University of Salford MANCHESTER

DYNASTEE

Using Smart Meter Data to Assist in Energy Performance Measurement

Todays session will cover the following:

- **The need for measurement of energy in buildings (briefly)**
- **Smart meters, context, capabilities and limits**
- **Examples of Smart meter – energy performance work**
- **How do we get the data**
- **Final closing remarks**

1

DYNASTEE

Energy House Labs: What we do and why….

Dr Richard Fitton

Reader in Energy

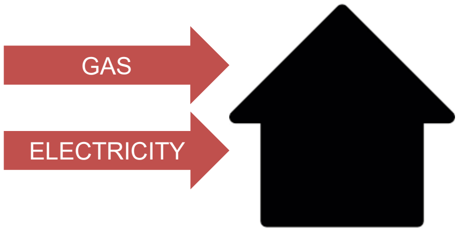
2

KEY ISSUES

- Energy consumption in the home and small commercial
- Changes to the energy system – decarbonisation, decentralisation and digitisation
- Everything from insulation, to heating systems, controls, smart meters and electric vehicles
- Domestic energy is getting complicated

OLD MODEL

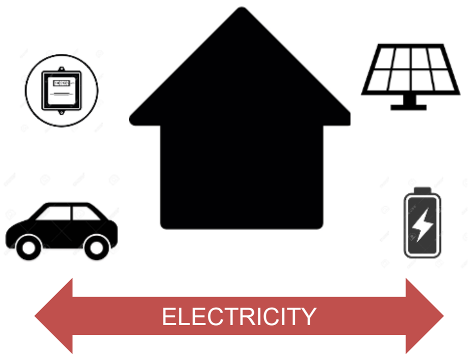
Simple model we were working to



SIMPLE CONSUMER

NEW MODEL

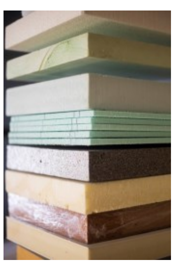
Emerging view of domestic energy



FACILITIES

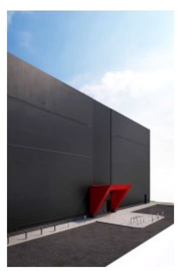
ENERGY HOUSE

THERMAL MEASUREMENT LABORATORY



SMART METERS SMART HOMES LAB

ENERGY HOUSE 2.0



DYNASTEE

ENERGY HOUSE 2

ENERGY HOUSE LABS

European Union
European Regional Development Fund

7

DYNASTEE

ENERGY HOUSE 2

ENERGY HOUSE LABS

European Union
European Regional Development Fund

8

PEOPLE

ENERGY HOUSE LABS

SCOPE

ENERGY HOUSE LABS

What do we do?

ENERGY HOUSE

Research:
- Retrofit of fabric (insulation air tightness etc)
- Smart controls
- Electric Vehicle Charging
- Building performance methodologies
- Large scale field trials

- Commercial testing of energy savings devices

THERMAL MEASUREMENT LABORATORY

Research:
- Ageing of insulation materials
- Recycled materials for use as insulation
- Insulation values of thatched properties

- Commercial testing of conductivity of insulation products

SMART METERS > SMART HOMES LAB

Research:
- Appliance disaggregation
- Smart meter/IOT linkage
- Data analytics with Field Trial team
- Innovative uses for SM data

University of Salford MANCHESTER

DYNASTEE

Context

Researchers and industry are quickly realising that the worlds of energy and environmental modelling and the real world often do not intersect. This has been shown in many studies globally. This is known as the "Performance Gap" (PG)

Country	Sample size (N)	Average Performance Gap	Country	Sample size (N)	Average Performance Gap	Reference
Canada	1	74%	Canada	1	74%	(Rouleau et al., 2018)
Germany	3400	30%	Germany	3400	30%	(Galvin, 2014)
United Kingdom	25	50%	United Kingdom	25	50%	(Johnston et al., 2015a)
Switzerland	50000	11%	Switzerland	50000	11%	(Cozza et al., 2020)
France			France			
Italy	6	45%	Italy	6	45%	(Ballarini & Corrado, 2009)

University of Salford MANCHESTER

DYNASTEE

Context

Performance gap can be found in either the positive or negative side of the modelled value.

- Many researchers have tried to state what may cause PG.
- The PG is actually caused by a number of different reasons.
- A typical reason could be the gap between a default value entered into an energy model that wasn't included in the actual building

Context

This can explain some issues, but as researchers we dig deeper:

- The way we model may not be perfect, but modelling should be accepted as a simplification of a building and not reality itself.
- Also many people will criticise a model after analysing results form a poorly measured /sensored measurement campaign, with consideration for uncertainty etc.

University of Salford MANCHESTER
DYNASTEE

How can smart meters help?

Historically. Some of the older members of the audience may remember these:

University of Salford MANCHESTER

DYNASTEE

How can smart meters help?

Historically.

In 2021 should we be gluing sensors to the front of gas and electric meters?

Plus, they fall off, occupants remove them, they need comms adding, and that is all expensive, and can be inaccurate.

So now we can do things a little “***smarter***” in some cases

University of Salford MANCHESTER

DYNASTEE

How can smart meters help?

So what is a smart meter?

In 2021 should we be gluing sensors to the front of gas and electric meters?

Plus, they fall off, occupants remove them, they need comms adding, and that is all expensive, and can be inaccurate.

So now we can do things a little “***smarter***” in some cases

In 2021 should we be gluing sensors to the front of gas and electric meters?

Plus, they fall off, occupants remove them, they need comms adding, and that is all expensive, and can be inaccurate.

So now we can do things a little “***smarter***” in some cases

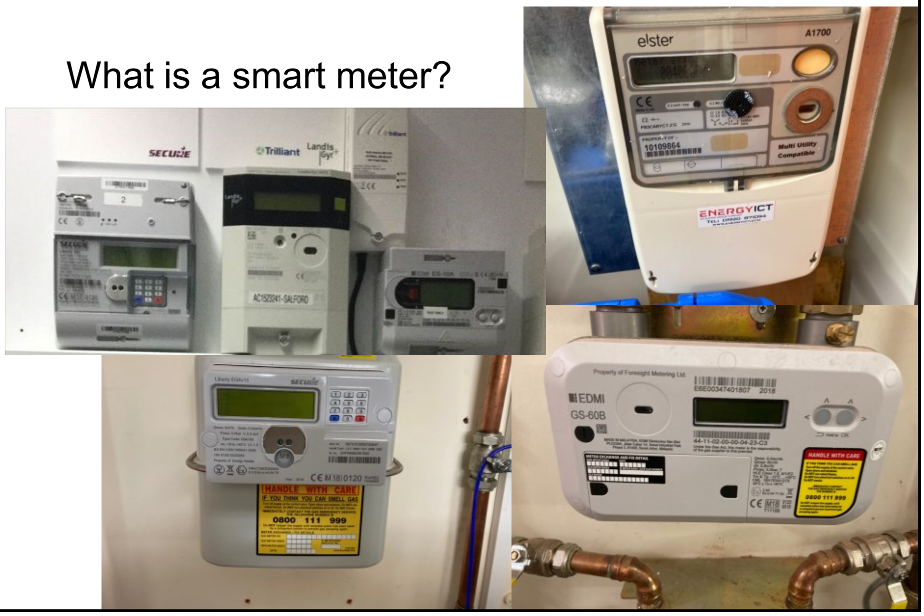
What is a smart meter?

No definition! / No harmonisation/ No common data patterns/ No standard comms!

European Smart Meter Association however do advise that smart meters should have the following characteristics:

- Automatic processing, transfer, management and utilisation of metering data
- Automatic management of meters
- 2 way data communication with meters
- Provides meaningful and timely consumption information to the relevant actors and their systems, including the energy consumer
- Supports services that improve the energy efficiency of the energy consumption and the energy system (generation, transmission, distribution and especially end use)
- Can be used on multiple utilities, such as water, gas and electric, generally one meter for each.

What is a smart meter?



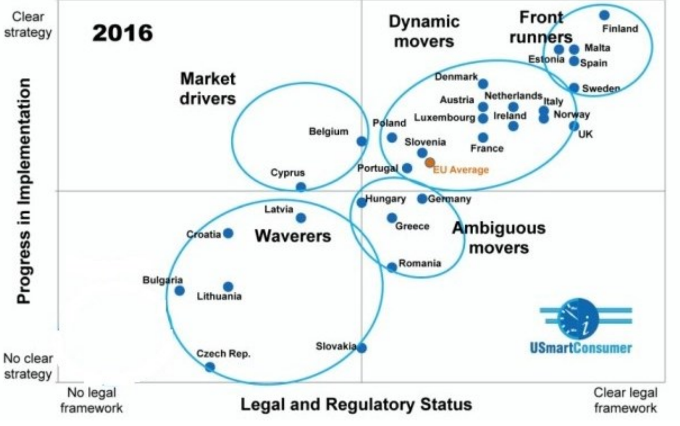
What is a smart meter?

There is also no harmonisation across the EU, either technically or even with regards to rollouts, although there is an aim:

It is an aim of the European Union (EU) to introduce smart meters across the union at a rapid rate with the aim of 72% coverage for electric and 40% for gas metering.

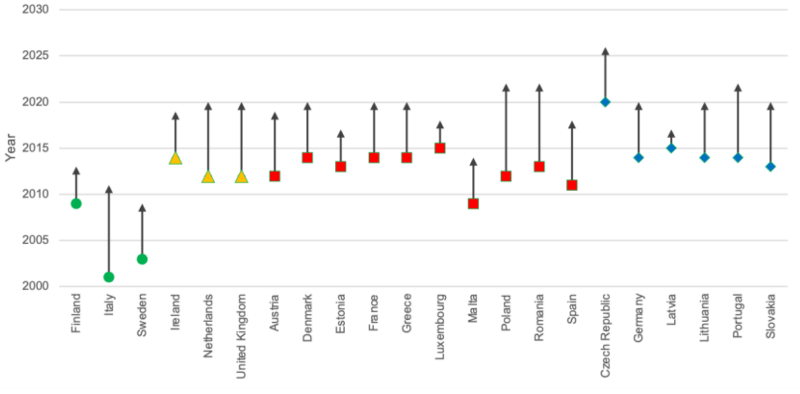
Rollout:

These figures are quite old but are the latest ones:



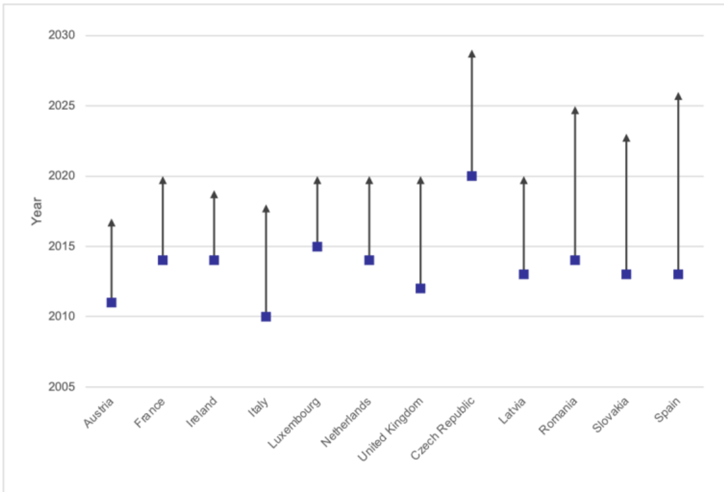
Rollout:

Electric



Rollout:

Gas



23

What can they do:

Another non harmonised issue!

- *Meter readings are taken over a range of 10 seconds to 60 minutes for electricity.*
- *Gas is less usually around 30 minutes and the device is remote from a power supply and is battery powered*

- *The key message is smart meters and the data offered is VERY country specific.*

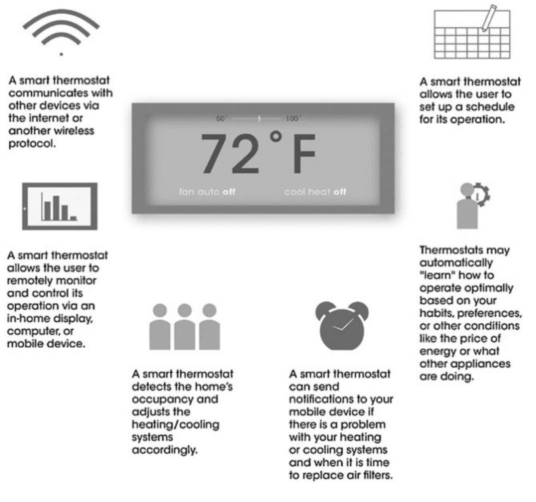
24

Can Smart Meters generate accurate energy ratings on their own, probably not!

Generally for energy efficiency works without going to detail, a source of inside and outside temperature is needed.

- *We can get smart meter data remotely so we need something similar*
- *IOT is our friend here, smart thermostats, boilers with comms, heat meters etc.*

Smart Devices:



Smart Devices:

- Actual sales figures are difficult to come by as these are commercially sensitive, but the trend is clear: Shin et al estimated that smart home technologies are currently present in 7.5% of homes globally with an annual market of $44.2 million in 2018 (Shin et al., 2018).
- The EU is a strong leader in this market place with a recent Berg Insight market research report claiming that at the end of 2017 the EU had around 22.5 million smart homes or 9.9% of households with France, Germany and the UK leading the market (Berg, 2018).

So if we presume that we can access the data (more of this later), then what can we do with that data:

A couple of high profile projects have recently worked with SM data, with well documented outputs. These are the latest ones:

- ***Annex 71 (pan EU)*** *https://bwk.kuleuven.be/bwf/projects/annex71*
- ***SMETERS (UK only)*** *https://www.gov.uk/guidance/smart-meter-enabled-thermal-efficiency-ratings-smeter-innovation-programme (data to be released October 2021)*

ANNEX 71 Final Meeting Youtube playlist contains many videos on this topic including, data collection, analysis, interpretation, policy implications and much more

This has just been released here

https://www.youtube.com/watch?v=h8nt_pEx36M&list=PL26pymJl0WS2e-GltP1eLa_bdcLOBVp0H&index=1

SMETER had an online workshop which is relevant to todays session. This is worth a view.

https://www.youtube.com/watch?v=kpwJaVek4Q0

The findings from these reports are interesting and state that this method of estimating energy perform well and have lots of scope for further work.

So how do we actually get this data?

This is highly country/ supplier specific. But lots of suppliers will have an interface to the data using a cloud platform.

This platform usually has an API:

*(API= **Application Programming Interface**, allows two applications to talk to each other)*

The Octopus has a number of possible data points

- Octopus Energy provides a REST API for customers and partner organisations to interact with our platform. Amongst other things, it provides functionality for:
- Browsing energy products, tariffs and their charges.
- Retrieving details about a UK electricity meter-point.
- Browsing the half-hourly consumption of an electricity or gas meter.
- Determining the grid-supply-point (GSP) for a UK postcode.
- Creating a quote.
- Creating a account.

University of Salford MANCHESTER

DYNASTEE

We will present an example as follows:

Apologies for UK centric angle, but this should still help.

Octopus Energy - (supplier)

Smart meter install - (Electric)

Data – energy consumption

Frequency - 30 mins

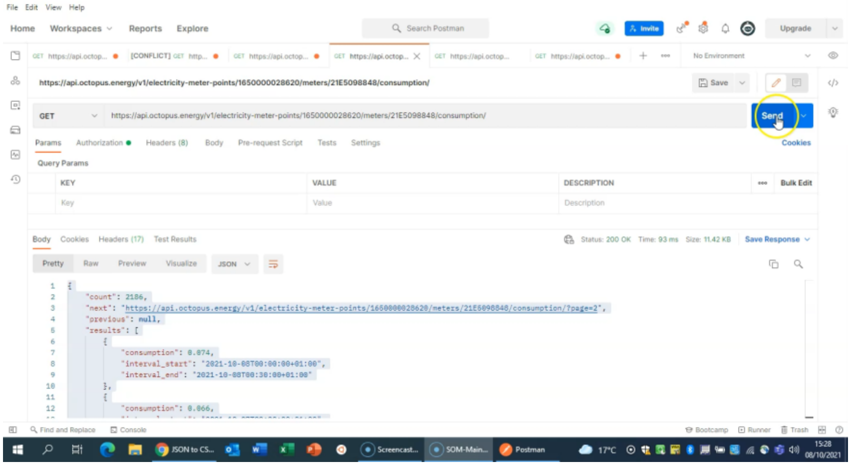
Domestic Installation

<u>Thanks to Dr I Paraskevas and A Sitmalidis for this presentation</u>

University of Salford MANCHESTER

DYNASTEE

Walkthrough



Walkthrough

Sample JSON files can be received by emailing R.fitton@Salford.ac.uk

Space Heating Data

A key component for any heat loss/ HTC estimation is the Delta component (Internal and external data)

Some easy to use (and working systems)

University of Salford MANCHESTER

DYNASTEE

Tado

https://www.openhab.org/addons/bindings/tado/

University of Salford MANCHESTER

DYNASTEE

Honeywell Evohome

https://developer.honeywellhome.com/content/getting-started-guide

Google Nest

https://developers.google.com/nest/device-access/api/thermostat

39

Air Source Heat Pumps

https://api.nibeuplink.com

SEE EXCEL SHEET

40

University of Salford MANCHESTER

DYNASTEE

No Energy rating estimation is possible without external conditions

Weather compensation sensors

Weather observations API

Forecasted weather API

BEMS Systems

University of Salford MANCHESTER

DYNASTEE

Data observation:

Of course data can have glitches, it can be correct or incorrect.

You should use some relevant mechanisms to check your data makes sense.

Just because something is smart does not make it right

Data often drops/has gaps etc, so you need to deal with this also.

Some excellent advice is found here https://iea-ebc.org/projects/project?AnnexID=58

University of Salford MANCHESTER

DYNASTEE

How can you start doing this:

Smart meter data is now fairly easy to achieve, however it is totally unique to the country and supplier, you are going to have to have to do some research.

Please note :Data and GDPR, this needs a significant amount of thought, people's life's are well documented through energy consumption and heating patterns and it should not be in the public domain.

It should never be linked to personal data such as addresses, peoples names etc, where it can be used for nefarious purposes.

43

University of Salford MANCHESTER

DYNASTEE

What else can we do:

This is just the start of an energy data revolution, more data will come on stream, EV, ASHP, batteries, Heat meters etc.

If we merge in other data sources we can extrapolate more sensing, if we know where the EV is we can map temp data etc.

Lets take a look, my journey home last night:

44

DYNASTEE DYNamic Analysis. Simulation and Testing applied to the Energy and Environmental performance of buildings

Free On-Line Training Webinars; 22 and 29 September, 6 and 13 October 2021

Dynamic Calculation Methods for Building Energy Performance Assessment

Technical University of Denmark DTU · GCU Glasgow Caledonian University · Universidad del País Vasco Euskal Herriko Unibertsitatea · Gobierno de España Ministerio de Economía y Competitividad · Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas · University of Salford Manchester

DYNASTEE

DYNASTEE

DYNASTEE

https://DYNASTEE.INFO website

Network for

- DYNamic
- Analysis
- Simulation and
- Testing of
- Energy and
- Environmental performance of buildings

DYNASTEE - OBJECTIVE

- Global leading network on dynamic testing and evaluation of Energy Performance in Buildings
- Consolidation of existing knowledge
- Bringing together academic, industry and governmental experts
  - on the **test environment and experimental setup** as well as on the **data analysis** and **performance prediction**.
- DYNASTEE - NoE: ST5 of IEA EBC Annex 71

TRAINING OBJECTIVES

Background to Course

The course has two main objectives:

- Train a dynamic methodology to assess the thermal performance of a building such as a wall, and a whole buildings' performance.
- Examine and understand the performance of nZEB and renewable energy technologies in built environment

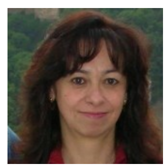
The approach to these will be a combination of **building physics, applied mathematics** and **statistical methods**

Overall Topic of Sessions

- Building physics to support the development of mathematical models for energy performance assessment.
- Knowledge of thermodynamic processes, heat transfer and the impact of solar radiation.
- Thermal conduction, convection, radiation and thermal mass.
- Using benchmark data for analysis
- Complexity of the physical process and how to translate the available information in mathematical models,
- Importance of model simplification of building physics represented by measured signals.
- Variability of the environments and the uncertainty of data
- Measured data and not-measured phenomena and how to build a mathematical model based on the available input.

5

The Experts



- María José Jiménez (CIEMAT, Spain),
- Irati Uriarte (UPV-EHU, Bilbao, Spain),
- Hans Bloem (INIVE-DYNASTEE, BE),
- Paul Baker (GCU, Glasgow, UK),
- Aitor Erkoreka (UPV-EHU, Bilbao, Spain)
- Peder Bacher (DTU, Lyngby, Denmark),
- Richard Fitton (University of Salford, UK)
- Luk Vandaele (INIVE-DYNASTEE, BE)

6

PRESENTATIONS

General

- Introduction to general approach of different analysis techniques used to perform the thermal characterisation for elements (walls, roofs etc) through to the whole building.

- Two software tools have been introduced that could be used during series of webinars; LORD, and CTSM-R. An easy exercise is available given with the correct result that may help you to build confidence in your analytics skills

- Introduction to measured data, specific sensors for buildings physics and energy performance and what is important to know about building physics, sensors and instruments

- The experimental set up and measurements at the Plataforma Solar de Almeria (PSA)has been presented, an explanation and demonstration of the data available have been given. Data series 16-17 have been presented.

PRESENTATIONS

Data that can be used be the participants *is available at the website dynastee.info;* zipped folder PSA_RRbox_DataSeries20

- An exercise that will allow of a study to be analysed with and without solar radiation.
- An introduction to dynamic analysis methods specifically LORD has been will provided
- A practical demonstration has been given of the software tool LORD on the PSA data series 16 and 17

- Introduction to discrete time and continuous time methods and how to use CTSM-R with statistical tools .
- Demonstration of the CTSM-R software applied to PSA data series 16,17.

- An introduction to the analysis of metering data, the specification and limitations of the data and analysis techniques.

CONCLUSION

"One needs a certain level of skill to perform well"

- Improve knowledge through a Training and Competition
- After >25 years DYNASTEE states:

Training make sense

CONCLUSION

- These webinars have been attended by 26 people
- Improve your knowledge and your skill through a Competition
- DYNASTEE investigates the organization of such a Dynamic Analysis Competition
- Could be whole building energy performance assessment based on metering data and/or real measurements

Future; 2022

Last year is atypical; the decision was made to postpone the complete Summer School for good reasons and substitute it by a series of webinars

However we are already planning the next summer school to take place in Almeria in Spain in 2022, this will be a full Summer School with classroom-based learning sessions and interactive sessions.

Future; 2022 and beyond

We will be using the forthcoming year to work on new topics for the summer school as follows:

- **Use of online data platforms such as weather API, renewable energy data**
- **Use of on-board systems such as connected thermostats**
- **Use of smart metering data for energy input**

Most countries now have access to at least most of these data, and some, all of it.

- The work and findings of IEA Annex 71 which focus on the data mentioned above to deem the energy performance of a dwelling. https://dynastee.info/new-iea-ebc-annex-71-building-energy-performance-assessment-based-on-in-situ-measurements/
- We will provide learning on not only the acquisition of this data using live API access to smart meter and controls, but the analytical tools to deem the energy performance.

DYNASTEE DYNamic Analysis, Simulation and Testing applied to the Energy and Environmental performance of buildings

Free On-Line Training Webinars; 22 and 29 September, 6 and 13 October 2021

THANKS TO

Webinar management

Maria Kapsalaki
(INIVE, BE)

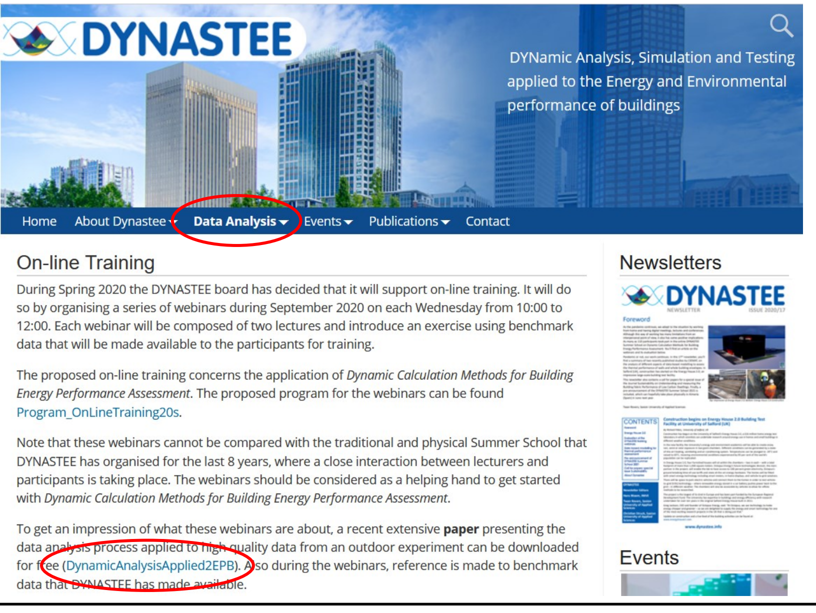
Valérie Leprince
(INIVE, BE)

Disclaimer: The sole responsibility for the content of presentations and information given orally during DYNASTEE webinars lies with the authors. It does not necessarily reflect the opinion of DYNASTEE. Neither DYNASTEE nor the authors are responsible for any use that may be made of information contained therein.

13

DYNASTEE

WWW.DYNASTEE.INFO



14

DYNASTEE

Introduction to measured data, instrumentation and sensors in relation to building physics and energy performance. What is important to know? QUESTIONS

Aitor Erkoreka
University of the Basque Country (UPV/EHU)

1 – QUESTIONS

DYNASTEE

Questions for Aitor Erkoreka

On the internal surface temperature measurement, how did you consider overheating on captors exposed to solar radiation ? Especially on captors on internal surface of glazing

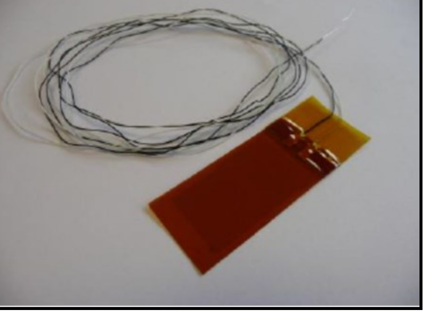
Considerations on temperature measurements: THERMORESISTANCE

The main advantages of the thermoresistances are:

- Accuracy
- Sensitivity

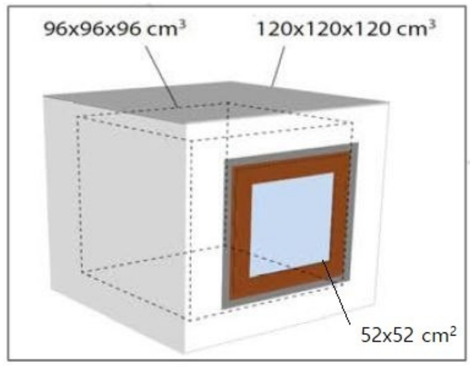
The main disadvantages of the thermoresistan

- Fragility
- Price (more expensive than thermoco



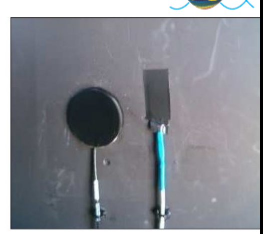
1 – QUESTIONS

DYNASTEE



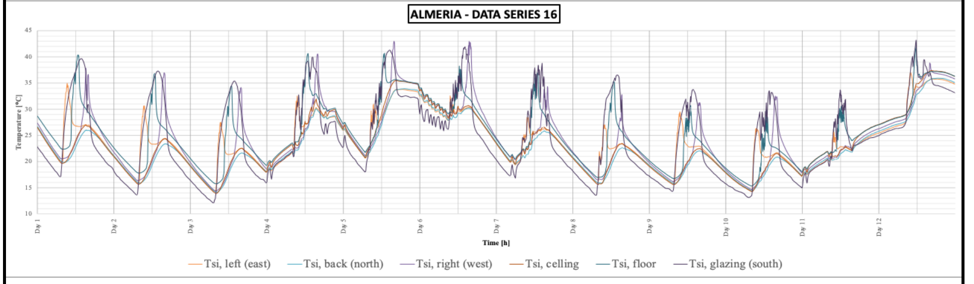
1 – QUESTIONS

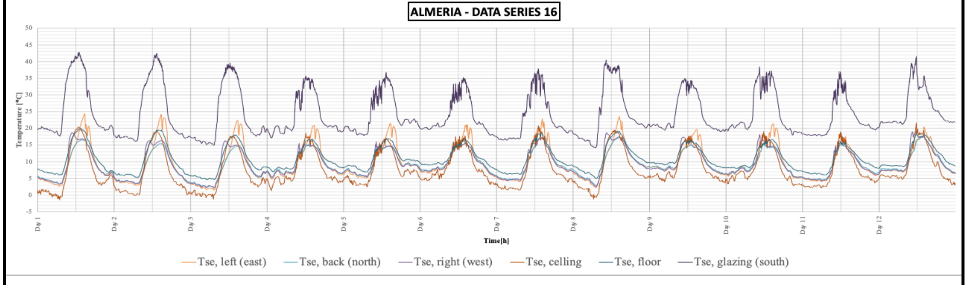
DYNASTEE



1 – QUESTIONS

DYNASTEE





Universidad del País Vasco Euskal Herriko Unibertsitatea

Introduction to ctsm-r

(Based on slides created by Rune Juhl)

Peder Bacher

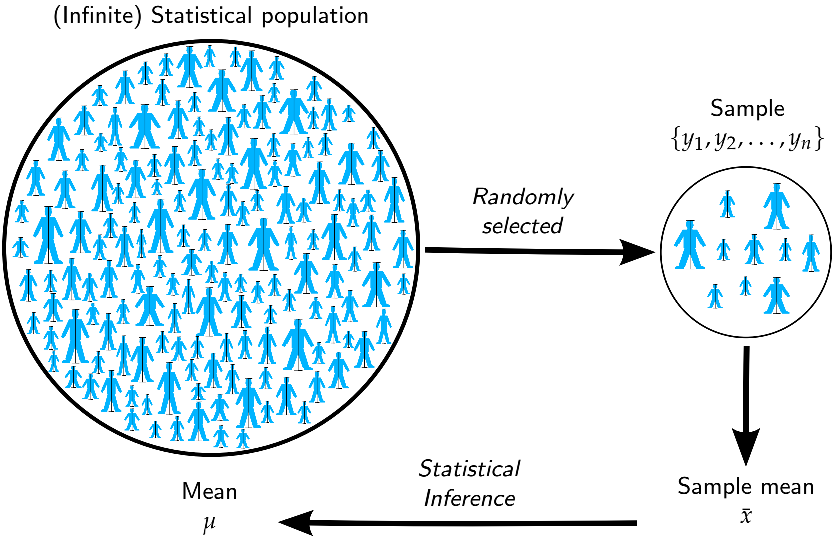
DTU Compute, Dynamical Systems
Building 303B, Room 010
DTU - Technical University of Denmark
2800 Lyngby – Denmark
e-mail: pbac@dtu.dk

Summer school 2021:

Time Series Analysis - with a focus on Modelling and Forecasting in Energy Systems

Overview

Population and sample



Parameter estimation with example

Simplest example: a constant model for the mean

- Model

$$Y_i = \mu + \varepsilon_i \quad \text{, where } \varepsilon_i \sim N(0, \sigma^2) \text{ and i.i.d.}$$

- i.i.d.: identically and independent distributed
- The parameters are: the mean μ and the standard error σ
- We take a sample of $n = 10$ observations

$$(y_1, y_2, \ldots, y_{10})$$

Likelihood

The likelihood is defined by the joint probability of the data

$$L(\mu, \sigma) \equiv p(y_1, y_2, \dots, y_{10} | \mu, \sigma)$$

Hence, it's a function of the two parameters (the sample is observed, so it is not varying).

Due to independence

$$= \prod_{i=1}^{10} p(y_i | \mu, \sigma)$$

We assume in our model that the error $\varepsilon_i = Y_i - \mu$ is normal distributed (Gaussian), so

$$p(y_i | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right) \tag{1}$$

Maximum likelihood estimation

Parameter estimation

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left(-\ln(L(\theta)\right)$$

where $\theta = (\mu, \sigma)$

Maximum likelihood estimation

Parameter estimation

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left(-\ln(L(\theta)\right)$$

where $\theta = (\mu, \sigma)$

Run the example in R

Likelihood for time correlated data

Given a sequence of measurements $\mathcal{Y}_N$

$$\begin{aligned} L(\theta) = p(\mathcal{Y}_N|\theta) &= p(y_N, y_{N-1}, \ldots, y_0|\theta) \\ &= \left(\prod_{k=1}^{N} p(y_k|\mathcal{Y}_{k-1}, \theta)\right) p(y_0|\theta) \end{aligned}$$

Parameter estimation

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left(-\ln(L(\theta))\right)$$

Likelihood for time correlated data

If Gaussian

$$\hat{y}_{k|k-1} = E[y_k|\mathcal{Y}_{k-1}, \theta]$$
$$R_{k|k-1} = V[y_k|\mathcal{Y}_{k-1}, \theta]$$
$$\varepsilon_k = y_k - \hat{y}_{k|k-1}$$

then the likelihood is

$$L(\theta) = \left(\prod_{k=1}^{N} \frac{\exp(-\frac{1}{2}\varepsilon_k^T R_{k|k-1}^{-1} \varepsilon_k)}{\sqrt{|R_{k|k-1}|}\sqrt{2\pi}^l} \right)$$

Maximised using quasi Newton

Kalman filter

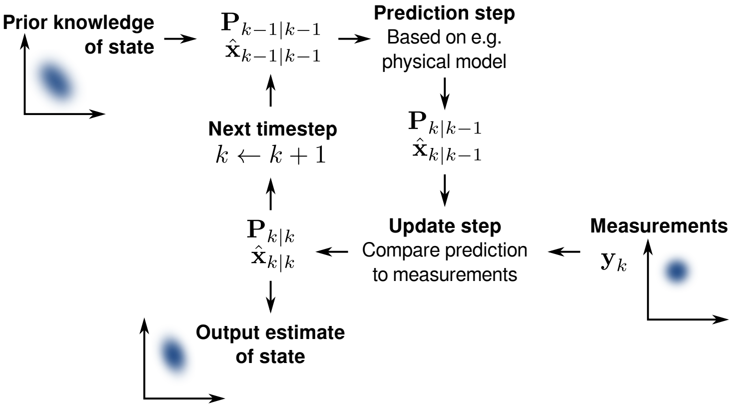


Figure: "Basic concept of Kalman filtering" by Petteri Aimonen. Wikipedia

Introduction to grey-box modelling and **ctsmr**

Grey-box modelling

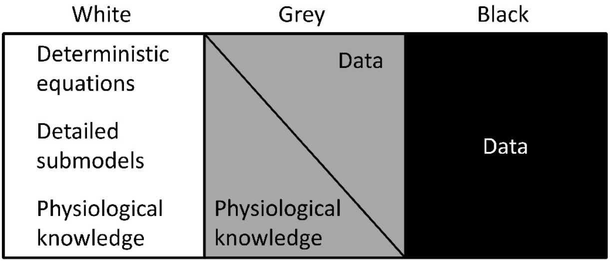


Figure: Ak et al. 2012

Grey-box modelling

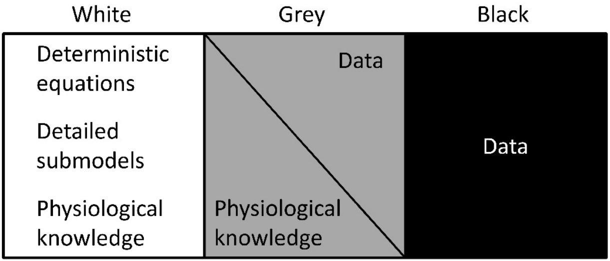


Figure: Ak et al. 2012

Bridges the gap between physical and statistical modelling.
THERE is a manual on `ctsm.info`

ctsmr

Continuous Time Stochastic Modelling in R

ctsmr

Continuous Time Stochastic Modelling in R

more correctly

Continuous-Discrete Time Stochastic Modelling in R

The model class

ctsmr implements a state space model with:

Continuous time stochastic differential system equations (SDE)

$$dX_t = f(X_t, U_t, t, \theta)dt + g(X_t, U_t, t, \theta)dB_t$$

Discrete time measurement equations

$$Y_{t_n} = h(X_{t_n}) + e_{t_n} \qquad e_{t_n} \in N(0, S(u_n, t_n, \theta))$$

- Underlying physics (system, states) modelled using continuous SDEs.
- Some (or all) states are observed in discrete time.

Features in CTSM-R

- Automatic classification (LTI or NL)
- Symbolic differentiation replaced AD (NL only)
  (Jacobians are computed faster.)
- Finite difference approximation of gradients are computed in parallel.
- Scriptable! Run multiple model during the night. Possible to use compute cluster.
- Direct access to plotting facilities from the R framework.

Loading the library

The R package is called **ctsmr**

R code

```
library(ctsmr)
```

Loading the library

The R package is called **ctsmr**

R code

```
library(ctsmr)
```

The model class is called **ctsm** - **C**ontinuous **T**ime **S**tochastic **M**odel.

R code

```
MyModel <- ctsm$new()
```

Class.. huh?

- **ctsm** is a ReferenceClass.
- The functions are methods attached to the class.

Class.. huh?

- **ctsm** is a ReferenceClass.
- The functions are methods attached to the class.

ctsm methods

Specifying the model:

- $addSystem()
- $addObs()
- $setVariance()
- $addInput()

Estimate and prediction:

- $setParameter()
- $setOptions()
- $estimate()

ctsmr defined functions

- predict
- simulate
- filter.ctsmr
- smooth.ctsmr

How to add System Equations

Use the `$addSystem` method to add a stochastic differential equation as a system equation.

R code

```
MyModel$addSystem( dX ~ (mu*X-F*X/V)*dt + sig11*dw1)
MyModel$addSystem( dS ~ (-mu*X/Y+F*(SF-S)/V) * dt + sig22*dw2)
MyModel$addSystem( dV ~ F*dt + sig33*dw3 )
```

Pay attention to the $\sim$. Do not use $=$.
The diffusion processes must be named `dw{n}`

How to add Observation Equations

Use the `$addObs` method to add a measurement/observation equation.

$$\boldsymbol{Y} = \begin{bmatrix} Y1 \\ Y2 \\ Y3 \end{bmatrix} = \begin{bmatrix} X \\ S \\ V \end{bmatrix}$$

R code

```
MyModel$addObs(y1 ~ X)
MyModel$addObs(y2 ~ S)
MyModel$addObs(y3 ~ V)
```

Pay attention to the $\sim$. Do not use $=$.

How to set the Variance structure of the Measurement Equations

The Example

Use the `$setVariance` method.

Example

```
MyModel$setVariance(y1y1 ~ s11)
```

How to set the Variance structure of the Measurement Equations

The Example

Use the `$setVariance` method.

Example

```
MyModel$setVariance(y1y1 ~ s11)
```

For `y1,,y2,y3` the size of the variance-covariance matrix is $3x3$.

$$S = \begin{bmatrix} \mathrm{s11} & & 0 \\ 0 & \mathrm{s22} & \\ & & \mathrm{s33} \end{bmatrix}$$

R code

```
MyModel$setVariance(y1y1 ~ s11)
MyModel$setVariance(y2 ~ s22)
MyModel$setVariance(y3^2 ~ s33)
```

Pay attention to the $\sim$. Do not use $=$.

Which variables are inputs?

Use the $addInput method to specify which variable is an input and not a parameter.

R code

```
MyModel$addInput(F)
```

How to specify initial values, boundaries and prior standard deviance (for MAP)?

Use the `$setParameter` method.

R code

```
MyModel$setParameter(X = c(init=1,lb=0,ub=2),
                      S0 = c(0.25,0,1))
MyModel$setParameter(V0 = c(1,lower=0,upperbound=2))
```

Pay attention to the $=$. Do not use $\sim$.

- Quite flexible.
- Named numbers (e.g. `init=3`) are processed first.
- Initial state values (e.g. X_0) can be named `X0` or `X`.
- `MyModel$ParameterValues` contains the parsed values.

How to change filtering and numerical optimisation options (advanced)?

Use the `$setOptions` method to change the options found in `MyModel$options`.

Specify the data

`ctsm` expects a data.frame containing time and all inputs and outputs.

Example

```
MyData <- data.frame(t = c(1,2,3), F = c(4,3,2), Y1 = c(7,6,5), Y2 =
...)
```

Multiple independent datasets can be given as a list of data.frames.

Example

```
AllMyData <- list(MyData1, MyData2, MyData2, ...)
```

Estimate the parameters

To estimate the parameters run:

```
fit <- MyModel$estimate(data = MyData)
```

Parameter inference

Like `lm()` use `summary()` on the fit for additional information.

- Parameter estimates alone:
  `fit`
- + standard deviance, t-statistics and p-values:
  `summary(fit)`
- + correlation of parameter estimates:
  `summary(fit, correlation=TRUE)`
- + additional information $\left(\frac{dF}{d\theta}, \frac{dPen}{d\theta}\right)$:
  `summary(fit, extended=TRUE)`

How to get k-step predictions

Use the `predict` function.

Usage

```
one.step.prediction <- predict(fit)
```

Available options:

- `n.ahead` number of steps ahead to predict.
- `newdata` to predict using a new dataset.

Diagnostics

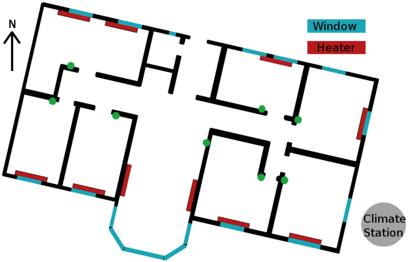
- k-step predictions `predict`
- filtered states `filter.ctsmr`
- smoothed states `smooth.ctsmr`
- simulations `simulate`
- likelihood ratio tests

Example: Selecting **a suitable grey-box model** for the heat dynamics of a building

Test case: One floored 120 m^2 building

Objective

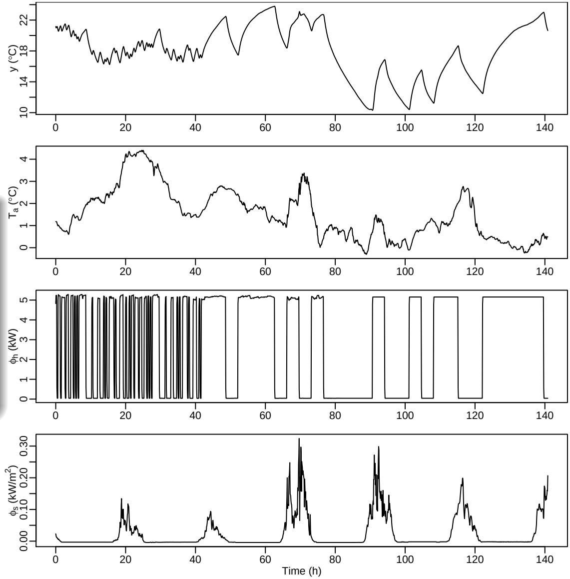
Find the best model describing the heat dynamics of this building



Data

Measurements of:

- y_t Indoor air temperature
- T_a Ambient temperature
- Φ_h Heat input
- Φ_s Global irradiance



Two big challenges when modelling with data

- **Model selection:** How to decide which model is most appropriate to use?
  We are looking for a model which gives us un-biased estimates of physical parameters of the system. This requires that the applied model is neither too simple nor too complex

- **Model validation:** How to validate the performance of a dynamical model?
  We need to asses if the applied model fulfill assumptions of white-noise errors, i.e. that the errors show no lag-dependence

Model selection

Likelihood ratio test: Test for model expansion

Say we have a model and like to find out if an expanded version will give a significantly better description of data

i.e. give an answer to: Should we use the expanded model instead of the one we have?

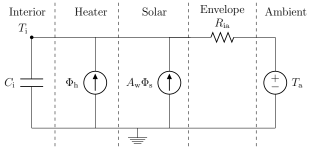
The likelihood ratio test

$$\lambda(\boldsymbol{y}) = \frac{L_{\text{sub}}(\hat{\boldsymbol{\theta}}_{\textbf{mle,sub}})}{L(\hat{\boldsymbol{\theta}}_{\textbf{mle}})}$$

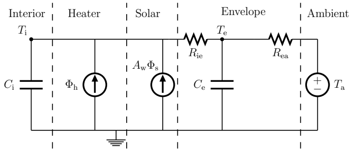
can be applied to test for significant improvement of the expanded model (with maximum likelihood $L_{\text{sub}}(\hat{\boldsymbol{\theta}}_{\textbf{mle,sub}})$) over the sub-model (with maximum likelihood $L(\hat{\boldsymbol{\theta}}_{\textbf{mle}})$)

Test for expansion

Simplest model

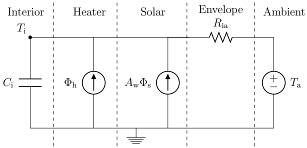


First extension: building envelope part (*TiTe*)

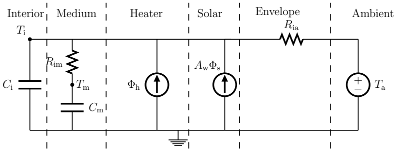


Test for expansion

Simplest model

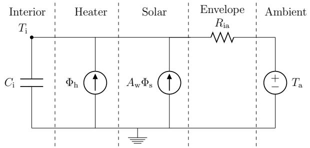


First extension: indoor medium part (*TiTm*)

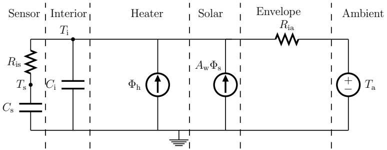


Test for expansion

Simplest model

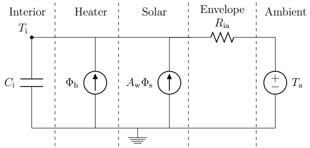


First extension: sensor part (*TiTs*)

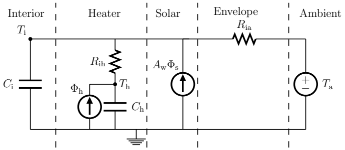


Test for expansion

Simplest model

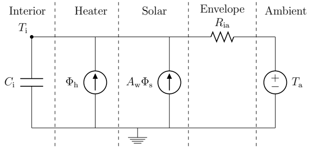


First extension: heater part (*TiTh*)



Test for expansion

Simplest model



First extension: Which one??

TiTe, *TiTm*, *TiTs*, or *TiTh* ?

Log-likelihoods

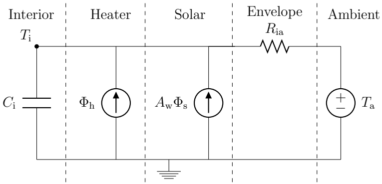
Simplest	*Ti*			
$l(\theta; \mathcal{Y}_N)$	**2482.6**			
m	6			
Expanded	*TiTe*	*TiTm*	*TiTs*	*TiTh*
$l(\theta; \mathcal{Y}_N)$	3628.0	3639.4	3884.4	**3911.1**
m	10	10	10	10

Likelihood-ratio test

Sub-model	**Model**	$m - r$	p-value
Ti	*TiTh*	4	$< 10^{-16}$

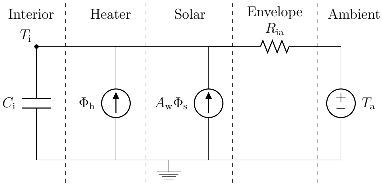
Identify the best physical model for the data

Simplest model

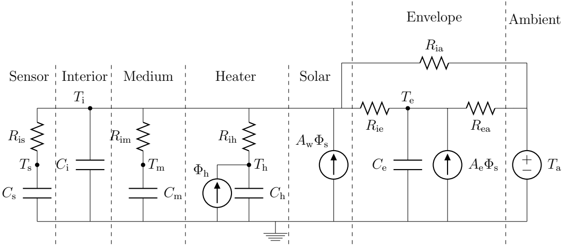


Identify the best physical model for the data

Simplest model

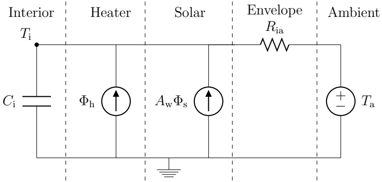


Most complex model applied



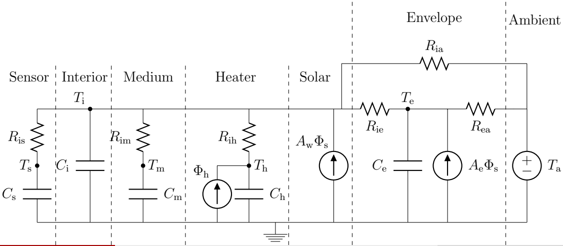
Identify the best physical model for the data

Simplest model



The best model for the given data is probably in between

Most complex model applied



Iteration	Models			
Start	*Ti*			
$l(\theta; \mathcal{Y}_N)$	**2482.6**			
m	6			
1	*TiTe*	*TiTm*	*TiTs*	*TiTh*
	3628.0	3639.4	3884.4	**3911.1**
	10	10	10	10
2	*TiThTs*	*TiTmTh*	*TiTeTh*	
	4017.0	5513.1	**5517.1**	
	14	14	14	
3	*TiTeThRia*	*TiTeThAe*	*TiTmTeTh*	*TiTeThTs*
	5517.3	5520.5	5534.5	**5612.4**
	15	15	18	18
4	*TiTeThTsRia*	*TiTmTeThTs*	*TiTeThTsAe*	
	5612.5	5612.9	**5614.6**	
	19	22	19	
5	*TiTmTeThTsAe*	*TiTeThTsAeRia*		
	5614.6	**5614.7**		
	23	20		

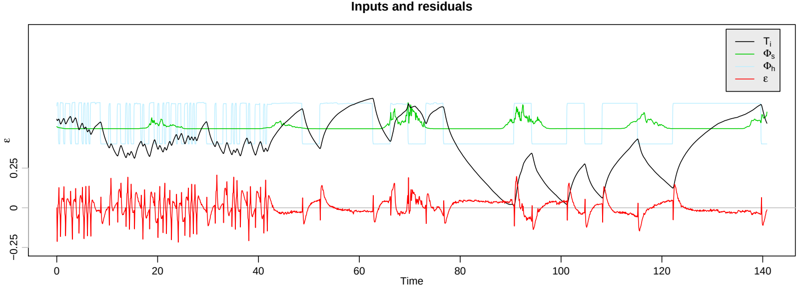
Iteration	Sub-model	Model	$m-r$	$-2log(\lambda(y))$	p-value
1	*Ti*	*TiTh*	4	4121	$< 10^{-16}$
2	*TiTh*	*TiTeTh*	4	4634	$< 10^{-16}$
3	*TiTeTh*	*TiTeThTs*	4	274	$< 10^{-16}$
4	*TiTeThTs*	*TiTeThTsAe*	1	6.4	0.011
5	*TiTeThTsAe*	*TiTeThTsAeRia*	1	0.17	0.68

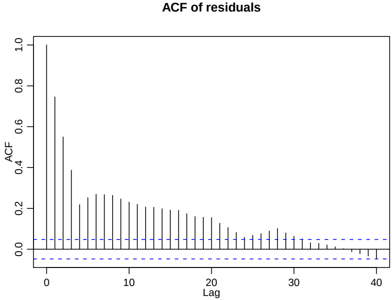
Model validation

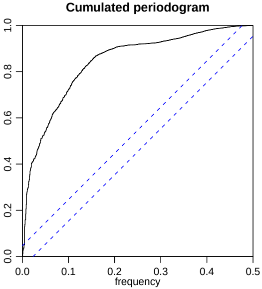
How can the performance of a dynamical model be evaluated?

- We assume that the residuals are i.i.d and normal
- Auto-Correlation Function (ACF) and Cumulated Periodogram (CP) of the errors are the basic tools
- Time series plots of the inputs, outputs, and the errors are valuable for pointing out model deficiencies

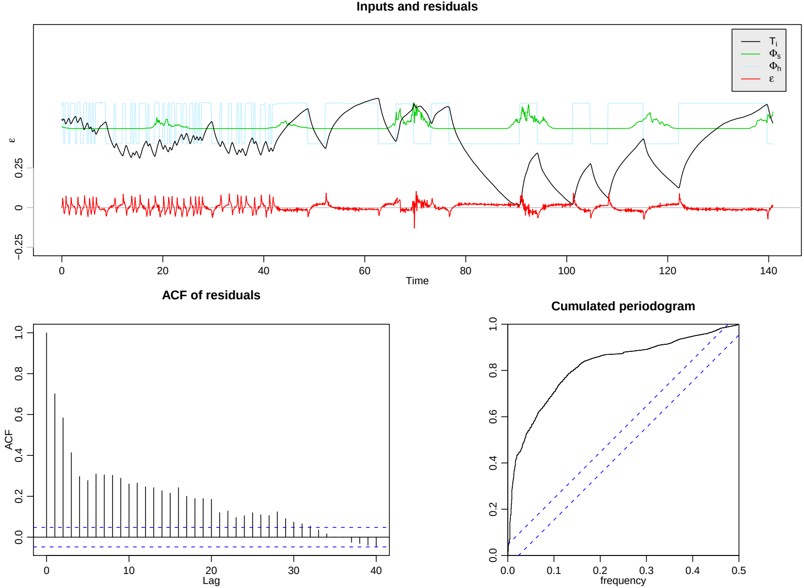
Evaluate the simplest model



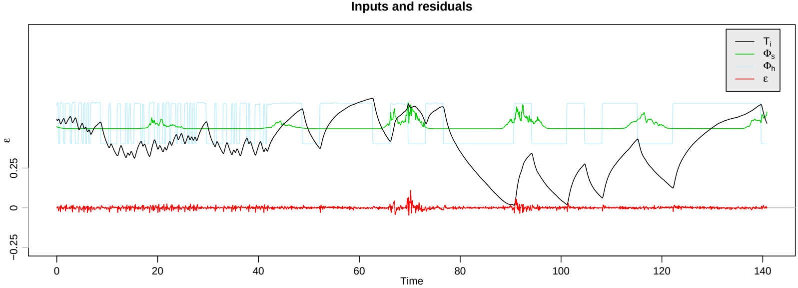


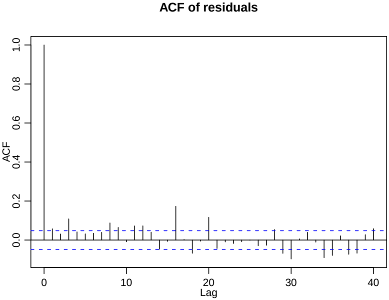


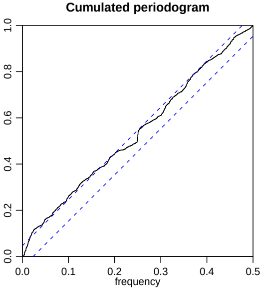
Evaluate the model selected in step one



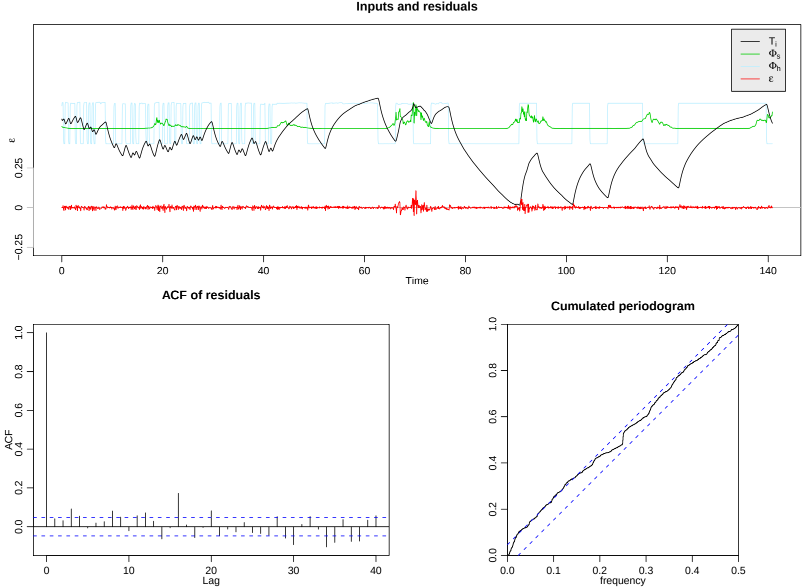
Evaluate the model selected in step two



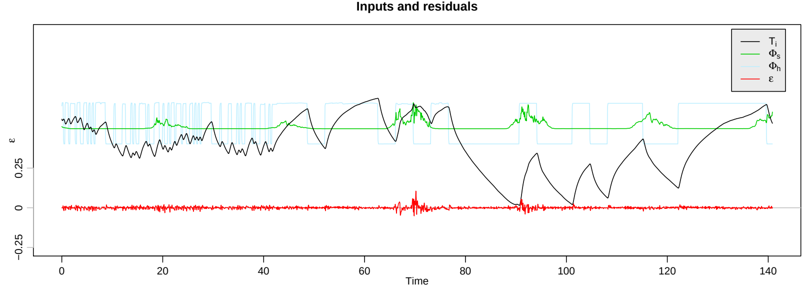


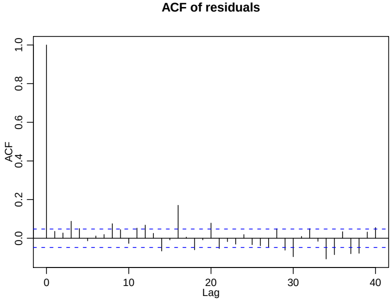


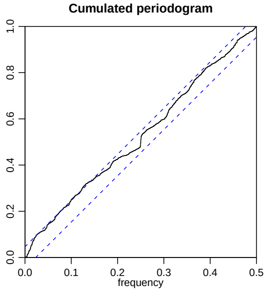
Evaluate the model selected in step three



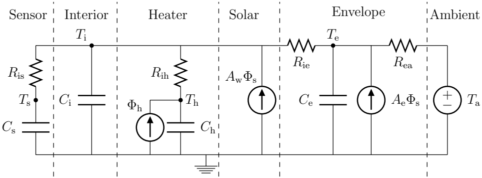
Evaluate the selected model in step four



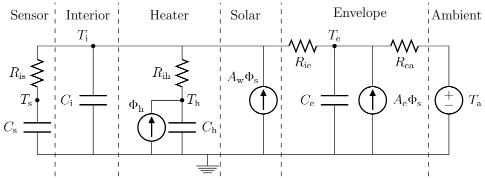




Selected model



Selected model



Estimated parameters

$\hat{C}_i$	0.0928	(kWh/°C)
$\hat{C}_e$	3.32	-
$\hat{C}_h$	0.889	-
$\hat{C}_s$	0.0549	-
$\hat{R}_{ie}$	0.897	(circC/kW)
$\hat{R}_{ea}$	4.38	
$\hat{R}_{ih}$	0.146	-
$\hat{R}_{is}$	1.89	-
$\hat{A}_w$	5.75	(m^2)
$\hat{A}_e$	3.87	-

Estimated time constants

$\hat{\tau}_1$	0.0102	hours
$\hat{\tau}_2$	0.105	-
$\hat{\tau}_3$	0.788	-
$\hat{\tau}_4$	19.3	-

Conclusions

- Applied Grey-box modelling, where a combination of *prior physical knowledge* and *data-driven modelling* is utilized
- Using a forward selection procedure with likelihood-ratio tests a suitable physical model is found
- The ability of the selected models to describe the heat dynamics are evaluated with the ACF, CP, and time series plots

Identifiability

Identifiability

Model identifiability is important for estimation in general (less important for prediction, very important for parameter interpretation).

There are two aspects of identifiability:

- **Structural identifiability:** the parameters in the model can never be estimated due to the structure of the model. Depends only on the model.
- **Practical identifiability:** there is not enough information in the data available to estimate the parameters in the model. Depends both on the model and the data.

Structural identifiability

State space model (innovation form)

$$\frac{d\hat{X}(t)}{dt} = A\hat{X}(t) + BU(t) + K\epsilon(t)$$
$$Y(t) = C\hat{X}(t) + DU(t) + \epsilon(t)$$

Apply the bilateral Laplace transformation (and after some voodoo)

$$\begin{aligned} Y(s) &= C(sI-A)^{-1}BU(s) + C(sI-A)^{-1}K\epsilon(s) + DU(s) + \epsilon(s) \\ &= \left(C(sI-A)^{-1}B + D\right)U(s) + \left(C(sI-A)^{-1}K + I\right)\epsilon(s) \end{aligned}$$

Focus on the input related transfer function

$$H_i(s) = C(sI-A)^{-1}B + D \tag{2}$$

Analyse the identifiability of an SDE model of a Wall

A lumped RC model of the wall

$$dT_w = \frac{1}{C_w}\left(\frac{T_a - T_w}{R_{aw}} + \frac{T_i - T_w}{R_{wi}}\right)dt + d\omega_1(t)$$

$$dT_i = \frac{1}{C_i}\left(\frac{T_w - T_i}{R_{wi}}\right)dt + d\omega_2(t)$$

$$y_{t_k} = Ti_{t_k} + \sigma_{t_k}$$

Transfer function

Apply equation **??** to obtain the input transfer function

$$H_{input}(s) = \frac{\frac{1}{C_i\, C_w\, R_{aw}\, R_{wi}}}{s^2 + \frac{R_{aw}\, C_i + C_i\, R_{wi} + R_{aw}\, C_w}{C_i\, C_w\, R_{aw}\, R_{wi}} \cdot s + \frac{1}{C_i\, C_w\, R_{aw}\, R_{wi}}}$$

Transfer function

Apply equation **??** to obtain the input transfer function

$$H_{input}(s) = \frac{\frac{1}{C_i\, C_w\, R_{aw}\, R_{wi}}}{s^2 + \frac{R_{aw}\, C_i + C_i\, R_{wi} + R_{aw}\, C_w}{C_i\, C_w\, R_{aw}\, R_{wi}} \cdot s + \frac{1}{C_i\, C_w\, R_{aw}\, R_{wi}}}$$

And compare it to

$$H(s) = \frac{b_0}{s^2 + a1 \cdot s + a0}$$

Transfer function

Apply equation **??** to obtain the input transfer function

$$H_{input}(s) = \frac{\frac{1}{C_i\,C_w\,R_{aw}\,R_{wi}}}{s^2 + \frac{R_{aw}\,C_i + C_i\,R_{wi} + R_{aw}\,C_w}{C_i\,C_w\,R_{aw}\,R_{wi}} \cdot s + \frac{1}{C_i\,C_w\,R_{aw}\,R_{wi}}}$$

And compare it to

$$H(s) = \frac{b_0}{s^2 + a1 \cdot s + a0}$$

Only two independent equations

$$a_0 = \frac{1}{C_i\,C_w\,R_{aw}\,R_{wi}}$$
$$a_1 = \frac{R_{aw}\,C_i + C_i\,R_{wi} + R_{aw}\,C_w}{C_i\,C_w\,R_{aw}\,R_{wi}}$$

Fit all four parameters?

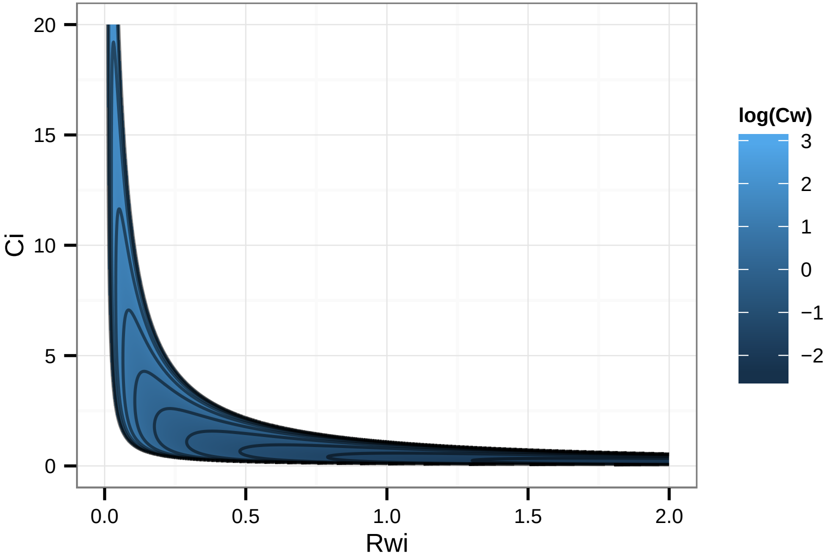
Solve two equations for four parameters.

$$
\begin{aligned}
Ci &= Ci \\
Rwi &= Rwi \\
Cw &= -\frac{C_i}{{C_i}^2{R_{wi}}^2 a_0 - a_1\, C_i\, R_{wi} + 1} \\
R_{aw} &= -\frac{{C_i}^2{R_{wi}}^2 a_0 - a_1\, C_i\, R_{wi} + 1}{{C_i}^2 R_{wi}\, a_0}
\end{aligned}
$$

Note: a_0 and a_1 are known when simulating data.

C_w is a function of other parameters

Below is the feasible C_w parameters: $C_w > 0$

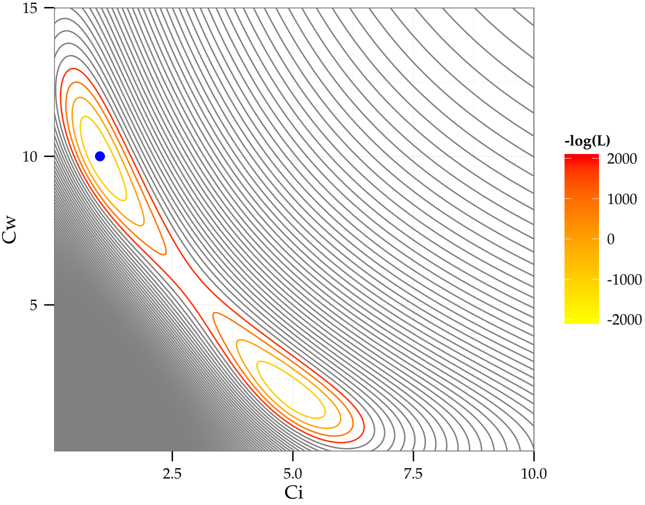


Estimate two parameters

We can estimate two.. So try fixing R_{wi} and R_{aw}

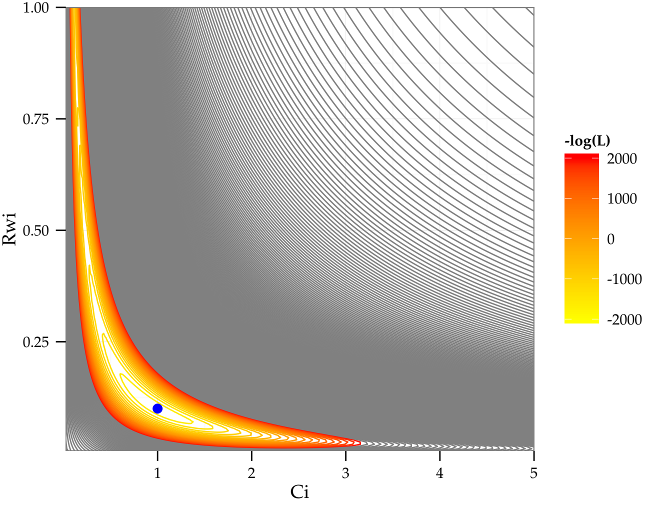
Estimate two parameters

We can estimate two.. So try fixing R_{wi} and R_{aw}



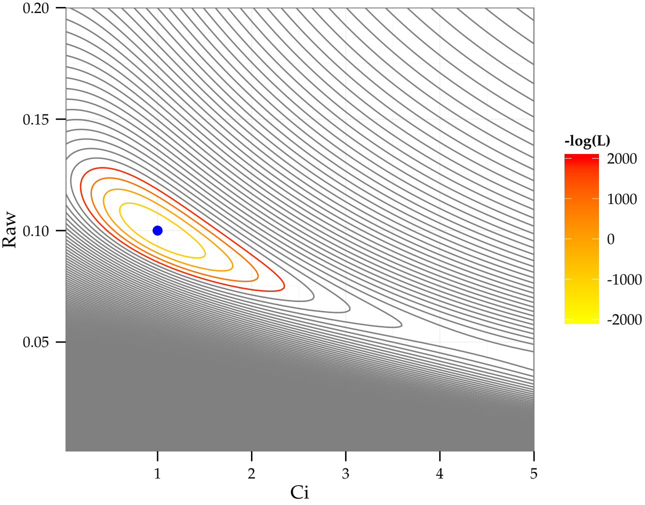
Estimate two parameters

We can estimate two.. So try fixing C_w and R_{aw}



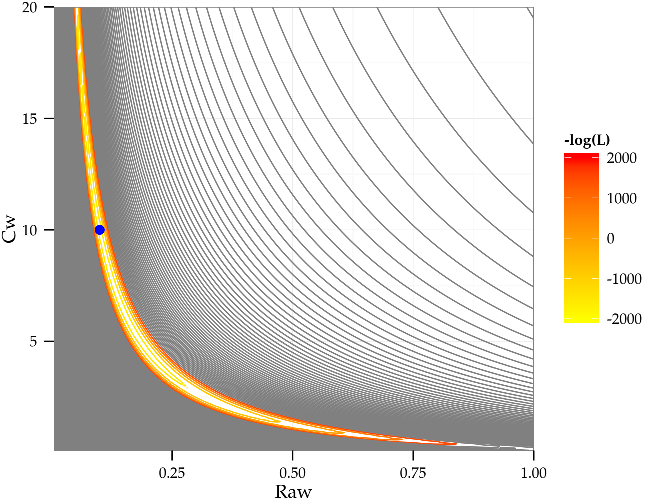
Estimate two parameters

We can estimate two.. So try fixing C_w and R_{wi}



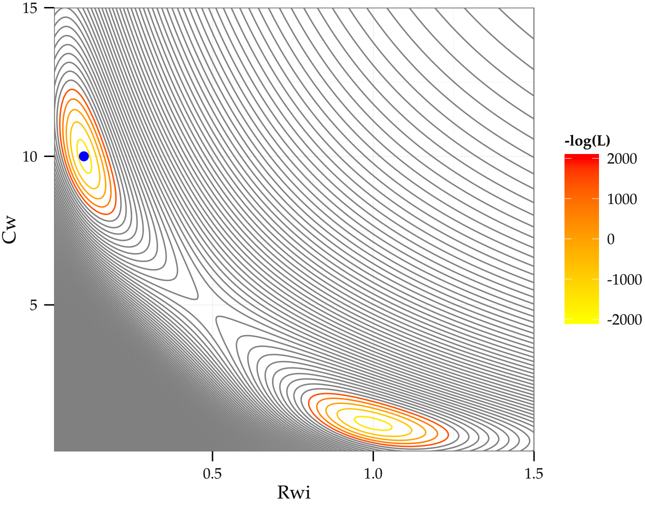
Estimate two parameters

We can estimate two.. So try fixing R_{wi} and C_i



Estimate two parameters

We can estimate two.. So try fixing R_{aw} and C_i



Estimate two parameters

We can estimate two.. So try fixing C_i and C_w