# 1 - Introduction

The rapidly growing electrification of buildings energy systems and appliances lead to an increasing electricity demand. On a global scale, direct and indirect CO2 emissions from buildings energy use reached its all-time high in 2019 \cite{IEEA2020}. In Europe \cite {EU2019} estimates that Buildings sector account for 40% of final energy use and 36% of CO2 emissions. The targets imposed by the European community to reduce greenhouse emissions by at least 55% by 2030 \cite{ ﻿EU2018} highlight the critical role played of buildings. Considering that \cite{ ﻿Ramesh2010 } estimates that almost 90% of the total energy consumed during the life cycle of a building depends on the building operation, reducing energy consumption, increasing appliances efficiency and prevent energy wastes is the key to meet climate change goals.

The opportunity that has been presented in the last years comes from the IT sector. A large amount of operation data is produced by the operation of buildings thanks to the increasingly widespread introduction of IOT devices. \cite{ ﻿Erhan2021 } estimates that building sector accounts for. reaching zettabyte of data making buildings not only and energy intensive but information intensive \cite{﻿Fan2021}.

Building data are heterogeneous and reflects the complex interaction that occurs between occupants, energy systems, the building envelope, and external conditions. Managing those data is not trivial \cite{ ﻿Molina-Solana2017 }, however if properly managed ingested and analysed, provide the opportunity to gain insight on the building operational behaviour discovering opportunities for savings.

Integration of IoT sensors and Machine learning approaches to automatically infer information about the energy usage in buildings has been proved to be effective to variety of tasks \cite{Molina-Solana2017}: pattern recognition, analysis for performing essential tasks in building energy management such as energy consumption forecasting, anomaly detection and diagnosis, advanced benchmarking, load profiling, and schedule optimization of building energy systems.

In this paper, we focus anomaly detection of electrical loads in buildings, which is a key application to aid decision makers to reduce wastes and promote sustainable behaviour of end users.

## 1.1 – Anomaly detection: related work

Anomaly detection can be employed in detecting abnormal behaviour of end users, detection of faulty appliance or energy subsystem and spotting technical and non-technical energy loss \cite{ ﻿Himeur2020}.

An anomaly is a point that deviates from normal expected behaviour

Many categorization have been proposed in literature and some are specific for building environments \cite{ Himeur2020, ﻿Erhan2021}. The scope of this paper is not to go deep into categorization, thus we adopted an anomaly classification based on type, level and algorithm as reported in figure.

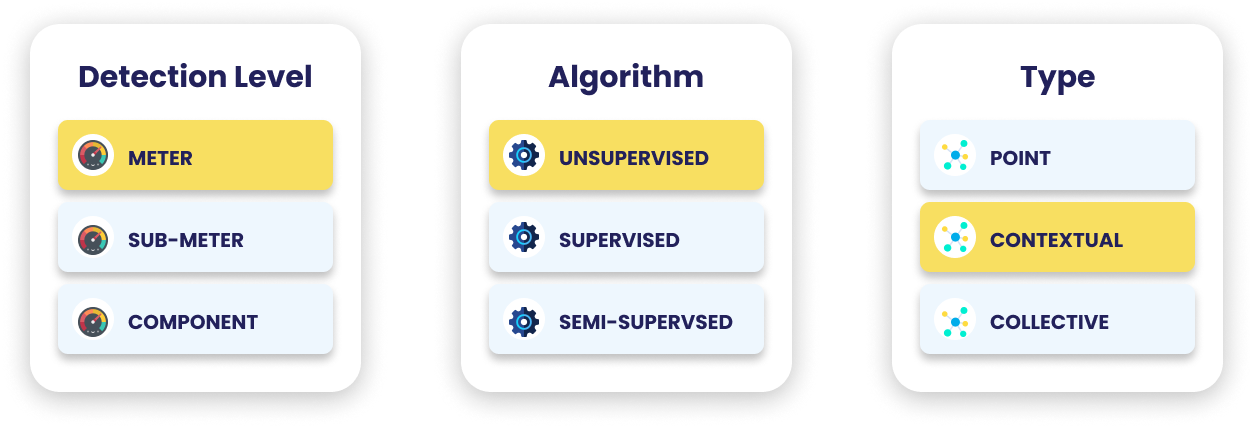


Figure 1

Regarding the classification of anomaly type: A *point anomaly* is one individual instance that can be considered anomalous when compared to the remaining data. *Context anomalies* start from the assumption of dividing the behaviour from the context: the same behaviour might not be considered an anomaly if it happens in a different context. In a *Collective anomaly* the event instance does not represent an anomaly per se, but only if considered within the collection of all the other events instances.

Depending on the detail of electrical load monitored the detection level can be performed at different levels: Meter level: consists in analysing the whole building electrical load measured at meter level, without having any information on the disaggregation of that load among the different sub loads or appliances. Sub-meter level disaggregated to energy system. Component level consists in identifying anomalies referring to a given appliance, it is performed thanks to sub metering infrastructure that measures every single sub system

A classification of data-driven anomaly detection techniques on algorithmic centric view in building energy systems: Supervised anomaly detection requires to train a machine learning algorithm using labelled dataset in order to classify anomalous consumption or not. Although supervised anomaly detection can achieve high-accuracy identification results as demonstrated in academic frameworks, its adoption in real-world is still limited compared to unsupervised methods, due to the absence of power consumption annotated datasets. \cite{Himeur2020}. Deep learning, ANN, Regression, Probabilistic models, Traditional classification

Unsupervised anomaly detection consists in detecting rare and unknown anomalous energy patterns without any a priori knowledge. It usually consists in modelling the normal behaviour and then identify patterns that deviates, under the assumption that the number of anomalies is low compared to the observations. Clustering, \cite{﻿Voltage2016} performs anomaly detection on smart grid though the use of clustering

Semi supervised.

## 1.3 - Anomaly detection using Matrix Profile

One of the most promising technique is Matrix Profile. Introduced by \cite{﻿Yeh2017d} it is a novel algorithm that performs the all-similarity-join-search among two time series, i.e. finds the nearest neighbour for each object of a data collection. This problem has always been tackled by a computation al time issue resulting, even for modest datasets, into not reasonable computational time. In literature this issue was addressed by reducing the dimensionality of dataset through PAA in order to speed up computation. However, this method causes loss of valuable information. MP proposes a ultra-fast similarity search under the z-Euclidean distance that does not reduce dimensionality, MP in fact calculates the full join, eliminating the need of setting a threshold, making the method almost parameter free. The exact and scalable algorithm allows the method to be incrementally maintainable, deterministic in time and so parallelizable on multicore processor to speed up even further the computations.

Given two time series and a given subsequence length, the matrix profile algorithm produces two new series: the matrix profile and matrix profile index. Matrix profile is a one-dimensional timeseries that stores the z-normalized Euclidean distance between each subsequence of the first series and the closest matching subsequence (i.e. nearest neighbour) of the second time series. Matrix profile index is a one-dimensional timeseries that contains the index of where the nearest neighbour is located in the second timeseries.

﻿

By finding the minimum value of the Matrix Profile is possible to find the best matching subsequence in a series, i.e. motif discovery, on the other side by finding the maximum value of the Matrix Profile it is possible to find the subsequence with the largest distance to its nearest match, i.e. discord discovery. In this sense discord discovery may be interpreted as an anomaly detection method that discovers the most unique sub sequences in a dataset. Discord discovery using MP as anomaly detection method has been employed with success in different fields.

In medical field \cite{﻿Alshaer2020} proposes an unsupervised real time anomaly detection method based on continuous learning of time series shaplets extracted though Matrix Profile algorithm. Those shaplets are extracted and stored in an anomaly library and then used for anomaly detection in an electro-cardiogram (ECG) time series (﻿MIT-BIH database [32]), using a in a sliding window.

An industrial application of anomaly detection is presented in\cite{﻿Anton2020} which combines the classical approach of Matrix profile with the hamming distance to automatically detect intrusions in the network of a water processing facility.

\cite{ ﻿Madrid2019 } applies a generalization of MP algorithm called Pan MPto find different length anomalies in ﻿automated pedestrian counting system developed in Taipei.

MP is an effective method used to identify anomalies in IT field. \cite{DinalHerath2019} introduces ﻿a real time anomaly detection framework based on matrix profile called RAMP (Real-Time Aggregated Matrix Profile), that is able to identify anomalies in scientific workflows. (building block). \cite{ ﻿DePaepe2020a} Applies a noise elimination technique based on Matrix Profile on real Yahoo! internet traffic metrics to detect anomalous behaviours.\cite{ ﻿DePaepe2019 } demonstrate how the elimination of noise can help in anomaly detection of noisy date by testing the algorithm on Numenta Benchmark \cite{ ﻿Ahmad2017 }.

In the energy field there are few implementations of MP algorithm. \cite{ ﻿Nichiforov2020 } Identifies anomalous patterns though a basic application of Matrix Profile on public building energy traces and then classifies the pattern.\cite{ ﻿Zhu2020} demonstrates how Matrix Profile can be useful in detecting anomalies in different fields in particular in meter swapping and earthquake monitoring. \cite{﻿DePaepe2020b} applied an implementation of the classic Matrix Profile, called Contextual Matrix Profile, in detection of anomalous energy consumption on a ventilation units of three households. \cite{ ﻿Park2020a } applies MP as a part of a automated load profile discord identification ( ALDI ) based on statistic comparison between normal and anomalous patterns in a large portfolio of buildings.

## 1.4 – Implication MP on energy domain

From the previous literature review it emerges that the MP method have been employed successfully in different fields for anomaly detection and the authors have proposed different implementations according to the field of interest. In fact, even if MP is an unsupervised method useful for discord discovery every field have different boundary conditions and different restrictions that cannot be overlooked. In the field of building, the most interesting timeseries are those related to energy consumption. Those timeseries are strictly correlated to many different variables such as occupation, weather conditions, energy systems and so on. A completely unsupervised method may fi by its nature to consider the relation with those variables and extract ineffective or trivial results, not useful for anomaly detection.

In building an anomaly is an unexpected behaviour that consists in a strongly different energy consumption. The classic MP (with z-score normalization) searches for each subsequence the nearest neighbour based on shape similarity, however, anomalous shapes not always correspond to anomalous energy consumption, as well as similar shapes in z score not always reflect similar behaviour. With reference to figure it is possible to appreciate how three sub sequences of electrical load timeseries are considered similar in z score while without normalization the amplitudes are very different reflecting very different energy consumption.

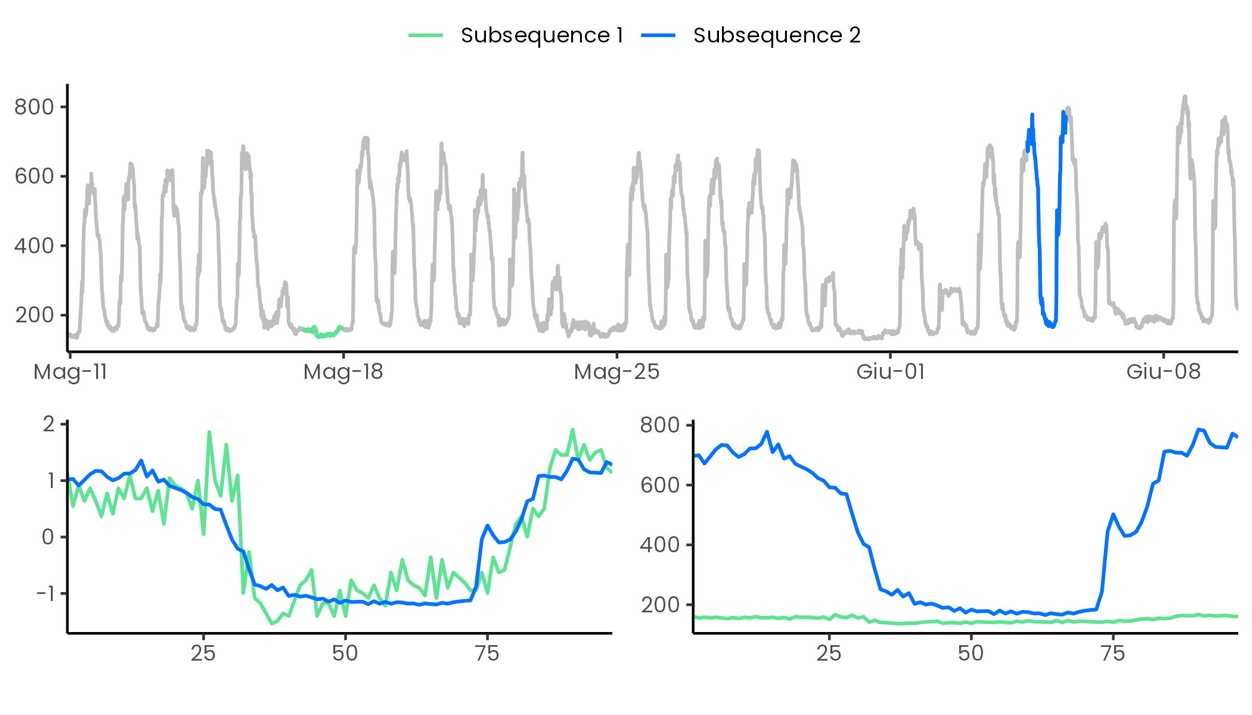


Figure 1

The effect of z-normalization not only avoid to take into account the magnitude of the time series but also tends to enhance any fluctuation of the timeseries. By comparing two relatively flat sub sequences under z-score normalization the resulting Euclidean distance is higher compared to non-flat sub sequences (see figure), this results into higher values of MP in flat regions of the timeseries. This issue have been largely analysed in \cite{ ﻿DePaepe2020a } where a smoothing is proposed as possible solution to this issue, beside the trivial solutions of discard flat regions or change the sub sequence length. However, in energy

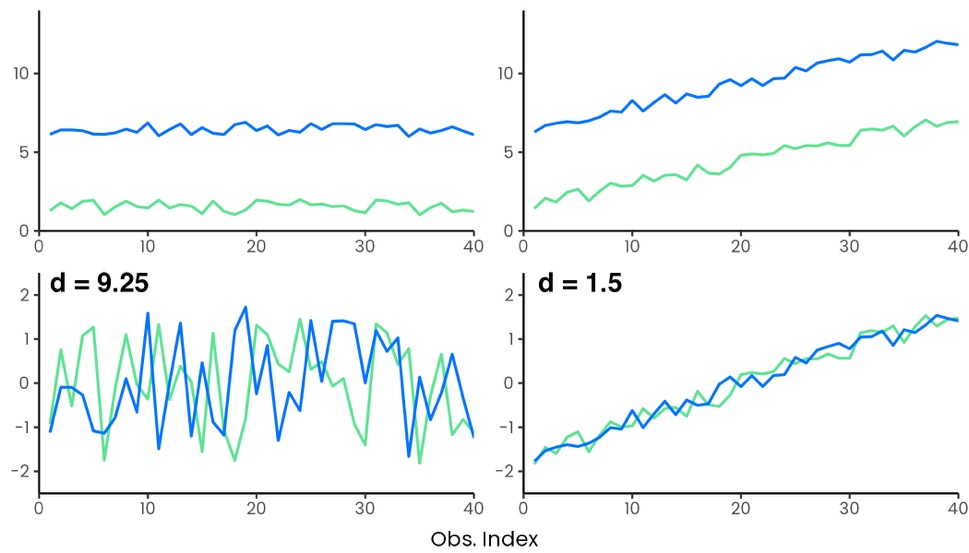


Figure 2

Moreover the energy consumption pattern changes between weekdays and weekends/holiday, so it would be unfair comparison to compare sub sequences pertaining to these groups, the same is to compare sub sequences of night hours and daily hours. Introducing domain knowledge to find discords only in sone subgroups of the timeseries became important.

﻿\cite{Dau2017} introduces the concept of annotation vector used to introduce domain knowledge in the process of motif and discord discovery, which allows to find results that follows users defined constraint and produce better results, closer to expectations of the analyst. This method has been proved to be effective to solve different issues: simplicity bias, actionability bias.

However, this method is a posteriori method that does not modify the way MP is calculated, in particular all-pairs similarity search is always performed and then some regions are excluded form motif/discord search. Sometimes it can be useful to exclude some region or to group sub sequences into different groups and then perform the similarity search in order to discover anomalies by comparing only the interesting regions and excluding others. A solution to this problem have been proposed by \cite{DePaepe2020b} where CMP algorithm permits to define ranges along T1 and T2 and look for the best matching subsequence among these ranges. This permits different grouping before the MP calculation thus providing different insights.

## 1.5- Contribution of the paper

The prompt and accurate discovery of anomalies in building electrical load is the key to reduce energy wastes and enhance energy management in buildings. To this aim the objective of this work is to detect anomalous electric load at building level in quasi real time through the unsupervised anomaly detection methods based on matrix profile method. According to the previous literature review and excursus on implication of MP as anomaly detection method, this paper intends to address the following issues by contributing as follows:

* Overcame issue of z-score in energy with the use of Euclidean distance
* Apply a contextual anomaly detection by applying an advanced contextual matrix profile method with the definition of context and groups through an unsupervised fashion
* Couple unsupervised automatic data driven methods, without setting any parameters
* Propose an almost real time anomaly detection method
* Define a clever anomaly score that talked into account only positive anomalies (higher power consumption)

The rest of the paper is organized as follows: in section 2 in section data analytics methods are presented, in section 3 the methodological framework, in section 4 results and discussions.

# 2.-Description of the Data Analysis Methods

## 3.1 - Matrix Profile

In the case the two time series are the same a self-join is performed, additional constraints are added in order to avoid trivial matches where sub sequences match themselves, in particular an exclusion zone is set to avoid neighbour search just before and after the reference subsequence.

Even if different robust distance metrics have been proposed \cite{Gharghabi2020, ﻿Zhu2020}, the original method uses the z-normalized Euclidean distance to calculate the MP, which provides a way to focus on shape instead of amplitude and permits to extract information without having domain knowledge.

Definitions:

* A time series $T \in R^n$ is a sequence of real-valued numbers $t\_i \in R : T = \{t\_1,t\_2 \dots t\_n \}$ with $1<i\leq n$ where $n$ is the length of $T$
* A subsequence $T\_{i,m} \in R^m$ is a continuous subset of values from $T$ of length $m$ starting from position $i$. Formally defined as $t\_i \in R : T\_{i,m} = \{t\_i,t\_{i+1} … t\_{i+m-1} \}$ with $1<i<n-m+1$
* An all-sub sequences set $A$ of a time series $T$ is an ordered set of all possible sub sequences of $T$ obtained by sliding a window of length $m$ across $T$. It is formally defined as follows: $A = \{T\_{1,m} , T\_{2,m} \dots T\_{n-m+1,m} \}$ where $m$ is a user-defined subsequence length. We use $A[i]$ to denote $T\_{i,m}$
* A distance profile $D\_i$ of a timeseries $T$ is a vector of distances between a given query (subsequence $T\_{i,m}$) and each subsequence in an all-subsequence set $A$. Formally, $D\_i = [d\_{i,1}, d\_{i,2},\dots, d\_{i,n-m+1}]$, where $d\_{i,j} = dist(T\_{i,m}, T\_{j,m})$ for all $j \in [1,2,\dots, n-m+1]$ where $i\neq j$ and $dist$ is the distance metric applied.

The concept behind MP calculation is presented in figure, where self-join MP calculation is performed. Given a timeseries T, by sliding the window of length m, the Euclidean distance d,I,j between the z-normalized sequence T\_i,m in the first time series T and z-normalized sequence T\_i,m in the second time series is calculated and then stored in a distance profile D. This process is performed for each subsequence of the first time series and the generated distance profile create the so called full distance matrix. By extracting the smallest value in each row/column (the smallest non-diagonal value for the self-join case, A=B) MP is finally generated.

The previously presented method is the naïve implementation of the MP calculation and is the most inefficient since the full distance matrix have to be calculated. Behind the MP algorithm there are more efficient algorithms such as stamp and stomp that provides the complete MP in a more efficient and time saving way

By joining information of MP index and MP many insights could be extracted. The main application is the motif and discord discovery that consists in finding respectively the minimum and the maximum of the MP and locating the nearest neighbour of the subsequence starting in that particular index.

The mMP is the column wise minimum over the entire full distance matrix, meaning that if finds the best matching T1 subsequence (i.e. minimum distance) for any subsequence in T2. On the other hand contextual matrix profile is defined as the minimum over rectangular regions of the full distance matrix, allowing to find the best matching subsequence in ranges dover T1 and T2 allowing to group data in custom quay comparing only portions of T1 with portions of T2

Contextual Matrix Profile

Immagine che contiene testo, elettronico

Descrizione generata automaticamente

Figure 3

## 2.2 – Decision tree

Classification is the task that assigns a class label to unlabelled instances through a classification model which is able to create a relationship between the predicted or target variable and the predictive variables.

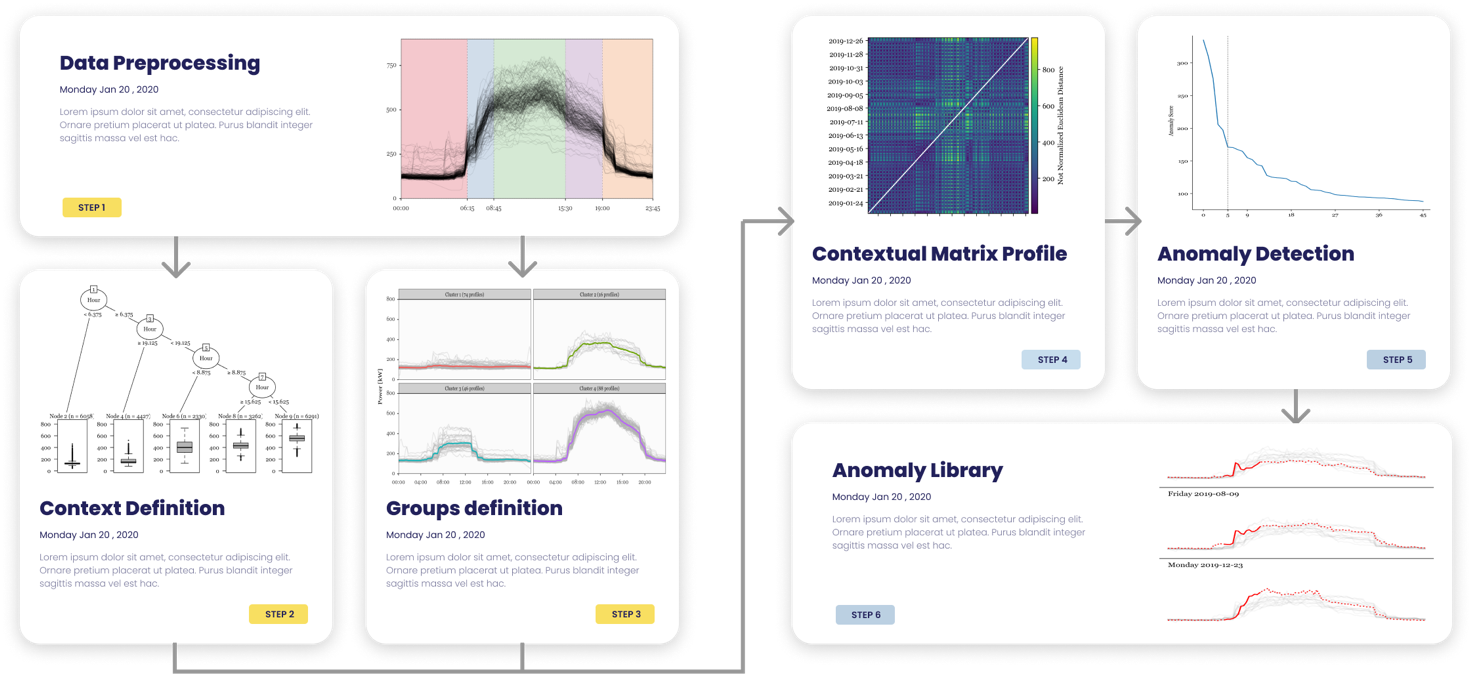
Among the numerous classification methods used to describe and explore complex data, the most used are those based on decision trees thanks to their capability to be easily translated in graphical form are commonly used in different fields \cite{﻿Grubinger2014}. Depending on the type of target variable a decision tree can be a classification tree (categorical target variable) or regression tree (numerical target variable) \cite{ ﻿Tan2011 }.

In this paper is employed the recursive partitioning decision tree called classification and regression tree CART. Starting from the root (all the available instances),this method proceeds through a binary decision fashion to split the instances in purer subsets (nodes) in a froward stepwise fashion maximizing at each step the purity of each node, yielding local optimum \cite{ ﻿Grubinger2014} once a stopping condition is satisfied. The purity of each node can be identified through an impurity measure \cite{ ﻿Yan2016a, ﻿Capozzoli2018a, ﻿Tan2011 }.

## 2.3 – Clustering

Clustering is the process of grouping instances based on similarity within some attributes, the goal is to create groups (i.e. clusters) in which objects

# 4. - Methodological Framework



## 6.1 – Pre Processing

## 6.1 - Context definition

## 6.1 - Groups definition

## 6.1 - Contextual matrix profile

## 6.1 - Anomaly detection

## 6.1 - Anomaly library

# 5. - Case study

The case study analysed refers to the energy consumption of a MV/LV transformer cabin identified as “substation C”, that serves a part of the main campus of Politecnico di Torino (PoliTo), an Italian university located in Turin. The measurement infrastructure provides the total electrical load with 15 min timestamps. In order to use a dataset that is large enough to capture the behaviour of the electrical load, with regular occupation patterns

# 6. – Results

## 6.1 - Preprocessing

## 6.1 - Context definition

## 6.1 - Groups definition

## 6.1 - Contextual matrix profile

## 6.1 - Anomaly detection

## 6.1 - Anomaly library

# 7. - Discussion

Twin freak

﻿For a given sub-sequence, Matrix Profile computes the Euclidean distance with respect to all other sub- sequences and identifies the minimum distance. Therefore, a repeated anomaly instance would cause false negatives due to the previous anomaly instance being part of the all sub- sequence set.

Specifically, frequent/rare sub sequences are defined as the ones with the smallest/largest 1-nearest neighbour distance, which are also known as motif/discord. However, discord fails

the ones with the smallest/largest 1-nearest neighbour distance, which are also known as motif/discord. However, discord fails to identify rare sub sequences when it occurs more than once in the time series, which is widely known as the twin freak problem.

\cite{DinalHerath2019} through a semi-supervised model permits to limits the number of sub sequences compared, considering for comparison only references with no anomalies.

\cite{﻿He2020} proposes a method called “Neighbour Profile” based on sampling and density estimation to perform anomaly detection and overcame the issue of twin freak.

# 8. - Conclusions and Future Work