



Classification d'images satellite

ALLAIRE Floriane, SABBAGH Guillaume, SUN Jian



Plan

- Présentation des données
- Analyse exploratoire
- Apprentissage non-supervisé
- Apprentissage supervisé
- Conclusion

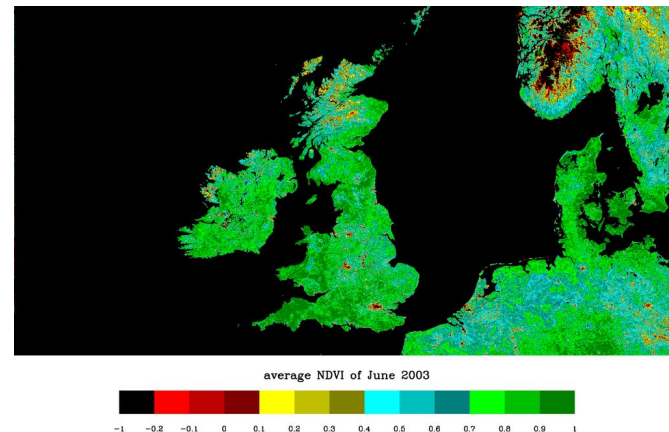
Le jeu de donnée

Séries temporelles de mesures NDVI sur plusieurs parcelles de terrain

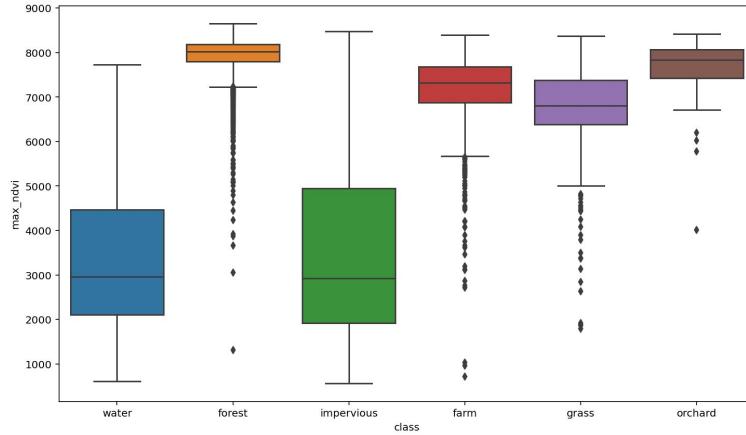
Permet de quantifier la végétation à distance

Classes : eau, verger, herbe, ferme, forêt et imperméable

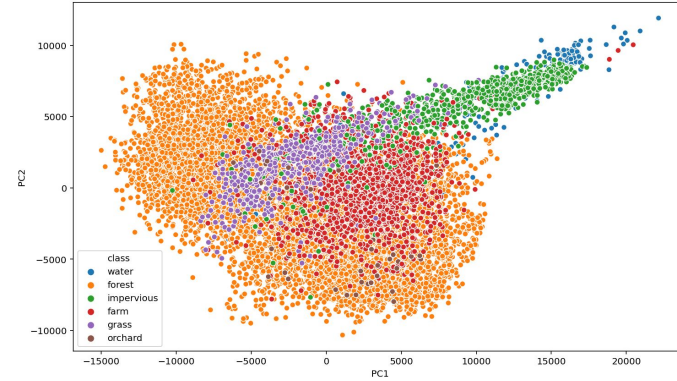
Données récoltées de 2014 à 2015 à intervalle d'environ 16 jours



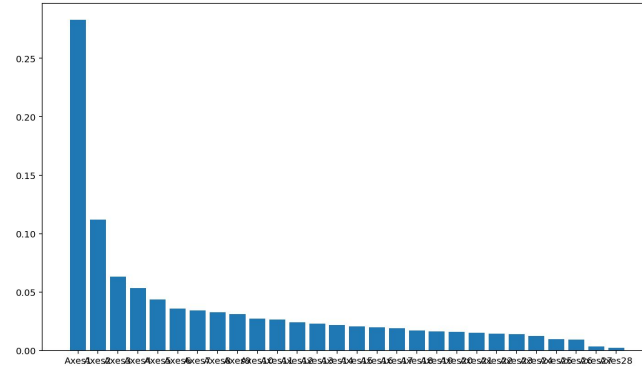
Analyse exploratoire



Répartition des valeurs NDVI max par classe

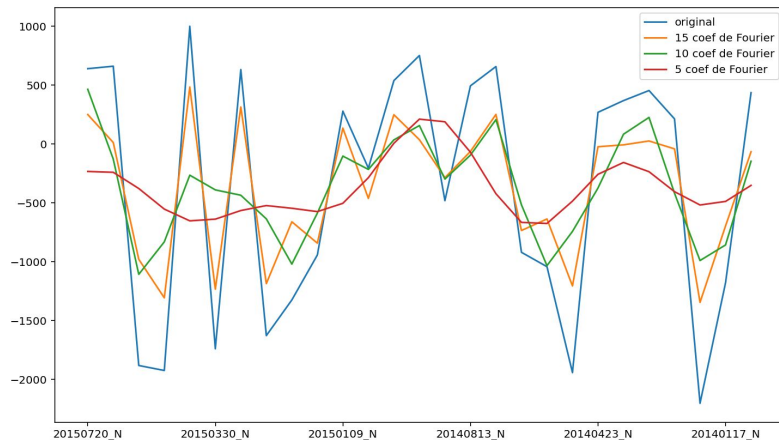


ACP des séries temporelles brutes sur deux axes

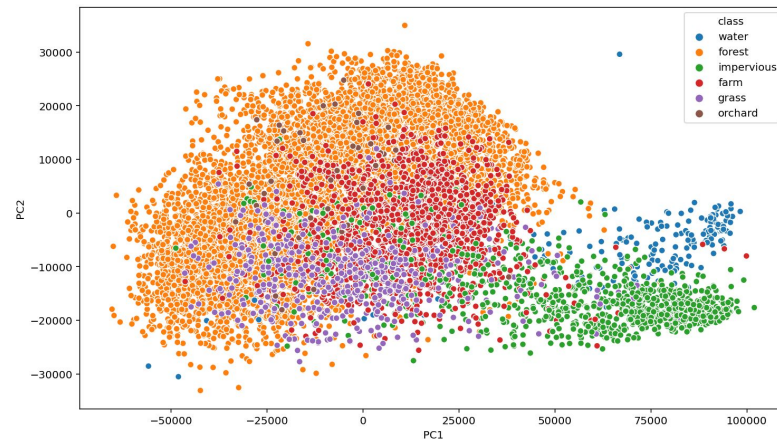


Variance expliquée

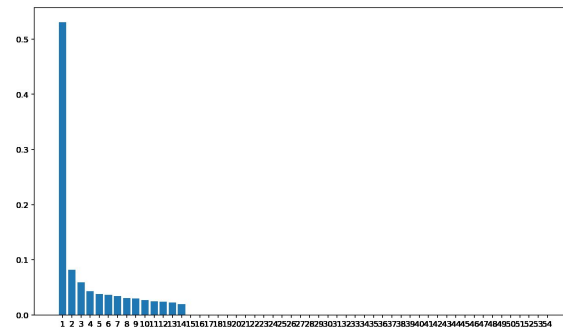
Analyse spectrale



Approximations de la série temporelle avec des filtres passe-bas



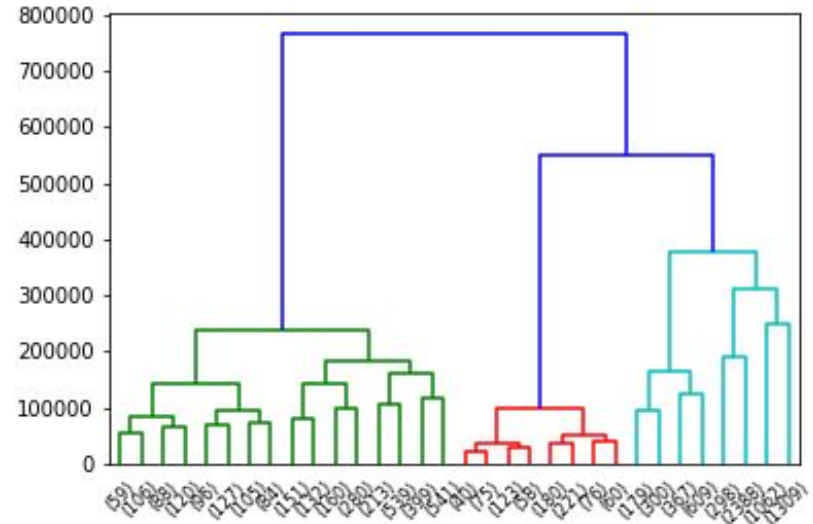
ACP après analyse spectrale



Variance expliquée

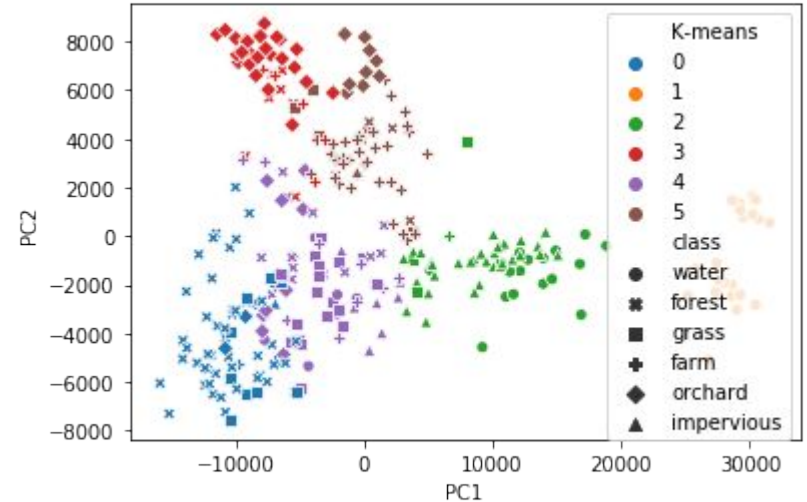
Classification ascendante hiérarchique

- Hyper-paramètres
 - Distance inter-classe
- Performance ARI
 - 13% avec méthode de Ward
- Causes
 - L'inertie faible inter-classe



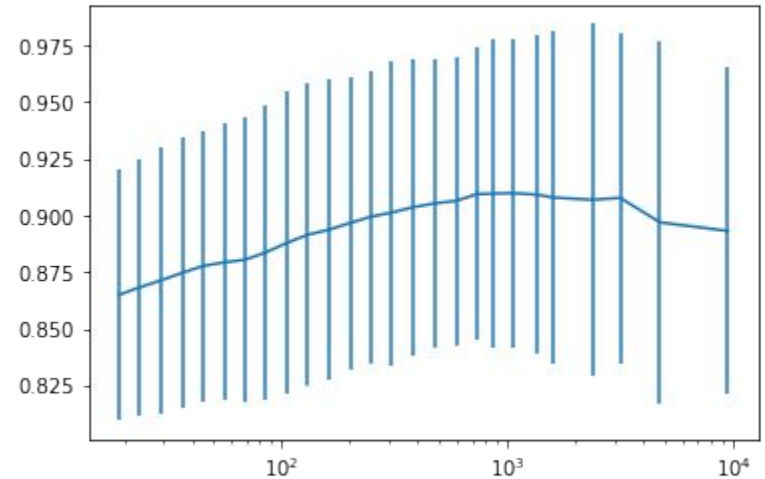
Classification K-means

- Nombre de classe pré-définie $k=6$
- Performance ARI
 - Comparaison des partitions : 33%
- Causes
 - Des formes géométriques peu distinctes
 - Des classes avec des points qui se superposent



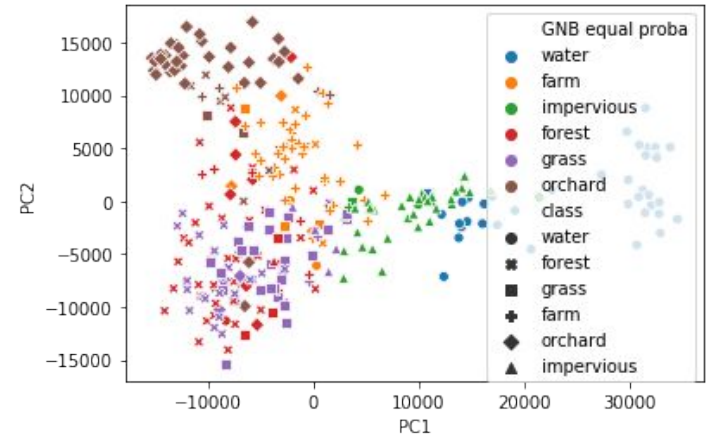
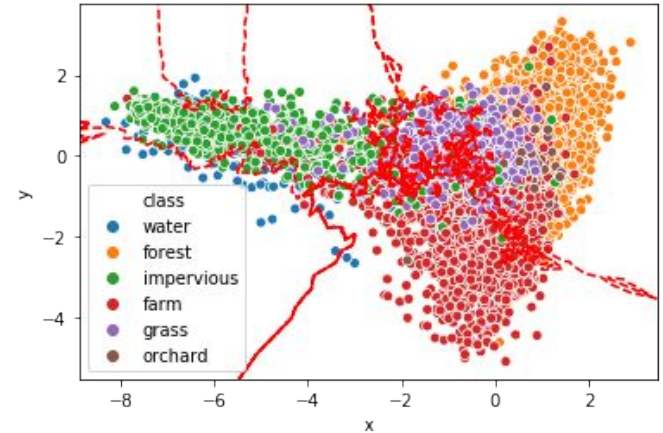
Les K plus proches voisins (K-PPV)

- Choix de K :
 - Trop petit : sur-apprentissage
 - Trop grand : trop grande généralisation
- Performance
 - $k=9$ avec une performance de 62.3%
 - Distance euclidienne
- Causes
 - Le fléau de la dimension
 - Les classes sont asymétriques (classes n'ayant pas toutes les mêmes proportions)
 - Correction envisageable avec pondération



Analyse discriminante

- Analyse Discriminante Linéaire
 - 59.33%
 - Hyper-paramètre : solveur svd
 - Réduction des dimensions
- Analyse Discriminante Quadratique
- Naïve Bayésienne
 - 70.00%
 - Minimiser les probabilités d'erreurs
 - Suppose que les variables sont indépendantes



Arbre

- Arbre de décision
 - Hyperparamètre
 - Rapport coût-complexité λ
 - Performance de prédiction
 - 56% avec $\lambda=0.00024$
 - Inconvénient
 - Sur-apprentissage
- Forêt aléatoire
 - Performance de prédiction
 - 63%
 - Inconvénient
 - Informations partiellement apprises
- 1000 x Forêts aléatoires
 - 1000 x retraits aléatoires et de nombre identique d'individus dans des classes
 - 1000 x votes
 - Performance de prédiction
 - 72%

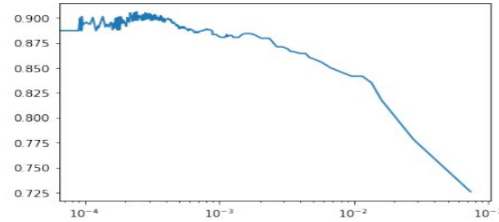


FIGURE 17 – Erreurs de validation en fonction des $\bar{\lambda}_k$

TABLE 2 – Matrice de confusion d'une forêt aléatoire sur les données brutes

	E	F	I	A	P	V
Eau	34	2	3	6	1	0
Forêt	0	69	1	6	2	0
Imper	0	1	36	2	1	0
Agri	0	11	0	42	0	0
Prairie	0	19	3	6	8	0
Verger	0	43	0	4	0	0

TABLE 3 – Matrice de confusion de 1000 répétition de forêt aléatoire sur les données brutes

	Fa.	Fo.	Gr.	Im.	Or.	Wa.
Farm	46	5	2	0	0	0
Forest	5	51	21	0	1	0
Grass	11	3	19	1	2	0
Impervious	1	1	0	38	0	0
Orchard	2	23	1	0	20	1
Water	3	1	2	2	0	38

SVM

- Normalisation
 - Z-score
 - Min-max
- Hyper-paramètres
 - régularisation C
- Performance de prédiction
 - 60% avec C= 50

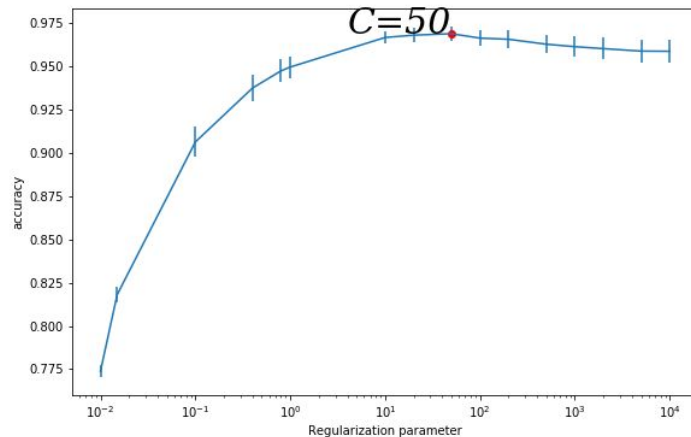
Min-max :

$$\text{value}^* = \frac{\text{value} - \min}{\max - \min}$$

Z-score:

$$\text{value}^* = \frac{\text{value} - \mu}{\sigma}$$

$$\text{Minimiser } \frac{1}{2} ||w||^2 + C \sum_{k=1}^p \xi_k, \quad C > 0$$





Conclusion

- Comparaison à une recherche similaire : “Land Cover Classification of Landsat Data”
- Difficultés
 - Bruit de donnée
 - Rapports temporels intervariables
 - Déséquilibre de classe
- Classifieurs
 - 70% avec **Gaussian NaiveBayes**
 - 72% avec **Random Forest**
 - 72% > 16% (choix aléatoire = 1/6)

Merci pour votre attention !



Plan

[10 min] => 3 min 30 par personne

- Présentation des données + Analyse Exploratoire [2min] ~ 2 Slides - *Guillaume*
- Coefficient de Fourier [1min30] ~ 1 Slide - *Guillaume*
- Apprentissage non-supervisé [1min30]
 - CAH ~ 1 Slide - *Sun*
 - K-Means ~ 1 Slide - *Floriane*
- Apprentissage supervisé [2min 30]
 - K-PPV ~ 1 Slide - *Floriane*
 - Analyse Discr ~ 1 Slide - *Floriane*
 - Support Vector Machine ~ 1 Slide - *Sun*
 - Arbre ~ 1 Slide - *Sun*
- Overture [30 sec] + Conclusion [1min] ~ 1 Slide - *Floriane & Sun*