

1

CONTEXTE ET BASE DE DONNÉES

PRÉSENTATION DU JEU DE DONNÉES

Base de données = Cours de l'électricité dans 5 pays pour produire les gaz (azote, oxygène...)

5 pays : France - Belgique - Allemagne - Pays Bas - Portugal

1er janvier 2013 - 26 juin 2016 sauf Portugal : jusqu'au 30 avril 2017

1 observation = prix de l'électricité à une date et une heure précise



1 fichier = 1 pays
pour 1 année



date	heure_start	heure_end	price
2013-12-31	0	1	15.15
2013-12-31	1	2	19.02



16 fichiers

Portugal



1 seul fichier



0h	1h	...	23h
48.01	45	...	48.57
44.25	29	...	48.01

VALIDATION ET ENRICHISSEMENT DES DONNÉES



1 fichier unique



id	date	day_of_week	country	price
0	2013-12-31-00	tuesday	FR	15.15
1	2013-12-31-00	tuesday	GE	19.02

Valeurs aberrantes et manquantes : variable price

Price = -200

→ énergies renouvelables
Souvent un dimanche :
beaucoup de production pour peu
d'utilisation

Aucune
modification

date	day_of_week	country	price
2013-06-01-14	saturday	BE	NA
2013-03-31-23	sunday	PT	NA
2015-03-29-23	sunday	PT	NA

&

13 dates non inscrites

→ dû au changement d'heure

Imputation par les voisins :
(prix de l'heure précédente + prix de l'heure suivante) / 2

2

MÉTHODOLOGIE ADOPTÉE

Outils statistiques utilisés

LES TRANSFORMÉES

Les données
"brutes"

id	date	day_of_week	country	price
0	2013-12-31-00	tuesday	FR	15.15
1	2013-12-31-01	tuesday	FR	19.02

= série

prix_0h_
01/01/2013

prix_0h
02/01/2013

prix_1h_
01/01/2013

prix_1h
02/01/2013

...

prix_23h_
01/01/2013

prix_23h
02/01/2013

prix_0h
26/06/2013

...

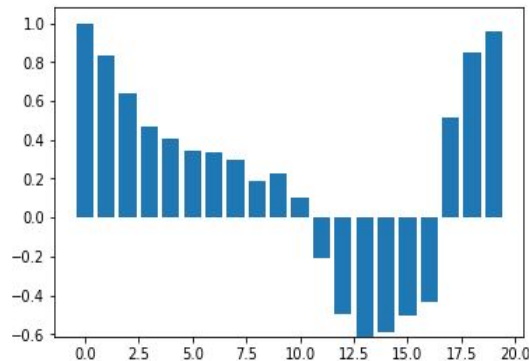
prix_23h
26/06/2016

Variance mobile

paramètre d'entrée : dispersion + ou - 2 heures



ACF



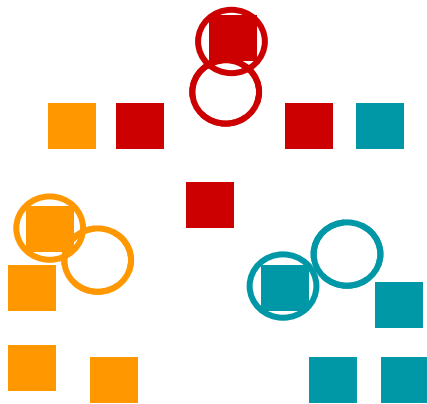
Transformée de Fourier discrète

CLUSTERING : L'ALGORITHME

Utilisation du clustering : classification non supervisée ⇒ **Clustering K-Means**

Algorithme itératif

- *Paramètre : nombre de clusters (k) à fixer à priori*
- *Critère d'arrêt : convergence des centroïdes*
- *Sortie : regroupement des séries temporelles similaires selon une distance choisie*



1.
**Choisir k centroïdes initiaux
aléatoirement**

CLUSTERING : LES MÉTRIQUES

⇒ Mesurer la similitude entre deux séries chronologiques

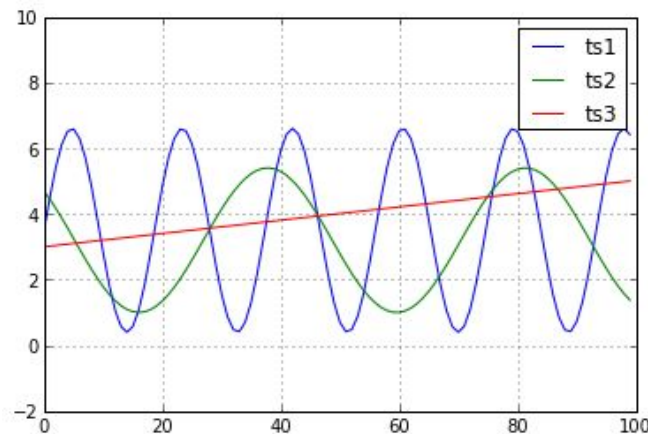
Utilisation d'une distance :

◆ **Euclidienne**
$$d(Q, C) = \sqrt{\sum_{i=1}^n [Q(i) - C(i)]^2}$$



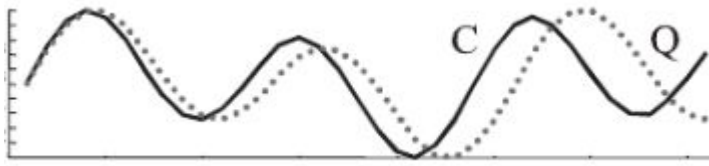
$$d(ts1, ts2) > d(ts1, ts3)$$

Contredit intuition → mesures de similarité pessimistes lors d'une distorsion dans l'axe des temps.



CLUSTERING : LES MÉTRIQUES

- ◆ **Dynamic Time Warping (DTW)** : trouver l'alignement non linéaire optimal entre deux séries temporelles

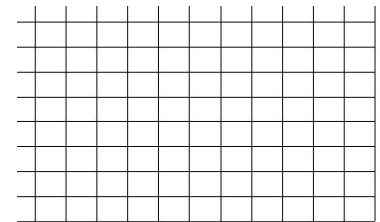
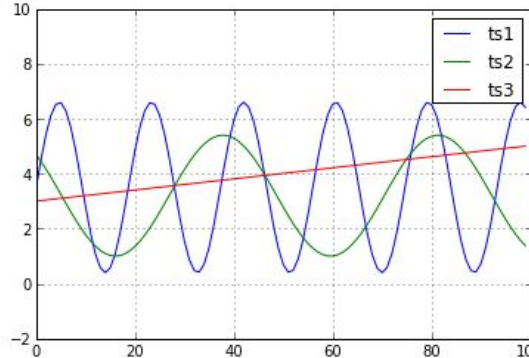
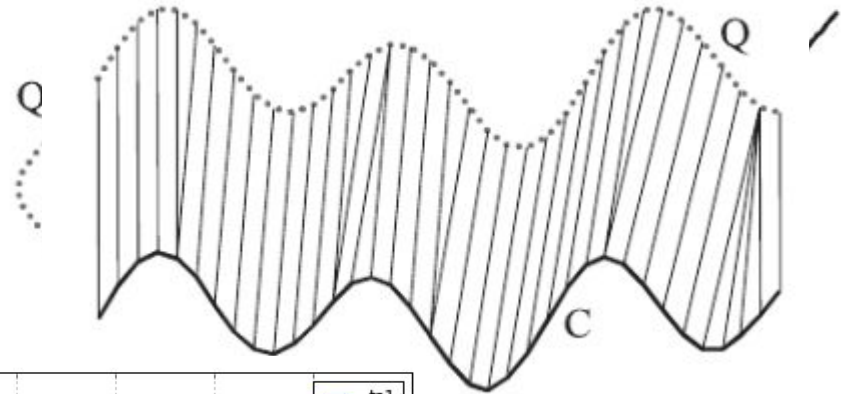


On cherche W^* tel que

$$W^* = \operatorname{argmin}_W \left(\sqrt{\sum_{k=1}^K w_k} \right) \quad w_k = (q_i - c_j)^2$$

En revenant à l'exemple précédent

$$\rightarrow d(ts1, ts2) < d(ts1, ts3)$$



3

IMPLÉMENTATION DU TABLEAU DE BORD

PRÉSENTATION DES DEUX BRIQUES DE VISUALISATION

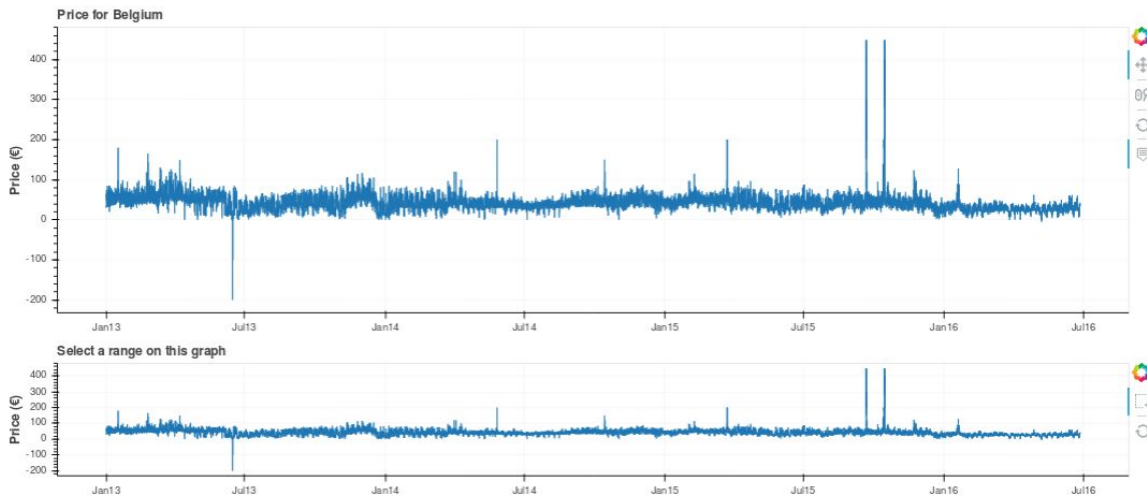
◆ Brique de visualisation générale : en fonction d'un pays et d'une plage de temps

Global visualization

Select the country of your choice and the desired range on the bottom graph.

Country :

Belgium



*Source de données :
Série complète (statique)*

PRÉSENTATION DES DEUX BRIQUES DE VISUALISATION

◆ Brique de clustering :

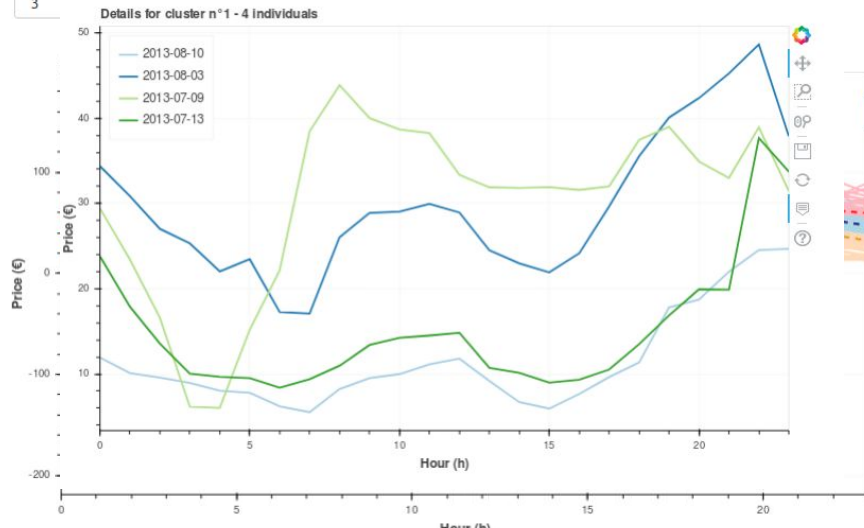
- Visualisation des profils quotidiens sur une plage spécifique
Sélection de différents paramètres :
 - Plage de temps
 - Pays
 - Type de jours
 - Nombre de cluster
 - Métrique
 - Transformée
- Visualisation d'un échantillon du cluster sélectionné

*Source de données :
Série filtrée selon plage (dynamique)*

Visu: Select a cluster to display its detail.

Select Cluster: Number of Individuals:
Start date:

Number:



STRUCTURE DES MODULES PYTHON

Initialisation et mise à jour

initialisation.py

Gestion de formats

module_selection.py

module_convert.py

module_treatment.py

module_data.py

Outils données

Outils graphiques

module_plot.py

module_widget.py

INTERACTION AVEC L'APPLICATION

initialisation.py

module_widget.py

module_plot.py

module_selection.py

module_selection.py

