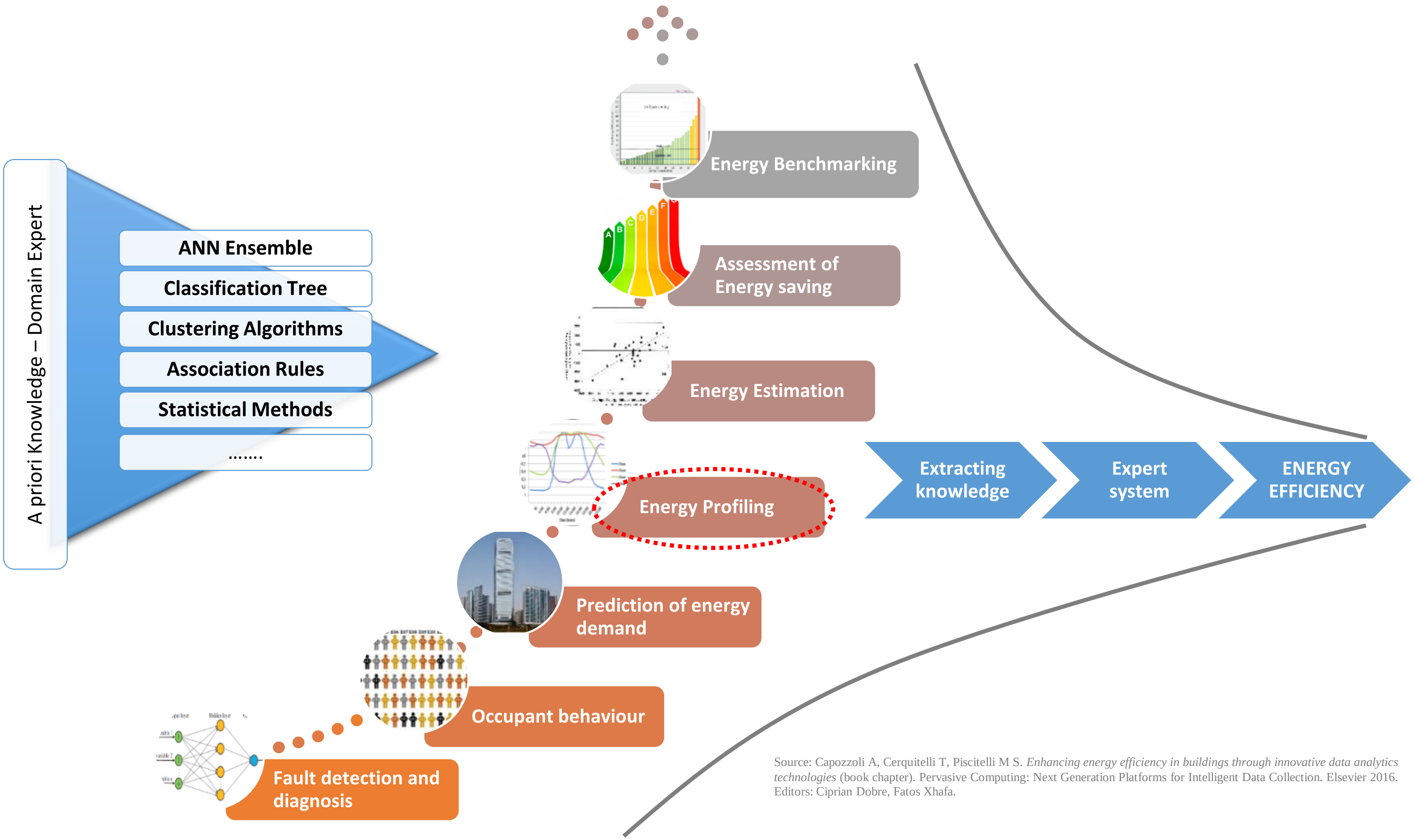
Mining typical load profiles in buildings to support energy management in the smart city context

Alfonso Capozzoli , Marco Savino Piscitelli, Silvio Brandi

Politecnico di Torino, DENERG, TEBE Research Group, corso Duca degli Abruzzi 24, Turin, 10129, Italy

DATA ANALYTICS IN BUILDINGS



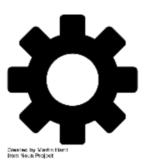
Source: Capozzoli A, Cerquitelli T, Piscitelli M S. *Enhancing energy efficiency in buildings through innovative data analytics technologies* (book chapter). Pervasive Computing: Next Generation Platforms for Intelligent Data Collection. Elsevier 2016. Editors: Ciprian Dobre, Fatos Xhafa.

TIME SERIES DATA GROWTH IN BUILDINGS

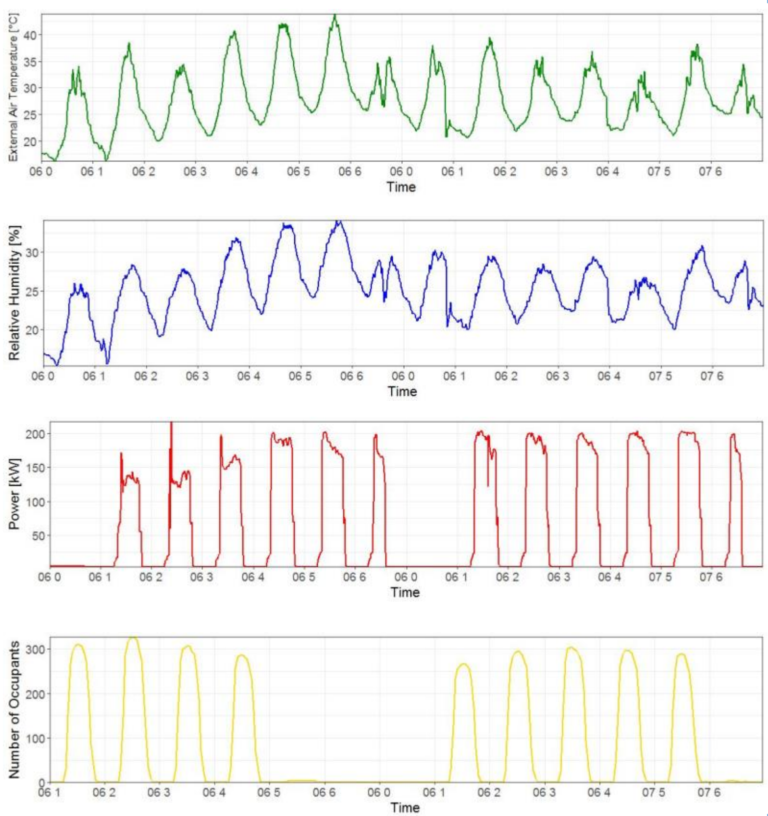
The increasing implementation of ICT and EMS in the current ***paradigm of smart buildings in smart cities*** has enabled an easier availability of a huge amount of heterogeneous and complex building-related data in form of time series.

Climatic data

Operational data



User related data



ENERGY PROFILING

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**.

ENERGY PROFILING OBJECTIVES

When a stock of buildings is analysed the main objective of energy profiling is ***load classification*** to discover homogeneous classes of buildings/customers according to the concept of load profiles similarity.

On the other hand, energy profiling at individual building level is aimed at supporting ***detailed diagnostic analyses*** (e.g., energy demand prediction, fault detection and diagnosis (FDD), energy benchmarking) performed at sub-system or whole building level.

OBJECTIVE

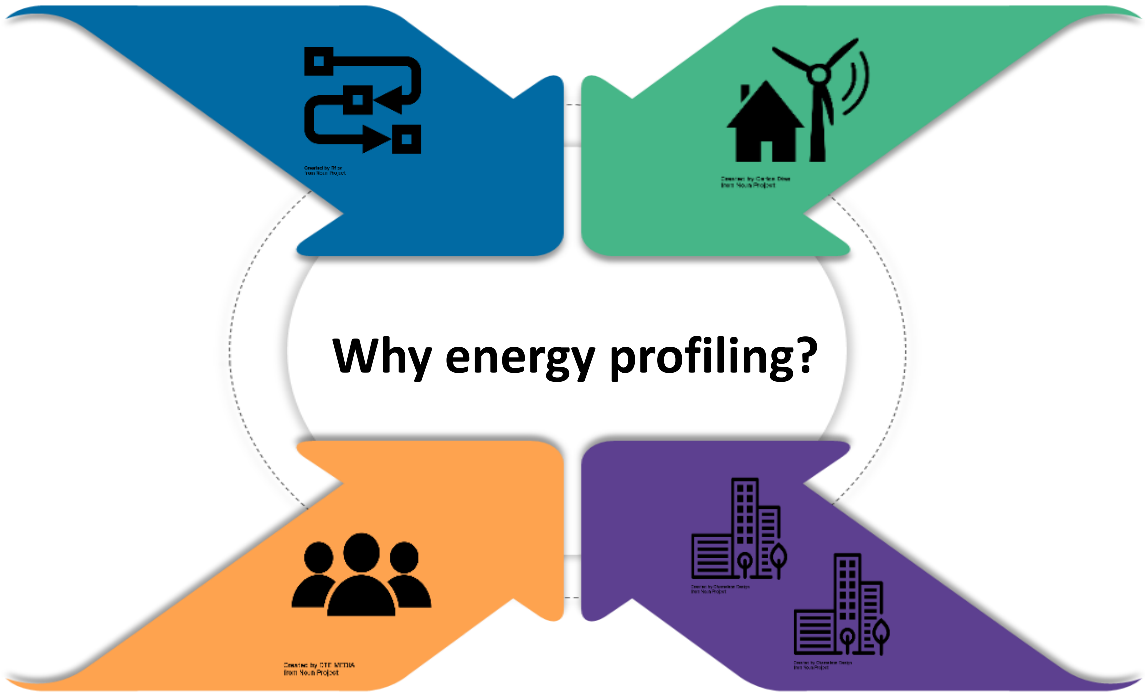
Detailed diagnostic analysis at single building/customer level

Load profiles classification at multiple buildings/customers level

KEY QUESTIONS TO BE EXPLORED

1 Which are the main **methodological steps** of energy profiling?

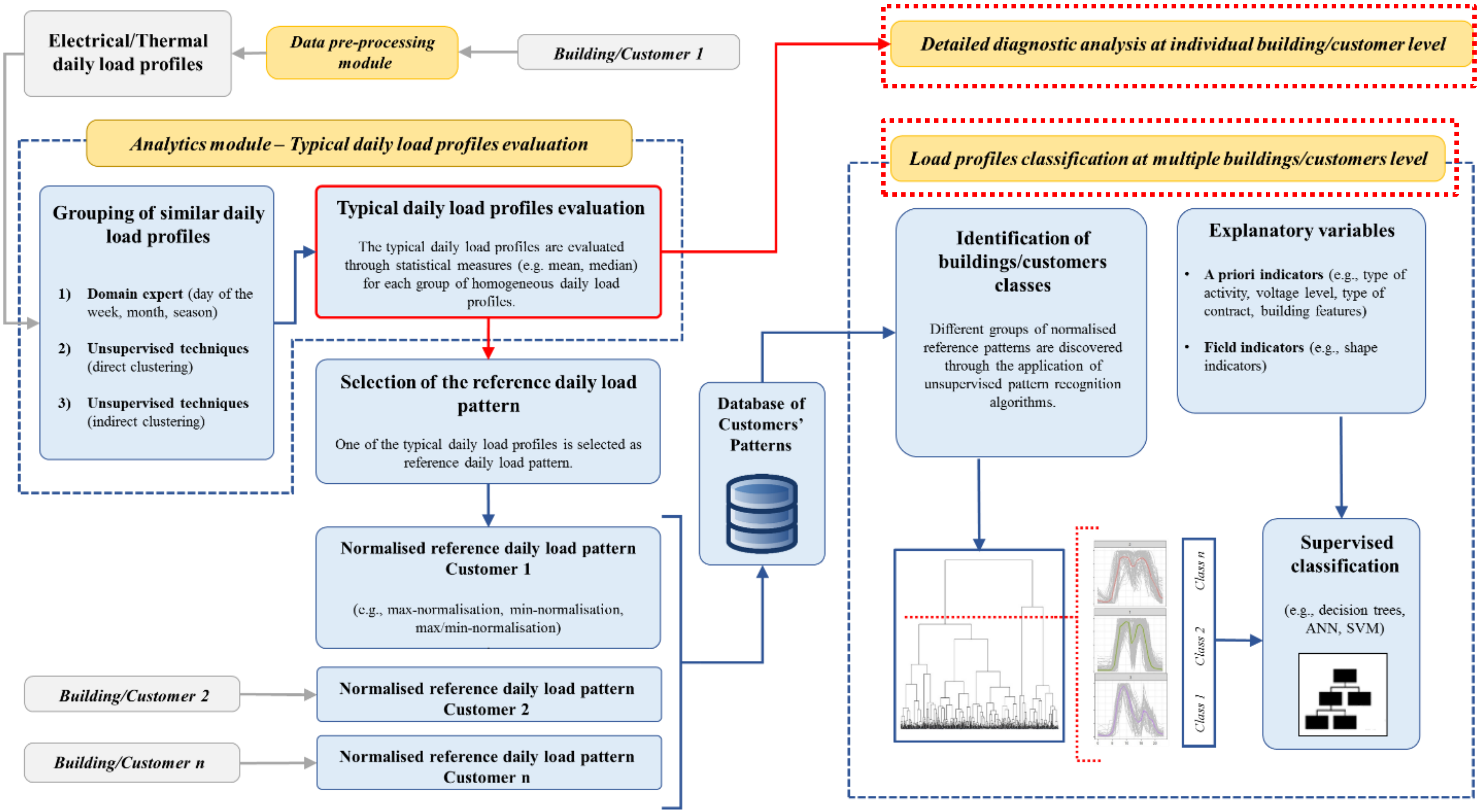
2 Which are the **possible implications** of the energy profiling process at **individual building**?



4 Who are the **different actors** in the smart city environment for which the process of energy profiling could be beneficial?

3 Which are the **possible implications** of the energy profiling process at **buildings stock level** in a smart city environment?

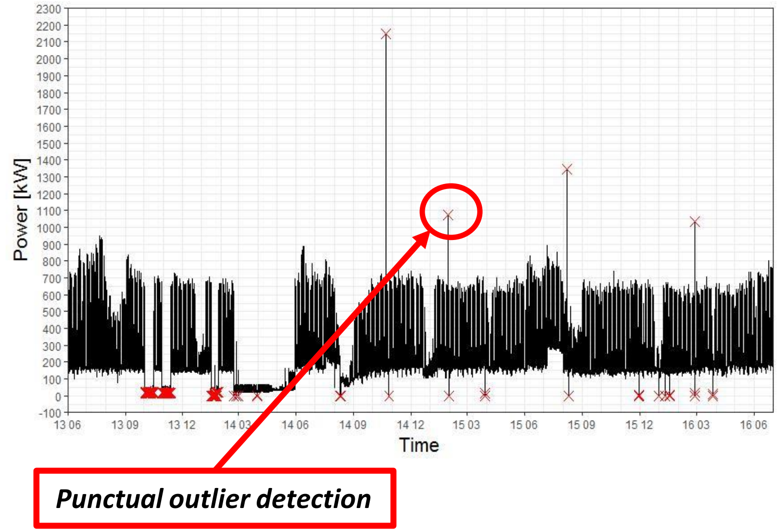
GENERAL FRAMEWORK FOR LOAD PROFILES CHARACTERISATION



DATA PRE-PROCESSING

In a first step, the collected raw data in form of time series are analysed through different statistical methods to identify potential **missing values and punctual outliers** that must be replaced or removed.

DATA CLEANING



TIME-SERIES CHUNKING

In a second step, the original **time series is chunked in fixed length windows** (sub-sequences). The sub-sequences, representing the daily load profiles, are **organized into a MxN matrix** where M is the number of daily load profiles while N depends from the data granularity.

TYPICAL DAILY LOAD PROFILES

This phase of the framework is performed at individual building/customer level and it is aimed at **identifying groups of homogenous profiles** through a data segmentation phase.

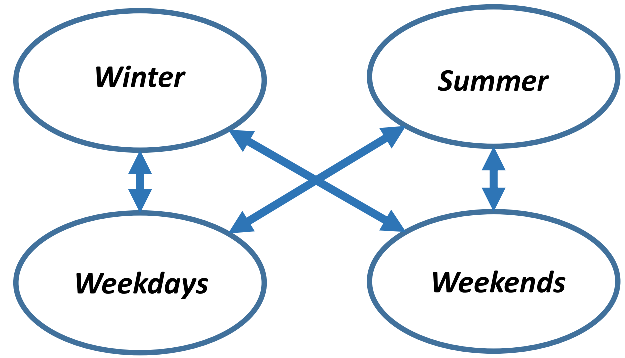
The typical profiles can be then evaluated through statistical measures (e.g. mean, median) calculated in each group of homogenous daily load profiles identified in the data segmentation phase. To this purpose, data segmentation may be performed following:

1. **Domain expert based approach.**
2. **Data mining approach by using unsupervised techniques.**
3. **Indirect clustering through data reduction methods.**

DATA SEGMENTATION

Domain expert based approach

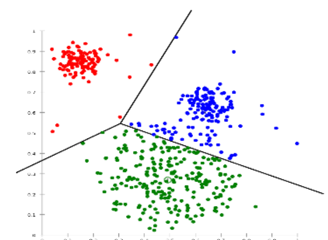
It is **completely driven by the domain knowledge of the analyst** and it is aimed at generating subsets of daily load profiles that are supposed to be subjected to homogenous boundary conditions.



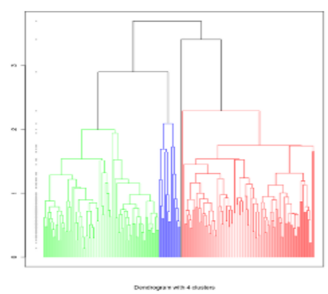
Data mining approach by using unsupervised techniques

Unsupervised pattern recognition techniques such as **cluster analysis** allows load patterns to be identified in a not pre-determined time domain.

K-means



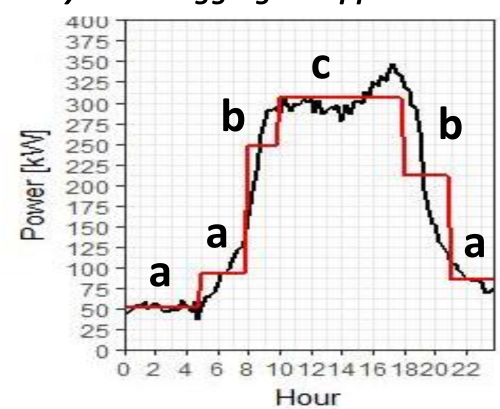
Hierarchical clustering



Indirect clustering through data reduction methods

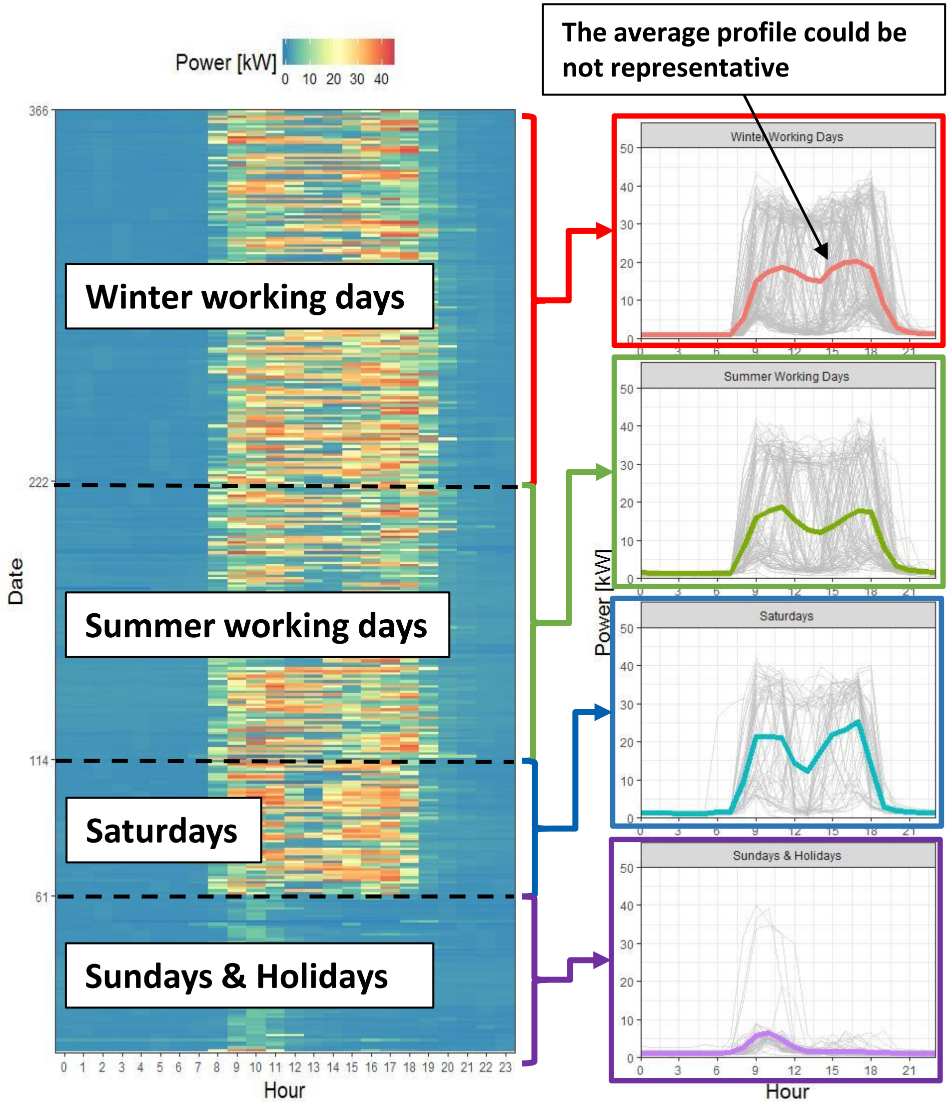
A further approach for the data segmentation and profiles extraction relies on indirect clustering, where the **object of clustering are features extracted from the load profiles.**

Symbolic Aggregate approXimation

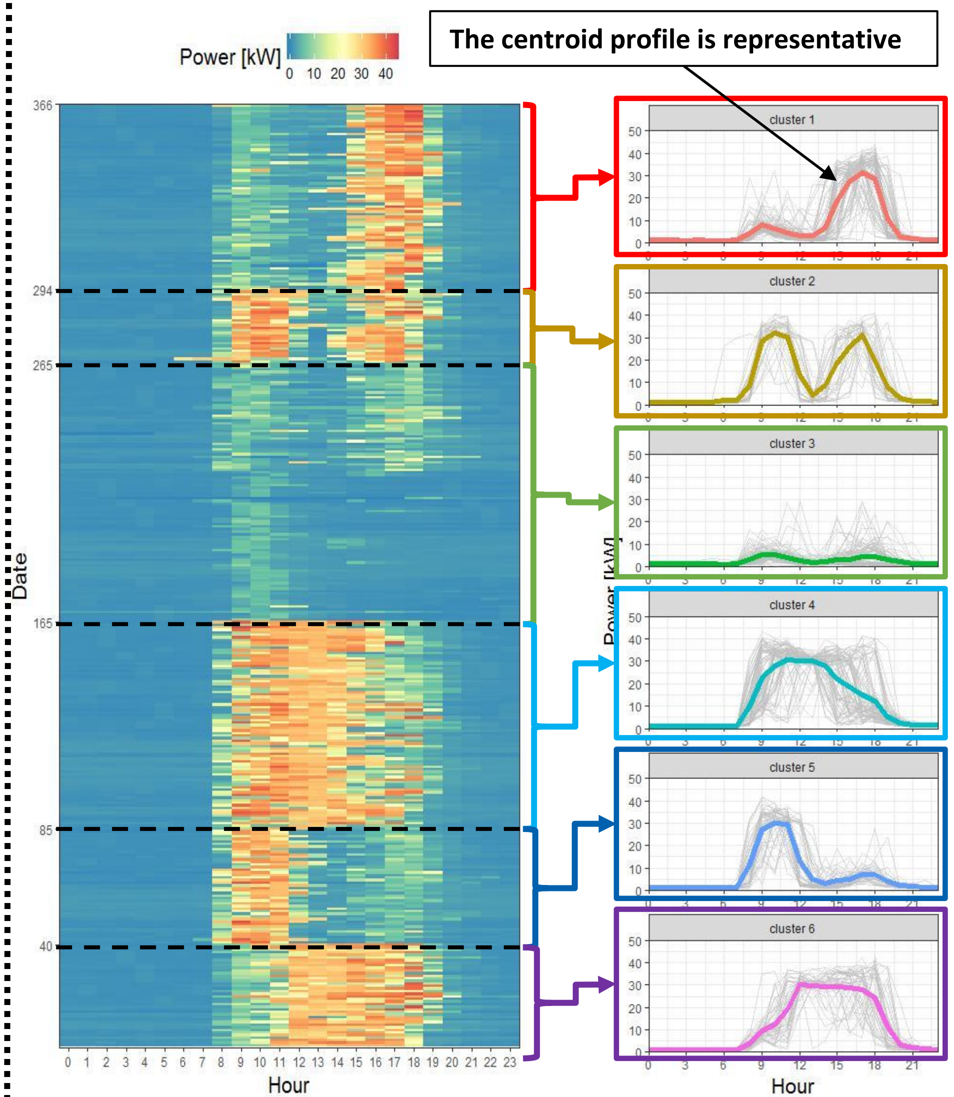


WHICH DATA SEGMENTATION?

Domain expert based approach



Unsupervised pattern recognition



Alfonso Capozzoli, Marco Savino Piscitelli, Silvio Brandi

Mining typical load profiles in buildings to support energy management in the smart city context

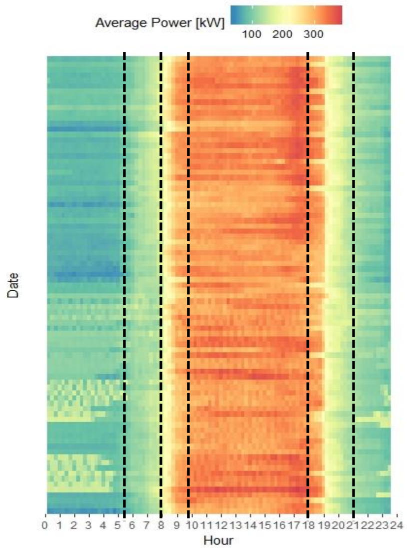
Ninth International conference on:
Sustainability in Energy and Buildings (SEB-17)

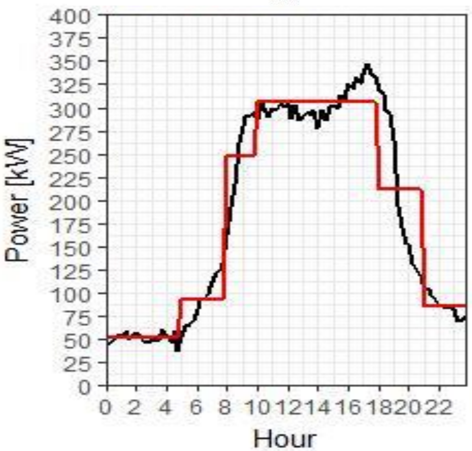
DETAILED DIAGNOSTIC ANALYSIS AT INDIVIDUAL BUILDING LEVEL

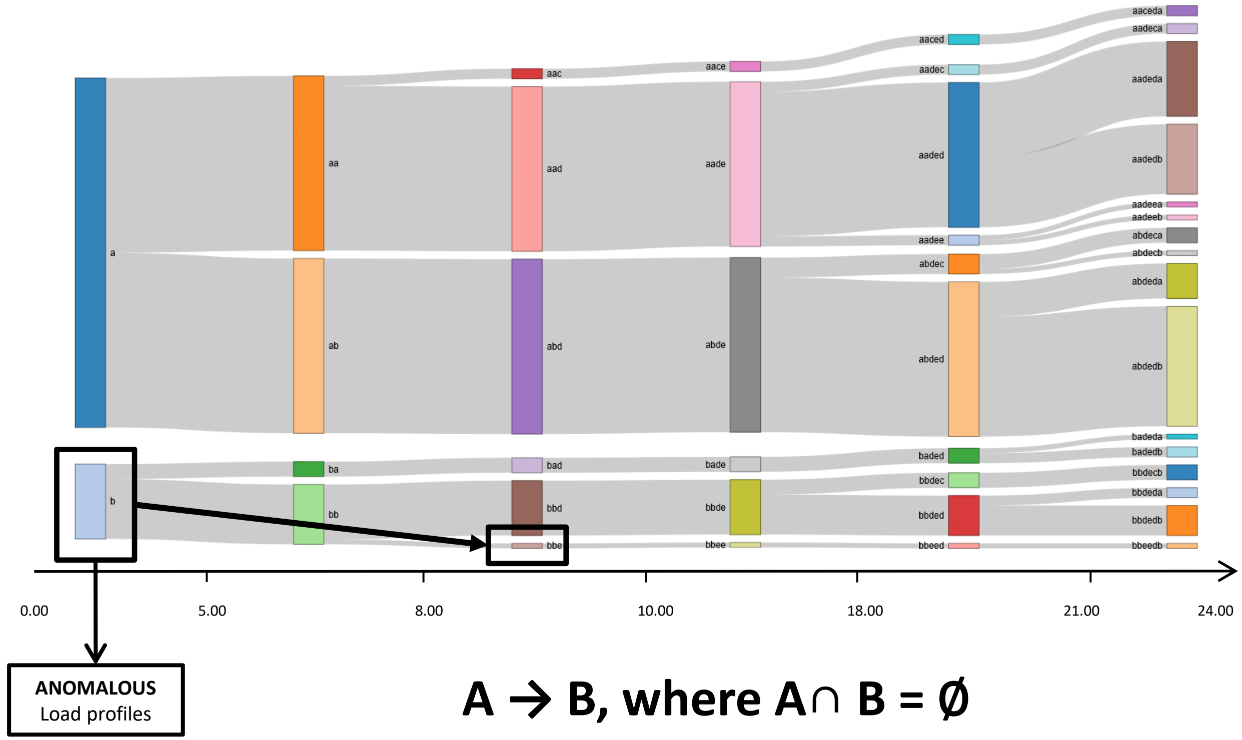
The knowledge of typical load profiles at single building/system level offers the opportunity to address complex emerging issues in **energy management at individual building level**.

- Improve the accuracy and robustness of energy consumption **forecasting models**.
- Provide relevant information for the **calibration** of simulation models.
- Implementation of **Fault Detection and Diagnosis** (FDD) strategies.
- Energy **benchmarking** over time.
- Exploitation of **on-site renewable energy** sources production.
- Support the **optimal operation** of a building at multiple levels Active **demand response** applications.

DETAILED DIAGNOSTIC ANALYSIS AT INDIVIDUAL BUILDING LEVEL





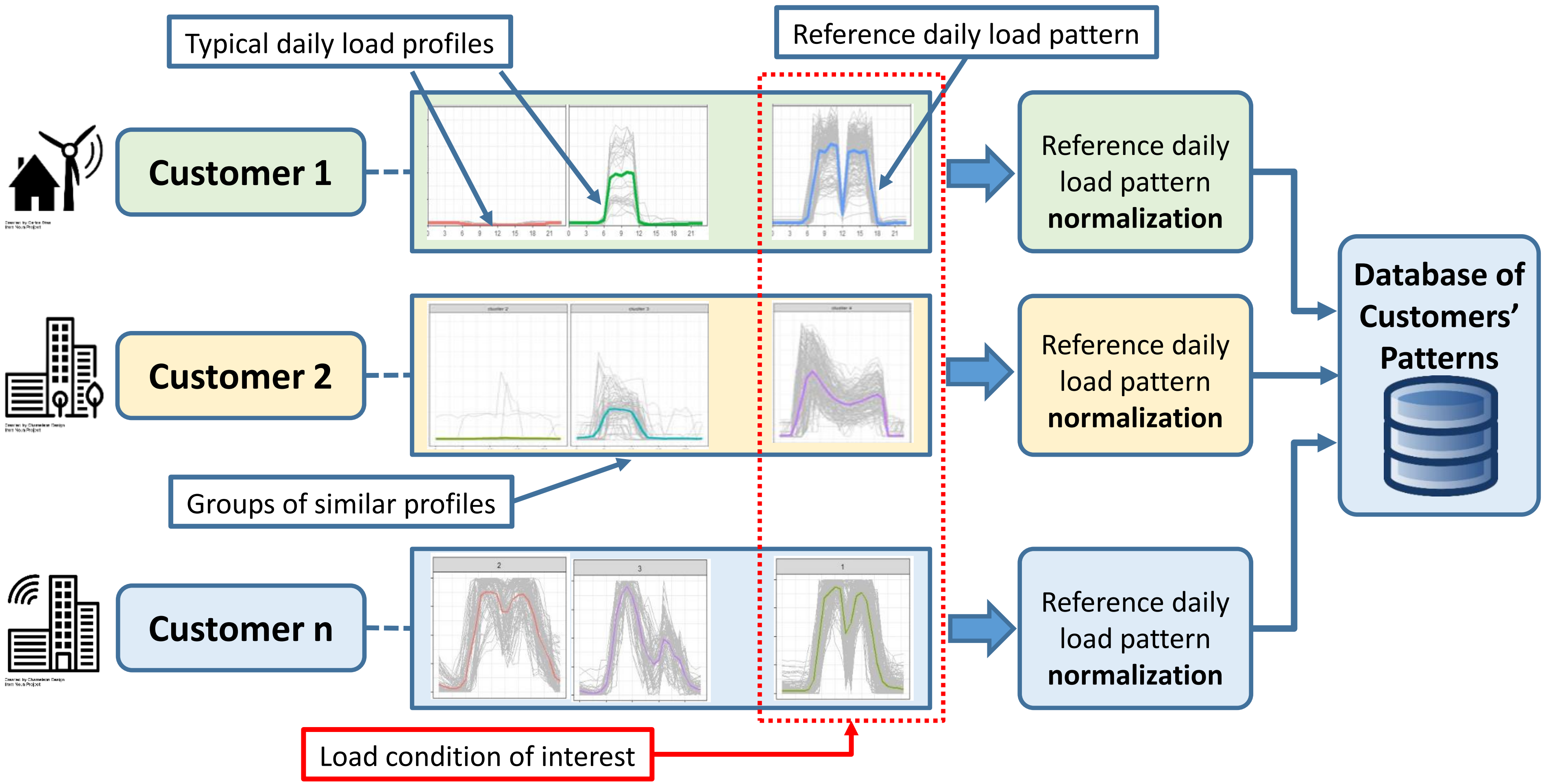


$$A \rightarrow B, \text{ where } A \cap B = \emptyset$$

Association Rule Mining is a data mining method to identify all associations and correlations between attribute values over time. The output is a set of rules that are used to represent patterns of attributes that are frequently associated together (if A happens, B will also happen, A is called **antecedent** and B is the **consequent**).

EXTENTION TO STOCK OF BUILDINGS

Moreover, the **mining of typical load profiles** in buildings could be considered a **preliminary phase in customers' classification**.



LOAD PROFILES CLASSIFICATION

After the selection and normalization/standardization of the reference load pattern for each customer/building, they are processed in order to **discover typical classes** of customers/buildings and classify them according to appropriate variables.

The whole process consists of three different steps:

- **Identification of *n* customer classes** of buildings/customers.
- Definition of the normalised **reference load pattern for each customers' class** (e.g. centroid).
- **Enrichment of the database** with additional attributes (categorical or numerical) for each load profile to perform a **supervised classification** process.

UNSUPERVISED PATTERN RECOGNITION

The identification of *n* customer classes of buildings/customers unfolds through the **application of unsupervised pattern recognition techniques** such as hierarchical or partitive cluster analysis.

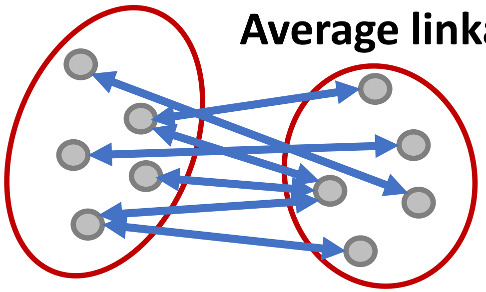
Task	Method	Reference
Unsupervised Pattern Recognition Algorithms	Kohonen Self-Organising-Maps (SOM)	[10, 18, 33]
	K-means clustering (KM)	[10, 11, 13, 14, 18, 22, 33]
	Fuzzy K-Means clustering (FKM)	[17, 18, 20, 22, 29]
	K-Shape clustering	[31]
	K-medoids / Partition Around Medoids (PAM)	[14]
	Hierarchical - Single-linkage (SL)	[17, 22, 30]
	Complete-linkage (CL)	[17, 22]
	Hierarchical - Average linkage - Unweighted Pair Group Method Average (UPGMA)	[17, 18, 22]
	Hierarchical - Weighted Pair Group Method Average (WPGMA)	[22]
	Hierarchical - Centroid linkage -Unweighted Pair Group Method Centroid (UPGMC)	[17, 22]
	Hierarchical - Median linkage - Weighted Pair Group Method Centroid (WPGMC)	[22]
	Hierarchical - Ward-linkage (WARD)	[17, 18, 22]
	Follow the Leader clustering (FLD)	[17, 18]

Hierarchical clustering algorithms

Refereces referred to the paper

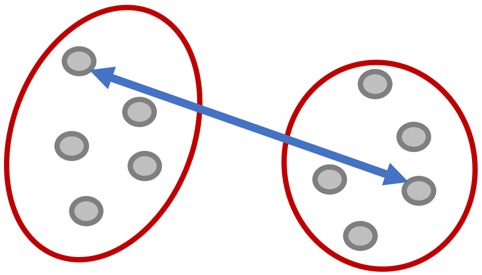
WHICH LINKAGE TYPE?

Average linkage



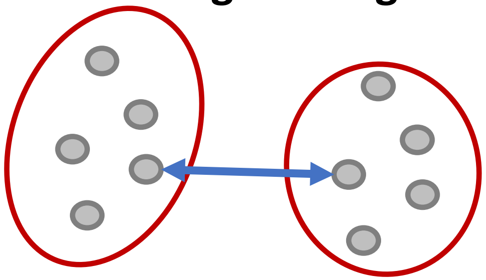
$$L(r,t) = \frac{1}{n_r n_t} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} D(x_{ri}, x_{tj})$$

Complete linkage



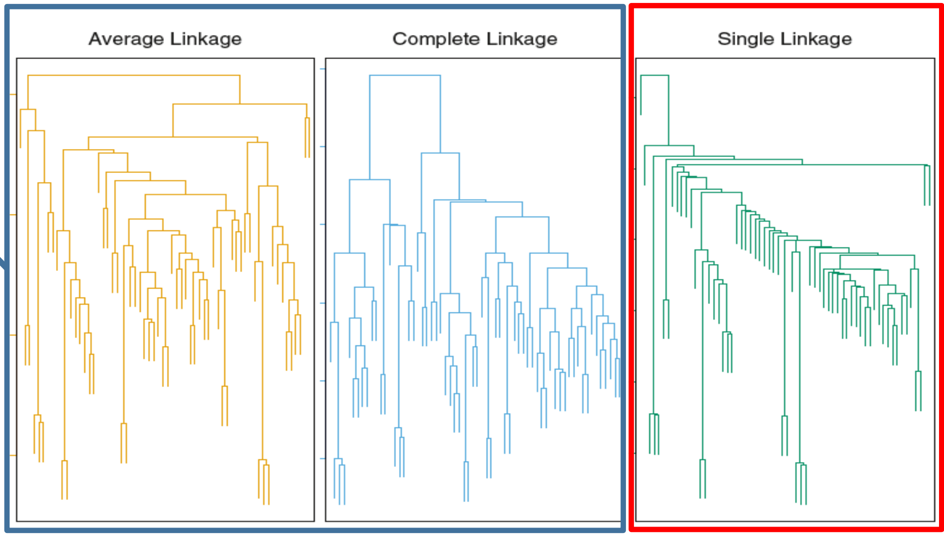
$$L(r,t) = \max(D(x_{ri}, x_{tj}))$$

Single linkage



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

Average and **Complete** produce more balanced dendrograms.

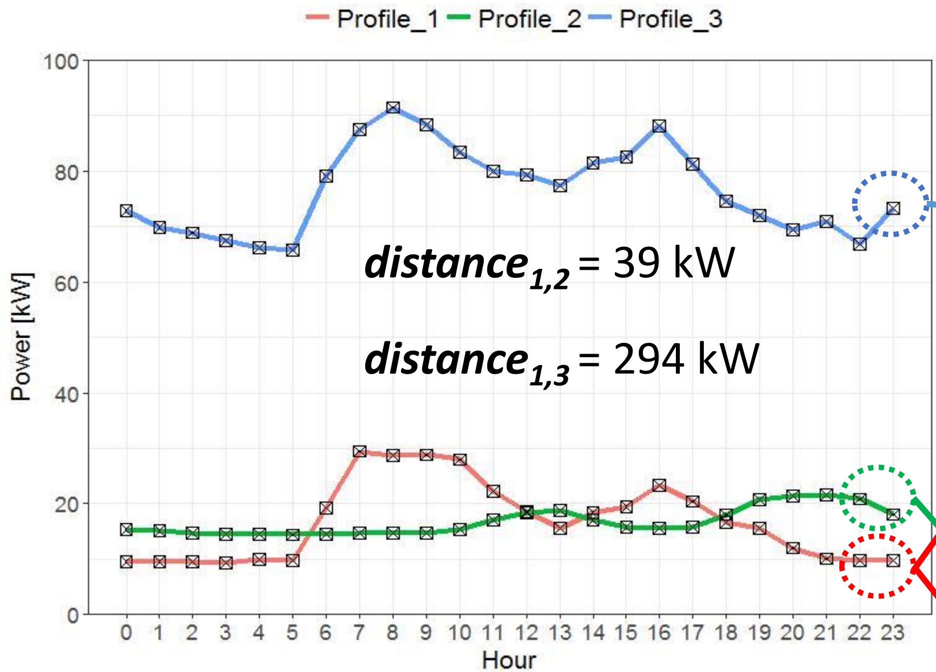


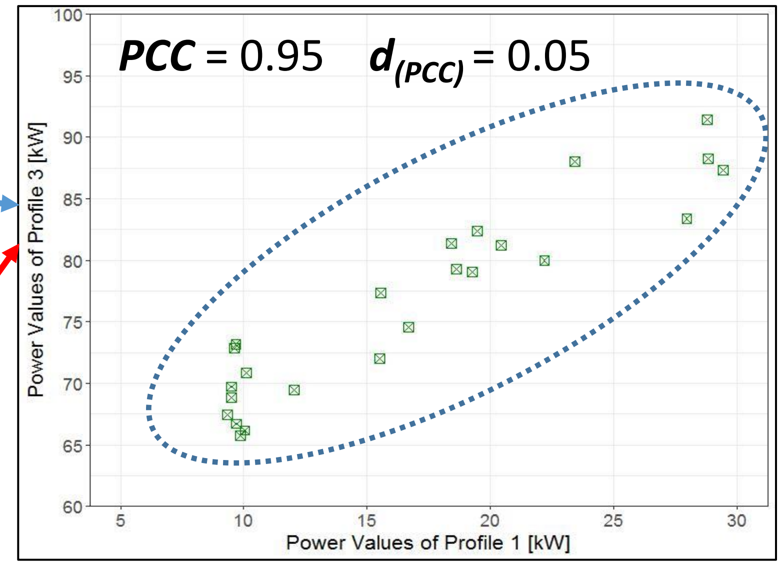
Single is more sensitive to outliers.

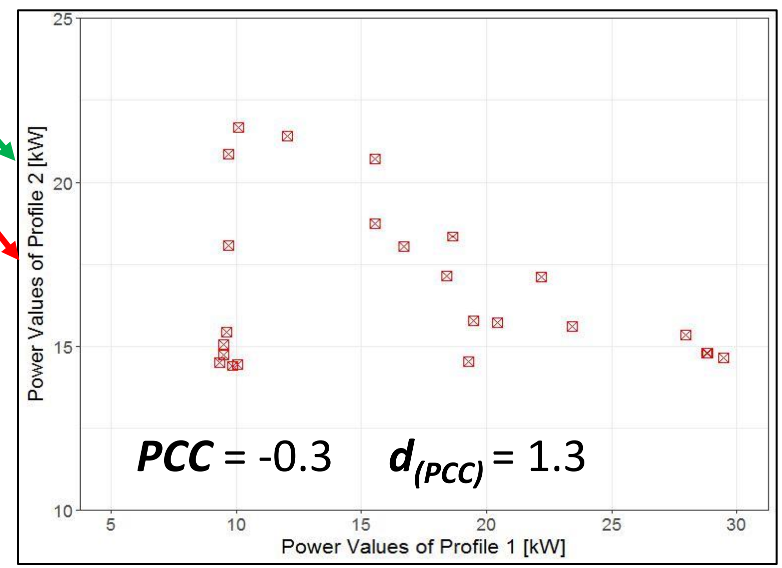
Alfonso Capozzoli, Marco Savino Piscitelli, Silvio Brandi
Mining typical load profiles in buildings to support energy management in the smart city context

Ninth International conference on:
Sustainability in Energy and Buildings (SEB-17)

WHICH DISSIMILARITY MEASURES?







Euclidean distance

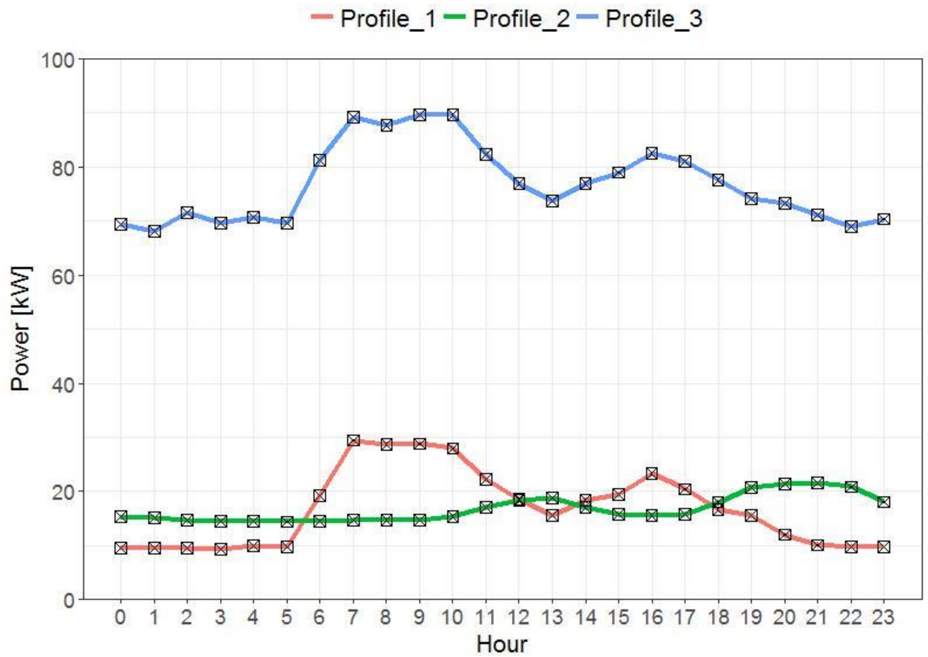
$$d_{ED}(X,Y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2}$$

Pearson correlation coefficient (PCC)

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

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

WHICH DISSIMILARITY MEASURES?

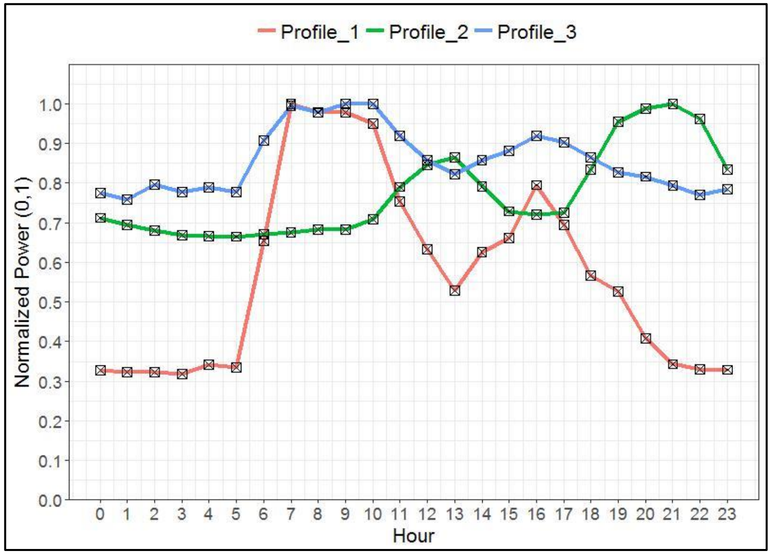


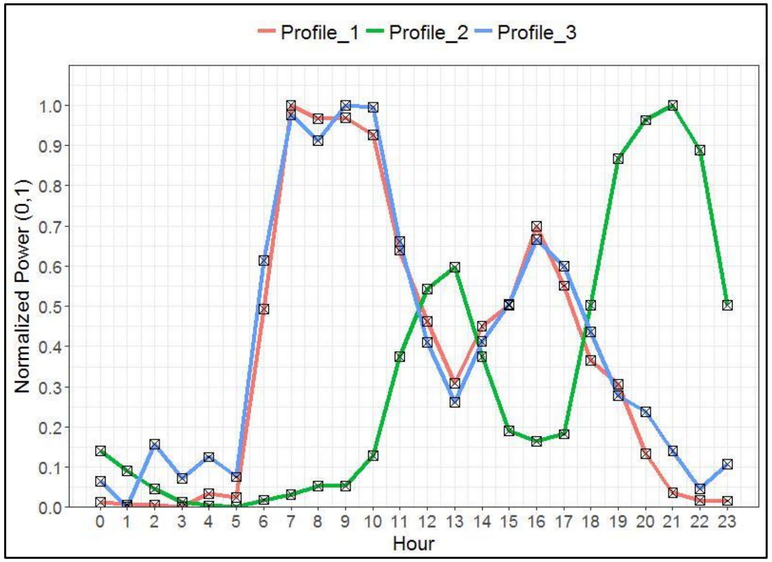
Normalization on daily maximum value

$$x_{i,norm,max} = \frac{x_i}{\max(x)}$$

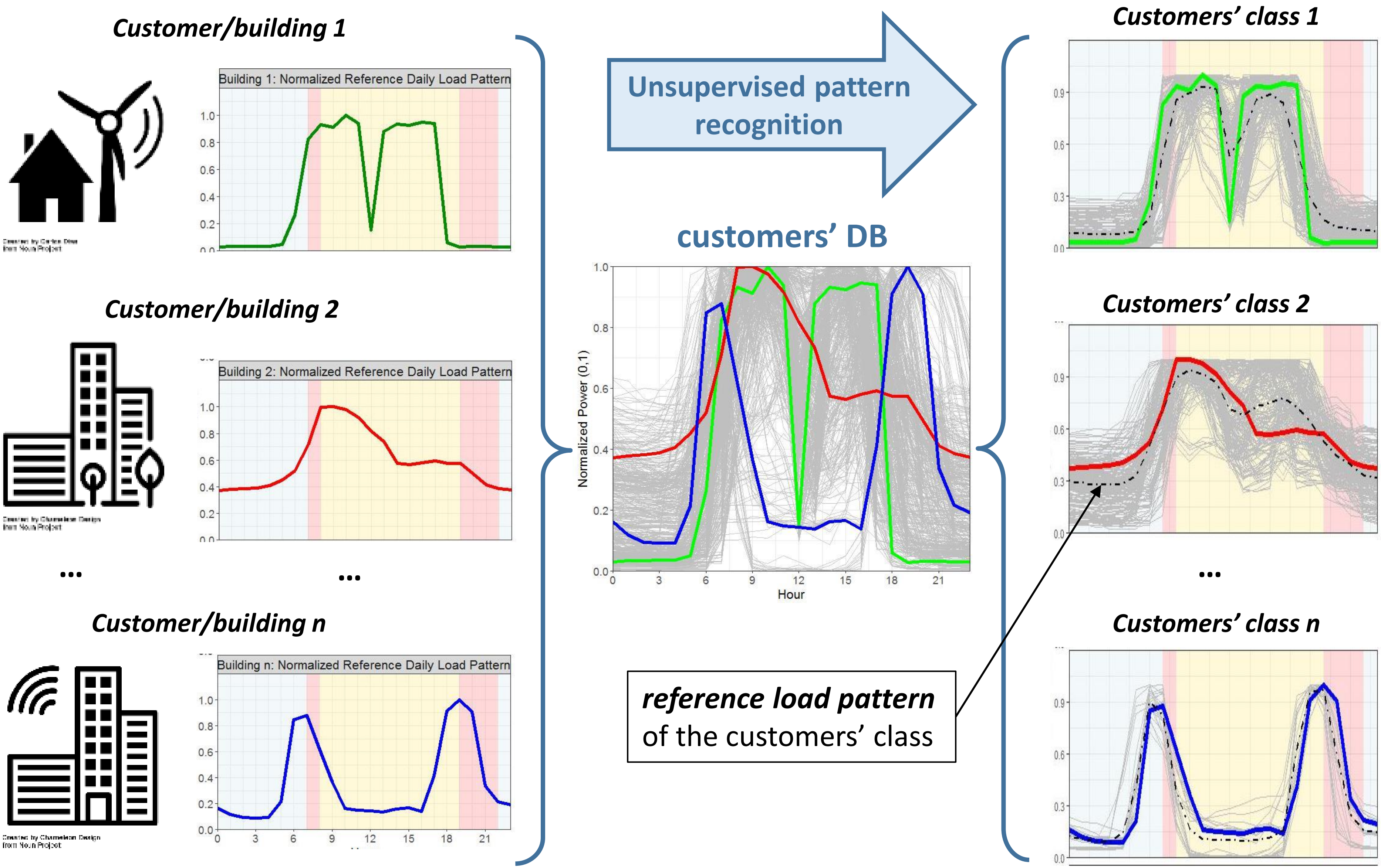
Normalization between daily maximum and minimum value

$$x_{i,norm,min-max} = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$





CUSTOMERS' CLASSES IDENTIFICATION



Alfonso Capozzoli, Marco Savino Piscitelli, Silvio Brandi
Mining typical load profiles in buildings to support energy management in the smart city context

Ninth International conference on:
Sustainability in Energy and Buildings (SEB-17)

SUPERVISED CLASSIFICATION

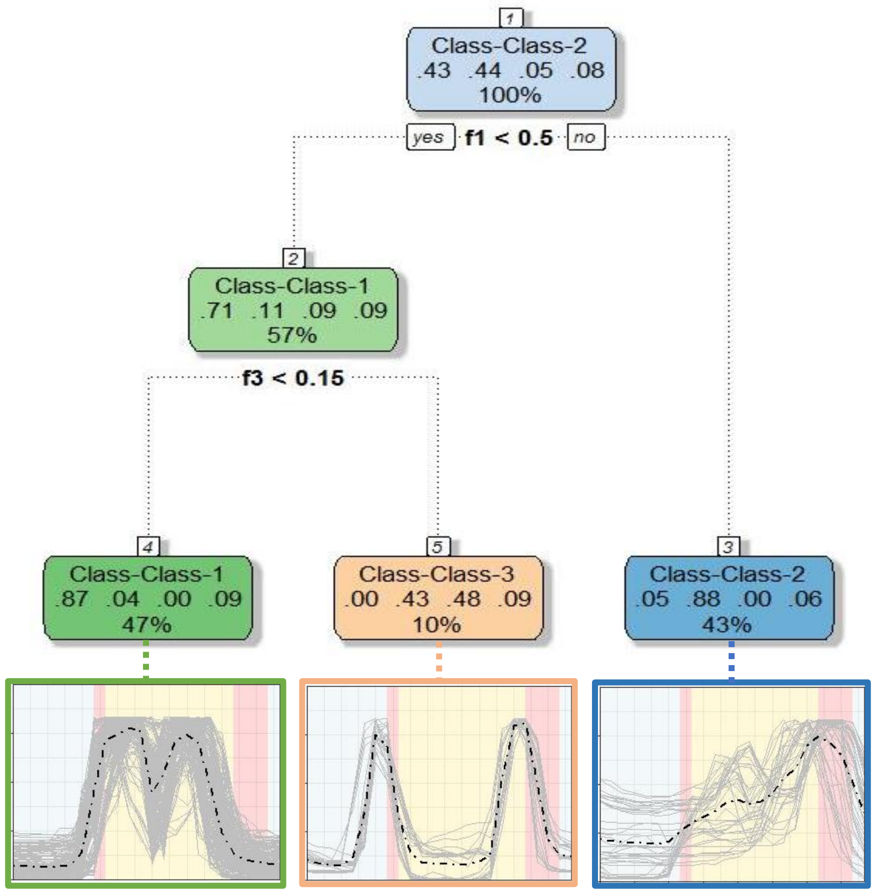
The **customers' class** label is defined **as a categorical dependent variable** which can be predicted with a classification model using additional attributes for the **supervised classification process**.

Explanatory variables

- **A priori indicators** (e.g., type of activity, voltage level, type of contract)
- **Field indicators** (e.g., shape indicators)

Parameter	Definition	Acquisition period
Daily P_{av}/P_{max}	$f_1 = P_{av,day}/P_{max,day}$	1 day
Daily P_{min}/P_{max}	$f_2 = P_{min,day}/P_{max,day}$	1 day
Night Impact	$f_3 = 1/3 P_{av,night}/P_{av,day}$	1 day (8 h night, from 11 p.m. to 6 a.m.)
Lunch impact	$f_4 = 1/8 P_{av,lunch}/P_{av,day}$	1 day (3 h lunch, from 12 a.m. to 15 p.m.)
Daily P_{min}/P_{av}	$f_5 = P_{min,day}/P_{av,day}$	1 day

Adapted from: Chicco G, Napoli R, Postolache P, Scutariu M, Toader C. Customer characterization options for improving the tariff offer. IEEE Trans Power Syst 2003; 18: 381–387.



ENERGY PROFILING IN STOCK OF BUILDINGS: IMPLICATIONS

The described methodology represents a **robust and useful tool** to easily estimate for a new statistical object its membership to a specific class of customers/buildings in order to:

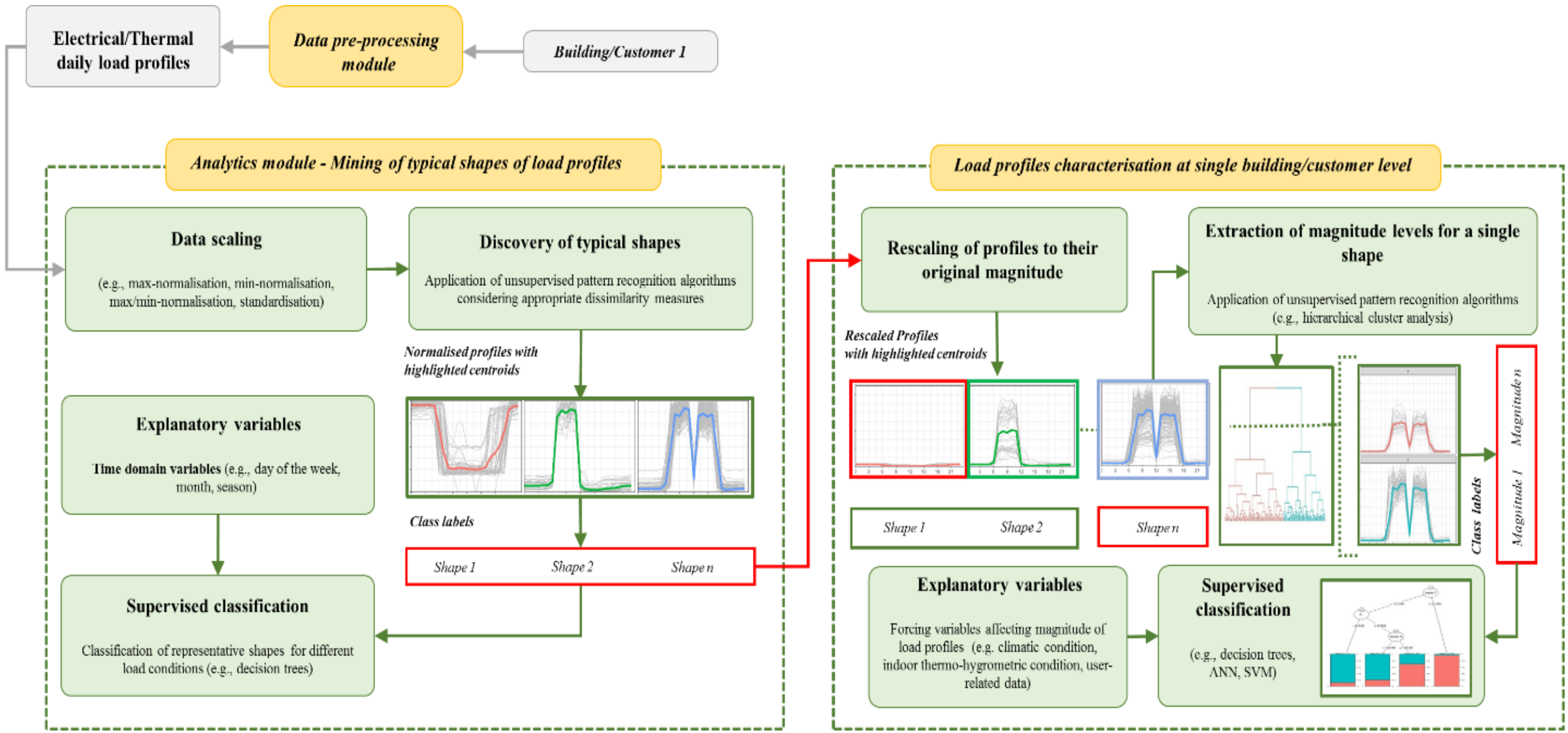
- Implementing more effective energy management strategies through ***targeted financial demand response programs.***
- Better manage the grid operation and the ***interactions*** between energy ***consumption and production.***
- Promote the modification of a load profile that allows the demand profile to be flat or to follow the generation pattern for ***grid stability.***
- Extract knowledge about building energy use patterns for fully exploiting the benefits of ***energy management also at micro grid level***.
- **Assess** the ***impact*** of DSM and DR initiatives over time.

APPLICATION FOR A CASE STUDY

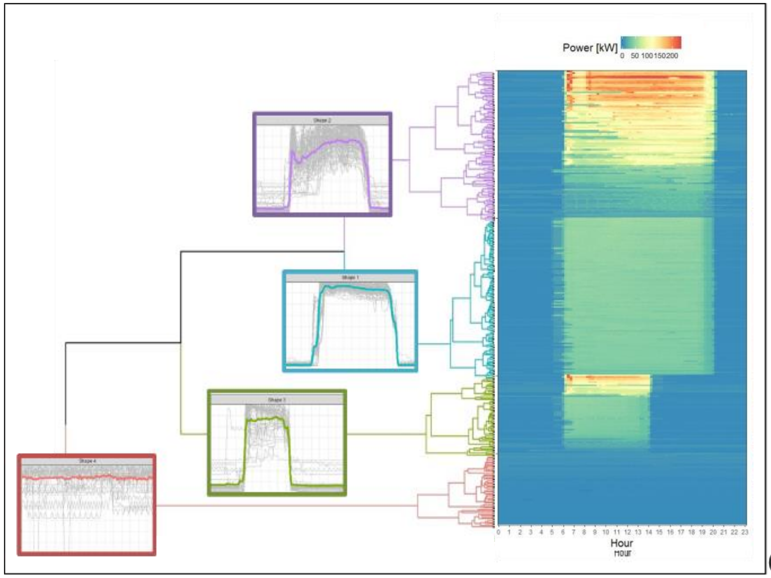
Application of a **pattern recognition procedure** applied to electrical consumption data related to a **heating/cooling mechanical room** of Politecnico di Torino campus in Turin (Italy)

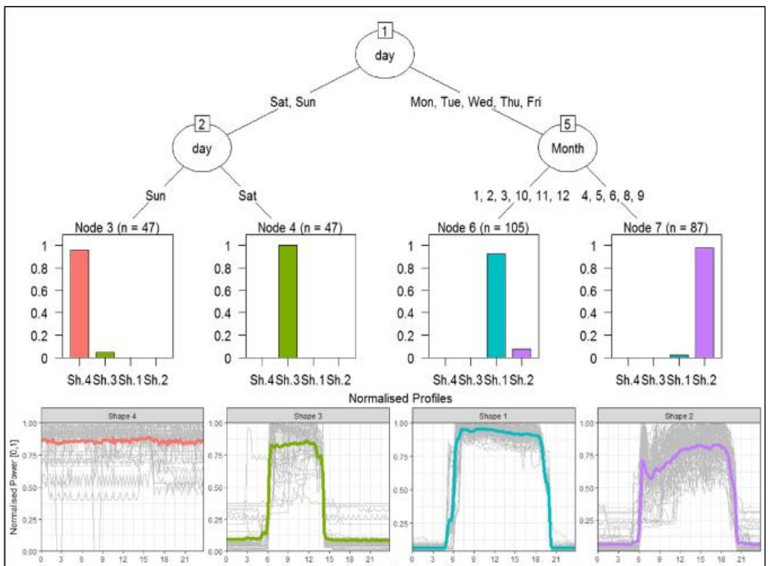
- The system includes both hot and chilled water circuits of the building with the corresponding auxiliary pumping systems.
- The **circulation pumps** installed are different for the two circuits and have an overall **designed electrical power of 120 kW**.
- The hot water is produced through a district heating heat exchanger located in separate area of the campus.
- **The chilled water is provided by two chillers** (with a total design electrical power of 220 kW and a rated cooling capacity of 1120 kW) and a water to water reversible heat pump (with a design electrical power of 165 kW and a rated capacity of 590 kW in cooling mode).
- **The two chillers and the heat pump are connected in parallel and the heat rejection is operated through a geothermal water source in a closed loop**. The operation of chillers is controlled according to the cooling load of building to maintain supply/return temperature of the chilled water at 7/12 ° C.

APPLICATION: METHODOLOGICAL FRAMEWORK



APPLICATION FOR A CASE STUDY





In a first phase the **daily load profiles were normalised in the (0,1) range on the maximum power** value of each profile. A **hierarchical clustering** algorithm with Ward linkage method was then implemented to group the normalised load profiles.

Four different clusters are discovered:

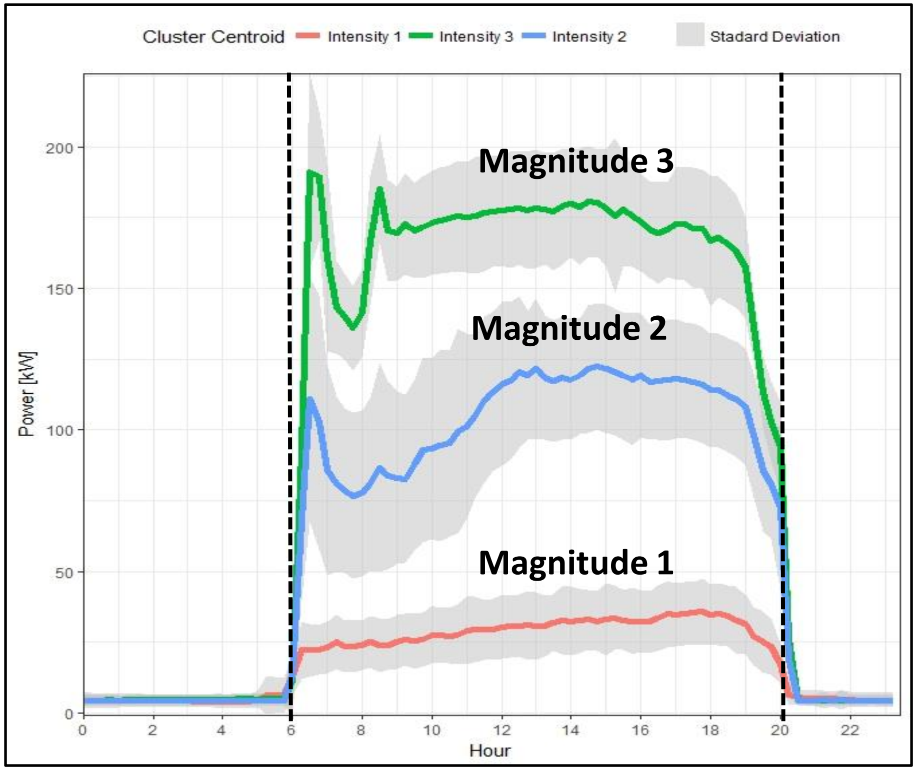
- "**Shape 1**" = Cluster where the electrical energy consumption of the system is due to the operation of the auxiliary pumping system of the hot water circuit
- "**Shape 2**" = Cluster where chillers, auxiliary pumping system and geothermal water pumps were working under different conditions.
- "**Shape 3**" = Cluster of saturdays
- "**Shape 4**" = Cluster of sundays

APPLICATION FOR A CASE STUDY

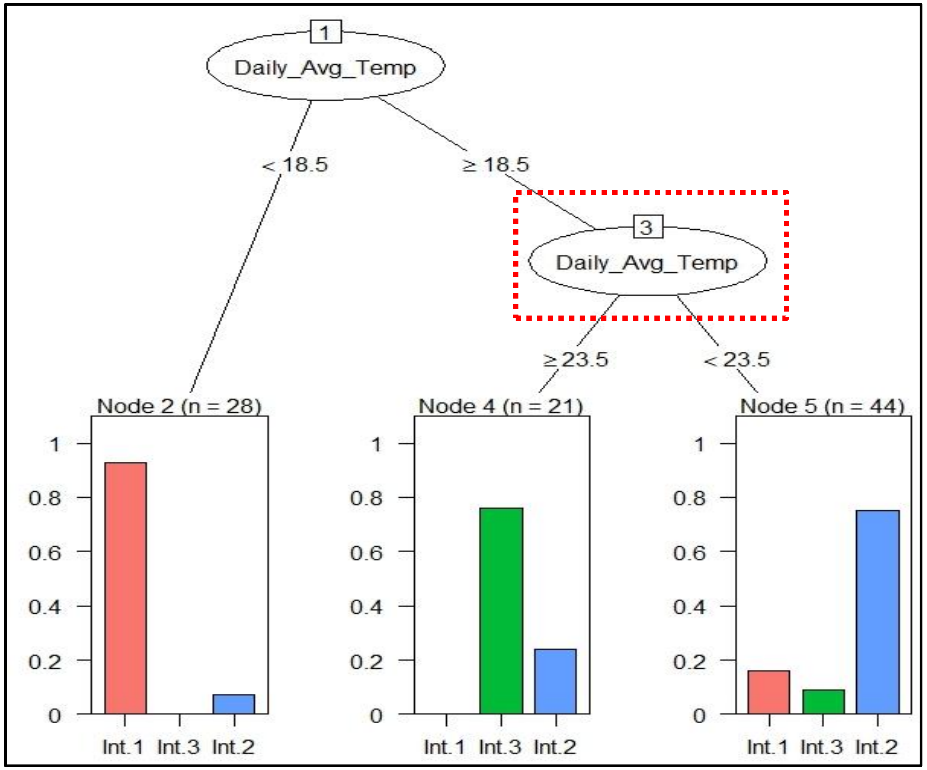
The normalized load profiles grouped in cluster "Shape 2" were rescaled to their original values to perform a further analysis.

A segmentation of the energy profiles belonged in this cluster was performed and **three different groups of daily profiles with magnitudes significantly different were discovered**.

Unsupervised pattern recognition



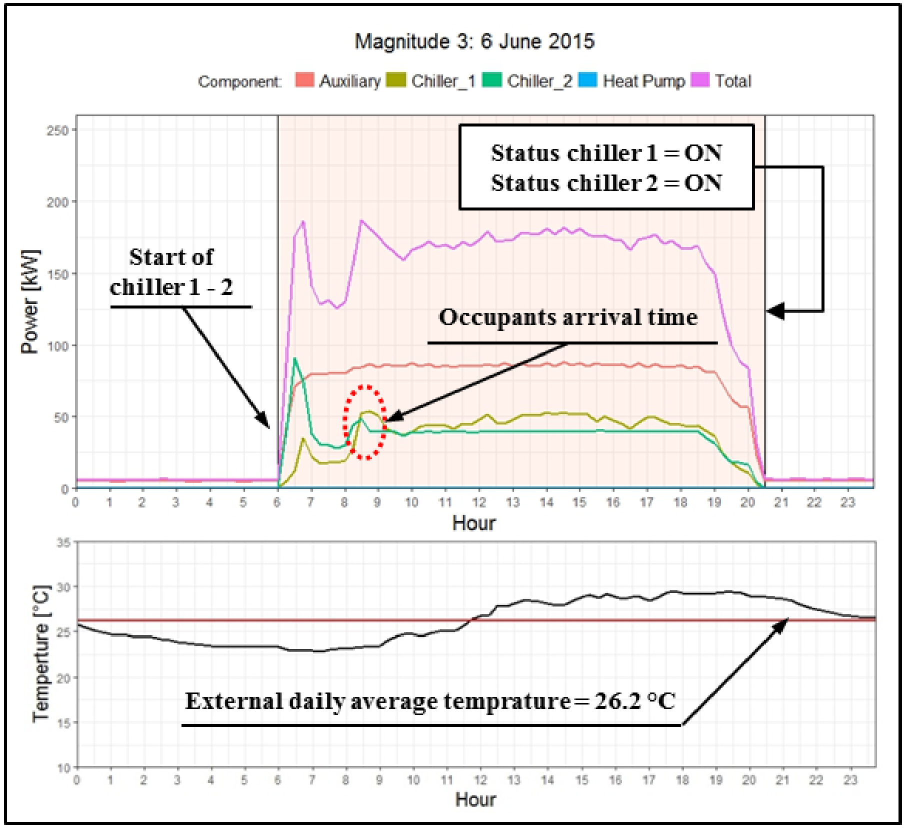
Supervised classification



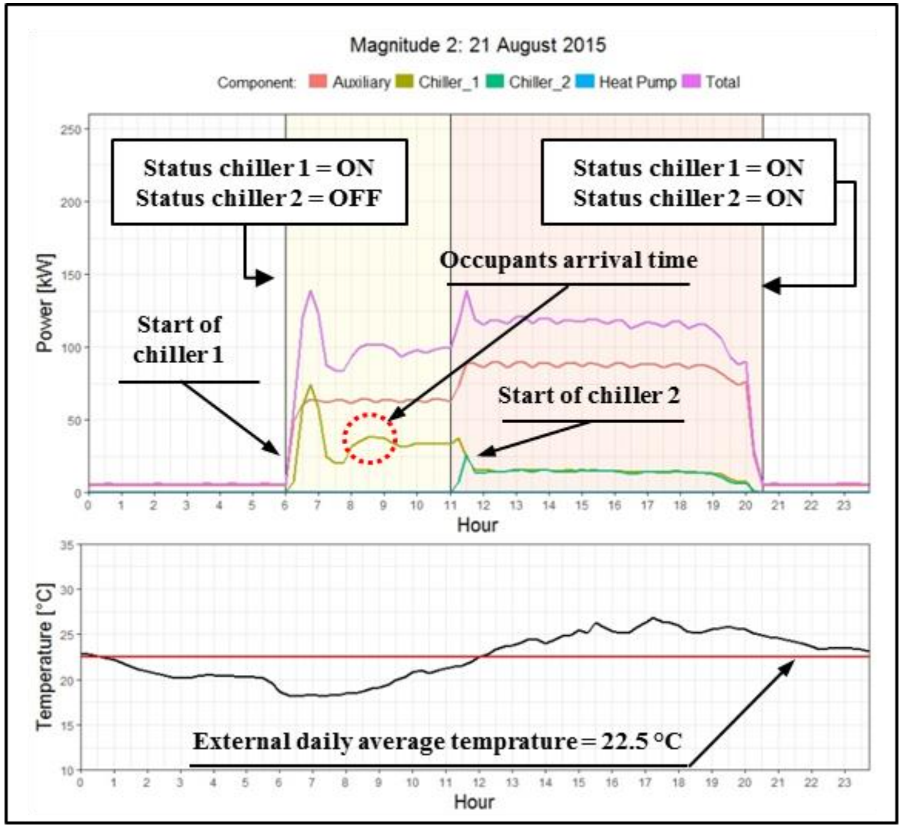
APPLICATION FOR A CASE STUDY

In order to characterise the typical operational patterns of the clusters "**Magnitude 2**" and "**Magnitude 3**", the two days with the **electrical load profiles closest to the centroids** were selected for a further investigation.

Daily profile closest to centroid of cluster "Magnitude 3"



Daily profile closest to centroid of cluster "Magnitude 2"



CONCLUSIONS: WHO CAN BENEFIT?

From single building to stock of buildings

Building managers take advantage in developing different strategies involving energy savings opportunities such as the installation of PV or a thermal/electrical storage systems. Information about typical daily patterns could help in the selection of the most appropriate tariff plan or proper DSM strategies and implementation of anomaly detection strategies. Fundamental aspect involves also the assessment of energy savings consequent to energy conservation measures that can be achieved comparing load patterns before and after a retrofit action.

Energy Service Companies (ESCO) could employ knowledge about building load profiles to develop energy savings and conservation measures along with other energy services.

Transmission System Operators (TSOs) and **Distribution System Operators** (DSOs) In the case of smart electricity grids or district heating installations could employ profiling tools as robust support to DSM strategies aimed at improving the grid balance and developing proper tariff plans for the different customers' categories.

Policy makers may take advantage from load profiles characterization to identify which actions could have the major effects over a specific group of consumers.

CONCLUSIONS & LESSONS LEARNED

- **Energy management systems** capable to exploit the potential of building related-data for energy management and operation by means of a data analytics technologies represent a powerful opportunity in the building physics sector.

- Extracting useful knowledge **by coupling building physics domain expertise with data analytics procedures** makes it possible to discover operational rules to support building operation and correlations that are not so obvious for experienced energy management.

- The diversity of data analytics techniques and their combination needs **robust frameworks.**

- The knowledge of energy consumption patterns at single system/building makes it possible to promote their **optimisation through changes in energy demand, load shifting, the detection and diagnosis of anomalies** related to uncorrected system operations or users' behaviour.

Mining typical load profiles in buildings to support energy management in the smart city context

Questions?

Alfonso Capozzoli

Marco Savino Piscitelli and Silvio Brandi

POLITECNICO DI TORINO

alfonso.capozzoli@polito.it