



L2S - Laboratoire des signaux et des systèmes

# Rapport de stage

Dylan FAGOT

Approches convexes non-supervisées pour la reconstruction d'images  
obtenues avec l'instrument SPIRE d'HERSCHEL

Encadrants : François ORIEUX, Maître de conférence de l'Université Paris Sud  
Aurélia FRAYSSE, Maître de conférence de l'Université Paris Sud

# Sommaire

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Présentation du laboratoire . . . . .	3
1.2	Problématique du stage . . . . .	3
1.2.1	Les enjeux de la reconstruction . . . . .	3
1.2.2	Principe de la reconstruction . . . . .	3
<b>2</b>	<b>Fonctionnement des algorithmes</b>	<b>6</b>
2.1	Approximation bayésien variationnel (VBA) . . . . .	6
2.1.1	Énergie libre négative . . . . .	6
2.1.2	Séparation . . . . .	6
2.2	Code SUPREME . . . . .	7
2.2.1	Lois a priori . . . . .	7
2.2.2	Application du VBA . . . . .	8
2.3	Code SUPREME non-quadratique . . . . .	8
2.3.1	Lois a priori . . . . .	8
2.3.2	Application du VBA . . . . .	9
<b>3</b>	<b>Implémentation</b>	<b>12</b>
3.1	Initialisation . . . . .	12
3.2	Résultats . . . . .	12
3.2.1	NGC 7023 bande PSW . . . . .	13
3.2.2	NGC 7023 bande PLW . . . . .	17
3.2.3	IC 4604 bande PSW . . . . .	20

# Chapitre 1

## Introduction

### 1.1 Présentation du laboratoire

Le Laboratoire des signaux et systèmes (L2S) est une unité mixte de recherche (UMR8506) du CNRS, de CentraleSupélec et de l'université Paris-Sud. Celui-ci est divisé en trois pôles :

- **Pôle systèmes**

Recherches axées sur l'automatique ainsi que sur les systèmes électriques et mécaniques.

- **Pôle télécoms et réseaux**

Recherches axées sur les télécommunications et la sécurité.

- **Pôle signal**

Recherches axées sur la modélisation et l'analyse du signal et les problèmes inverses.

Le pôle signal est lui-même divisé en deux groupes : un spécialisé en modélisation statistique et l'autre orienté vers les problèmes inverses (groupe problème inverse ou GPI). C'est dans ce dernier que j'ai effectué mon stage.

### 1.2 Problématique du stage

#### 1.2.1 Les enjeux de la reconstruction

Le sujet de stage a vu le jour suite à la thèse soutenue par une doctorante du GPI l'année dernière. Cette thèse présente notamment un algorithme de reconstruction semi-quadratique non-supervisé qui peut être appliqué à la reconstruction d'images astronomiques.

La reconstruction d'image consiste à ici obtenir une image haute résolution à partir de plusieurs images basse résolution. Disposer d'images haute résolution est crucial, notamment pour les astrophysiciens qui étudient la matière stellaire et qui ont besoin de bien la localiser. Le L2S collabore depuis maintenant plus de dix ans avec l'institut d'astrophysique spatiale (IAS), qui utilise le code de reconstruction SUPREME pour reconstruire des images acquises par le télescope HERSCHEL. Cependant, les images reconstruites présentent des oscillations au niveau des discontinuités de l'image (phénomènes de Gibbs) ainsi que des tâches ("grumeaux") indésirables.

L'objectif du stage était donc de modifier le code SUPREME à partir d'éléments développés dans la thèse afin d'obtenir une image reconstruite plus fidèle à la réalité.

#### 1.2.2 Principe de la reconstruction

##### Le télescope HERSCHEL

HERSCHEL est un télescope spatial infrarouge développé par l'ESA. Il a été en activité de 2009 à 2013 et a notamment permit d'étudier la formation des étoiles et la matière stellaire. Il embarquait trois instruments :

- **HIFI** : Instrument hétérodyne pour l'infrarouge lointain.

- **PACS** : Spectromètre et caméra à réseau de photodétecteurs.

– **SPIRE** : Imageur photométrique et spectral.

Nous nous intéressons ici au dernier instrument qui fonctionne sur les trois bandes PSW, PMW et PLW, respectivement centrées sur 250  $\mu\text{m}$ , 350  $\mu\text{m}$  et 500  $\mu\text{m}$ .

### Acquisition des images

L'acquisition de l'image se fait en balayant la zone d'intérêt avec une matrice de bolomètres suivant les deux axes notés  $\alpha$  et  $\beta$ . On suppose connu le déplacement  $\Delta\alpha$  et  $\Delta\beta$  entre deux instants consécutifs  $t$  et  $t + 1$ .

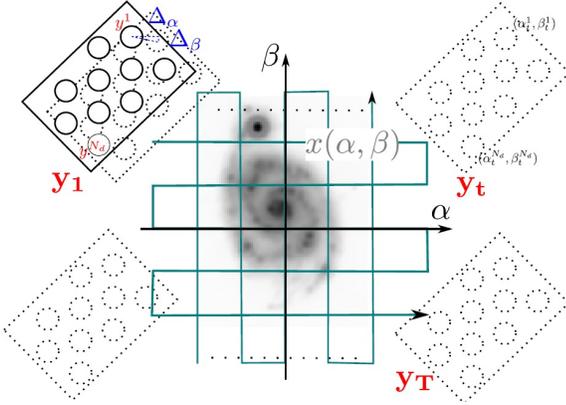


FIGURE 1.1 – Balayage suivant les axes  $\alpha$  et  $\beta$ , extrait de l'article [Ha12].

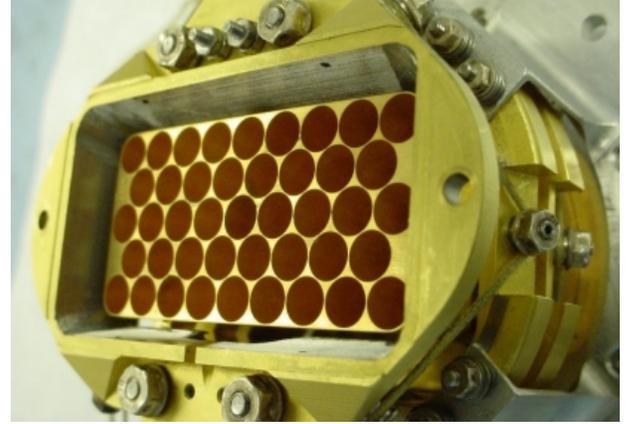


FIGURE 1.2 – Matrice de bolomètres avec les guides d'onde PLW (ESA)

Le résultats de ce balayage est une matrice de taille  $N_{\text{bolomètres}} \times N_{\text{instants}}$  contenant les différentes mesures. A titre d'exemple, on obtient une matrice  $131 \times 11584$  pour une image  $301 \times 302$  de NGC 7023 (nébuleuse de l'iris) dans la bande PSW avec une résolution de 6 secondes d'arc.

Remarque :

La taille des bolomètres est proportionnelle à la longueur d'onde. Par conséquent, le nombre de bolomètres décroît en montant en longueur d'onde, c'est pour cette raison qu'une matrice PSW contient 139 bolomètres contre seulement 43 pour une matrice PLW.

### Formulation du problème inverse

Un problème inverse se pose sous la forme

$$y = G(x) \oplus n, \quad (1.1)$$

où  $y \in \mathbf{R}^M$  représente les mesures,  $G$  un opérateur représentant le fonctionnement des capteurs,  $x \in \mathbf{R}^N$  les paramètres de cet objet et  $n \in \mathbf{R}^M$  le bruit.

En supposant que  $G$  est linéaire et que le bruit est additif, cela se réécrit

$$y = Hx + n. \quad (1.2)$$

Un tel problème est dit bien posé si il vérifie les trois conditions d'Hadamard : la solution doit exister, être unique et dépendre continûment des données. Souvent, le problème est mal posé car il fait intervenir une matrice  $H$  mal conditionnée voire non-inversible. Une façon de contourner ce problème est de rechercher une solution  $\hat{x}_{MC}$  minimisant la norme 2 du résidu du modèle : c'est la méthode des moindres carrés. Cette solution vérifie

$$\hat{x}_{MC} = \arg \min_x \|y - Hx\|^2. \quad (1.3)$$

L'expression de  $\hat{x}_{MC}$  est alors

$$\hat{x}_{MC} = (H^T H)^{-1} H^T y. \quad (1.4)$$

En pratique, le problème est souvent sous-déterminé, c'est à dire que l'on a  $M < N$ . La matrice  $H^T H$  n'est pas inversible et la solution n'est donc pas unique.

La reconstruction peut être vue comme un problème inverse : les mesures  $y$  sont les données obtenues avec les bolomètres,  $n$  est un bruit blanc et  $H$  de la forme  $H = PC$ .  $P$  est une matrice de pointage faisant le lien entre instant de mesure  $t$  et position  $(\alpha, \beta)$  dans l'espace et  $C$  représente le flou introduit par le système optique (PSF). Ce flou étant modélisé comme une convolution, la reconstruction est donc un problème de super-résolution et de déconvolution.

Ce problème inverse étant mal posé, nous allons utiliser une approche statistique pour choisir la solution la plus adaptée. Cela consiste à fixer un modèle a priori pour  $x$ .

### Coaddition

La méthode la plus simple pour obtenir une image exploitable est la coaddition ( $x_{coadd} = (P^T P)^{-1} P^T y$ ) mais elle a l'inconvénient de donner une image de faible résolution car on ne modélise pas le flou. De plus, les offsets au niveau des capteurs créent des bandes indésirables que l'on souhaiterait éliminer.

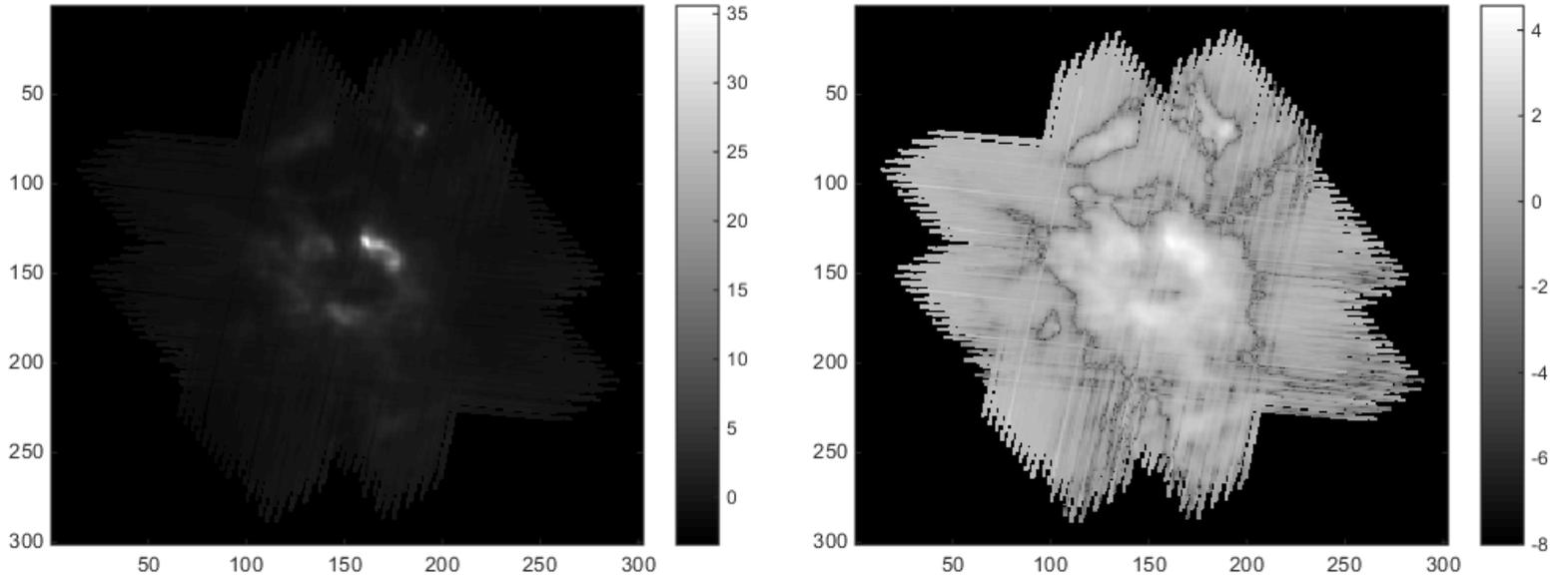


FIGURE 1.3 – Images de NGC7023 obtenue par coaddition.

Il est intéressant de visualiser l'image en échelle linéaire (à gauche) et logarithmique (à droite). La deuxième échelle permet notamment de mettre en évidence les bandes causées par les offsets ainsi que les zones où il manque des données. Les pixels situés à ces endroits seront interpolés.

Remarque : Tout comme la taille des bolomètres, la largeur de la PSF est elle aussi proportionnelle à la longueur d'onde. Les images obtenues par coaddition dans la bande PLW seront donc moins résolues que dans la bande PSW.

# Chapitre 2

## Fonctionnement des algorithmes

### 2.1 Approximation bayésien variationnel (VBA)

#### 2.1.1 Énergie libre négative

La reconstruction consiste ici à faire une modélisation probabiliste du vecteur de paramètres inconnus  $x$  (les pixels de l'image). Elle se fait à partir des mesures  $y$  et d'information a priori que l'on a sur l'objet cherché. La relation de Bayes fait le lien entre probabilité a priori et a posteriori

$$P(x | y) = \frac{P(y | x)P(x)}{P(y)}. \quad (2.1)$$

$P(x | y)$  : loi a posteriori       $P(x)$  : loi a priori  
 $P(y | x)$  : vraisemblance       $P(y)$  : évidence

Cette formule n'est en pratique pas utilisée car on ne peut calculer la constante de normalisation  $P(y)$ . Pour contourner ce problème, on recherche une loi  $Q(x)$  approchant au mieux  $P(x | y)$ . Cela se traduit par la minimisation d'une mesure de dissemblance. On définit la *KL*-divergence par

$$KL [Q \| P(\cdot | y)] = \int Q(x) \ln \frac{Q(x)}{P(x | y)} dx. \quad (2.2)$$

En utilisant la règle de Bayes, on peut la réécrire

$$KL [Q \| P(\cdot | y)] = \ln P(y) - F(Q). \quad (2.3)$$

Avec l'énergie libre négative  $F(Q)$  donnée par

$$F(Q) = \int Q(x) \ln \frac{P(x, y)}{Q(x)} dx. \quad (2.4)$$

La minimisation de la *KL*-divergence va donc se faire en maximisant  $F(Q)$ .

#### 2.1.2 Séparation

Prenons le cas d'un a priori gaussien de variance  $\rho_x^{-1}$  sur  $x$  et d'un bruit blanc de variance  $\rho_n^{-1}$ .

##### Séparabilité partielle

En pratique, on recherche la loi approchante  $Q(x)$  sous la forme d'un produit de lois simples

$$Q(x, \rho_x, \rho_n) = Q_x(x) Q_{\rho_n}(\rho_n) Q_{\rho_x}(\rho_x). \quad (2.5)$$

Il suffit ensuite de maximiser l'énergie libre négative en alternant l'optimisation de  $Q_x, Q_{\rho_x}$  et  $Q_{\rho_n}$ .

## Séparabilité totale

On suppose cette fois que la loi  $Q_x(x)$  est elle-même séparable. Elle peut donc s'écrire

$$Q(x, \rho_x, \rho_n) = \prod_{i=1}^N Q_i(x_i) Q_{\rho_n}(\rho_n) Q_{\rho_x}(\rho_x). \quad (2.6)$$

L'intérêt de cette séparation est de simplifier les équations de mise à jour, raison pour laquelle les deux codes de reconstruction utilisent cette approximation. En revanche, on perd en précision car cela revient à considérer que les variables  $x_i$  sont totalement décorréliées.

## 2.2 Code SUPREME

Ce code est l'implémentation de la méthode présentée dans l'article [Ha12]. Elle est basée sur l'utilisation du bayésien variationnel et a la particularité d'être non-supervisée : elle estime la puissance du bruit et les offsets.

### 2.2.1 Loïs a priori

#### Loi a priori pour l'objet d'intérêt

Contrairement à l'article, nous nous concentrons uniquement sur la reconstruction de la partie lisse de l'image.

Un a priori markovien est adapté pour reconstruire une image lisse. L'expression de la loi sur le modèle est donc

$$P(x | \rho_x) \propto \exp\left(-\frac{\rho_x (\|D_v x\|_2^2 + \|D_h x\|_2^2)}{2}\right), \quad (2.7)$$

où  $D_v$  et  $D_h$  sont respectivement des matrices de différences finies sur la verticale et sur l'horizontale. Le terme  $\rho_x$  (précision du modèle a priori) influe sur l'homogénéité : plus il est faible, plus l'image est homogène.

Remarque : On vient pénaliser la norme 2 des gradients pour que la différence entre deux pixels soit petite, le problème serait sinon mal posé. On dit dans ce cas que la méthode de reconstruction est quadratique.

#### Loi a priori pour les hyperparamètres

Les données sont obtenues par des bolomètres qui vont chauffer en présence de rayonnement, chauffe ensuite traduite en tension. Cependant, la température n'est pas tout à fait uniforme sur l'ensemble des bolomètres ce qui va entraîner une tension d'offset non uniforme. On va donc modéliser le bruit par un bruit blanc de variance  $\rho_n^{-1}$  et de moyenne  $o$  (offset) non nulle. Cet offset suit lui-même une loi normale  $N(m_o, \rho_o^{-1}\mathbf{I})$ . L'expression de l'a priori sur le bruit est donc

$$P(n | \rho_n, o) \propto \exp\left(-\frac{\rho_n \|n - o\|_2^2}{2}\right). \quad (2.8)$$

La relation (1.2) nous permet d'en déduire la vraisemblance

$$P(y | x, \rho_x, \rho_n, o) \propto \exp\left(-\frac{\rho_n \|y - Hx - o\|_2^2}{2}\right). \quad (2.9)$$

On choisit de faire suivre à la précision du bruit  $\rho_n$  et à la précision sur le modèle  $\rho_x$  une loi gamma qui est souvent utilisée dans les applications bayésien variationnel

$$\rho_n \sim \Gamma(\phi_n, \gamma_n) \quad \rho_x \sim \Gamma(\phi_x, \gamma_x).$$

Cette loi est intéressante car c'est une loi conjuguée dont on peut modifier la forme en jouant sur deux degrés de liberté. De plus, l'expression de sa moyenne est très simple.

## 2.2.2 Application du VBA

La loi a posteriori est obtenue en appliquant la règle de Bayes

$$P(u | y) \propto P(y | x, \rho_x, \rho_n, o) P(x | \rho_x) P(\rho_x | \gamma_x, \phi_x) P(\rho_n | \gamma_n, \phi_n) P(o | m_o, \rho_o) \propto \exp(J) \quad (2.10)$$

avec  $u = \{x, \rho_n, \rho_x, o\}$  et

$$J = -0.5 \left( \rho_n \|y - Hx - o\|_2^2 + \rho_x (\|D_v x\|_2^2 + \|D_h x\|_2^2) - N_y \ln(\rho_n) + \frac{\rho_n}{\gamma_n} - (\phi_n - 1) \ln(\rho_n) - N_x \ln(\rho_x) + \frac{\rho_x}{\gamma_x} - (\phi_x - 1) \ln(\rho_x) + \rho_o \|o - m_o\|_2^2 \right) \quad (2.11)$$

où  $N_x$  représente le nombre de pixels de l'image et  $N_y$  le nombre de données.

La solution au problème de maximisation de l'énergie libre négative s'écrit

$$Q(u_i) \propto \exp \left( \langle J \rangle_{\prod_{j \neq i} Q(u_j)} \right). \quad (2.12)$$

Le problème étant de grande taille, on utilise une variante du bayésien variationnel : le bayésien variationnel de type gradient (voir [Fr11]). La solution s'écrit ici

$$\check{Q}^k(u_i) \propto (\check{Q}^{k-1}(u_i))^{1-\lambda_i} \exp \left( \lambda_i \langle J \rangle_{\prod_{j \neq i} \check{Q}^{k-1}(u_j)} \right). \quad (2.13)$$

L'intérêt de cette méthode est de réduire le temps de calcul, notamment pour la mise à jour de  $x$  qui contient plusieurs milliers voire millions de pixels. En revanche, l'introduction du pas  $\lambda$  complexifie l'expression des équations de mise à jour. Nous décidons donc de fixer  $\lambda_\varrho = 1$  pour l'ensemble  $\varrho = \{\rho_n, \rho_x, o\}$ .

## 2.3 Code SUPREME non-quadratique

Le code ressemble à SUPREME dans le sens où il utilise l'approximation bayésien variationnel. Sa particularité est de faire intervenir des variables cachées  $b = [b_v, b_h] \in \mathbf{R}^{2N}$  qui le rend non-quadratique. Concrètement, celles-ci vont s'activer aux niveau des forts gradients de luminosité dans l'image de façon à moins les pénaliser que dans le cas quadratique de façon à préserver les contours tout en réduisant le bruit.

### 2.3.1 Lois a priori

#### Loi a priori pour l'objet d'intérêt

L'a priori sur le modèle est semblable à celui utilisé dans le code précédent. On s'autorise cependant quelques gradients importants dans l'image. Les forts gradients sur l'horizontale et sur la verticale sont respectivement stockés dans  $b_h$  et  $b_v$ . L'expression de la loi a priori sur le modèle est donc

$$P(x | b, \gamma_d) = Z_{x|b}^{-1} \exp \left( -\frac{\gamma_d (\|D_v x - b_v\|_2^2 + \|D_h x - b_h\|_2^2)}{2} \right). \quad (2.14)$$

On admet que la fonction de partition  $Z_{x|b}^{-1}$  peut être approchée par  $\gamma_d^{N_x/2}$ .

La loi introduite pour les variables cachées  $b$  est une loi de Laplace qui permet de pénaliser non plus la norme 2 mais la norme 1 des gradients supérieurs à un certain seuil. L'expression de cette loi est donc

$$P(b | \gamma_b) \propto \gamma_b^\xi \exp \left( -\frac{\gamma_b \|b\|_1}{2} \right). \quad (2.15)$$

Remarque : Le terme  $\xi \in [0, 2N_x]$  est introduit pour remplacer  $2N_x$ . On privilégie des gradients parcimonieux donc la grande majorité des éléments de  $b$  seront à zéro.

A partir de ces deux lois, on peut en déduire la distribution a priori de  $x$

$$P(x | \gamma_d, \gamma_b) = \int P(x | b, \gamma_d) P(b | \gamma_b) db \\ \propto \int \gamma_d^{N_x/2} \exp\left(-\frac{\gamma_d(\|D_v x - b_v\|_2^2 + \|D_h x - b_h\|_2^2)}{2}\right) \gamma_b^\xi \exp\left(\frac{\gamma_b \|b\|_1}{2}\right) db. \quad (2.16)$$

On remarque que la loi de  $x$  dépend de deux paramètres  $\gamma_b$  et  $\gamma_d$  que l'on devrait estimer. On préfère introduire un paramètre de forme  $\nu = \frac{\gamma_b}{\sqrt{\gamma_d}}$  que l'on fixe et un paramètre d'échelle  $\kappa_d = \sqrt{\gamma_d}$  que l'on estime. La loi jointe se réécrit alors

$$P(x, b | \gamma_d, \nu) \propto \kappa_d^{N_x} \exp\left(-\frac{\kappa_d^2(\|D_v x - b_v\|_2^2 + \|D_h x - b_h\|_2^2)}{2}\right) (\nu \kappa_d)^\xi \exp\left(\frac{\nu \kappa_d \|b\|_1}{2}\right). \quad (2.17)$$

### Loi a priori pour les hyperparamètres

Les lois a priori des offsets  $o$  et de la précision du bruit  $\rho_n$  ont été conservées dans le code. En revanche, la loi a priori du terme  $\kappa_d = \sqrt{\gamma_d}$  ne peut être une loi gamma à cause de la racine carrée. En conséquence, on introduit une autre densité de probabilité  $\psi$  paramétrée par trois termes  $(\alpha, \beta, \mu)$

$$\psi(z | \alpha, \beta, \mu) = Z_{\alpha, \beta, \mu}^{-1} z^\alpha \exp(-\beta(z + \mu)^2). \quad (2.18)$$

On utilise cette distribution comme loi a priori de  $\kappa_d$  en fixant  $(\tilde{\alpha}, \tilde{\beta}, \tilde{\mu}) = (-1, 0, 0)$  pour que l'a priori soit non informatif.

La fonction de partition  $Z_{\alpha, \beta, \mu}$  se calcule en effectuant le changement de variable  $z' = z + \mu$  et en utilisant la formule du binôme de Newton

$$Z_{\alpha, \beta, \mu} = \int z^\alpha \exp(-\beta(z + \mu)^2) dz = \sum_{i=0}^{\alpha} \binom{\alpha}{i} (-\mu)^i \frac{1}{2\beta^{\frac{\alpha-i+1}{2}}} \Gamma\left(\frac{\alpha-i+1}{2}, \beta\mu^2\right) \quad (2.19)$$

où  $\Gamma$  représente la fonction gamma incomplète

$$\Gamma(r, u) = \int_u^\infty t^{r-1} e^{-t} dt. \quad (2.20)$$

Remarque :

L'expression du moment d'ordre  $n$  se déduit aisément

$$\langle z^n \rangle = \int z^n \left( Z_{\alpha, \beta, \mu}^{-1} z^\alpha \exp(-\beta(z + \mu)^2) \right) dz = Z_{\alpha, \beta, \mu}^{-1} \int z^{n+\alpha} \exp(-\beta(z + \mu)^2) dz = \frac{Z_{\alpha+n, \beta, \mu}}{Z_{\alpha, \beta, \mu}}. \quad (2.21)$$

Cependant, le calcul direct de ces moments est numériquement impossible compte tenu de la valeur de  $\alpha$ . Nous préférons donc les estimer avec un algorithme MCMC présenté un peu plus loin.

### 2.3.2 Application du VBA

La loi a posteriori est obtenue en appliquant la règle de Bayes

$$P(u | y, \nu) \propto P(y | x, \rho_n, o) P(x | b, \kappa_d) P(b | \kappa_d, \nu) P(\rho_n | \gamma_n, \phi_n) P(o | m_o, \rho_o) P(\kappa_d), \quad (2.22)$$

avec  $u = \{x, \rho_n, \rho_x, o\}$ .

Là encore, on exploite l'hypothèse de séparabilité totale de la loi  $Q(u)$ . La maximisation de  $F(Q)$  se fait donc en jouant alternativement sur  $Q_x, Q_b, Q_{\rho_n}, Q_{\kappa_d}$  et  $Q_o$ .

Cependant, le choix d'une loi de Laplace pour  $P(b | \gamma_b)$  non conjuguée avec  $P(x | b, \gamma_b)$  nous oblige à se placer dans l'approximation bayésien variationnelle restreinte. On fait le choix d'approcher la loi de Laplace par une distribution de Dirac

$$Q_b^k = \delta(b - \hat{b}^k). \quad (2.23)$$

La maximisation de l'énergie libre négative  $F(Q)$  se fera normalement au détail près que l'on considérera  $Q_b$  comme un Dirac au moment de l'optimiser.

### Optimisation de $Q_x$

Ici, nous utilisons une autre variante du bayésien variationnel : le bayésien variationnel à mémoire de gradient.

Il faut dans un premier temps calculer les paramètres d'une loi intermédiaire  $Q_R$

$$\sigma_R^2 = [\langle \rho_n \rangle^k \text{diag}(H^T H) + \langle \kappa_d^2 \rangle^k \text{diag}(D^T D)]^{-1} \quad (2.24)$$

$$m_R^2 = \sigma_R^2 [\langle \rho_n \rangle^k (H^T (y - Hm_k - o) + \text{diag}(H^T H) \circ m_k) + \langle \kappa_d^2 \rangle^k (D_h^T b_h + D_v^T b_v - D^T D m_k + \text{diag}(D^T D) \circ m_k)] \quad (2.25)$$

avec  $D^T D = D_h^T D_h + D_v^T D_v$  et  $\circ$  produit matriciel d'Hadamard.

Ces paramètres vont ensuite servir à calculer ceux de  $Q_x$

$$\sigma_k^2 = \left[ \frac{1}{\sigma_k^2} + s_1^k \left( \frac{1}{\sigma_R^2} - \frac{1}{\sigma_k^2} \right) + s_2^k \left( \frac{1}{\sigma_k^2} - \frac{1}{\sigma_{k-1}^2} \right) \right]^{-1} \quad (2.26)$$

$$m_k = \sigma_k^2 \left[ \frac{m_k}{\sigma_k^2} + s_1^k \left( \frac{m_R}{\sigma_R^2} - \frac{m_k}{\sigma_k^2} \right) + s_2^k \left( \frac{m_k}{\sigma_k^2} - \frac{m_{k-1}}{\sigma_{k-1}^2} \right) \right] \quad (2.27)$$

où

$$s^k = (s_1^k, s_2^k)^T$$

représente le pas sous-optimal.

Ce pas sous-optimal vérifie

$$s^k = \begin{cases} s_{opt}^k & \text{si } g(s^k) > g(0) \\ 2^{-t} s_{opt}^k & \text{si } g(s^k) < g(0) \text{ et } g(2^{-t} s_{opt}^k) > g(0) \end{cases} \quad (2.28)$$

$s_{opt}^k$  est le pas optimal qui est choisit de façon à augmenter le plus possible notre critère  $g = F[Q(u)]$  à chaque itération. Son expression est

$$s_{opt}^k = - \left( \frac{\partial^2 g}{\partial s \partial s^T} \Big|_{s=0} \right)^{-1} \frac{\partial g}{\partial s} \Big|_{s=0}. \quad (2.29)$$

Remarques :

1. Nous avons directement utilisé le pas optimal qui ne nécessite pas de calculer l'énergie libre.
2. Dans le cas où la hessienne de  $g$  s'avère non-inversible à l'itération  $k$ , on prend

$$s_{opt}^k = \left( - \frac{\partial^2 g}{\partial s_1^2} \Big|_{s=0}^{-1} \frac{\partial g}{\partial s_1} \Big|_{s=0}, 0 \right)^T. \quad (2.30)$$

L'utilisation d'un seul pas réduira simplement la vitesse de convergence.

### Optimisation de $Q_b$

Sous l'hypothèse bayésien variationnel restreint présentée précédemment, les équations de mise à jour des variables cachées sont

$$\hat{b}_h^{k+1} = \text{signe}(D_h m_{k+1}) \max \left( |D_h m_{k+1}| - \frac{\nu \langle \kappa_d \rangle^k}{2 \langle \kappa_d^2 \rangle^k}, 0 \right), \quad (2.31)$$

$$\hat{b}_v^{k+1} = \text{signe}(D_v m_{k+1}) \max \left( |D_v m_{k+1}| - \frac{\nu \langle \kappa_d \rangle^k}{2 \langle \kappa_d^2 \rangle^k}, 0 \right). \quad (2.32)$$

Remarque :

Les variables cachées s'activent lorsque des gradients dépassent le seuil  $s = \frac{\nu \langle \kappa_d \rangle^k}{2 \langle \kappa_d^2 \rangle^k}$ . Si ce seuil est trop grand, on empêche les variables cachées de s'exprimer et on se ramène dans le cas quadratique.

## Optimisation de $Q_{\kappa_d}$

Les équations de mise à jour des paramètres sont

$$\begin{cases} \alpha^{k+1} = N_x + \lfloor \xi \rfloor + \tilde{\alpha} \\ \beta^{k+1} = 0.5 \left( \|D_h m_{k+1} - \hat{b}_h^{k+1}\|_2^2 + \|D_v m_{k+1} - \hat{b}_v^{k+1}\|_2^2 + \sum_i \text{diag}(D^T D)_i (\sigma_k^2)_i \right) + \tilde{\beta} \\ \mu^{k+1} = \frac{\nu \|\hat{b}^{k+1}\|_1 + 4\tilde{\beta}\tilde{\mu}}{4\beta^{k+1}} \end{cases} \quad (2.33)$$

Après avoir mit à jour  $Q_{\kappa_d}$ , il faut calculer ses moments d'ordre 1 et 2 pour mettre à jour  $Q_x$  et  $Q_b$ . Ces moments peuvent se calculer sur des échantillons de  $\kappa_d$  générés par un algorithme d'échantillonnage de type Metropolis-Hastings. Cet algorithme se résume à

1. Initialiser par  $\kappa_d^{(0)} \sim U_{[0, \kappa_d^{(max)}]}$
2. Générer  $\epsilon^{(t)} \sim U_{[-\tau/2, \tau/2]}$  puis calculer  $\kappa_d^{(*)} = \kappa_d^{(t)} + \epsilon^{(t)}$
3. Prendre

$$\kappa_d^{(t+1)} = \begin{cases} \kappa_d^{(*)} & \text{avec la probabilité } p = \min \left( 1, \frac{\psi(\kappa_d^{(*)} | \alpha^{k+1}, \beta^{k+1}, \mu^{k+1})}{\psi(\kappa_d^{(t)} | \alpha^{k+1}, \beta^{k+1}, \mu^{k+1})} \right) \\ \kappa_d^{(t)} & \text{avec la probabilité } 1 - p \end{cases}.$$

On répète ensuite les étapes 2 et 3 jusqu'à la convergence.

$\kappa_d^{(max)}$  est le mode de la loi  $Q_{\kappa_d}$  qui se calcule en annulant la dérivée de sa densité. Son expression est

$$\kappa_d^{(max)} = -\frac{\mu}{2} + \sqrt{\frac{\mu^2}{4} + \frac{\alpha}{2\beta}}. \quad (2.34)$$

$\tau$  représente la largeur de l'intervalle dans lequel on prend  $\epsilon$ . D'après les la thèse, cette largeur peut être fixée à  $\kappa_d^{(max)}/50$ .

Remarque : L'intérêt de cet algorithme est de générer des échantillons d'une loi que l'on connaît à sa fonction de partition près ( $Z_{\alpha, \beta, \mu}^{-1}$  dans (2.18)). Celle-ci n'apparaît pas car on calcule la probabilité de transition  $p$  à partir du rapport entre deux valeurs de  $\psi$ .

# Chapitre 3

## Implémentation

### 3.1 Initialisation

La plupart des variables et paramètres du code développé sont initialisés aux mêmes valeurs que celles du code SUPREME original. Il reste toutefois à initialiser les paramètres introduits dans le cas non-quadratique. La thèse suggère de fixer

$$\begin{cases} \nu = 0.8 \\ \xi = 0.1N_x \\ \kappa_d^{(0)} = \sqrt{\rho_n^{(0)}} \end{cases} . \quad (3.1)$$

Les variables cachées  $b_h$  et  $b_v$  sont respectivement initialisées avec les  $\frac{\xi}{2N_x}$  plus grandes valeurs de  $D_h x_{coadd}$  et  $D_v x_{coadd}$ . On se retrouve ainsi avec des variables cachées privilégiant des gradients parcinomieux.

Dans notre cas, il ne faut pas choisir  $\xi$  en fonction de la taille de l'image  $N_x$  mais en fonction de la taille de la zone balayée  $N_{covered}$ . En effet, la partie extérieure de l'image étant mise à zéro, il ne peut y avoir de gradients et donc pas de variables cachées. Par conséquent, on fixe  $\xi = 0.1N_{covered}$ . Dans le cas où  $N_{covered} = 0.5N_x$ , cela revient à fixer près de 98 % des valeurs de  $b$  à zéro.

### 3.2 Résultats

Nous nous sommes intéressés à la reconstruction de deux objets : la nébuleuse de l'iris (NGC 7023) dans les bandes PSW et PLW et la nébuleuse rho ophiuchi (IC 4604) dans la bande PSW.

Les images résultantes de la reconstruction sont affichées avec les images initiales (coaddition) de façon à bien voir le gain en résolution, l'influence de l'estimation des offsets ainsi que l'interpolation des pixels manquants.

### 3.2.1 NGC 7023 bande PSW SUPREME

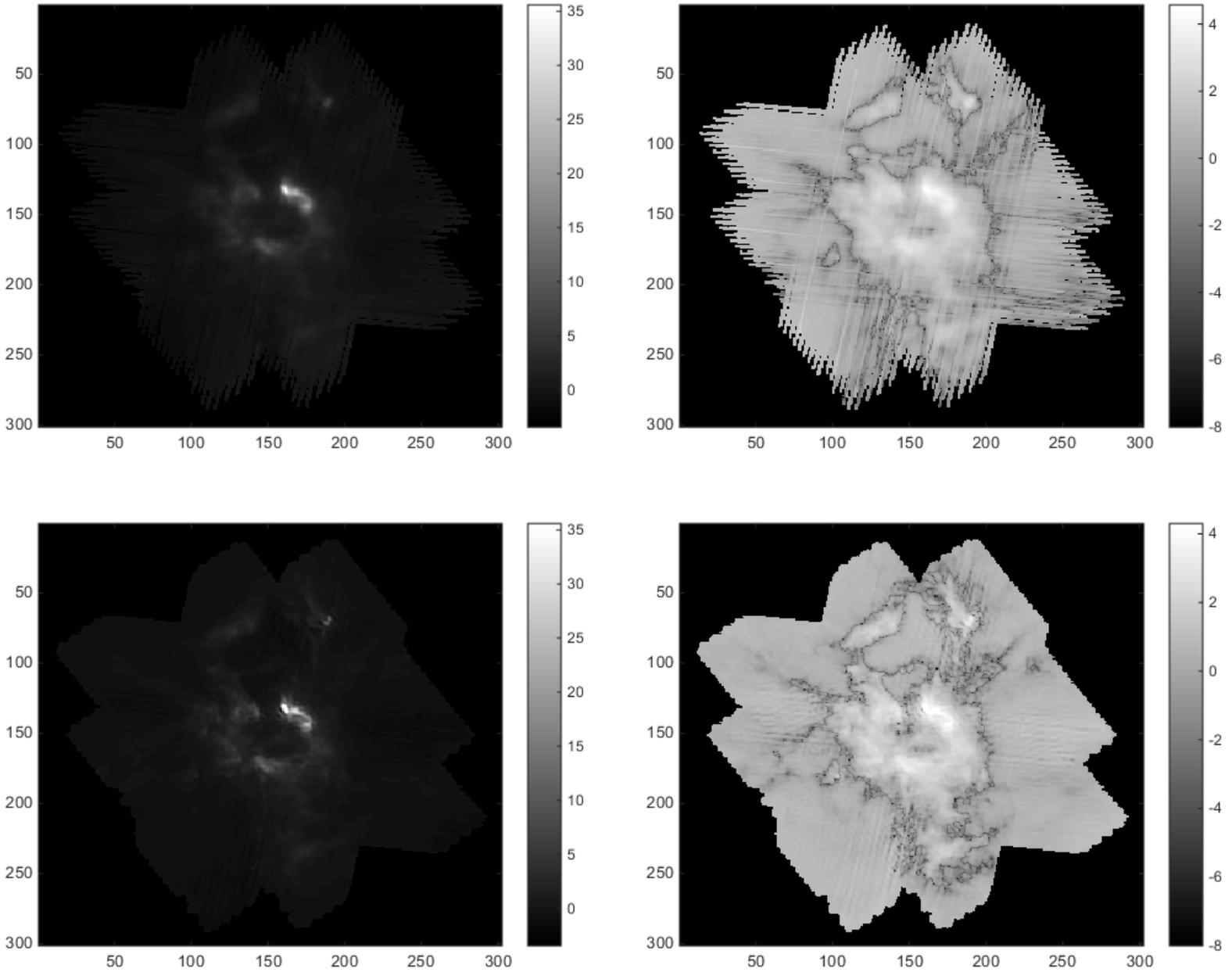


FIGURE 3.1 – Ligne du haut : images obtenues par coaddition. Ligne du bas : résultat de la reconstruction (code SUPREME)

On remarque un net gain en résolution par rapport à l'image obtenue par coaddition. De plus, l'estimation des offsets permet bien de limiter l'apparition de bandes sur l'image. Les pixels situés au niveau de données manquantes ont été interpolés.

### SUPREME non-quadratique

L'estimation des moments de  $\kappa_d$  étant peu efficace, nous avons choisis de ne pas les estimer. Le choix de la valeur du paramètre est important car il permet de fixer le seuil sur les gradients. Il se détermine de façon empirique et il s'avère qu'une valeur de  $\kappa_d^2 = 6.445$  donne des résultats intéressants pour NGC7023.

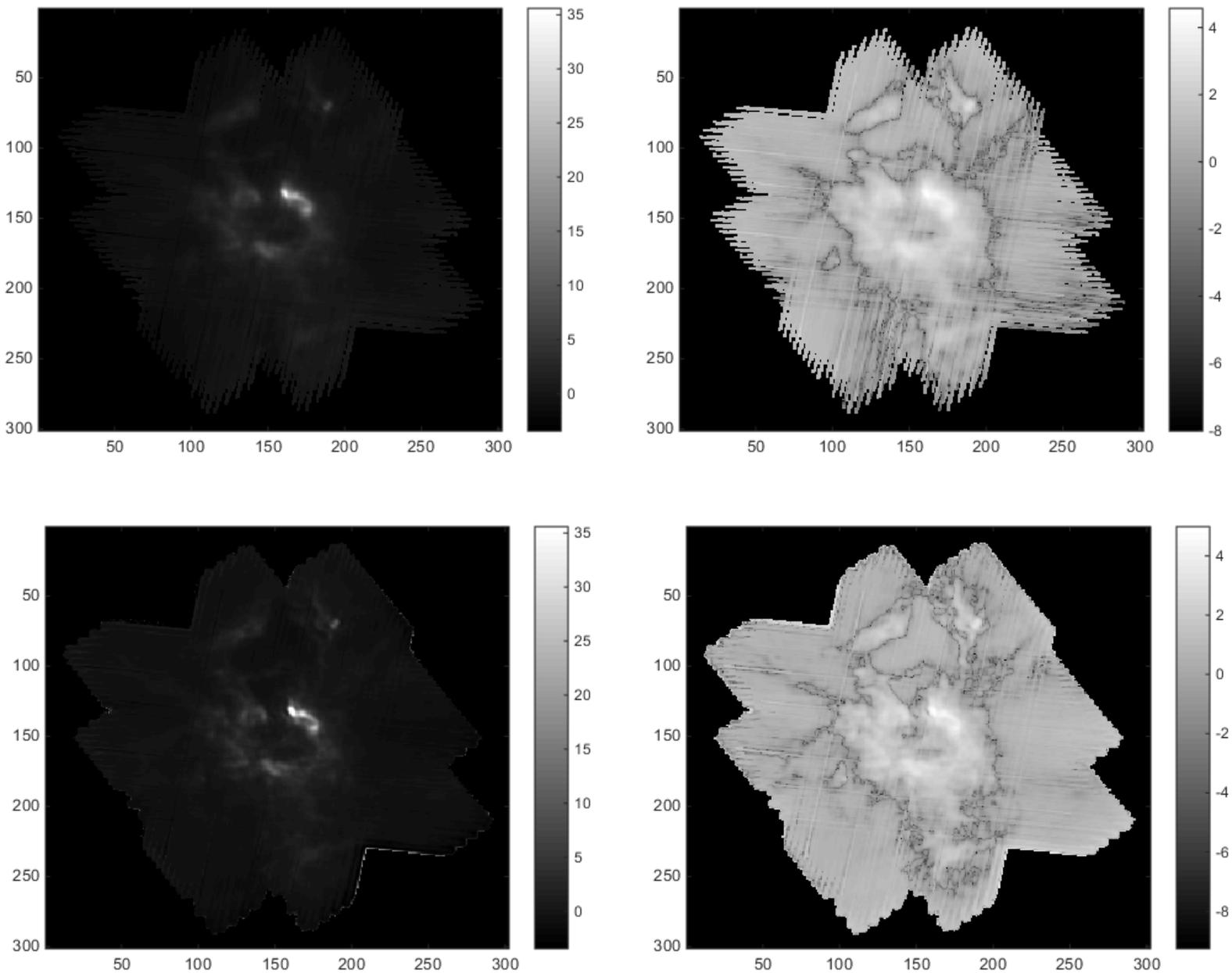


FIGURE 3.2 – Ligne du haut : images obtenues par coaddition. Ligne du bas : résultat de la reconstruction (code SUPREME non-quadratique)

Nous constatons ici encore un net gain en résolution. Cependant, on remarque que les bords ne sont pas toujours bien reconstruits.

## Tracés des hyperparamètres

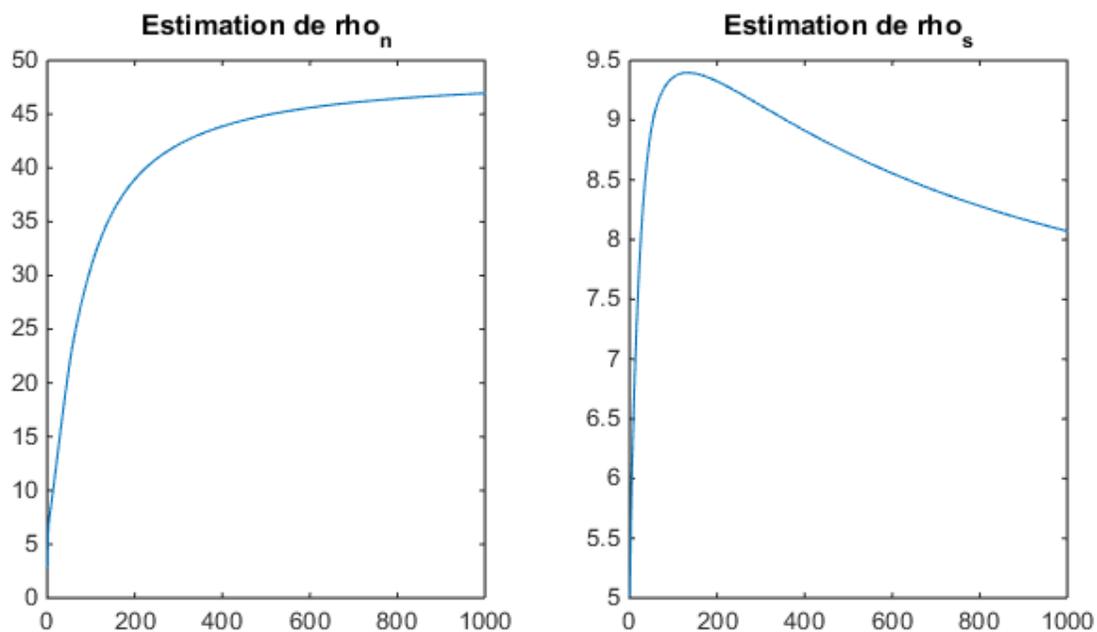


FIGURE 3.3 – Tracés des hyperparamètres (SUPREME). Précision sur le bruit (gauche) et précision sur le modèle a priori (droite)

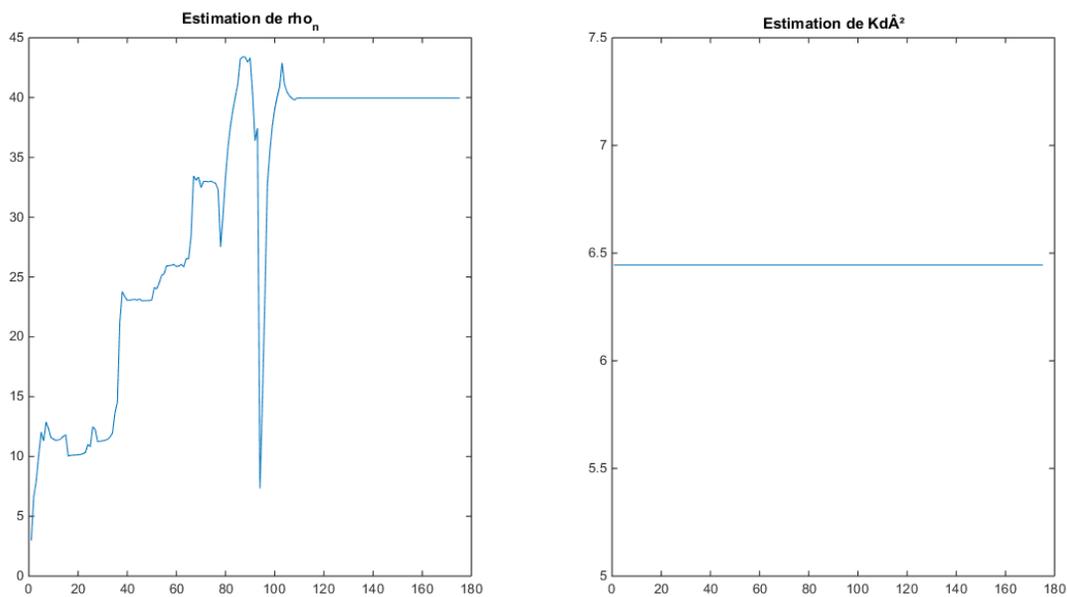


FIGURE 3.4 – Tracés des hyperparamètres (SUPREME). Précision sur le bruit (gauche) et précision sur le modèle a priori (droite)

On voit ici que l’algorithme bayésien variationnel à mémoire de gradient se montre bien plus rapide, notamment par l’utilisation de deux pas.

### Vues en coupes

Pour mieux voir les différences entre le résultat des deux codes, nous avons effectué deux coupes : une suivant la colonne 161 et une autre suivant la ligne 133. Ces coupes présentent plusieurs discontinuités, endroits où on peut s'attendre à voir des oscillations.

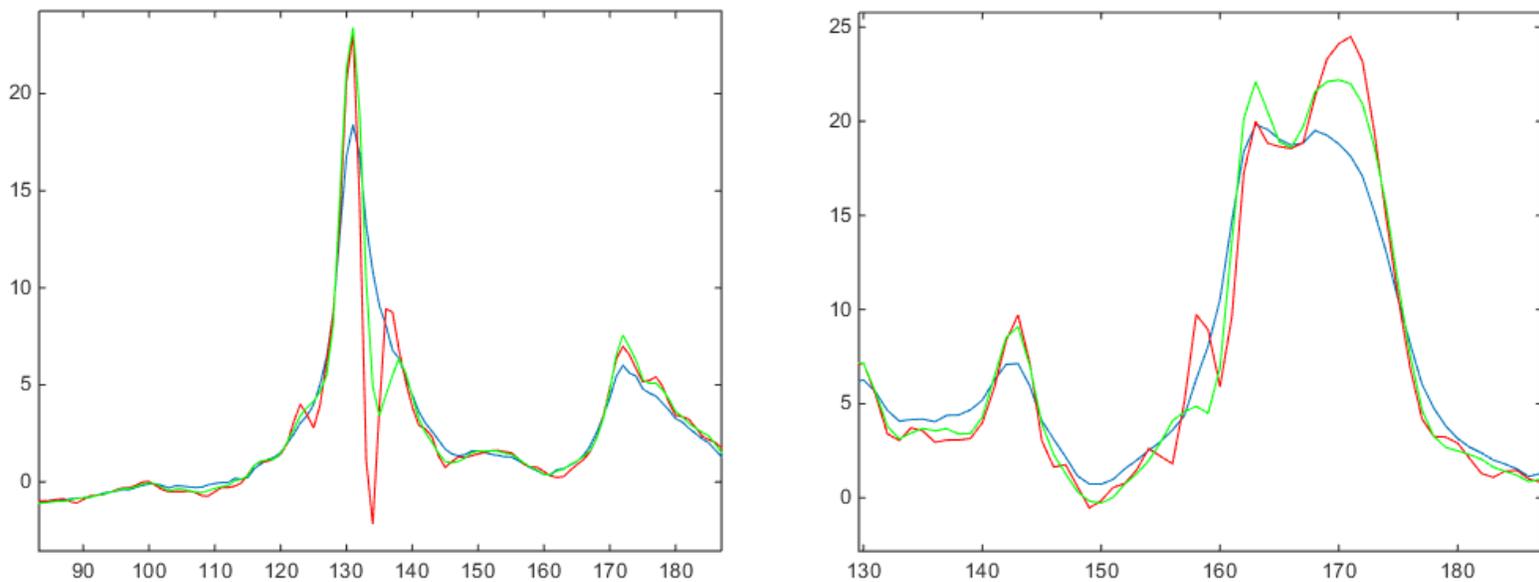


FIGURE 3.5 – Vues en coupe de la colonne 161 (gauche) et de la ligne 133 (droite). Image obtenue par coaddition (bleu), résultats de SUPREME (rouge) et SUPREME non-quadratique (vert).

On remarque que la reconstruction non-quadratique permet de diminuer significativement les rebonds (phénomènes de Gibbs) au niveau des discontinuités. Sur la première coupe, on remarque de forts rebonds dans le cas quadratique, surtout au niveau des pixels des lignes 123, 134 et 138. Sur la seconde coupe, on retrouve des rebonds au niveau des pixels des colonnes 155, 160 et 180.

### 3.2.2 NGC 7023 bande PLW SUPREME

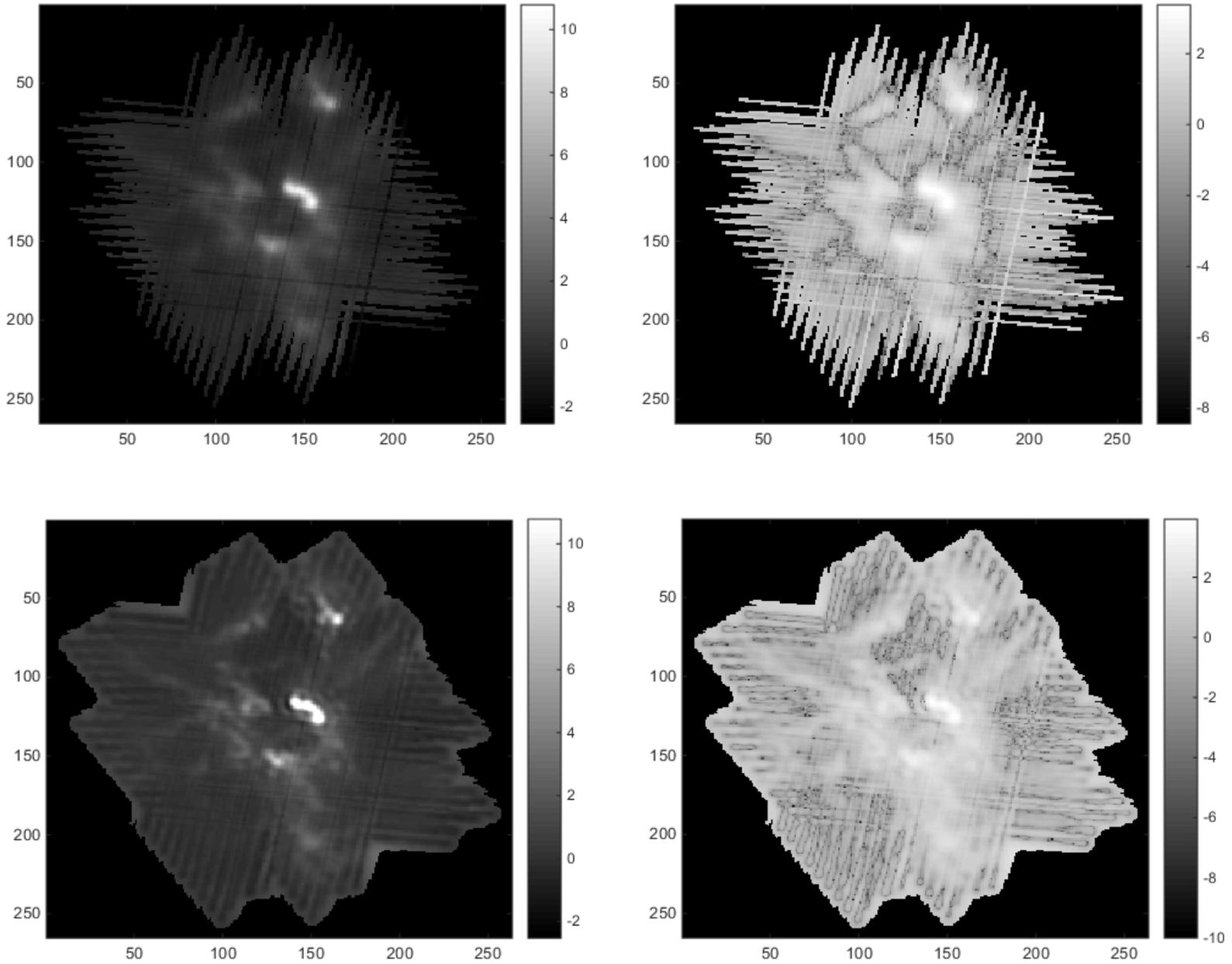


FIGURE 3.6 – Ligne du haut : images obtenues par coaddition. Ligne du bas : résultat de la reconstruction (code SUPREME)

Comme attendu, l'image dans la bande PLW est moins résolue que dans la bande PSW.

### SUPREME non-quadratique

Nous avons ici encore fixé la valeur de  $\kappa_d^2$ . Cette fois, sa valeur est de 13.

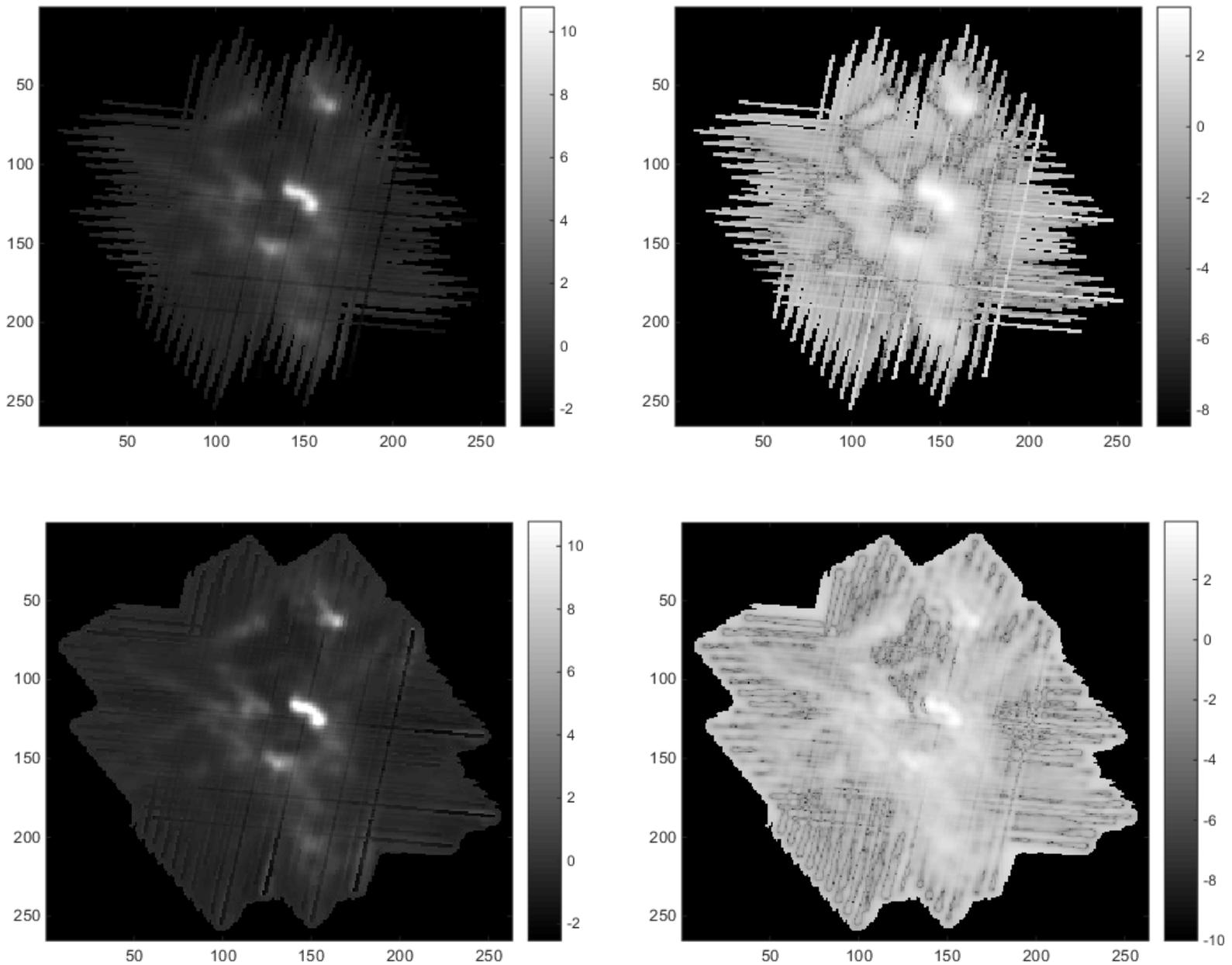


FIGURE 3.7 – Ligne du haut : images obtenues par coaddition. Ligne du bas : résultat de la reconstruction (code SUPREME non-quadratique)

Les bords sont ici correctement reconstruits dans le cas non-quadratique. Cependant, on remarque que les bandes sont toujours présentes sur l'image, ce qui laisse penser que les offsets sont mal estimés.

### Vues en coupes

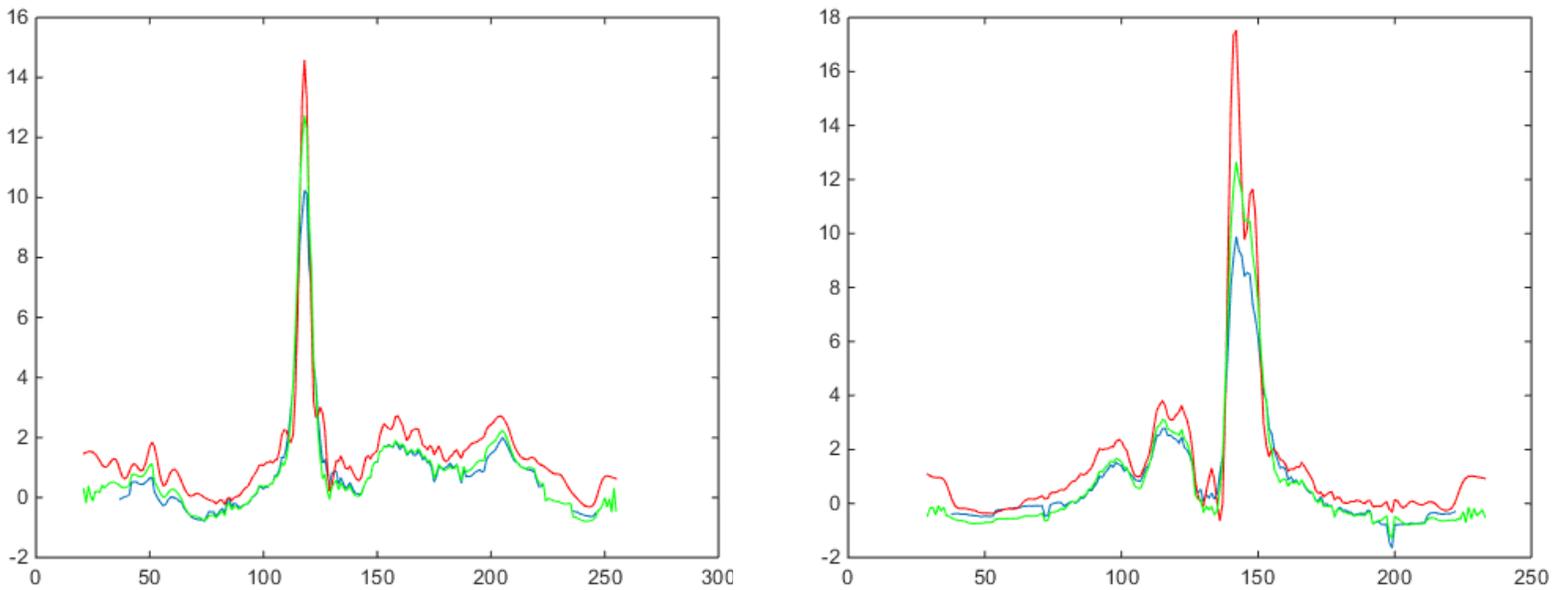


FIGURE 3.8 – Vues en coupe de la colonne 161 (gauche) et de la ligne 133 (droite). Image obtenue par coaddition (bleu), résultats de SUPREME (rouge) et SUPREME non-quadratique (vert).

Là encore, on observe encore une diminution des oscillations à proximités des forts gradients, ce qui était attendu. On remarque par ailleurs que les bords sont un peu mieux reconstruits avec des valeurs plus proches de zéro.

### 3.2.3 IC 4604 bande PSW SUPREME

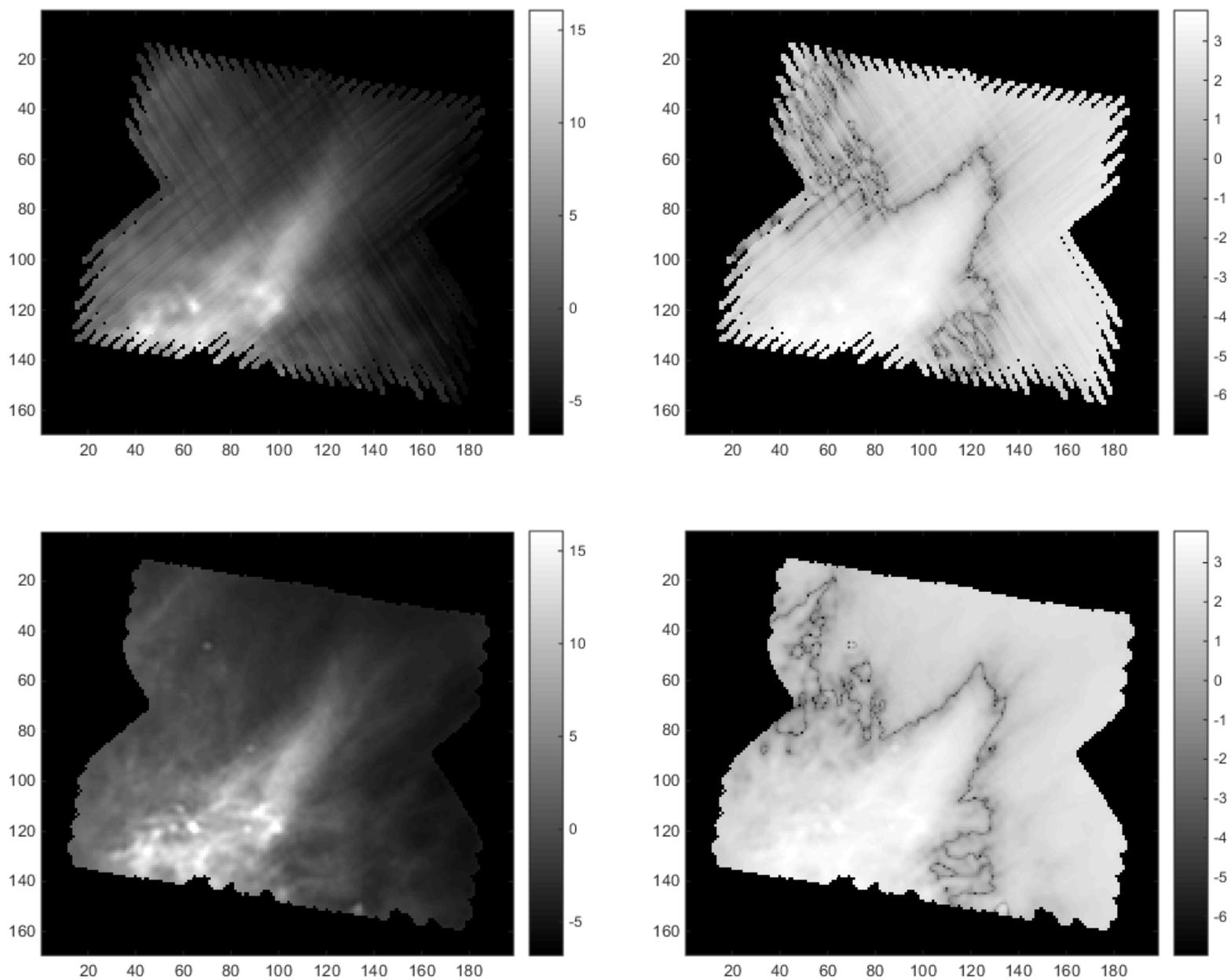


FIGURE 3.9 – Ligne du haut : images obtenues par coaddition. Ligne du bas : résultat de la reconstruction (code SUPREME)

### SUPREME non-quadratique

Nous avons ici encore fixé la valeur de  $\kappa_d^2$ . Cette fois, sa valeur est de 1,82.

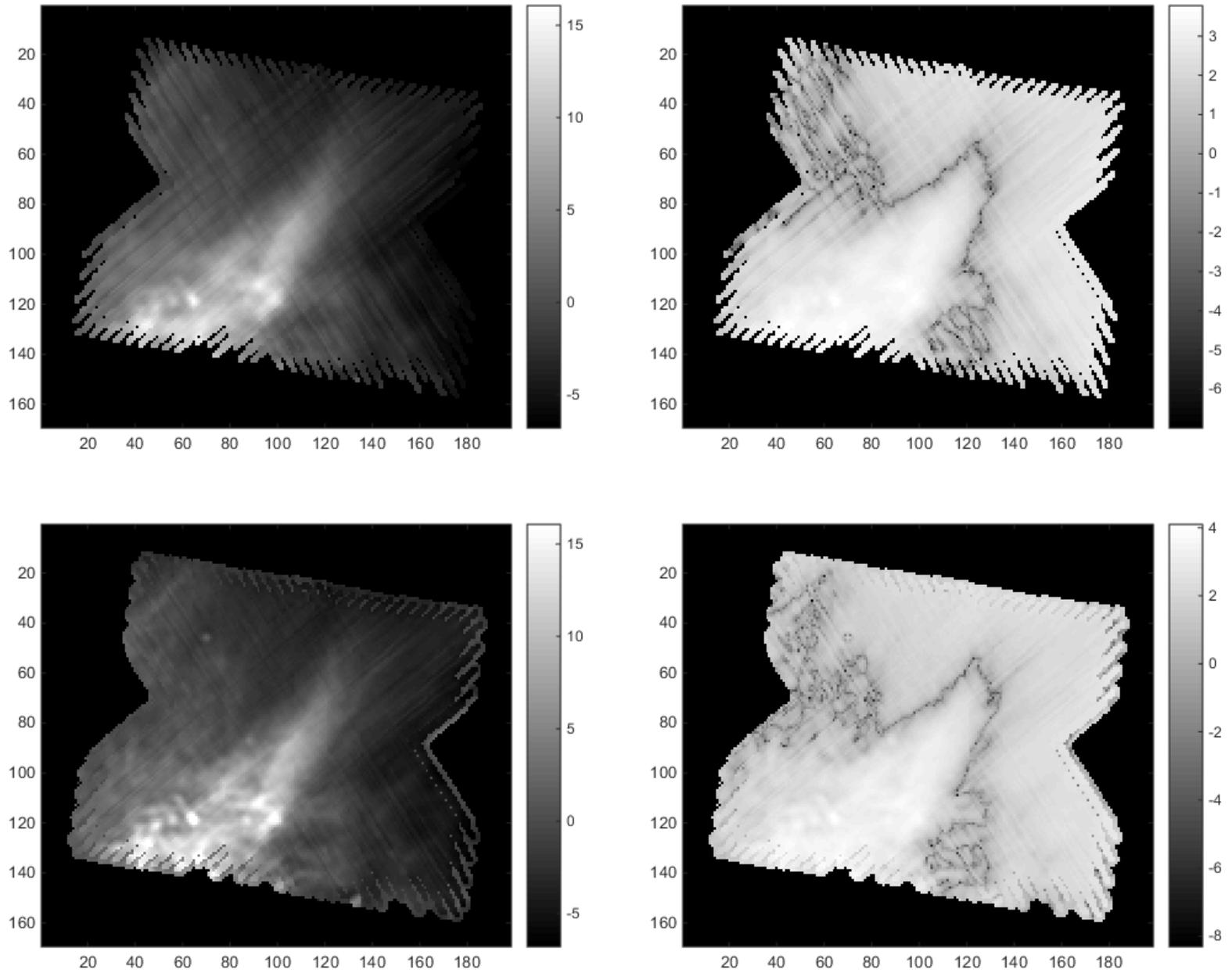


FIGURE 3.10 – Ligne du haut : images obtenues par coaddition. Ligne du bas : résultat de la reconstruction (code SUPREME non-quadratique)

On remarque de nouveau que les bords sont mal reconstruits dans le cas non-quadratique. Le problème des bandes est également moins bien corrigé.

## Vues en coupes

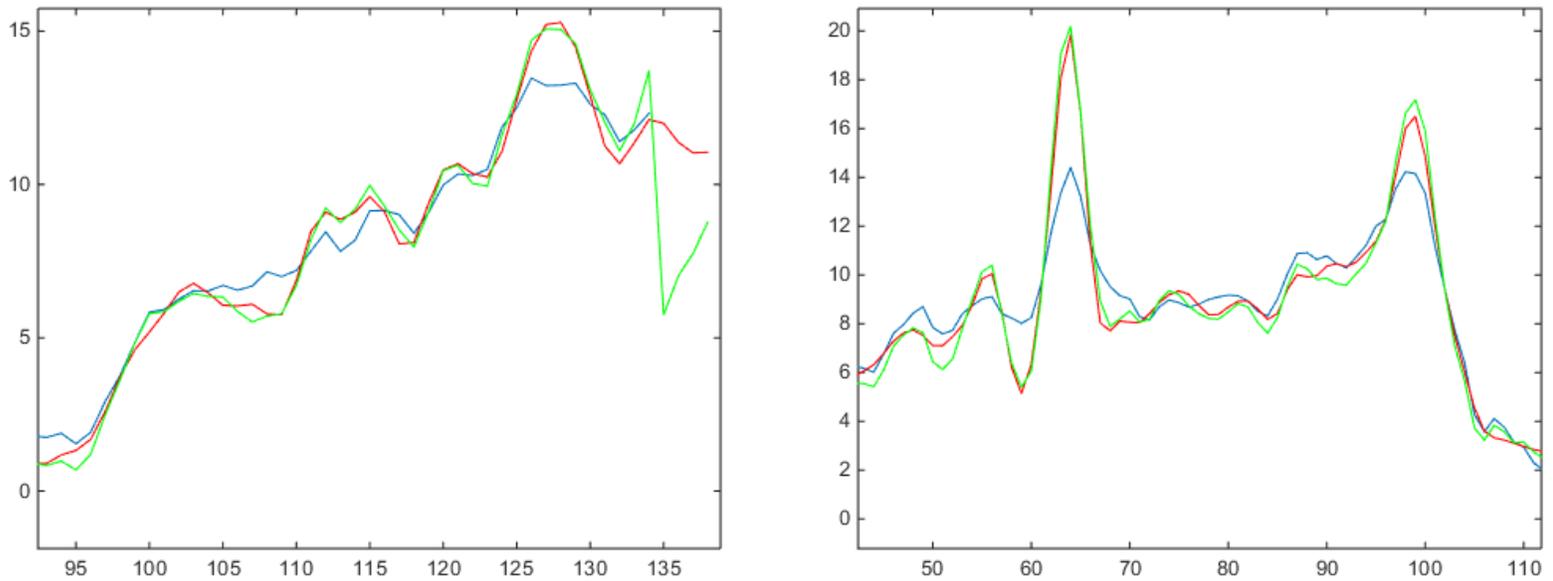


FIGURE 3.11 – Vues en coupe de la colonne 70 (gauche) et de la ligne 118 (droite). Image obtenue par coaddition (bleu), résultats de SUPREME (rouge) et SUPREME non-quadratique (vert).

La coupe suivant la colonne 70 et permet de bien voir le problème de reconstruction sur les bords de l'image. Cette fois, on obtient autant voire plus de rebonds que dans le cas quadratique, ce qui signifie que l'on doit trouver une meilleure valeur de  $\kappa_d^2$  plus adaptée ou jouer sur la valeur du facteur de forme  $\nu$ .

# Conclusion

L'introduction des variables cachées a bien permit de limiter les phénomènes de Gibbs sur l'image reconstruite. On observe parfois un problème de reconstruction sur les bords de la zone d'intérêt, problème probablement causé par la détection d'un gradient impropre à l'image à reconstruire. Bien que nous nous soyons concentrée sur la reconstruction de la partie lisse de l'image ( $s$  dans l'article présentant SUPREME), il serait possible d'ajouter au code la reconstruction de la partie parcimonieuse de l'image ( $p$  dans l'article). L'image ci-dessous permet de bien voir ces deux parties.

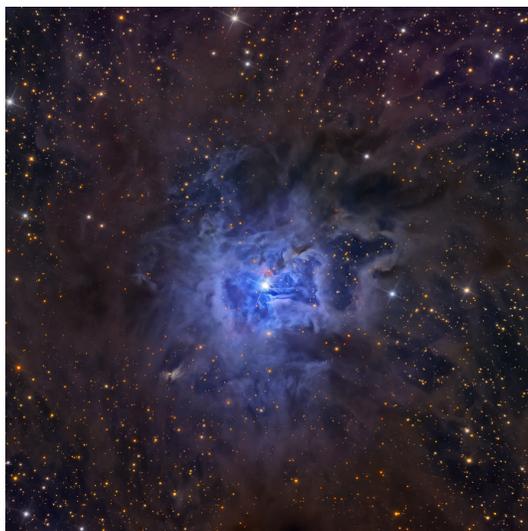


FIGURE 3.12 – Image de NGC 7023 (NASA)

De plus, il serait possible de faire un code non-supervisé en calculant directement les moments d'ordre 1 et 2 de la loi de  $\kappa_d$  à partir de ses paramètres  $(\alpha, \beta, \mu)$ . Cela permettrait de s'affranchir de l'algorithme à marche aléatoire. Il resterait cependant à jouer sur la valeur du facteur de forme  $\nu$  qui dépend a priori de l'image à traiter.

# Bibliographie

[Fr11] : Fraysse A and Rodet T 2011 A gradient-like variational Bayesian algorithm IEEE Statistical Signal Processing Workshop pp 605–8

[Or11] : Orioux F, Giovannelli J, Rodet T, Abergel A, Ayasso H and Husson M 2012 Super-resolution in map-making based on a physical instrument model and regularized inversion. Application to SPIRE/Herschel Astron. Astrophys. 539 16

[Gio08] : Giovannelli, J. F. (2008). Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function. Image Processing, IEEE Transactions on, 17(1), 16-26.

[GY95] : D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. IEEE Trans. Image Process., 4(7) :932946, July 1995.

[Yu14] : Yuling Zheng. Algorithmes bayésiens variationnels accélérés et applications aux problèmes inverses de grande taille. Image Processing. Université Paris Sud - Paris XI, 2014. French. <NNT : 2014PA112354>. <tel-01135180>