Classification Morphologique d'un Echantillon Optique d'Amas de Galaxies

Florent Rostagni*

Master Physique Fondamentale et Appliquée ΩMEGA
Université de Nice-Sophia Antipolis

Christophe Benoist† and Sophie Maurogordato‡

Laboratoire Cassiopée

(Dated: 11 Juin 2009)

Une méthode de classification automatique d'amas de galaxies du Sloan Digital Sky Survey (SDSS) a été mise en place : détection des amas et classification selon leur morphologie. Cette classification est essentielle pour se servir des amas comme traceurs cosmologiques et pour étudier la physique des amas de galaxies. Des 211 champs présents, 218 amas de galaxies ont été extraits puis classés. 24 % des amas sont en interaction, 31 % des amas en cours d'accrétion et 45 % sont dans un état d'équilibre.

Contents

I. Introduction 1

II. Quelques propriétés physiques des amas 2
A. Distribution de redshifts des amas 2
B. Masse des amas 3

III. Description de l'échantillon d'amas 3

IV. Méthodologie 4
A. Travaux antérieurs 4
B. Construction de champs de densité 5
C. Un programme de décomposition en mixtures gaussiennes : EMMIX 6
1. Détection des gaussiennes 7
2. Affectation des points 10
D. Extraction d'un catalogue de structures 10
E. Algorithme d'identification et de classification des amas 10

V. Analyse du catalogue 12
A. Identification des amas 12
B. Classification morphologique 13

VI. Conclusion 13

VII. Perspectives 14
A. Comparaison avec les observations en X 14
B. Evolution de la morphologie avec le redshift 14
C. Etat dynamique des amas et formation stellaire de leurs galaxies 14

Remerciements 14

Glossaire 14

Références 15

*Electronic address: rostagni@oca.eu
†Electronic address: benoist@oca.eu
‡Electronic address: maurogor@oca.eu

I. INTRODUCTION

Les amas de galaxies sont les structures les plus massives gravitationnellement liées de l'univers. Dans le cadre du modèle cosmologique standard, ils se sont formés relativement récemment par l'assemblage d'unités moins massives attestant du processus de formation hiérarchique des structures. Comme l'ont révélé les grands relevés tri-dimensionnels de galaxies (SDSS, 2dF,...), la distribution de la matière à grande échelle présente une topologie complexe caractérisée par de grands vides, de grands « murs » ou encore des filaments aux noeuds desquels sont localisés les amas. Outre l'accrétion continue de galaxies à travers ces filaments, les amas connaissent encore à l'époque actuelle des épisodes fréquents de fusion avec des groupes ou autres amas de galaxies. Une des signatures de ces processus est la présence de sous-structures ou une morphologie perturbée et fortement asymétrique. De telles signatures sont reproduites par les simulations numériques cosmologiques les plus récentes.

Comme le montre la figure 1, le spectre de puissance de la distribution de la matière et la vitesse de croissance des structures dépend du modèle cosmologique considéré. Ainsi, si la structure des amas permet d'appréhender la physique de formation des structures, elle est aussi une sonde cosmologique de premier plan (22). Par exemple, dans des univers de faible densité (en haut de la figure 1), les amas se forment plus tôt et seront en moyenne plus relaxés à l'époque actuelle. Dans le cadre de ce stage, nous proposons de réaliser la classification morphologique d'un échantillon de plus de 200 amas issus du grand relevé de galaxies Sloan Digital Sky Survey (SDSS) qui est aujourd'hui la référence pour l'univers local. La classification automatique mise en œuvre s'appuie sur une méthode originale appliquée pour la première fois sur des données d'imagerie optique et de spectroscopie.

Dans un premier temps (section II), nous décrivons quelques propriétés physiques des amas de galaxies utiles au présent travail. Dans la section III, nous présentons

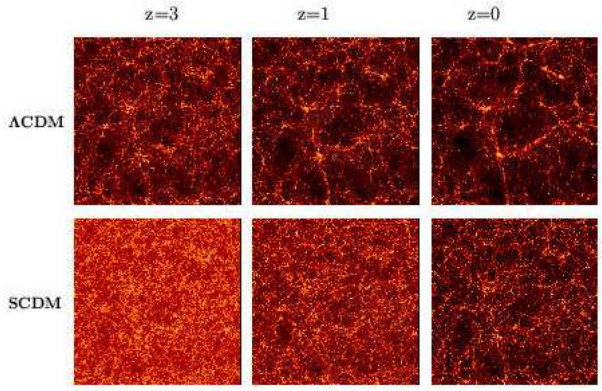


FIG. 1 Simulations de deux modèles d'univers à trois époques différentes. Les densités sont normalisées par rapport à la densité maximale pour un univers à une date donnée. Selon le modèle utilisé la formation des structures est plus rapide. Par exemple, le modèle SCDM correspond à un univers plus dense que celui du ΛCDM et les amas de galaxies aux noeuds des filaments se forment plus lentement.

l'échantillon d'amas et les données utilisées dans cette étude. Les outils et la méthode de classification sont présentés dans la section IV, et leur application dans la section suivante. Enfin, les sections VI et VII présentent la conclusion et les perspectives de ce travail.

II. QUELQUES PROPRIÉTÉS PHYSIQUES DES AMAS

Les amas de galaxies relaxés sont des regroupements réguliers de plusieurs centaines de galaxies et ont une masse comprise entre $10^{13} M_\odot$ et $10^{15} M_\odot$[1].

Les galaxies représentent environ 5 % de la masse totale des amas. Elles sont observées dans le domaine optique mais aussi à d'autres longueurs d'onde (UV, Infrarouge, mm, radio . . .). La distribution dans le plan du ciel des galaxies d'un amas relaxé suit un profil de King (15) :

$$\Sigma(r) = \Sigma(0) \left(1 + \frac{r^2}{R_C^2}\right)^{-1},$$

où R_C est le rayon de cœur. Ce profil permet de définir une taille caractéristique d'un amas de l'ordre de 0.5 Mpc[2]. Les amas contiennent aussi du gaz chaud, émetteur dans le domaine X et représentant environ 15 % de la masse. Le gaz et les galaxies suivent les mêmes distributions en densité dans les amas relaxés. Enfin, 80 % de la masse se trouve sous forme de « matière noire », de nature inconnue. Elle peut être mise en évidence par l'effet de lentille gravitationnelle induit par sa masse. La matière noire représentant la majeure partie de la masse des amas, c'est son potentiel qui détermine les orbites des galaxies. Ainsi, étudier la distribution des positions et vitesses des galaxies peut permettre de remonter au potentiel gravitationnel de la matière noire.

A. Distribution de redshifts des amas

Le redshift z d'un objet astrophysique est la valeur du décalage de son spectre (densité spectrale d'énergie lumineuse émise par l'objet) vers le rouge. Du fait de l'expansion de l'univers, les objets lointains s'éloignent les uns des autres. En faisant une analogie avec l'effet Doppler-Fizeau, on sait relier le décalage spectral d'un objet avec sa vitesse de déplacement par rapport à l'observateur :

$$z = \frac{\lambda - \lambda_{ref}}{\lambda_{ref}} = \frac{v}{c} \qquad (\text{si } z \ll 1)$$

donc $v = cz$. A bas z, la vitesse radiale est donc proportionnelle au redshift ; sinon le calcul de la vitesse dépend des constantes cosmologiques. Grâce au spectre, on peut donc déterminer le redshift et la vitesse dite de récession. Ensuite, la loi de Hubble relie vitesse de récession et distance physique de l'objet :

$$v = HD$$

où H est la constante de Hubble et vaut environ 75 km/s/Mpc. Donc, plus la vitesse d'un objet est importante, plus la distance qui nous sépare de lui est grande et il correspond à une époque reculée. Par exemple, $z = 1$ correspond à la moitié de l'âge de l'univers. En combinant cette information avec les coordonnées dans le plan du ciel (2D), la localisation de l'objet dans l'espace (3D) peut en être déduite. Toutefois, le redshift observé ne permet pas strictement de remonter à la distance réelle nous séparant d'un objet. En effet, le redshift observé $z_{\rm obs}$ est la somme de deux termes :

- le redshift cosmologique $z_{\rm cosmo}$ qui détient l'information sur la distance ;
- le mouvement propre de l'objet $z_{\rm mvt} = v_{\rm mvt}/c$.

Les vitesses particulières entrainent un biais dans l'utilisation du redshift comme indicateur de distance. Pour les objets proches, à bas z, le redshift cosmologique et le redshift dû au mouvement propre sont comparables (il peut même y avoir un « blueshift » ($z < 0$) si l'objet s'approche), donc déterminer une distance à partir du redshift observé n'a pas beaucoup de sens. De même, pour les amas de galaxies, la dispersion de vitesses des galaxies provoque un effet dit de « doigt de Dieu ». Sur la figure 2, les galaxies d'une petite région du ciel sont représentées dans un diagramme ascension droite (une coordonnée dans le plan du ciel) - redshift. Un amas de galaxies est mis en évidence par l'ellipse verte. L'effet de doigt de Dieu allonge artificiellement l'amas sur la ligne de visée, alors qu'à l'échelle de cette figure les galaxies devraient être à la même distance.

Pour un amas de galaxies relaxé, les vitesses des galaxies suivent à priori une distribution gaussienne (24).

[1] $1 M_\odot \simeq 2 \times 10^{30}$kg (masse solaire).
[2] 1 parsec (pc) = 3.26 années-lumière (a.l.)

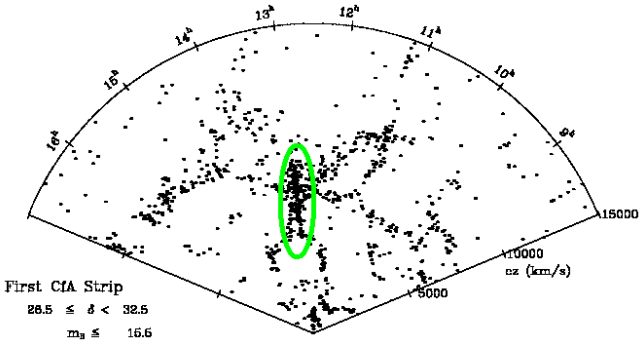


FIG. 2 Illustration de l'effet de « doigt de Dieu ». Les galaxies sont représentées dans un diagramme ascension droite-vitesse. Les galaxies d'amas ont une vitesse propre due à leurs orbites dans le potentiel gravitationnel de l'amas ce qui induit un effet de doigt de Dieu. Ici, un amas est mis en évidence par l'ellipse verte : l'amas est allongé dans la direction radiale.

Bande	λ_0 (nm)	$\Delta\lambda$ (nm)
u	359	59.9
g	481	137.9
r	623	138.2
i	764	153.5
z	906	137

TAB. I Bandes photométriques du SDSS centrées en λ_0 et de largeur à mi-hauteur $\Delta\lambda$

B. Masse des amas

A l'équilibre, la masse des groupes et des amas de galaxies peut être calculée à partir de la distribution de vitesses des galaxies en utilisant le théorème du Viriel : à l'équilibre, la moitié de l'énergie gravitationnelle (reliée à la masse) est égale à l'opposé de l'énergie cinétique. L'énergie cinétique, dans l'hypothèse de l'isotropie des vitesses, est proportionnelle au carré de la dispersion des vitesses :

$$E_c = \frac{3}{2} M \langle \sigma_v^2 \rangle.$$

Dans l'hypothèse d'un système sphérique, l'énergie gravitationnelle s'exprime ainsi :

$$E_g = -\frac{GM^2}{r_g},$$

où r_g est le rayon gravitationnel. La relation entre la masse et la dispersion des vitesses peut donc s'exprimer de la façon suivante :

$$M = 3\frac{r_g \langle \sigma_v^2 \rangle}{G}.$$

III. DESCRIPTION DE L'ÉCHANTILLON D'AMAS

Nous nous appuyons dans cette étude sur le Sloan Digital Sky Survey(SDSS) qui couvre 25% du ciel. Les observations ont débuté en 2000 et sont réalisées par un télescope de 2.5m à l'Observatoire d'Apache Point au Nouveau Mexique (USA). Le SDSS est un relevé photométrique en cinq bandes, u, g, r, i, z (voir Table I), complet jusqu'à la magnitude 24 dans la bande r. Le SDSS est aussi un relevé spectroscopique complet à 99 % dans son ensemble (26) avec une magnitude dans la bande r comprise entre $14.5 < r < 17.77$. Cependant, cette complétude est en général de l'ordre de 60-70 % (27) dans les régions les plus denses. En effet, les régions denses en galaxies, comme les amas, sont plus difficiles à échantillonner de fa çon complète pour des raisons de contraintes géométriques liées aux observations spectroscopiques.

L'avantage d'un grand relevé est l'homogénéité des données, ce qui est essentiel pour mener des études statistiques sur les amas de galaxies. En effet, d'une part, les amas de galaxies sont rares et d'autre part quand ils sont proches, un grand champ est nécessaire pour les observer en entier. Par exemple, un amas dont la taille est 1 Mpc à $z = 0.05$ couvre environ 18 arcmin alors que le champ typique couvert par une caméra est de 10 arcmin. De nombreuses données relatives aux amas de galaxies provenant de divers instruments sont disponibles. Toutefois, l'hétérogénéité de la qualité des mesures rend l'analyse et l'interprétation physique difficiles.

Un catalogue d'amas de galaxies a été extrait du SDSS à partir des couleurs des galaxies, grâce à un algorithme nommé « C4 » (19). L'échantillon d'amas C4 contient environ 2500 amas de galaxies proches, le redshift médian des amas étant 0.1046. Dans ce catalogue, nous avons sélectionné les amas ayant une couverture spectroscopique avec un échantillonnage supérieur à 80 % pour pouvoir réaliser une étude fine de leur distribution de vitesses et de leur morphologie. 211 champs ont ainsi été sélectionnés pour cette étude. Un champ correspondant à 3 Mpc au niveau de chacun des amas a été extrait du SDSS.

Chaque catalogue d'amas contient des galaxies caractérisées par leur position dans le plan du ciel, leur magnitude de Petrosian dans les bandes u, g, r, i, z. Pour les galaxies, dont le spectre a été mesuré, le redshift spectroscopique est également donné.

Dans le champ correspondant à un amas C4, des galaxies d'avant ou d'arrière plan sont également présentes et elles peuvent appartenir à d'autres structures que l'amas. La figure 8 illustre, pour deux champs, cet empilement de structures sur la ligne de visée.

Type	Description
I	amas contenant une galaxie de type cD
I-II	cas intermédiaire
II	amas contenant une ou plusieurs galaxies de type D et de type elliptiques géantes Virgo
II-III	cas intermédiaire
III	amas sans galaxie dominante

TAB. II Classification des amas de galaxies par Bautz et Morgan (3)

IV. MÉTHODOLOGIE

Grâce au C4, on sait que le champ contient un amas et grâce aux couleurs des galaxies, un redshift photométrique est évalué. Il s'agit ici d'étudier en détail la morphologie et la distribution de vitesses des amas C4. Dans le domaine X, les observations fournissent directement la densité du gaz qui est continue, contrairement aux galaxies pour lesquelles leur distribution est discontinue. Pour caractériser leur morphologie, nous allons construire des champs de densité à partir de la distribution de galaxies.

Après avoir présentées quelques méthodes de classification développées dans des études antérieures, je vais présenter une nouvelle méthode basée sur une analyse en ondelettes des distributions de vitesses puis des champs en densité des galaxies dans le plan du ciel.

A. Travaux antérieurs

Les premières classifications d'amas de galaxies ont été réalisées avec des données optiques. Elles étaient basées sur le contenu en galaxie des amas (3; 23; 28). Par exemple, Bautz et Morgan (1970) (3) ont établi une classification à partir de la présence d'une ou plusieurs galaxies dominantes (cD) (voir table II).

Les études en X ont débuté plus tard que celles réalisées en optique. En effet, il a fallu attendre le premiers satellite X Einstein dans les années 1970-1980 pour avoir des données dans l'X. Toutefois elles sont plus développées qu'en optique car les effets de projection sont moins importants qu'en optique. En 1993, Mohr, Fabricant et Geller (20) ont développé une technique permettant de mettre en évidence les asymétries dans la distribution de gaz. Une série d'anneaux circulaires est définie autour de la position du centre de l'amas. Le barycentre pondéré par la densité du gaz est alors calculé pour chaque anneau. Ensuite, la moyenne des barycentres est évaluée et la distance entre le centre de l'amas et la moyenne des barycentres est calculée. Cette distance est le *centroid shift*. Plusieurs méthodes basées sur la statistique de la distribution de densité du gaz ont été réalisées (7; 8; 14; 25). Par exemple, Buote et Tsai (7; 8) ont eu une approche basée sur les moments statistiques de la distribution de densité du gaz : ils se sont servi du rapport entre les moments d'ordre supérieurs à 1 et le moment d'ordre 0. Une approche par l'ellipticité de la distribution de densité du gaz a été tentée (16), mais elle ne permet pas de distinguer les amas relaxés des amas perturbés. En effet, un amas relaxé a *a priori* une forme sphérique et un amas perturbé une forme plutôt elliptique, mais à cause de possibles alignements le long de la ligne de visée, ce schéma n'est pas toujours respecté.

Plus récemment, des études similaires en optique ont été développées grâce aux relevés multi-longueurs d'onde et spectroscopiques. En 2006, Flin et Krywult ont mis en place une méthode de reconstruction de champ continu appliqué aux amas de galaxies et ont classé les amas en trois catégories : les amas unimoduax, les bimodaux et les complexes. Ramella et al. (21) ont utilisé une méthode détection de sous-structuration basées sur des procédures de kernel adaptatif (DEDICA, 2007).

En 2007, Hashimoto et al. (13) ont comparé les analyses optique et X. Ils ont montré que les morphologies optique et X peuvent être corrélées si les galaxies faibles sont prises en compte dans l'analyse optique. Pour réaliser cette comparaison, ils disposaient d'un échantillon de 100 amas pour lesquels les deux informations sont disponibles. Pour caractériser la distribution de gaz, ils ont appliqué les techniques du *centroid shift*, des moments et de la concentration dans les ellipses.

Des méthodes à une dimension dans le domaine optique existent également. Ces méthodes étudient la présence de sous-structures dans un amas à partir de la distribution de vitesses des galaxies, comme *KMM* ou *ROSTAT* (1; 4). Ces méthodes reposent essentiellement sur le caractère non gaussien de la distribution de vitesses : une distribution non gaussienne signifie que l'amas n'est pas dans un état relaxé.

Avec le relevé du SDSS, l'étude de l'environnement à grande échelle des amas *et* une étude statistique sur la morphologie des amas sotn possibles. Les données du SDSS contiennent également l'information spectroscopique d'un grand nombre de galaxies d'amas ce qui va permettre pour la première fois une analyse dynamique fine d'un grand échantillon d'amas. En outre, la méthode employée ici utilise les cartes en densité des galaxies d'un amas, comme ce qui est fait dans le domaine X, bien qu'il ne s'agisse pas d'une technique de classification développée pour les observations en X. Elle est basé sur une analyse de la distribution de vitesses puis de la distribution de densité des galaxies dans le plan du ciel (1D+2D), car une étude directement 3D est impossible[3]. Un glossaire en fin d'article décrit brièvement les logiciels et les notions que je vais utiliser.

[3] voir II.A et le problème du redshift pour caractériser une distance physique.

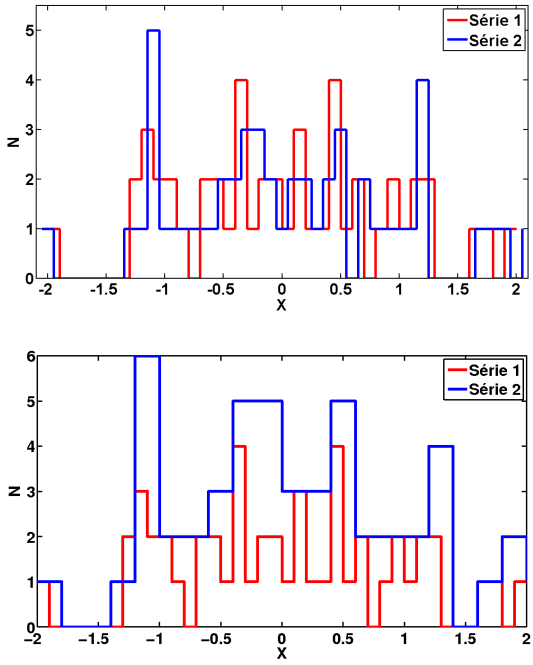


FIG. 3 Histogrammes d'une même réalisation d'une variable aléatoire gaussienne. En haut, les histogrammes ont la même classe mais lese points de départ sont différents. En bas, les points de départ des histogrammes sont les mêmes mais les classes sont différentes.

B. Construction de champs de densité

Lorsque l'on étudie un échantillon de données statistiques à une dimension, l'outil le plus simple et le plus utilisé pour représenter des données est le diagramme en histogramme. Pourtant, les histogrammes présentent au moins quatre défauts majeurs :

- l'utilisteur doit choisir *lui-même* la classe, c'est-à-dire la largeur des histogrammes ;
- l'utilisateur doit choisir également la valeur où il fait débuter ses histogrammes. ;
- la définition d'une taille caractéristique au travers de la classe ;
- le « shot noise ».

Lors de l'utilisation des histogrammes pour représenter un échantillon de données, l'utilisateur doit choisir de manière arbitraire ces deux paramètres, classe et départ. Malheureusement, différents choix de paramètres peuvent impliquer différentes représentations des histogrammes. Comme illustré sur la figure 3, selon le choix de ces deux paramètres, les histogrammes révèlent deux structures différentes pour les mêmes données. Le troisième défaut des histogrammes est de ne pas mettre en évidence les structures plus petites que la classe. Enfin, le dernier défaut des histogrammes vient du shot noise, qui est dû au nombre fini de points d'un échantillon. En effet, deux réalisations d'une même variable aléatoire donne deux histogrammes différents et il peut donc

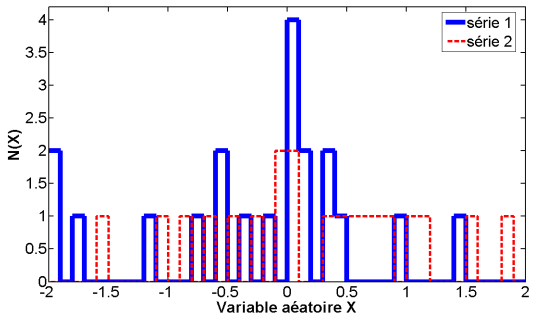


FIG. 4 Histogrammes de deux réalisations de la même variable aléatoire. Les histogrammes ont les mêmes paramètres pour les deux série. Les histogrammes de ces deux réalisations sont très différents et il est difficile de les comparer.

être délicat de comparer les deux histogrammes (voir figure 4). Ainsi, les histogrammes peuvent faire apparaître des structures qui n'ont pas de sens physique et en masquer d'autres. La confusion possible entre deux distributions de points est un autre problème de ce type de représentation. Par exemple, prenons deux lois de probabilités :

- une loi normale de moyenne 0 et d'écart-type 1 :

$$f_g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \; ;$$

- une loi de Cauchy définie de la manière suivante :

$$f_c(x) = \frac{0.85}{0.4\pi} \frac{0.4}{1 + (0.85x)^2}.$$

Si l'on trace les histogrammes d'une réalisation de ces deux distributions avec 20 points pour la loi gaussienne et 40 pour la loi de Cauchy (voir figure 5), l'allure globale est similaire alors que les alors que les lois de probabilité sous-jacentes sont différentes.

Pour s'affranchir de ces problèmes, une représentation continue pour des points distribués selon une certaine loi de probabilité a été utilisée. Une reconstruction d'un champ de densité à partir d'une série de transformations en ondelettes est réalisée par le logiciel *w1* (10), mis en œuvre à l'OCA. De cette manière, une approximation de la fonction de distribution de l'échantillon de points est obtenue. Avec cette approche, l'utilisateur n'a pas à définir de manière arbitraire les paramètres nécessaires pour visualiser ses données. La figure 6 montre les fonctions de distribution des points, issus de l'exemple précedent, calculées par *w1*. Les deux histogrammes, qui se superposaient, se révèlent être deux fonctions de distribution différentes lorsqu'elles sont visualisées en représentation continue. Un autre avantage de la représentation continue à partir d'ondelettes est la possibilité de filtrage des fréquences spatiales. Le filtrage

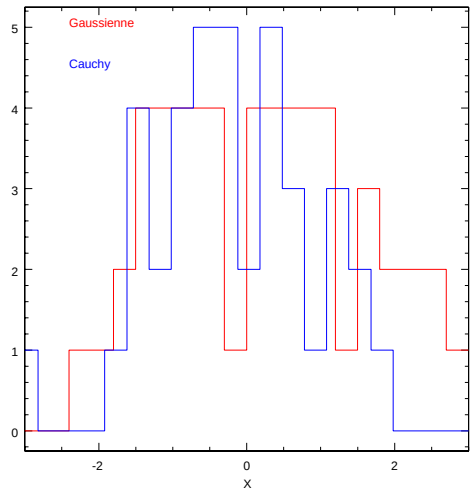


FIG. 5 Distribution de points issus d'une loi de Cauchy ($f_c(x) = (0.85/0.4\pi) \times 0.4/(1 + 0.72x^2)$) et d'une loi normale ($f_g(x) = 1/\sqrt{2\pi}e^{-x^2/2}$) avec une classe de 0.3 et un point de départ de -2.5.

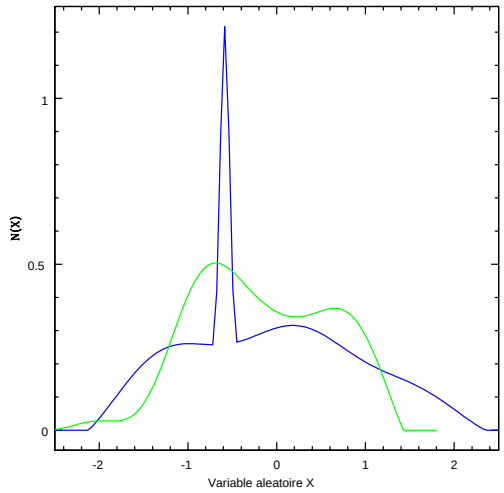


FIG. 6 Fonction de distribution pour les réalisations des lois de distribution normale (vert) et de Cauchy (bleu). La séparation des distributions est plus claire qu'avec une représentation en histogrammes (voir figure 5).

fréquentiel est intéressant car si l'œil est un bon détecteur de structures, il n'est pas quantitatif.

Pour les problèmes à deux dimensions, la même technique a été utilisée. La représentation continue à deux dimensions est plus efficace pour visualiser les structures qu'un simple nuage de points. Sur la figure 7(a), les points distribués dans le plan sont épars. Un regroupement autour de la position $(328.5, -8.7)$ est visible. La représentation continue (voir figure 7(b)) permet de mieux caractériser et quantifier les surdensités.

Le programme *wdensity* (11) est une extension de *w1* à deux dimensions et présente donc les mêmes avantages. De plus, *wdensity* offre la possibilité de ne s'intéresser qu'à certaines fréquences spatiales. Ceci est

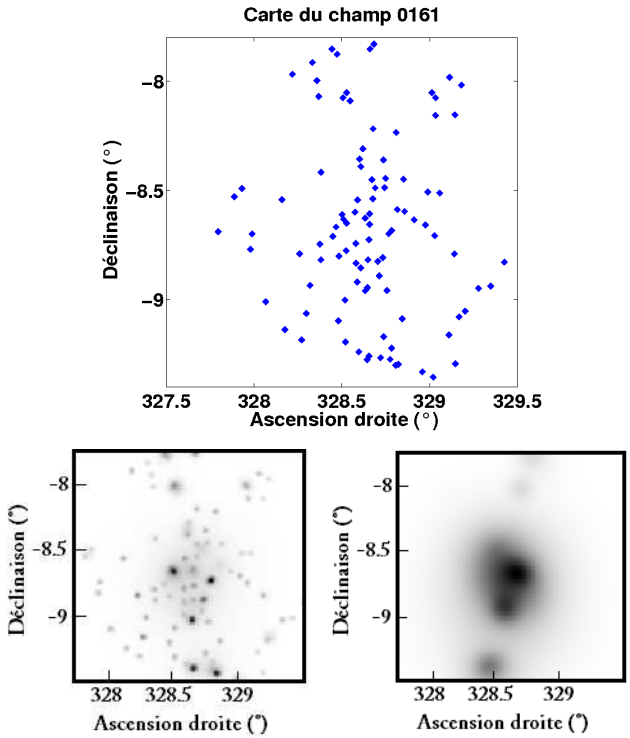


FIG. 7 Une distribution de points et sa reconstruction en champ de densité. En haut (a), les points sont représentés par des croix. En bas à gauche (b), la carte en densité est représentée, la partie noire indiquant une surdensité. La représentation continue (en bas à droite, c) permet d'isoler plus facilement les structures que les croix : les hautes fréquences spatiales ont été retirées de la reconstruction pour faire apparaître les structures à grande échelle.

particulièrement intéressant pour savoir si une structure est significative ou non à une certaine échelle spatiale en effectuant un filtrage de fréquences. Un filtrage supprimant les hautes fréquences a été appliqué à la figure 7(c). Les petites structures ont été enlevées et les structures à grande échelle sont observées, ce qui n'était pas aisément observable sur la figure 7(b), où toutes les échelles spatiales sont prises en compte.

C. Un programme de décomposition en mixtures gaussiennes : EMMIX

Pour un amas relaxé de galaxies, les vitesses des galaxies de l'amas suivent une distribution normale (24). Par contre, la distribution de vitesses des galaxies d'une structure en interaction est plus complexe. La figure 8 montre deux exemples de distributions de redshifts issues du SDSS. Ces deux distributions sont bien différentes et ne sont pas comparables à une distribution normale. Cependant, le caractère non gaussien d'une distribution dans l'espace des redshifts ne permet pas de conclure sur l'état dynamique de la structure. En effet, la distribution de redshift de deux amas relaxé proches dans le plan du

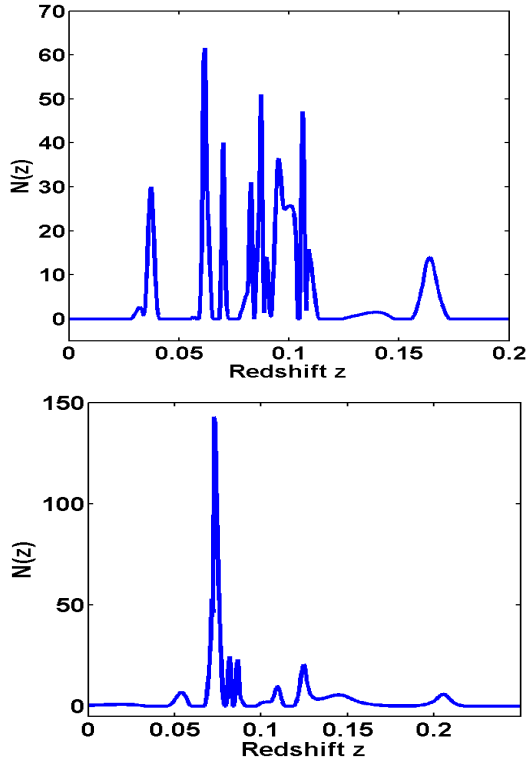


FIG. 8 Exemples réels d'écarts à la normalité de la fonction de distribtion des redshifts pour des champs issus de l'échantillon étudié.

ciel et avec des redshifts voisins peut présenter deux pics résultant de la somme de deux gaussiennes. Or classer cette distribution comme non gaussienne est erronée : un amas en interaction semble être détecté alors qu'il s'agit de deux amas relaxés. Même si la représentation des galaxies dans l'espace des redshifts ne permet pas formellement de différencier les structures en interaction des structures relaxées, cette méthode est essentielle pour révéler les *paquets*. Il faut donc utiliser un logiciel capable de séparer les différentes gaussiennes dans l'espace des redshifts. Pour cela, le logiciel *EMMIX* a été utilisé.

Le logicel *EMMIX* (18) a été développé pour être utilisé en biologie et en médecine. C'est un logiciel permettant de modéliser une distribution de points par une mixture de distributions gaussiennes. Les paramètres calculés par *EMMIX* sont la moyenne, l'écart-type et le nombre de points appartenant à la gaussienne. La détection des gaussiennes est faite à l'aide d'un algorithme Espérance-Maximisation. *EMMIX* ajuste une distribution de points par un nombre de distributions gaussiennes compris entre deux valeurs fixées au préalable par l'utilisateur. Pour chaque nombre de gaussiennes, il calcule des valeurs de tests (la vraisemblance, le Bayesian Information Criterion (BIC), l'Akaike Information Criterion (AIC)...) pour caractériser la qualité de l'ajustement correspondant à ce nombre de gaussiennes. L'utilisateur n'a plus qu'à regarder le nombre de gaussiennes qui minimise les critères pour connaître le nombre de distributions gaussiennes présentes dans son échantillon de points.

Le point fort d'*EMMIX* est de ne pas nécessiter de paramètres d'entrés fournis par l'utilisateur. Les autres logiciels traditionnellement utilisés en astrophysique (*KMM* (1) ou *ROSTAT* (4)) étudiant les distributions normales ont besoin de savoir à l'avance le nombre de gaussiennes et une estimation des moyennes ou ne travaillent que sur une seule gaussienne. Quand on doit traiter un seul ensemble de points, un champ contenant un amas par exemple, il est possible d'évaluer le nombre de gaussiennes et leur moyenne à l'œil nu pour ensuite donner ces informations au logiciel d'ajustement. Mais une telle approche est inappropriée lorsque l'on veut traiter un grand nombre d'amas, comme dans le cas du SDSS. C'est là que l'utilisation d'*EMMIX* devient pertinente : la seule information à donner à *EMMIX* est l'intervalle du nombre de gaussiennes (par exemple, 2 à 4 gaussiennes). *EMMIX* est aussi, voire plus, performant que les autres logiciels. En effet, l'utilisateur doit seulement indiquer à*EMMIX* le nombre minimal et le nombre maximal de gaussiennes entre lesquels il doit chercher et non pas le nombre exact de gaussiennes qu'il doit ensuite ajuster. En revanche, *EMMIX* nécessite un temps de calcul plus important puisqu'il utilise la technique du bootstrap, très performante, pour estimer de manière plus efficace les valeurs qui serviront à déterminer le mixture optimale de distributions gaussiennes. *EMMIX* n'a jamais été utilisé dans le cadre des amas et de la distribution de vitesses, car il n'est sans doute pas connu de la communauté.

1. Détection des gaussiennes

Pour déterminer la pertinence des résultats d'*EMMIX* vis-à-vis de données semblables à celles du SDSS, j'ai donc effectué plusieurs séries de tests. De plus, *EMMIX* calculant plusieurs tests statistiques, cela permettra de savoir quel test utiliser pour avoir des détections les plus pertinentes.

La distribution de redshifts d'un amas relaxé suit une loi gaussienne. Il faut donc savoir si *EMMIX* est capable de détecter correctement une distribution gaussienne. De plus, *EMMIX* a été utilisé sur les paquets dont les ailes sont tronquées. En effet, les paquets sont définis à partir des zéros de la fonction de distribution calculée avec *w1*. Or *w1* a un seuil de détection de trois galaxies par pas de résolution ce qui implique que les zéros de la fonction de distribution ne sont pas de vrais zéros : une troncature des ailes de la distribution peut se produire lors du regroupement des galaxies par paquets. Un test sur des distributions gaussiennes tronquées est donc nécessaire. En outre, les amas en cours d'accrétion de galaxies le long d'un filament présentent une distribution de redshifts de la forme : une gaussienne de petit écart-type et avec un grand nombre de points superposée à une gaussienne de très grand écart-type et avec peu de points. Cette superposition se traduit par un surpeuplement des ailes de la

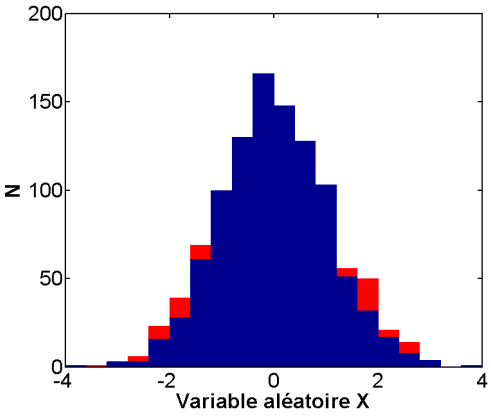


FIG. 9 Exemple de surpeuplement des ailes d'une distribution de points gaussienne. En bleu, 1000 points tirés dans une gaussienne. En rouge, les points rajoutés par surpeuplement au-delà de $1.5\,\sigma$

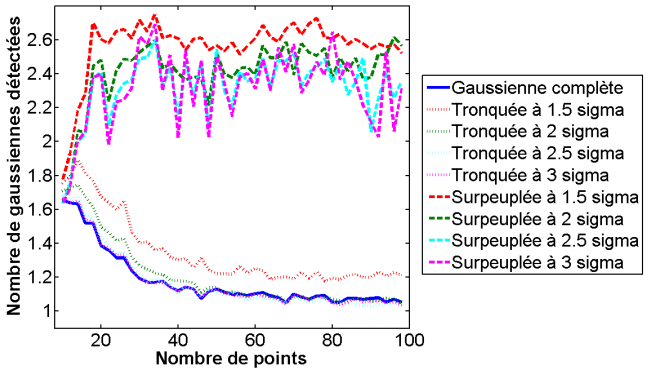


FIG. 10 Résultats des tests *EMMIX* sur les distributions gaussiennes complète (bleu), tronquées (en pointillés) et surpeuplées (en tirets). Le rouge correspond à des surpeuplements ou à des troncatures effectués à $1.5\,\sigma$, le vert à $2\,\sigma$, le cyan à $2.5\,\sigma$ et le magenta à $3\,\sigma$. La troncature des ailes n'a pas beaucoup d'influence sur la détection d'une gaussienne. Par contre, le surpeuplement des ailes d'une distribution gaussienne entraîne la détection de plusieurs gaussiennes (entre 2 et 3).

gaussienne (voir figure 9).

En général, un amas en formation peut avoir plusieurs sous-structures. Un cas typique est la somme de plusieurs gaussiennes dans l'espace des redshifts. Il est donc important de connaître la réaction d'*EMMIX* face à une distribution contenant des gaussiennes identiques mais de moyenne différente.

Un autre test intéressant est le cas de deux gaussiennes d'écarts-type et de nombre de points différents. Ce mélange est plus général que le précédent.

a. Séries de tests à une gaussienne

Cette série de tests vise à appréhender comment *EMMIX* traite des points distribués selon une loi normale dont les paramètres sont : une moyenne $m = 1$, un écart-type $\sigma = 1$ et le nombre de points N varie de 10 à 100 de 2 en 2. Il est aussi essentiel de savoir quelle détection *EMMIX* va effectuer en présence d'une distribution normale dont les ailes sont soit tronquées soit surpeuplées. Pour tronquer (respectivement surpeupler) les ailes, on enlève (respectivement rajoute) les points au-delà d'un certain nombre d'écart-types. Pour surpeupler les ailes, j'ai refait un tirage aléatoire de la même loi normale, dont je n'ai gardé que les points dans les ailes et j'ai rajouté ces points aux autres. Ici, les troncatures et les surpeuplements ont été pratiqués à $1.5\,\sigma$, à $2\,\sigma$, à $2.5\,\sigma$ et à $3\,\sigma$.

La figure 10 montre les résultats de ces tests. Pour une gaussienne complète (bleu), la détection d'une seule gaussienne est très performante pour $N > 30$. Avec moins de points, *EMMIX* peut déteter plusieurs gaussiennes. La troncature des ailes n'influe pas de manière effective la détection d'une seule gaussienne. Cependant, le surpeuplement des ailes d'une distribution gaussienne entraîne la détection de plusieurs gaussiennes. Plus le surpeuplement augmente, plus le nombre de gaussiennes détectées augmente : la distribution étudiée s'écarte de plus en plus d'une distribution gaussienne. Ces tests sur des distributions de points typiques de notre échantillon permettent de dire si les détections que va faire *EMMIX* sur nos données sont significatives.

Pour 20 objets distribués selon une loi gaussienne, *EMMIX* détecte au plus, en moyenne, 1.4 distributions gaussiennes. Pour $N > 20$, le nombre de gaussiennes détectées diminue lorsque N augmente. Les distributions tronquées présentent le même « profil » de détection, seule la troncature à 1.5σ influe de manière notable sur la détection par rapport au cas où la distribution est complète. Les erreurs sur l'estimation des paramètres sont les suivantes : 8 % pour la moyenne et 13 % pour l'écart-type. Concernant le surpeuplement, avec $N > 20$, *EMMIX* trouve entre deux et trois gaussiennes. En moyenne, lorsqu'*EMMIX* détecte deux gaussiennes, elles ont sensiblement la même moyenne (1.02 ± 0.15) près) mais les écarts-type sont différents :

1. une gaussienne a un écart-type proche de 1 à 20 % près et contient environ 80 % des points de l'échantillon ;
2. l'autre a un écart-type plus grand, de l'ordre 2 et comprenant environ 20 % des points.

Cette caractérisation est très intéressante pour l'étude des amas car une répartition symétrique des galaxies selon ce schéma peut s'interprêter de la manière suivante : la gaussienne principale correspond à un amas de galaxies alors que la gaussienne avec le grand écart-type correspond à un filaments ou un feuillet de galaxies et l'amas est en train d'accréter des galaxies le long de cette structure.

En moyenne, lorsqu'*EMMIX* détecte trois gaussienne dans une distribution gaussienne surpeuplée, elles ont les caractéristiques suivantes :

1. la gaussienne avec le plus de points a une moyenne 1.09 ± 0.21, un écart-type 0.98 ± 0.23 et un nombre de l'ordre de 60 % ;
2. une gaussienne de moyenne 0.21 ± 0.47, un écart-type de l'ordre de 1 et contient environ 20 % des points ;
3. une gaussienne de moyenne 2.04 ± 0.51, un écart-type de l'ordre de 1 et contient environ 20 % des points.

Là encore, cette caractérisation est intéressante car une configuration de ce type représente une distribution réelle de redshifts. Il s'agit d'un amas principal et deux autres structures d'avant ou d'arrière plan dans l'espace des redshifts.

b. Séries de tests à deux gaussiennes identiques

Pour les tests sur des distributions de points gaussiennes identiques mais translatées, j'ai réalisé des tirages de nombres aléatoires distribués selon deux lois normales. Les deux lois ont le même écart-type $\sigma = 1$ et on tire le même nombre de points N dans chacune des gaussiennes. Les deux lois ont une moyenne différente ; l'écart entre les deux vaut Δ. Les tests effectués portent sur le nombre de points générés N et sur la distance Δ entre les deux gaussiennes. Pour un couple (N, Δ), j'ai réalisé cent échantillons de N points, afin de réduire les erreurs, dues en particulier à une réalisation particulière de la variable aléatoire.

Mon point de départ était deux gaussiennes de cinq points chacunes avec une séparation $\Delta = 1.5\,\sigma = 1.5$. Ensuite, j'ai augmenté le nombre de points de cinq en cinq jusqu'à cent points et la séparation de $0.1\,\sigma$ en $0.1\,\sigma$ jusqu'à atteindre $4\,\sigma$. Pour un couple (N, Δ) donné, cent tirages de nombres aléatoires ont été réalisés et j'ai utilisé *EMMIX* pour savoir le nombre de gaussiennes qu'il détectait. Pour un couple (N, Δ), ces cent tirages ont permis de calculer le nombre moyen de gaussiennes détectées ainsi que la probabilité de détection de deux gaussiennes (voir figure 11). Il apparaît que la détection des gaussiennes est meilleure lorsque le nombre de points augmente et lorsque la séparation entre les deux gaussiennes est élevée.

La détection pertinente par *EMMIX* de deux gaussiennes nécessite une séparation au moins supérieure à $3\,\sigma$. Au delà de $3\,\sigma$, l'influence du nombre de points diminue avec la séparation.

c. Séries de tests à deux gaussiennes différentes

Dans cette série de tests, je me suis intéressé à la détection de deux gaussiennes différentes mélangées : la première a une moyenne $m_1 = 2$, un écart-type $\sigma_1 = 1$ et n_1 points ; la seconde a une moyenne m_2, un écart-type $\sigma_2 = 0.2$ et n_2 points. n_1 et n_2 sont tels que $n_1 = 20 \times n_2$. On fait varier le nombre de points n_2 de 1 en 1, à partir de $n_2 = 1$ et jusqu'à $n_2 = 20$. La séparation entre les deux gaussiennes vaut $\Delta = m_2 - m_1$. Initialement, on part de $\Delta = 0$, cas où les deux gaussiennes sont superposées, et on augmente Δ de 0.1 en 0.1 jusqu'à $\Delta = 5$. Pour

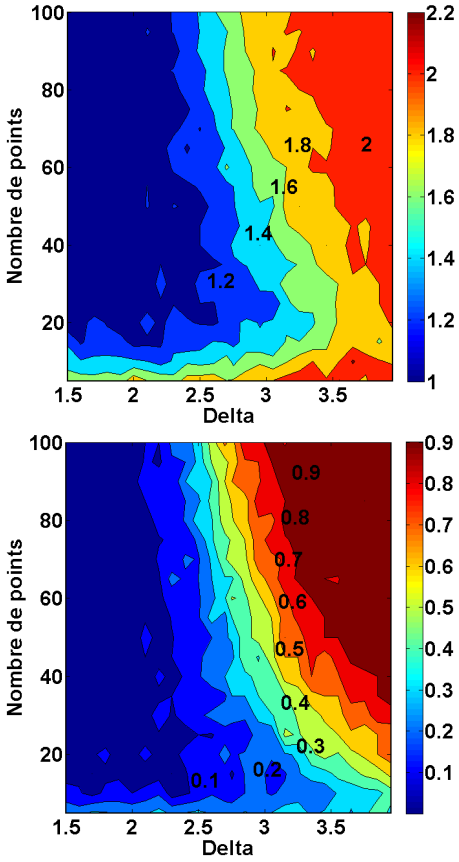


FIG. 11 Réultats des tests *EMMIX* pour la détection de deux gaussiennes de mêmes nombre de points et écart-type mais de moyennes distantes de Δ. A gauche, est représenté le nombre moyen de gaussiennes détectées et à droite la probabilité de détection de deux gaussiennes.

un couple (n_2, Δ) donné, on fait cent réalisations de la distribution de points correspondante, toujours pour minimiser les erreurs dues aux fluctuations de réalisations particulières.

Les résultats de détection de deux gaussiennes aux paramètres différents sont très satisfaisants : la probabilité de détection est comprise entre 0.6 et 0.9. La détection de deux gaussiennes dépend faiblement de la séparation et varie fortement avec le nombre de points. Cette conclusion est très utile puisque quelque soit le nombre de points, la probabilité de détection des deux gaussiennes est toujours supérieure à 0.6. En outre, si le nombre de points est assez élevé, nous savons qu'*EMMIX* détectera parfaitement les gaussiennes quelque soit leur séparation. *EMMIX* trouve que dans 89 % des cas la gaussienne la plus peuplée est celle dont la moyenne est la plus proche de $m - 1 = 0$

Ces tests ont également permis de déterminer le meilleur test calculé par *EMMIX*. Il s'agit du *Bayesian Information Criterion* ou BIC : la probabilité de détecter le bon nombre de gaussiennes minimisant ce critère est plus importante que celle des autres tests dans les différents tests réalisés. Les résultats de cette partie

sont ceux obtenues avec le BIC.

2. Affectation des points

Après avoir calculé pour un nombre de gaussiennes donné les différents paramètres, *EMMIX* distribue les points de l'échantillon dans les différentes gaussiennes : selon le nombre de gaussiennes détectées *EMMIX* répartira chaque point dans une des gaussiennes. Cette affectation est une étape particulièrement sensible de l'étude puisqu'*EMMIX* va classer les galaxies dans une certaine structure plutôt que dans une autre. Comme *EMMIX* n'est utilisé qu'à une dimension, la répartition des galaxies dans les sous-structures est aléatoire dans les zones de recouvrements de plusieurs gaussiennes. Quand la superposition des sous-structures est modérée, l'affectation des galaxies est correcte et n'influe pas sur l'analyse à 2D. Mais si les sous-structures sont trop proches, alors il faut regarder plus précisemment la répartition des galaxies dans le plan du ciel. La figure 12 illsutre un cas où l'affectation des galaxies n'est pas satisfaisante. Sur la figure 12(a), le paquet a été divisé en deux structures gaussiennes par *EMMIX* et ensuite les galaxies ont été réparties dans ces deux structures. Dans le plan du ciel, figure 12(b), deux structures sont visibles : un amas sur la droite et une structure de type feuillet-filament sur la gauche. La figure 12(c) représente la fonction de distribution de redshifts des galaxies avec une ascension droite $\alpha > 207.6°$ (partie droite du champ) et celle pour les galaxies avec $\alpha < 207.6°$ (partie gauche du champ). Ces deux fonctions de distribution tirées des observations sont les mêmes que celles détectées par *EMMIX*. L'affectation des galaxies n'est pas optimale mais cela n'influe pas sur la détection des deux structures ni sur l'évaluation des paramètres des deux gaussiennes qui sont très satisfaisantes.

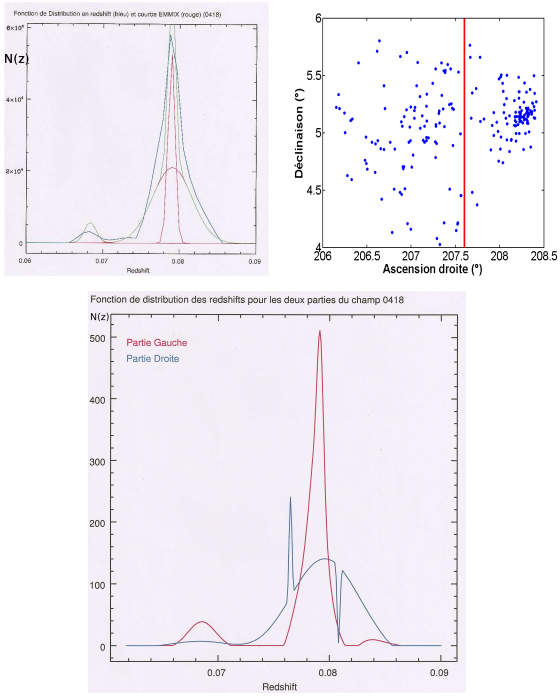


FIG. 12 Illustration de problèmes liés à l'affectation des points par *EMMIX*. La figure (a) représente la division d'un paquet en trois structures gaussiennes par *EMMIX*. Les deux gaussiennes les plus importantes se situent au même redshift, la troisième étant une petit groupe de galaxie d'avant-plan. La figure (b) montre la répartition des galaxies dans le plan du ciel : les différentes couleurs illsutrent les différentes affectations. L'axes des abscisses correspond à l'ascension droite (en degrés) et l'axes des ordonnées représente la déclinaison (en degrés). Les galaxies des deux sous-structures peuvent se séparer en deux ensembles dans le plan du ciel : les galaxies avec une ascension droite $\alpha < 207.5°$ formant une structure de type feuillet-filament et celles avec $\alpha > 207.5°$ formant un structure de type amas. Les fonctions de distribution des redshifts associées à ces deux ensembles sont représentées sur la figure (c) : ce sont les mêmes que les gaussiennes extraites par *EMMIX*.

D. Extraction d'un catalogue de structures

Dans chaque champ étudié, les galaxies peuvent être membres de l'amas principal, ou bien appartenir à d'autres structures d'avant ou d'arrière plan. Les objets d'étude étant les amas de galaxies, il faut donc pouvoir enlever les autres structures de la classification. De plus, une structure gaussienne en z peut être bimodale dans le plan du ciel. Pour le déterminer, il faut analyser les concentrations de densité dans le plan du ciel qui nous permettront de différencier les différentes structures (c.f. II.C) détectées dans l'espace des redshifts. En effet, les groupes et les amas de galaxies sont des structures piquées dans le plan du ciel. Pour faire le tri de ces structures, j'ai utilisé le logiciel *SExtractor* (6) qui construit un catalogue d'objets à partir d'une image. *SExtractor* ajuste les objets par des ellipses et fournit les paramètres de forme des objets comme le demi-grand axe, l'angle d'orientation du demi-grand axe, le flux total dans l'ellipse. Le catalogue *SExtractor* d'une carte en densité va donc permettre de distinguer les structures.

E. Algorithme d'identification et de classification des amas

La méthode de caractérisation et de classification des amas que j'ai développée est la suivante pour chaque champ.

1. La définition de la magnitudede coupure de l'échantillon est déterminée par le meilleur taux e complétude pout tous les champs.
2. La fonction de distribution de redshifts est calculée avec *w1*, ce qui me permet de définir des paquets : un paquet est un ensemble de galaxies ayant un redshift compris entre deux zéros consécutifs de la fonction de distribution (la figure 13).
3. Un filtre sur le nombre et la concentration des galaxies dans le paquet est appliquée : il faut au moins 20 galaxies et la fonction de distribution

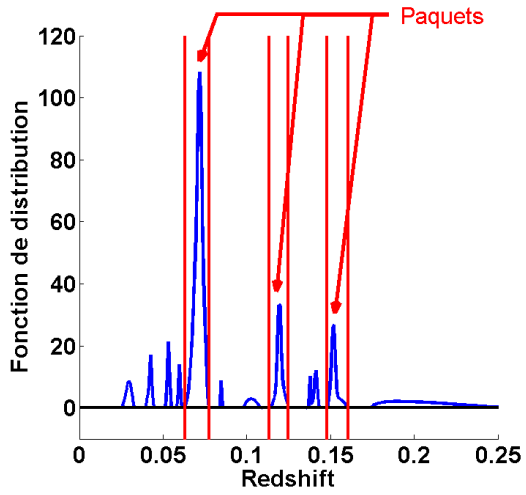


FIG. 13 Exemple de fonctions de distribution dans l'espace des redshifts avec les limites des *paquets* matérialisés par les traits verticaux rouges.

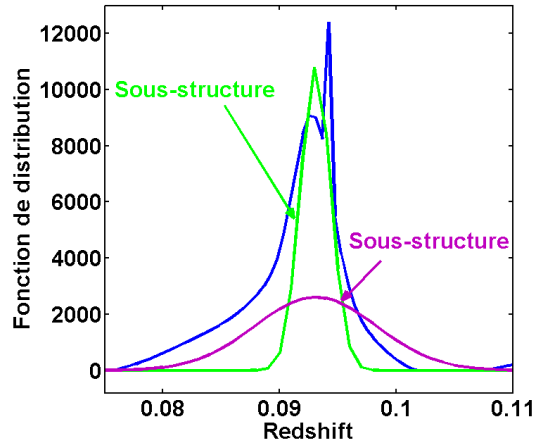


FIG. 14 Exemples de deux *sous-structures* présentes dans un paquet.

doit dépasser 25 dans les paquets pour qu'ils soient conservés. Ces paramètres constituent un premier choix de travail que j'ai effectué.

4. Chaque paquet est décomposé en une mixture de gaussiennes avec *EMMIX*.
5. Comme les objets d'étude sont les amas de galaxies, ils contiennent beaucoup de galaxies et sont massifs, un filtre sur ces deux quantités doit donc être appliqué. Une sous-structure doit contenir au moins 15 galaxies et avoir une dispersion des vitesses $\sigma_v > 100\text{km.s}^{-1}$. Si une sous-structure ne remplit pas l'un de ces deux critères alors, on regarde le partitionnement obtenu avec une sous-structure de moins, jusqu'à ce que les deux critères soient satisfaits. Un exemple de sous-structure est montré sur la figure 14).
6. La carte en densité dans le plan du ciel pour chaque sous-structure est construite avec *wdensity* en ne considérant que des basses fréquences. Les cartes en densité ont une taille de 128×128 pixels, correspondant à 3×3 Mpc.
7. *SExtractor* (6) extrait les sources en densité de ces cartes pour pouvoir différencier les structures de type amas aux structures de type feuillets.
8. Un filtre basé sur la largeur à mi-hauteur des sources est appliqué pour éliminer les feuillets. En effet, les structures de type feuillet sont plus étendues que les amas. Un amas a un taille typique de 0.5 Mpc ce qui correspond à environ 21 pixels. Les feuillets doivent donc avoir une taille typique supérieure à 25 pixels, ce qui équivaut à une largeur à mi-hauteur $\Lambda > 2 \times 25 = 50$. Si la source de flux maximal en densité a une largeur à mi-hauteur $\Lambda > 50$ alors la sous-structure est classée comme feuillet, sinon elle est classée comme amas.

Les amas sont ainsi caractérisés de manière plus fine que ne l'avait fait l'algorithme du C4 et nous pouvons classer toutes les sous-structures selon des critères morphologiques. Une classification, quelle que soit sa nature, nécessite en premier lieu de définir des catégories générales et d'y classer les objets. Ensuite, la classification doit être affinée avec des sous-catégories plus précises. Cette itération est répétée jusqu'à obtenir le degré de classification souhaité. La première classification morphologique de l'échantillon d'amas distingue trois groupes :

1. les amas potentiellement en interaction : l'amas est composé de plusieurs sources avec un nombre important de galaxies dans chacune d'elles ;
2. les amas potentiellement à l'équilibre : la distribution de redshifts est gaussienne et la répartition des galaxies dans la plan du ciel est symétrique ;
3. les amas potentiellement en cours d'accrétion : un amas principal avec des fluctuations en densité, traduisant un état dynamique non relaxé.

Le critère employé ici pour classer les amas selon ce schéma est basé sur le rapport entre les deux plus grands flux maximaux des sources : $R = f_{max_1}/f_{max_2}$, où f_{max_1} est le plus grand flux maximal et f_{max_2} est le deuxième plus grand flux maximal. Si $R < 2$ alors les deux sources principales de densité de cet amas sont équivalentes en nombre de galaxies. Cet amas est classé comme un amas en cours d'interaction. Si le rapport est plus grand que 5 alors l'amas possède une composante principale qui domine les autres : l'amas est classé comme relaxé. Enfin si le rapport est compris entre 2 et 5 alors l'amas n'est ni relaxé ni en interaction, il est classé dans la catégorie accrétion : soit il accrète des galaxies individuelles ou en couple, soit il a subi une fusion plus massive récemment. Les différents intervalles définis ici sont également un premier choix de travail. Le dernier filtre appliqué est un filtre de masse pour ne s'intéresser qu'aux amas. Seules les structures avec une masse $M > 10^{13} M_\odot$ sont conservées.

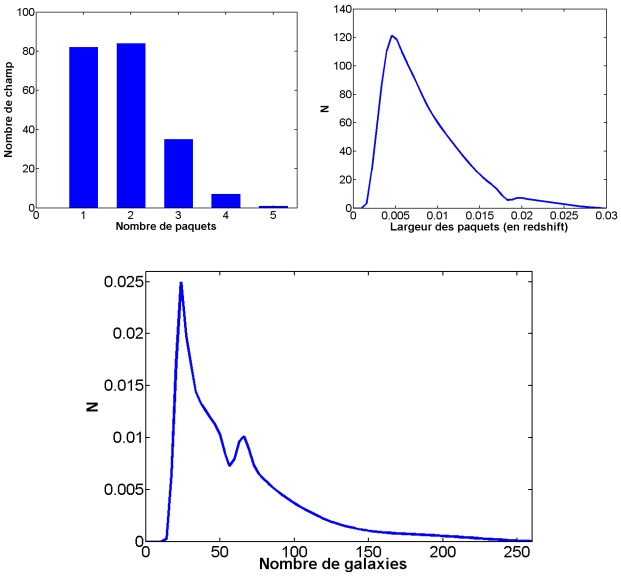


FIG. 15 La figure de gauche représente l'histogramme du nombre de paquets par champ. 81 champs ne possède qu'un seul paquet, 84 champs possèdent deux paquets, 35 champs en possèdent 3, 7 champ en ont 4 et un champ possède 5 paquets. La figure de droite représente la fonction de distribution des largeurs en redshift des paquets. Enfin, la figure du bas représente la fonction de distribution du nombre de galaxies par paquet.

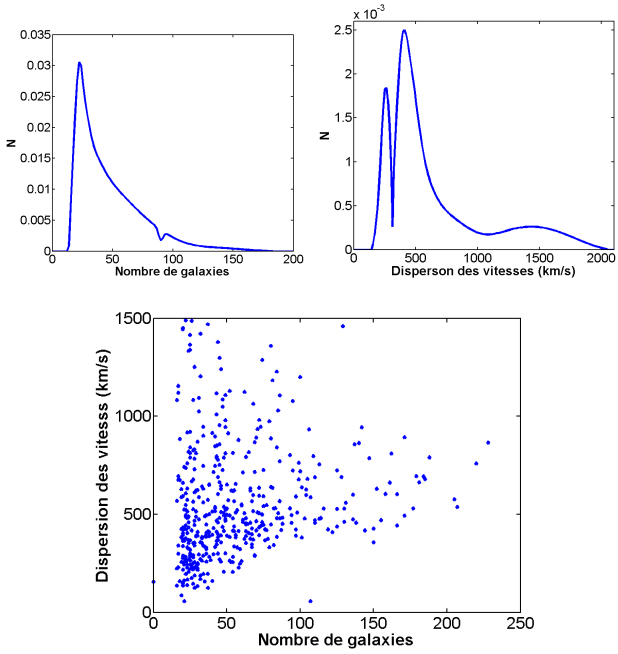


FIG. 16 Statistiques générales sur les sous-structures. La figure a représente la fonction de distribution du nombre de galaxies dans les sous-structures *EMMIX*. La figure b représente la fonction de distribution des dispersions des vitesses des sous-strucures *EMMIX*. La figure c représente les sous-structures *EMMIX* dans un diagramme nombre de galaxies - dispersions des vitesses. Plus le nombre de galaxies augmente, plus la dispersion des vitesses (donc la masse) augmente.

V. ANALYSE DU CATALOGUE

A. Identification des amas

L'analyse du sous-échantillon d'amas du C4 a été faite avec la méthode décrite dans la partie IV.B. Dans cette partie, je vais détailler les résultats de chaque étape de la méthode appliquée à l'échantillon.

Le premier pas de l'analyse du catalogue consiste à dégager les structures en ne tenant compte que du redshift des galaxies. Il faut détecter les structures contenant plus de vingt galaxies et ayant une densité suffisamment importante. Des 211 champs que j'avais à ma disposition, 389 paquets ont été décelés. La figure 15 représente la distribution du nombre de paquets et la distribution de largeurs en redshift des paquets.

L'étape suivante consiste à analyser les *sous-structures* présentes dans chaque paquet. Des 389 paquets, *EMMIX* a extrait 774 sous-structures. Ensuite, le filtre sur le nombre de galaxies présentes dans la sous-structure a été appliquée. Toutes les sous-structures avec moins de quinze galaxies ont été redistribuées. Puis, le filtre sur la dispersion des vitesses a été utilisé, mais il s'est avéré que toutes les sous-structures avec $\sigma_v < 100\text{km.s}^{-1}$ ont été filtrées avec le nombre de galaxies. Sur la figure 16, j'ai représenté des statistiques générales sur les sous-structures.

Le troisième pas de l'identification des structures est l'analyse des cartes en densité. Les structures plus étendues que les amas sont retirées de l'analyse. Des 774 sous-structures extraites par *EMMIX*, il reste 496 sous-structures qui sont potentiellement des amas.

Le quatrième pas consiste à enlevées toutes les structures qui ne sont pas suffisamment massives pour être des amas. Seules les sous-structures avec une masse supérieure à $10^{13}M_\odot$ sont considérées comme étant des amas. Ainsi, 218 amas ont été extrait des 211 champs de départ.

a. Elimination des feuillets

Pour savoir si une sous-structure est significative, *SExtractor* a été employé sur toutes les cartes en densité à l'échelle qui nous intéresse. J'ai décidé d'utiliser la largeur à mi-hauteur (FWHM) afin de déterminer si une structure est significative ou non. La valeur seuil est définie par rapport à la taille typique d'un amas (21 pixels) et vaut 50 pixels. Les sous-structures avec une largeur à mi-hauteur $\Lambda > 50$ pixels correspondent à des structures plus étendues que les amas de galaxies : il s'agit donc de structures de type filament ou feuillet ; celles avec $\Lambda < 50$ sont de type amas. Enfin, le filtre sur la masse est appliqué : les structures avec une masse $M < 10^{13}M_\odot$ sont éliminés car elles ne sont pas assez massives pour être des amas.

Avec ces critères, un catalogue de 218 structures de type amas a donc été extrait à partir des données du

SDSS. On peut maintenant passer à l'étude des amas et à leur classification morphologique.

B. Classification morphologique

Pour classer les amas dans les différentes classes définies dans la partie IV.E, je me suis intéressé aux deux structures contenant le plus de galaxies dans chaque amas. En tenant compte des grandes échelles spatiales, *SExtractor* a établi un catalogue d'objets. Pour des amas en interaction, les deux objets avec le plus de flux en densité ont des flux comparables (avec au plus un facteur 2 d'écart). Les amas en cours d'accrétion ont un objet dominant, mais ils ont également d'autres objets avec des densités plus faibles : le rapport des flux maximaux est compris entre 2 et 5. Enfin, les amas relaxés ont uniquement un objet avec une forte densité, le rapport des flux maximaux est supérieur à 5. Il s'agit ici de résultats préliminaires en conservant toutes les structures avec une masse supérieure à $10^{13} M_{\odot}$.

Les proportions d'amas dans les différentes classes sont les suivantes :

1. 53 amas sont potentiellement en interaction, soit 24 % des amas de l'échantillon ;
2. 97 amas sont relaxés, soit 45 % des amas de l'échantillon ;
3. 68 amas sont en cours d'accrétion, soit 31 % des amas de l'échantillon.

La figure 17 représente la distribution des vitesses et la carte en densité d'un amas de chaque classe.

Cette classification met en évidence que 55 % des amas présentent des sous-structures. Les résultats d'autres classifications, aussi bien dans l'optique que dans l'X, sont disparates. Dans l'X, Jones et Forman (14) ont étudié 368 amas proches ($z < 0.2$) pour trouver une fraction de 40 % d'amas présentant une sous-structuration. Schuecker et al. (25), avec plusieurs tests statistiques, ont trouvé que 52 % des 470 amas de leur échantillon étaient sous-structurés. Dans l'optique, Burgett et al (**?**) ont étudié 25 amas par méthode basée sur les moments à une, deux et trois dimensions. Il en est ressorti que 21 amas (84 %) présentent une sous-structuration. Flin et Krywult (12) ont utilisé une méthode utilisant des ondelettes sur 183 amas pour trouver 34 % d'amas avec des sous-structures. Ramella et al. (21) ont utilisé une procédure à kernel adaptatif (DEDICA) sur 77 amas proches ($0.04 < z < 0.07$) et obtiennent un taux de présence de sous-structures de 73 %.

Le taux de sous-structuration dérivé de la classification déceloppée ici est quelque peu différent des résultats obtenus dans l'optique (34 %, 73 % et 84 %) mais se rapproche plus de ceux obtenus dans l'X (40 % et 52 %). En effet, l'utilisation de cartes en densité de la distribution des galaxies est de même nature que les cartes en densité du gaz.

Il s'agit ici d'un résultat préliminaire car l'étude n'est pas achevée. En effet, les sous-structures sont issues de

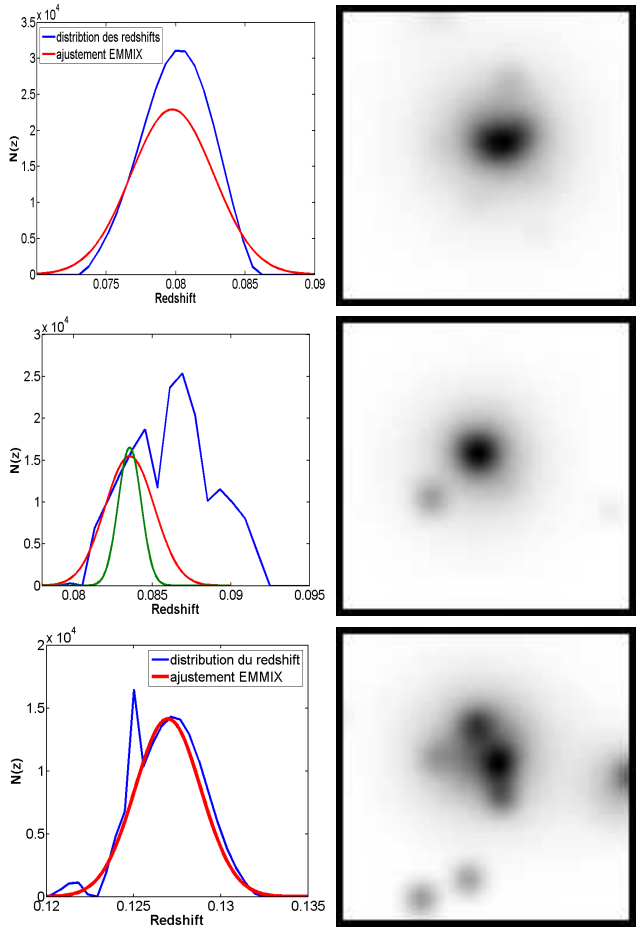


FIG. 17 Exemple des trois classes d'amas. Pour chaque exemple, la fonction de distribution de redshifts est représentée à gauche et la carte en densité dans le plan du ciel à droite. En haut, un amas relaxé : la distribution de redshifts est gaussienne et il y a une seule composante principale dans l'amas. Au milieu, un amas en interaction : la distribution des redshifts présentent un écart à la normalité et une composante principale et une secondaire sont visibles sur la carte en densité. En bas, un amas de galaxies potentiellement en interaction : la distribution de redshifts est gaussienne mais l'amas présente une structure complexe dans le plan du ciel.

paquets, elles peuvent donc interagir entre elles. De plus, l'étude est conditionnée par les choix de paramètres qui sont à validés, notamment en regardant la stabilité des résultats sur la classification lorsque ces paramètres varient.

VI. CONCLUSION

Une classification morphologiques des 218 amas de galaxies du SDSS a été réalisée de manière automatique après la mise en place d'une technique de détection et de caractérisation des amas du relevé. Les amas ont été classés en trois grands groupes : les amas potentielle-

ment en interaction (24 %), les amas en cours d'accrétion (31 %) et les amas relaxés (45 %).

Cette classification est essentielle pour pouvoir se servir des amas comme sondes cosmologiques et étudier de manière précise la physique des amas de galaxies (formation stellaire, mécanisme de fusion...). De plus, les amas du SDSS ont servi d'étalon à la classification morphologique, qui va être utilisée sur les amas plus distants du CFHT Legacy Survey. Cette analyse optique doit être complétée en utilisant l'information dans tout le spectre électromagnétique pour avoir accès à l'ensemble l'information disponible.

VII. PERSPECTIVES

A. Comparaison avec les observations en X

Trente amas de l'échantillon d'étude ont été observés dans le domaine X. Les galaxies et le gaz ayant des comportements différents (notamment par rapport aux collisions), il serait intéressant de comparer la morphologie des amas observée dans l'optique à celle observée en X, comme l'a fait Hashimoto (13).

B. Evolution de la morphologie avec le redshift

Toutes les galaxies de notre échantillon ont un $z < 0.3$. Les objets que l'on observe se situent donc à environ moins de 3 milliards d'années-lumière : ils se situent dans l'univers local. Les conditions qui règnent dans leur environnement sont semblables aux nôtres. Il serait donc très intéressant de regarder des amas plus distants donc vivants dans un univers plus jeune que le nôtre. Dans l'univers jeune, les conditions physiques sont différentes de celles d'aujourd'hui et on s'attend à avoir des structures de galaxies de plus petites tailles mais qui subissent plus de phénomènes de fusion.

C. Etat dynamique des amas et formation stellaire de leurs galaxies

Des simulations numériques récentes, aussi bien que des travaux observationnels, ont montré que la coalescence entre amas peut fortement affecter les propriétés physiques internes de leurs galaxies, en particulier augmenter leur taux de formation stellaire (5; 11). Ce résultat est quand même fortement débattu (2). En outre, le taux de formation stellaire des galaxies d'amas pourrait dépendre non seulement de l'état dynamique des amas observés, mais aussi de la phase du processus de coalescence que l'on observe (17). Comprendre l'effet de la coalescence entre amas sur la formation stellaire des galaxies s'insere dans le long débat sur l'origine physique de l'évolution observée des galaxies en fonction de la densité de leur environnement (9).

Grâce à mon travail de stage, je dispose d'un catalogue d'amas classés sur la base de leur morphologie. De l'autre côté, nous pouvons mesurer le taux de formation stellaire des galaxies membres de ces amas à partir des spectres de la base de donées du SDSS. Les galaxies avec formation stellaire en cours sont en effet caractérisées par la présence de raies d'émission dans leurs spectres optiques. Nous envisageons ainsi d'étendre l'étude du lien entre formation stellaire et état dynamique des amas à travers l'analyse d'un échantillon statistiquement significatif d'amas de galaxies.

Remerciements

Je tiens à remercier tout particulièrement Christophe Benoist pour m'avoir donné la chance de pouvoir intégrer l'équipe de Cosmologie et d'Astrophysique Extragalactique du Laboratoire Cassiopée. Je le remercie également pour tout le temps qu'il m'a accordé tout au long de ce stage. Je remercie Chiara Ferrari, Sophie Maurogordato, Alberto Cappi et Jean-Christophe Mauduit, membres de l'équipe, pour leur collaboration à mon travail.

Je tiens également à remercier toutes les personnes de l'Observatoire pour leur accueil chaleureux.

Enfin, je remercie les enseignants du Master Ωmega de m'avoir suivi et conseillé tout au long de ce stage et je remercie les personnes qui vont lire ce travail pour l'attention qu'elles y porteront.

Glossaire

- *w1* : logiciel de reconstruction de fonction de distribution à une dimension d'un échantillon de points, basée sur une série de transformations en ondelettes.
- *paquet* : ensemble de galaxies d'un champ ayant des redshifts compris entre deux zéros consécutifs de la fonction de distribution de redshifts du champ correspondant.
- *EMMIX* : logiciel d'extraction de distribution gaussienne dans un échantillon de points de dimension quelconque, il n'est utilisé quà une seule dimension dans cette étude.
- *sous-structure* : composante gaussienne extraite par *EMMIX* d'un *paquet*.
- *wdensity* : logiciel de reconstruction de champ de densité à deux dimensions, permet de faire du filtrage d'échelles spatiales : l'utilisateur choisit la taille (une puissance de 2) en pixels de la carte qu'il veut obtenir et *wdensity* permet de filtrer certaines échelles spatiales définies en puissance 2, un exemple de filtrage d'une image de 256 pixels est couper les échelles spatiales supérieures à $2^4 = 16$ pixels, couper les échelles spatiales supérieures à $2^8 = 256$ pixels n'a pas de sens puisque la carte ne fait que 256 pixels.
- *SExtractor* : logiciel d'extraction de source d'une image, de manière plus générale, à partir d'un fichier de points à trois dimensions, il va extraire les objets des deux premières dimensions en fonction de la valeur de la troisième dimension.

Références

[1] K. M. Ashman, C. M. Bird, and S. E. Zepf. Detecting bimodality in astronomical datasets. *AJ*, 108 :2348–2361, December 1994.

[2] S. Bardelli, E. Zucca, and A. Baldi. A study of the core of the Shapley Concentration - V. The A3528 complex : a young merger event ?*. *MNRAS*, 320 :387–400, January 2001.

[3] L. P. Bautz and W. W. Morgan. On the Classification of the Forms of Clusters of Galaxies. *ApJ*, 162 :L149+, December 1970.

[4] T. C. Beers, K. Flynn, and K. Gebhardt. Measures of location and scale for velocities in clusters of galaxies - A robust approach. *AJ*, 100 :32–46, July 1990.

[5] K. Bekki. Group-Cluster Merging and the Formation of Starburst Galaxies. *ApJ*, 510 :L15–L19, January 1999.

[6] E. Bertin and S. Arnouts. SExtractor : Software for source extraction. *A&AS*, 117 :393–404, June 1996.

[7] D. A. Buote and J. C. Tsai. Quantifying the Morphologies and Dynamical Evolution of Galaxy Clusters. I. The Method. *ApJ*, 452 :522–+, October 1995.

[8] D. A. Buote and J. C. Tsai. Quantifying the Morphologies and Dynamical Evolution of Galaxy Clusters. II. Application to a Sample of ROSAT Clusters. *ApJ*, 458 :27–+, February 1996.

[9] H. Butcher and A. Oemler, Jr. The evolution of galaxies in clusters. I - ISIT photometry of C1 0024+1654 and 3C 295. *ApJ*, 219 :18–30, January 1978.

[10] D. Fadda, E. Slezak, and A. Bijaoui. Density estimation with non-parametric methods. *A&AS*, 127 :335–352, January 1998.

[11] C. Ferrari, C. Benoist, S. Maurogordato, A. Cappi, and E. Slezak. Dynamical state and star formation properties of the merging galaxy cluster Abell 3921. *A&A*, 430 :19–38, January 2005.

[12] P. Flin and J. Krywult. Substructures in Abell clusters of galaxies. *A&A*, 450 :9–14, April 2006.

[13] Y. Hashimoto, J. P. Henry, H. Boehringer, and G. Hasinger. Comparison between X-ray morphology and optical characteristics of clusters of galaxies. *A&A*, 468 :25–31, June 2007.

[14] C. Jones and W. Forman. Einstein Observatory Images of Clusters of Galaxies. *ApJ*, 511 :65–83, January 1999.

[15] I. King. The structure of star clusters. I. an empirical density law. *AJ*, 67 :471–+, October 1962.

[16] V. Kolokotronis, S. Basilakos, M. Plionis, and I. Georgantopoulos. Searching for cluster substructure using APM and ROSAT data. *MNRAS*, 320 :49–60, January 2001.

[17] T. Kronberger, W. Kapferer, C. Ferrari, S. Unterguggenberger, and S. Schindler. On the influence of ram-pressure stripping on the star formation of simulated spiral galaxies. *A&A*, 481 :337–343, April 2008.

[18] Geoff McLachlan and David Peel. The emmix algorithm for the fitting of normal and t-components. *Journal of Statistical Software*, 4(2) :1–14, 7 1999.

[19] C. J. Miller, R. C. Nichol, D. Reichart, R. H. Wechsler, A. E. Evrard, J. Annis, T. A. McKay, N. A. Bahcall, M. Bernardi, H. Boehringer, A. J. Connolly, T. Goto, A. Kniazev, D. Lamb, M. Postman, D. P. Schneider, R. K. Sheth, and W. Voges. The C4 Clustering Algorithm : Clusters of Galaxies in the Sloan Digital Sky Survey. *AJ*, 130 :968–1001, September 2005.

[20] J. J. Mohr, D. G. Fabricant, and M. J. Geller. An X-ray method for detecting substructure in galaxy clusters - Application to Perseus, A2256, Centaurus, Coma, and Sersic 40/6. *ApJ*, 413 :492–505, August 1993.

[21] M. Ramella, A. Biviano, A. Pisani, J. Varela, D. Bettoni, W. J. Couch, M. D'Onofrio, A. Dressler, G. Fasano, P. Kjørgaard, M. Moles, E. Pignatelli, and B. M. Poggianti. Substructures in WINGS clusters. *A&A*, 470 :39–51, July 2007.

[22] D. Richstone, A. Loeb, and E. L. Turner. A lower limit of the cosmic mean density from the ages of clusters of galaxies. *ApJ*, 393 :477–483, July 1992.

[23] H. J. Rood and G. N. Sastry. "Tuning Fork" Classification of Rich Clusters of Galaxies. *PASP*, 83 :313–+, June 1971.

[24] C. L. Sarazin. X-ray emission from clusters of galaxies. *Reviews of Modern Physics*, 58 :1–115, January 1986.

[25] P. Schuecker, H. Böhringer, T. H. Reiprich, and L. Feretti. A systematic study of X-ray substructure of galaxy clusters detected in the ROSAT All-Sky Survey. *A&A*, 378 :408–427, November 2001.

[26] M. A. Strauss, D. H. Weinberg, R. H. Lupton, V. K. Narayanan, J. Annis, M. Bernardi, M. Blanton, S. Burles, A. J. Connolly, J. Dalcanton, M. Doi, D. Eisenstein, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, S. Kent, R. S. J. Kim, G. R. Knapp, R. G. Kron, J. A. Munn, H. J. Newberg, R. C. Nichol, S. Okamura, T. R. Quinn, M. W. Richmond, D. J. Schlegel, K. Shimasaku, M. SubbaRao, A. S. Szalay, D. Vanden Berk, M. S. Vogeley, B. Yanny, N. Yasuda, D. G. York, and I. Zehavi. Spectroscopic Target Selection in the Sloan Digital Sky Survey : The Main Galaxy Sample. *AJ*, 124 :1810–1824, September 2002.

[27] J. H. Yoon, K. Schawinski, Y.-K. Sheen, C. H. Ree, and S. K. Yi. A Spectrophotometric Search for Galaxy Clusters in SDSS. *ApJS*, 176 :414–423, June 2008.

[28] F. Zwicky. *Morphological astronomy*. 1957.