France regional electricity consumption clustering using Generalised Cross Correlation.

Pierre Mercatoris

<2018-05-19 Sat>

1 Introduction

1.1 Cluster electricity consumption using GCC

1.2 Clustering time series

2 Methodology

2.1 Data description

The electricity consumption was available at a 30 minutes frequency for each of the 12 regions of France from 2013 to 2017. Each year of each region can be downloaded from the French transmission operator (Rte) download portal[1].

Consumption from January 2013 to September of 2017 were downloaded for each of the 12 metropolitan mainland regions of France (excluding Corsica).

However, those regions are still very young, as before 2016, those were 21 separate regions. Regions in France lack separate legislative power, but can manage a considerable part of their budget for main infrastructures such as education, public transport, universities and research, and help to businesses. It is therefore expected to find some interesting clusters, where we might see some reminiscence of the old regions.

2.2 Data preparation

2.2.1 Cleaning

1. Append all regions and years together

2. Clean the region names

3. Format each column to appropriate data type

4. Set UTC time to correct summer/winter time changes

---

[1] http://www.rte-france.com/en/eco2mix/eco2mix-telechargement-en

5. Pivot table so that the columns are the regions and the rows are consumption values

6. Resample the date as 30 minutes intervals

7. Pivot the table again so that we get daily value for each row

8. Save to access from R

The complete data set was spread across 60 different tables that were merged into one large table.

Périmètre	Nature	Date	Heures	Consommation
Grand-Est	Données définitives	2016-01-01	00:00	5130
Grand-Est	Données définitives	2016-01-01	00:15	
Grand-Est	Données définitives	2016-01-01	00:30	5130
Grand-Est	Données définitives	2016-01-01	00:45	
Grand-Est	Données définitives	2016-01-01	01:00	5014
. . . ..				

As data rarely comes clean, there were some imperfections in the names of the data. Here for some days the regions were named after the old regions e.g. Languedoc-Roussillon et Midi-Pyrénées instead of Occitanie, or Aquitaine, Limousin et Poitou-Charentes instead of Nouvelle-Aquitaine.

With the raw data cleaned of imperfection, each column was formatted to required data type. The date needing to be set as UTC in order to avoid problems at the summer/winter time change. The data was resampled from 15 minutes to 30 minutes using a mean so as to deal with a problem of Centre-Val de Loire in September 2017 where 30 minutes consumption was give every 15 minutes.

Périmètre Datetime	Auvergne-Rhône-Alpes	Bourgogne-Franche-Comté	. . .
2013-01-01_00:00:00+00:00	NaN	NaN	. . .
2013-01-01_00:30:00+00:00	8173.0	2357.0	. . .
2013-01-01_01:00:00+00:00	7944.0	2289.0	. . .
2013-01-01_01:30:00+00:00	7896.0	2326.0	
2013-01-01_02:00:00+00:00	7882.0	2409.0	

The region with the highest consumptions are observed in Iles-de-France and the lowest in the Centre-Val de Loire. We can also clearly see yearly seasonality with higher consumption during winter times.
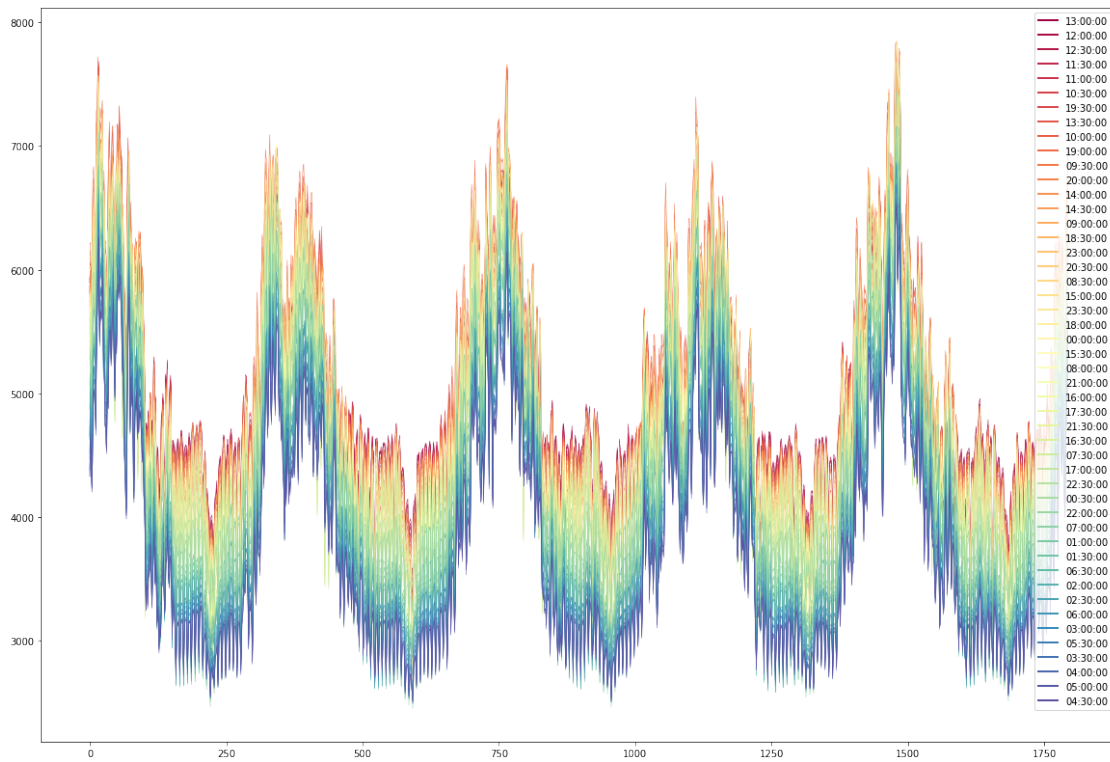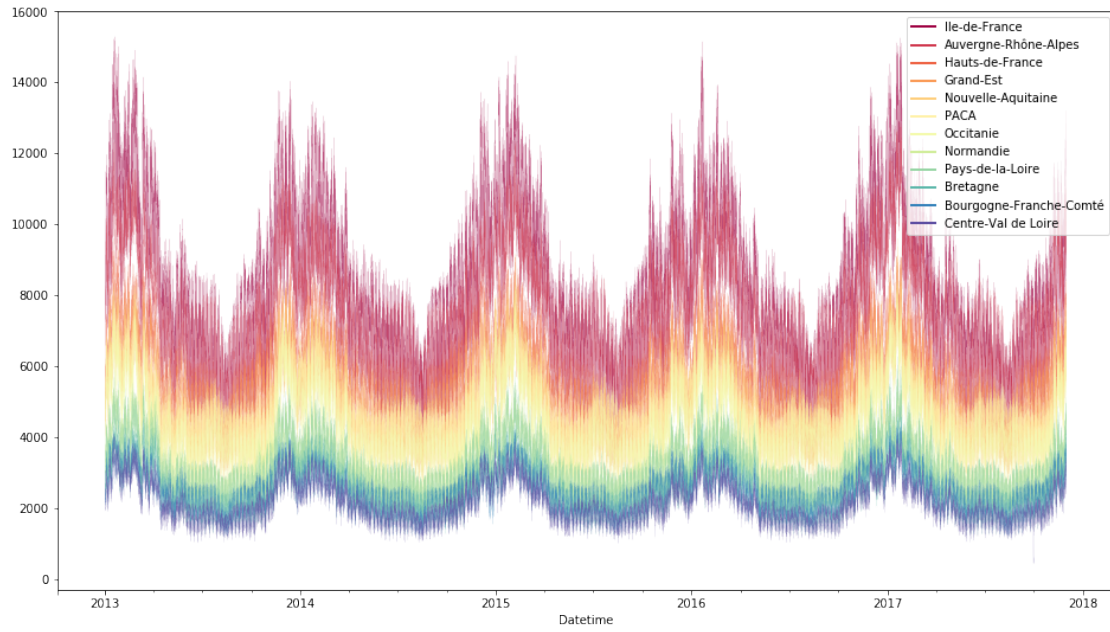
This resulted in a table of 576 columns (48 x 12 regions) and 1794 rows/days.

Périmètre	Auvergne-Rhône-Alpes					
time		00:00:00	00:30:00	01:00:00	01:30:00	02:00:00	02:30:00
2013-01-02		7847.0	7674.0	7427.0	7441.0	7467.0	7550.0
2013-01-03		9028.0	8839.0	8544.0	8560.0	8569.0	8667.0
2013-01-04		8982.0	8754.0	8476.0	8480.0	8453.0	8554.0
2013-01-05		8625.0	8465.0	8165.0	8134.0	8087.0	8149.0
2013-01-06		8314.0	8097.0	7814.0	7791.0	7785.0	7842.0

2.2.2 Transformation

2.3 GCC description

2.4 Distance calculation

1. Selecting k

2. Distance matrix

3 Results

3.1 Clustering

3.2 Cluster analysis

4 Conclusion