Deux méthodes de déconvolution de mélanges de composantes; application à la reconstruction des amas de galaxies

S. ANTHOINE[1], E. PIERPAOLI[2], I. DAUBECHIES[1]

[1]Program in Applied and Computational Mathematics
Princeton University, Princeton, NJ, 08544, USA

[2]California Institute of Technology
Mail Code 130–33, Pasadena, CA, 91125, USA

anthoine@princeton.edu, pierpa@caltech.edu
ingrid@math.princeton.edu

Résumé – Nous présentons deux approches pour résoudre le problème de séparation et de déconvolution simultané de mélanges de composantes. La première est basée sur une description statistique des coefficients d'ondelettes des signaux. La seconde est nouvelle et consiste à minimiser une fonctionnelle variationnelle. Nous appliquons ces deux méthodes à la reconstruction des amas de galaxies par l'effet Sunyaev-Zeldovich dans le cadre de la mission d'observation du fond diffus cosmique *ACT*. Nous trouvons que pour des paramètres appropriés, les deux méthodes donnent des résultats comparables et que prendre en compte le caractère non-gaussien des amas très intenses améliore nettement leur reconstruction.

Abstract – Two approaches are presented to solve the problem of simultaneously deconvolving and separating mixtures of components. The first one uses a statistical description of the wavelet coefficients of the signals. The second one is new and consists in minimizing a variational functional. Both methods are applied to the reconstruction of Sunyaev-Zeldovich galaxy clusters from Cosmic Microwave background experiments such as *ACT*. We find that both methods, when tuned, yield similar results and that the reconstruction of intense clusters is substantially improved when their non-gaussianity is taken into account.

1 Introduction

Le problème que nous nous proposons de résoudre est le suivant : nous souhaitons estimer plusieurs composantes, notées $x^1, x^2, \cdots, x^M$, à partir de plusieurs observations, notées $y^1, y^2, \cdots, y^L$. Chaque observation est une version flue et bruitée d'un mélange (combinaison linéaire) des composantes originales. Autrement dit, y^l, l'observation numéro l, s'écrit :

$$y^l = \sum_{m=1}^{M} f^{m,l}\, x^m * b^l + n^l \qquad (1)$$

où $*$ désigne l'opérateur de convolution, les scalaires $f^{m,l}$ ainsi que les noyaux de convolution b^l sont supposés connus, et n^l est un bruit de variance connue : σ_l^2. Les observations du fond diffus cosmique (rayonnement fossile) en astrophysique sont un exemple de données correspondant à ce modèle. Typiquement, on acquiert plusieurs observations de la même portion de ciel à des longueurs d'onde différentes (y^l). Dans chaque observation coexistent différentes composantes : le fond diffus cosmique (x^1), des amas de galaxies (x^2), des points sources (x^3)... La contribution de chaque composante dépend de la longueur d'onde d'observation ($f^{m,l}$). Le processus d'acquisition de l'image sur le télescope se traduit par une convolution ($*b^l$). Enfi n les observations sont imparfaites (n^l).

Le fi ltre de Wiener, habituellement utilisé pour estimer une composante à partir d'une observation, peut facilement être adapté à ce cas. Optimal si les composantes peuvent être décrites comme des processus gaussiens, il donne de bons résultats lorsque cette hypothèse est une bonne approximation. Si le modèle gaussien est approprié pour le fond diffus cosmique, ce n'est pas le cas pour les amas de galaxies, qui sont des structures spatialement localisées avec un pic d'intensité au centre et des structures fi lamenteuses vers l'extérieur. En conséquence, utiliser une méthode de type fi ltre de Wiener donne des estimations peu précises des amas de galaxies. De plus, ce manque de précision pollue l'estimation du fond diffus cosmique. Il semble donc utile de développer des méthodes qui prennent en compte les caractéristiques non gaussiennes des signaux.

Dans la section 2, nous décrivons deux méthodes de ce type. La méthode présentée au paragraphe 2.1, inspirée de [5] et déjà présentée dans [3], exploite le fait que les coeffi cients d'ondelettes varient de manière cohérente à travers les échelles et à des positions proches les unes des autres. Chaque composante est caractérisée par les statistiques de voisinages de coeffi cients en ondelettes et une estimation bayesienne aux moindres carrés est calculée.

Nous décrivons au paragraphe 2.2 une nouvelle méthode qui consiste à trouver les minimiseurs d'une fonctionnelle variationnelle composée d'un terme d'erreur par observation et d'un terme de pénalisation par composante. Chaque terme de pénalisation est adapté à la composante correspondante : c'est la norme dans un espace fonctionnel approprié. Par exemple, pour le fond diffus cosmique, qui est un signal régulier à variations lentes, nous utilisons la norme dans un espace de type Sobolev. Un algorithme itératif permet de minimiser cette fonctionnelle.

Enfi n dans la section 3, nous présentons des résultats pour chacune des méthodes sur des données simulées avec les paramètres du projet *ACT*.

2 Deux méthodes de déconvolution de mélanges de composantes :

2.1 Estimation bayesienne de voisinages de coefficients d'ondelettes

2.1.1 Modèle statistique

Considérons une décomposition de type ondelettes pour le signal x : $x = \sum_{q\in\mathbb{Z}}\langle x,\phi_q\rangle\phi_q + \sum_{j=0}^{J}\sum_{q\in 2^{-j}\mathbb{Z}^2}\langle x,\psi_{j,q}\rangle\psi_{j,q}$ où ϕ est la fonction d'échelle, $\phi_q(.) = \phi(.-q)$, les $\psi_{j,q} = \psi(2^j.-q)$ sont les ondelettes et $\langle,\rangle$ désigne le produit scalaire. Notons $x_{j,q} = \langle x,\psi_{j,q}\rangle$. Si le signal x a une structure particulière au point q_0 et à l'échelle j_0, alors les coeffi cients d'ondelettes à des échelles ou des positions proches sont en général aussi affectés. Par exemple, si x a une discontinuité au point q_0, tous les coeffi cients d'ondelettes $\{x_{j,q_0}\}_j$ sont grands. On peut donc espérer faire une meilleure estimation du coeffi cient $x_{j,q}$ en considérant un ensemble de coeffi cients à des échelles proches de j et des points proches de q, autrement dit un « voisinage » . Un exemple de « voisinage » de $x_{j,q}$ est : $\mathbf{x} \equiv \mathbf{x}_{j,q} = (\ x_{j,q}\ ,\ x_{j,q-1}\ ,\ x_{j,q+1}\ ,\ x_{j-1,q}\ )$. Comme dans [5], nous modélisons le voisinage $\mathbf{x}_{j,q}$ par une mixture de gaussiennes dépendant uniquement de l'échelle j :

$$\forall q,\ \mathbf{x}_{j,q} \equiv \sqrt{z}\ \mathbf{u} \textit{ en distribution} \qquad (2)$$

où z est une variable aléatoire scalaire, indépendante de $\mathbf{u}$, de moyenne 1 et de distribution p_z et $\mathbf{u}$ est un vecteur gaussien centré, de covariance $\mathbf{C_u}$. p_z et $\mathbf{C_u}$ dépendent tous deux du signal x et de l'échelle j. Ce modèle étant donné pour chaque composante x^m, l'estimation du coeffi cient $x^m_{j,q}$ est le premier coeffi cient de l'estimateur bayesien du voisinage $\mathbf{x}^m_{j,q}$; le bruit étant supposé gaussien. (Il peut être blanc ou coloré).

2.1.2 Débruitage d'une composante

Ici, on estime une composante x à partir d'une observation bruitée $y = x + n$. Si de plus, on suppose $p_z = \delta_{\{z=z_0\}}$, alors les voisinages sont supposés gaussiens et l'estimation bayesienne de $\mathbf{x}$ n'est autre que le fi ltre de Wiener appliqué aux voisinages d'ondelettes :

$$E\{\mathbf{x}_{j,q}|\mathbf{y}_{j,q}, z = z_0\} = z_0\mathbf{C}_{x,j}(z_0\mathbf{C}_{x,j} + \mathbf{C}_{n,j})^{-1}\mathbf{y}_{j,q}. \quad (3)$$

où $\mathbf{C}_{x,j}$ désigne la matrice de covariance des voisinages d'ondelettes du signal x à l'échelle j. Pour p_z quelconque, on obtient donc une combinaison linéaire de fi ltres de Wiener, les poids étant donnés par la règle de Bayes :

$$E\{\mathbf{x}_{j,q}|\mathbf{y}_{j,q}\} = \int_0^\infty p(z = z_0|\mathbf{y}_{j,q})E\{\mathbf{x}_{j,q}|\mathbf{y}_{j,q}, z = z_0\}\ dz_0 \qquad (4)$$

$$\text{où}\quad p(z = z_0|\mathbf{y}_{j,q}) = \frac{p(\mathbf{y}_{j,q}|z = z_0)p_z(z_0)}{\int p(\mathbf{y}_{j,q}|z = z')p_z(z')dz'} \qquad (5)$$

et $p(\mathbf{y}_{j,q}|z = z')$ est une gaussienne centrée de covariance $z'\mathbf{C}_{x,j} + \mathbf{C}_{n,j}$. Ce cas a été présenté dans [5].

2.1.3 Déconvolution de mélanges de composantes

Nous avons étendu le résultat prédédent au cas des mélanges de composantes décrits par l'éq. (1) (cf. [3], [4]). A échelle j et position q fi xées, les M voisinages $\mathbf{x}^m_{j,q}$ sont estimés simultanément pour toutes les composantes x^m à partir des L voisinages observés $\{\mathbf{y^l}_{j,q}\}_{l=1,..,L}$. Pour chaque m, le voisinage $\mathbf{x}^m_{j,q}$ est modélisé par une mixture de gaussiennes : $\mathbf{x}^m_{j,q} \overset{dist.}{\equiv} \sqrt{z^m}\mathbf{u^m}$, qui est déterminée par la probabilité p_{z^m} et la matrice de covariance $\mathbf{C}^m_u$. Comme précédemment, si les p_{z^m} sont des fonctions δ pour toutes les composantes, on retrouve l'hypothèse gaussienne. Dans ce cas, l'estimation des $\mathbf{x}^m_{j,q}$ est donnée par le fi ltre de Wiener multi-objets/multi-observations sur les voisinages de coeffi cients d'ondelettes. Notons que ces fi ltres tiennent en compte le mélange par les coeffi cients $f_{m,l}$ et les convolutions par les b_l. L'estimation du modèle complet est une combinaison linéaire de tels fi ltres, les poids étant donnés par les probabilités à postériori :

$$p(\{z^m\}_{m=1,..,M}|\{\mathbf{y^l}_{j,q}\}_{l=1,..,L}). \qquad (6)$$

Comme pour le cas monodimensionnel, les probabilités à postériori sont calculées en utilisant la règle de Bayes, et en conséquence sont des combinaisons de gaussiennes multidimensionnelles $p(\{\mathbf{y^l}_{j,q}\}_{l=1,..,L}|\{z^m\}_{m=1,..,M})$ avec les probabilités à priori p_{z_m}, $m = 1,..,M$. Nous notons ici encore, que les coeffi cients de mélange $f_{m.l}$ et les convolutions par les b_l interviennent dans les probabilités $p(\{\mathbf{y^l}_{j,q}\}_l|\{z^m\}_m)$. Une description détaillée des formules obtenues se trouve dans [3].

La taille des voisinages considérés doit dépendre de l'échelle j (car la transformée a un pas de translation dépendant de j), ainsi que des noyaux de convolution b_l (puisque la convolution corrèle les coeffi cients). D'autre part, le calcul de l'estimateur bayesien requiert les matrices de covariances des voisinages à chaque échelle pour chaque composante et pour le bruit ainsi que des probabilités p_{z^m}. Nous expliquons dans le paragraphe 3.2 comment déterminer ces paramètres pour l'application à l'extraction des amas de galaxies.

2.2 Minimisation d'une fonctionnelle variationnelle par un algorithme itératif

2.3 Modèle

Posons $X = (x^1, x^2, .., x^M)$, $Y = (y^1, y^2, .., y^L)$ et $N = (n^1, n^2, .., n^L)$, alors l'équation (1) s'écrit : $Y = KX + N$ et K est un opérateur linéaire continu de $(L^2)^M$ dans $(L^2)^L$. Nous cherchons donc à « inverser » K tout en contrôlant la taille du terme dû au bruit après inversion ($K^{-1}N$). Comme K n'a pas nécessairement un inverse borné, le mieux que l'on puisse espérer est un schéma régularisant i.e.

$$\forall X, \lim_{\epsilon\to 0}\ \sup_{||N||<\epsilon}\ ||K^{-1}(KX + N) - X|| \to 0. \qquad (7)$$

La minimisation d'une fonctionnelle variationnelle permet de défi nir un tel schéma. Par exemple, la minimisation de

$$J_{L^2}(X) = ||KX - Y||^2_{(L^2)^L} + \gamma||X||^2_{(L^2)^M} \qquad (8)$$

$$J_{L^2}(X) = \sum_{l=1}^{L}||(KX)^l - y^l||^2_{L^2} + \gamma\sum_{m=1}^{M}||X^m||^2_{L^2} \qquad (9)$$

avec γ strictement positif est une méthode régularisante pour notre problème. Minimiser cette fonctionnelle revient à chercher des composantes X^m dont l'image par l'opérateur K est aussi près que possible des observations Y^l (premier terme)

et de norme aussi petite que possible dans L^2 (second terme). Toutefois, elle traite chacune des composantes de la même façon. Qui plus est, le choix de la norme L^2 n'est pas nécessairement adéquat pour toutes les composantes X^m. Par exemple, ajouter des oscillations change peu la norme quadratique d'un objet. Si K est un fi ltre basse fréquence, alors ajouter des oscillations hautes fréquences à une composante X^m change peu la fonctionnelle. Il est donc très possible d'obtenir une estimation X^m oscillant à haute fréquence. Ceci n'est pas acceptable par exemple pour la composante représentant les amas de galaxies puisqu'on sait qu'ils sont localisés spatialement.

Nous proposons donc de minimiser une fonctionnelle variationnelle du type suivant :

$$J_S(X) = ||KX - Y||^2_{(L^2)^L} + \sum_{m=1}^{M} \gamma_m ||X^m||^{p^m}_{\mathcal{G}^m} \quad (10)$$

$$J_S(X) = \sum_{l=1}^{L} ||(KX)^l - y^l||^2_{L^2} + \sum_{m=1}^{M} \gamma_m ||X^m||^{p^m}_{\mathcal{G}^m} \quad (11)$$

où les γ_m sont strictement positifs pour tout m et $\mathcal{G}^m$ est un espace fonctionnel choisi de telle manière que les éléments ayant les mêmes propriétés que x^m ont une petite norme dans cet espace. Par exemple, on choisira un espace de Sobolev pour un signal régulier à variations lentes comme le fond diffus cosmique et un espace de Besov pour un signal localisé avec des transitions rapides comme les amas de galaxies.

2.4 Estimation

En toute généralité, nous ne savons pas minimiser une telle fonctionnelle. Mais s'il existe une base orthonormée $\{\phi_\lambda\}_{\lambda\in\Lambda}$ telle que pour chaque espace fonctionnel $\mathcal{G}^m$, il existe des poids strictement positifs $\mathbf{w}^m = \{w^m_\lambda\}_{\lambda\in\Lambda}$ et p^m tels que :

$$||.||_{\mathcal{G}^m} = \left(\sum_{\lambda\in\Lambda} w^m_\lambda |\langle ., \phi_\lambda\rangle|^{p^m}\right)^{\frac{1}{p^m}}, \quad 1 \le p^m \le 2 \quad (12)$$

alors il existe un algorithme itératif permettant de trouver des minimiseurs. Cet algorithme à été introduit par exemple dans [1] dans le cas d'une composante et d'une observation. Nous l'avons généralisé au cas de plusieurs composantes et observations. La généralisation étant assez directe, les propriétés de convergence et de régularisation prouvées dans [1] s'en suivent. L'algorithme itératif suivant converge fortement vers un minimiseur de l'équation (11) :

- X_o quelconque
- $\forall n > 0$:

$$X_{n+1} = \sum_{\lambda\in\Lambda} S_{\theta^m_\lambda, p_m}\left(\left\langle \left[cX_n + K^*g - K^*KX_n\right]^m, \varphi_\lambda\right\rangle\right)\varphi_\lambda, \quad (13)$$

où K^* est l'adjoint de K, c vérifi e $c > \|K^*K\|$ et $\theta^m_\lambda = \gamma_m w^m_\lambda$. Les fonctions $S_{\theta,p}$ sont défi nies par :

$$S_{\theta,p}\left(x + \frac{\theta}{2}\ \mathrm{sign}(x)\ |x|^{p-1}\right) = x \quad (14)$$

Nous utilisons en particulier $p_m = 1$ ou $p_m = 2$. Pour $p = 1$, $S_{\theta,1}$ est le seuillage doux de paramètre $\frac{\theta}{2}$:

$$S_{\theta,1}(x) = \begin{cases} x + \mathrm{sign}(x)\frac{\theta}{2} & \text{si } x \ge \frac{\theta}{2} \\ 0 & \text{sinon} \end{cases} \quad (15)$$

Et pour $p = 2$:

$$S_{\theta,2}(x) = \frac{x}{1 + \frac{\theta}{2}} \quad (16)$$

3 Application : reconstruction des amas de galaxies

3.1 Données

Les deux méthodes précédentes ont été appliquées à des données simulées avec les paramètres du projet d'observation du fond diffus cosmique *ACT* (voir [3] pour plus de précisions sur les simulations). Nous considérons le cas de deux composantes : x^1 est le fond diffus cosmique et x^2 les amas de galaxies. Nous disposons de trois observations : à 145, 217 et 265 GHz. Les noyaux de convolution sont gaussiens, d'écart-type 1.7, 1.1 et 0.93 $arcminute$. Le bruit est blanc, d'écart type 3.4, 3.63 et 4.37μK par arcminute pixel. Les simulations couvrent une portion de ciel de 1.19x1.19 degrés et comptent 512 pixel par côté.

Une example d'observation à 265 GHz est donné dans la Figure 1 (haut), ainsi que la contribution du fond diffus cosmique (au bas à gauche) et des amas de galaxies (en bas à droite).

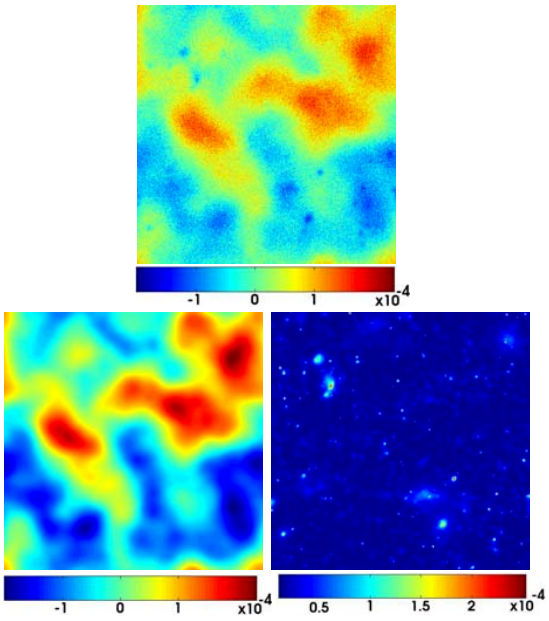

FIG. 1 – Données : en haut, une observation g_l, en bas, la contribution de deux composantes à cette observation : à gauche celle du fond diffus cosmique, à droite, celle des amas de galaxies. Échelle de couleur en Kelvin.

3.2 Paramètres de la méthode statistique

Les matrices de covariance du bruit sur les voisinages sont calculées théoriquement puisque la covariance du bruit est supposée connue. Le signal venant des amas de galaxies est en revanche trop peu puissant par rapport au fond diffus cosmique ou au bruit pour que les matrices de covariance soient estimées précisément à partir des observations. Elles sont calculées à partir de simulations de chaque composante.

La probabilité p_z est fi xée à $\delta_{\{z=1\}}$ pour le fond diffus cosmique car il est modélisé par un signal gaussien. Pour les amas de galaxies, nous avons comparé quatre possibilités pour p_z

dans [3] et en présentons ici deux : celle dérivée des simulations et $\delta_{\{z=1\}}$. Dans le cas où l'on utilise $p_z = \delta_{\{z=1\}}$, les voisinages pour les amas de galaxies sont supposés gaussiens. Au contraire, la probabilité p_z dérivée des simulations induit un modèle non-gaussien.

3.3 Paramètres de la méthode variationnelle

Le fond diffus cosmique est pénalisé par une norme de type Sobolev avec des poids ajustés à sa puissance spectrale. Plus précisément, nous utilisons une norme l^2 sur les coeffi cients d'ondelettes, avec des poids $w_{j,q}$ ne dépendant que de l'échelle j, inversement proportionnels à la puissance spectrale à celle échelle. Cela permet de pénaliser les variations rapides et de prendre en compte la puissance spectrale du fond diffus cosmique. Pour les amas de galaxies, nous utilisons une norme l^1 sur les coeffi cients d'ondelettes (ce qui est équivalent à une norme dans un espace de Besov B_1^1), favorisant la sparsité du signal et les transitions rapides.

Enfi n les coeffi cients γ_m sont choisis de manière à ce que chacun des termes dans la fonctionnelle (11) aient le même ordre de grandeur.

3.4 Résultats

Dans cette section, nous comparons les résultats obtenus sur une série de 24 simulations avec les paramètres de *ACT*. Les trois types de reconstructions considérés sont celles de la méthode variationnelle, celles de la méthode statistique gaussienne avec $p_z = \delta_{\{z=1\}}$ pour les amas de galaxies, et celles de la méthode statistique non-gaussienne avec p_z dérivée des simulations pour les amas de galaxies.

Nous avons implémenté chacune des deux méthodes, avec deux transformées en ondelettes différentes : la *steerable pyramid* décrite dans [5] et la transformée en ondelettes complexes de [2]. Toutes deux sont des transformées redondantes, qui permettent de décomposer les images en sous-bandes orientées. L'avantage sur une base d'ondelette ordinaire est que l'on obtient une meilleure selectivité directionnelle (le nombre d'orientations peut être choisi), et que l'on se rapproche de l'invariance par translation. Les deux transformées donnent des résultats de qualité égale. Nous donnons ici les résultats avec la *steerable pyramid* et 4 orientations.

Nous trouvons que les différentes reconstructions du fond diffus cosmique sont de qualité égale. En revanche, pour les amas de galaxies, les différences sont notables. La fi gure 2 montre que la méthode variationnelle et la méthode statistique non-gaussienne donnent des résultats comparables et que ceux-ci sont nettement meilleurs que les résultats de la méthode statistique gaussienne en terme de reconstruction de la valeur nominale d'amas très intenses. Dans [3], nous avons établi que les reconstructions par la méthode statistique non-gaussienne sont en moyenne 20% meilleurs que ceux de la méthode statistique gaussienne pour les amas très brilliants. Le résultat est similaire pour la méthode variationnelle. Les différences entre la méthode variationnelle et la méthode statistique non-gaussienne sont plus fi nes. Nous trouvons que la méthode variationnelle résout les amas plus compacts et permet de reconstruire les amas plus intenses plus précisément que la méthode statistique non-gaussienne. En revanche, les structures fi lamenteuses sont mieux préservées par cette dernière.

Ainsi chaque approche, fonctionnelle ou statistique, présente ses propres avantages dans le cadre de *ACT*. Il convient d'étudier plus en détails l'influence des paramètres de la mission considérée (longeurs d'onde d'observation, niveau de bruit...) pour choisir l'approche à utiliser. Par ailleurs une étude plus poussée incorporant les points sources, une composante non modélisée ici, est en cours.

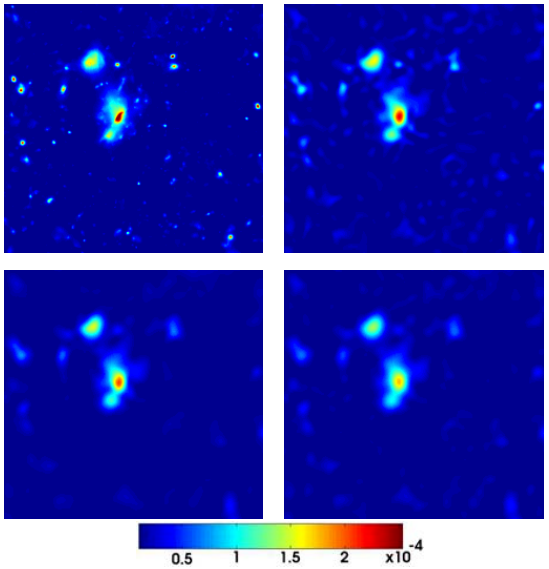

FIG. 2 – Zoom sur un amas étendu. En haut à gauche l'original, puis 3 reconstructions : en haut à droite par la méthode variationnelle ; en bas à gauche par la méthode statistique avec p_z dérivée des simulations ; en bas à droite par la méthode statistique avec $p_z = \delta_{\{z=1\}}$. Les cartes ont la même échelle de couleur (en Kelvin).

Références

[1] I. Daubechies, M. Defrise, C. De Mol. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*. Comm. Pure Appl. Math., 2004, vol. 57, issue 11, p. 1413–1457.

[2] N. G. Kingsbury, *Complex wavelets for shift invariant analysis and filtering of signals*, Journal of Applied and Computational Harmonic Analysis, vol 10, no 3, May 2001, pp. 234-253.

[3] E. Pierpaoli, S. Anthoine, K. Huffenberger, I. Daubechies, *Reconstructing Sunyaev-Zeldovich clusters in future CMB experiments*. Mon. Not. Roy. Astron. Soc., 2005, vol. 359, issue 1, p. 261-271.

[4] E. Pierpaoli, S. Anthoine, *Finding SZ clusters in the ACBAR maps*. JASR, COSPAR 2004, à paraître.

[5] J. Portilla, V. Strela, M. Wainwright, E. Simoncelli. *Image denoising using a scale mixture of Gaussians in the wavelet domain*. IEEE Trans. Image Proc., 2003, vol. 12, issue 11, p. 1338–1351.