Statistical Models and Methods;
Thermal performance and Occupant behavior

REBUS WP2B
Saint-Gobain, May 2017

Henrik Madsen

www.henrikmadsen.org
George Box:

All models are wrong – but some are useful
Modeling made simple

Suppose we have a time series of data:

\[ \{X_t\} = X_1, X_2, \ldots , X_t, \ldots \]

The purpose of any modeling is to find a nonlinear function \( h(\{X_t\}) \) such that

\[ h(\{X_t\}) = \varepsilon_t \]

Where \( \{\varepsilon_t\} \) is white noise – ie no autocorrelation
Thermal performance characterization using time series data - statistical guidelines

IEA EBC Annex 58

Henrik Madsen and Peder Bacher¹
DTU Compute, Technical University of Denmark
Lyngby, Denmark

Geert Bauwens, An-Heleen Deconinck, Glenn Reynders and Staf Roels
KU Leuven, Civil Engineering Department, Building Physics Section
Heverlee, Belgium

Eline Himpe
Ghent University, Department of Architecture and Urban Planning, Building Physics Group
Gent, Belgium

Guillaume Lethé
BBRI, Belgian Building Research Institute
Brussels, Belgium

November 28, 2016
Methods in Annex 58 Guidelines

- **Linear regression**  
  (steady state approach)

- **ARX model**  
  (dynamical, linear, time-invariant)

- **Grey-box model** (RC-network model + )  
  (dynamical, linear or nonlinear, time-varying)

The Annex 58 Guidelines contains recipes as well as examples are in R (open source stat package)
Static and dynamic conditions: estimate the Heat Loss Coefficient (HLC) and gA-value from 'simple' data:

- Constant indoor temperature
- Model input: ambient temperature and global radiation (wind not included in guideline models)
- Model output: heat load

Grey-box models for detailed building behavior characterization:

- Varying indoor temperature (turn the heating on/off)
- Model input: ambient temperature, global radiation, wind
- Model output: indoor air temperature

Procedures (recipes) for model selection and validation, with examples in R
Contents

1. A single sensor (a smart meter)
2. Several sensors (and grey-box modelling)
3. Special sensors (model for occupant behavior)
Part 1
A single sensor (smart meter)

- Smart Meters and data splitting
- Smart Meters and Thermal Characteristics
  - Problem setting
  - Simple tool
Case Study No. 1

Split of total readings into space heating and domestic hot water using data from smart meters
Data

- 10 min averages from a number of houses

<table>
<thead>
<tr>
<th>House</th>
<th>Year build</th>
<th>House size</th>
<th>Occupants</th>
</tr>
</thead>
<tbody>
<tr>
<td>House 1</td>
<td>1963</td>
<td>119 m²</td>
<td>2</td>
</tr>
<tr>
<td>House 2</td>
<td>1937</td>
<td>86 m²</td>
<td>2</td>
</tr>
<tr>
<td>House 3</td>
<td>1963</td>
<td>140 m²</td>
<td>2</td>
</tr>
<tr>
<td>House 4</td>
<td>1967</td>
<td>137 m²</td>
<td>5</td>
</tr>
</tbody>
</table>
Data separation principle

- House Characteristic
  - e.g. size, insulating power, solar absorption

- Occupants Characteristic
  - e.g. open/close windows, turn up/down the heating, night-time drop

- Heating Consumption

- Hot Water Consumption
  - e.g. shower, dishwashing

- Raw Data
Holiday period

House: 2, Occupants: 2

REBUS 2B meeting,
Saint-Gobain, May 2017
Non-parametric regression

\[
\hat{g}(x) = \frac{\sum_{s=1}^{N} Y_s k\left\{ \frac{x-X_s}{h} \right\}}{\sum_{s=1}^{N} k\left\{ \frac{x-X_s}{h} \right\}}
\]

\[
k(u) = \frac{1}{2\pi} \exp\left\{ -\frac{u^2}{2} \right\}
\]

Weighted average

Every spike above \( 1.25 \cdot \hat{g}(x) \) is regarded as hot water use.
Robust Polynomial Kernel

To improve the kernel method

Rewrite the kernel smoother to a Least Square Problem

\[ \arg \min_\theta \frac{1}{N} \sum_{s=1}^{N} w_s(x) (Y_s - \theta)^2 \quad w_s(x) = \frac{k\{x - X_s\}}{\frac{1}{N} \sum_{s=1}^{N} k\{x - X_s\}} \]

Make the method robust by replacing \((Y_s - \theta)^2\) with

\[ \rho_{\text{Huber}}(\varepsilon) = \begin{cases} \frac{1}{2\gamma} \varepsilon^2 & \text{if } |\varepsilon| \leq \gamma \\ |\varepsilon| - \frac{1}{2}\gamma & \text{if } |\varepsilon| > \gamma \end{cases} \]

\[ \varepsilon_s = Y_s - \theta \]

Make the method polynomial by replacing \(\theta\) with

\[ P_s = \theta_0 + \theta_1(X_t - x) + \theta_2(X_t - x)^2 \]
Case Study No. 2

Identification of Thermal Performance using Smart Meter Data
Example

\[ U = 0.86 \text{ W/m}^2\text{K} \]

\[ U = 0.21 \text{ W/m}^2\text{K} \]
Examples (2)

Whole House Heat Loss - Measured versus Predicted for New Build UK Dwellings (n=18)

Measured versus predicted energy consumption for different dwellings
Characterization Smart Meter Data

- Energy labelling
- Estimation of UA and gA values
- Estimation of energy signature
- Estimation of dynamic characteristics
- Estimation of time constants
Simple estimation of UA-values

Consider the following model (t=day No.) estimated by kernel-smoothing:

\[ Q_t = Q_0(t) + c_0(t)(T_{i,t} - T_{a,t}) + c_1(t)(T_{i,t-1} - T_{a,t-1}) \]  \hspace{1cm} (1)

The estimated UA-value is

\[ \hat{U}A(t) = \hat{c}_0(t) + \hat{c}_1(t) \]  \hspace{1cm} (2)

With more involved (but similar models) also gA and wA values can be estimated.
Estimated UA-values
## Results

<table>
<thead>
<tr>
<th>UA</th>
<th>$\sigma_{UA}$</th>
<th>$gA_{\text{max}}$</th>
<th>$wA_{E,\text{max}}$</th>
<th>$wA_{S,\text{max}}$</th>
<th>$wA_{W,\text{max}}$</th>
<th>$T_i$</th>
<th>$\sigma_{T_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4218598</td>
<td>211.8</td>
<td>10.4</td>
<td>597.0</td>
<td>11.0</td>
<td>3.3</td>
<td>8.9</td>
<td>23.6</td>
</tr>
<tr>
<td>4381449</td>
<td>228.2</td>
<td>12.6</td>
<td>1012.3</td>
<td>29.8</td>
<td>42.8</td>
<td>39.7</td>
<td>19.4</td>
</tr>
<tr>
<td>4711160</td>
<td>155.4</td>
<td>6.3</td>
<td>518.8</td>
<td>14.5</td>
<td>4.4</td>
<td>9.1</td>
<td>22.5</td>
</tr>
<tr>
<td>4836681</td>
<td>155.3</td>
<td>8.1</td>
<td>591.0</td>
<td>39.5</td>
<td>28.0</td>
<td>21.4</td>
<td>23.5</td>
</tr>
<tr>
<td>4836722</td>
<td>236.0</td>
<td>17.7</td>
<td>1578.3</td>
<td>4.3</td>
<td>3.3</td>
<td>18.9</td>
<td>23.5</td>
</tr>
<tr>
<td>4986050</td>
<td>159.6</td>
<td>10.7</td>
<td>715.7</td>
<td>10.2</td>
<td>7.5</td>
<td>7.2</td>
<td>20.8</td>
</tr>
<tr>
<td>5069878</td>
<td>144.8</td>
<td>10.4</td>
<td>87.6</td>
<td>3.7</td>
<td>1.6</td>
<td>17.3</td>
<td>21.8</td>
</tr>
<tr>
<td>5069913</td>
<td>207.8</td>
<td>9.0</td>
<td>962.5</td>
<td>3.7</td>
<td>8.6</td>
<td>10.6</td>
<td>22.6</td>
</tr>
<tr>
<td>5107720</td>
<td>189.4</td>
<td>15.4</td>
<td>657.7</td>
<td>41.4</td>
<td>29.4</td>
<td>16.5</td>
<td>21.0</td>
</tr>
</tbody>
</table>
Based on measurements from the heating season 2009/2010 your typical indoor temperature during the heating season has been estimated to 24 °C. If this is not correct you can change it here 24 °C.

If your house has been left empty in longer periods with a partly reduced heat supply you have the possibility of specifying the periods in this calendar.

According to BBR the area of your house is 155 m² and from 1971.

Based on BBR information it is assumed that you do not use any supplementary heat supply. If this is not correct you can specify the type and frequency of use here:

- Wood burning stove used [0] times per week in cold periods.
- Solar heating [y/n], approximate size of solar panel [0] × [0] meters.

Based on the indoor temperature 24 °C, the use of a wood burning stove 0 times per week, and no solar heating installed, the response of your house to climate is estimated as:

- The response to outdoor temperature is estimated to 200 W/°C which given the size and age of your house is expectable.

- On a windy day the above value is estimated to increase with 60 W/°C when the wind blows from easterly directions. This response to wind is relatively high and indicates a problem related to the air sealing on the eastern side of the house.

- On a sunny day during the heating season the house is estimated to receive 800 W as an average over 24 hours. This value is quite expectable.

*aMany kind of different recommendations can be given here.*
Perspectives for using Smart Meters

- Reliable Energy Signature.
- Energy Labelling
- Time Constants (eg for night setback)

Proposals for Energy Savings:
- Replace the windows?
- Put more insulation on the roof?
- Is the house too untight?
- ......

- Optimized Control

- Integration of Solar and Wind Power using DSM
Part 2
Several sensors

- Introduction to Grey-Box Modelling (a continuous-discrete state space models)
- A model for the thermal characteristics of a small office building
- Models for control
Introduction to Grey-Box modelling
Traditional Dynamical Model

- Ordinary Differential Equation:

\[ \frac{dA}{dt} = -KA \]
\[ Y = A + \epsilon \]
Stochastic Dynamical Model

Stochastic Differential Equation:

\[ dA = -KAdt + \sigma dw \]
\[ Y = A + \epsilon \]
The grey box model

Notation:

\( X_t \): State variables
\( u_t \): Input variables
\( \theta \): Parameters
\( Y_k \): Output variables
\( t \): Time
\( \omega_t \): Standard Wiener process
\( e_k \): White noise process with \( N(0, S) \)
Grey-box modelling concept

- Combines prior physical knowledge with information in data
- Equations and parameters are physically interpretable
Forecasting and Simulation

Grey-Box models are well suited for ...

- One-step forecasts
- K-step forecasts
- Simulations
- Control
- … of both observed and hidden states.

- It provides a framework for pinpointing model deficiencies – like:
  - Time-tracking of unexplained variations in e.g. parameters
  - Missing (differential) equations
  - Missing functional relations
  - Lack of proper description of the uncertainty
Grey-Box Modelling

- Bridges the gap between physical and statistical modelling
- Provides methods for model identification
- Provides methods for model validation
- Provides methods for pinpointing model deficiencies
- Enables methods for a reliable description of the uncertainties, which implies that the same model can be used for k-step forecasting, simulation and control
Grey-Box Modelling

- Bridges the gap between physical and statistical modelling
- Provides methods for model identification
- Provides methods for model validation
- Provides methods for pinpointing model deficiencies
- Enables methods for a reliable description of the uncertainties, which implies that the same model can be used for k-step forecasting, simulation and control
Grey box model building framework

- Initial model
- Transform to SDE model
- Non parametric modelling
- Tracking variations
- Estimate parameters
- Model evaluation
- Extend model
- Estimate parameters
- Final model

Yes/No

Case study

Model for the thermal characteristics of a small office building
Test case: One floored 120 m² building

Objective

Find the best model describing the heat dynamics of this building ([1], [4]).
Data

Measurements of:

$y_t$ Indoor air temperature

$T_a$ Ambient temperature

$\Phi_h$ Heat input

$\Phi_s$ Global irradiance
Selection Procedure

Iterative procedure using statistical tests

Start $l(\theta; Y_N)$

$\begin{array}{c|c|c|c|c}
1 & Model_{T_{i}e} & Model_{T_{i}m} & Model_{T_{i}T_{s}} & Model_{T_{i}T_{h}} \\
10 & 3628.0 & 3639.4 & 3884.4 & 3911.1 \\
\end{array}$

Simplest model

First extension: heater part
EVALUATE THE SIMPLEST MODEL

Inputs and residuals

ACF of residuals

Cumulated periodogram

---

Saint-Gobain, May 2017
REBUS 2B meeting,
Saint-Gobain, May 2017
Grey-box modelling

Continuous time models (grey-box: stochastic state-space model)

\[
\begin{align*}
\text{States} & = \text{Fun}_1(\text{States, Inputs}) + \text{Fun}_2(\text{Inputs}) \cdot \text{SystemError} \\
\text{Measurements} & = \text{Fun}_3(\text{States, Inputs}) + \text{Fun}_4(\text{Inputs}) \cdot \text{MeasurementError}
\end{align*}
\]

- Used for buildings (single- and multi-zone), walls, systems (hot water tank, integrated PV, heat pumps, heat exchanger, solar collectors, ...)
- Formulate the model based on physical knowledge
- Maximum likelihood estimation (we have the entire statistical framework available)
- Description of the system noise is part of the model provides some very useful possibilities (e.g. control the weight of data in the estimation depending on input signals)
- Software, for example our R package \textsc{CTSM-R} \footnote{http://ctsm.info}
Case study

Models for Smart Control
(Ex: Control of Heat Pumps)
Existing Markets - Challenges

- Dynamics
- Stochasticity
- Nonlinearities
- Many power related services (voltage, frequency, balancing, spinning reserve, congestion, ...)
- Speed / problem size
- Characterization of flexibility
- Requirements on user installations
Different possibilities can be investigated for the coordination of the flexible resources:

**Market-based approach**

- Market
  - Energy Supplier
  - VPP
  - PA
  - GA

*Market operation* is intended all the way down to the prosumers’ level.

**Control-based approach**

- Market
  - Energy Supplier
  - Aggregator
  - Sub Aggregator
  - Controller

*Control problem* is formulated at the prosumers’ level.
Suggested ‘Market’ Setup (Smart-Energy OS)

Space

Country
Region
City
District
House

Time

Bidding +
Market clearing

Purpose based
Stochastic Control

Economic Model
Predictive Control

Bidding &
Clearing

Control based

\[
\min_p (U - U_{ref})^2
\]

\[
\min \sum (pU)
\]
Proposed methodology
Control-based methodology

\[
\begin{align*}
\min_p & \quad \mathbb{E}\left[ \sum_{k=0}^{N} w_{j,k} \left| \hat{z}_k - z_{\text{ref},k} \right| + \mu \left| p_k - p_{\text{ref},k} \right| \right] \\
\text{s.t.} & \quad \hat{z}_{k+1} = f(p_k)
\end{align*}
\]

We adopt a control-based approach where the price becomes the driver to manipulate the behaviour of a certain pool flexible prosumers.
Aggregation (over 20 houses)
Response on Price Step Change

![Graph showing consumption step response (Olympic Pen.) with a 5-hour mark.](image-url)
Control of Power Consumption

Model parameters

Price generator (controller)

Consumption references

Prices

Price-response estimator

Price-responsive consumption

Aggregated consumption
Control performance

- Considerable reduction in peak consumption
PilotB SN-10 signal overview
revision 1.0 (CITIES add-on)
3.2 Optimization problem

The MPC controller solves the following mixed integer linear optimization problem:

\[
\begin{align*}
\min_u & \quad \sum_{k=0}^{N-1} c_k u_k \\
\text{s.t.} & \quad x_{k+1} = A_d(T_o, w, T_a)x_k + B_d(T_o, w, T_a)u_k \\
& \quad y_k = C_d(T_o, w, T_a)x_k \\
& \quad u_k \in \{0, 1\} \\
& \quad y_{\min} \leq y_k \leq y_{\max}
\end{align*}
\]

(3.2a) \hspace{1cm} (3.2b) \hspace{1cm} (3.2c) \hspace{1cm} (3.2d) \hspace{1cm} (3.2e)

where (3.2b) and (3.2c) is discretized state-space model of (2.6); \(u_k\) is the valve position (1 - open; 0 - closed); \(y_k = [T_{in,k} \ T_{out,k}]^T\); \(N\) is the predictive horizon; \(c_k\) is the electricity price.
MPC Results
Part 3
Special Data (eg Non-Gaussian)

Identification of Occupant Behavior

- Use of CO2 measurements to model occupant behavior in summer houses
Today’s situation for indoor climate in buildings

**Design**
- Regulations
- Specifications
- Designers experience
- Standards

**Service**
- User
- Telephone contact
- Janitor (subjective assessment)

**Performance**
- Temperature
- Air Quality
- Humidity
- Draft
- Radiation
- Symmetry

**Percent Persons Dissatisfied**

**Regulations**
- Specifications
- Designers experience
- Standards

**Adjustments to indoor climate**

**Designed based on assumptions**
- Focus on in-direct parameters to user satisfaction
- Little knowledge of true user preferences

**No info of total user satisfaction**
- Subjective responses and actions
- Building controlled after in-direct parameters to user satisfaction
- No systematic logging or learning
Digital revolution

When we use their product, they gather information on how we use it, our preferences etc. Crucial information regarding how the product can be further developed.

The «Internet Of Things» (IoT) and sensor technology enables us to do the same for buildings. But how?
Building automation can be made more accurate to the user’s needs. Rather than using theoretical values and «ad-hoc» user feedback, Smart Technology enables use of Big Data from sensors and continuous user feedback.
Collect data on three levels

<table>
<thead>
<tr>
<th>Physical environment</th>
<th>Sensed environment</th>
<th>Total user satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Skanska Comfort Control App</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smartphone or room tablet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Please submit your feedback:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cooler</td>
<td>Warmer</td>
</tr>
<tr>
<td></td>
<td>Bad air</td>
<td>Draft</td>
</tr>
<tr>
<td></td>
<td>Indoor climate?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Measurement of user satisfaction at entrance door</td>
<td></td>
</tr>
</tbody>
</table>

By using a simple system for data collection via existing room automation systems, new smart sensors, smart phones with IoT and cloud computing we can achieve a high degree of accuracy for the automation system. Collecting data about the indoor environment and user at the same time.
Example:

Existing product/service enabling the user to give feedback regarding thermal comfort via smartphone app. The system uses this feedback to find the best temperature profile regarding economy and user satisfaction for the building. Claims to give 20% reduction in energy costs as well as improved user satisfaction. Could REBUS do the same, using the system to keep our customers happy and at the same time collect data helping us innovate and improve our buildings?
Summer houses represent a special challenge

- Large variation in the number of people present in the house
- Power Grids in summer house areas represent a special problem for some DSOs
- Time series of CO2 measurements are the key to the classification
The Model Space

\[
\begin{align*}
\theta & \sim f \left( \beta_{\text{fixed}}, t, \cdots \right) + g \left( U_{\text{random}}, t, \cdots \right) \\
\mathbf{d}X_t & \sim \text{Dynamical model} \left( \theta \right) \\
Y_t^{(1)} & = \text{Electrical consumption} \\
Y_t^{(2)} & = \text{Noise (indoor)} \\
Y_t^{(3)} & = \text{CO}_2 \ (\text{indoor}) \\
\ddots & \\
\end{align*}
\]

- \( \theta \) parameter vector for population/hierarchical model
  - Time, weather, demographics
- \( \mathbf{d}X_t \) state vector described by some dynamical model depending on \( \theta \)
  - People, consumption, windows
- \( Y \)'s: Observed measurements related to occupancy behavior, including measurements inside and outside the building and smart metering data
Hidden Markov Model

First Order Markov Property

\[
p(X_t | X_{t-1}) = p(X_t | \mathcal{X}^{(t-1)}), \quad t \in \mathbb{N}
\]

\[
p(Y_t | X_t) = p(Y_t | \mathcal{X}^{(t)}, \mathcal{Y}^{(t-1)}), \quad t \in \mathbb{N}
\]

**Figure:** Directed graph of basic HMM. The index denotes time.
Markov Chains

Discrete state vector at time \( t \), \( X_t \), with \( m \) states.

Transition probability

\[
p(X_t = j|X_{t-s} = i)
\]  \hspace{1cm} (4)

One-step transition probability

\[
\gamma_{ij,t} = p(X_t = j|X_{t-1} = i)
\]  \hspace{1cm} (5)

One-step transition probability matrix from time \( t - 1 \) to \( t \)

\[
\Gamma_t = \begin{pmatrix}
\gamma_{11,t} & \cdots & \gamma_{1m,t} \\
\vdots & \ddots & \vdots \\
\gamma_{m1,t} & \cdots & \gamma_{mm,t}
\end{pmatrix}
\]  \hspace{1cm} (6)

where the row must sum to 1.
Homogen Hidden Markov Model

Setting

\[ y_t = h(CO_{2,t}) \]
\[ p(x_t|x_{t-1}) \sim \Gamma \]
\[ p(y_t|x_t) \sim \mathcal{N}(\mu_i, \sigma_i^2) \text{ for } i = 1, 2, \ldots, m \]

Note that there is no time dependence in the transition probabilities in the homogen case.
Table 8.4: Comparison of univariate (log transformed $CO_2$) homogen HMMs for 2 to 5 states.

<table>
<thead>
<tr>
<th>States</th>
<th>$\mathcal{L}$</th>
<th>$p$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 states</td>
<td>-9378</td>
<td>6</td>
<td>18768</td>
<td>18814</td>
</tr>
<tr>
<td>3 states</td>
<td>-4292</td>
<td>12</td>
<td>8609</td>
<td>8701</td>
</tr>
<tr>
<td>4 states</td>
<td>-800</td>
<td>20</td>
<td>1640</td>
<td>1795</td>
</tr>
<tr>
<td>5 states</td>
<td>2181</td>
<td>30</td>
<td>-4303</td>
<td>-4071</td>
</tr>
</tbody>
</table>
Figure 8.7: Global Decoding of the HMM (log $CO_2$) with 5 states.
Inhomogen Hidden Markov Model

Setting

\[ y_t = h(CO_{2,t}) \]
\[ p(x_t|x_{t-1}) \sim \Gamma_t \]
\[ p(y_t|x_t) \sim \mathcal{N}(\mu_i, \sigma_i^2) \text{ for } i = 1, 2, \ldots, m \]

Note that there is time dependence in the transition probabilities in the inhomogen case.
Inhomogen Markov-switching with auto-dependent observations

Figure 8.10: Directed graph of Markov switching AR(1).
Inhomogen Markov-switching AR(1)

Setting

\[ y_t = h(CO_{2,t}) \]
\[ p(x_t|x_{t-1}) \sim \Gamma_t \]
\[ p(y_t|x_t,y_{t-1}) \sim N(c_i + \phi_i y_{t-1}, \sigma^2_i) \text{ for } i = 1,2,\ldots,m \]

Note that there is time dependence in the transition probabilities in the inhomogen case.
Interpretation of the states

- State 1: Absence or sleeping
- State 2: Long term absence
- State 3: Outdoor interaction
- State 4: Presence (high activity)
- State 5: Presence (long term, low activity)
Figure 8.11: Model diagnostics of the final model.
Figure 8.16: Transition probabilities over the day of the final model. The lower right plot is the stationary distribution.
Some conclusions:

That the low activity state 5 is not very likely from 10 am to 11 pm. The high activity is seen in the late afternoon.
Some references


Some references (cont.)


Some 'randomly picked' books on modeling ....
Thanks ...

- For more information
  www.ctsm.info
  www.henrikmadsen.org
  www.smart-cities-centre.org

- ...or contact
  - Henrik Madsen (DTU Compute)
    hmad@dtu.dk

- Acknowledgement CITIES (DSF 1305-00027B)