# Statistical features and identification tests for time series analysis

Introduction

- Statistical time series features extraction and discrimination tests are very useful in order to apply analytical algorithms over those time series because it solves lot of problems related with time series analysis by giving us a summarization of the time series itself:
  - How to apply classical machine learning algorithms?
  - How to choose the most suitable algorithm in order to satisfy a determined problem?
  - It makes sense to use this algorithm for analysing this time series?
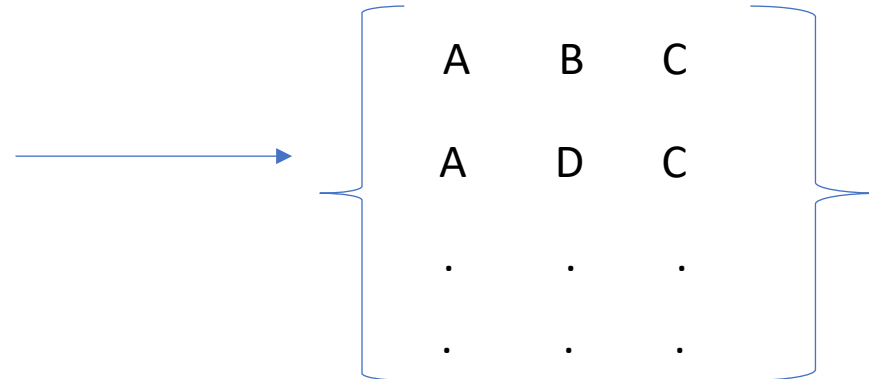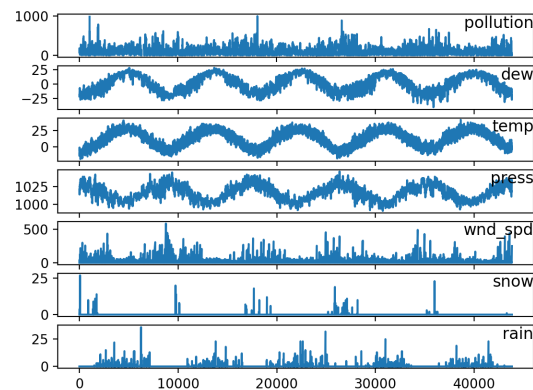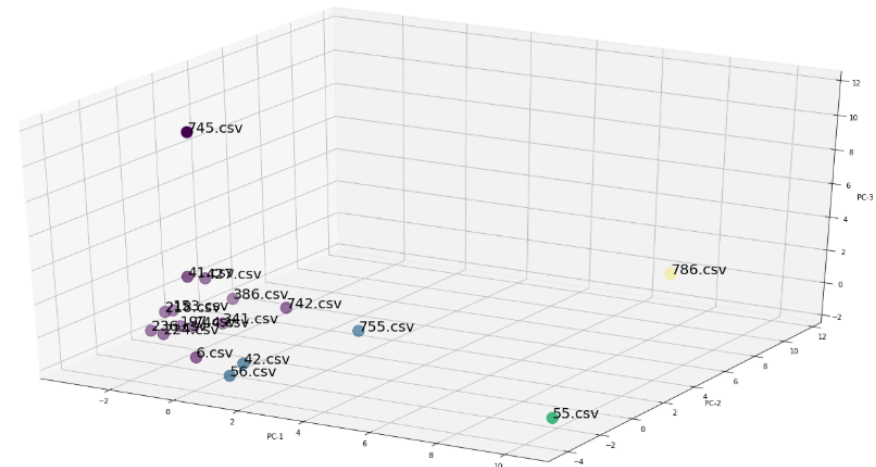
Features extraction

- The features of some time series are just statistical values or associated concepts that summarizes a determined time series in order to manage all the candidates of an application in a standard way without dealing with the typical time series analysis problems.

- Those features are usually statistical values obtained from the data of the time series itself.

- Examples: mean, maximum, minimum, median, kurtosis, number of points above the mean, etc.

- Libraries: Khiva, tsfresh, etc.

SHAPELETS

Features extraction. Use cases

- Once all the time series which are candidates to analyse are standardized, those can be processed by classic machine learning techniques and then clusterize or classify the time series in order to get insights, as simple as that.

- Some recommended techniques are the pre-processing of the resulting summarizations of the time series in order to remove unneeded features.

S H A P E L E T S

Statistical tests

- Statistical tests are another statistical feature that could be used in the features extraction part, but it bring as more properties.

TIME SERIES TYPE	STATISTICAL TEST
Regular/Irregular	Time axis comprobation.
Trend/Non-Trend	Person correlation trend test and its p-value.
	Daniels trend and its p-value.
	Mann-Kendall trend test and its p-value.
Ergodic/Non-Ergodic	Means of autocorrelation function evaluation. Augmented Dickey Fuller test.
Periodic/Non-Periodic	Fisher's test for periodicity.
Seasonal/Non-Seasonal	Chi-Square Goodness-of-fit test.
	Kolmorov Simrnov type statistic.
	Harmonic analyses based on the Edward type statistics.

SHAPELETS

Statistical tests. Use Cases.

- Discriminate non applicable time series algorithms

- More suitable selection of algorithms: filtering, distances, forecasting, clustering, segmentation, motif and discord discovery, anomaly detection, query by content, rule discovery, causality...

- Examples:

ALGORIHTM	EXPLANATION
ARIMA - FORECASTING	Most suitable for stationary time series.
SIMPLE EXPONENTIAL SMOOTHING - FORECASTING	Most suitable for time series which have no trend with and are no seasonal.
SELF-ORGANIZING MAPS - CLUSTERING	Most suitable for periodic time series
ROBUST ANOMALY DETECTION – ANOMALY DETECTION	Most suitable for stationary time series.

S H A P E L E T S

Exercise

- **Dataset:** 100 time-series related with energy consumption provided by commercial sites during 2012, which are tagged by subindustry. The dataset has in excess of 10,000,000 points.

- **Target:** Apply classical machine learning algorithms in order to predict the tag of the time series.

- **Work to do:** Explore the khiva's features documentation in order to vary the features used to summarize the time series.