



What can we learn from data?

Summer School University of Granada, July 2017

Henrik Madsen, www.henrikmadsen.org



DTU Compute

Department of Applied Mathematics and Computer Science

George Box:

All models are wrong – but some are useful



Modeling made simple

Suppose we have a time series of data:

$$\{X_t\} = X_1, X_2, \dots, X_t, \dots$$

The purpose of any modeling is to find a nonlinear function $h({X_{+}})$ such that

$$h({X_t}) = \varepsilon_t$$

Where $\{\epsilon_{t}\}$ 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

Annex 71 meeti Loughborough,



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)

GUIDELINES FROM ANNEX 58

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)

Annex 71 meeting, Loughborough, April 2017

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





Annex 71 meeting, Loughborough, April 2017

Data separation principle



Holiday period



Annex 71 meeting, Loughborough, April 2017

Non-parametric regression

$$\hat{g}(x) = \frac{\sum_{s=1}^{N} Y_s k\{\frac{x - X_s}{h}\}}{\sum_{s=1}^{N} k\{\frac{x - X_s}{h}\}} \qquad \qquad k(u) = \frac{1}{2\pi} \exp\{-\frac{u^2}{2}\}$$

Weighted average

Every spike above $1.25 \cdot \hat{g}(x)$ — Is regarded as hot water use.



Annex 71 meeting, Loughborough, April 2017

Robust Polynomial Kernel

To improve the kernel method

Rewrite the kernel smoother to a Least Square Problem

$$\arg\min_{\theta} \frac{1}{N} \sum_{s=1}^{N} w_s(x) \left(Y_s - \theta\right)^2 \qquad w_s(x) = \frac{k\{x - X_s\}}{\frac{1}{N} \sum_{s=1}^{N} k\{x - X_s\}}$$

Make the method robust by replacing $\left(Y_s- heta
ight)^2$ with

$$\rho_{\text{Huber}}(\varepsilon) = \begin{cases} \frac{1}{2\gamma} \varepsilon^2 & \text{if } |\varepsilon| \le \gamma \\ |\varepsilon| - \frac{1}{2}\gamma & \text{if } |\varepsilon| > \gamma \end{cases} \qquad \varepsilon_s = Y_s - \theta$$

Make the method polynomial by replacing θ with

$$P_{s} = \theta_{0} + \theta_{1}(X_{t} - x) + \theta_{2}(X_{t} - x)^{2}$$

Robust Polynomial Kernel



Annex 71 meeting, Loughborough, April 2017

Case Study No. 2

Identification of Thermal Performance using Smart Meter Data



Example



Consequence of good or bad workmanship (theoretical value is U=0.16W/m2K)

Examples (2)



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:

$$Q_t = Q_0(t) + c_0(t)(T_{i,t} - T_{a,t}) + c_1(t)(T_{i,t-1} - T_{a,t-1})$$
(1)

The estimated UA-value is

$$\hat{UA}(t) = \hat{c}_o(t) + \hat{c}_1(t)$$
 (2)

With more involved (but similar models) also gA and wA values can be stimated

Estimated UA-values



Annex 71 meeting, Loughborough, April 2017

Results

	UA	σ_{UA}	gA^{max}	wA_E^{max}	wA^{max}_S	wA_W^{max}	T_i	σ_{T_i}
	$W/^{\circ}C$		W	$W/^{\circ}C$	$W/^{\circ}C$	$W/^{\circ}C$	°C	
4218598	211.8	10.4	597.0	11.0	3.3	8.9	23.6	1.1
4381449	228.2	12.6	1012.3	29.8	42.8	39.7	19.4	1.0
4711160	155.4	6.3	518.8	14.5	4.4	9.1	22.5	0.9
4836681	155.3	8.1	591.0	39.5	28.0	21.4	23.5	1.1
4836722	236.0	17.7	1578.3	4.3	3.3	18.9	23.5	1.6
4986050	159.6	10.7	715.7	10.2	7.5	7.2	20.8	1.4
5069878	144.8	10.4	87.6	3.7	1.6	17.3	21.8	1.5
5069913	207.8	9.0	962.5	3.7	8.6	10.6	22.6	0.9
5107720	189.4	15.4	657.7	41.4	29.4	16.5	21.0	1.6

•

.

.

•

.

. . .

.

Based on measurements from the heating season 2009/2010 your typical indoor temperature during the heating season has been estimated to $24 \ ^{o}C$. If this is not correct you can change it here $\boxed{24} \ ^{o}C$.

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 $155 m^2$ 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 ^{o}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 $W/{}^{o}C$ which given the size and age of your house is expectable^{*a*}.
- On a windy day the above value is estimated to increase with 60 $W/{}^{o}C$ 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?
 - <u>،</u>
- Optimized Control
- Integration of Solar and Wind Power using DSM





Part 2 Several sensors



- Introduction to Grey-Box Modelling (a continuousdiscrete 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



Stochastic Dynamical Model



The grey box model



Drift term



Diffusion term

System equation Observation equation

Observation noise

Notation:

- X_t : State variables
- u_t : Input variables
- θ : Parameters
- Y_k : Output variables
- t: Time
- ω_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



7/5/17

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
- Φ_h Heat input
- Φ_s Global irradiance



SELECTION PROCEDURE

Iterative procedure using statistical tests Begin with the simplest model Model fitting

OK Likelihood-ratio tests of extended models Not OK Evaluate the selected model

Simplest model



First extension: heater part

10



10

10

т

10

EVALUATE THE SIMPLEST MODEL

Inputs and residuals



Model found using Grey-box modelling (using CTSM-R and a RC-model) Here we estimate the physical parameters



Inputs and residuals



Annex 71 meeting, Loughborough, April 2017 ۰.

GREY-BOX MODELLING

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

 $States = Fun_1(States, Inputs) + Fun_2(Inputs) \cdot SystemError$ Measurements = Fun_3(States, Inputs) + Fun_4(Inputs) \cdot MeasurementError

- Used for buildings (single- and multi-zone), walls, systems (hot water tank, integrated PV, heat pumpts, 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 CTSM-R¹

¹http://ctsm.info Annex 71 meeting, Loughborough, April 2017

Modelling the thermal dynamics of a building integrated and ventilated PV module



Several nonlinear and timevarying phenomena.

Consequently linear RC-network models are not appropriate.

A grey-box approach using CTSM-R is described in Friling et.al. (2009)

Part 3 Special Data (eg Non-Gaussian)

Identification of Occupant Behavior



Use of CO2 measurements to model occupant behavior in summer houses

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

$$oldsymbol{ heta} \sim f\left(eta_{\mathsf{fixed}}, t, \cdots
ight) + g\left(U_{\mathsf{random}}, t, \cdots
ight)$$
 (1a)

$$d\boldsymbol{X}_t \sim \mathsf{Dynamical\ model}\left(\boldsymbol{ heta}
ight)$$
 (1b)

$$Y_t^{(1)} = \text{Electrical consumption}$$

$$egin{aligned} Y_t^{(2)} &= ext{Noise (indoor)} \ Y_t^{(3)} &= ext{CO}_2 \ (ext{indoor}) \end{aligned}$$

•
$$heta$$
 parameter vector for population/hierarchical model

- Time, weather, demographics
- dX_t state vector described by some dynamical model depending on θ
 - 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

$$p(X_t|X_{t-1}) = p(X_t|\mathcal{X}^{(t-1)}), \quad t \in \mathbb{N}$$
(2)
$$p(Y_t|X_t) = p(Y_t|\mathcal{X}^{(t)}, \mathcal{Y}^{(t-1)}), \quad t \in \mathbb{N}$$
(3)



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

$$p(X_t = j | X_{t-s} = i) \tag{4}$$

One-step transition probability

$$\gamma_{ij,t} = p(X_t = j | X_{t-1} = i)$$
(5)

One-step transition probability matrix from time t - 1 to t

$$\mathbf{\Gamma}_{t} = \begin{pmatrix} \gamma_{11,t} & \cdots & \gamma_{1m,t} \\ \vdots & \ddots & \vdots \\ \gamma_{m1,t} & \cdots & \gamma_{mm,t} \end{pmatrix}$$
(6)

where the row must sum to 1.

Homogen Hidden Markov Model

Setting

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

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.

	\mathcal{L}	p	AIC	BIC
2 states	-9378	6	18768	18814
3 states	-4292	12	8609	8701
4 states	-800	20	1640	1795
5 states	2181	30	-4303	-4071



54





Figure 8.7: Global Decoding of the HMM (log CO_2) with 5 states.

State 1.0 -۲ 2 ۲ 0.8 -3 • 4 0.6 5 ¢ • sum 0.4 -0.2 -0.0 7 6 3 5 8 4 9 CO2 In

Density



DTU



58

Inhomogen Hidden Markov Model

Setting

$$\begin{aligned} y_t &= h(CO_{2,t}) \\ p(x_t | x_{t-1}) &\sim \Gamma_t \\ p(y_t | x_t) &\sim \mathcal{N}\left(\mu_i, \sigma_i^2\right) \text{ for } i = 1, 2, \cdots, m \end{aligned}$$

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

$$y_t = h(CO_{2,t})$$

$$p(x_t | x_{t-1}) \sim \Gamma_t$$

$$p(y_t | x_t, y_{t-1}) \sim \mathcal{N}\left(c_i + \phi_i y_{t-1}, \sigma_i^2\right) \text{ for } i = 1, 2, \cdots, m$$

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

DTU

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.



DTU

Time

64



Figure 8.16: Transition probabilities over the day of the final model. The lower right plot is the stationary distribution.



Figure 8.17: Profile of the states over the course of the day. I.e. Stacked stationary probabilities over the course of the day of the final model.

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

- Madsen, Henrik; Holst, Jan. Estimation of Continuous-Time Models for the Heat Dynamics of a Building. In: Energy and Buildings, Vol. 22, 1995
- Andersen, Klaus Kaae; Madsen, Henrik; Hansen, Lars Henrik. Modelling the heat dynamics of a building using stochastic differential equations. In: Energy and Buildings, Vol. 31, No. 1, 2000, p. 13-24.
- Kristensen, Niels Rode; Madsen, Henrik; Jørgensen, Sten Bay. Using continuous time stochstic modelling and nonparametric statistics to improve the quality of first principles models. In: Computer Aided Chemical Engineering, Vol. 10, 2002, p. 901-906
- Kristensen, N.R.; Madsen, Henrik; Jørgensen, Sten Bay. A unified framefork for systematic model improvement. In: Process Systems Engineering, Vol. 15, 2003, p. 1292-1297.
- Kristensen, Niels Rode; Madsen, Henrik; Jørgensen, Sten Bay. Parameter Estimation in Stochastic Grey-Box Models. In: Automatica, Vol. 40, No. 2, 2004, p. 225-237.
- Nielsen, Henrik Aalborg; Madsen, Henrik. Modelling the Heat Consumption in District Heating Systems using a Grey-box approach. In: Energy and Buildings, Vol. 38, No. 1, 2006, p. 63-71.
- Friling, N.; Jimenez, M.J.; Bloem, H.; Madsen, Henrik. Modelling the heat dynamics of building integrated and ventilated photovoltaic modules. In: Energy and Buildings, Vol. 41, No. 10, 2009, p. 1051-1057.
- Bacher, Peder; Madsen, Henrik. Identifying suitable models for the heat dynamics of buildings. In: Energy and Buildings, Vol. 43, No. 7, 2011, p. 1511-1522.
- Lodi, C.; Bacher, Peder; Cipriano, J.; Madsen, Henrik. Modelling the heat dynamics of a monitored Test Reference Environment for Building Integrated Photovoltaic systems using stochastic differential equations. In: Energy and Buildings, Vol. 50, 2012, p. 273-281.
- Morales González, Juan Miguel; Pinson, Pierre; Madsen, Henrik. A Transmission-Cost-Based Model to Estimate the Amount of Market-Integrable Wind Resources. In: I E E E Transactions on Power Systems, Vol. 27, No. 2, 2012, p. 1060-1069.
- Halvgaard, Rasmus; Bacher, Peder; Perers, Bengt; Andersen, Elsa; Furbo, Simon; Jørgensen, John Bagterp; Poulsen, Niels Kjølstad; Madsen, Henrik. Model predictive control for a smart solar tank based on weather and consumption forecasts. In: Energy Procedia, Vol. 30, 2012, p. 270-278.

Some references (cont.)

- Morales González, Juan Miguel; Pinson, Pierre; Madsen, Henrik. A Transmission-Cost-Based Model to Estimate the Amount of Market-Integrable Wind Resources. In: I E E E Transactions on Power Systems, Vol. 27, No. 2, 2012, p. 1060-1069
- Dorini, Gianluca Fabio ; Pinson, Pierre; Madsen, Henrik. Chance-constrained optimization of demand response to price signals. In: I E E E Transactions on Smart Grid, Vol. 4, No. 4, 2013, p. 2072-2080.
- Bacher, Peder; Madsen, Henrik; Nielsen, Henrik Aalborg; Perers, Bengt. Short-term heat load forecasting for single family houses. In: Energy and Buildings, Vol. 65, 2013, p. 101-112.
- Corradi, Olivier; Ochsenfeld, Henning Peter; Madsen, Henrik; Pinson, Pierre. Controlling Electricity Consumption by Forecasting its Response to Varying Prices. In: I E E E Transactions on Power Systems, Vol. 28, No. 1, 2013, p. 421-430.
- Zugno, Marco; Morales González, Juan Miguel; Pinson, Pierre; Madsen, Henrik. A bilevel model for electricity retailers' participation in a demand response market environment. In: Energy Economics, Vol. 36, 2013, p. 182-197.
- Meibom, Peter; Hilger, Klaus Baggesen; Madsen, Henrik; Vinther, Dorthe. Energy Comes Together in Denmark: The Key to a Future Fossil-Free Danish Power System. In: I E E E Power & Energy Magazine, Vol. 11, No. 5, 2013, p. 46-55.
- Andersen, Philip Hvidthøft Delff ; Jiménez, María José ; Madsen, Henrik ; Rode, Carsten. Characterization of heat dynamics of an arctic low-energy house with floor heating. In: Building Simulation, Vol. 7, No. 6, 2014, p. 595-614.
- Andersen, Philip Hvidthøft Delff; Iversen, Anne; Madsen, Henrik; Rode, Carsten. Dynamic modeling of presence of occupants using inhomogeneous Markov chains. In: Energy and Buildings, Vol. 69, 2014, p. 213-223.
- Madsen, H, Parvizi, J, Halvgaard, RF, Sokoler, LE, Jørgensen, JB, Hansen, LH & Hilger, KB 2015, 'Control of Electricity Loads in Future Electric Energy Systems'. in AJ Conejo, E Dahlquist & J Yan (eds), Handbook of Clean Energy Systems: Intelligent Energy Systems. vol. 4, Wiley.
- Halvgaard, RF, Vandenberghe, L, Poulsen, NK, Madsen, H & Jørgensen, JB 2016, Distributed Model Predictive Control for Smart Energy Systems IEEE Transactions on Smart Grid, vol 7, no. 3, pp. 1675-1682.
- Bacher, P, de Saint-Aubain, PA, Christiansen, LE & Madsen, H 2016, Non-parametric method for separating domestic hot water heating spikes and space heating Energy and Buildings, vol 130, pp. 107-112.

Some 'randomly picked' books on modeling





International Series in Operations Research & Management Science

Juan M. Morales - Antonio J. Conejo Henrik Madsen - Pierre Pinson Marco Zugno

Integrating Renewables in Electricity Markets

Operational Problems



🖄 Springer

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)