Statistical Modelling of Physical Systems

An introduction to Grey Box modelling

Henrik Madsen

ESI 101 Course,
NREL, Colorado

hmad.dtu@gmail.com

Outline

1. Introduction
2. Selection of modelling framework
3. Nonlinear versus linear modelling
4. Why Stochastic Differential Equations?
5. Grey Box Modelling
   - The Stochastic State Space Model
6. Model Identification, Estimation, Selection and Validation
   - Identification of model order
   - Identification of functional relation
7. Estimation – The maximum likelihood principle
   - The information matrix
   - Distribution of the ML estimator
   - Tests for individual parameters
   - Likelihood ratio tests
8. Model Validation
9. Software – CTSM-R
10. Summary

Introduction

- Various methods of advanced modelling are needed for an increasing 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 for noise due to approximation and measurement errors.
- Calls for methods that are capable of **bridging the gap between physical and statistical modelling**.

Which type of model to use?

- Simple / Complex
- Lumped / Distributed
- Linear / Non-linear
- Time-invariant / Time-varying
- Discrete / Continuous time
- Deterministic / Stochastic
- Black box / Grey box / White box
- Parametric / Non-parametric

Base the decision on

- Purpose
- Prior knowledge
- Available data
- Tools

Nonlinear versus linear modelling

- The aim of the modelling effort may be generally expressed as follows: Find a **nonlinear** function h such that $\{\epsilon_t\}$ defined by

$$h(X_t, X_{t-1}, \dots) = \epsilon_t \tag{1}$$

  is a sequence of independent random variables.
- Suppose also that the model is *causally invertible*, i.e. the equation above may be 'solved' such that we may write

$$X_t = h'(\epsilon_t, \epsilon_{t-1}, \dots). \tag{2}$$

Nonlinear vs. linear model buiding (cont.)

- Suppose that h' is sufficiently well-behaved to be expanded in a Taylor series

$$\begin{aligned} X_t &= \mu + \sum_{k=0}^{\infty} g_k \epsilon_{t-k} + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} g_{kl} \epsilon_{t-k} \epsilon_{t-l} \\ &\quad + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} g_{klm} \epsilon_{t-k} \epsilon_{t-l} \epsilon_{t-m} + \dots \end{aligned} \tag{3}$$

- The functions

$$\mu = h'(0), \; g_k = \left(\frac{\partial h'}{\partial \epsilon_{t-k}}\right), \; g_{kl} = \left(\frac{\partial^2 h'}{\partial \epsilon_{t-k} \partial \epsilon_{t-l}}\right), \text{ etc.} \tag{4}$$

are called the *Volterra series* for the process $\{X\}$. The sequences $g_k, g_{kl}, \dots$ are called the *kernels* of the Volterra series.

Non-linear vs. linear model building (cont.)

- For **linear systems**

$$g_{kl} = g_{klm} = g_{klmn} = \ldots = 0 \qquad (5)$$

- Hence the system is completely characterized by either

$$\begin{aligned} \{g_k\} \quad &: \quad \text{Impulse response function} \\ \text{or} \quad & \\ \mathcal{H}(\rightarrow) \quad &: \quad \text{Frequency response function} \end{aligned}$$

Non-linear vs. linear model building (cont.)

- In general there is *no such thing as a transfer function* for non-linear systems.
- However, an *infinite sequence of generalized transfer functions* may be defined as

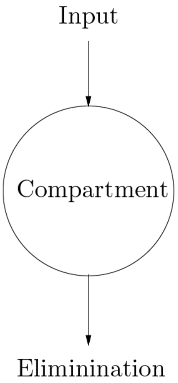
$$
\begin{aligned}
H_1(\omega_1) &= \sum_{k=0}^{\infty} g_k e^{-i\omega_1 k} \\
H_2(\omega_1, \omega_2) &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} g_{kl} e^{-i(\omega_1 k + \omega_2 l)} \\
H_3(\omega_1, \omega_2, \omega_3) &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} g_{klm} e^{-i(\omega_1 k + \omega_2 l + \omega_3 m)} \\
&\vdots
\end{aligned}
$$

Non-linear vs. linear model building (cont.)

Let U_t and X_t denote the input and the output of a given system.

- For **linear** systems it is well known that
  - L1 If the input is a single harmonic $U_t = A_0 e^{i\omega_0 t}$ then the output is a single harmonic of *the same frequency*, but with the amplitude scaled by $|H(\omega_0)|$ and the phase shifted by $\arg H(\omega_0)$.
  - L2 Due to the linearity, the *principle of superposition* is valid, and the total output is the sum of the outputs corresponding to the individual frequency components of the input. (Hence the system is completely described by knowing the response to all frequencies – that is what the transfer function supplies).
- For **non-linear** systems, however, neither of the properties (L1) or (L2) hold.
  - NL1 For an input with frequency ω_0, the output will, in general, contain also components at the frequencies $2\omega_0, 3\omega_0, \ldots$ (*frequency multiplication*).
  - NL2 For two inputs with frequencies ω_0 and ω_1, the output will contain components at frequencies $\omega_0, \omega_1, (\omega_0 + \omega_1)$ and all harmonics of the frequencies (*intermodulation distortion*).

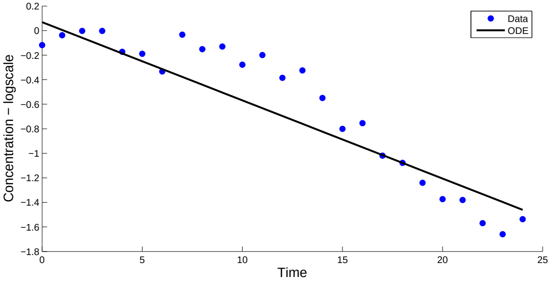
Why Stochastic Differential Equations?
Problem Scenario



Ordinary differential equation

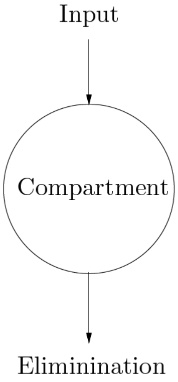
$$\mathrm{d}A = -KA\,\mathrm{d}t$$
$$Y = A + \epsilon$$

ODE



- Autocorrelated residuals!!

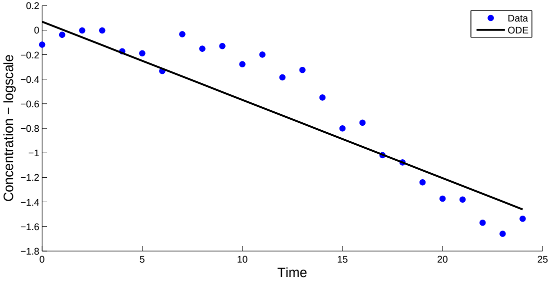
Problem Scenario



Stochastic differential equation

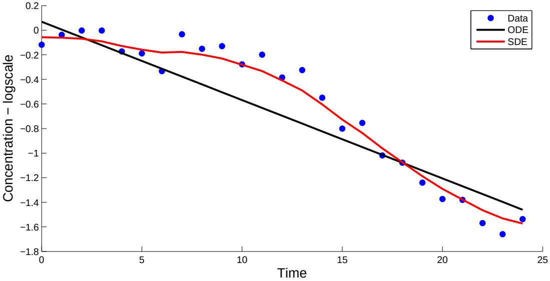
$$dA = -KA\,dt + dw$$
$$Y = A + e$$

ODE vs SDE



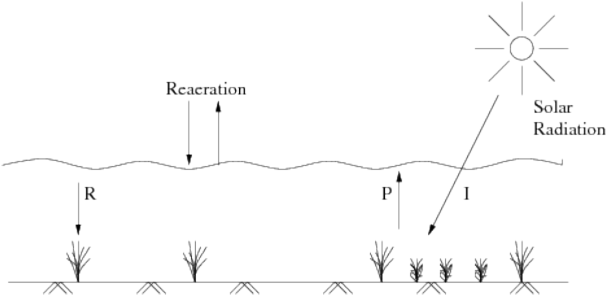
- Uncorrelated residuals
- System noise
- Measurement noise

ODE vs SDE



- Uncorrelated residuals
- System noise
- Measurement noise

Grey box modelling of oxygen concentration
- A sketch of the physical system



Grey box modelling of oxygen concentration - A white box model

Model found in the literature:

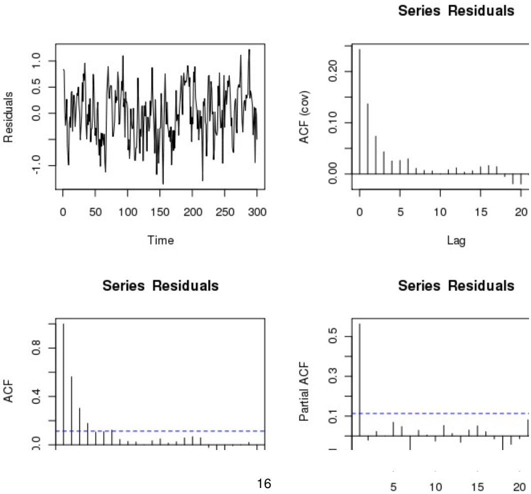
$$\frac{dC}{dt} = \frac{K}{h\sqrt{h}}(C_m(T) - C) + P(I) - R(T)$$

$$\begin{aligned} P(I) &= P_m E_0 \frac{I}{P_m + E_0 I} (= \beta I) \\ R(T) &= R_{15}\theta^{T-15} \quad [mg/l] \\ C_m(T) &= 14.54 - 0.39T + 0.01T^2 \quad [mg/l] \end{aligned}$$

- Simple - however, a non-linear model.
- Uncertainty of prediction does not depend on horizon.

Model validation

The **autocorrelation** and **partial autocorrelation** function for the residuals from the first order model



Grey box model of oxygen concentration

The following nonlinear state space (Hidden Markov) model has been found:

***The system equation**:

$$\begin{bmatrix} dC \\ dL \end{bmatrix} = \begin{bmatrix} \frac{K}{h\sqrt{h}} & -K_c \\ K_3 & -K_l \end{bmatrix} \begin{bmatrix} C \\ L \end{bmatrix} dt + \begin{bmatrix} \beta & \frac{\sqrt{C}K_b}{h} \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} I \\ P_r \end{bmatrix} dt$$

$$+ \begin{bmatrix} \frac{K}{h\sqrt{h}} C_m(T) - R(T) \\ 0 \end{bmatrix} dt + \begin{bmatrix} dW_1(t) \\ dW_2(t) \end{bmatrix}$$

***The observation equation**:

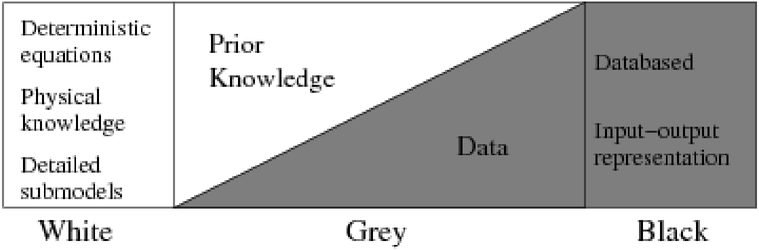
$$C_r = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} C \\ L \end{bmatrix} + e$$

Model types

- **White box models**
  - the model structure is known and deterministic.
  - uncertainty is discarded and the model tends to be overspecified.
- **Black box models**
  - data based models - input/output models.
  - the model and its parameters have little physical significance.
- **Grey box models**
  - between white and black box models

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

The system equation (set of Itô stochastic differential eqs.)

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

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
$\boldsymbol{W}_t$	Wiener process of dimension, d, with incremental covariance $\boldsymbol{Q}_t$
$\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$	Unknown parameter vector

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 : $t_1 < \ldots < t_i < \ldots < t_N$

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

Advantages of HMM with the dynamics described by SDE's

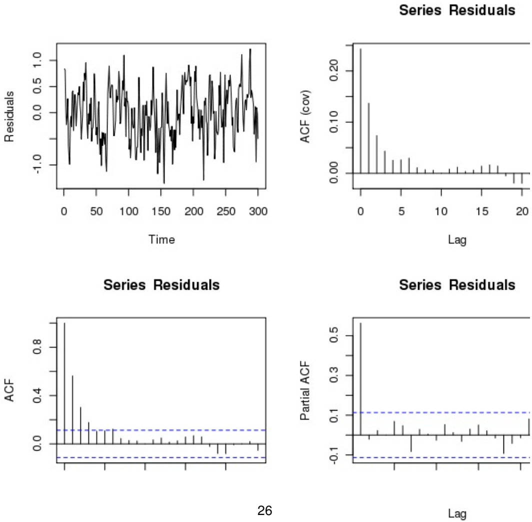
- Provides a decomposition of the total error into process error and measurement error.
- Facilitates use of statistical tools for model validation.
- Provides a systematic framework for pinpointing model deficiencies – will be demonstrated later on.
- Covariances of system error and measurement error 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.
- **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 model order (here: number of states)

Use the **autocorrelation** and **partial autocorrelation** functions



Identification input variables and eg. time delays

Use **the Pre-whitening procedure (pp. 223-226 in Time Series Analysis book)** or **Ridge regression (pp. 227-228 in TSA)**.

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**:

$$
\begin{aligned}
\mathrm{d}\boldsymbol{X}_t &= f(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{\theta})\,\mathrm{d}t + G(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{\theta})\,\mathrm{d}\boldsymbol{W}_t, \quad \boldsymbol{X}_{t_0} = \boldsymbol{X}_0 \\
\mathrm{d}r_t &= \mathrm{d}W_t^* \\
\boldsymbol{Y}_{t_i} &= h(\boldsymbol{X}_{t_i}, \boldsymbol{U}_{t_i}, \boldsymbol{\theta}) + \boldsymbol{e}_{t_i}
\end{aligned}
\tag{6}
$$

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 $\widehat{r}_{t|N}$ (use for instance the software tool CTSM-R - see slides by Niamh O'Connell).
- 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.

The information matrix

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.

Invariance property

[redacted]

Assume that $\widehat{\boldsymbol{\theta}}$ *is a maximum likelihood estimator for* $\boldsymbol{\theta}$*, and let* $\boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{\theta})$ *denote a one-to-one mapping of* $\Omega \subset \mathbb{R}^k$ *onto* $\Psi \subset \mathbb{R}^k$*. Then the estimator* $\boldsymbol{\psi}(\widehat{\boldsymbol{\theta}})$ *is a maximum likelihood estimator for the parameter* $\boldsymbol{\psi}(\boldsymbol{\theta})$*.*

The principle is easily generalized to the case where the mapping is not one-to-one.

Distribution of the ML estimator

We assume that $\widehat{\boldsymbol{\theta}}$ is consistent. Then, under some regularity conditions,

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \to \mathrm{N}(0, \boldsymbol{i}(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{i}(\boldsymbol{\theta})$ is the expected information or the information matrix.

The results can be used for inference under very general conditions. As the price for the generality, the results are only asymptotically valid.

- Asymptotically the variance of the estimator is seen to be equal to the Cramer-Rao lower bound for any unbiased estimator.
- The practical significance of this result is that the MLE makes efficient use of the available data for large data sets.

Distribution of the ML estimator

In practice, we would use

$$\widehat{\theta} \sim \mathrm{N}(\theta, \boldsymbol{j}^{-1}(\widehat{\boldsymbol{\theta}}))$$

where $\boldsymbol{j}(\widehat{\boldsymbol{\theta}})$ is the observed (Fisher) information.

This means that asymptotically

i) $E[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$

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

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 $\widehat{\theta}_j$.

The Wald Statistic

A test of an individual parameter

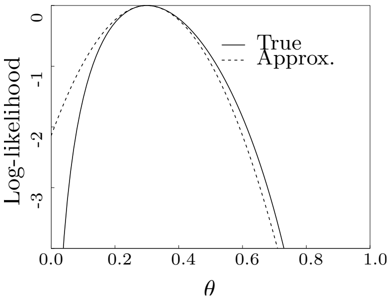
$$\mathcal{H}_0 : \theta_i = \theta_{i,0}$$

is given by the *Wald statistic*:

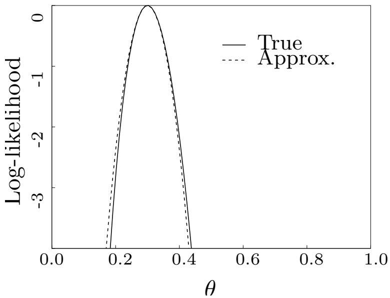
$$Z_i = \frac{\widehat{\theta}_i - \theta_{i,0}}{\widehat{\sigma}_{\hat{\theta}_i}}$$

which under $\mathcal{H}_0$ is approximately N(0, 1)-distributed.

Example: Quadratic approximation of the log-likelihood



(a) $n = 10, y = 3$



(b) $n = 100, y = 30$

Figure: Quadratic approximation of the log-likelihood function.

Likelihood ratio tests

- Want to know the distribution of D when assuming $\mathcal{H}_0$ (model B).
- It is sometimes possible to calculate the exact distribution. This is for instance the case for the General Linear Model for Gaussian data.
- In most cases, however, we must use following important result regarding the asymptotic behavior.

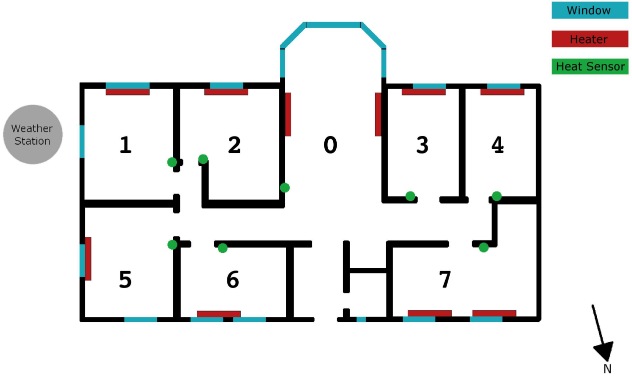
[redacted]

The random variable $D = 2(\ell_A(\widehat{\boldsymbol{\theta}_A}, \boldsymbol{Y}) - \ell_B(\widehat{\boldsymbol{\theta}_B}, \boldsymbol{Y}))$ *converges in law to a* χ^2 *random variable with* $f = (\dim(\Omega_A) - \dim(\Omega_B))$ *degrees of freedom, i.e.,*

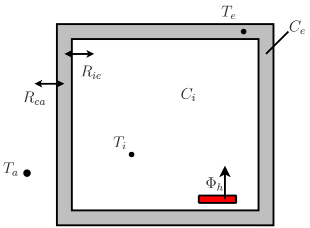
$$D \to \chi^2(f)$$

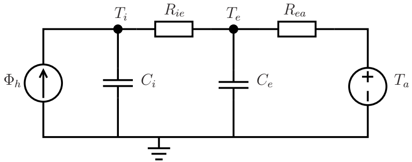
under $\mathcal{H}_0$*.*

Flexhouse layout

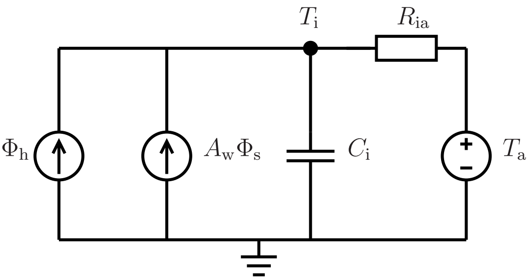


RC-diagram ofte used for illustrating linear models





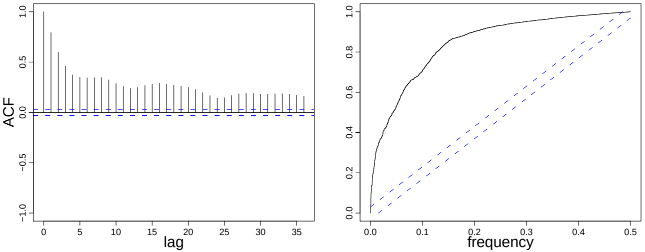
Model A



A_w is the effective window area.

Model Validation for Model A

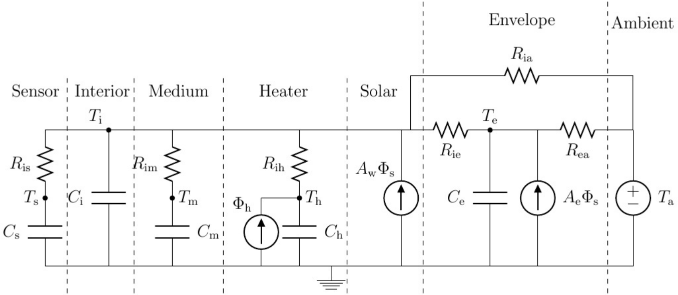
Autocorrelation function and Periodogram for the residuals.



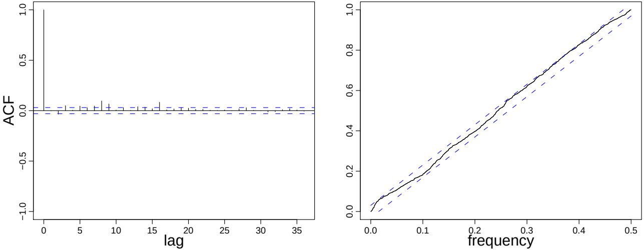
Model is seen not to be adequate.

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-R)

- The parameter estimation is performed by using the software CTSM-R.
- The software has been developed at DTU Compute
- Download from www.ctsm.info (see also slides by Niamh O'Connell)
- The program returns the uncertainty of the parameter estimates as well.

The estimation procedure (CTSM-R)

CTSM-R is based on

- The Extended Kalman Filter
- Approximate likelihood estimation

The estimation procedure (CTSM-R)

CTSM-R is based on

- The Extended Kalman Filter
- Approximate likelihood estimation

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 $E[\boldsymbol{X}_t|\mathcal{Y}_{t_0}]$

The continuous-discrete time stochastic state space formulation

General formulation

$$
\begin{aligned}
d\boldsymbol{x}_t &= \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)d\boldsymbol{w}_t \\
\boldsymbol{y}_k &= \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, \boldsymbol{e}_k, t_k),
\end{aligned}
$$

where

- $\boldsymbol{x}_t$ is the continuous time state and $\boldsymbol{y}_k \in \mathbb{R}^l$ is the discrete time observations.
- $\boldsymbol{u}_t \in \mathbb{R}^r$ is the inputs
- $\boldsymbol{\theta} \in \mathbb{R}^p$ is a parameter vector
- $\boldsymbol{e}_k \in \mathbb{R}^l$ is a random observation error.

The estimation procedure (CTSM) - Limitations

Most general set up in CTSM

$$
\begin{aligned}
d\boldsymbol{x}_t &= \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, \boldsymbol{\theta}, t)d\boldsymbol{w}_t \\
\boldsymbol{y}_k &= \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, t_k) + \boldsymbol{e}_k,
\end{aligned}
$$

where

- $\boldsymbol{\sigma} \in \mathbb{R}^{n \times n}$ is a quadratic matrix, independent of the state
- $\boldsymbol{e}_k \sim N(\boldsymbol{0}, \boldsymbol{S}_k(\boldsymbol{\theta}, \boldsymbol{u}_k))$ is a Gaussian random variable.

Transformation of the State Space 1

Consider the system equation

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)\boldsymbol{R}(\boldsymbol{u}_t, \boldsymbol{\theta}, t)d\boldsymbol{w}_t,$$

where $\boldsymbol{R}(\boldsymbol{u}_t, \boldsymbol{\theta}, t) \in \mathbb{R}^{n\times n}$ is any matrix function and $\boldsymbol{\sigma} \in \mathbb{R}^{n\times n}$ is a diagonal matrix

$$\sigma_{ii}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t) = \sigma_i(x_{i,t}, \boldsymbol{u}_t, \boldsymbol{\theta}, t)$$

Transformation of the State Space 2

Choose the transformation

$$z_{i,t} = \psi^i(x_{i,t}, \boldsymbol{u}_t, \boldsymbol{\theta}, t) = \int \frac{d\xi}{\sigma_i(\xi, \boldsymbol{u}_t, \boldsymbol{\theta}, t)}\bigg|_{\xi = x_i},$$

Transformation of the State Space 2

Choose the transformation

$$z_{i,t} = \psi^i(x_{i,t}, \boldsymbol{u}_t, \boldsymbol{\theta}, t) = \int \frac{d\xi}{\sigma_i(\xi, \boldsymbol{u}_t, \boldsymbol{\theta}, t)}\Bigg|_{\xi = x_i},$$

then by Itô's lemma z_i is also an Itô process given by

$$dz_{i,t} = \frac{\partial}{\partial t}\psi^i(\cdot, t)dt + \frac{f_i(\cdot)}{\sigma_i(\cdot)}dt - \frac{1}{2}\sigma_i(\cdot)\sum_{j=1}^{n}[\boldsymbol{R}(\cdot)]^2_{i,j}\,dt + \sum_{j=1}^{n}[\boldsymbol{R}(\cdot)]_{i,j}\,dw_j,$$

where the diffusion term is now independent of the state z_i.

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.
- we obtain more accurate predictions and more realistic prediction intervals.
- we obtain realistic simulations (ODE based models do not provide a reasonable framework for simulations)

Some References

- H. Madsen: *Time Series Analysis*, Chapman and Hall, 392 pp, 2008.
- J.M. Morales, A.J. Conejo, H. Madsen, P. Pinson, M. Zugno: *Integrating Renewables in Electricity Markes*, Springer, 430 pp., 2013.
- H.Aa. Nielsen, H. Madsen: *A generalization of some classical time series tools*, Computational Statistics and Data Analysis, Vol. 37, pp. 13-31, 2001.
- H. Madsen, J. Holst: *Estimation of Continuous-Time Models for the Heat Dynamics of a Building*, Energy and Building, Vol. 22, pp. 67-79, 1995.
- J.N. Nielsen, H. Madsen, P.C. Young: *Parameter Estimation in Stochastic Differential Equations; An Overview*, Annual Reviews in Control, Vol. 24, pp. 83-94, 2000.
- N.R. Kristensen, H. Madsen, S.B. Jørgensen: *A Method for systematic improvement of stochastic grey-box models*, Computers and Chemical Engineering, Vol 28, 1431-1449, 2004.
- N.R. Kristensen, H. Madsen, S.B. Jørgensen: *Parameter estimation in stochastic grey-box models*, Automatica, Vol. 40, 225-237, 2004.
- J.B. Jørgensen, M.R. Kristensen, P.G. Thomsen, H. Madsen: *Efficient numerical implementation of the continuous-discrete extended Kalman Filter*, Computers and Chemical Engineering, 2007.
- K.R. Philipsen, L.E. Christiansen, H. Hasman, H. Madsen: *Modelling conjugation with stochastic differential equations*, Journal of Theoretical Biology, Vol. 263, pp. 134-142, 2010.
- P. Bacher, H. Madsen: *Identifying suitable models for the heat dynamics of buildings*, Vol. 43, pp. 1511-1522, 2011.