Functional modeling of hyperspectral data with heteroscedastic noise

Application to statistical estimation and classification

Anthony Zullo [1,2]   Frédéric Ferraty[2]   Mathieu Fauvel[1]

[2016-05-12 Thu]

[1]Dynafor, Université de Toulouse, INRA, INPT, INP-EI Purpan, Castanet-Tolosan, France

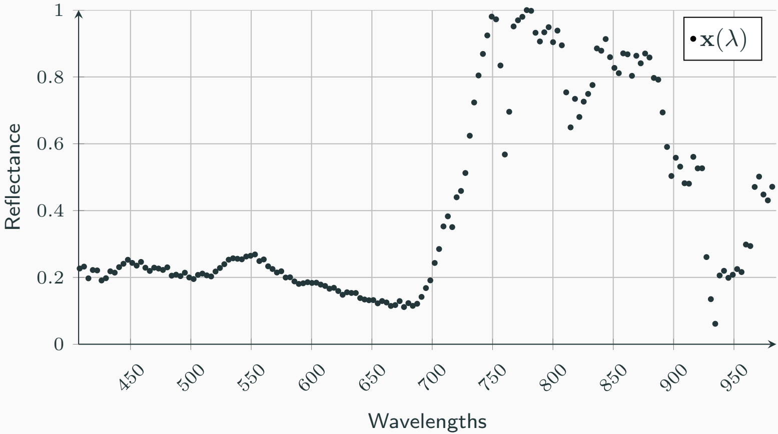
[2]UMR 5219 Institute of Mathematics, University of Toulouse, France

Outline

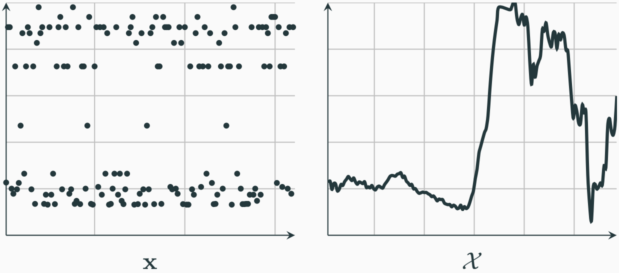
Introduction

Functional modeling of hyperspectral images

- Properties of hyperspectral images increase on a regularly basis
  - Higher spatial resolution,
  - Higher temporal resolution,
  - *Higher number of spectral channels*.



From spectral variables to spectral curves



Spectral variables

We observe a random vector $\mathbf{x}$ of $\mathbb{R}^d$ and our statistical processing (classification, unmixing, ...) is invariant to a random permutation of the spectral variables.

Spectral curves

We observe a random curve $\mathcal{X}$ of $\mathcal{F}$ and we can integrate some curves properties (derivative, smoothness) in the statistical processing.

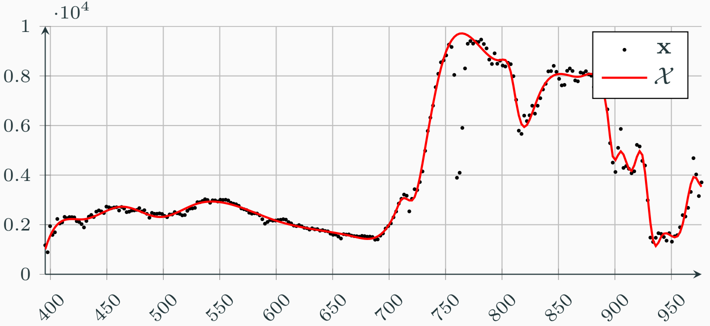
Multivariate versus functional modeling

Multivariate

- $\mathbf{x} = \left[\mathbf{x}(\lambda_1), \ldots, \mathbf{x}(\lambda_d)\right]$
- Card($\mathbb{R}^d$)=d
- $\mathbf{x}^{(m)} \rightarrow$ numerical differences
- $\langle \mathbf{x}, \theta \rangle = \sum_{i=1}^{d} \mathbf{x}(\lambda_i)\theta(\lambda_i)$

Functional

- $\mathcal{X} = \left\{\mathcal{X}(\lambda), \lambda \in [\lambda_{\min}, \lambda_{\max}]\right\}$
- Card($\mathcal{F}$)=∞
- $\mathcal{X}^{(m)} \rightarrow$ explicit formulae
- $\langle \mathcal{X}, \theta \rangle = \int_{\lambda_{\min}}^{\lambda_{\max}} \mathcal{X}(\lambda)\theta(\lambda) d\lambda$



Introduction

Contributions

Contributions of this work

- Nonparametric functional regression (and classification) applied to hyperspectral imagery.
- Heteroscedastic noise assumption in the observed spectra.
- Nested kernels estimator.
- Application to

Regression	Simulated PROSAIL data,
Classification	HYSPEX hyperspectral data.

Nested kernels estimator

Functional nonparametric regression

Non parametric model

- Learning set : $(\mathcal{X}_i, y_i)_{i=1}^n$.
- $y = r(\mathcal{X}) + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
- r is a regression operator with some regularity-type conditions

$$\hat{r}(\mathcal{X}) = \frac{\sum_{i=1}^n y_i K_r\left(h_r^{-1}\delta(\mathcal{X}, \mathcal{X}_i)\right)}{\sum_{i=1}^n K_r\left(h_r^{-1}\delta(\mathcal{X}, \mathcal{X}_i)\right)}$$

  - K_r is an asymmetric kernel,
  - $h_r \in \mathbb{R}_+^*$ is a smoothing parameter,
  - δ is a proximity measure between two curves.
- $\mathcal{X}$ is supposed noise-free, but in practice we only observe $\mathcal{X}^*$, *a contaminated version of the spectra* :

$$\mathcal{X}^*(\lambda) = \mathcal{X}(\lambda) + \eta(\lambda)$$

  where η is a random process independent of $(\mathcal{X}, y)$ such as
  - $E[\eta(\lambda)] = 0$,
  - $E[\eta(\lambda)\eta(\lambda')] = \sigma_\eta^2(\lambda)\mathbf{1}_{\lambda=\lambda'}$,
  - σ_η^2 twice differentiable.

Nested kernels approach

- Estimate both the noise-free spectra **and** the regression operator.
- Nesting two kernel estimators

$$\hat{\hat{r}}(\mathcal{X}) = \frac{\sum_{i=1}^{n} y_i K_r\Big(h_r^{-1}\delta(\mathcal{X}, \hat{\mathcal{X}}_i)\Big)}{\sum_{i=1}^{n} K_r\Big(h_r^{-1}\delta(\mathcal{X}, \hat{\mathcal{X}}_i)\Big)}$$

where $\hat{\mathcal{X}}_i$ is obtained through

$$\hat{\mathcal{X}}_i(\lambda) = \frac{\sum_{j=1}^{d} \mathcal{X}_i^*(\lambda_i) K_s\Big(h_s(\lambda_j)^{-1}(\lambda - \lambda_j)\Big)}{\sum_{j=1}^{d} K_s\Big(h_s(\lambda_j)^{-1}(\lambda - \lambda_j)\Big)}.$$

- h_s is a smoothing parameter depending on the variable λ.

Nested kernels estimator

Properties

Main theoretical results

Under some mild conditions, it is possible to prove

1. The nested estimator $\hat{\hat{r}}$ converges to the true operator r.
2. The rate of convergence is not decreased in comparison to the situation where noise-free samples are observed as soon as d is much larger than n.
3. The rate of convergence is increased when the ratio d/n is increased.

Application to the statistical analysis of hyperspectral images

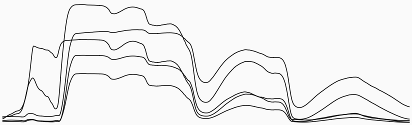
Estimation of Chlorophyll content

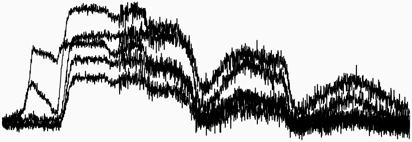
Data set

- Simulated data using PROSAIL (J.B. Féret).
- 5000 samples (n=5000) and 2101 wavelengths (d=2101) from 400 to 2500 nm.
- Heteroscedastic noise has been added.
- 500 spectra were randomly used to build $\hat{\hat{r}}$.
- 500 spectra were randomly used to compute the *relative mean square error* :

$$RMSE = \frac{\sum_{i=1}^{500}(y_i - \hat{\hat{r}}(\mathcal{X}_i))^2}{\sum_{i=1}^{500}(y_i - \bar{y})^2}$$

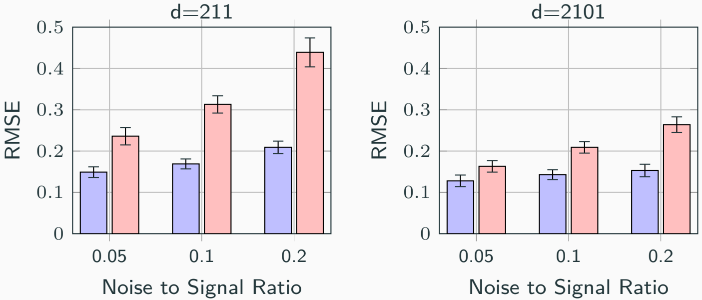
- Smoothing parameters have been optimized with 5-CV.
- 50 repetitions.





Results

- Proximity measure : L_2 *norm on the first derivative* w.r.t. the spectral variable of the spectra.
- Results : Nested kernels estimator and kernel estimator.



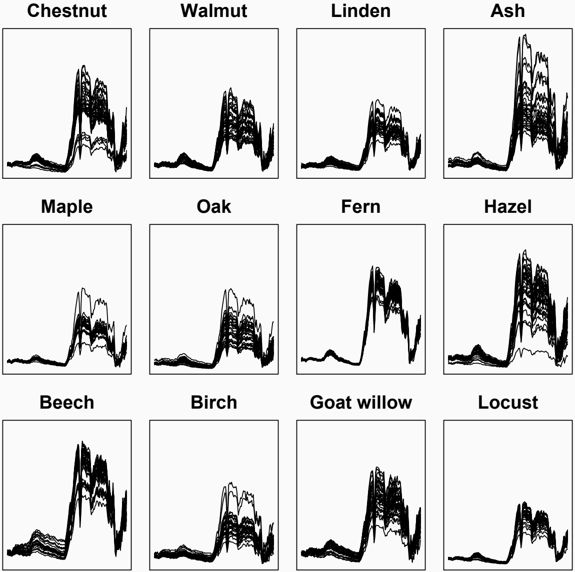
Application to the statistical analysis of hyperspectral images

Classification of hyperspectral images

Data set 1/2

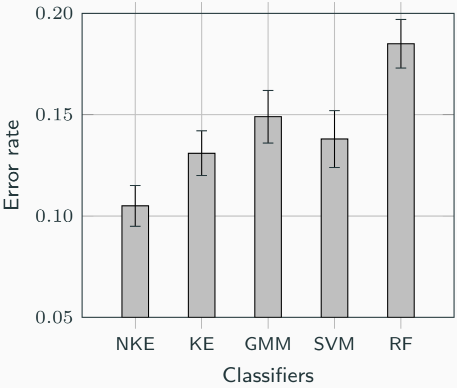
- Data set acquired with HYSPEX sensor.
- 32224 pixels (n=32224), with 50 cm as spatial resolution and 160 spectral bands (d=160).
- 12 woody species have been identified during field campaigns.
- Competitive methods were :
  - SVM,
  - GMM with ridge regularization,
  - Random Forest,

  B-Splines expansion has been used for the multivariate methods.
- 30 spectra per class used to build $\hat{\hat{r}}$, the remaining are used to compute the *error rate*.
- 50 repetitions.


Chestnut
Walmut
Linden
Ash
Maple
Oak
Fern
Hazel
Beech
Birch
Goat willow
Locust

Results

- Proximity measure : *PLS basis*.
- Results



Conclusions

✓ Functional modelling of spectral curves.

✓ Heteroscedastic noise model.

✓ Good performances w.r.t multivariate methods.

✓ Flexible framework to define proximity measures

- Derivatives,
- Subspaces (PCA, PLS . . .).

× Computing time.