Improving energy management in buildings through data analytics: challenges and opportunities

alfonso.capozzoli@polito.it

marco.piscitelli@polito.it

Speakers:

Prof. Alfonso Capozzoli

Eng. Marco Savino Piscitelli

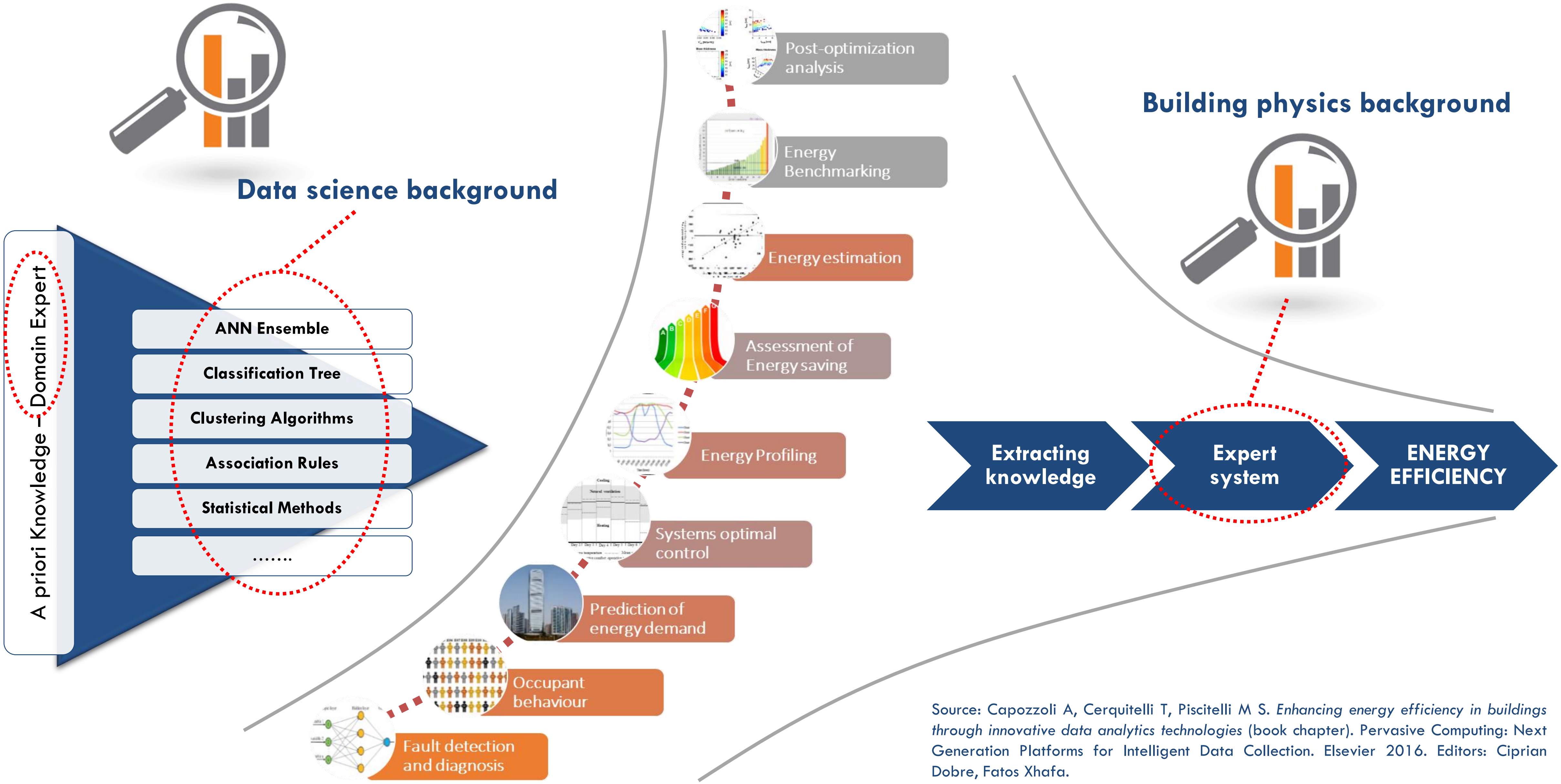
BAEDA Lab – Building Automation Energy Data Analytics

TEBE – Technology Energy Building Environment

Hong Kong

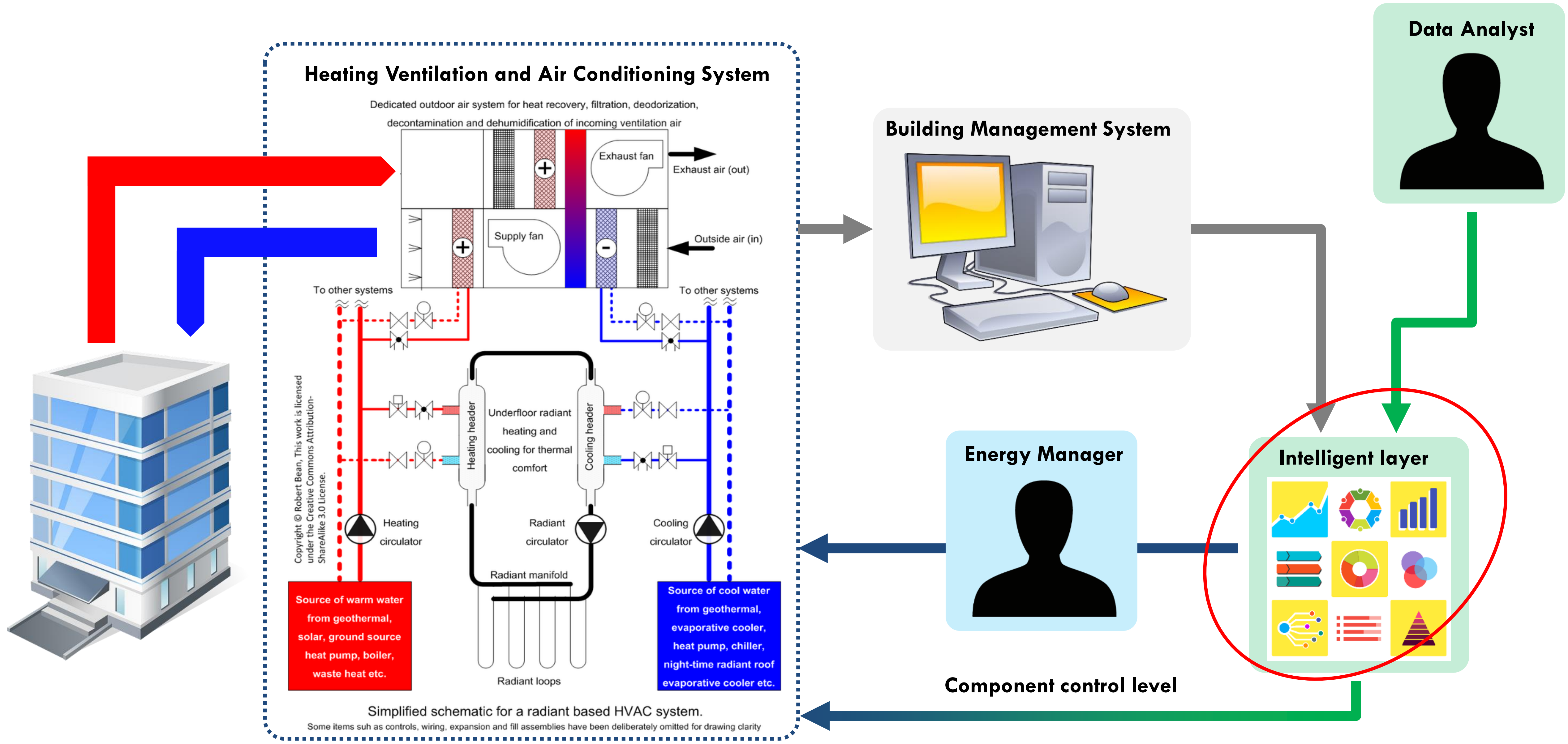
05/12/2018

A new professional figure?

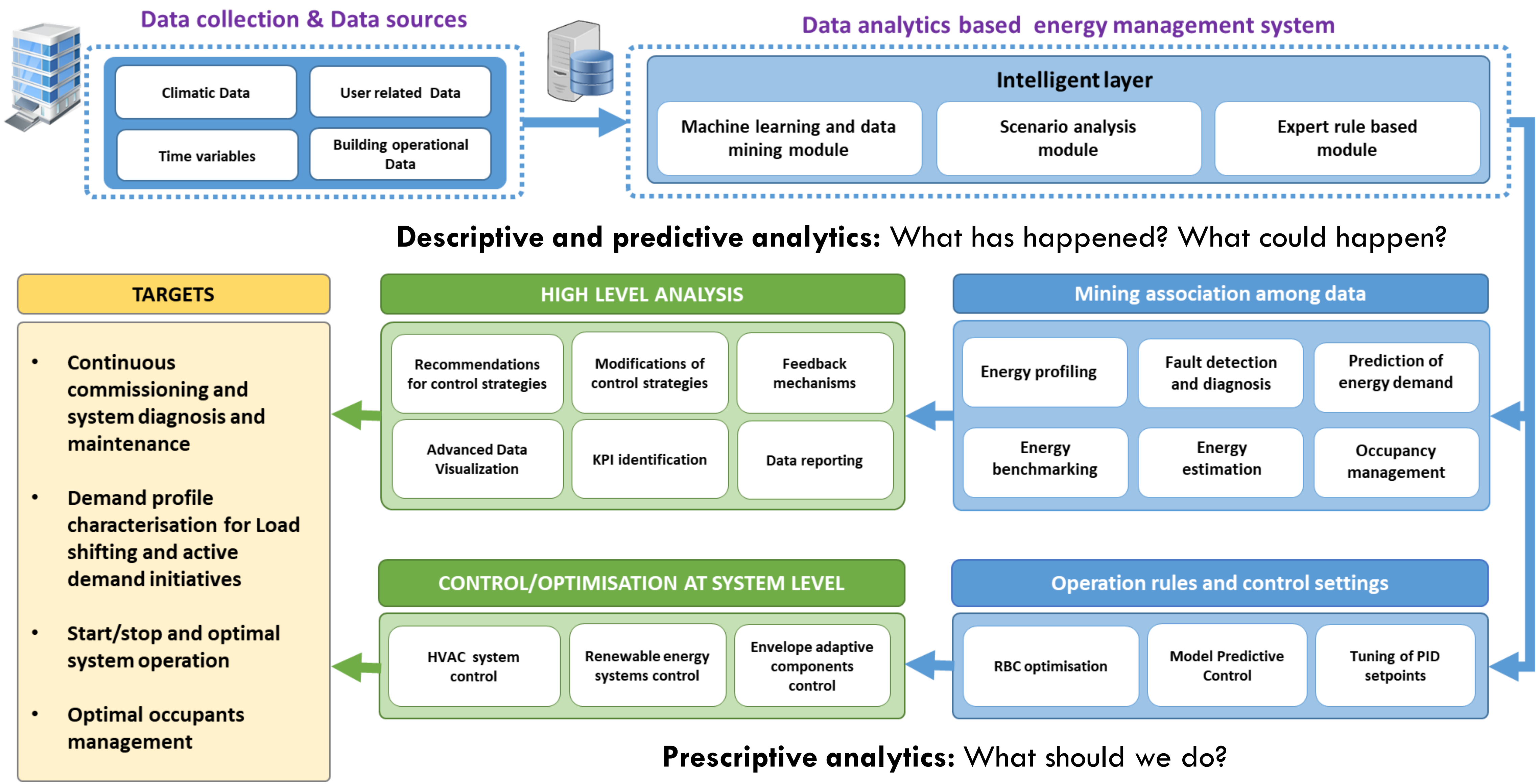


Source: Capozzoli A, Cerquitelli T, Piscitelli M S. *Enhancing energy efficiency in buildings through innovative data analytics technologies* (book chapter). Pervasive Computing: Next Generation Platforms for Intelligent Data Collection. Elsevier 2016. Editors: Ciprian Dobre, Fatos Xhafa.

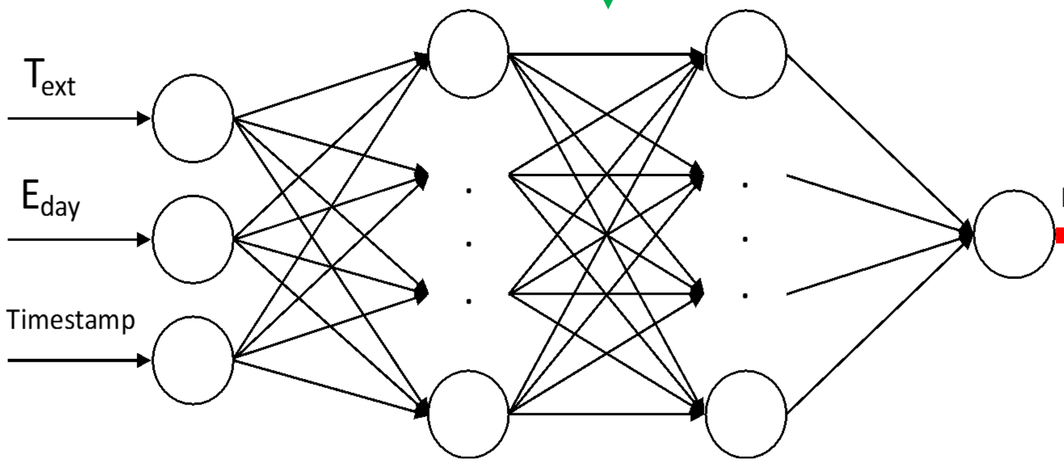
Advanced data analytics in BEMS: which barriers?



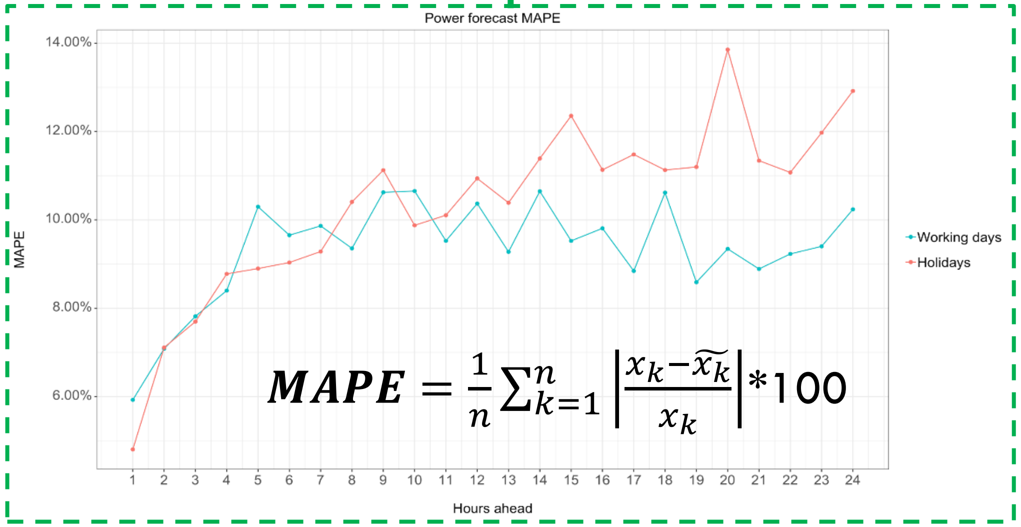
Multilevel building energy management system in buildings



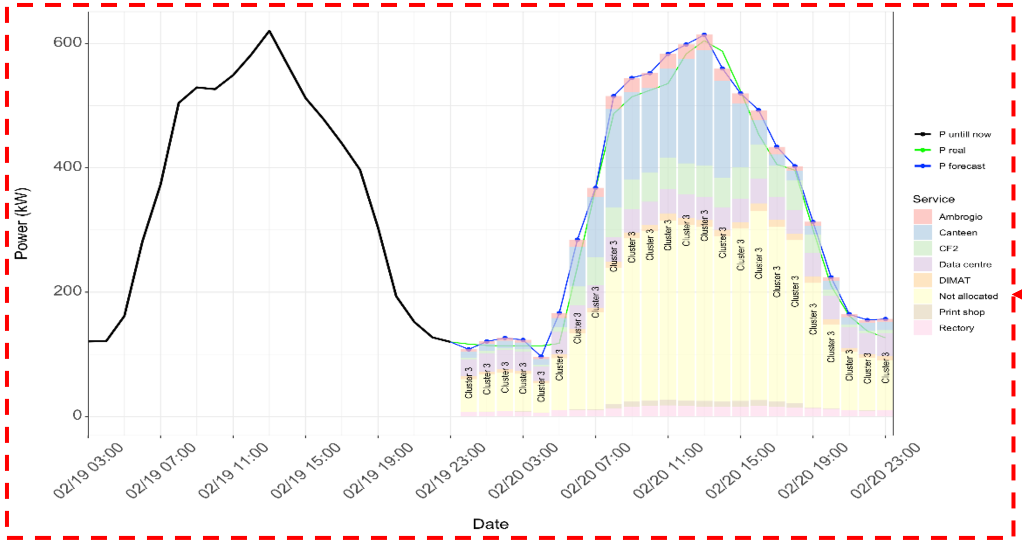
Predictive analytics: development of advanced prediction models



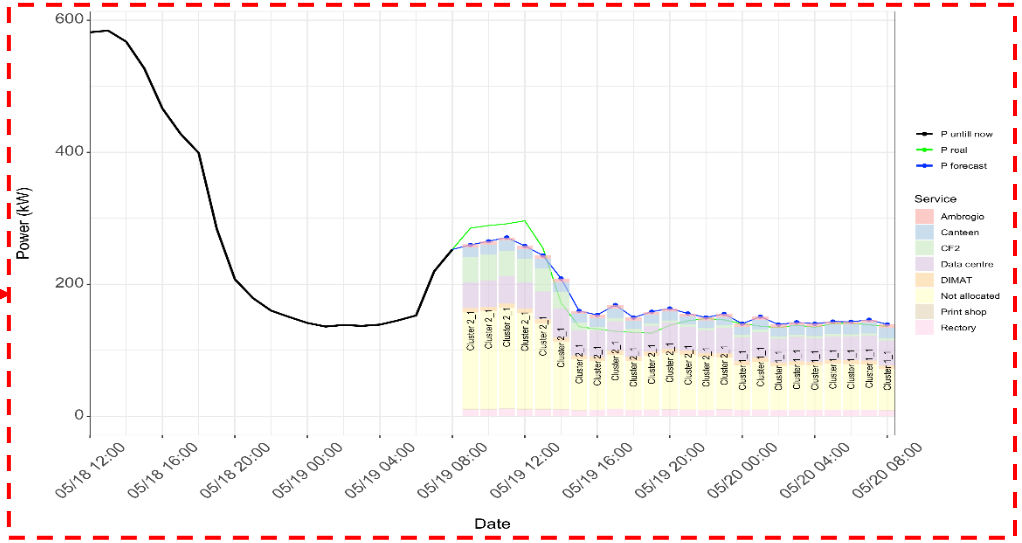
ANN model structure



MAPE of the forecasting model at each time horizon



Power load forecasting for a working day



Power load forecasting for a weekend

Prescriptive analytics: Optimal and adaptive control in buildings

Predicted output

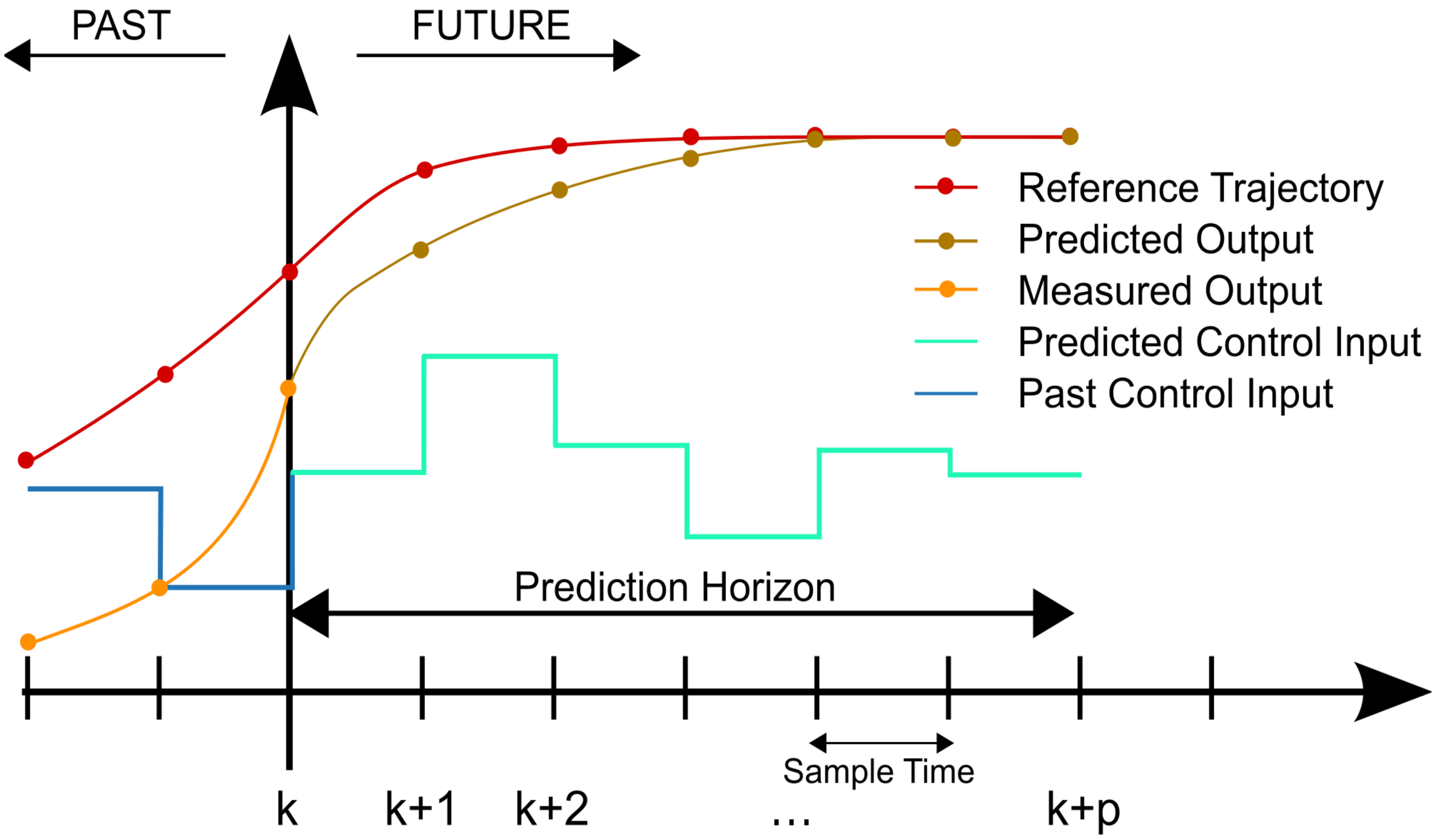
- A model predicts plant behavior in the future on the basis of predicted disturbances

Optimization

- The best control input sequence is solved

Control intput

- The signal at the first timestep is applied



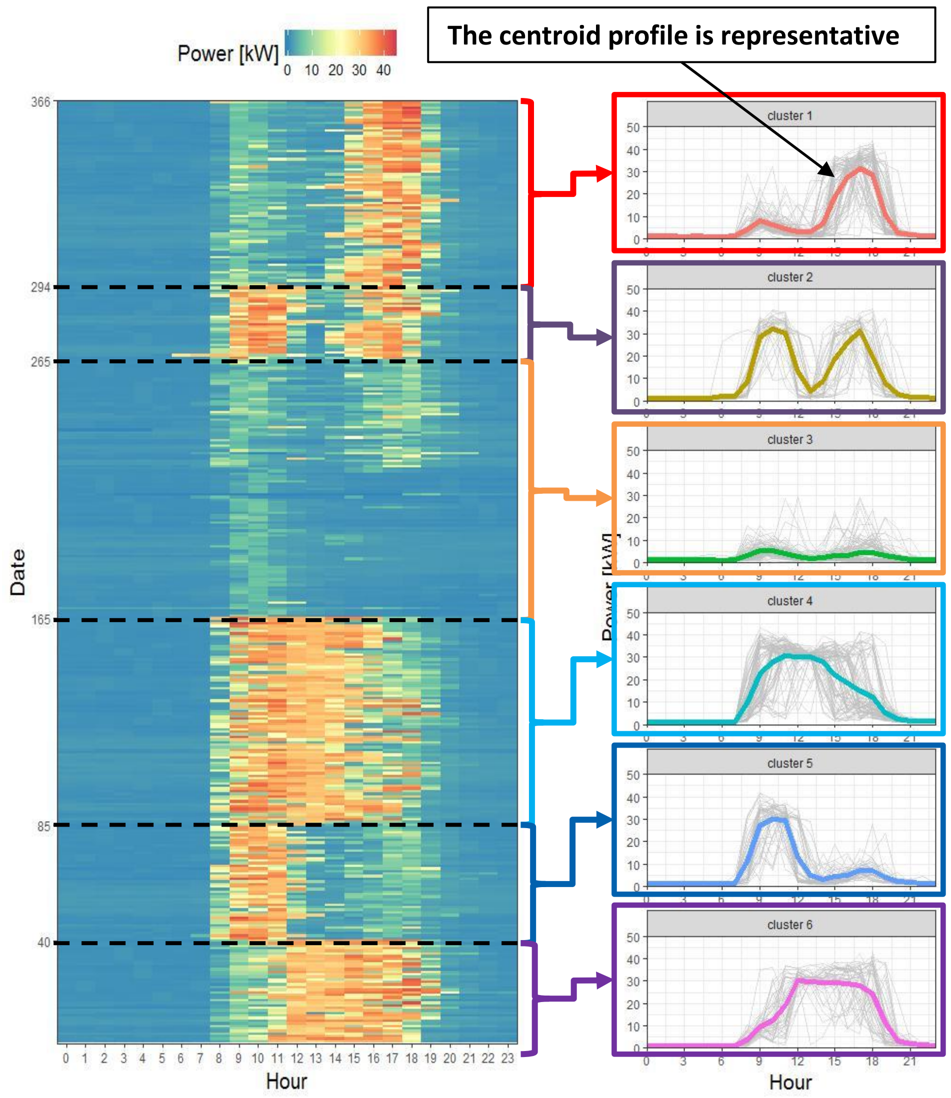
Challanges in descriptive analytics for energy and buildings

The mining of ***time series data*** has recently gained high attention as a way to **describe** and deeply ***characterise*** typical operational patterns and trends of ***energy consumption in buildings.***
Time series of building related variables have to be analyzed with an integrated approach preserving their ***association in the time domain***.

A novel paradigm based on temporal data analytics is needed to characterize dynamics at whole building or component level through:

- ***Sequential and recurrent pattern mining***
- ***Causality analysis***
- ***Time series segmentation and trend analysis***

Unsupervised pattern recognition



Source: Capozzoli A, Piscitelli M. S., Brandi S. Mining typical load profiles in buildings to support energy management in the smart city context. Energy Procedia 2017;0:5–7. doi:10.1016/j.egypro.2017.09.545.

Time series analytics at different scales

Detail of analysis

Energy time series analytics at system scale

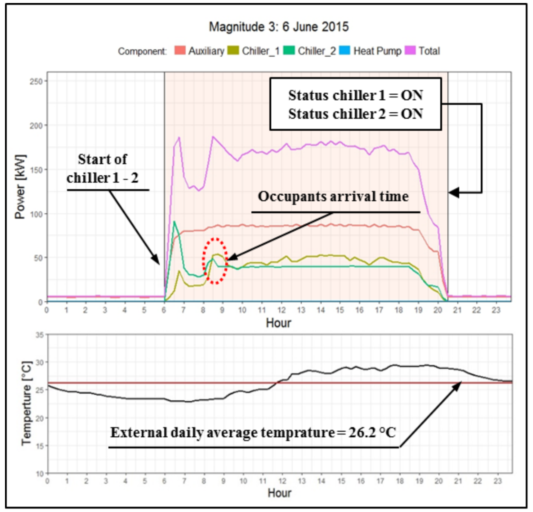
- Implementation of (FDD) strategies (e.g., AHU).
- Support the **optimal operation of chillers**

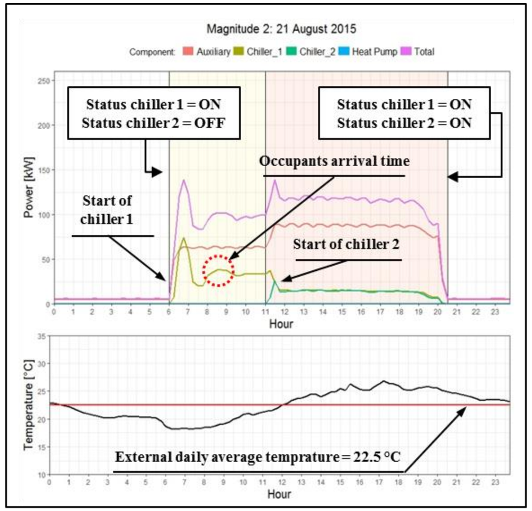
Energy time series analytics at single building scale

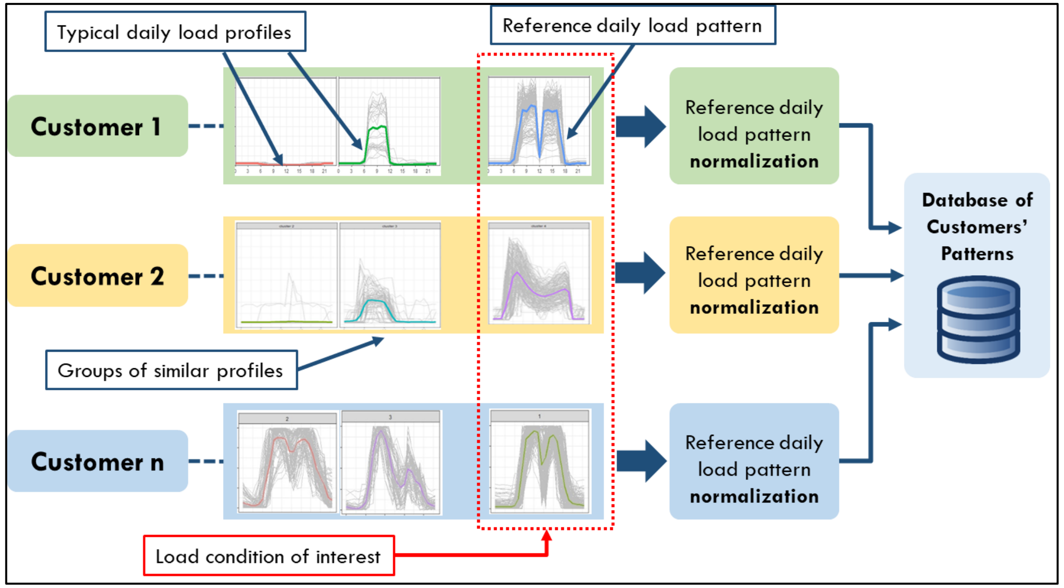
- Improve the accuracy of energy consumption forecasting models.
- Provide information for the **calibration** of simulation models.
- Implementation of (FDD) strategies.
- Energy **benchmarking** over time.
- Promote active **demand response** programs.

Energy time series analytics at multiple building scale

- Set rules for the automatic classification of new consumers
- Implementing targeted financial demand response programs.
- Better manage the grid operation
- Promote the modification of a load profile
- Fully exploit the benefits of energy management also at micro grid level.
- Assess the impact of DSM and DR initiatives over time.





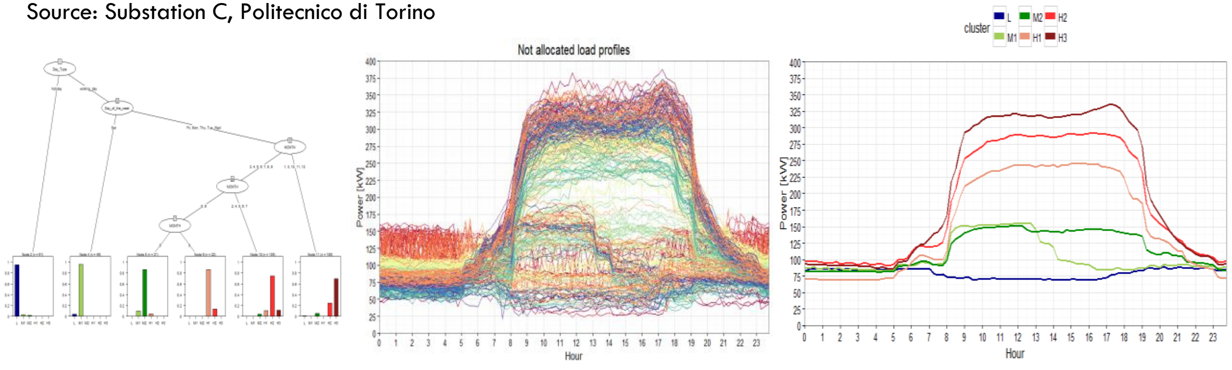


Source: Capozzoli A, Piscitelli M. S., Brandi S. Mining typical load profiles in buildings to support energy management in the smart city context. Energy Procedia 2017;0:5–7. doi:10.1016/j.egypro.2017.09.545.

Applications that benefit from time series analytics

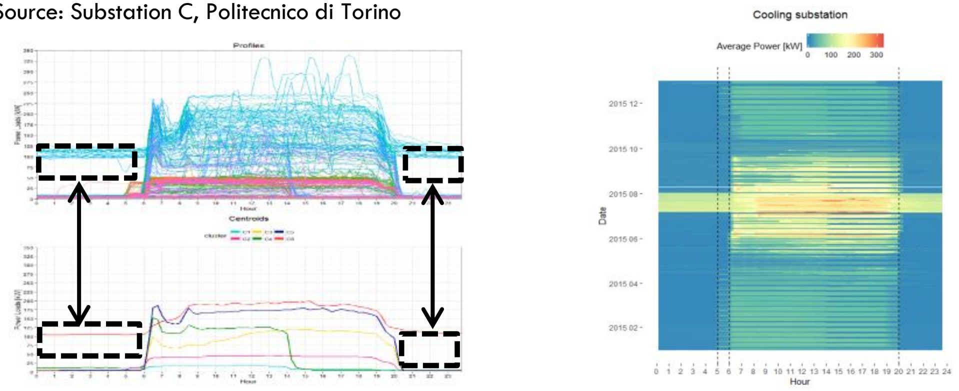
Demand profile characterisation

Source: Substation C, Politecnico di Torino

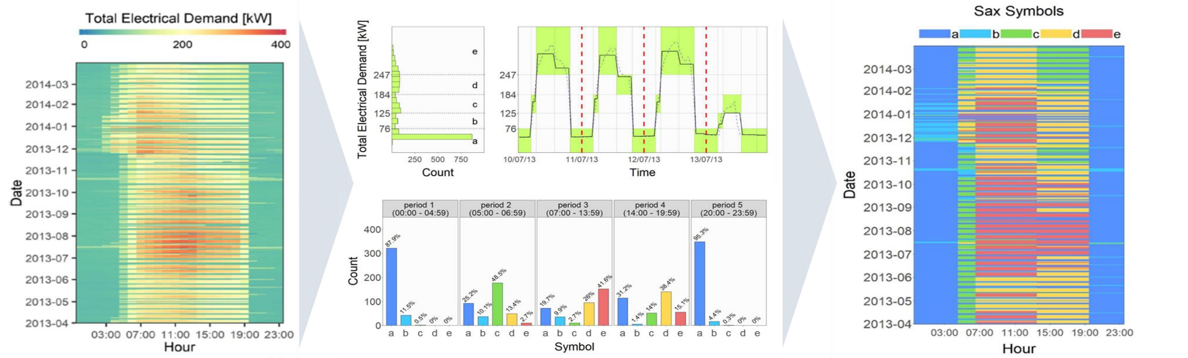


Anomalous trend detection

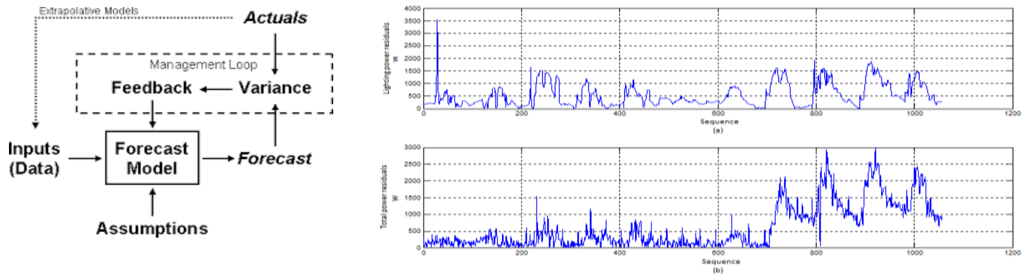
Source: Substation C, Politecnico di Torino



Integrated data analytics based energy management strategies



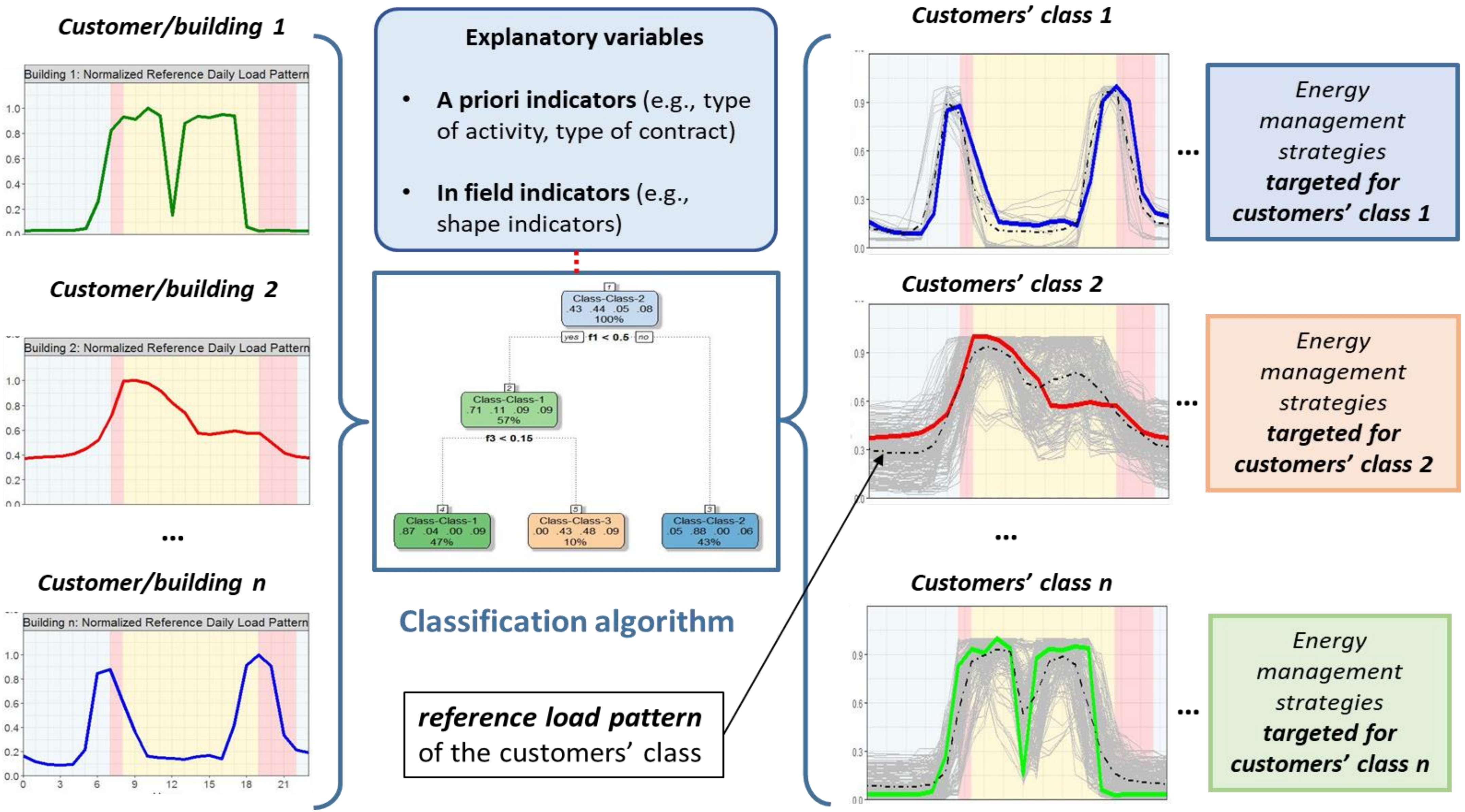
Data visualisation



Source: Capozzoli A., Lauro F., Khan I. (2015), Fault detection analysis using data mining techniques for a cluster of smart office buildings, Expert Systems with Applications, vol. 42, Issue 9, Elsevier,

Fault detection & diagnosis

Demand profile characterization: customers' classification



Source: Capozzoli A, Piscitelli M. S., Brandi S. Mining typical load profiles in buildings to support energy management in the smart city context. Energy Procedia 2017;0:5–7. doi:10.1016/j.egypro.2017.09.545.

Time series segmentation process

MxN matrix

N - dimension

M - dimension

date	00:00	06:00	12:00
20/09/2018	10	20	34
21/09/2018	6	15	67
22/09/2018	9	12	21
23/09/2018	10	20	34
24/09/2018	6	15	67
25/09/2018	9	12	21

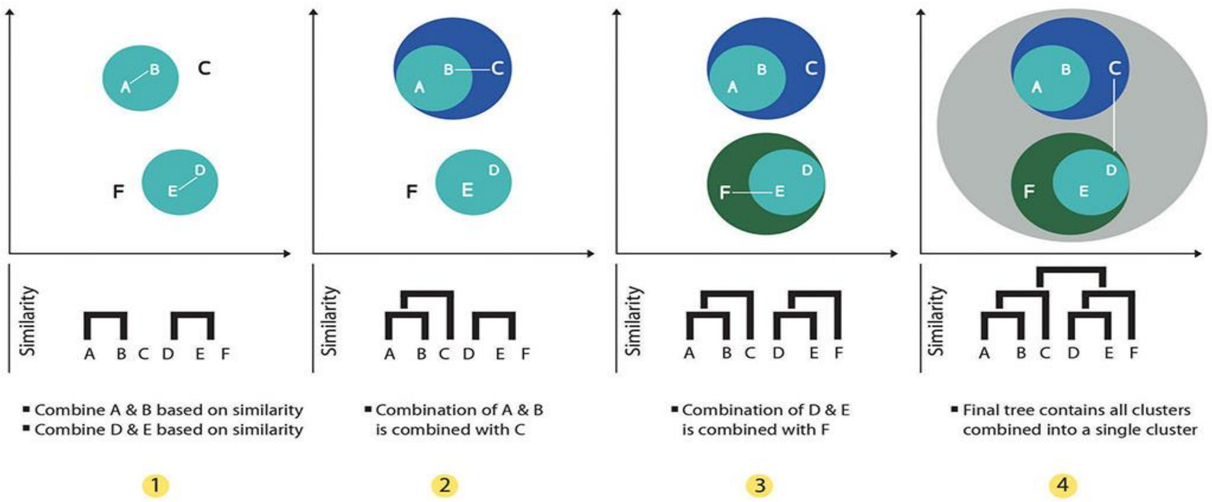
distance matrix

Euclidean distance

$$d_{ED}(X,Y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2}$$

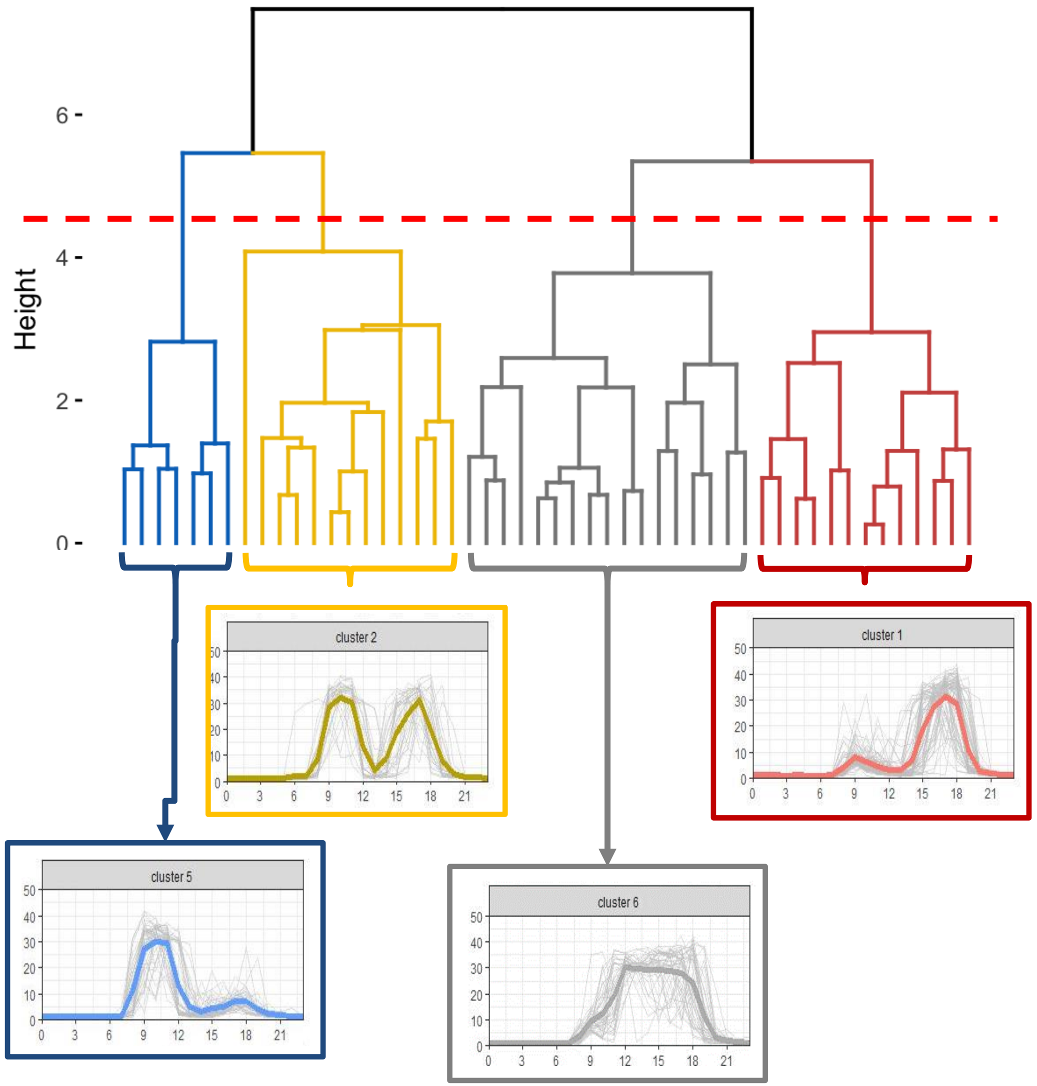
$$A = \begin{bmatrix} 0 & d_{12}^2 & d_{13}^2 & \dots & d_{1n}^2 \\ d_{21}^2 & 0 & d_{23}^2 & \dots & d_{2n}^2 \\ d_{31}^2 & d_{32}^2 & 0 & \dots & d_{3n}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1}^2 & d_{n2}^2 & d_{n3}^2 & \dots & 0 \end{bmatrix}$$

Selection of the clustering algorithms



https://www.brandidea.com/hierarchicalclustering.html

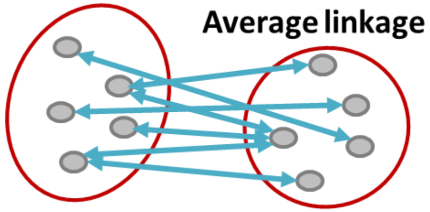
Identification of the optimal number of clusters



Managing similarity in time series segmentation

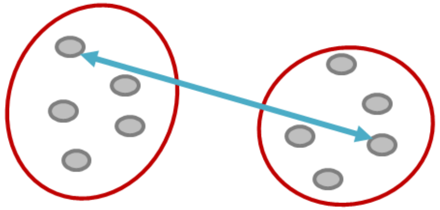
Which algorithm?

Average linkage



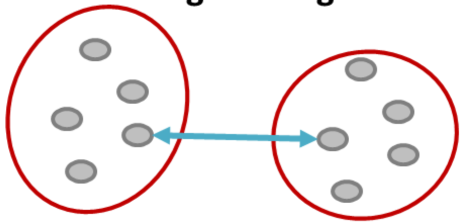
$$L(r,t) = \frac{1}{n_r n_t} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} D(x_{ri}, x_{tj})$$

Complete linkage



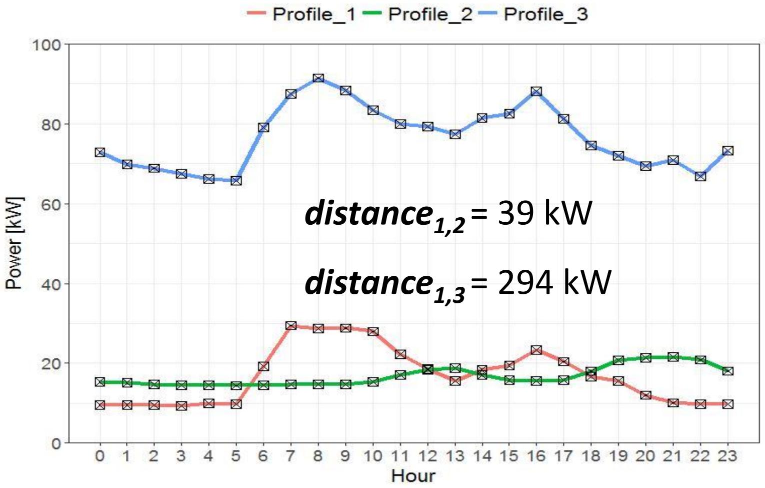
$$L(r,t) = \max(D(x_{ri}, x_{tj}))$$

Single linkage



$$L(r,t) = \min(D(x_{ri}, x_{tj}))$$

Which distance?



Euclidean distance

$$d_{ED}(X,Y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2}$$

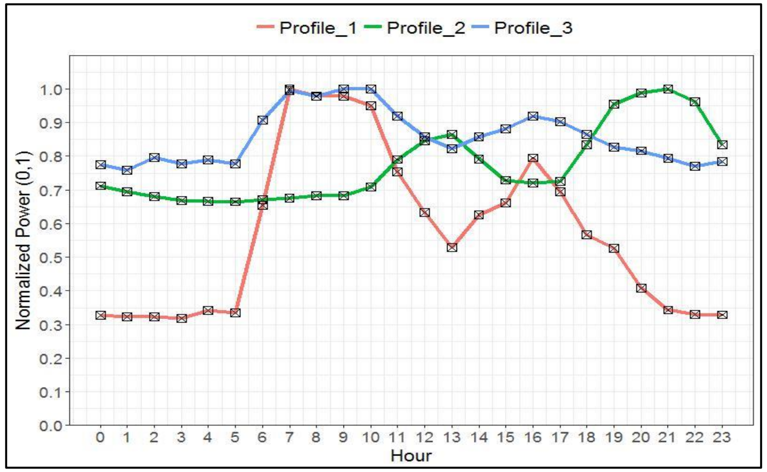
Pearson correlation coefficient (PCC)

$$PCC(X,Y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

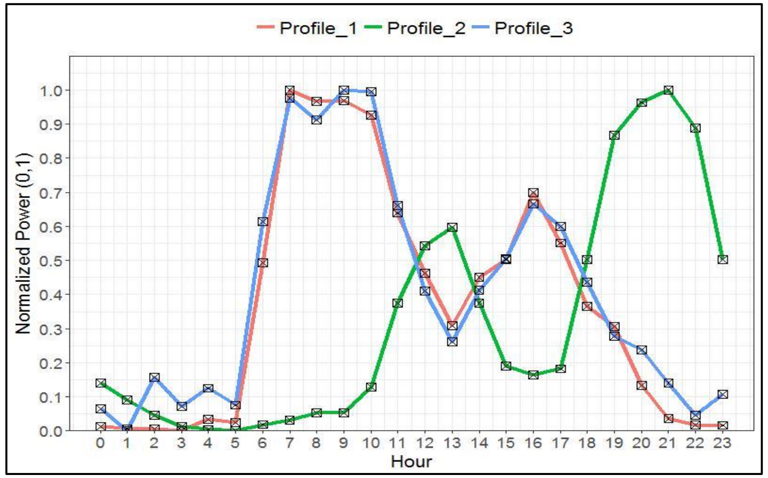
$$d_{PCC}(X,Y) = 1 - PCC$$

Dynamic Time Warping (DTW)

Which data normalization?



Normalization on daily maximum value



Normalization between daily maximum and minimum value

Energy time series analytics at single building/system scale

Data visualisation and anomalous trend detection

Objectives

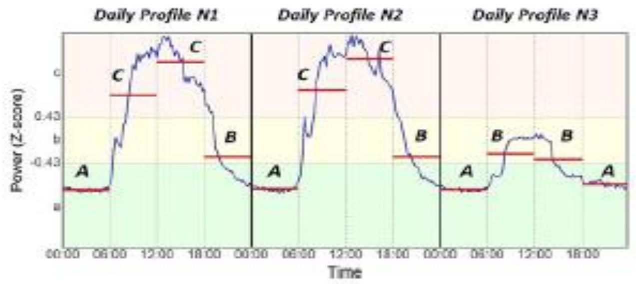
1. **Characterization of whole building energy consumption** patterns over time (univariate problem);
2. Development of **INFREQUENT PATTERN DETECTION** procedure in quasi real-time;
3. Enhanced **data visualization**

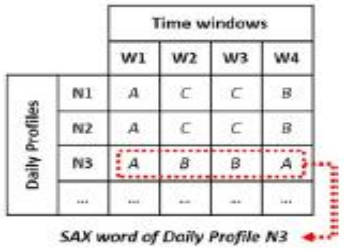
Fault detection and Diagnosis in AHU

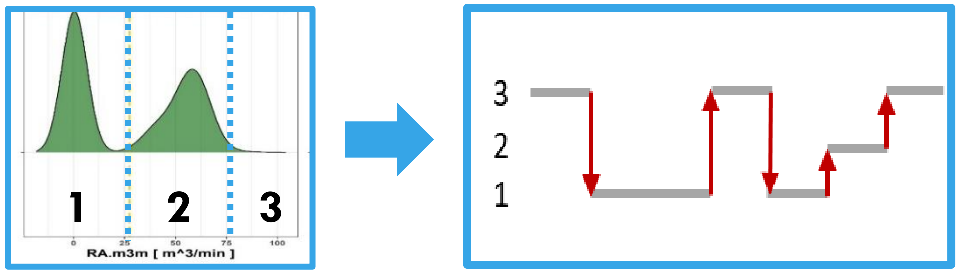
Objectives

1. Create a Model-free and Unsupervised methodology for **FAULT DETECTION and DIAGNOSIS in AHU**
2. The methodology has to exploit **Temporal Association Rules** Mining to correlate different variables with a time lag between the events, to consider even inertial effect in the response (multivariate problem).

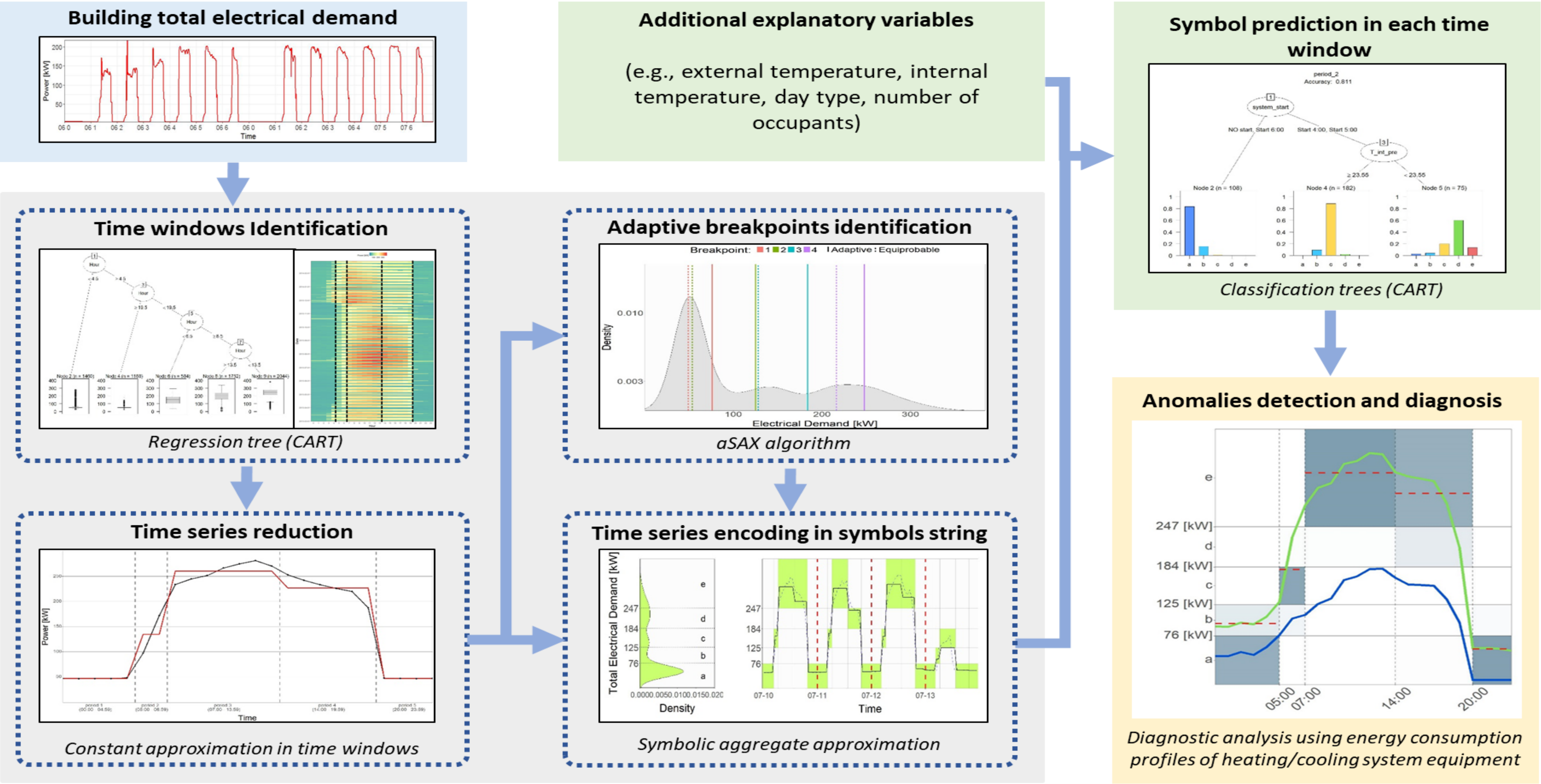
Both procedures rely on **temporal abstraction** as a preprocessing stage for knowledge extraction







Data visualisation and anomalous trend detection

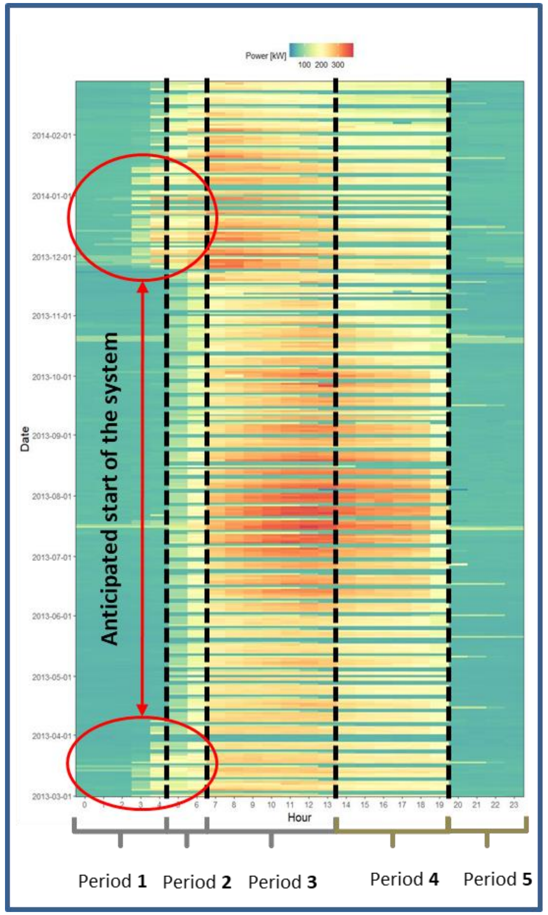


Source: Capozzoli A, Piscitelli M S, Brandi S., Grassi D., Chicco G. Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. Energy 2018

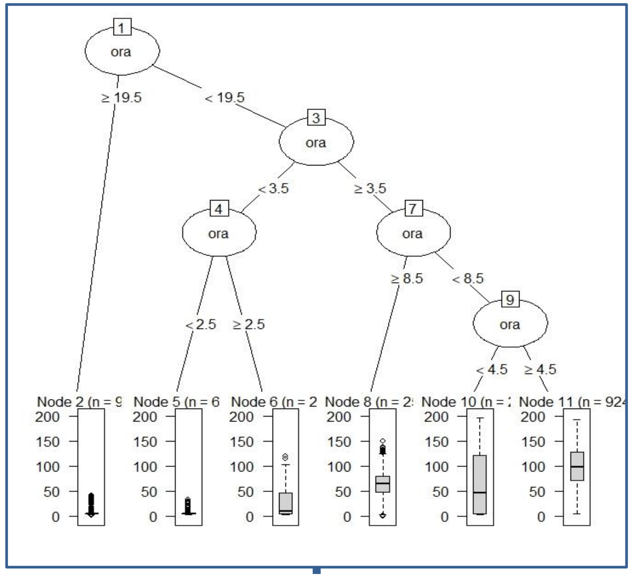
Temporal abstraction in univariate time series

Time window identification

The optimal number and size of the time windows for piecewise approximated aggregation is evaluated through a regression tree.



non-overlapped time windows with unequal lengths

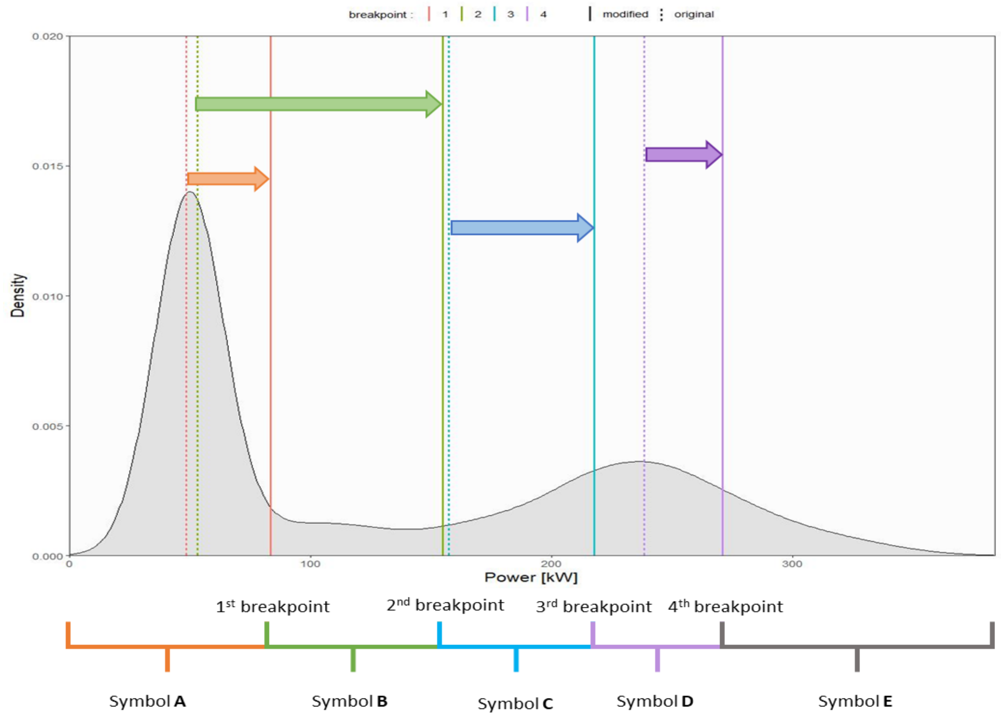


Symbol identification with adaptive breakpoints

1) The hypothesis of normal distribution is not verified.

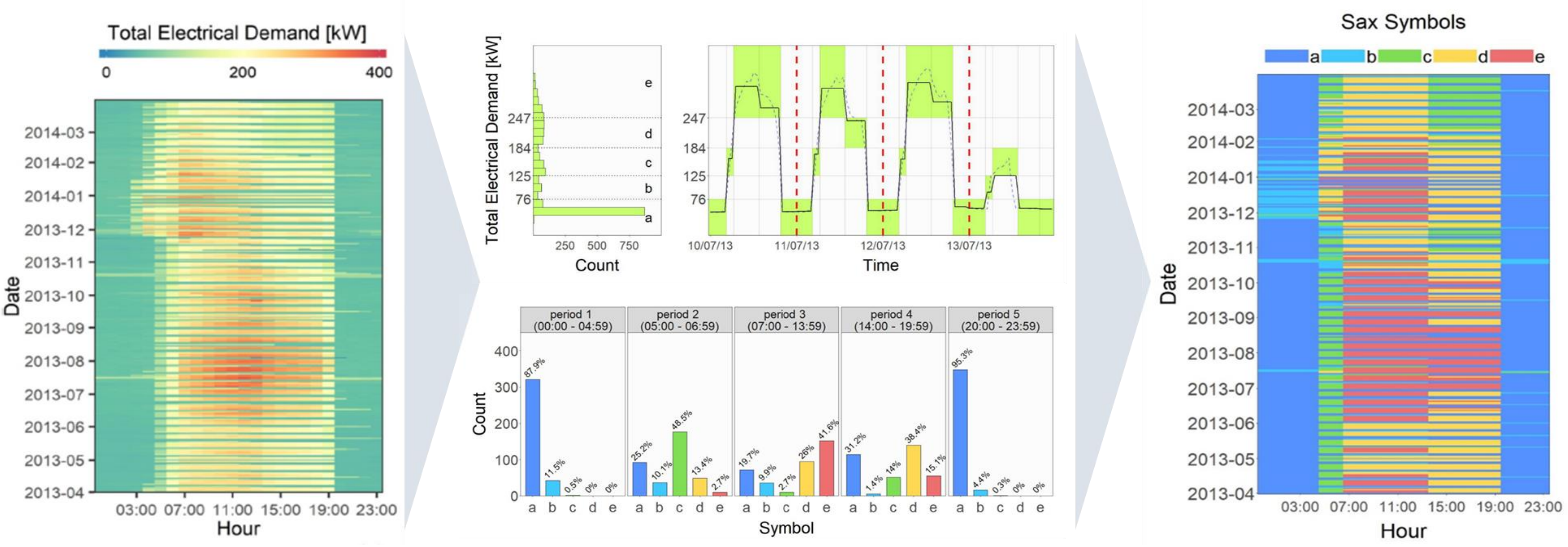
2) The identification of equi-probable symbols on real distribution results in very narrow ranges for low power values.

*adaptive **S**ymbolic **A**ggregate appro**X**imation (aSAX)*



Temporal abstraction in time series

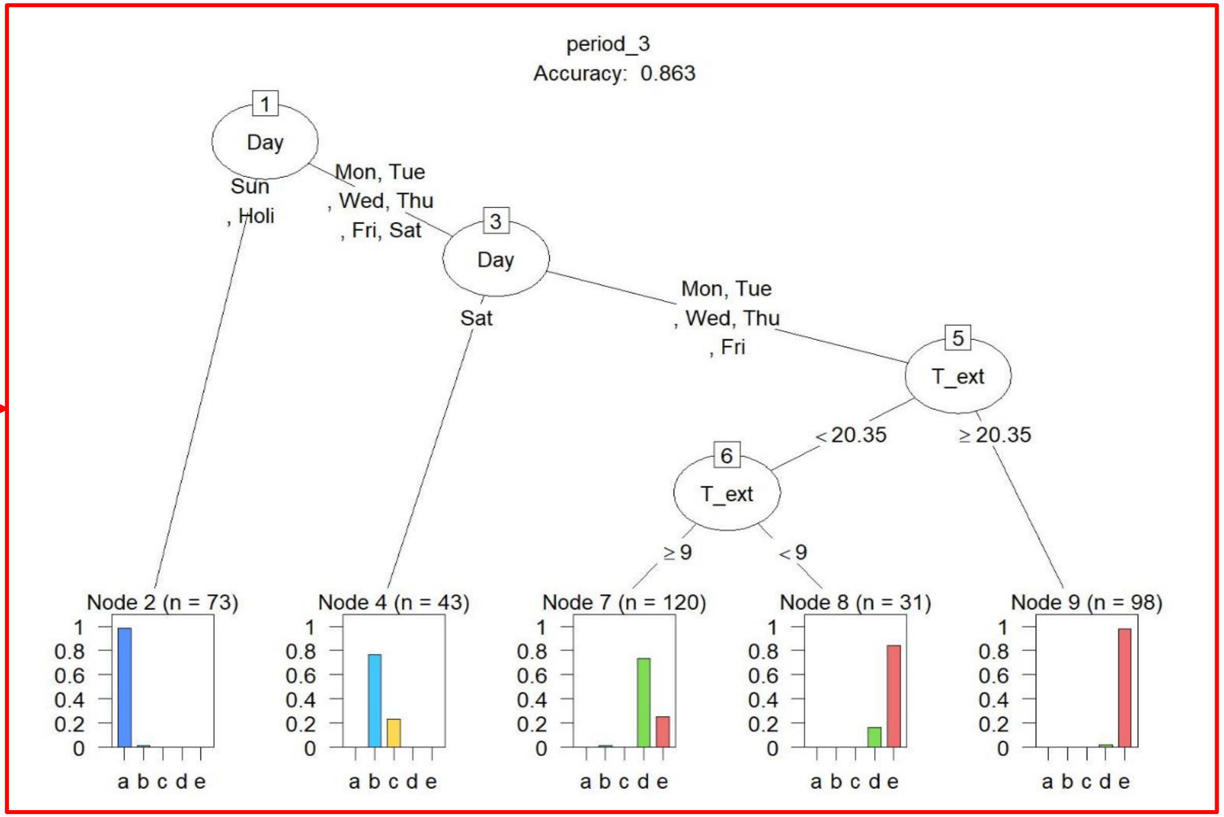
Temporal abstraction consists in transforming time series from numerical sequences to discrete state sequences by means of the reduction and the transformation of data.



Source: Capozzoli A, Piscitelli M S, Brandi S., Grassi D., Chicco G. Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. Energy 2018

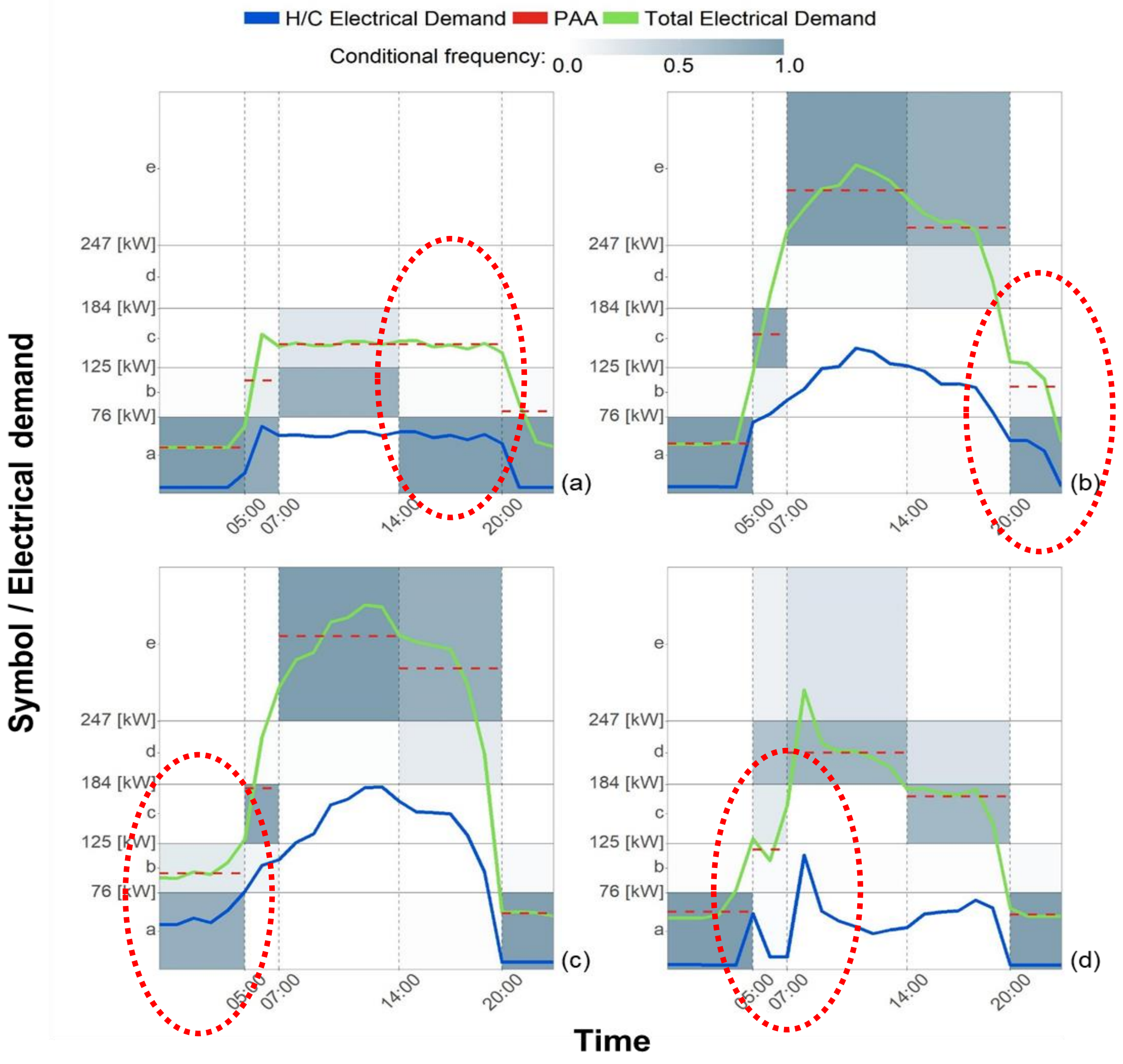
In this way the **discrete states, called events**, or their transaction/sequence could be associated to the observation of physical phenomena over time

Infrequent pattern recognition and first level diagnostic



Decision rules for Case study 1.

Time window	Decision rules	Symbol	Accuracy
Period 1 (00:00–04:59)	IF ***system_start*** = is turned OFF	→ ***a***	98%
	IF ***system_start*** = is turned ON at 04:00 a.m. AND ***T_int*** ≥ 23,43 °C	→ ***a***	80%
	IF ***system_start*** = is turned ON at 04:00 a.m. AND ***T_int*** < 23,43 °C	→ ***b***	79%
Period 2 (05:00–06:59)	IF ***Day*** = Holiday OR Sunday OR Saturday	→ ***a***	83%
	IF ***Day*** = Monday OR Tuesday OR Wednesday OR Thursday OR Friday AND ***T_int_pre (period 1)*** ≥ 23,55 °C	→ ***c***	88%
	IF ***Day*** = Monday OR Tuesday OR Wednesday OR Thursday OR Friday AND ***T_int_pre (period 1)*** < 23,55 °C	→ ***d***	60%
Period 3 (07:00–13:59)	IF ***Day*** = Holiday OR Sunday	→ ***a***	99%
	IF ***Day*** = Saturday	→ ***b***	77%
	IF ***Day*** = Monday OR Tuesday OR Wednesday OR Thursday OR Friday AND 9 °C ≤ ***T_ext*** < 20,35 °C	→ ***d***	73%
	IF ***Day*** = Monday OR Tuesday OR Wednesday OR Thursday OR Friday AND ***T_ext*** ≥ 20,35 °C	→ ***e***	98%
	IF ***Day*** = Monday OR Tuesday OR Wednesday OR Thursday OR Friday AND ***T_ext*** < 9 °C	→ ***e***	84%
Period 4 (14:00–19:59)	IF ***Sym_pre*** = a OR b OR c	→ ***a***	96%
	IF ***T_ext*** < 24,1 °C AND ***Sym_pre (period 3)*** = "d" AND ***T_int*** < 25,55 °C	→ ***c***	69%
	IF ***T_ext*** < 24,1 °C AND ***Sym_pre (period 3)*** = "d" AND ***T_int*** ≥ 25,55 °C	→ ***d***	75%
	IF ***T_ext*** < 24,1 °C AND ***Sym_pre (period 3)*** = "e"	→ ***d***	94%
	IF ***Sym_pre*** = "d" OR "e" AND ***T_ext (period 3)*** ≥ 24,1 °C	→ ***e***	79%
Period 5 (20:00–23:59)	–	→ ***a***	95%

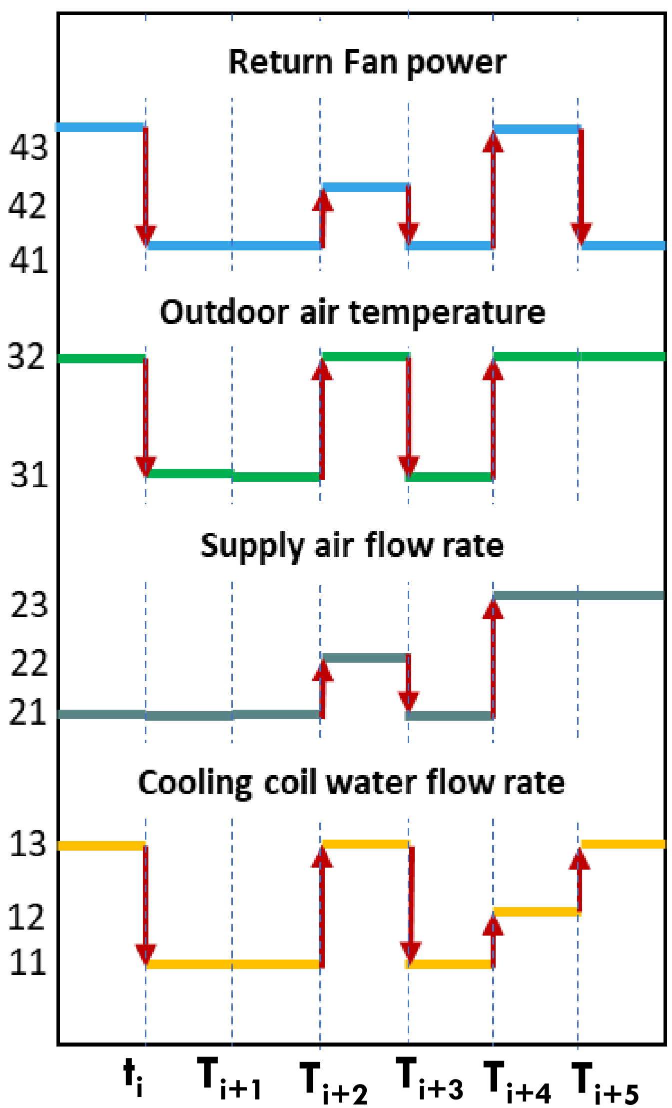


Source: Capozzoli A, Piscitelli M S, Brandi S., Grassi D., Chicco G. Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. Energy 2018

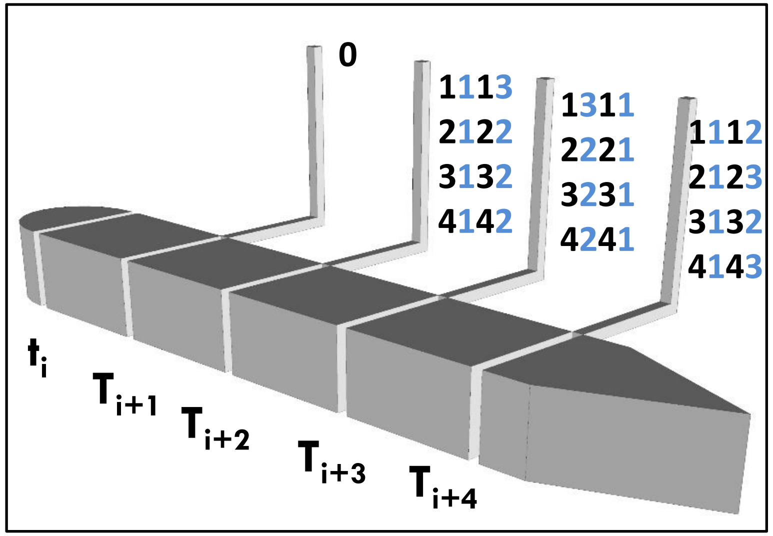
Temporal abstraction in multivariate time series

Sub-hourly aggregation

Encoded time series (aSAX)

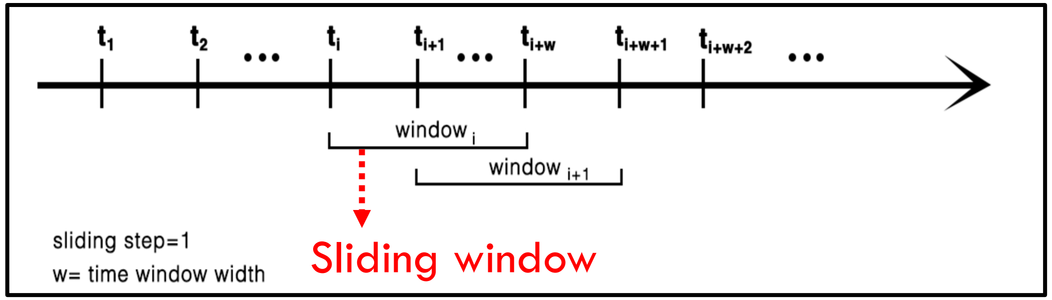


Transitions time series

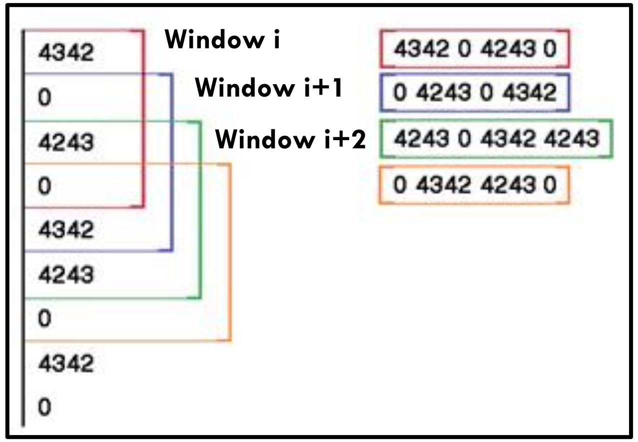


The transitions of the variables are gathered in a **unique time series,** reporting the encoding of the events present instant by instant. The 0 represents the absence of events, while the number identifies the transition specifying the **amplitude and the direction of the change.**

Transactions identification trhough Sliding windows



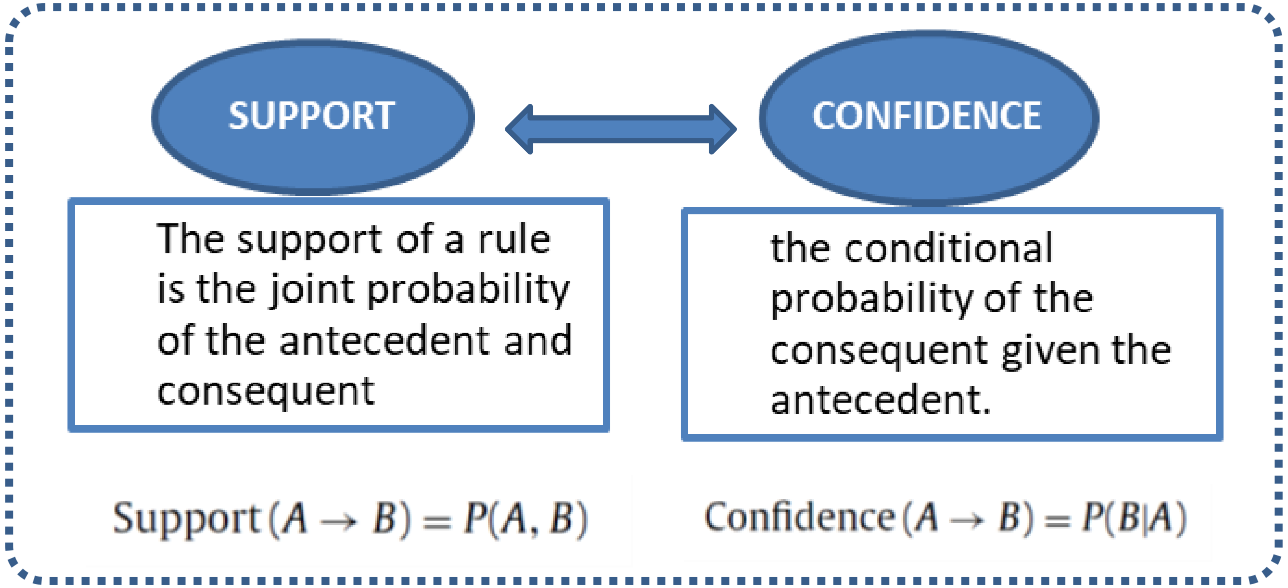
Transactions Databse



Temporal Association Rule Mining (TARM)

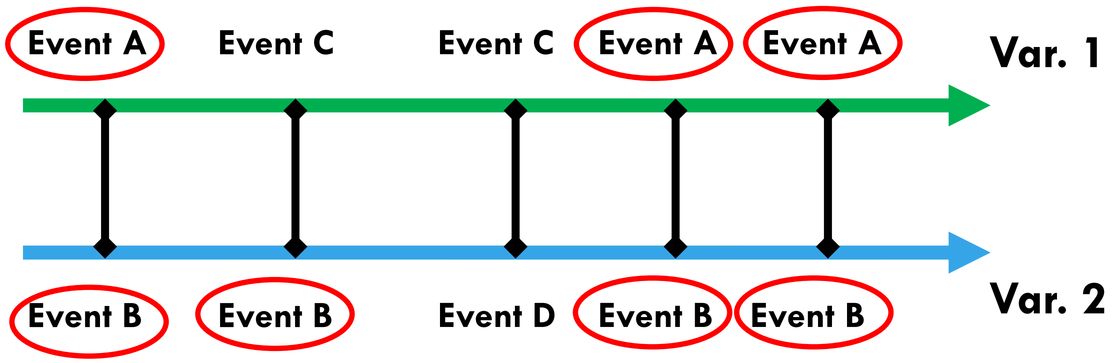
In the context of **discrete-state-transactions,** association rules can be used for mining **co-occurrences or implications between events** in the time domain that are frequently associated together.
The output is a set of rules that are used to represent patterns (if A happens, B will also happen, A is called **antecedent** and B is the **consequent**).

Rules quality metrics



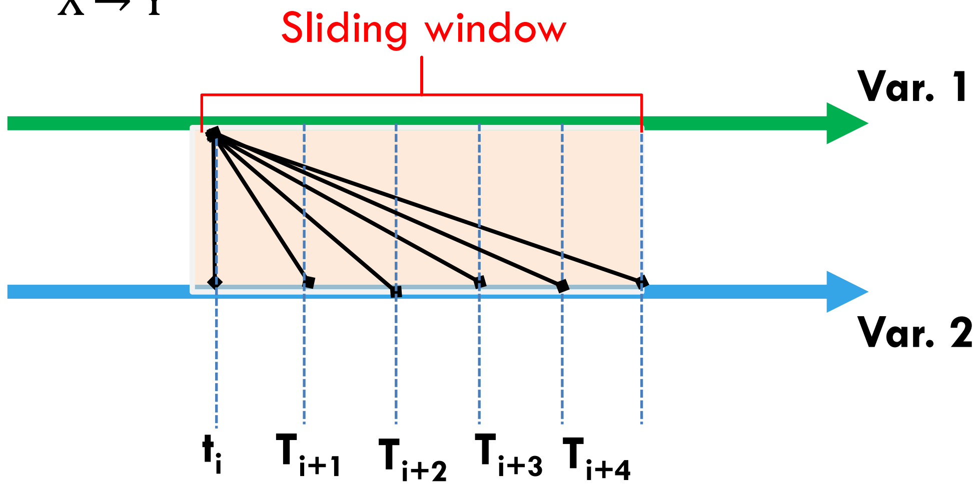
Co-occurrence

$$X \xrightarrow{t} Y$$



Temporal implication

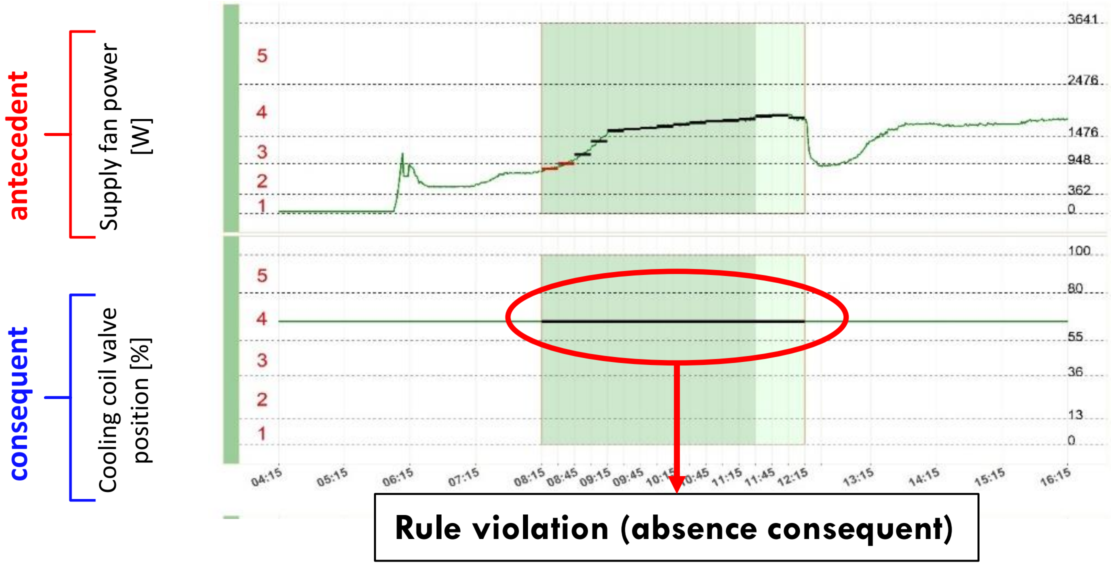
$$X \xrightarrow{t} Y$$



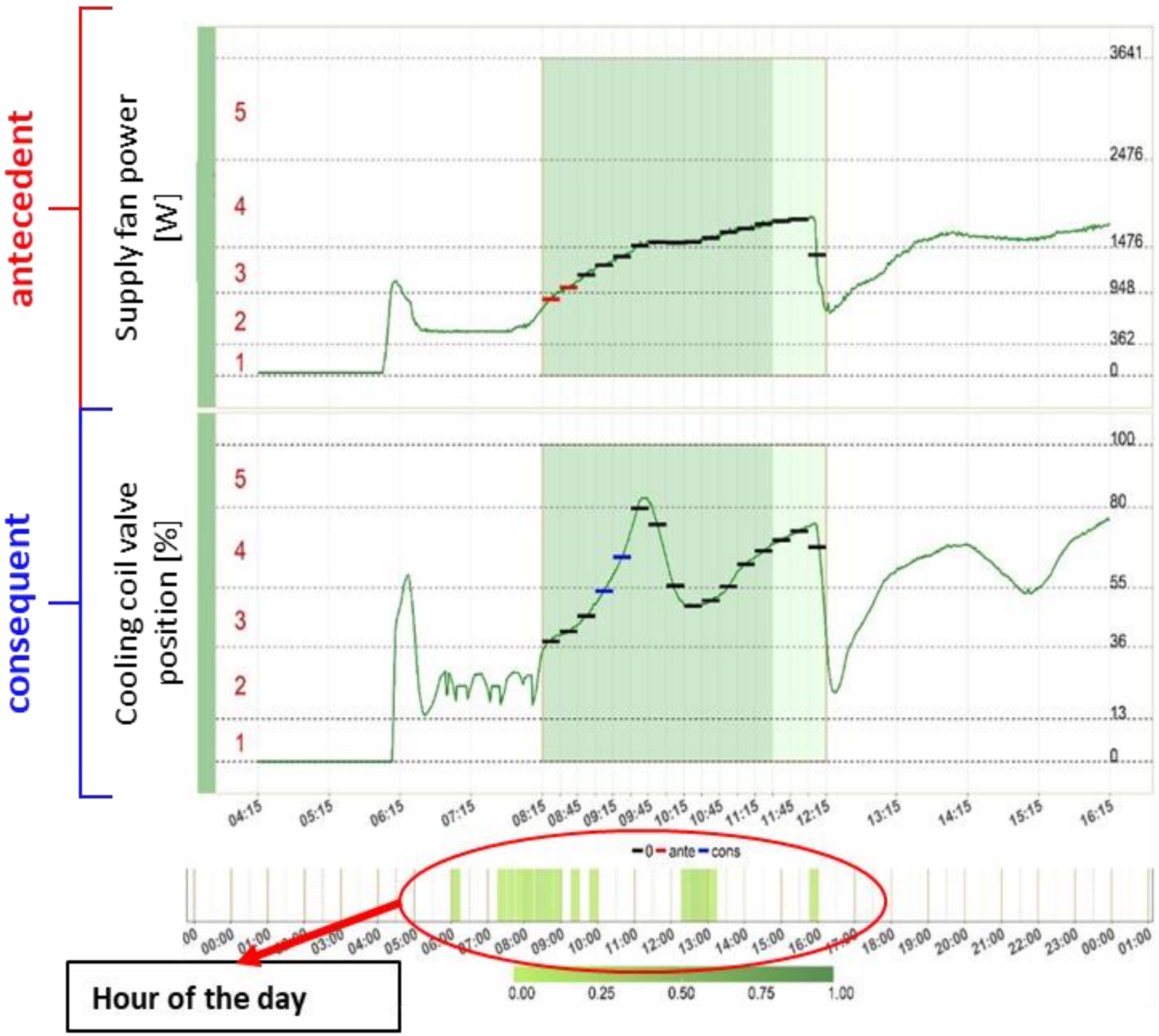
TARM for Fault detection and diagnosis in AHU

1. In fault detection and diagnosis, TARM makes it possible to extract patterns in the time domain representative of the relation between disturbances and energy demand and then detecting deficiencies when these patterns are **violated over time**.

2. The great advantage of TARM is its capability in learning and **characterizing thermal dynamics features for unforeseen systems without a-priori knowledge** of their configuration.



Temporal Association Rule n. 48 – system general operation rule



Conclusions & Final Remarks

Building physics vs data science or data science for building physics?
In the field of energy and buildings, a robust background in building physics represents a cornerstone in order to extract useful and non-trivial knowledge

EMS and Data Analysis = energy saving: is it always true?
The development of a smart energy management system if not correctly designed can result in a high cost investment with incompatible payback.

Rapid executions but ...
The preparation and pre-processing of data represent complex operations that require very long time and experience to the analyst (Data pre-processing might take 80% of the total data mining effort)

"In theory, there is no difference between theory and practice….But, in practice, there is".
Real world cases of implementation data analytics based frameworks in BEMS still remain a demanding task

Smart city: Everything belongs to everyone but nothing belongs to anybody
In the smart city context data are shared between different systems and stakeholders. Security and privacy issues are fundamental aspects to be considered with extreme care.

From reactive to predictive energy management in buildings. A predictive management provides a scheduled optimisation opportunities which can enable high level supporting recommendations to be implemented.

Improving energy management in buildings through data analytics: challenges and opportunities

alfonso.capozzoli@polito.it

marco.piscitelli@polito.it

Speakers:

Prof. Alfonso Capozzoli

Eng. Marco Savino Piscitelli

Hong Kong

05/12/2018