Computer Sciences Faculty
University of Murcia

# Data driven modeling for energy consumption prediction in smart buildings

Big Data 2017. BDA-IoT Workshop

Aurora González-Vidal, Alfonso P. Ramallo-González, Fernando Terroso-Sáenz, Antonio F. Skarmeta
{aurora.gonzalez2, alfonsop.ramallo, fterroso, skarmeta}@um.es

# Content

# Introduction

# Objective



- To develop a methodology in order to generate predictive models related to the **energy efficiency** in smart buildings.
- Application: evaluation (testing the effectiveness) of an efficiency plan (EM&V)
- Comparison of two approaches: black box (data driven) and grey box

# Smart environments

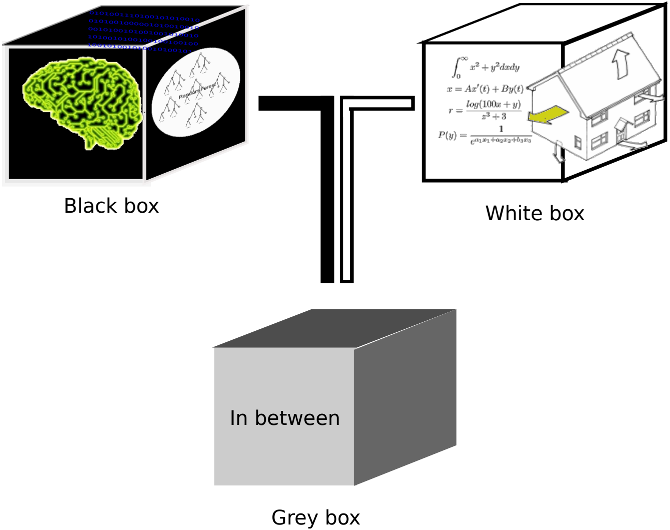


The **Internet of Things** (IoT) has provided a great scenario where large amounts of data can be collected and analysed using **big data analytic** techniques allowing the emergence of:

- Smart cities,
- Smart buildings

# Related work

# Related work



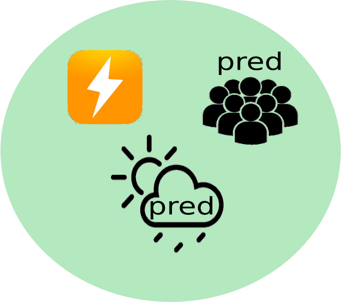
# Methodology

# Inputs

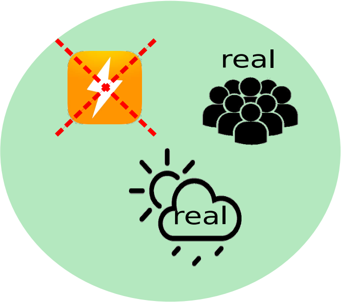


The considered inputs are goal dependent

Forecast future




**Quantify savings**

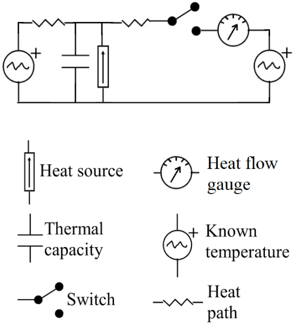


# Grey box



## Thermal network model

### Resistor-capacitor network




### State-space model

$$\begin{cases} x'(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases}$$

# Black box



Our process includes:

- Cleaning and description of the data
- **Model definition:** time series as input
- Preprocessing
- **Validation and training**
  1. Support Vector Regressor (SVR)
  2. Regression Forest (RF)
  3. Extreme Gradient Boosting (XGB)
- Evaluation

# Baseline black box



- Time of Week and Temperature (TWT): time of the week effect + piecewise continuous temperature effect
    - Accuracy
    - Low complexity
    - Low computational cost

| $T$ | $T_{c,1}$ | $T_{c,2}$ | $T_{c,3}$ | $T_{c,4}$ | $T_{c,5}$ | $T_{c,6}$ |
|---|---|---|---|---|---|---|
| 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| 18 | 10 | 8 | 0 | 0 | 0 | 0 |
| 32 | 10 | 10 | 10 | 2 | 0 | 0 |
| 47 | 10 | 10 | 10 | 10 | 7 | 0 |
| 58 | 10 | 10 | 10 | 10 | 10 | 8 |

$$load(i) = \alpha_i + \sum_{j=1}^{6} \beta_j T_{c,j}(i)$$

- Gaussian Process modelling
    - High flexibility: it uses the covariance matrix rather than the algebraic structure of the input–output
    - Bayesian setting: quantify uncertainty

# Metrics



- Model accuracy

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \bar{y}_i}{y_i}\right| \times 100,$$

$$CVRMSE = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}}{\bar{y}} \times 100,$$

- Saving metrics
  Whole building metering

# Use case

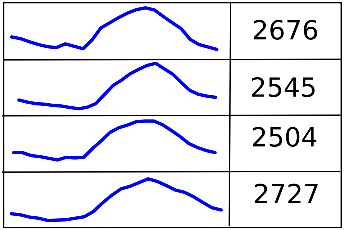
# Building and data



## Chemistry Faculty

## 1 year of time series

(hourly temperature in a day) —> (Daily consumption Wh)



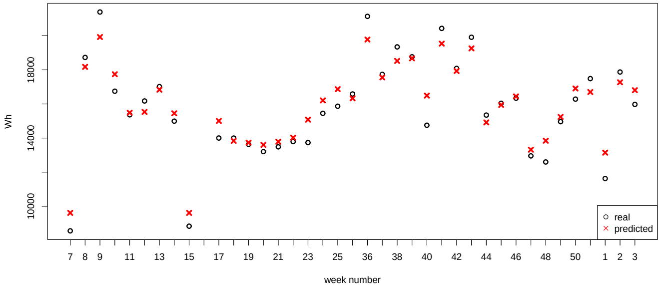
# Table of results



| | | Models | | | | | |
|---|---|---|---|---|---|---|---|
| | | SVR | **RF** | XGB | TWT | Gauss | Grey |
| Daily | CVRMSE | 12.4 | **9** | 11 | 14.9 | 17.45 | 33.57 |
| | MAPE | 7.2 | **6** | 7.3 | 12.3 | 15.01 | 43.02 |
| Weekly | CVRMSE | 6.4 | **5** | 6.2 | 11.1 | 16.3 | 19.53 |
| | MAPE | 5.2 | **4.5** | 5.5 | 9.4 | 12.3 | 15.48 |

# Weekly predictions graph

## Random Forest



# Discussion and conclusions

# Discussion and conclusions



- Black box models outperform the rest
- Behavioural patterns are out of the scope of grey-box models
- The use of time series for energy prediction compared to the use of instantaneous measurements provides a better result.

# Future Work



- Using more frequent measurements and comparing
- Applying feature transformation to the series
- Use a transfer learning approach for scaling the deployment of EM&V