

AWT – Clustering using an Aggregated Wavelet Tree: A novel automatic unsupervised clustering and outlier detection algorithm for time series

Christina Pacher, Irene Schicker, Rosmarie DeWit, and Claudia Plant

Introduction

Data processing in meteorology & climatology can be considered as a high dimensional problem. Thus, to identify the transitions and relations in (non-)stationary data methods capable of working with high dimensional time series are needed. Using non-standard data such as privately owned weather sites makes the usage of additional (fast) data processing tools unavoidable.

AWT – a novel (comp. fast) algorithm for clustering/outlier detection

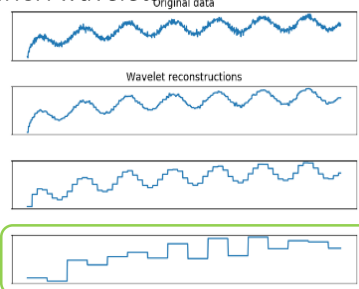
Basic idea:

creating a clustering feature tree from wavelet approximations with I-Kmeans applied to the approximations to improve the clustering result

Basis:

BIRCH and I-Kmeans with following changes and adaptations:

BIRCH wavelets



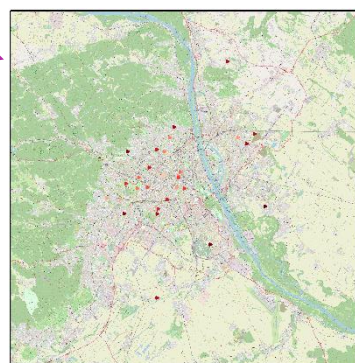
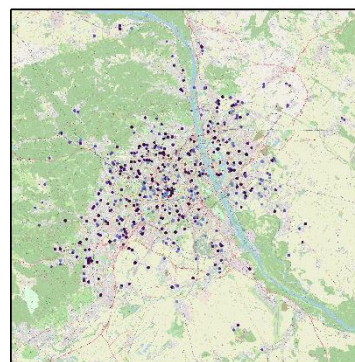
+ discard an approp. #of highest wavelet resolution levels
+ threshold for tightness clusters

cluster

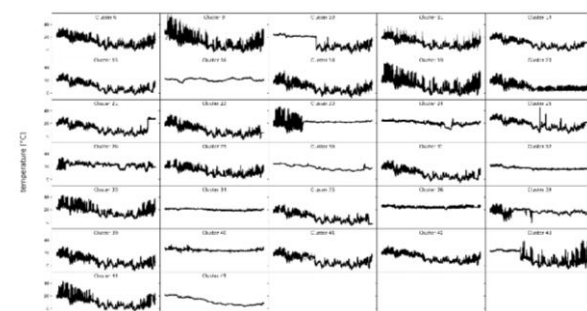
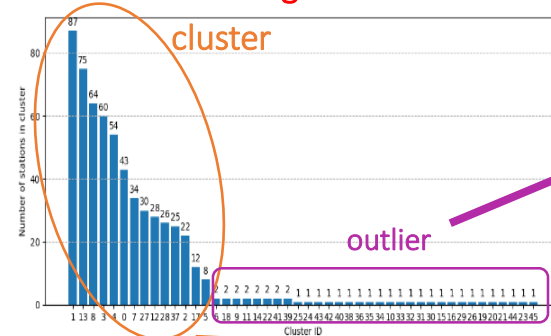
outlier via threshold

Extension to multivariate

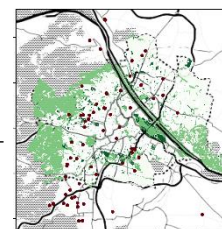
$$dist(X, Y) = \sum_{i=1}^n \sum_{j=1}^r (X[i][j] - Y[i][j])^2$$



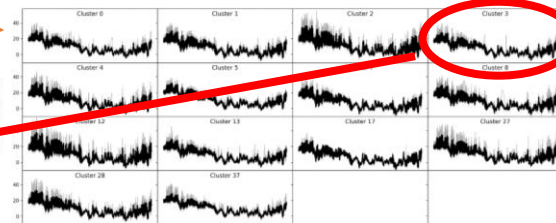
Results clustering and outlier detection for NetATMO sites in Vienna



Cluster 3 corresponds to places in Vienna with “green” land-use (e.g. parks, leisure)

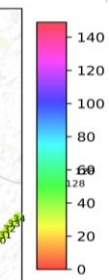
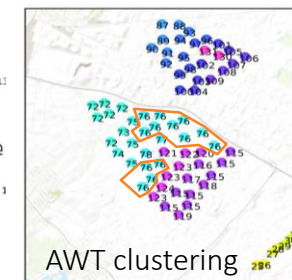
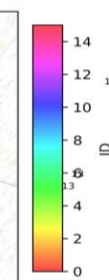
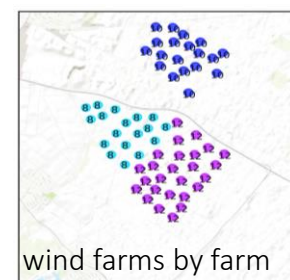


cluster 3



NetATMO temperatures September 2018 – February 2019

Results clustering wind turbines (sub-region shown)



Clustering of wind farm turbines for a period of 4 years of observations. AWT cluster indicate turbines with similar characteristics. → can be used to (i) enhance training pool for post-processing or (ii) if one turbine is missing replacing values be others of cluster.

I-KMEANS

(here, for comp. purposes using n = 46)