

# A LA RECHERCHE DES FILAMENTS INTERGALACTIQUES

Raphaël Bacher, Florent Chatelain, Olivier Michel, Roland Bacon

Financé par l'ERC MUSICOS



gipsa-lab



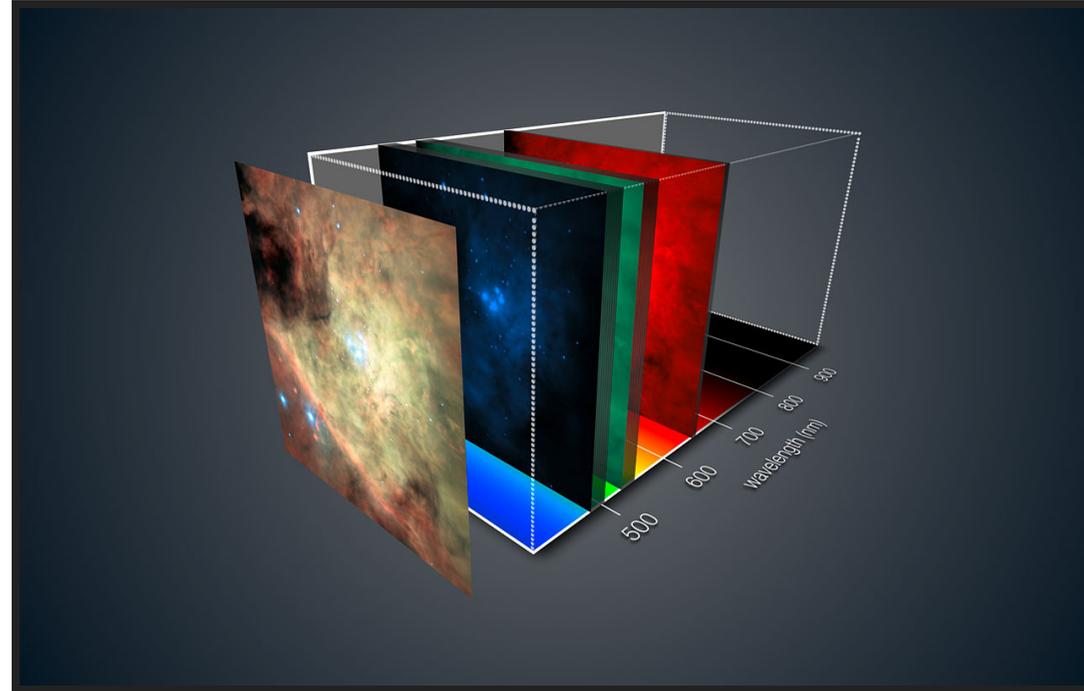
- Contexte
- Detection du Circum-Galactic-Medium (CGM)
- Pré-traitement : Estimation et soustraction spectre continu
- Résultats

# CONTEXTE

# LE PROJET MUSE

- Spectrographe 3D
- Consortium européen, initié par le CRAL à Lyon
- Fonctionnel depuis janvier 2014, déjà l'un des instruments les plus demandés du VLT

# DE L'HYPERSPECTRAL PAS COMME LES AUTRES



- On regarde dans l'autre sens...
- Vraiment "hyper"-spectral : 3600 longueurs d'onde (de 470 à 930 nm)
- Champ de 1 arcmin<sup>2</sup> (300 par 300 pixels)
- RSB extrêmement faible
- Pas de vérité terrain
- Problème du redshift

# DETECTION DU CIRCUM-GALACTIC-MEDIUM (CGM)

# LE CGM

On s'intéresse aux ensembles Galaxie + Halo de gaz

Caractéristiques d'une galaxie

- Quasi-ponctuelle (de la taille de la PSF)
- Contient de l'hydrogène + d'autres composants
- Donc émet la raie Lyman + une émission continu (+ éventuellement d'autres raies)
- On connaît sa position et son redshift/la position raie de Lyman

Caractéristiques du halo:

- Nuage de gaz (quasi-exclusivement hydrogène)
- Donc émet dans la raie de Lyman (pas de composante continue)
- S'étend à partir de la galaxie
- La raie Lyman proche spectralement de celle de la galaxie

# MÉTHODE DE DÉTECTION

A partir d'une galaxie :

- Recherche de la raie Lyman dans le voisinage spatial
- Spectre recherché : proche du spectre de la galaxie (faible décalage/déformation possible)
- Approche par test d'hypothèse

# TEST DE DÉTECTION

→ Construction d'un spectre de référence  $r_0$ : Moyenne spatiale des spectres du centre de la galaxie.

→ Test à partir de la Spectral Angular Distance (SAD)

$$SAD(\mathbf{r}_0, \mathbf{x}) = \frac{\mathbf{r}_0 \cdot \mathbf{x}}{\|\mathbf{r}_0\| \|\mathbf{x}\|}$$

Hypothèse :

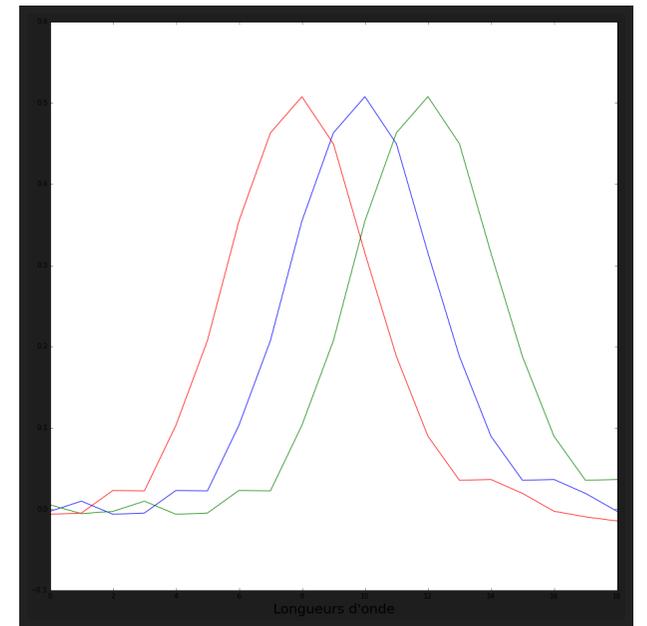
$$\begin{cases} H_{0i} : T_0(\mathbf{x}_i) = SAD(\mathbf{r}_0, \mathbf{x}_i) = 0 \\ H_{1i} : T_0(\mathbf{x}_i) \neq 0 \end{cases}$$

# TEST DE DÉTECTION SUR UNE FAMILLE DE RÉFÉRENCE

Pour prendre en compte la variabilité spectrale au sein du halo :

- A partir de  $r_0$  on construit une famille de spectres translatés  $\{\mathbf{r}_j\}_{1 \leq j \leq k}$ .
- Calcul du SAD pour chaque pixel sur tous les éléments de ce "dictionnaire".
- Test pour chaque pixel sur le maximum sur l'ensemble du dictionnaire du SAD.

$$\begin{cases} H_{0i} : T_{max}(\mathbf{x}_i) = \max_j SAD(\mathbf{r}_j, \mathbf{x}_i) = 0 \\ H_{1i} : T_{max}(\mathbf{x}_i) \neq 0 \end{cases}$$



# LOI SOUS $H_0$

Si on suppose que sous  $H_0$  le bruit est symétrique :

Sous  $H_0$ , la loi des opposés des minimums de corrélations devrait être identique à la loi des maximums.

On prend donc la loi empirique de l'opposé des mins comme loi pour  $H_0$ .

En notant  $F$  la loi empirique de distribution des opposés des minimums, on peut alors calculer la p-valeur associée au pixel  $x_i$ ,  $p(x_i) = 1 - F(x_i)$ .

# CONTRÔLE DE L'ERREUR

- Situation de tests multiples ( ~2500 pixels testés)
- La PFA n'est pas donc adaptée → Utilisation du False Discovery Rate [Benjamini & Hochberg 1995]

Pour avoir un FDR de  $\alpha$  lorsque qu'on fait  $m$  tests indépendants:

- Trouver le plus grand  $k$  tel que  $p_{(k)} \leq \frac{k}{m} \alpha$
- Rejeter l'hypothèse nulle pour tout  $x_{(i)}$  *for*  $i = 1, \dots, k$

On s'assure alors qu'au plus  $\alpha \times m$  tests sont rejetés à tort (faux positifs).

Dans notre cas (dépendance faible des tests ) : contrôle asymptotique assuré [Storey 2004]

# **PRÉ-TRAITEMENT : ESTIMATION ET SOUSTRACTION SPECTRE CONTINU**

# ENJEU

On cherche à soustraire la composante continu de la composante "raie" d'émission. Ainsi :

- Un pixel de la galaxie étudiée ne contient alors plus que la raie
- Un pixel de bruit reste un pixel de bruit
- Un pixel d'une galaxie autre (définie par le fait qu'elle ne contient pas la même raie spectrale) ne contient plus que du bruit résiduel

Pour l'instant, estimation du continu par un filtre médian

# MÉTHODE PROPOSÉE

Regression locale robuste : Adaptive Trimmed Least Squares

Dans l'approche Least Trimmed Squares (LTS), on cherche non pas à minimiser  $J = \sum_{i=1}^N r_i$  (moindres carrés) mais plutôt

$$J = \sum_{i=1}^k r_{(i)}$$

avec  $k$  fixé et  $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(k)} \dots \leq r_{(N)}$  les résidus.

A  $k$  fixé, le problème se met sous la forme :

$$\left\{ \begin{array}{l} \min_{\beta, w} \sum_{i=1}^n w_i (y_i - x_i^T \beta)^2 \\ \|\mathbf{1} - w\|_0 \leq k, w_i \in \{0, 1\} \end{array} \right.$$

# ALGORITHME

*Entrée* :  $X, \mathbf{y}$  (données),  $k$  (nombre de non-outliers)

*Boucle* :

Tant que  $\hat{\beta}_j \neq \hat{\beta}_{j-1}$

$$\hat{\beta}_j = (X^T W_j X)^{-1} X^T W_j \mathbf{y}, W_j = \text{diag}(\{w_{j,i}\}_i)$$

Calcul des résidus  $r_j = \mathbf{y} - X\beta_j$

Mise à jour  $w_j$

*Fin*

$$\text{Sortie } \hat{\mathbf{y}} = X\beta_j$$

Transformation d'un problème non-convexe en une alternance d'une étape convexe et d'une étape combinatoire simple (algorithme Prédiction-Correction type KMeans)

# ADAPTATIVITÉ : ESTIMATION DU NOMBRE DE NON-OUTLIERS $k$

On estime à chaque fois le nombre de non-outliers dans la fenêtre :

- Initialisation de  $\sigma$  à l'aide du Median Absolute Deviation et d'un LTS à 50%
- On cherche le nombre de résidus qui suivent cette variance

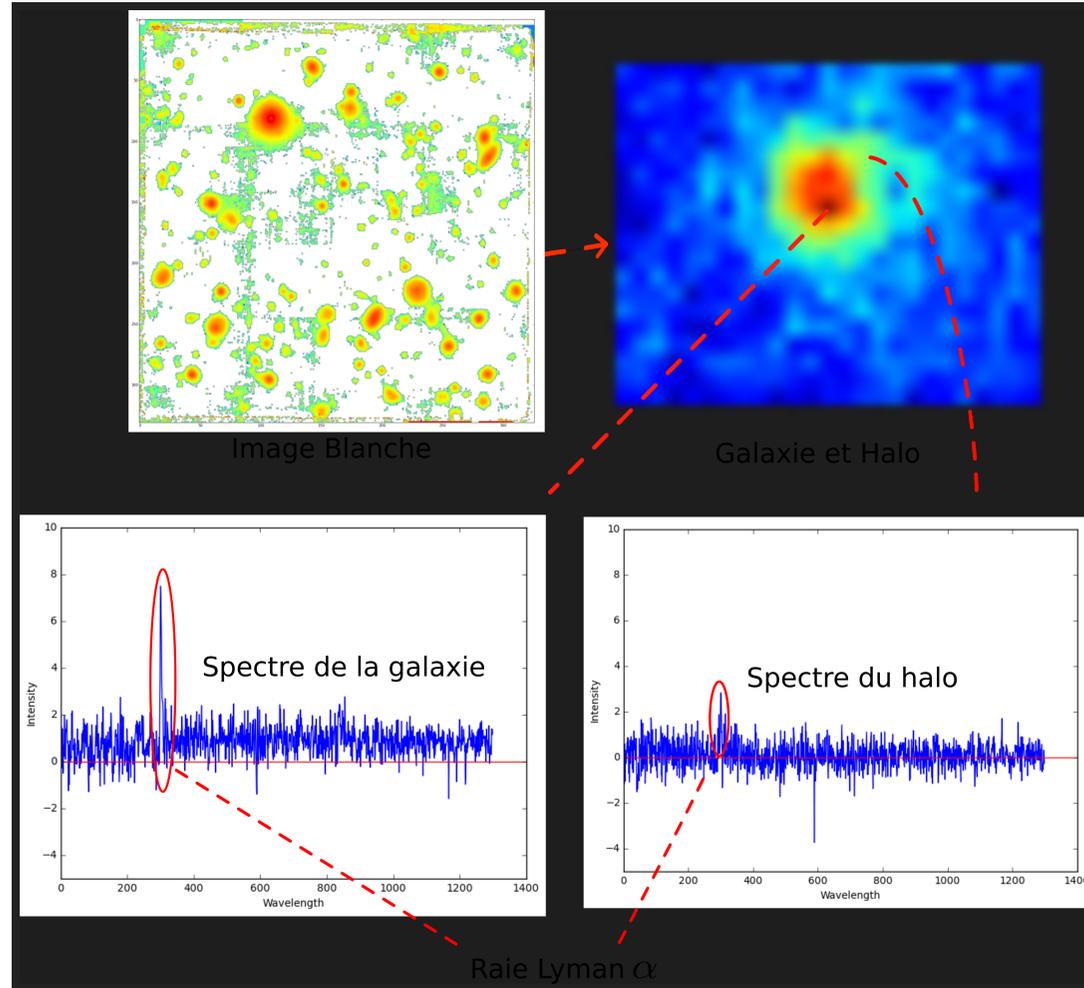
A chaque itération on choisit

$$k = \max \left\{ h; \sqrt{\frac{\sum_{i=1}^h r_{(i)}^2}{h}} \leq \hat{\sigma} \right\}$$

On peut ensuite réitérer le processus en ré-estimant  $\sigma$  à l'aide du Median Absolute Deviation et d'un LTS avec le  $k$  estimé (algorithme type point fixe).

# RÉSULTATS

# CHAMPS HDF5



# PRÉ-TRAITEMENTS

Chaîne de pré-traitements:

1. Soustraction du continu
2. Réduction par le cube de variance fourni par MUSE (1 variance par voxel)
3. Deuxième réduction avec estimation robuste de  $m$ ,  $\sigma$  en chaque longueur d'onde (cf Thèse de C. Meillier)

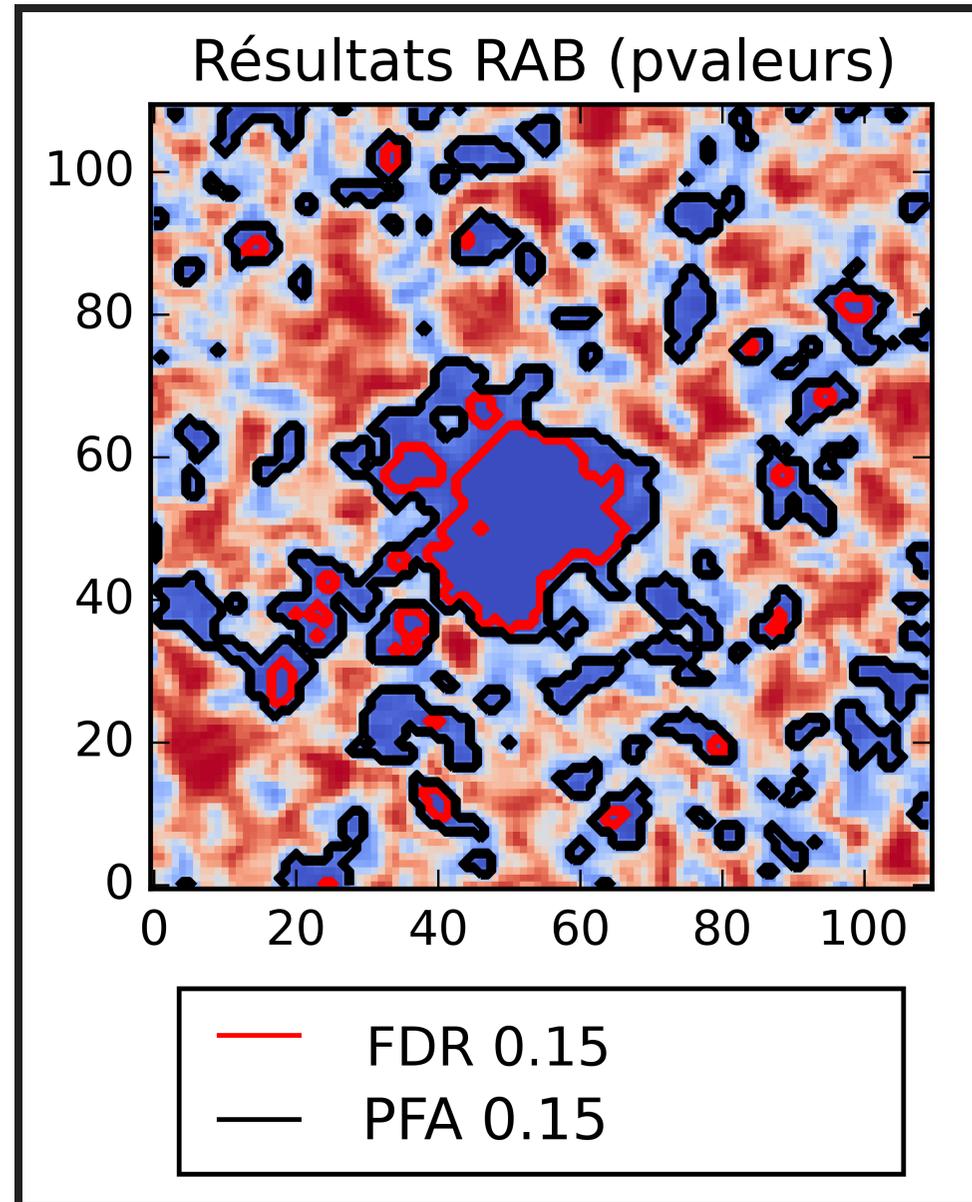
→ Bruit  $\mathcal{N}(0, 1)$  + Sources

Puis

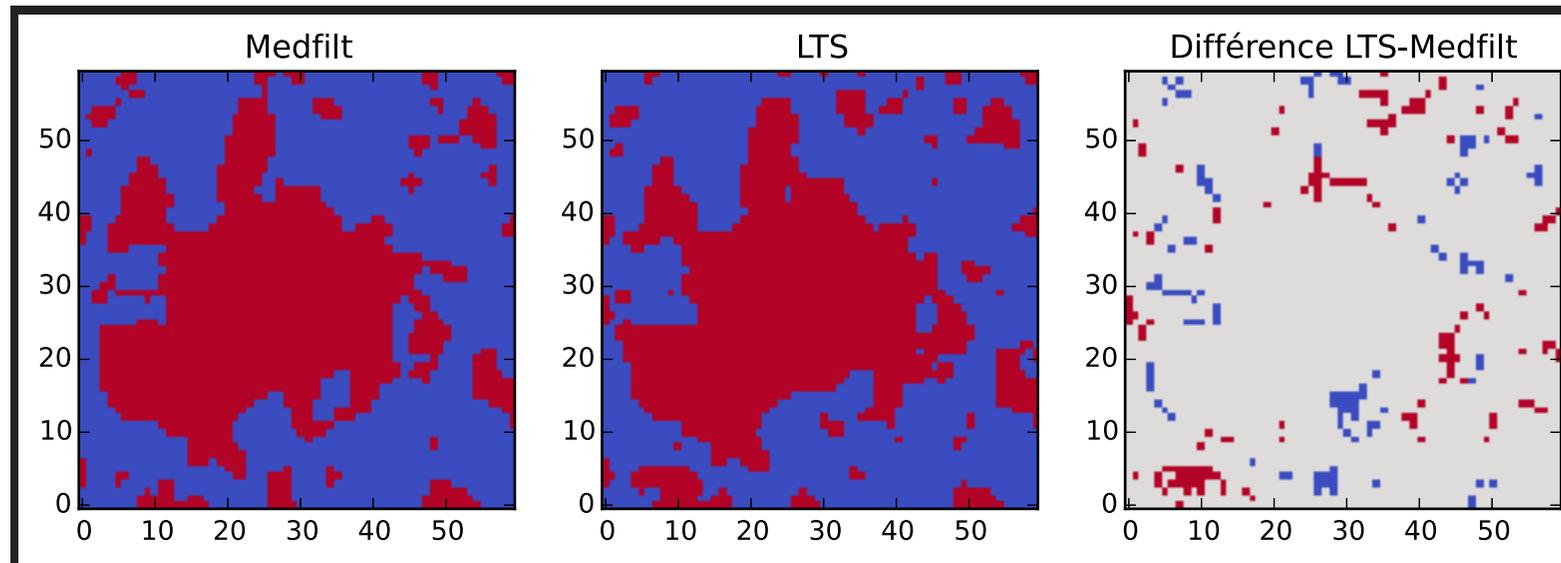
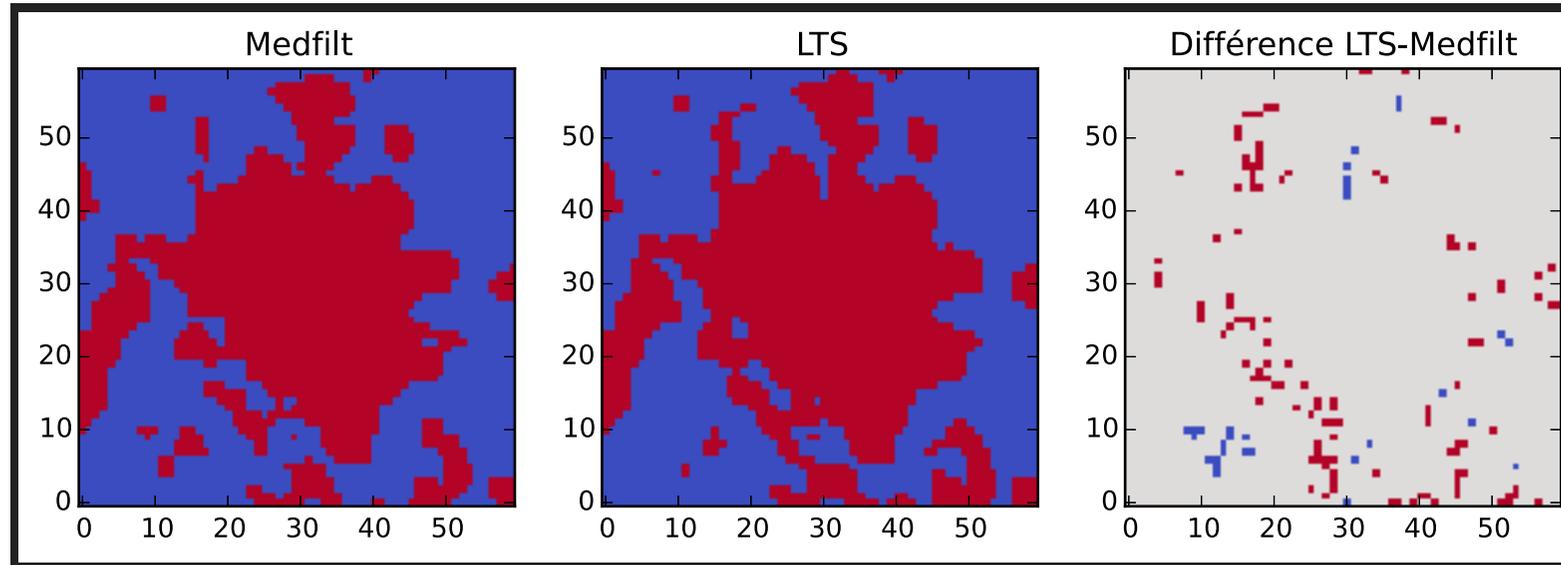
4. Application du filtre adapté à la FSF

→ Bruit + Sources magnifiées

# DÉTECTION



# SOUSTRACTION DU CONTINU



# LA SUITE

- Fusion avec les données du télescope Hubble (pansharpening/démélange)
- Construction un algorithme de poursuite des filaments
- Seul a priori : globalement linéique

# CONCLUSION

- L'hyperspectral c'est pas que "l'hyperspectral"...
- On cherche à détecter un faible signal hypothétique...
- Toute suggestion est bienvenue