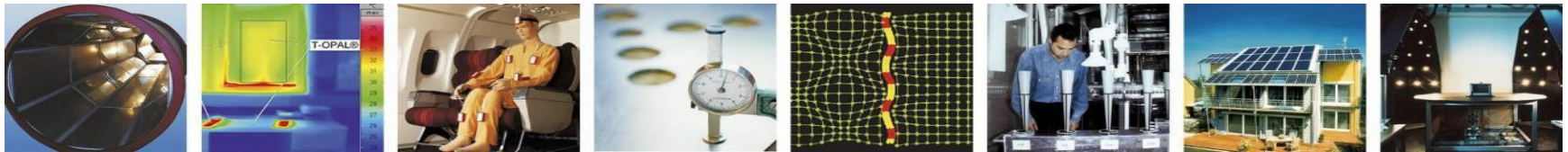

Modelling and prediction of buildings energy consumption using Machine Learning techniques

Georgios Kontes and Dimitrios Rovas

Integration of Sustainable Energy Conference (iSEneC 2016)

11 - 12 July 2016, Nuremberg, Germany

Auf Wissen bauen



Contributors

Georgios Kontes

*Fraunhofer Institute for Building Physics & Technische Hochschule Nürnberg
Georg Simon Ohm*

Dimitrios Rovas

Faculty of the Built Environment, University College London

Acknowledgements



Motivation

Fault Detection and Diagnosis (FDD):

- Detect and report “unusual” energy consumption of:
 - Specific parts of the Heating, Ventilation and Air Conditioning (HVAC) system;
 - Whole building.

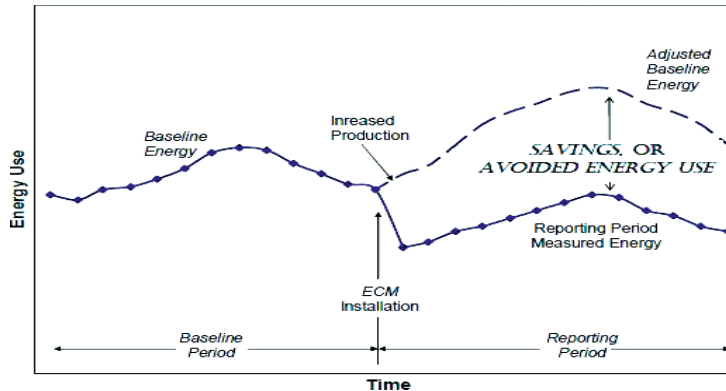
Post-retrofit evaluation:

- Apply an Energy Conservation Measure (ECM) and evaluate the achieved energy savings – performance-based contracts from Energy Services Companies (ESCOs).

Energy consumption prediction

Baselining – Reference Energy Consumption is constructed using a model;

- Can be used to detect faulty operation (e.g. excessive energy consumption during unoccupied periods);
- Can be used as a baseline model for a **M&V process**.



- IPMVP (International Performance Measurement and Verification Protocol): Concepts and Options for Determining Energy and Water Savings. vol. 1, Technical Report

Energy consumption prediction

Baselining – Reference Energy Consumption is constructed using a model;

- Model types:
 - First principles models (using detailed thermal simulation software);
 - Grey-box models (e.g. RC networks);
 - Black-box models:
 - Support Vector Regression (SVR);
 - Neural Networks (NN);
 - ...

Energy consumption prediction

Baselining – Reference Energy Consumption is constructed using a model;

- Model types:

- First principles models (using detailed thermal simulation software);

- Grey-box models (e.g. RC networks);

- Black-box models

- Support Vector Regression (SVR);

- Neural Networks (NN);

- ...

Example Application – Data used

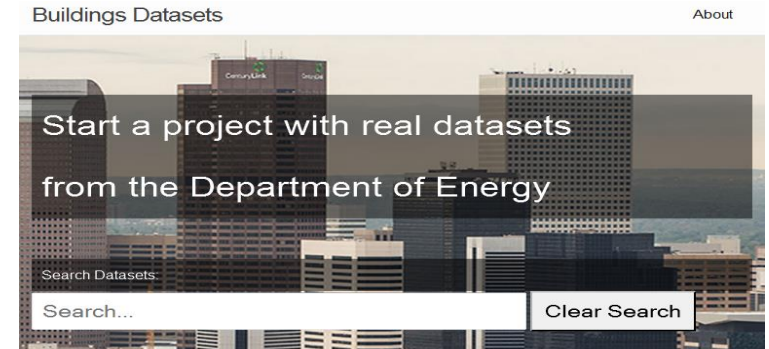
Data available online (Office Building #7 dataset)

Baseline and post-retrofit hourly data:

- Energy Consumption (kWh);
- Outside Air Temperature (OAT)

No knowledge on:

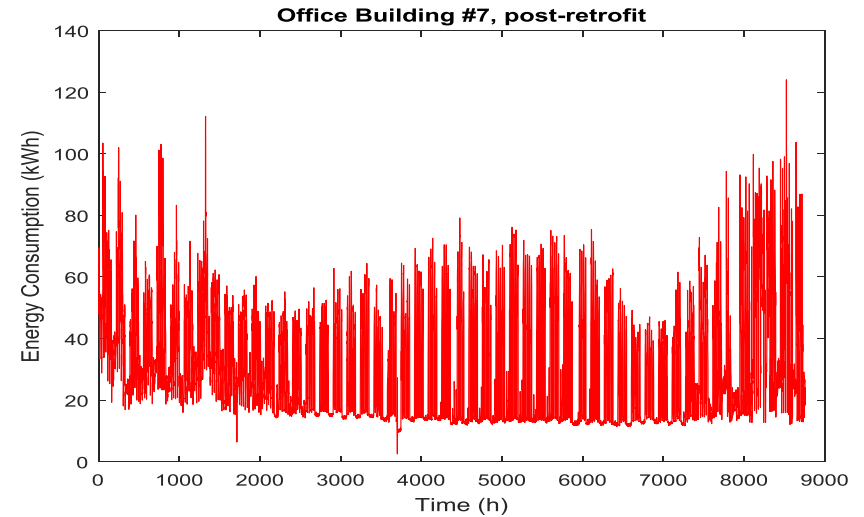
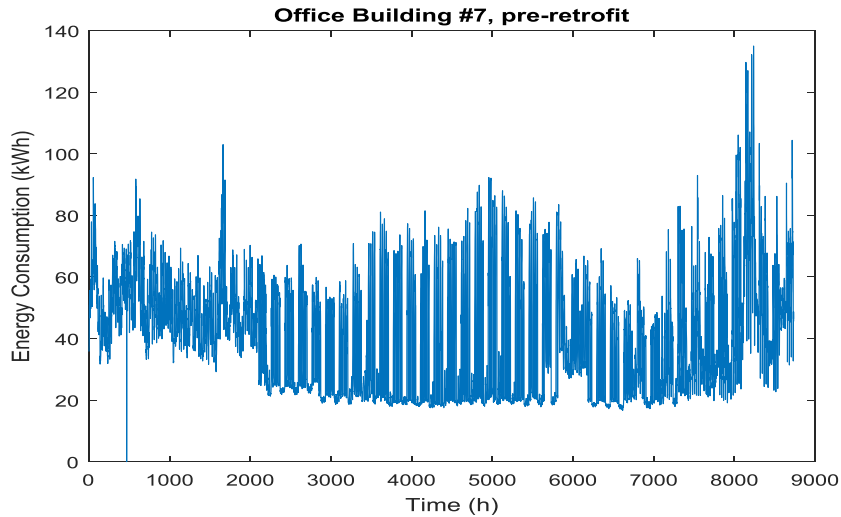
- Construction;
- Number of Offices;
- Systems;
- Retrofit;
- ...



<https://trythink.github.io/buildingsdatasets/>

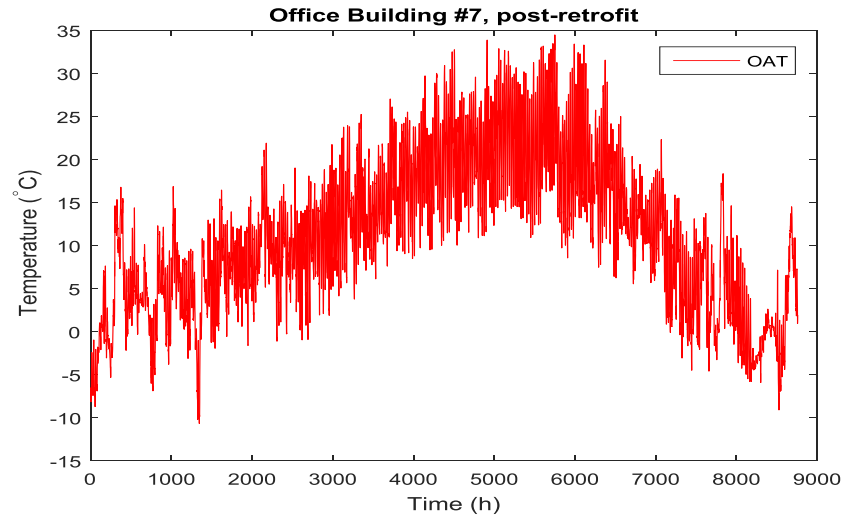
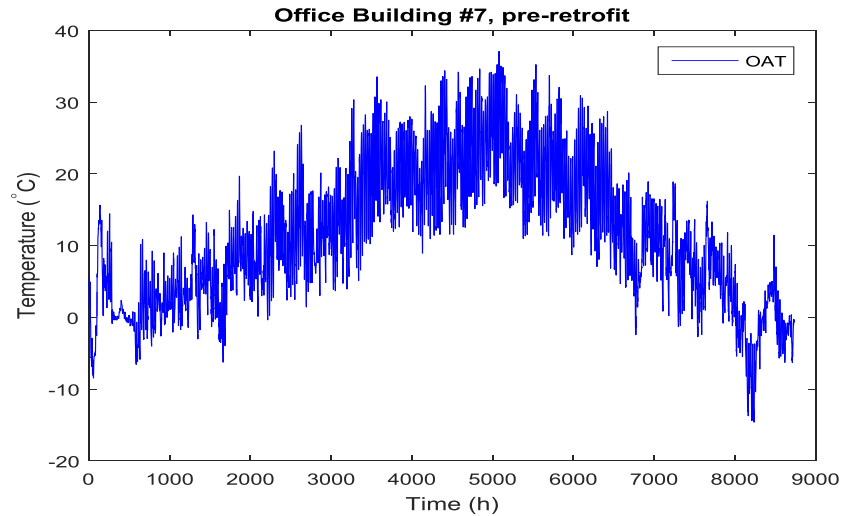
Problem definition

Baseline and post-retrofit demand for Office Building #7



Problem definition

Baseline and post-retrofit weather conditions for Office Building #7 location



The baseline model

Constructing the baseline model:

- $y = f(x) + \epsilon$, with:
 - $y \in \mathbb{R}$, the target variable (e.g. Total Energy Consumption); and
 - $x \in \mathbb{R}^n$, the features (e.g. Outside temperature, occupancy, etc.);
- Goal: given historical data for the target variable and the features, construct a model that can predict the energy consumption of the building.

The baseline model

Constructing the baseline model:

- $y = f(x) + \epsilon$, with:

Predictor:

- Able to approximate complex functions;
- Good statistical properties (robust to outliers, avoid overfitting, etc.);
- **Gaussian Processes for Regression**
(**GPML toolbox**, <http://www.gaussianprocess.org/gpml/code/matlab/doc/>).

The baseline model

Constructing the baseline model:

- $y = f(x) + \epsilon$, with:

Features:

- Meaningful for the task at hand;
- Feature selection requires expert knowledge.

The baseline model

Constructing the baseline model:

- $y = f(x) + \epsilon$, with:

Time-step:

- 1 hour for FDD applications;
- 1 day for post-retrofit evaluation.

Office Building #7 hourly data – The selected features

Feature vector \mathbf{x} :

- x_1 : Day of week;
- x_2 : Time of Day;
- x_3 : Outside Air Temperature (OAT).

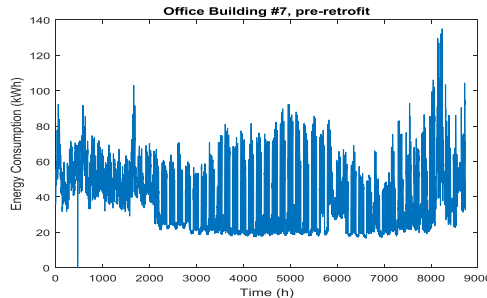
x_1^1	x_2^1	x_3^1	y^1
x_1^2	x_2^2	x_3^2	y^2
...			
x_1^N	x_2^N	x_3^N	y^N

} N samples

Office Building #7 hourly data – Training/Test sets

From initial baseline set:

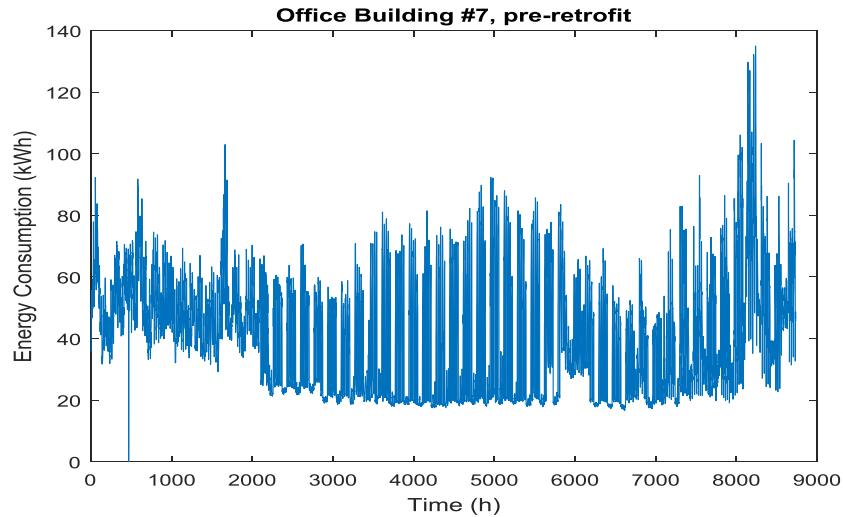
- Use 90% of the samples for training;
- Use 10% of the samples for validation;
- Use the coefficient of determination (or R^2) to determine the goodness-of-fit.



$$R^2 = 1 - \frac{\sum_1^N (y_{\text{pred}}^i - y^i)^2}{\sum_1^N (\bar{y} - y^i)^2}$$

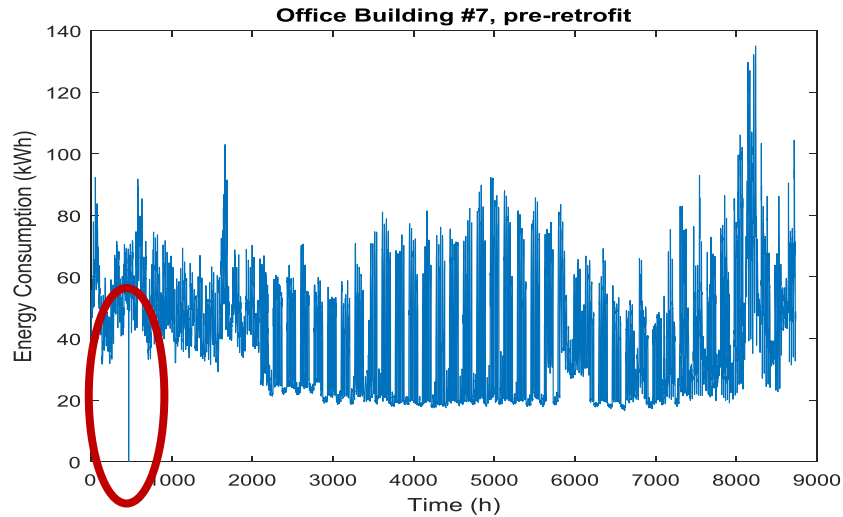
Office Building #7 hourly data – Data pre-processing

Eliminate outliers



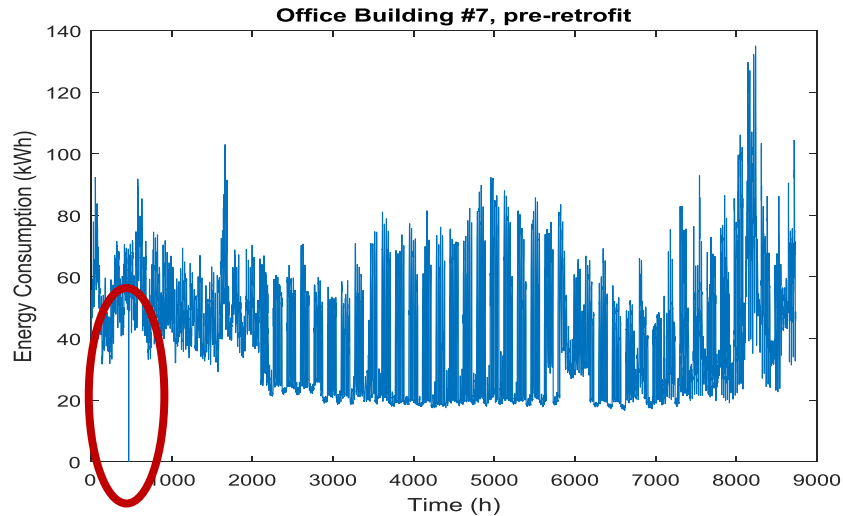
Office Building #7 hourly data – Data pre-processing

Eliminate outliers



Office Building #7 hourly data – Data pre-processing

Eliminate outliers

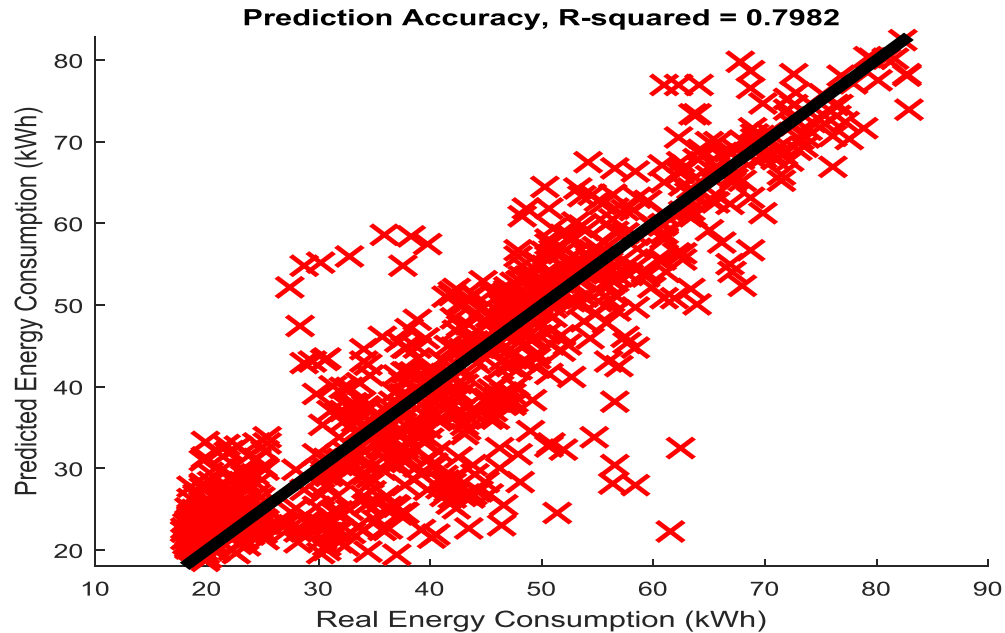


Normalize data

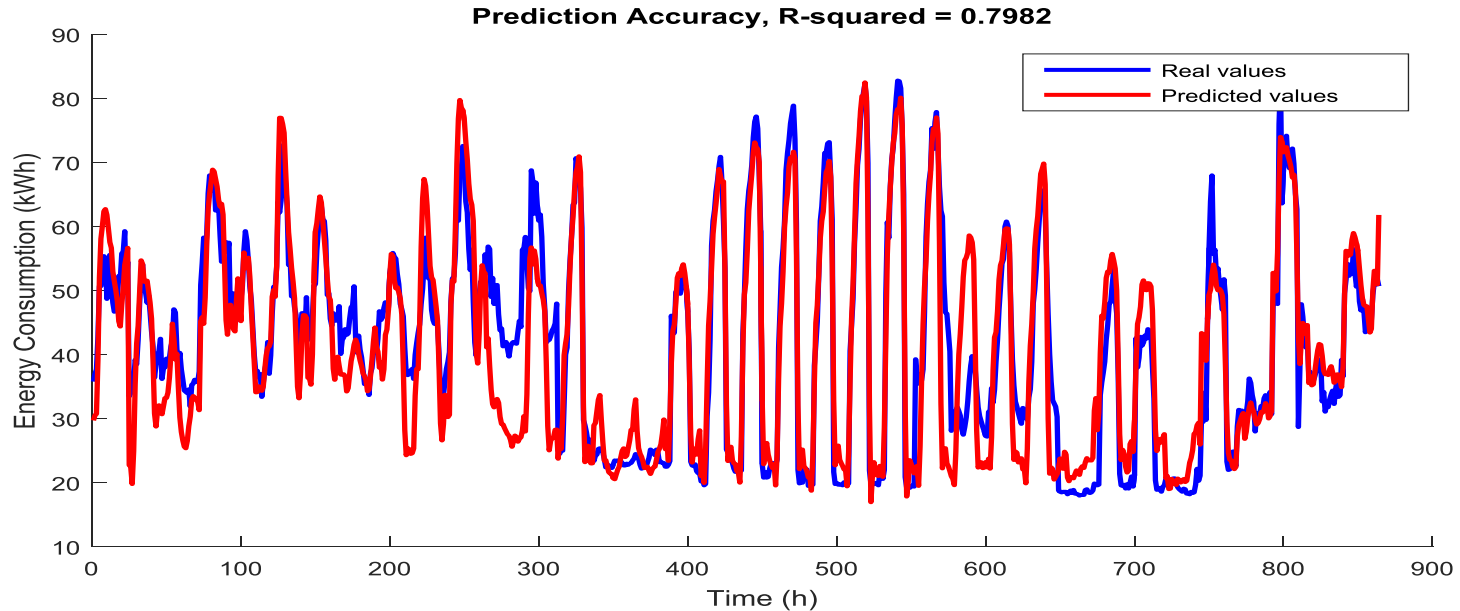
$$\mathbf{x} = \frac{\mathbf{x}}{\max(\mathbf{x})}$$

$$\mathbf{y} = \frac{\mathbf{y}}{\max(\mathbf{y})}$$

Office Building #7 hourly data – Prediction Accuracy



Office Building #7 hourly data – Prediction Accuracy



Office Building #7 daily data – The selected features

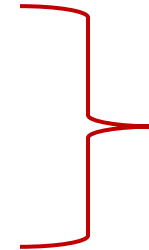
Feature vector \mathbf{x} :

- x_1 : Day of week;
- x_2 : Minimum OAT;
- x_3 : Maximum OAT;
- x_4 : Mean OAT;
- x_5 : Heating Degree Days (HDD);
- x_6 : Cooling Degree Days (CDD).

Office Building #7 daily data – The selected features

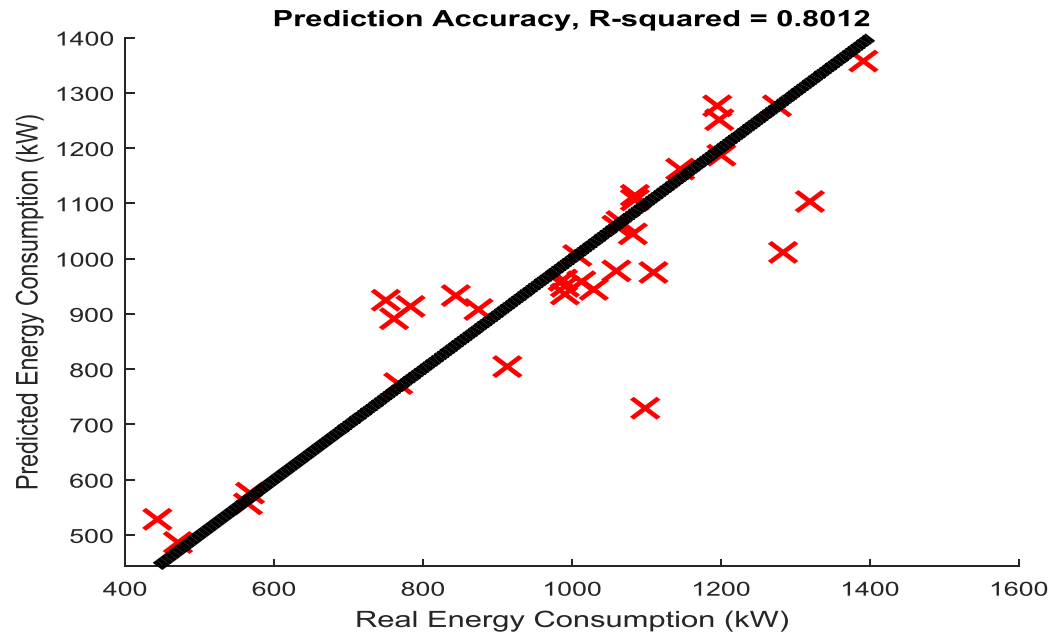
Daily dataset

x_1^1	x_2^1	...	x_6^1	y^1
x_1^2	x_2^2	...	x_6^2	y^2
...				
x_1^N	x_2^N	...	x_6^N	y^N

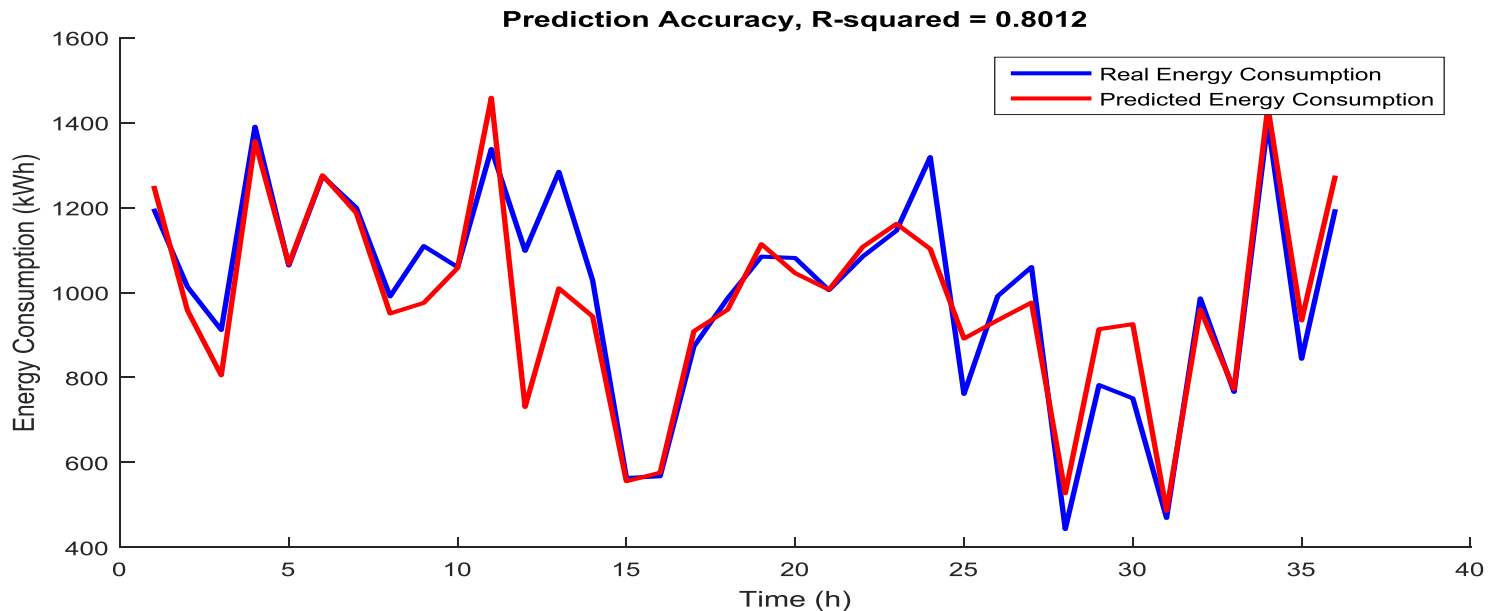


N/24
samples

Office Building #7 daily data – Prediction Accuracy

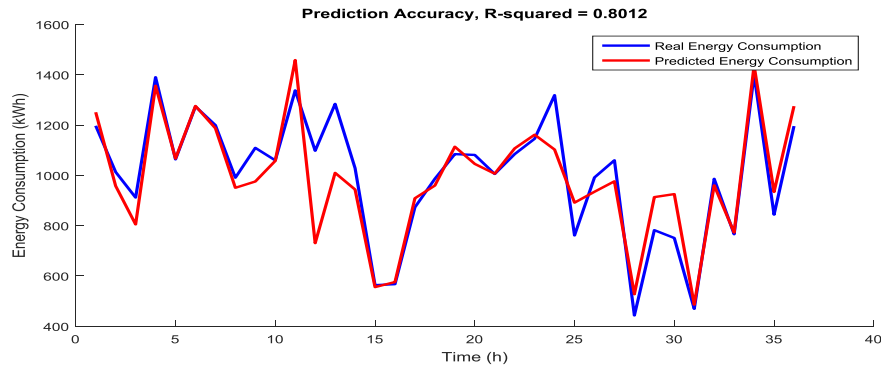


Office Building #7 daily data – Prediction Accuracy

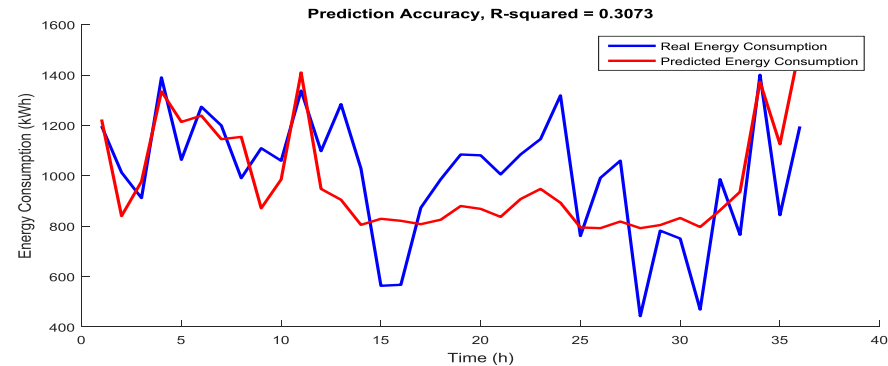


Office Building #7 daily data – Importance of Feature Selection

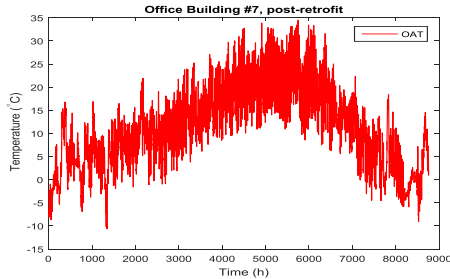
All features



Mean OAT



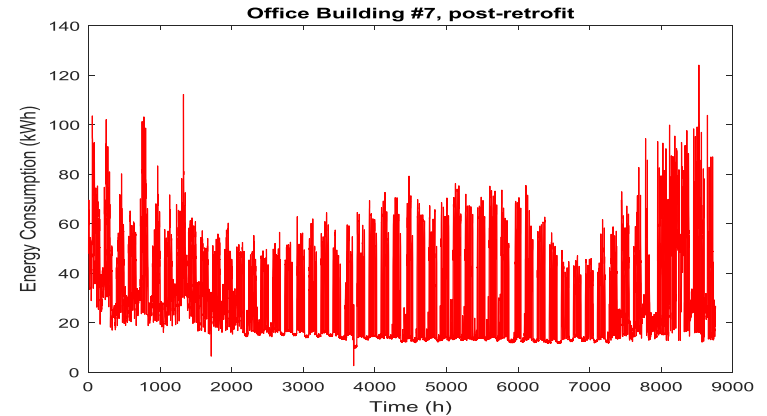
Office Building #7 hourly data – Estimation of energy savings



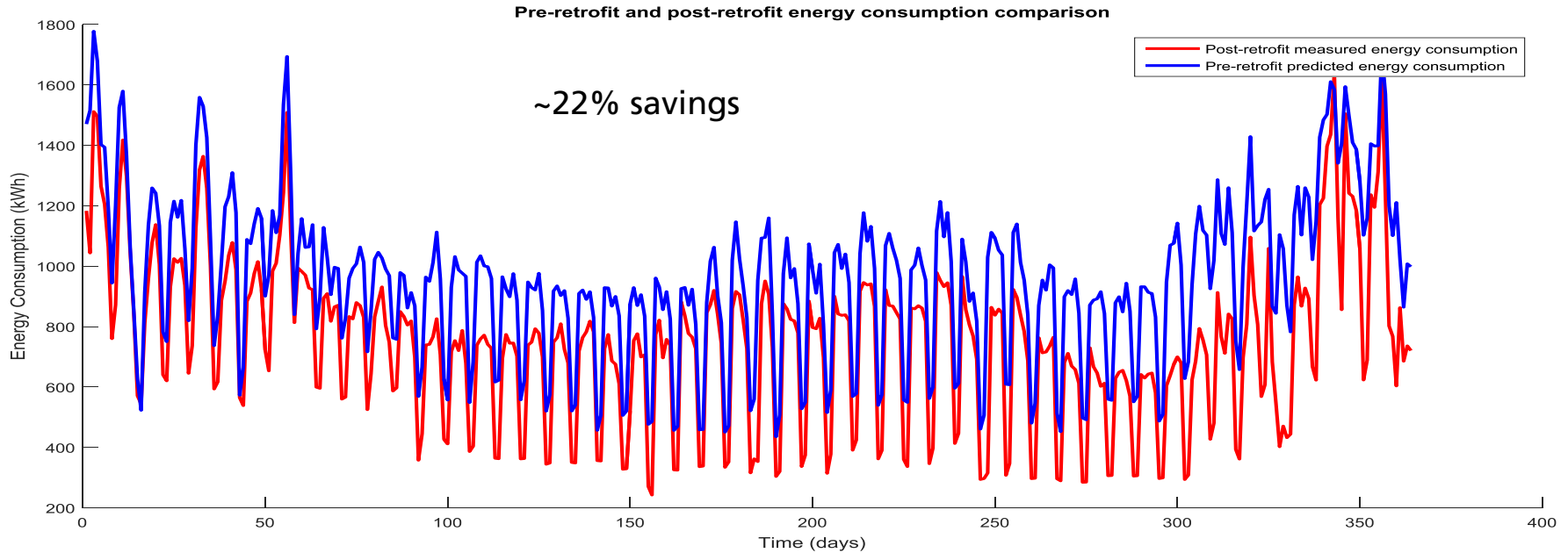
$$[x_1 - x_6]$$



$$y = f(x) + \epsilon$$



Office Building #7 daily data – Estimation of energy savings



Conclusions and Future Work

Properties of the methodology:

- Robust and accurate prediction of energy consumption;
- Selecting informative features is crucial;
- Historical data of good quality are required.

Next steps:

- Test with more datasets from different building types;
- Offer energy prediction/benchmarking services as part of a collection of cloud-based services for efficient building operation.

References

- IPMVP (International Performance Measurement and Verification Protocol): Concepts and Options for Determining Energy and Water Savings. vol. 1, Technical Report
- Heo, Yeonsook, and Victor M. Zavala. "Gaussian process modeling for measurement and verification of building energy savings." *Energy and Buildings* 53 (2012): 7-18.
- Park, Sumee, et al. "Methodological approach for calibration of building energy performance simulation models applied to a common "measurement and verification" process." *Bauphysik* 35.4 (2013): 235-241.
- Zhang, Yuna, et al. "Comparisons of inverse modeling approaches for predicting building energy performance." *Building and Environment* 86 (2015): 177-190.
- Hsu, David. "Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data." *Applied Energy* 160 (2015): 153-163.

Thank you for your attention 😊



Calculation of HDD and CDD

<http://www.vesma.com/ddd/ddcalcs.htm>

HDD

Condition	Formula used
$T_{\min} > T_{\text{base}}$	$D_h = 0$
$(T_{\max} + T_{\min})/2 > T_{\text{base}}$	$D_h = (T_{\text{base}} - T_{\min})/4$
$T_{\max} \geq T_{\text{base}}$	$D_h = (T_{\text{base}} - T_{\min})/2 - (T_{\max} - T_{\text{base}})/4$
$T_{\max} < T_{\text{base}}$	$D_h = T_{\text{base}} - (T_{\max} + T_{\min})/2$

CDD

Condition	Formula used
$T_{\max} < T_{\text{base}}$	$D_c = 0$
$(T_{\max} + T_{\min})/2 < T_{\text{base}}$	$D_c = (T_{\max} - T_{\text{base}})/4$
$T_{\min} \leq T_{\text{base}}$	$D_c = (T_{\max} - T_{\text{base}})/2 - (T_{\text{base}} - T_{\min})/4$
$T_{\min} > T_{\text{base}}$	$D_c = (T_{\max} + T_{\min})/2 - T_{\text{base}}$