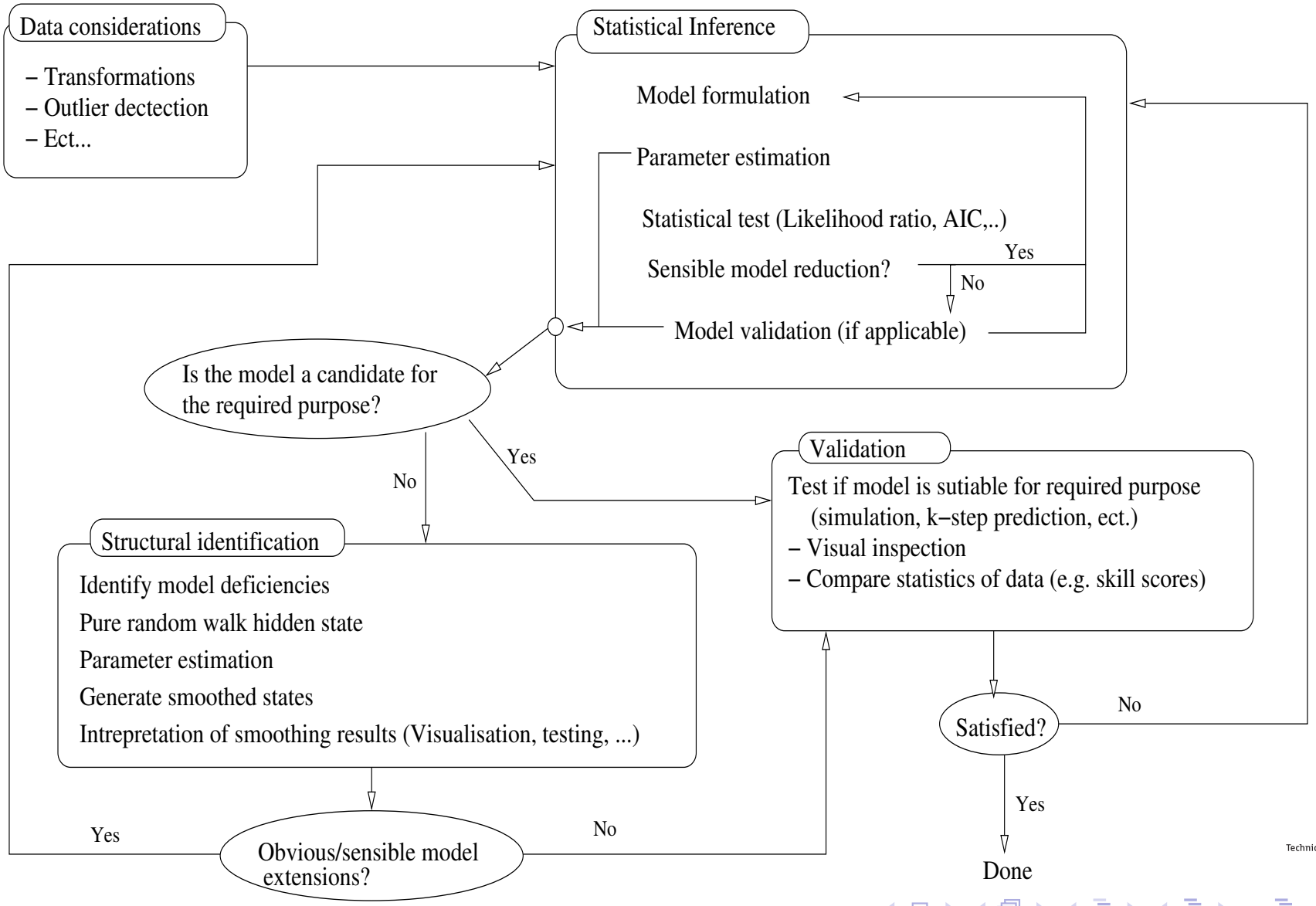
- Average global radiation per day (GR_t) (DMI).
- Freshwater runoff (Q_t) (NERI)
- Nitrogen-input by runoff ($N_{ex,t}$) (NERI)

Observations

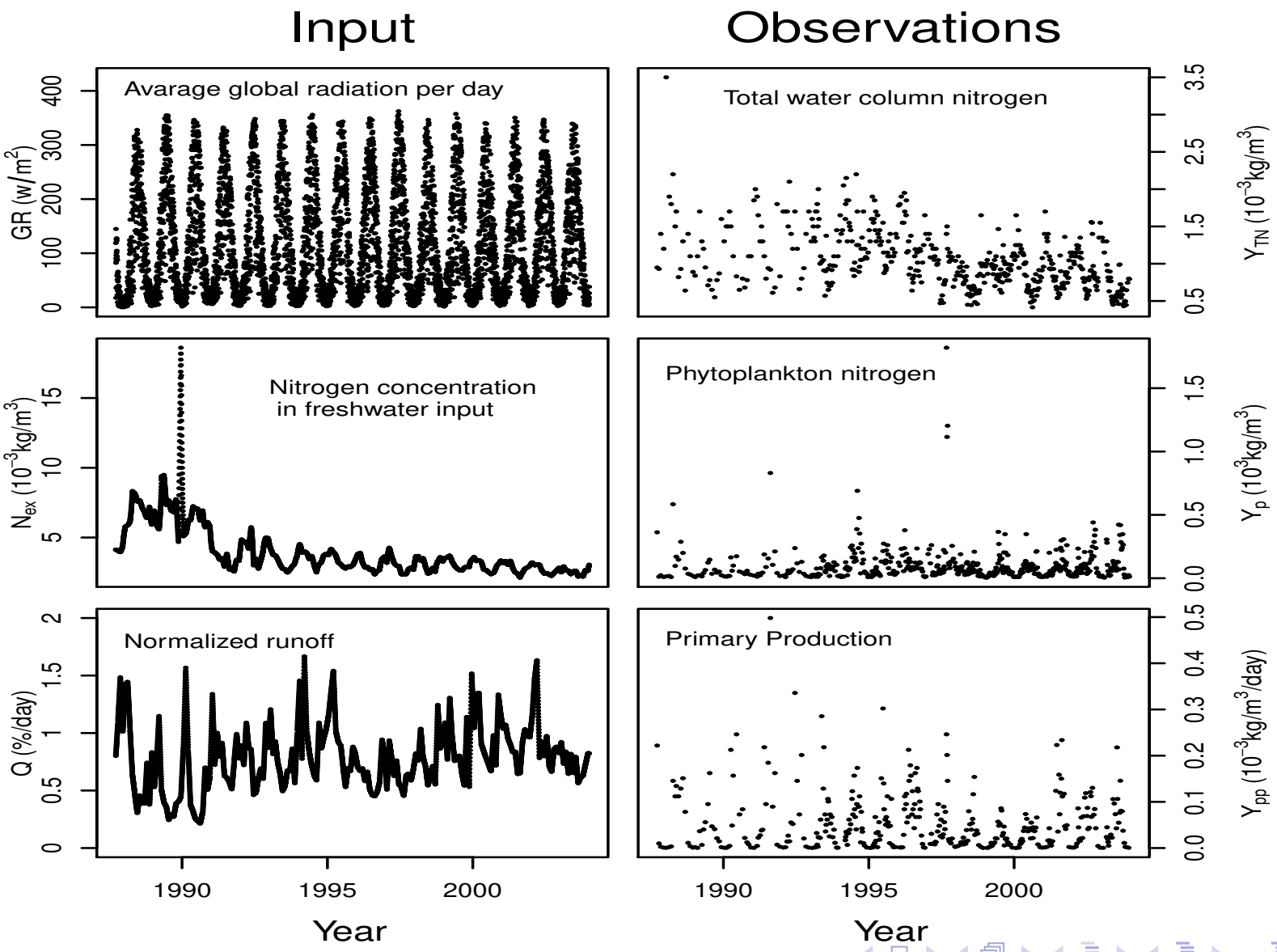
- Total Nitrogen in the water column given as mgN/l, converted to kgN assuming total mixing ($Y_{TN,t}$) (NERI).
- Phytoplankton nitrogen given as $mg \cdot chla/l$ converted to kgN assuming chlorophyll to carbon ratio of 1:50 (weight) and the Redfield ratio C:N (106:16 molar) and total mixing ($Y_{p,t}$) (NERI).
- Primary production (phytoplankton growth) given as $mgC/m^2/d$ converted to kg assuming Redfield ratio and total mixing ($Y_{pp,t}$) (NERI).

Technical University of Denmark DTU

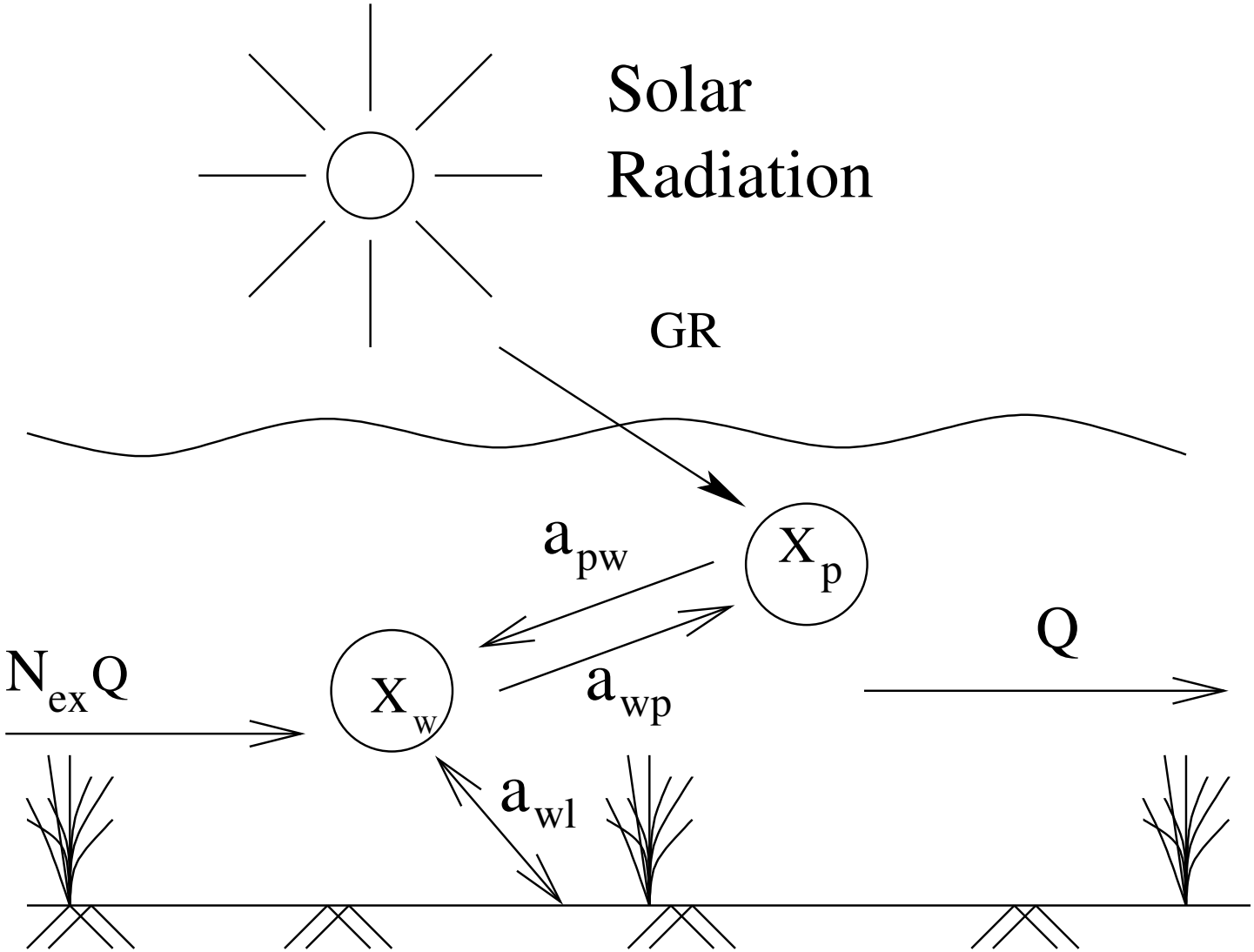
Methodology



Data


Input
Observations
Avarage global radiation per day
GR (w/m^2)
Total water column nitrogen
Y_{TN} (10^{-3}kg/m^3)
Nitrogen concentration in freshwater input
N_{ex} (10^{-3}kg/m^3)
Phytoplankton nitrogen
Y_p (10^3kg/m^3)
Normalized runoff
Q (%/day)
Primary Production
Y_{pp} (10^{-3}kg/m^3/day)
1990
1995
2000
Year

Conceptual model



Model 0: Initial formulation

Continuous time stochastic state space formulation

$$d\begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} = \begin{bmatrix} N_{ex,t}Q_t \\ 0 \end{bmatrix} dt + \begin{bmatrix} -Q_t - a_{wp} - a_{wl} & a_{pw} \\ a_{wp} & -a_{wp} - Q_t \end{bmatrix} \begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} dt + \begin{bmatrix} \sigma_w X_{w,t} & 0 \\ 0 & \sigma_p X_{p,t} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} d\boldsymbol{w}_t,$$

discrete time observation

$$\begin{bmatrix} \log(Y_{TN,k}) \\ \log(Y_{p,k}) \\ \log(Y_{pp,k}) \end{bmatrix} = \begin{bmatrix} \log(X_{w,t_k} + X_{p,t_k}) \\ \log(X_{p,t_k}) \\ \log(a_{wp}X_{w,t_k}) \end{bmatrix} + \begin{bmatrix} e_{TN,k} \\ e_{p,k} \\ e_{pp,k} \end{bmatrix},$$

Choose the transformation

$$Z_{i,t} = \frac{\log(X_{i,t})}{\sigma_i} \quad \Rightarrow \quad X_{i,t} = e^{\sigma_i Z_{i,t}}.$$

The transformed system is given by

$$d\begin{bmatrix} Z_{w,t} \\ Z_{p,t} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_w}\left(\frac{N_{ex,t}Q_t + a_{pw}X_{p,t}}{X_{w,t}} - \left(Q_t + a_{wl} + a_{wp}\right)\right) & -\frac{1}{2}\sigma_w(1+r_{12}^2) \\ \frac{1}{\sigma_p}\left(\frac{a_{wp}}{X_{p,t}} - (a_{pw} + Q_t)\right) & -\frac{1}{2}\sigma_p(1+r_{12}^2) \end{bmatrix} dt + \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} d\boldsymbol{w}_t.$$

Estimation results

Estimation results for Model 0, value of log-likelihood equal -1445.8

	θ_{min}	θ_0	θ_{max}	$\hat{\theta}$	std. dev	$P(x > \|t\|)$
a_{wp}	0.0	0.1	1.0	0.016	0.002	0.000
a_{wl}	-1.0	0.0	1.0	0.007	0.003	0.032
a_{pw}	0.0	0.1	1.0	0.053	0.034	0.122
σ_w	0.0	1.0	10.0	0.061	0.004	0.000
$\boldsymbol{\sigma_p}$	0.0	1.0	10.0	**0.625**	0.085	0.000
r_{12}	-1.0	0.0	1.0	-0.164	0.046	0.000
s_w^2	0.0	1.0	10.0	0.013	0.002	0.000
$\boldsymbol{s_p^2}$	0.0	1.0	10.0	**0.304**	0.148	0.041
$\boldsymbol{s_{pr}^2}$	0.0	1.0	10.0	**3.546**	0.285	0.000

Technical University of Denmark DTU

Structural identification

Consider primary production as a function of the state space and input

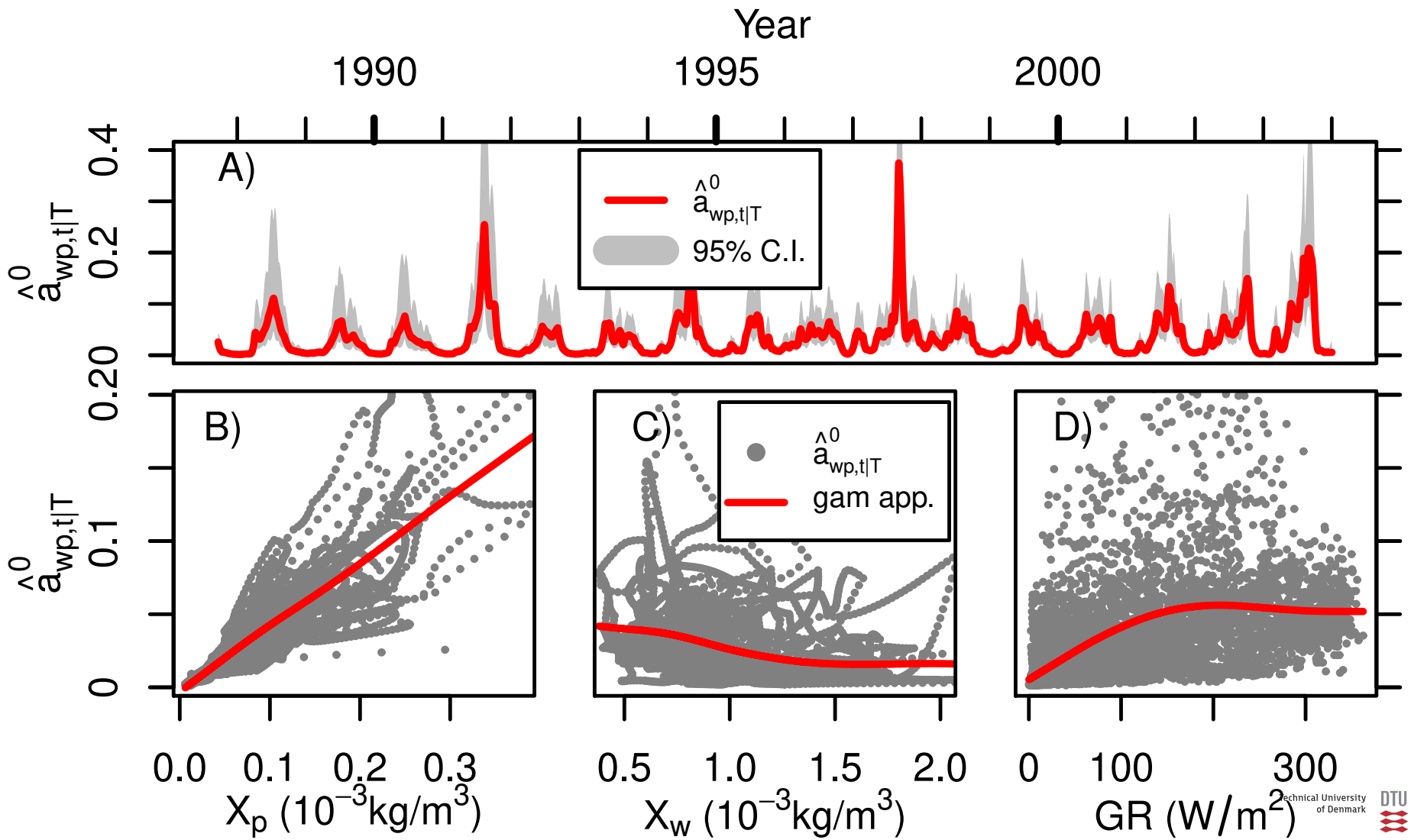
$$a_{wp,t} = a^0_{wp} X_{w,t} f_i(\boldsymbol{X}_t, \boldsymbol{u}_t),$$

and formulate as a random walk

$$\log(a^0_{wp,t}) = \sigma_{a_{wp}} d\boldsymbol{w}_{a_{wp},t}$$

Reestimate

Structural identification



Structural identification

Formulate possible model improvements

$$H_1: \quad \hat{a}^0_{wp,t|T} = \alpha_{gr,0} + \alpha_{gr,1} Gr_t + \epsilon_{Gr,t} \qquad AIC = -22928$$

$$H_2: \quad \hat{a}^0_{wp,t|T} = \alpha_{p,0} + \alpha_{p,1} \hat{X}_{p,t|T} + \epsilon_{p,t} \qquad AIC = -28410$$

$$H_3: \quad \hat{a}^0_{wp,t|T} = \alpha_{w,0} + \alpha_{w,1} \hat{X}_{w,t|T} + \epsilon_{w,t} \qquad AIC = -23998$$

Include linear dependence on phytoplankton and choose

$$a_{wp,t} = a^0_{wp} X_{p,t} X_{w,t}.$$

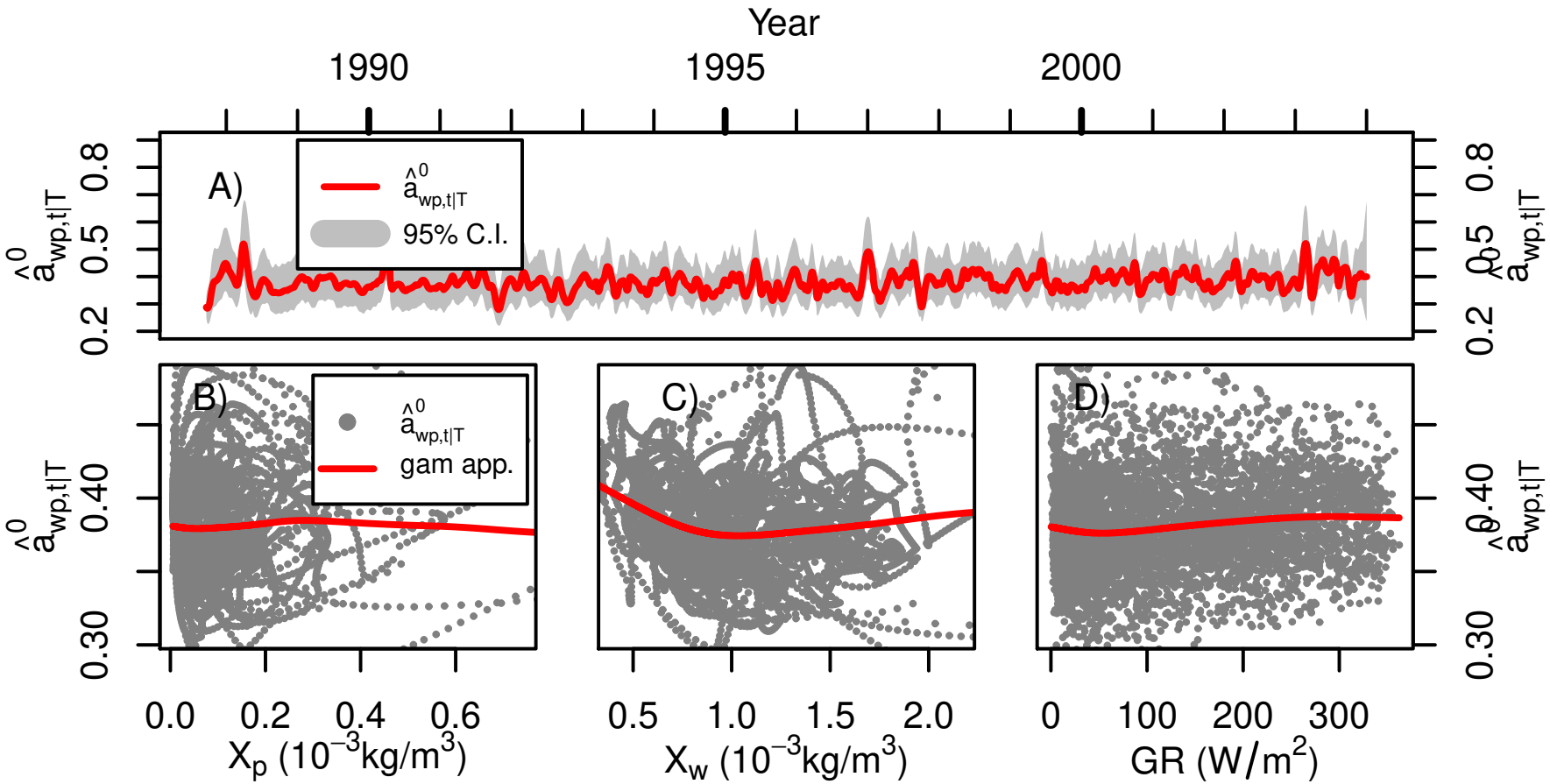
Model 1: Estimation results

Estimation results for Model 1, value of log-likelihood equal -1374.1

Par	θ_{min}	θ_0	θ_{max}	$\hat{\theta}$		std. dev	$P(x > \|t\|)$
a^0_{wp}	0.0	0.35	1.0	0.159		0.006	0.000
a_{wl}	-1.0	0.02	1.0	0.018		0.001	0.000
a_{pw}	0.0	0.32	1.0	0.124		0.007	0.000
σ_w	0.0	0.04	10.0	0.055		0.003	0.000
σ_p	0.0	0.06	10.0	0.252	0.625	0.015	0.000
r_{12}	-1.0	0.95	1.0	-0.135		0.034	0.000
s^2_w	0.0	0.01	10.0	0.029		0.004	0.000
s^2_p	0.0	0.19	10.0	0.141	0.304	0.047	0.003
s^2_{pr}	0.0	1.11	10.0	2.193	3.546	0.179	0.000

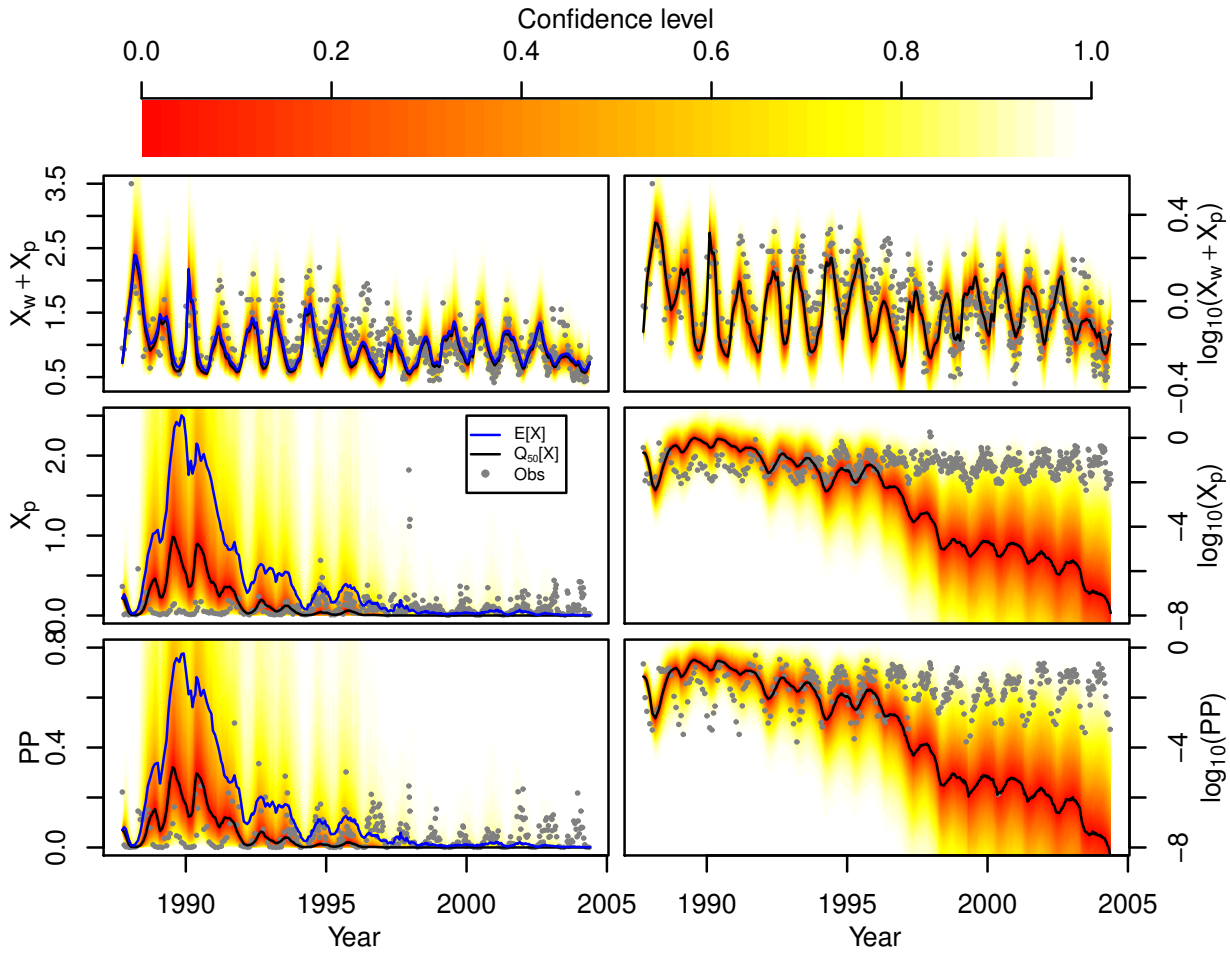
Technical University of Denmark DTU

After some iterations...



$$a_{wp,t} = a^0_{wp} X_{p,t} \frac{X_{w,t}}{k_w + X_{w,t}} \frac{Gr_t}{k_{gr} + Gr_t}.$$

Validation


Confidence level
0.0
0.2
0.4
0.6
0.8
1.0
$X_w + X_p$
X_p
PP
E[X]
Q_{50}[X]
Obs
$\log_{10}(X_w + X_p)$
$\log_{10}(X_p)$
$\log_{10}(PP)$
1990
1995
2000
2005
Year

Model formulation

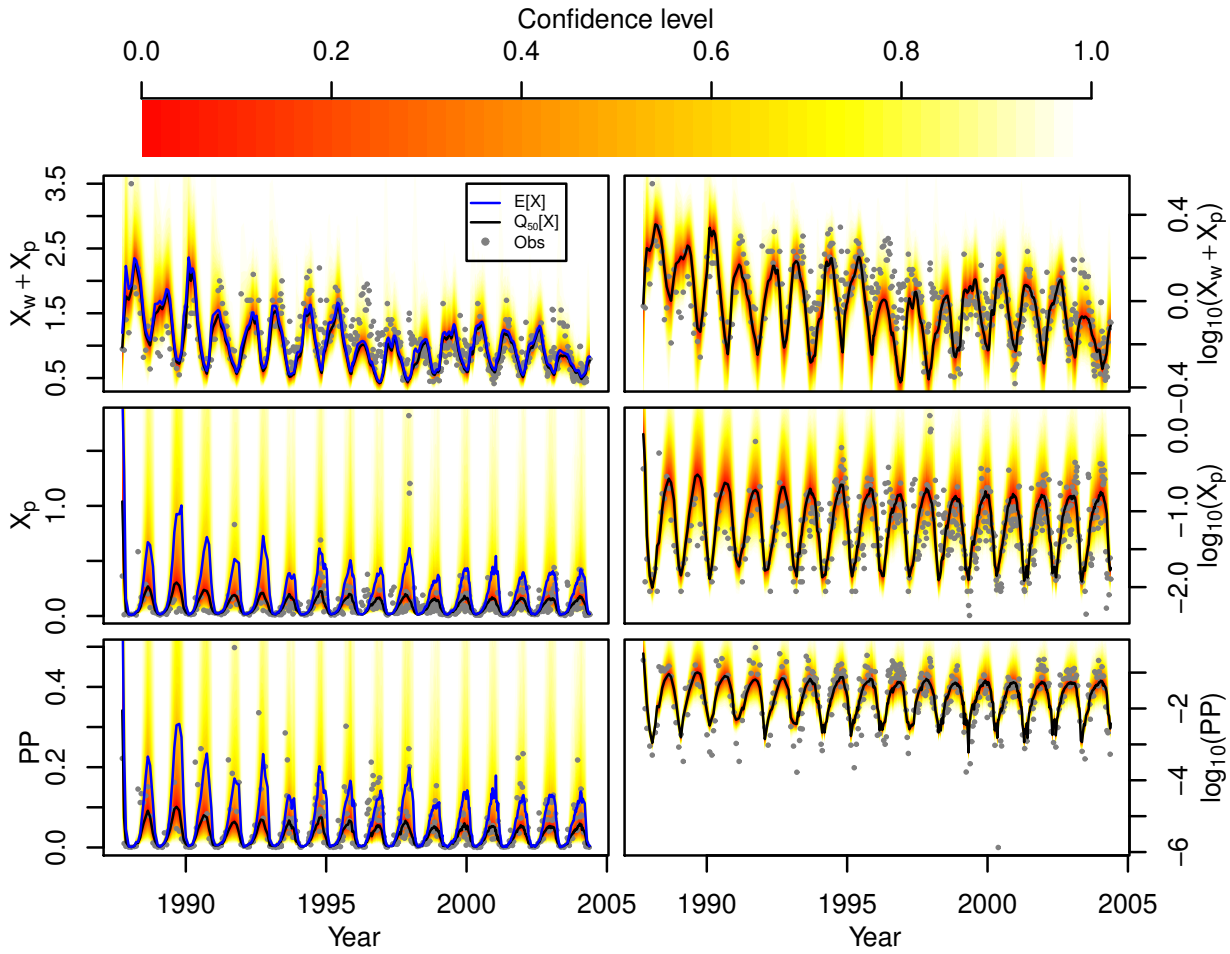
Inspection or phytoplankton equation

$$dX_{p,t} = \left(a_{wp} \frac{X_{w,t} Gr_t}{(k_w + X_{w,t})(k_{gr} + Gr_t)} - Q_t - a_{pw} \right) X_{p,t} dt + \sigma_p X_{p,t} (r_{12} dw_{1,t} + dw_{2,t})$$

Including a constant inoculum

$$dX_{p,t} = a_{p0} dt + \left(a_{wp} \frac{X_{w,t} Gr_t}{(k_w + X_{w,t})(k_{gr} + Gr_t)} - Q_t - a_{pw} \right) X_{p,t} dt + \sigma_p X_{p,t} (r_{12} dw_{1,t} + dw_{2,t}).$$

Model 4: Validation



Model formulation

Distribution controlled by the diffusion term

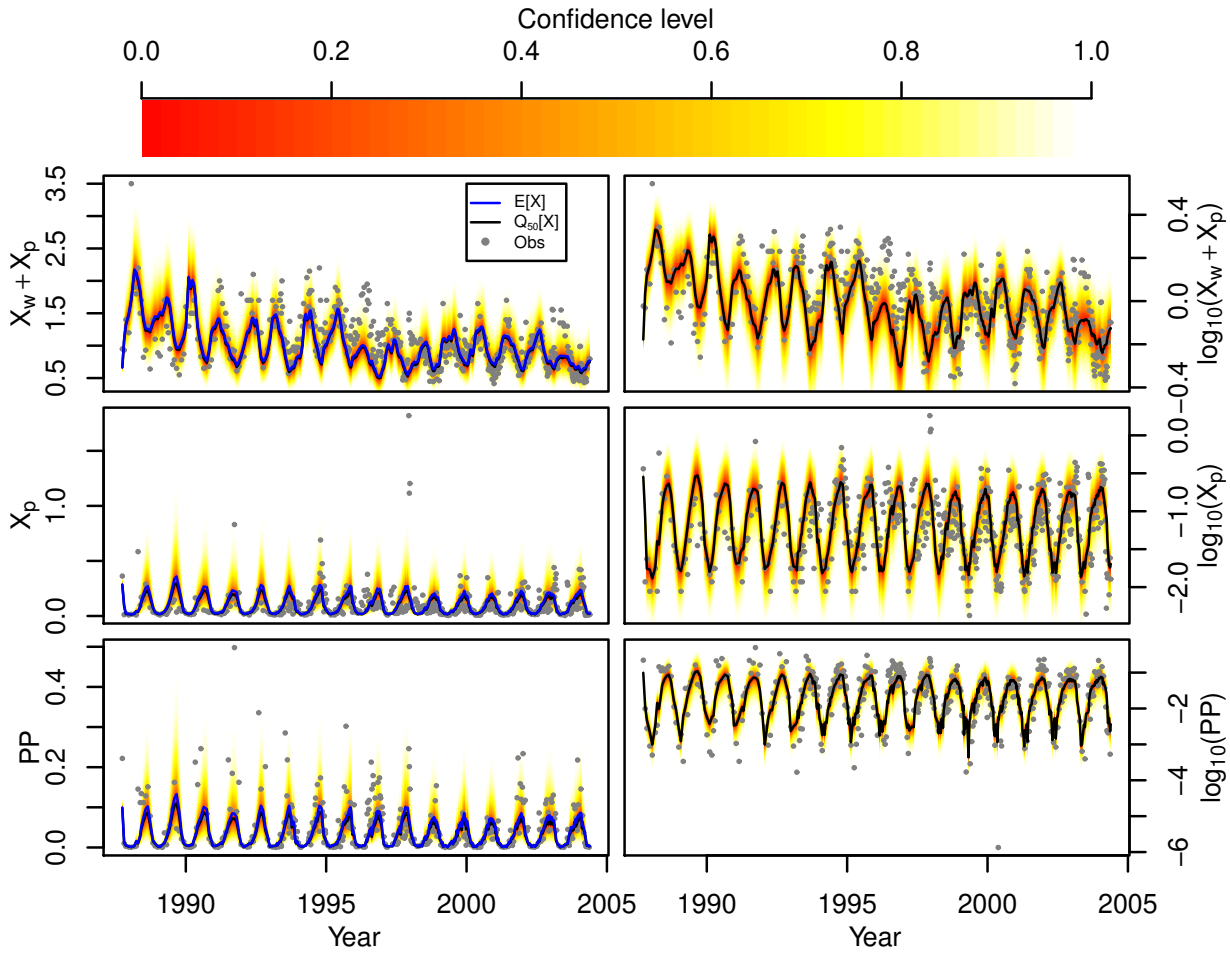
$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_w X_{w,t} & 0 \\ 0 & \sigma_p X_{p,t} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix},$$

variance proportional to X_i^2. A common assumption is that diffusion proportional to $\sqrt{X_i}$, here we estimate the coefficients

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_w X_{w,t}^{\gamma w} & 0 \\ 0 & \sigma_p X_{p,t}^{\gamma p} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}.$$

Reformulate transformation and reestimate.

Validation


Confidence level
0.0
0.2
0.4
0.6
0.8
1.0
E[X]
Q50[X]
Obs
Xw+Xp
Xp
PP
log10(Xw+Xp)
log10(Xp)
log10(PP)
Year
Year
1990
1995
2000
2005

Test summary

Table: Likelihood table

	f	a_{p0}	γ_i	$\log(L)$	DF	AIC	BIC	$P(x > -2Dl)$
Model 1	$f_1 = 1$	0	1	-1446	11	2914	2971	
Model 2	$f_2 = f_1 X_{p,t}$	0	1	-1374	11	2770	2828	
Model 3	$f_3 = \frac{f_2}{k_w + X_{w,t}}$	0	1	-1079	12	2182	2245	
Model 4	$f_4 = \frac{f_3 GR_t}{k_{gr} + GR_t}$	0	1	-1067	13	2159	2227	
Model 5	$f_5 = f_4$	free	1	-1007	14	2042	2115	0.0000
Model 6	$f_6 = f_4$	free	free	-986	16	2004	2088	0.0000

Summary

1. Introduction
2. SDE recap
3. Grey-box models
4. Estimation – The maximum likelihood principle
5. State Dependent diffusion
6. A case study

Further reading

[1] Møller J.K., Carstensen J., Madsen H. (2011) Structural Identification and Validation in Stochastic Differential Equation based Models - With application to a Marine Ecosystem NP-model, *DTU*

[2] Møller J.K. and, Madsen H. (2010) From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform, *DTU*

[3] Thygesen U.H. (2023) Stochastic Differential Equations for Science and Engineering, *Taylor and Francis*

[4] Iacus S. M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples *Springer Series in Statistics*