



SHAPELETS

Time Series

Features and Statistical tests

Oscar Torreno - @oscartorreno

David Cuesta - @Dcuesta7

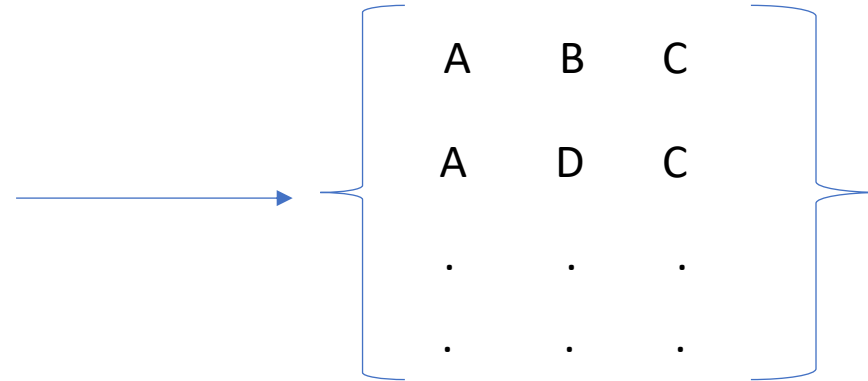
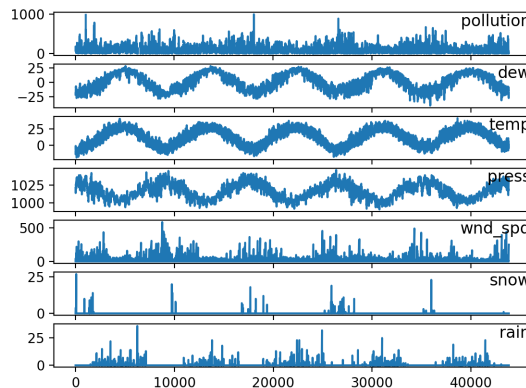
Features extraction

- Features are typically statistical values or associated concepts that summarize a given time series.
- Typically used to: accelerate computations, have a set of common features for series of different lengths, etc.
- Examples: mean, maximum, minimum, median, kurtosis, number of points above the mean, symmetry looking, etc.
- Existing libraries: Khiva, tsfresh, etc.

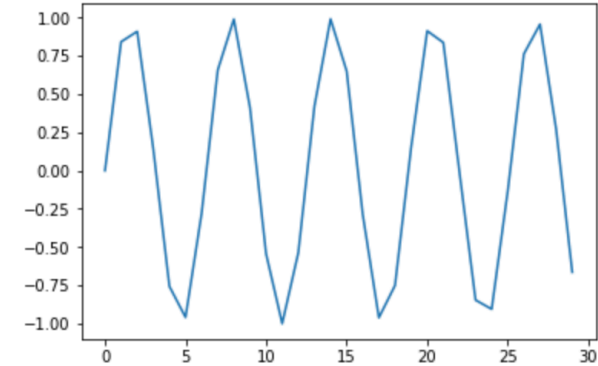
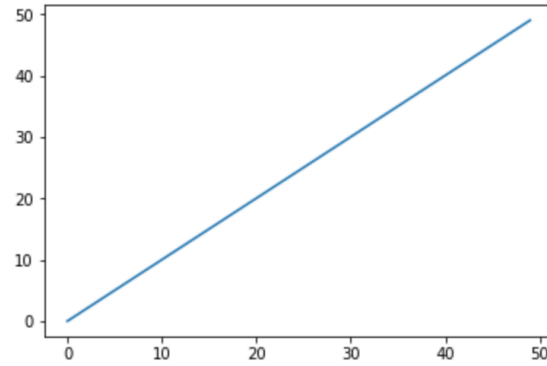
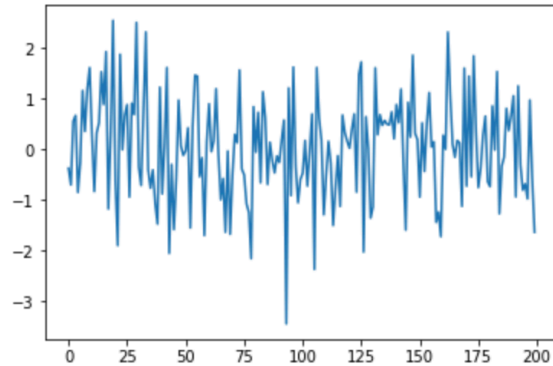


What can we use features for?

- To apply classical machine learning algorithms mainly for clustering and prediction/forecasting



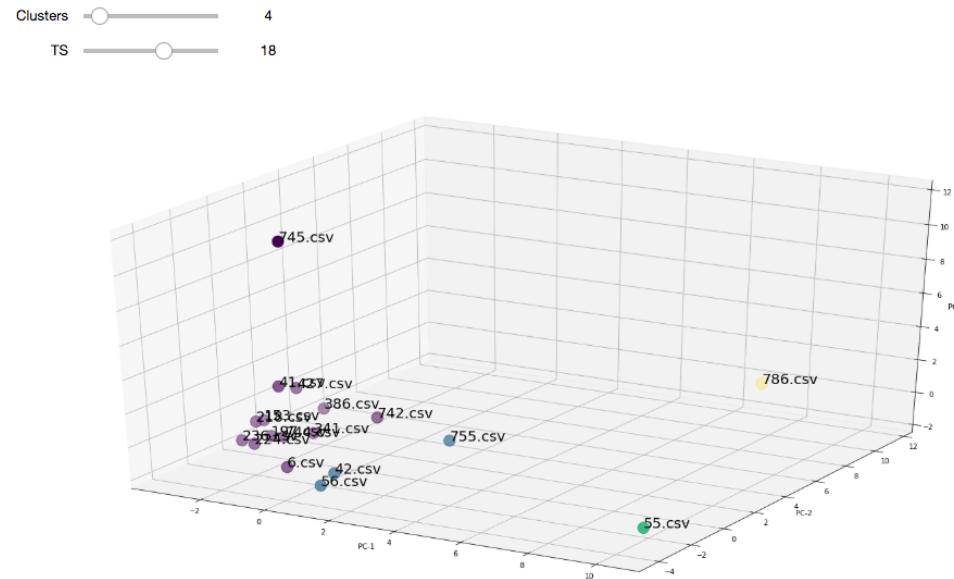
Features. Example



	mean	kurtosis	skewness	symmetry_looking
0	0.039324	0.106709	-0.113297	True
1	24.499994	-1.200000	0.000001	True
2	0.042269	-1.509307	-0.082693	False

Features extraction. Use cases

- Based on the features matrix, classic machine learning techniques are able to clusterize or classify the time series.
- If the feature vector is too long, PCA or equivalent algorithms might be needed.



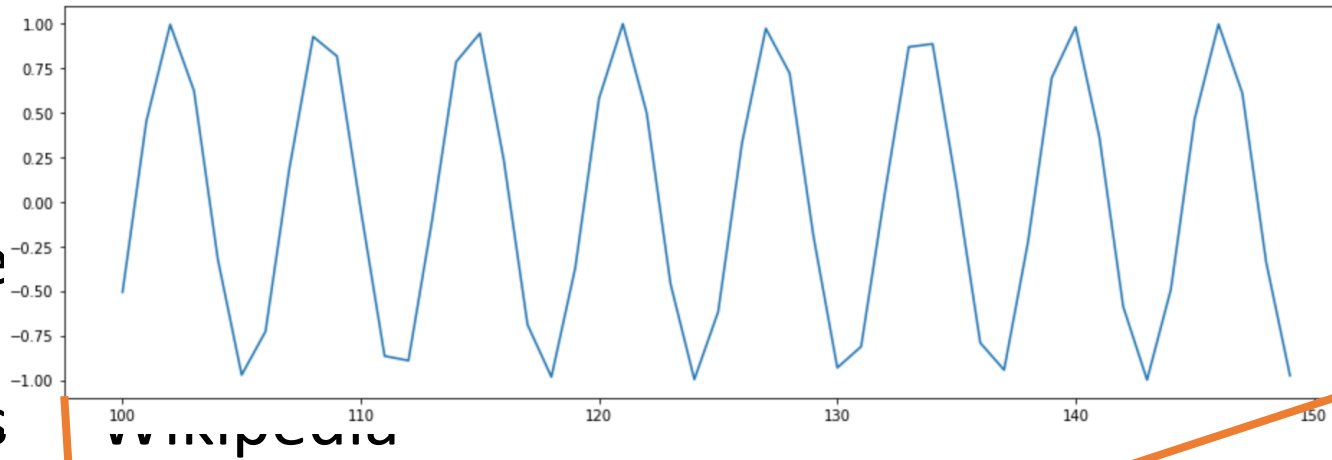
Statistical tests

- Statistical tests check statistical properties of the series

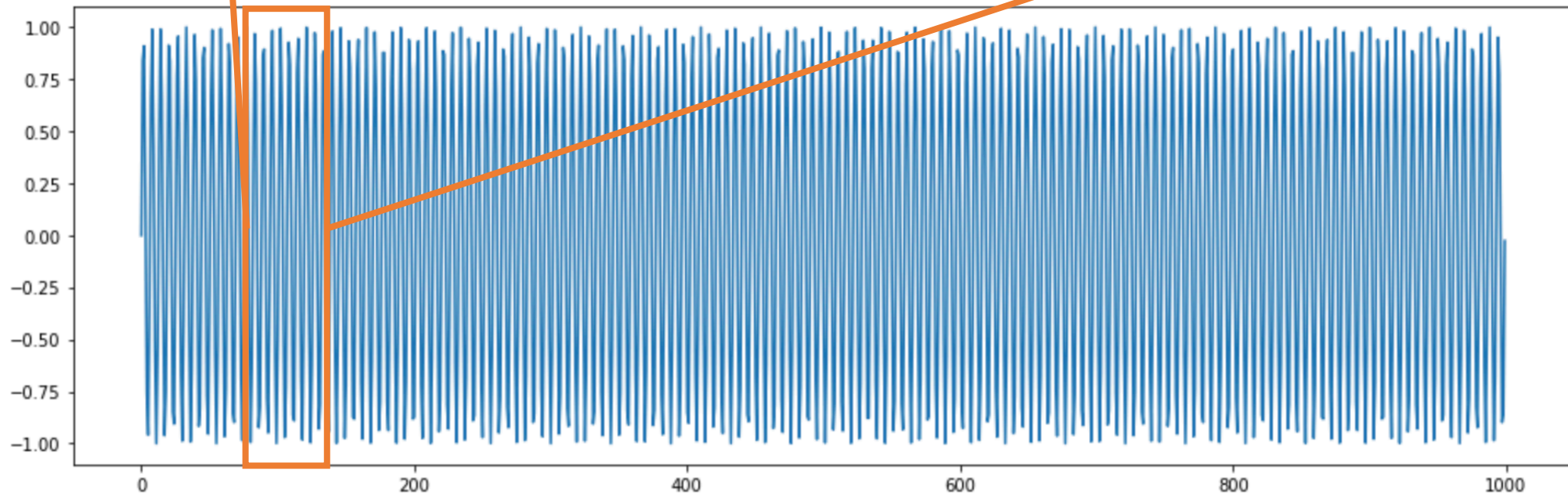
TIME SERIES PROPERTY	STATISTICAL TEST
Regular/Irregular	Time axis check.
Trend/Non-Trend	Pearson correlation trend test and its p-value.
	Daniels trend and its p-value.
	Mann-Kendall trend test and its p-value.
Ergodic/Non-Ergodic	Means of autocorrelation function evaluation. Augmented Dickey Fuller 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.

Ergod

- “A series deduce process



ies can be
ple of the



What can we use statistical tests for?

- To choose the most suitable algorithm in order to satisfy a determined problem
- To determine if an algorithm is not recommended to analyse a time series

Statistical tests. Use Cases.

- Discriminate non applicable time series algorithms
- More suitable selection of algorithms: filtering, distances, forecasting, clustering, etc.

- Examples:

ALGORITHM	RECOMMENDATION
ARIMA - FORECASTING	Most suitable for stationary time series.
SIMPLE EXPONENTIAL SMOOTHING - FORECASTING	Most suitable for time series which have no trend and do not contain a seasonal component.
SELF-ORGANIZING MAPS - CLUSTERING	Most suitable for periodic time series
ROBUST ANOMALY DETECTION – ANOMALY DETECTION	Most suitable for stationary time series.

Statistical tests. Use Cases.

ANALISYS	REGULAR	PERIODIC	...
METHOD 1	YES/NO	YES/NO	
METHOD 2			
METHOD 3			
METHOD 4			

Exercise

- **Dataset:** 100 energy consumption time series of commercial sites during 2012. The series are tagged by subindustry.
- **Target:** Apply classical machine learning algorithms in order to predict the subindustry of the time series.
- **Work to do:** Clustering of time series using khiva's time series features algorithms.