# Bayesian fusion of hyperspectral astronomical images

A. Jalobeanu[1], M. Petremand[2], C. Collet[2]

[1] CGE, University of Evora, Portugal
[2] LSIIT UMR CNRS 7005, University of Strasbourg, France

DAHLIA project (ANR-08-BLAN-0253)

MaxEnt 2010

# Outline

# Outline

# MUSE: a new Integral Field Spectrograph (IFS)

## Observing the universe

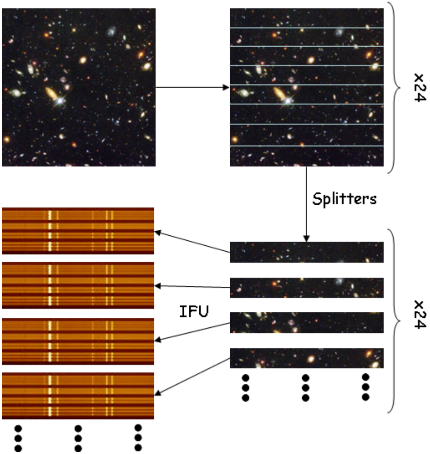
### Instrument specifications

- Mainly dedicated to the observation of distant galaxies
- Wide-field IFS: high spectral and spatial resolutions $\Rightarrow$ hyperspectral observations
- Spectral axis: 465 to 930nm, step 0.13nm $\sim$ 4000 samples
- Spatial axes: $1' \times 1'$ field of view $\sim 300 \times 300$ samples
- One observation: $300 \times 300 \times 4000$ pixels $\sim 1.2GB$

*Muse will be operational in 2012 on the VLT at Paranal, Chile*

# Inside MUSE

## Observation acquisition




### MUSE optics

A MUSE raw observation IS NOT a data cube $u = (x, y, \lambda)$ but a set of interlaced samples $p = (s, t, k)$:

- $s \Rightarrow$ spatial dimension ($\sim$ 4000 p.)
- $t \Rightarrow$ spectral dimension ($\sim$ 4000 p.)
- $k \Rightarrow$ IFU (24 CCD)

### Mapping & reconstruction

- Mapping between $(s, t, k)$ and $(x, y, \lambda)$ positions $\Rightarrow$ pixtable
- Sensor space $\Rightarrow$ Model space : reconstruction
- MUSE default reconstruction: DRS (*Data Reduction Software*)

# Using MUSE

## Galaxy observation

### Observing distant galaxies

- Study of such faint galaxies requires a long exposure time $\sim$ 80 hours
- Because of cosmic rays, an acquisition session cannot be longer than 1 hour! $\Rightarrow$ 80 observations of 1 hour each
- 80 observations $= 80 \times 300 \times 300 \times 4000$ p. $= 80 \times 1.2$GB
- Quite complicated to handle and analyze... $\Rightarrow$ Let's compute the average!

### Simple average: a very bad idea!

Between each acquisition, observational parameters have changed :

- Atmospheric conditions: PSF and spatial shifts
- Geometric fluctuations: spatial and spectral shifts
- Noise, cosmic rays, bad pixels
- Exposure time, sky transparency
- Sampling grids

One needs an optimal fusion algorithm $\Rightarrow$ Bayesian framework

# Bayesian Fusion

## Main features

### Specifications

- Data fusion: combine the raw observations by inverting a forward model ⇒ The knowledge of instrument design and parameters is crucial
- Bayesian framework ⇒ optimal data fusion
- Estimation of uncertainties on the fused image

### Issues to deal with...

- Resampling for the reconstruction of the observations over a common grid
- Preserving astrometry and photometry
- Size of the data (∼ 1.2GB/obs.) ⇒ critical issue for Bayesian approach
- Set of related acquisition parameters: PSF, variances, shifts, calibration, sampling grids... (∼ 5GB/obs.)
- Compromise between computing time and accuracy

# Outline

# Image formation

## From scene to sensor

The underlying "ground truth" $T$ is disturbed by :

### Atmosphere

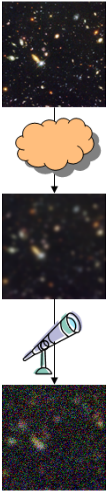
- variable spatial convolution (blur operator)

### Instrument & CCD sensor

- variable spatial convolution
- variable spectral convolution
- variable spectral and spatial shifts (due to IFU)
- spatial and spectral samplings
- acquisition noise
- missing data: dead pixels (known), cosmic rays (unknown locations)...

### And...

- integration time, sensor offset, sensitivity... (compensated by the radiometric correction)
- spatial shifts of the telescope between acquisitions



# Image formation

## From ground truth to observation

### From $T$ to $Y^i$ (after radiometric correction)

$$Y_p^i = (T \star h^i_{u_p^i})(u_p^i) + B_p^i$$

- $u_p^i$: 3D spatial-spectral sampling grid defined by the sampling geometry (shift, orientation)
- $h^i_{u_p^i}$ : 3D separable convolution kernel ($PSF \times LSF$) depending on $i$ and $u_p^i$
- $B_p^i \sim \mathcal{N}(0, \sigma_p^i)$ where $\sigma_p^i$ is a signal-dependent standard deviation
- $+$ cosmic rays (unknown locations) and bad pixels (known locations) $\Rightarrow$ setting $\frac{1}{\sigma_p^i} = 0$

### *Assumption*: $Y^i$ are band-limited

- Assumption: $Y^i$ are band-limited in space and wavelength and recovering the ground truth $T$ (not band-limited) from a set of $Y^i$ is therefore not possible!
- Our target: a band-limited version of $T \Rightarrow F = T \star \varphi$
- Spatial and spectral resolutions of $F$ are finite and fixed by the 3D kernel $\varphi = \varphi_x \times \varphi_y \times \varphi_\lambda$
- $\varphi$ corresponds to the PSF of an ideal instrument (better than MUSE) but how to choose $\varphi$?

# Image formation

## Choice of $\varphi$

### $\varphi$ = B-Spline function because:

- Finite and small footprint $\Rightarrow$ Fast implementation
- Nearly band-limiting functions meaning that $F$ is a good approximation of a band-limited signal [Unser]
- Third degree (cubic) B-Splines $\varphi$ : good compromise between accuracy and complexity

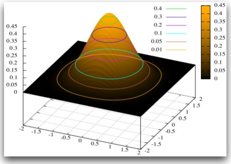
### Application

- Band-limiting: $F = T \star \varphi$
- Interpolation theory :

$$F(z) \simeq \sum_m L_m \varphi(z - m), z \in \mathbb{R}^3, m \in \mathbb{Z}^3$$

- $L$ is a discrete set of interpolation coefficients
- Our target $\Rightarrow$ discrete version of $F$ :

$$X_p = F(p) = (L \star \varphi)(p), p \in \mathbb{Z}^3$$



Target PSF (B-Spline 3)

# Image formation

## Application

### *Assumption*: PSFs are bandlimited

- In practice, PSFs are wider than the B-Spline
- The PSF can be written as a discrete sum of kernels weighted by B-Spline coefficients
- Then:

$$Y_p^i = (T \star h^i_{u_p^i})(u_p^i) + B_p^i = \sum_m L_m \alpha^i_{pm} + B_p^i \text{ where } \alpha^i_p = h^i_{u_p^i}(u_p^i - m)$$

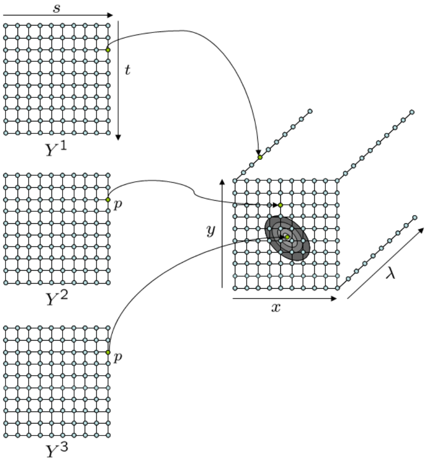
- The set $\alpha^i_p$ encodes, for each $p$, PSF, geometry and sampling grids and acts like a blur kernel
- $\alpha^i_p$ is almost perfectly known from calibration

### Linear forward problem: matrix notation

- $\mathbf{Y}^i = \boldsymbol{\alpha}^i \mathbf{L} + \mathbf{B}^i$ and $\mathbf{B}^i \sim \mathcal{N}(0, \mathbf{P}^{i\,-1})$ where $\mathbf{P}^i$ is the inverse covariance matrix of $\mathbf{Y}^i$
- $\mathbf{X} = \mathbf{SL}$ where $\mathbf{S}$ is the spline operator

# Image formation

## Understanding rendering coefficients




### $\alpha$ : Principle

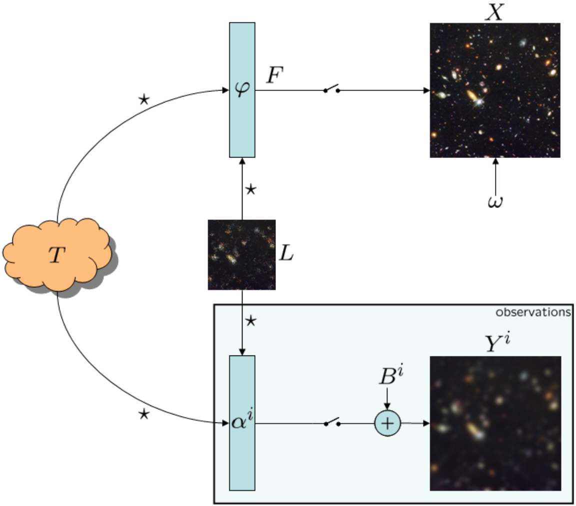
- Each $Y_p^i$ is a noisy combination of model space parameters $\boldsymbol{\alpha}^i \mathbf{L} + \mathbf{B}^i$

### $\alpha$ : Computation

- For each $p$ and depending on $\Theta^i$ $\Rightarrow$ a set of $\alpha^i$ for each $Y^i$
- Each parameter set $\Theta^i$ is included in $\alpha^i$: PSF, samplings, calibration...
- Theoretically $\Rightarrow$ huge number of coefficients (for 1 MUSE observation: 750 PB)
- Thresholding (for 1 MUSE observation: still 1.2 TB)

# Image formation

## Summary



# Outline

# Fusion

## Bayesian inference

Bayesian fusion ⇔ Maximize the *a posteriori* probability

$$P(\mathbf{L}|\{\mathbf{Y}^i\}_i, \omega) \propto \prod_i P(\mathbf{Y}^i|\mathbf{L}) \times P(\mathbf{L}|\omega)$$

### Bayesian inference

- $P(\mathbf{Y}^i|\mathbf{L}) \Rightarrow$ Likelihood (data driven term) $\Rightarrow \mathbf{Y}^i|\mathbf{L} \sim \mathcal{N}(\alpha^i \mathbf{L}, \mathbf{P}^{i\,-1})$
- $P(\mathbf{L}|\omega) \Rightarrow$ Prior on $\mathbf{X} = \mathbf{SL}$. For now, we use a simple first-order Markov Random Field but one could use more realistic priors (sparse, astronomical objects)

### Fusion: infer $\hat{\mathbf{L}}$ from the set $\{\mathbf{Y}^i, \alpha^i\}$ then $\hat{\mathbf{X}} = \mathbf{SL}$

- Minimize the energy function $U(\mathbf{L}) = -log\left(P(\mathbf{L}|\{\mathbf{Y}^i\}_i, \omega)\right)$
- Conjugate gradient algorithm (iterative minimization)

# Fusion

Deterministic, gradient-based energy minimization

$$\nabla_{\mathbf{L}} U(\mathbf{L}) = \underbrace{\left(\sum_i {\alpha^i}^T \mathbf{P}^i \alpha^i\right)}_{②} \mathbf{L} - \underbrace{\sum_i {\alpha^i}^T \mathbf{P}^i \mathbf{Y}^i}_{①} + 2\omega \mathbf{Q}\mathbf{L}$$

## Dealing with large datasets

- Solving $\nabla_{\mathbf{L}} U(\mathbf{L}) = 0$ ⇔ Evaluation of sums ① and ② for each iteration ⇒ time consuming!
- Implementation: pre-compute ① and re-compute ② to avoid storage issues

## ① Drizzling-like term

$$\mathbf{\Lambda}^f = \sum_i {\alpha^i}^T \mathbf{P}^i \mathbf{Y}^i$$

- Applying $\mathbf{P}^i$ ⇒ Inverse variance weighting
- Applying ${\alpha^i}^T$ ⇒ Shift cancellation and re-blurring to form a geometrically consistent result

# Fusion

## Energy minimization

$$\nabla_{\mathbf{L}} U(\mathbf{L}) = \underbrace{\left(\sum_i \boldsymbol{\alpha}^{i\,T} \mathbf{P}^i \boldsymbol{\alpha}^i\right)}_{②} \mathbf{L} - \underbrace{\sum_i \boldsymbol{\alpha}^{i\,T} \mathbf{P}^i \mathbf{Y}^i}_{①} + 2\omega \mathbf{Q}\mathbf{L}$$

### Dealing with large datasets

- Solving $\nabla_{\mathbf{L}} U(\mathbf{L}) = 0$ ⇔ Evaluation of sums ① and ② for each iteration ⇒ time consuming!
- Implementation: pre-compute ① and re-compute ② to avoid storage issues

### ② Data precision matrix $\boldsymbol{\Lambda}^f$

$$\boldsymbol{\alpha}^f = \sum_i \boldsymbol{\alpha}^{i\,T} \mathbf{P}^i \boldsymbol{\alpha}^i$$

- Computation of $\boldsymbol{\alpha}^f$ is highly time consuming and mainly depends on the size of the PSF
- Size of $\boldsymbol{\alpha}^f$ is higher than the size of each $\boldsymbol{\alpha}^i$

# Deconvolution

## Estimation of $\hat{\mathbf{X}}$ and uncertainties

### Minimization of $U(\mathbf{L})$

- Conjugate gradient
- Fixed $\omega$ (weight of the prior): estimation from complete data or ideal image. *Automatic estimation may be highly time-consuming due to the size of the data (under investigation)*
- After convergence, we get $\hat{\mathbf{X}} = \mathbf{S}\hat{\mathbf{L}}$

### Estimation of uncertainties on $\hat{\mathbf{X}}$: precision matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$

- Approximation: posterior distribution of $\mathbf{X}$ is a multivariate Gaussian: $\mathbf{X}|\{\mathbf{Y}^i\}_i, \omega \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}})$
- Inverse covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \Rightarrow$ Second derivatives of the log-pdf at the optimum $\Rightarrow \nabla^2_{\mathbf{X}} U(\mathbf{X})$
- With $\mathbf{L} = \mathbf{S}^{-1}\mathbf{X}$ :

$$\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} = {\mathbf{S}^{-1}}^T \alpha^f \mathbf{S}^{-1} + 2\omega {\mathbf{S}^{-1}}^T \mathbf{Q} \mathbf{S}^{-1}$$

# Uncertainties

## Use of uncertainties

### More about $\mathbf{\Sigma}_{\mathbf{X}}^{-1}$

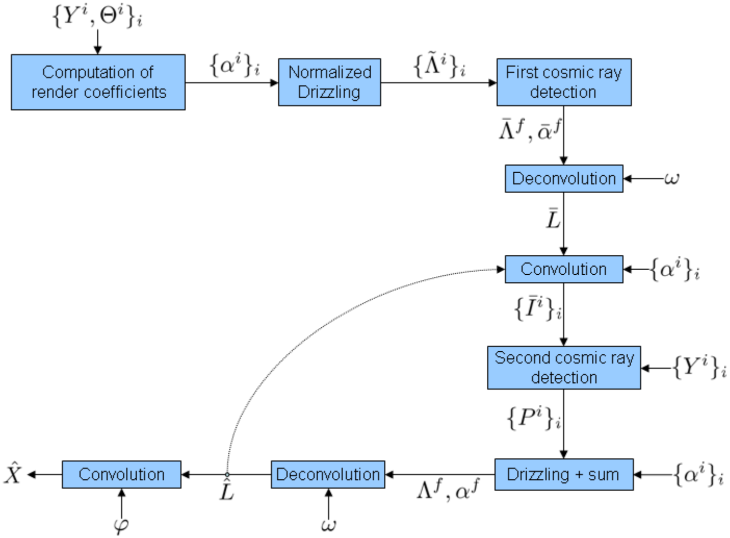
- Large sparse matrix: closely related to $\alpha^f$
- Same storage as $\alpha^f$: list of non-zero values
- The inverse covariance matrix is computed after the deconvolution
- Use of $\mathbf{\Sigma}_{\mathbf{X}}^{-1}$ for further investigations: denoising, new fusion...

### More about $\mathbf{\Sigma}_{\mathbf{X}}$

- Require the inversion of the large matrix $\mathbf{\Sigma}_{\mathbf{X}}^{-1}$
- Can be performed for the neighborhood of the desired pixel $i$ using a conjugate gradient algorithm
- One can only focus on variances and nearest neighbour covariances
- Additional information can be found in [Jalobeanu, Gutierrez]

# Fusion pipeline

## Fusion diagram



# Outline

# Preliminary results

## Results ♯1

### Simulated data using simple astronomical objects

- For the moment, we do not have access to real data, but accurate simulations of the MUSE instrument will be available in a few weeks
- We have developed a little "toy model" allowing us to simulate raw astronomical observations with variable parameters (spatial and spectral shifts, variable PSF, noise, IFU number...) containing simple gaussian objects (stars and galaxies)

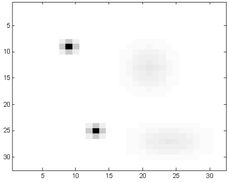
### Dataset

- The ground truth **T** is composed of 4 objects : two stars (spatial dirac with a spectrum composed of a gaussian/dirac mixture) and two galaxies (gaussian spatial profile with a spectrum composed of a gaussian/dirac mixture)
- Four $32 \times 32 \times 32$ observations with different PSF, variable spatial shifts, constant noise :

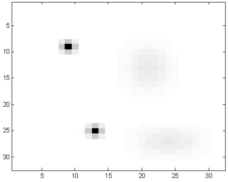
| ♯ | $PSF_{\lambda_0}$ | $PSF_{\lambda_n}$ | $LSF_{\lambda_0}$ | $LSF_{\lambda_n}$ | Spatial shifts $(x, y)$ | SNR (Star, Galaxy, Total) |
|---|---|---|---|---|---|---|
| 1 | 1.4 | 1.96 | 1.8 | 1.9 | $(0, 0)$ | (57, 38, 44) |
| 2 | 1.6 | 2.24 | 1.4 | 1.46 | $(1.2, 1.4)$ | (56, 38, 43) |
| 3 | 1.4 | 1.96 | 1.4 | 1.46 | $(0.4, 0.5)$ | (57, 38, 44) |
| 4 | 1.7 | 2.38 | 1.7 | 1.8 | $(0.2, 0.3)$ | (55, 38, 43) |

# Results ♯1

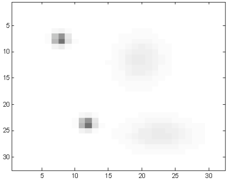
## Band 1



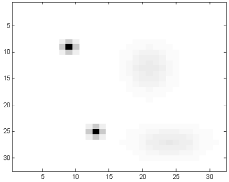

(a) $X$



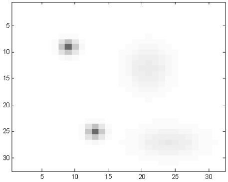

(b) $Y^1$



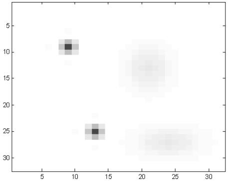

(c) $Y^2$




(d) Bayesian fusion



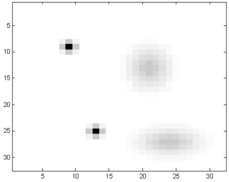

(e) Linear interp.



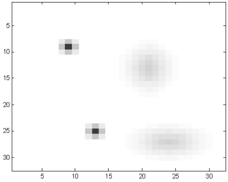

(f) B-Spline interp.

# Results ♯1

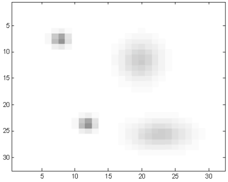
## Band 13



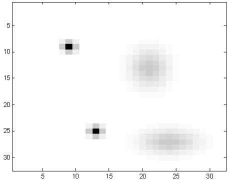

(a) $X$



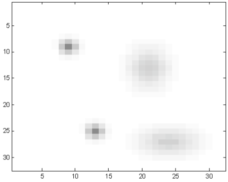

(b) $Y^1$



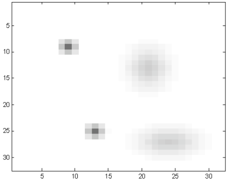

(c) $Y^2$




(d) Bayesian fusion



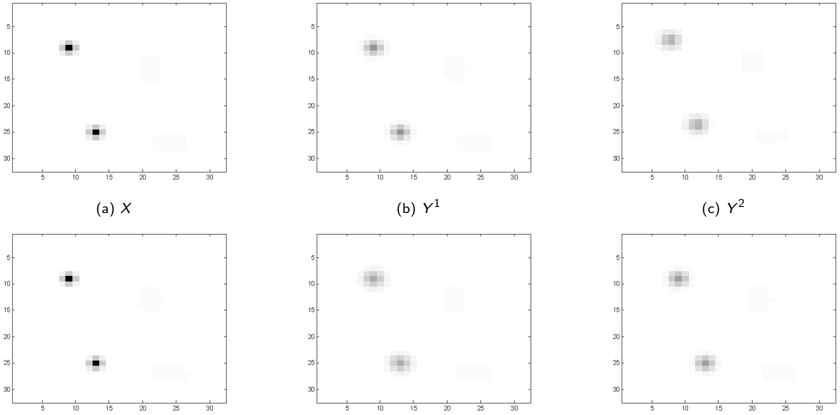

(e) Linear interp.




(f) B-Spline interp.

# Results ♯1

## Band 32

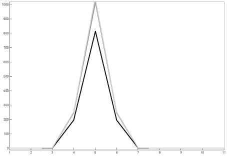



(a) $X$ (b) $Y^1$ (c) $Y^2$

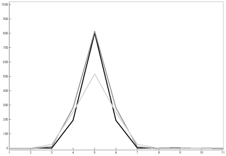
(d) Bayesian fusion (e) Linear interp. (f) B-Spline interp.

# Results ♯1

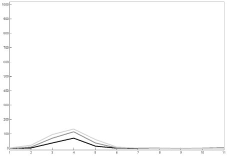
Star profiles. black : band 1. gray : band 13. light gray : band 32



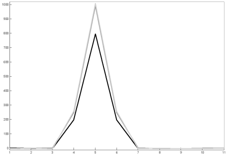
(a) $X$



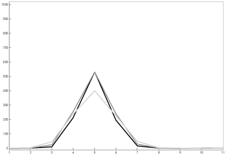
(b) $Y^1$



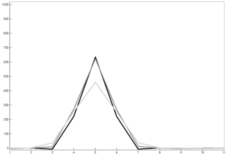
(c) $Y^2$



(d) Bayesian fusion



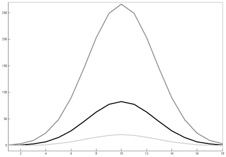
(e) Linear interp.



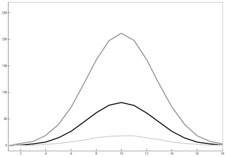
(f) B-Spline interp.

# Results ♯1

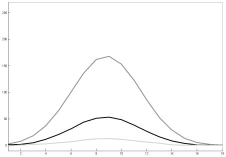
Galaxy profiles. black : band 1. gray : band 13. light gray : band 32



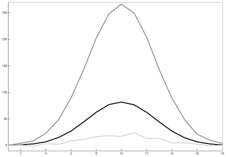
(a) $X$



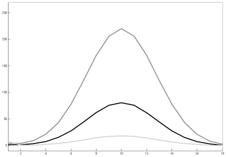
(b) $Y^1$



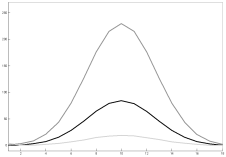
(c) $Y^2$



(d) Bayesian fusion



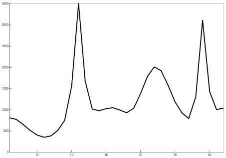
(e) Linear interp.



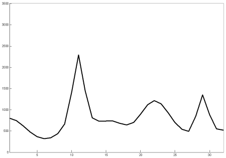
(f) B-Spline interp.

# Results ♯1

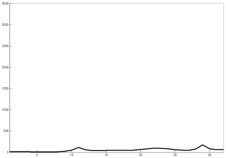
## Star spectra



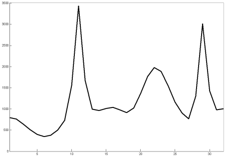
(a) $X$



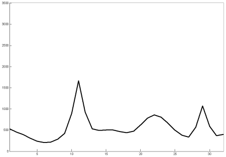
(b) $Y^1$



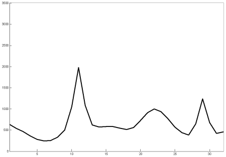
(c) $Y^2$



(d) Bayesian fusion



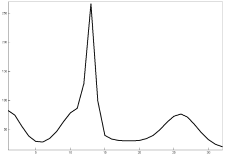
(e) Linear interp.



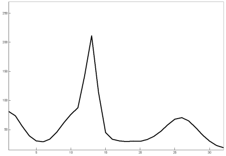
(f) B-Spline interp.

# Results ♯1

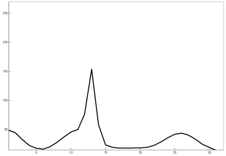
## Galaxy spectra



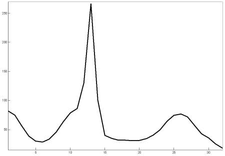
(a) $X$



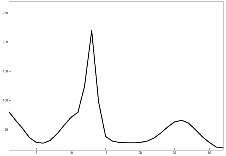
(b) $Y^1$



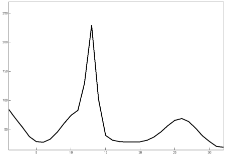
(c) $Y^2$



(d) Bayesian fusion



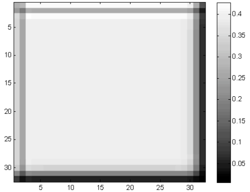
(e) Linear interp.



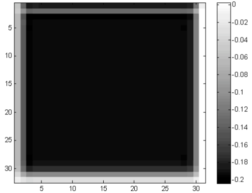
(f) B-Spline interp.

# Results ♯1

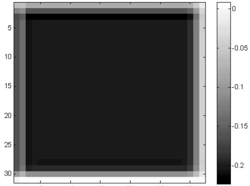
## Covariances : band 16



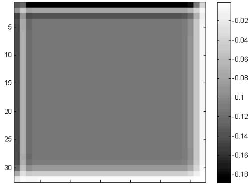

(a) Var.




(b) Covar. : right neighbors



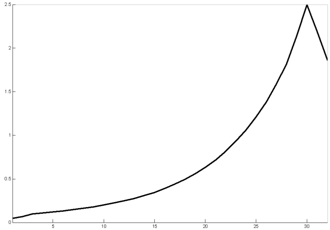

(d) Covar. : bottom neighbors



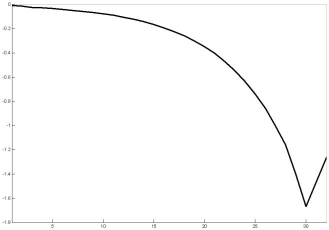

(e) Covar. : front neighbors

# Results ♯1

## Covariances : spectrum at (16, 16)




(a) Var.




(b) Covar. : right neighbors

# Conclusion and perspectives

## Conclusion

- Fusion and reconstruction of complex hyperspectral observations (with various PSF, shifts...) within a rigorous Bayesian framework
- Uncertainty computation using a deterministic approach
- Management of large datasets (raw data and parameters)
- Ability to deal with additional observations

## Perspectives

- Implementation of a 2-step detection of cosmic rays
- "Play" with the simulations: add observations, spectral shifts, higher noise, larger blur size and check the robustness of the method
- Development of the pipeline for real observations (scaling)
- Visualization of the variances
- Improve the prior on $X$