



**HAL**  
open science

# Etude de la morphologie de galaxies lointaines à partir de grands relevés en imagerie infrarouge depuis le sol

Marc Huertas

► **To cite this version:**

Marc Huertas. Etude de la morphologie de galaxies lointaines à partir de grands relevés en imagerie infrarouge depuis le sol. Astrophysique [astro-ph]. Observatoire de Paris, 2008. Français. NNT : . tel-00349783

**HAL Id: tel-00349783**

**<https://theses.hal.science/tel-00349783>**

Submitted on 4 Jan 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OBSERVATOIRE DE PARIS-MEUDON  
École Doctorale d'Astronomie et d'Astrophysique d'Ile-de-France  
UNIVERSIDAD DE GRANADA  
Departamento de Física Teórica y del Cosmos

## THÈSE EN CO-TUTELLE

présentée pour obtenir les diplômes de  
DOCTEUR DE L'OBSERVATOIRE DE PARIS  
Spécialité : Astronomie - Astrophysique  
DOCTOR EN CIENCIAS FÍSICAS POR LA UNIVERSIDAD DE GRANADA

par

Marc HUERTAS-PORTOCARRERO COMPANY

# **Etude de la morphologie de galaxies lointaines à partir de grands relevés en imagerie infrarouge depuis le sol**

Soutenue le 30 septembre 2008 devant le Jury composé de :

Didier Pelat	Président
Roberto Abraham	Rapporteur
Laurence Tresse	Rapporteur
Isabel Márquez	Examinatrice
Eduardo Battaner	Examineur
Daniel Rouan	Co-directeur de thèse
Geneviève Soucail	Co-directrice de thèse
Mariano Moles	Co-directeur de thèse



# Remerciements

*Llegó el momento de echar la vista atrás.  
Tres años de trabajo, cinco años en París.  
Una infinidad de recuerdos  
Dulces y amargos  
Y tantas personas en el camino.*

Ce n'est pas juste un point final à une thèse que je m'apprête à mettre avec ces dernières lignes, mais aussi le point final à une étape. C'est ainsi que je le sens du moins. Je suis venu à Paris pour un an, pour finir mes études d'ingénieur en télécommunications. J'y suis resté cinq et j'ai fait un DEA et une thèse en astronomie. Je ne peux pas m'empêcher de jeter un regard en arrière. Je garde une infinité de souvenirs, souvent sucrés, parfois amers, mais surtout le sentiment d'avoir forgé des relations solides. Une thèse n'est jamais le travail d'une seule personne. Celle-ci n'est pas une exception bien entendu même si parfois j'ai ressenti un certain isolement scientifique. Je n'ai pas envie de parler de cela ici. Il s'agit au contraire de remercier toutes les personnes qui m'ont aidé d'une façon ou d'une autre pour que je sois ici, maintenant, entraîné d'écrire ces dernières lignes.

Merci Daniel. Toujours présent malgré tes nombreuses occupations et la « nouveauté » du sujet. Merci pour ton soutien, pour ta confiance, pour me pousser à continuer dans les moments de faiblesse. Un plaisir d'apprendre auprès de toi. Un plaisir de travailler avec toi. Sincèrement j'espère, que ce n'est que le début.

Une pensée aussi pour Geneviève. Merci pour tes conseils, merci de m'avoir accueilli à Toulouse à plusieurs reprises même si ce n'est pas toujours facile de travailler à distance.

*Gracias Mariano. Gracias por haber confiado en mi sin apenas conocerme, por haberme propuesto trabajo interesante y ayudado a conocer a la comunidad científica española.*

Lidia, merci pour ta patience, pour tout le temps consacré à m'apprendre les subtilités du métier, pour l'accueil chaleureux dès ma première visite à Marseille, pour répondre toujours à mes courriels désespérés en te demandant des données. Cette thèse te doit beaucoup.

Merci Olivier de m'avoir accueilli au sein de l'équipe. Merci pour la confiance que tu as éprouvée en me proposant du travail dès le début. Merci pour les conseils pertinents, pour les critiques constructives qui m'ont permis de me remettre en question, d'avancer, d'apprendre le métier de chercheur en définitive. J'espère pouvoir continuer à travailler avec toi.

Jean-Paul merci de m'avoir proposé du travail intéressant, de m'avoir invité à Marseille à plusieurs reprises, de m'avoir donné des conseils pour mon avenir. Merci.

Merci à toutes les personnes avec qui j'ai pu discuter et qui m'ont offert la possibilité de travailler et d'apprendre. Didier, je te remercie pour tes conseils mathématiques.

*Gracias Alfonso por haberme invitado a Canarias. Un placer trabajar contigo. Todavía nos quedan bastantes cosas por hacer.*

Merci aux membres du jury. Merci d'abord d'avoir accepté de relire les pages qui suivent, même si pour beaucoup d'entre vous le français n'est pas votre langue maternelle. *Eduardo, gracias por haber llevado a cabo todos los procedimientos administrativos en Granada para que pueda optar al título en España.*

Merci Léo, Xavier, collègues de bureau, collègues de bières, collègues de capoeira, collègues de voyages, *amigos vaya*, pour longtemps j'espère. Savoir que vous étiez là, m'a motivé pour venir travailler tous les jours, même dans les moments plus difficiles.

Une pensée également pour les habitants du bâtiment Lyot, Yann, Zoltan, Marie, Arnaud, Fabrice, Thibaut...Je ne vois pas comment j'aurais pu finir ma thèse sans les discussions philosophiques autour de la machine à café et la recherche Wikipédia du jour.

*Mamá, papá, gracias por estar ahí, siempre. No puedo decir nada más, a parte de que os quiero mucho.*

*Lula, pobrecilla, soportando todos estos meses de tesis por aquí, tesis por allá. Gracias. Parece que se acabó por fin.*

**RÉSUMÉ :** Dans l'Univers local, la distribution de galaxies est bimodale, reflétant la relation existante entre la couleur et la morphologie. Une question clef en cosmologie observationnelle est de comprendre comment cette dichotomie est apparue au cours de l'histoire de l'Univers. Les relevés de galaxies ont pour but d'isoler des populations de galaxies à différentes époques cosmiques et de suivre leur évolution. Quantifier de manière fiable les morphologies de galaxies lointaines demeure cependant un obstacle majeur.

Dans le visible, les progrès durant les dix dernières années proviennent essentiellement de données acquises par le télescope spatial Hubble (HST) grâce à la haute résolution angulaire délivrée. Au delà de  $z \sim 1$  pourtant, les caméras visibles du HST sondent le flux UV des galaxies ce qui peut biaiser la morphologie estimée vers des types tardifs. De ce point de vue, les observations dans le proche infrarouge au delà de  $z \sim 1$  sont particulièrement importantes puisque la lumière analysée provient d'étoiles évoluées et est donc moins affectée par des épisodes récents de formation stellaire.

Dans cette thèse, on s'interroge sur la possibilité de quantifier les morphologies de galaxies lointaines dans l'infrarouge proche à l'aide de télescopes au sol afin de quantifier les effets de k-correction morphologique.

Dans un premier temps, l'utilisation d'optique adaptative pour corriger les effets de la turbulence atmosphérique est explorée. On présente l'analyse de 79 galaxies observées avec le système d'optique adaptative NaCo du *Very Large Telescope* (VLT). Les limites sont obtenues grâce à des simulations exhaustives et des comparaisons avec le télescope spatial. Une première estimation de l'évolution morphologique depuis  $z \sim 1.2$  est obtenue à partir de cet échantillon. La comparaison avec les morphologies obtenues avec le télescope spatial laisse entrevoir des différences mais l'échantillon est trop petit pour extraire des conclusions robustes. Un programme d'observations plus large est proposé.

Dans un deuxième temps, des améliorations aux méthodes classiques non-paramétriques d'estimation morphologique sont proposées (nombre illimité de paramètres simultanés, frontières non linéaires) afin de pouvoir analyser des données observées à travers l'atmosphère. On démontre que ces améliorations permettent une estimation morphologique en deux grands types (elliptique, spirale) avec un taux de réussite de  $\sim 80\%$  sur des données limitées par le *seeing*. Le code source a été mis à disposition de la communauté.

La méthode développée est ensuite employée pour analyser un échantillon de  $\sim 50\,000$  galaxies dans l'infrarouge proche. L'évolution morphologique depuis  $z \sim 2$  est obtenue et une quantification des effets de k-correction morphologique est présentée par comparaison avec des données spatiales. On observe une augmentation de la fraction de galaxies elliptiques depuis  $z \sim 2$  ( $\sim 20\%$ ) jusqu'à nos jours ( $\sim 30\%$ ). La classification obtenue est globalement en bon accord avec celle obtenue avec des données spatiales dans le visible jusqu'à  $z \sim 1$ . Au delà, la classification infrarouge a tendance à trouver  $\sim 1.5$  fois plus de galaxies précoces.

**MOTS-CLEFS :** galaxies à grand redshift – morphologies – haute résolution angulaire – optique adaptative – k correction morphologique



MORPHOLOGIES OF DISTANT GALAXIES FROM LARGE COSMOLOGICAL SURVEYS  
USING GROUND BASED NIR IMAGING

**ABSTRACT** : In the local Universe, the distribution of galaxies is bimodal, primarily reflecting a relationship between color and morphology. A capital question in observational cosmology is how this bimodality appears across the history of the Universe. Quantifying morphologies of high redshift objects with a few simple, reliable measurements is however a major obstacle. In the visible, progress over the last ten years have come in particular from the Hubble Space Telescope (HST). At  $z > 1$ , the HST visible cameras probe however the UV flux, dominated by the emission of young stars, which could bias the estimated morphologies towards late-type systems. In this context, near infrared (NIR) observations are particularly important because the K band flux at  $z \sim 1$  is less dependent on the recent history of star formation.

We investigate whether it is possible to obtain morphologies of high redshift galaxies in the NIR using ground based facilities in order to properly quantify the morphological k-correction effects.

We first use adaptive optics at the *Very Large Telescope* to obtain a high resolution NIR sample of 79 galaxies in the Ks band. We obtain the limits of the morphological analysis by means of exhaustive simulations and comparisons with space observations. We then present a first estimate of the morphological evolution of the three main morphological types from  $z \sim 1.2$ . The comparison with morphologies obtained with HST/ACS reveals some discrepancies but the small number of objects does not allow to conclude safely about the nature of these differences. A Large Programme proposal has been submitted to ESO to significantly increase our sample.

Second, some improvements to the classical non-parametric methods are proposed (non-linear boundaries and unlimited parameters) using support vector machines. We show that these improvements enable a determination of the morphological type (in two types) with an  $\sim 80\%$  accuracy on seeing limited data at  $z \sim 1$ . The code is now freely available for download.

The method is then used to analyze a sample of  $\sim 50\,000$  galaxies in the NIR within the COSMOS area. We obtain the morphological evolution from  $z \sim 2$  and quantify the k-correction effects by comparing with HST data. We observe an increase of the early-type population from  $z \sim 2$  ( $\sim 20\%$ ) to the present ( $\sim 30\%$ ) probably reflecting a progressive building up of the red sequence from late-type objects. The classification is globally in good agreement with the one obtained using HST/ACS for  $z < 1$ . Above  $z \sim 1$ , the I-band classification tends to find less early-type galaxies than the Ks-band one by a factor  $\sim 1.5$  which might be a consequence of morphological k-correction effects.

**KEYWORDS** : high redshift galaxies – morphologies – high angular resolution – adaptive optics – morphological k-correction



ESTUDIO DE LAS MORFOLOGÍAS DE GALAXIAS LEJANAS A PARTIR DE GRANDES  
CARTOGRAFIADOS CÓSMICOS EN EL INFRARROJO CERCANO MEDIANTE  
TELESCOPIOS TERRESTRES

RESUMEN : En el Universo local, la distribución de galaxias es bimodal, reflejo de la relación existente entre el color y la morfología. Una cuestión central en cosmología observacional es la comprensión de los mecanismos que han llevado a dicha dicotomía, conocida como secuencia de Hubble. Con el fin de aportar una respuesta, los cartografiados cósmicos aíslan poblaciones de galaxias en épocas diferentes permitiendo así seguir su evolución. La estimación de la morfología de galaxias lejanas de manera fiable sigue siendo sin embargo un problema a la hora de analizar e interpretar los datos.

En el dominio óptico, los progresos en éstos últimos años se han producido esencialmente gracias al telescopio espacial Hubble (HST). Sin embargo a  $z > 1$ , las cámaras ópticas del HST sondan el espectro ultravioleta de las galaxias lo que puede sesgar la morfología obtenida hacia tipos más tardíos.

Las observaciones en el infrarrojo cercano en torno a  $z \sim 1$  son desde este punto de vista especialmente interesantes ya que la luz analizada proviene de estrellas tardías y se ve por consiguiente menos afectada por episodios recientes de formación estelar que las observaciones con cámaras visibles.

En la presente tesis se plantea pues la posibilidad de emplear telescopios en tierra para el análisis de morfologías de galaxias lejanas. En un primer lugar se explora la posibilidad del uso de óptica adaptativa para corregir los efectos de la turbulencia atmosférica en la resolución angular de la imagen. Para ello se analiza una muestra de 79 galaxias observadas con el sistema de óptica adaptativa del *Very Large Telescope* (VLT). Se obtienen los límites actuales de la técnica mediante simulaciones exhaustivas y comparaciones con datos espaciales y se presenta un primer análisis de la evolución morfológica de los tres principales tipos (espiral, elíptica, irregular) desde  $z \sim 1.5$ .

En un segundo lugar se presentan una serie de mejoras aportadas a las técnicas de análisis morfológico no paramétricas como son el uso de un número ilimitado de parámetros y fronteras no lineales. Se demuestra que estas mejoras permiten la determinación de la morfología sobre datos limitados por el *seeing* con una tasa de acierto del  $\sim 80\%$ . Dicho método se emplea a continuación para el análisis de  $\sim 50\,000$  galaxias de campo en el área de COSMOS. Se obtiene la evolución morfológica desde  $z \sim 2$  de los dos tipos principales y se estudian los efectos de k-corrección morfológica mediante comparación con datos del HST. Se observa en particular un aumento de la fracción de galaxias de tipo precoz desde  $z \sim 2$  ( $\sim 20\%$ ) hasta hoy ( $\sim 30\%$ ). La morfología obtenida está en buen acuerdo con la obtenida con el telescopio espacial en el dominio óptico hasta  $z \sim 1$ . Más allá, la clasificación infrarroja tiende a encontrar  $\sim 1.5$  veces más de galaxias precoces.

PALABRAS CLAVE : galaxias a alto *redshift* – morfología – alta resolución angular – óptica adaptativa – k-corrección morfológica



*A mi abuela,  
La primera doctora de la familia.*



# Table des matières

<b>Avant-propos</b>	<b>9</b>
<b>1 Introduction : Cosmologie et galaxies</b>	<b>13</b>
1.1 Vision actuelle de l'Univers : le modèle $\Lambda$ CDM	14
1.1.1 Un Univers en expansion : la constante de Hubble	14
1.1.2 Une origine chaude et dense	15
1.1.3 Matière sombre	16
1.1.4 Énergie sombre : un Univers accéléré	17
1.2 Le modèle de concordance	18
1.2.1 L'Univers primordial	18
1.2.2 Premières étoiles - réionisation	18
1.2.3 Croissance hiérarchique	19
1.2.4 Un Univers modélisé	20
1.3 Histoire des baryons	21
1.3.1 Une distribution de galaxies bimodale : la séquence de Hubble	21
1.3.2 Archéologie galactique : l'époque des grands relevés	25
1.3.3 Quand les galaxies se forment-elles ? Histoire de la formation stellaire	26
1.3.4 Où les galaxies acquièrent-elles leur morphologie ? Le rôle de l'environnement	32
1.3.5 Comment les galaxies acquièrent-elles leur morphologie ? Interactions, <i>Downsizing</i> , noyaux actifs	33
Bibliographie	35
<b>2 Morphologies de galaxies lointaines</b>	<b>43</b>
2.1 D'une classification visuelle à une « science morphologie »	44
2.1.1 Les premières classifications	44
2.1.2 Morphologie ou traceur de la structure physique des galaxies	45
2.2 Morphologies de galaxies lointaines	45
2.2.1 Une automatisation des méthodes	46
2.2.2 Le télescope spatial Hubble	52
2.2.3 Une difficulté d'interprétation : la k-correction morphologique	55
2.3 Le travail de cette thèse	57
Bibliographie	58
<b>I Morphologies de galaxies à grand redshift avec optique adaptative</b>	<b>63</b>
<b>3 Cosmologie et optique adaptative</b>	<b>65</b>
3.1 Formation d'une image au foyer d'un télescope	66

3.1.1	Cas idéal . . . . .	66
3.1.2	L'effet de l'atmosphère . . . . .	67
3.2	L'optique adaptative . . . . .	68
3.2.1	Principe . . . . .	68
3.2.2	FEP d'un système d'optique adaptative . . . . .	70
3.2.3	Les limites de l'OA classique . . . . .	71
3.3	Relevés cosmologiques, morphologie et optique adaptative . . . . .	72
3.3.1	A la limite de la technique . . . . .	72
3.3.2	Revue des travaux précédents . . . . .	73
	Bibliographie . . . . .	74
<b>4</b>	<b>Morphologies avec OA</b>	<b>77</b>
4.1	Description des données . . . . .	79
4.1.1	Quelques mots sur l'instrument NACO . . . . .	79
4.1.2	Observations . . . . .	80
4.1.3	Réduction . . . . .	82
4.2	Construction du catalogue . . . . .	84
4.2.1	Détection . . . . .	84
4.2.2	Nettoyage . . . . .	84
4.2.3	Complétude . . . . .	86
4.2.4	Comptages . . . . .	86
4.2.5	<i>Redshifts</i> photométriques . . . . .	88
4.2.6	Magnitudes absolues au repos . . . . .	88
4.3	Morphologies automatiques . . . . .	90
4.3.1	Considérations préliminaires . . . . .	90
4.3.2	Détection des galaxies irrégulières . . . . .	91
4.3.3	Analyse des galaxies régulières . . . . .	94
4.4	Résultats . . . . .	105
4.4.1	Statistiques globales . . . . .	105
4.4.2	Evolution morphologique . . . . .	106
4.5	Quantification des effets de k-correction morphologique . . . . .	109
4.5.1	Comparaison globale . . . . .	109
4.5.2	Evolution morphologique . . . . .	111
4.6	Conclusions de l'étude . . . . .	112
	Bibliographie . . . . .	114
<b>5</b>	<b>Déconvolution</b>	<b>117</b>
5.1	Principe . . . . .	118
5.2	Méthodes de déconvolution . . . . .	118
5.2.1	Maximum de vraisemblance . . . . .	119
5.2.2	Maximum à posteriori . . . . .	119
5.2.3	Déconvolution myope . . . . .	121
5.3	Application au champ NACO . . . . .	121
5.4	Sommaire et conclusions . . . . .	123

Bibliographie . . . . .	124
<b>6 Travaux en cours et perspectives</b>	<b>125</b>
6.1 Construction de la séquence de Hubble . . . . .	126
6.1.1 Caractéristiques principales du programme . . . . .	126
6.1.2 Objectifs scientifiques . . . . .	127
6.1.3 Des outils prêts . . . . .	127
6.1.4 Du neuf par rapport à NICMOS ? . . . . .	128
6.1.5 Une nouvelle approche, l'étoile laser . . . . .	128
6.1.6 Réactions aux premières soumissions . . . . .	130
6.2 Nouveaux enjeux de l'optique adaptative . . . . .	130
Bibliographie . . . . .	132
<b>Conclusions de la première partie</b>	<b>135</b>
<b>II Morphologies automatiques non paramétriques sur des données au sol non corrigées</b>	<b>137</b>
<b>7 Morphologies non paramétriques</b>	<b>139</b>
7.1 Survol historique . . . . .	140
7.2 Plus de deux paramètres ? . . . . .	140
7.3 Etalonnage de l'espace de paramètres ? . . . . .	141
7.4 Frontières optimales et estimation des erreurs ? . . . . .	141
Bibliographie . . . . .	142
<b>8 Machines à vecteurs support</b>	<b>145</b>
8.1 Machines d'apprentissage : définition . . . . .	146
8.2 Les machines à vecteurs support . . . . .	146
8.2.1 Machines linéaires, données séparables . . . . .	147
8.2.2 Machines linéaires, données non-séparables . . . . .	149
8.2.3 Machines non linéaires, données non-séparables . . . . .	149
8.3 Remarques sur l'entraînement et le réglage des paramètres . . . . .	152
8.4 Probabilités . . . . .	152
8.5 Bilan . . . . .	153
Bibliographie . . . . .	153
<b>9 GalSVM</b>	<b>155</b>
9.1 Validation sur un échantillon bien résolu . . . . .	157
9.1.1 C/A classique et SVM 2-D . . . . .	157
9.1.2 C/A classique et SVM 4-D . . . . .	158
9.2 Application à des galaxies lointaines . . . . .	159
9.2.1 Description de la méthode . . . . .	161
9.2.2 L'échantillon réel . . . . .	162
9.2.3 L'étalonnage . . . . .	163

9.2.4	Mesure des paramètres de classification . . . . .	164
9.2.5	Résultats . . . . .	165
9.2.6	Quelques mots sur le mélange morphologique . . . . .	169
9.3	Mise à disposition du code . . . . .	170
9.4	Bilans et conclusions . . . . .	170
	Bibliographie . . . . .	171
<b>10</b>	<b>Morphologies dans le champ COSMOS</b>	<b>173</b>
10.1	Le relevé COSMOS . . . . .	175
10.1.1	Description générale . . . . .	175
10.1.2	Données multi-longueur d'onde . . . . .	176
10.1.3	Principaux objectifs scientifiques . . . . .	176
10.2	Motivations du travail . . . . .	177
10.3	Les données . . . . .	178
10.3.1	Description . . . . .	178
10.3.2	Nettoyage . . . . .	179
10.3.3	Séparation étoiles/galaxies . . . . .	179
10.3.4	<i>Redshifts</i> photométriques . . . . .	179
10.3.5	Le catalogue morphologique final . . . . .	180
10.4	Analyse morphologique . . . . .	181
10.4.1	L'échantillon d'entraînement . . . . .	182
10.4.2	Classification . . . . .	182
10.4.3	Remarques sur l'influence du nombre de paramètres . . . . .	183
10.4.4	Catalogue de sortie et qualité des classifications . . . . .	183
10.5	Résultats et discussions . . . . .	186
10.5.1	Estimations des statistiques globales . . . . .	186
10.5.2	Comptages par type morphologique . . . . .	190
10.5.3	Evolution morphologique . . . . .	191
10.6	Autour des effets de k-correction morphologique . . . . .	191
10.6.1	Comparaison galaxie par galaxie . . . . .	193
10.6.2	Evolutions morphologiques comparées . . . . .	193
10.7	Conclusions et perspectives . . . . .	195
	Bibliographie . . . . .	198
<b>11</b>	<b>Autres applications</b>	<b>201</b>
11.1	Morphologies ALHAMBRA . . . . .	203
11.1.1	Le relevé ALHAMBRA . . . . .	203
11.1.2	Objectifs scientifiques de l'étude . . . . .	203
11.1.3	Les données analysées . . . . .	204
11.1.4	Catalogue photométrique et <i>redshifts</i> photométriques . . . . .	204
11.1.5	Sélection de l'échantillon morphologique . . . . .	206
11.1.6	Analyse morphologique . . . . .	207
11.1.7	Bilan et perspectives . . . . .	211
11.2	Morphologies dans des amas à <i>redshift</i> intermédiaire ( $z \sim 0.5$ ) . . . . .	213

---

11.2.1 Motivations du travail réalisé . . . . .	213
11.2.2 Description des données . . . . .	214
11.2.3 Sélection des galaxies appartenant à l'amas . . . . .	215
11.2.4 Analyse morphologique . . . . .	215
11.2.5 Inspection visuelle . . . . .	216
11.2.6 Relation <i>couleur–magnitude</i> . . . . .	216
11.2.7 Relation <i>morphologie–distance radiale</i> . . . . .	219
11.2.8 Conclusions et travaux en cours . . . . .	219
Bibliographie . . . . .	221
<b>Conclusion de la deuxième partie</b>	<b>219</b>
<b>Conclusions</b>	<b>217</b>
<b>Conclusiones (español)</b>	<b>221</b>
<b>Summary and conclusions (english)</b>	<b>225</b>
<b>A Redshifts photométriques</b>	<b>221</b>
<b>B Morphologies des 79 galaxies observées avec NACO</b>	<b>223</b>
<b>C Proposition de Large Programme ESO</b>	<b>229</b>
<b>D Documentation pour galSVM</b>	<b>241</b>
<b>E Articles publiés dans des revues à comité de lecture</b>	<b>253</b>



# Avant-propos

La cosmologie vise à étudier l'Univers comme un tout ; sa forme, son contenu et son histoire. Certains disent qu'il ne peut s'agir d'une science proprement dite puisqu'on ne dispose que d'un seul et unique échantillon. Pourtant, probablement depuis que l'Homme est Homme et possède des yeux pour regarder le ciel, il essaye de trouver une logique à ces milliers de points lumineux qui le guêtent chaque nuit. Ainsi, entre les premières interprétations mythologiques et le modèle du Big-Bang on trouve des personnes telles qu'Aristote, Ptolémée, Copernic, Galilée, Newton, Einstein qui par leurs pensées ont fait évoluer le regard de l'Homme sur l'Univers. Aujourd'hui, il semblerait que celui-ci est en expansion accélérée à cause d'une énergie de nature inconnue que l'on nomme *énergie sombre*, que la plupart de la matière qu'il abrite est invisible (*matière sombre*) et que tout aurait commencé il y a 13 milliards d'années par une phase chaude et très dense. La seule partie que l'on voit, la matière baryonique, se trouve essentiellement sous forme d'étoiles et de gaz rassemblés à leur tour dans des *galaxies*. C'est justement ce dont il s'agit dans ce manuscrit, de formation et d'évolution de galaxies et plus particulièrement de l'évolution de leurs formes ou *morphologies*. Comment la matière a-t-elle acquis sa structure actuelle ? Comment est-on passé de cet état de forte densité, presque homogène, à l'Univers d'aujourd'hui avec toute sa diversité de formes ?

Quand Hubble découvrit que les galaxies étaient des objets très distants, dans les années 30, il les classa en deux grands types d'après leur apparence. D'un côté des galaxies plutôt fines, circulaires et présentant des bras spiraux (*les galaxies spirales*), de l'autre des galaxies ovales présentant peu de structures (*les galaxies elliptiques*). Cette description purement phénoménologique s'est avérée être le reflet d'une vraie dichotomie cachant des propriétés physiques très différentes. Alors que les galaxies spirales possèdent essentiellement des étoiles jeunes et du gaz à partir duquel elles continuent à en former des nouvelles, les galaxies elliptiques semblent avoir achevé la transformation de gaz en étoiles et contiennent, en grande majorité, des étoiles vieilles. Ceci n'est qu'un exemple des nombreuses différences entre les deux populations. Comment cette dichotomie est-elle apparue ? Quels sont les chemins évolutifs suivis par chacun de ces objets ? Comment interagissent-ils entre eux ?

Le cosmologiste dispose d'un avantage remarquable par rapport à un archéologue ou un historien. Grâce à la vitesse finie de la lumière, il lui suffit de regarder de plus en plus loin pour remonter le temps à la recherche des ancêtres des galaxies actuelles. Les grands relevés de galaxies font en effet ce travail. De ce point de vue, étudier les morphologies de galaxies lointaines donne un accès direct à la structure de la matière dans le passé et donc une information précieuse sur son assemblage et son évolution.

Et pourtant les études des galaxies éloignées ne sont que très récentes et malgré les énormes progrès des dernières années il reste toujours un grand nombre de problèmes non résolus. Il ne s'agit pas en effet d'une tâche facile. L'information y est, certes, mais les objets lointains apparaissent petits et très faibles dans le ciel. Il a fallu attendre d'avoir des instruments suffisamment performants pour les détecter et les analyser. De plus, si l'accès à l'information semble plus facile que pour un archéologue, son analyse est cependant plus compliquée étant donné que le seul moyen est à travers l'étude indirecte de la lumière émise.

Le télescope spatial Hubble (HST) a permis de faire des avancées extraordinaires dans notre compréhension de l'évolution de l'Univers. Sa résolution angulaire sans précédents ainsi que sa très haute sensibilité, ont permis de sonder les populations de galaxies de l'Univers quand celui-ci avait la moitié de l'âge actuel ( $z \sim 1$ ). Il a permis d'isoler des populations de différents types morphologiques à différentes époques afin d'étudier leurs propriétés. On a pu se rendre compte ainsi que les galaxies dans le passé apparaissent plus perturbées, moins symétriques que dans l'Univers local et que leur activité de formation stellaire était plus importante. De même, on sait maintenant également que l'environnement, ou plutôt la densité de celui-ci, joue un rôle important dans l'apparence de la galaxie. Les fusions de galaxies sont effectivement de plus en plus fréquentes si on remonte le temps, indiquant une croissance hiérarchique des structures. Malgré cela, il semblerait que les galaxies les plus massives de l'Univers, généralement des elliptiques, étaient déjà présentes à cette époque comme si elles avaient subi un épisode de formation stellaire très rapide qui s'est arrêté brutalement. Pour expliquer ce comportement *anti-hiérarchique* il est nécessaire de faire intervenir l'activité de supernovae ou de noyaux actifs pour mettre un frein à la naissance d'étoiles. Cependant, les mécanismes exactes restent inconnus. De plus, expliquer cet assemblage rapide de la masse demeure toujours un problème.

Certaines questions requièrent un sondage de galaxies encore plus lointaines, *les progéniteurs des progéniteurs*. Lorsque l'on observe cependant des galaxies au delà de  $z \sim 1$  avec les caméras visibles du HST, les effets de décalage vers le rouge dus à l'expansion de l'Univers impliquent que nous recevons en réalité l'émission UV de la galaxie qui n'est pas forcément représentative de la forme intrinsèque de celle-ci. La solution passe par des observations dans des filtres plus rouges. Le télescope spatial possède bien une caméra de ce type (NICMOS), mais la résolution obtenue, nettement inférieure qu'avec ses caméras visibles, fait que jusqu'à présent aucune étude quantitative sur les morphologies des galaxies n'a été menée. Les relevés de galaxies futurs sont donc demandants d'une analyse morphologique dans l'infrarouge.

Qu'en est-il des télescopes au sol ? Est-il possible de les employer pour mener cette étude ? Cette thèse tente en partie de répondre à ces questions en explorant deux voies distinctes.

Au premier abord, la réponse semble être plutôt négative. Les effets dévastateurs de l'atmosphère sur la résolution de l'image obtenue rendent très difficile l'estimation des formes des galaxies très lointaines. Cependant les grands télescopes sur Terre sont pratiquement tous équipés de systèmes d'optique adaptative permettant de corriger les aberrations introduites par l'atmosphère et qui ont justement un fonctionnement optimal dans l'infrarouge. En quelle mesure l'emploi de ceux-ci apporte-t-il un gain pour l'étude des morphologies ? Ce sera l'objet de la première partie de ce manuscrit.

Le volume de données d'optique adaptative disponible actuellement reste cependant encore assez limité alors que des données en infrarouge non corrigées dans de très bons sites sont déjà exploitables. Est-il possible d'améliorer les techniques d'estimation pour extraire une information morphologique même à partir de ces données basse résolution ? J'essaierai d'apporter une réponse dans la deuxième partie.

Le mémoire débute ainsi par un court exposé de l'état d'art de nos connaissances en cosmologie et évolution de galaxies d'un point de vue observationnel afin de situer le cadre dans lequel se déroule cette thèse. Ce premier chapitre d'introduction est suivi d'un deuxième chapitre introductif plus focalisé sur l'estimation de morphologies de galaxies lointaines et sur les difficultés rencontrées.

Par la suite, la première partie de la thèse aborde l'utilisation de l'optique adaptative pour des fins cosmologiques.

Le troisième chapitre décrit les principales caractéristiques d'un système d'optique adaptative et les difficultés rencontrées lorsque il est employé pour des fins cosmologiques. Une revue des travaux précédents ayant utilisé l'optique adaptative pour étudier l'évolution de galaxies est également présentée.

Dans le quatrième chapitre nous montrons les résultats d'un programme d'observations dans l'infrarouge proche mené au *Vey Large Telescope* (VLT) avec le système d'optique adaptative NAOS/CONICA. Une analyse morphologique de 79 galaxies à  $z \sim 1$  avec deux méthodes automatiques d'estimation est présentée. On présente également une comparaison avec des données spatiales sur les mêmes objets.

Le cinquième chapitre montre des résultats préliminaires de la déconvolution appliquée aux données présentées.

La première partie finit ainsi par un exposé des travaux en cours. On présente en particulier les objectifs principaux d'un programme d'observations au VLT.

Dans le deuxième partie, une méthode non-paramétrique améliorée d'estimation morphologique est présentée, ainsi que des applications de celle-ci à des données observées depuis le sol. La méthode est basée sur un type particulier de machines

d'apprentissage. La deuxième partie débute ainsi par un exposé des principales limitations des méthodes paramétriques.

Par la suite, les principes de base de ces machines sont présentées dans le deuxième chapitre de cette seconde partie.

Dans le troisième chapitre, on présente les caractéristiques principales de la méthode développée et les tests réalisés pour quantifier les performances obtenues.

Les deux derniers chapitres du manuscrit décrivent des résultats scientifiques obtenus en appliquant la méthode décrite auparavant. En particulier, un chapitre est dédié à l'analyse de  $\sim 50\,000$  galaxies de champ dans l'aire du relevé COSMOS observées avec l'instrument WIRC*am* dans l'infrarouge proche. Le dernier chapitre illustre les travaux en cours : la méthode est en train d'être employée pour étudier des galaxies de champ dans le cadre du relevé ALHAMBRA et des galaxies dans des amas observées avec le *Canada France Hawaii Telescope*.

# Introduction : Cosmologie et galaxies

---

## Sommaire

<b>1.1</b>	<b>Vision actuelle de l'Univers : le modèle <math>\Lambda</math>CDM</b>	<b>14</b>
1.1.1	Un Univers en expansion : la constante de Hubble	14
1.1.2	Une origine chaude et dense	15
1.1.3	Matière sombre	16
1.1.4	Energie sombre : un Univers accéléré	17
<b>1.2</b>	<b>Le modèle de concordance</b>	<b>18</b>
1.2.1	L'Univers primordial	18
1.2.2	Premières étoiles - réionisation	18
1.2.3	Croissance hiérarchique	19
1.2.4	Un Univers modélisé	20
<b>1.3</b>	<b>Histoire des baryons</b>	<b>21</b>
1.3.1	Une distribution de galaxies bimodale : la séquence de Hubble	21
1.3.1.1	Propriétés des bulbes (ou galaxies elliptiques)	22
1.3.1.2	Propriétés des disques	23
1.3.1.3	Bimodalité	24
1.3.2	Archéologie galactique : l'époque des grands relevés	25
1.3.3	Quand les galaxies se forment-elles ? Histoire de la formation stellaire	26
1.3.3.1	Taux de formation stellaire	26
1.3.3.2	Fonction de luminosité	28
1.3.3.3	Couleur-magnitude	31
1.3.4	Où les galaxies acquièrent-elles leur morphologie ? Le rôle de l'environnement	32
1.3.5	Comment les galaxies acquièrent-elles leur morphologie ? Interactions, <i>Downsizing</i> , noyaux actifs	33
	<b>Bibliographie</b>	<b>35</b>

---

**Abstract :** This chapter briefly presents recent progresses in the fields of observational cosmology and galaxy evolution in order to situate the framework of this PhD thesis. We first present the accepted vision of the Universe ( $\Lambda - CDM$  model)

and we then describe last advances in galaxy evolution from an observational point of view using in particular results from recent galaxy surveys.

---

**Resumen :** El siguiente capítulo presenta de manera breve los conocimientos actuales y las problemáticas futuras en el campo de la cosmología y la evolución de galaxias con el fin de situar el marco en torno al cual se desarrolla la presente tesis doctoral. Se presenta en un primer lugar la visión actual del Universo (modelo  $\Lambda - CDM$ ) y se abordan posteriormente los avances de los últimos años en la comprensión de los mecanismos de formación de galaxias desde un punto de vista observacional.

---

Pendant plus de 10 siècles, le modèle cosmologique de notre civilisation situait la Terre statique au centre de l'Univers, le Soleil, les planètes et les étoiles décrivant des orbites circulaires autour de celle-ci. L'énoncé d'une théorie de la gravitation satisfaisante (Newton puis Einstein) ainsi que des progrès dans la conception d'instruments ont permis de passer de ce modèle géocentrique à un Univers en expansion accélérée dominé par une énergie sombre et où la majeure partie de la matière est sous forme de matière non-baryonique ou matière noire. Ce chapitre vise à revoir brièvement les principales caractéristiques de notre vision actuelle de l'Univers afin de soulever les questions autour desquelles se déroule le travail de cette thèse.

## 1.1 Vision actuelle de l'Univers : le modèle $\Lambda$ CDM

Notre vision actuelle de l'Univers repose essentiellement sur quatre piliers observationnels, inattendus lors de leur découverte et qui ont révolutionné le rapport de l'Homme à l'Univers.

### 1.1.1 Un Univers en expansion : la constante de Hubble

L'histoire de la cosmologie observationnelle moderne commence probablement avec les observations de Slipher et Hubble dans les années 1917-1925. Grâce à des mesures sur des étoiles Céphéïdes (étoiles pulsantes dont la période est précisément liée à la luminosité intrinsèque), ils démontrèrent que ces objets nébuleux rassemblant d'immenses quantités d'étoiles, étaient en réalité des objets physiques indépendants et très éloignés de nous : c'était la découverte des *galaxies*. Ils prouvèrent également que la lumière provenant de ces galaxies était d'autant plus décalée vers le rouge que les galaxies étaient lointaines, indiquant qu'elles s'éloignaient de nous avec une vitesse d'autant plus grande que leur distance était importante<sup>1</sup>. *L'Univers semblait*

---

<sup>1</sup>Ce décalage vers le rouge cosmologique est généralement désigné avec le terme anglais *redshift* et est désormais une mesure de distance et de temps dans l'Univers. Ce sera ce terme que l'on

*donc être en expansion.* Ce résultat mit longtemps à être accepté, et même Einstein continua à promouvoir l'idée d'un Univers statique, malgré les prédictions de sa propre théorie, jusqu'aux années 30.

Depuis, l'un des plus grands objectifs de la cosmologie observationnelle a été de mesurer la constante caractéristique de cette expansion appelée aussi constante de Hubble ( $H_0$ ). Les premières mesures effectuées en utilisant des télescopes terrestres datent en effet des années 1970 et donnèrent une valeur moyenne de  $67 \pm 15 \text{ km.s}^{-1}.\text{Mpc}^{-1}$ , très proche de la valeur actuelle.

Elle a été par la suite mesurée avec de plus en plus de précision en utilisant en particulier le télescope spatial Hubble (Freedman *et al.*, 2001 et Fig. 1.1) et les données du satellite *Wilkinson Microwave Anisotropy Probe* (WMAP, Spergel, 2003) et semble converger vers une valeur de  $H_0 = 70 \text{ km.s}^{-1}.\text{Mpc}^{-1}$ .

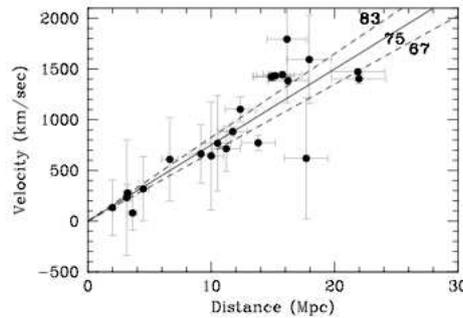


FIG. 1.1 – Mesure de la constante de Hubble via la mesure de distance de galaxies en utilisant des Céphéides avec le télescope spatial (Freedman *et al.*, 2001).

### 1.1.2 Une origine chaude et dense

En réalisant des mesures avec une antenne radio, Penzias et Wilson (1965), détectèrent un bruit résiduel qu'ils n'arrivaient pas à éliminer. Ce rayonnement isotrope à  $\sim 3K$  s'est avéré être la preuve la plus solide en faveur de la théorie du Big-Bang. La découverte du rayonnement fossile (CMB de l'anglais *Cosmic Microwave Background*) semblait prouver en effet que l'Univers aurait un début et que durant ses premières années, *il a dû être très chaud et extrêmement dense.* Les théories alternatives qui prédisaient un Univers statique (*Steady State*) comme une solution plus naturelle devaient ainsi être progressivement abandonnées.

La compréhension de cette émission a beaucoup progressé durant les 15 dernières années grâce aux données du satellite COBE (*Cosmic Background Explorer*). En particulier on a pu confirmer avec une extrême précision la nature de son spectre de corps noir (Mather *et al.*, 1990), preuve de son origine de type thermique et

---

emploiera tout le long du manuscrit.

mettre en évidence les premières fluctuations (Smoot *et al.*, 1992) à l'origine probablement des structures que l'on connaît actuellement (amas, galaxies). Les dernières mesures de spectre de puissance ont été réalisées avec WMAP (Spergel, 2003) : la localisation précise du premier pic dans le spectre de puissance angulaire a permis de contraindre significativement entre autres le contenu énergétique total de l'Univers ( $\Omega_{TOT}$ ).

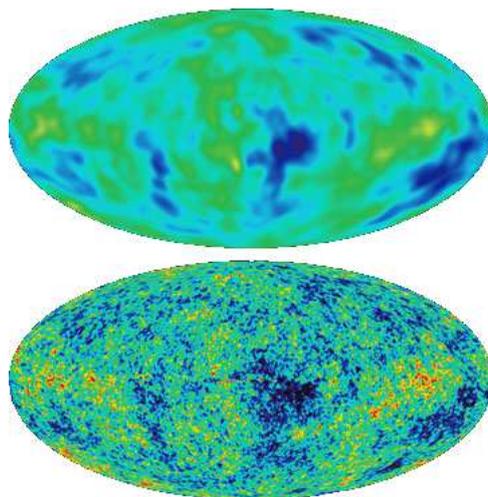


FIG. 1.2 – Fond diffus cosmologique (CMB) : en haut, mesure par le satellite COBE en 1992, en bas, mesure par le satellite WMAP 10 ans plus tard. Les fluctuations de température sont à l'origine des galaxies actuelles [Source : NASA/COBE/WMAP Science Team].

### 1.1.3 Matière sombre

La présence de *matière sombre* dans l'Univers a été constatée dès les années 30 grâce à des mesures des vitesses de dispersion ( $\sigma$ ) des galaxies dans l'amas de Coma (Zwicky, 1933). En supposant que l'amas était en équilibre et que les orbites des galaxies se distribuaient uniformément, il trouvait une masse dynamique contenue dans un rayon caractéristique  $R_{cl}$  de  $M = 3 \langle \sigma^2 \rangle R_{cl}/G$ , largement supérieure à la masse contenue dans les étoiles visibles. Jugée peut-être pertinente à l'époque mais redécouverte, cette mesure indiquait que les amas semblent donc être contenus dans un halo de matière non visible dont sa seule interaction avec son environnement est de type gravitationnel.

Aux échelles galactiques, la mesure de la courbe de rotation des étoiles dans les parties externes des galaxies spirales (Rubin, 2000) permet d'en déduire la masse contenue dans un rayon  $R$  en supposant des orbites circulaires pour les étoiles :  $\frac{GM(<R)}{R^2} = \frac{V^2}{R}$ . Ces courbes de rotation se sont révélées plates ( $V \sim \text{constante}$ ) jusqu'à des distances très grandes du centre de la galaxie. Le seul moyen d'expliquer

cela est en admettant que la masse contenue dans un rayon  $R$  croît proportionnellement à celui-ci ( $M(< R) \propto R$ ). Les galaxies spirales semblent ainsi elles aussi se situer dans des halos de matière sombre. La mise en évidence de ces halos autour des galaxies elliptiques est moins convaincante car il n'existe pas de traceurs de la masse aux échelles désirées pour ce type de galaxies à cause essentiellement de leur faible contenu en gaz et du fait que les étoiles ne décrivent pas des orbites circulaires (Gerhard *et al.*, 2001). Cependant, récemment, en combinant des mesures de lentilles gravitationnelles avec la dynamique des étoiles sur des galaxies elliptiques, Koopmans & Treu (2003) et Treu *et al.* (2006) ont réalisé une cartographie de la matière noire autour de ces galaxies et confirmé la présence du halo.

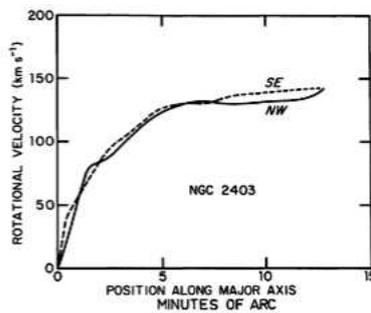


FIG. 1.3 – Courbe de rotation de NGC 2403 (Shostak, 1973)

Aux plus grandes échelles, la technique du *weak lensing* permet d'estimer le contenu global de l'Univers en matière sombre (Refregier, 2003).

En définitive, la matière « normale » (baryonique) que l'on connaît n'est pas la matière dominante dans l'Univers. Au contraire, une matière dite sombre dont on ignore la nature et dont la seule expression est à travers la force de gravitation a été mise en évidence à toutes les échelles de l'Univers et domine la distribution de matière.

#### 1.1.4 Énergie sombre : un Univers accéléré

A cause de la nature attractive des forces de gravitation, il semble assez logique de penser que l'expansion de l'Univers décélère sous l'effet de ces forces. Ce paramètre de décélération ( $q_0$ ) issu des équations de la relativité générale a pu être estimé notamment, en analysant la lumière provenant des supernovae de type Ia qui sont d'autres chandelles que les Céphéïdes, mais bien plus lumineuses (Astier *et al.*, 2006). Toutes les tentatives de mesure ont résulté en des valeurs négatives de  $q_0$ , ce qui correspond à une accélération et non à une décélération de l'Univers, comme s'il y avait une énergie répulsive s'opposant aux effets de la gravitation. Le seul moyen d'autoriser une accélération dans un Univers de Friedmann est d'introduire une constante cosmologique non nulle :  $q_0 = \frac{\Omega_M}{2} - 3\frac{\Omega_\Lambda}{2}$  avec  $\Omega_\Lambda$  étant l'énergie associée à cette constante, appelée aussi *énergie sombre*.

## 1.2 Le modèle de concordance

Les preuves observationnelles exposées précédemment ont permis de converger vers un modèle d'Univers dit *de concordance*  $\Lambda$ CDM ( $\Lambda$  pour énergie sombre et CDM pour *Cold Dark Matter*). Les derniers résultats de WMAP (Spergel, 2006) confirment que nous habitons dans un Univers dominé par l'énergie sombre ( $\sim 73\%$ ) et dont la plupart de la matière est sous forme de matière noire ( $\sim 22\%$ ) contre seulement  $\sim 5\%$  de matière baryonique. Par conséquent seul la nature du constituant le moins abondant est connue. Il est donc pertinent de se demander pourquoi ce modèle est considéré comme l'un des plus grands succès de la cosmologie moderne.

C'est un fait qu'il a réussi à expliquer avec une précision remarquable la croissance des structures depuis le CMB jusqu'aux galaxies locales. Je décris par la suite les principales étapes de l'histoire de l'Univers telle qu'on la conçoit actuellement.

### 1.2.1 L'Univers primordial

Il y a  $13.6 \text{ Gyr}^2$  l'Univers était dense et chaud. Des réactions nucléaires entre les protons et les neutrons dans l'Univers en expansion a résulté la synthèse des éléments plus légers (D, He, Li) : c'est la nucléosynthèse primordiale. Cet état de forte densité assurait l'équilibre thermique entre les espèces baryoniques et bosoniques. Le constituant principal de l'Univers à ce stade est la rayonnement électromagnétique. L'expansion de l'Univers conduit inexorablement à la rupture de cet équilibre : la densité des espèces diminuant, les taux de collisions ne suffisent plus à maintenir l'équilibre thermique entre les espèces qui se découplent progressivement les unes des autres. Le découplage matière-rayonnement est le dernier événement de ce type et s'est produit à  $z \sim 1000$ , environ 380 000 ans après le Big-Bang. C'est la lumière qui baignait la matière au moment de ce découplage que l'on reçoit lorsqu'on observe le CMB. A partir de cette époque, la matière se retrouve libre de toute interaction autre que gravitationnelle et on entre dans une époque où la matière domine la densité globale de l'Univers et non plus la radiation. Les électrons n'interagissent plus avec les photons et s'associent aux noyaux atomiques formés durant la nucléosynthèse primordiale, c'est la recombinaison. L'Univers entre alors dans une phase dite « sombre ».

### 1.2.2 Premières étoiles - réionisation

A partir de ce moment, le moteur principal de la formation des structures à grande-échelle est constitué des instabilités gravitationnelles. La croissance des structures dépend des inhomogénéités primordiales de la distribution de matière. L'étude détaillée du CMB révèle que l'Univers primordial n'était pas tout à fait homogène,

---

<sup>2</sup>1Gyr=10<sup>9</sup> années

il abritait en effet des petites inhomogénéités qui seraient à l'origine des galaxies actuelles. C'est à ce stade que la matière noire joue un rôle fondamental. En effet, si la matière baryonique était la seule matière de l'Univers, la faible densité n'aurait permis la croissance de structures que bien après le découplage et l'Univers d'aujourd'hui continuerait à être presque homogène. La matière noire agit donc comme un catalyseur : étant donné qu'elle n'interagit pas avec la radiation, les inhomogénéités dans la distribution de matière sombre commencent à croître rapidement à une époque très proche de la fin du découplage en formant ce que l'on appelle des halos de matière sombre. Ces halos croissent de manière hiérarchique par fusion d'entités de petite taille en formant des puits de potentiel gravitationnel de plus en plus forts.

### 1.2.3 Croissance hiérarchique

En même temps que les halos de matière noire fusionnent entre eux de manière hiérarchique, les baryons tombent dans les puits de potentiel formés par ceux-ci, perdent du moment angulaire par friction puis par radiation et s'effondrent pour former les premières étoiles. Les photons des premières étoiles réionisent le milieu et donnent fin à la période « sombre ». Les halos de matière sombre continuent à fusionner en provoquant la fusion des concentrations de baryons à l'intérieur qui forment donc des les premières galaxies et amas de galaxies.

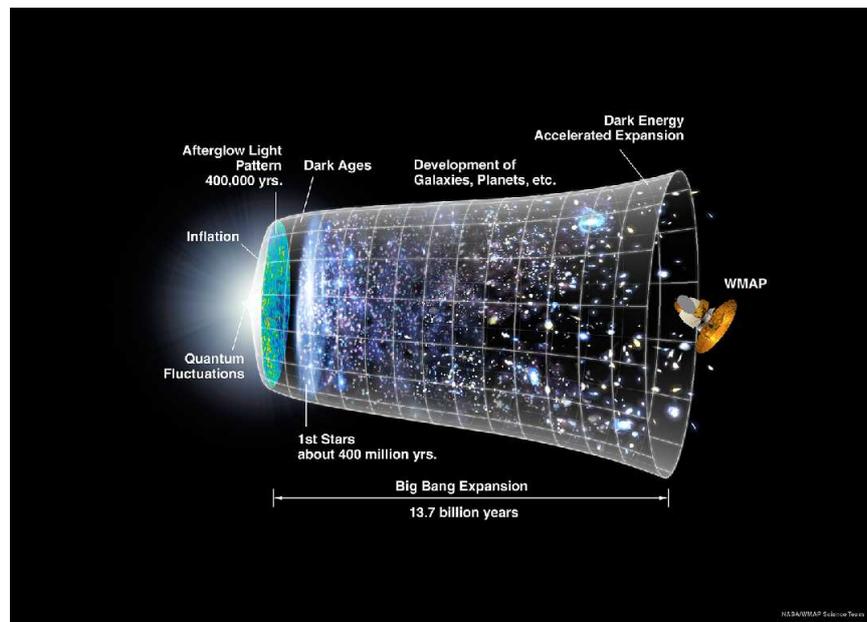


FIG. 1.4 – Le modèle du Big-Bang (c.f. texte pour plus de détails) [Source : NASA]

### 1.2.4 Un Univers modélisé

C'est ainsi, présenté d'une façon extrêmement simplifiée, que l'on comprend aujourd'hui la formation de notre Univers. Mais qu'est-ce qui nous prouve que ce scénario est réaliste ?

On est désormais capable actuellement, grâce en particulier aux progrès informatiques, de réaliser des simulations numériques qui utilisent de l'ordre de  $10^{10}$  particules (Springel *et al.*, 2005). On peut donc simuler ce scénario de croissance hiérarchique des halos de matière noire à plusieurs échelles physiques et prédire ainsi une distribution actuelle de matière dans l'Univers, directement comparable aux observations. Il en résulte un accord indiscutable entre la distribution de galaxies prédite par les simulations et les observations. La matière sombre non-collisionnelle ainsi qu'une constante cosmologique non nulle s'avèrent donc des ingrédients indispensables pour expliquer la formation de l'Univers.

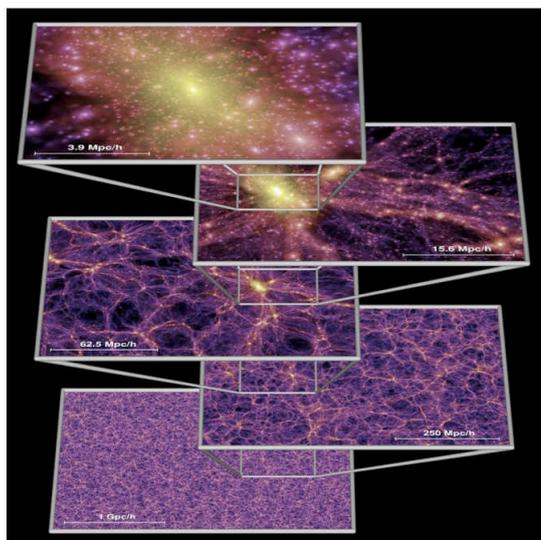


FIG. 1.5 – Modélisation du champ de densité de matière sombre, d'après Springel *et al.* (2005).

Le modèle  $\Lambda$ CDM est, malgré son succès, loin d'être parfait. Comme il a été dit précédemment, nous ignorons la nature de ses deux principales composantes (matière et énergie sombre). Comprendre les ingrédients reste l'un des enjeux majeurs dans l'avenir du modèle. De plus, les simulations ont certainement du mal à reproduire les observations aux petites échelles (galaxies ou amas). La plupart de ces difficultés apparaissent lorsque la physique des baryons commence à jouer un rôle important.

En effet, l'inclusion des baryons nécessite une physique beaucoup plus complexe du fait de son interaction avec le rayonnement électromagnétique en particulier. Par conséquent, introduire de façon réaliste cette physique dans les modèles est le

problème principal à résoudre pour faire le pont entre les observations et la théorie et pouvoir ainsi valider ou rejeter définitivement le modèle  $\Lambda$ CDM. Les modèles semi-analytiques (Kauffmann *et al.*, 1993; Cole *et al.*, 2000; Croton *et al.*, 2006) tentent de faire ce travail : ces codes font intervenir des baryons dans les simulations de croissances de structures en utilisant des lois simples de formation d'étoiles. Les prédictions restent cependant limitées lorsqu'on les compare aux observations.

### 1.3 Histoire des baryons : Formation et évolution de la séquence de Hubble

La matière baryonique ne représente donc que  $\sim 5\%$  du contenu énergétique de l'Univers et se trouve essentiellement au sein des galaxies sous forme de gaz ou d'étoiles. C'est cependant la seule composante à laquelle nous avons directement accès en tant qu'observateurs car c'est la seule à émettre de la lumière. Il s'agit par chance (ou pas) de la partie plus complexe à appréhender du fait de ses nombreuses interactions avec l'environnement et de la multitude de processus physiques qu'elle implique.

Je passe en revue donc par la suite nos principales connaissances sur la formation et l'évolution des baryons dans l'Univers, d'un point de vue observationnel et leurs implications dans les modèles.

#### 1.3.1 Une distribution de galaxies bimodale : la séquence de Hubble

Regardons d'abord comment se structurent les baryons de l'Univers aujourd'hui. Observées dans le visible, les galaxies abritant les baryons présentent essentiellement deux composantes appelées *bulbe* et *disque*. Le bulbe domine les parties centrales de la galaxie et a une forme presque sphérique. Le disque se situe dans les parties externes et est plutôt plat et circulaire. En fonction de la proportion entre le bulbe et le disque au sein de la galaxie on peut définir ainsi deux grands types morphologiques :

- si la galaxie ne présente pas de disque du tout, elle est appelée galaxie elliptique (E) ou galaxie dite précoce. Elle est alors classée en fonction de son ellipticité  $e = (a - b)/a$  où  $a$  et  $b$  sont respectivement les grand et petit axes de l'ellipse formée par la projection de la galaxie dans le plan du ciel. Ainsi une galaxie notée  $En$  a un rapport d'axes  $(b/a) = (1 - n/10)$ ,  $n$  pouvant varier de 0 à 7 ;
- si la galaxie présente un disque elle est alors appelée galaxie spirale (S) ou galaxie tardive. Elle est classée en fonction de l'importance du bulbe dans la partie centrale, pouvant aller de Sa (bulbe important) à Sd (bulbe inexistant). On dit que la galaxie est de plus en plus tardive au fur et à mesure que son bulbe diminue.

- Une dernière classe minoritaire de galaxies dites irrégulières est utilisée pour décrire les objets n’entrant dans aucune des ces deux classes.

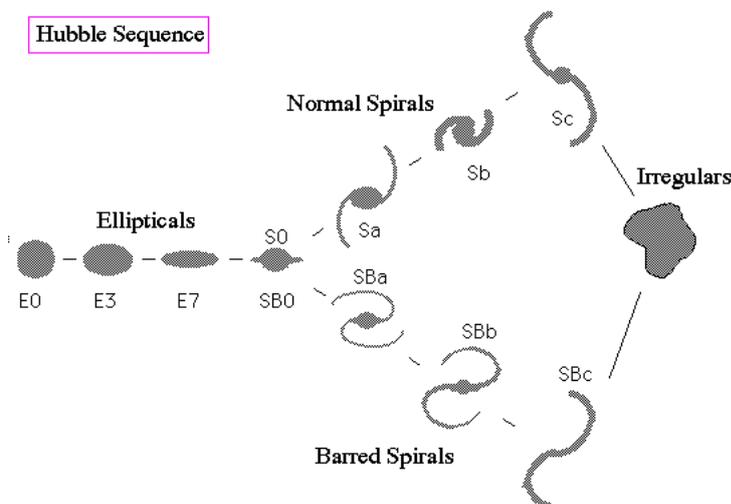


FIG. 1.6 – Schéma illustratif de la séquence de Hubble [Source : <http://www.hopkins.k12.mn.us/>]

Cette classification de galaxies est connue sous le nom de séquence de Hubble (Hubble, 1926) (Fig. 1.6). Même si à l’origine il s’agissait d’une description purement phénoménologique, il s’est avéré que ces deux types de galaxies présentent également des propriétés physiques très différentes.

### 1.3.1.1 Propriétés des bulbes (ou galaxies elliptiques)

La brillance de surface des galaxies de type elliptique est bien décrite par un profil appelé de *de Vaucouleurs* (de Vaucouleurs, 1959) :

$$I(r) = I_e \exp\left[-7.6697 \cdot \frac{r^{0.25}}{r_e} - 1\right] \quad (1.1)$$

où  $r_e$  est le rayon effectif, rayon à l’intérieur duquel se trouve la moitié du flux de la galaxie et  $I_e$  la brillance de surface moyenne. La luminosité totale est obtenue par intégration pour tout rayon  $r$  :  $L_{tot} = 22.7 I_e r_e^2$ .

Ces galaxies sont pauvres en gaz, l’essentiel de leur masse étant sous forme d’étoiles évoluées de population II. Ceci leur confère des couleurs rouges ( $B - V \sim 0.8$ ) et implique que l’essentiel de leur activité de formation stellaire s’est éteinte. Du point de vue de leur dynamique on a pendant longtemps cru que leur forme oblongue était la conséquence d’une rotation rapide. Aujourd’hui on sait que ce n’est pas le cas, le mouvement des étoiles dans les galaxies elliptiques est de nature aléatoire avec une dispersion de vitesses caractéristique ( $\sigma_0$ ). Il existe pour ces galaxies, une relation

empirique liant le rayon effectif ( $r_e$ ), la dispersion de vitesses ( $\sigma_0$ ) et la brillance de surface moyenne ( $I_e$ ) qui décrit ce que l'on appelle le plan fondamental des galaxies elliptiques :

$$r_e \propto \sigma_0^{1.49} I_e^{-0.86} \quad (1.2)$$

### 1.3.1.2 Propriétés des disques

Les galaxies de type disque sont bien décrites par un profil de luminosité exponentiel (Freeman, 1970) :

$$\Sigma(r) = \Sigma_0 \exp\left[-\frac{r}{r_d}\right] \quad (1.3)$$

avec  $R$ , la distance radiale au centre de la galaxie et  $\Sigma_0$  et  $r_d$  des constantes caractéristiques. Une intégration sur le rayon  $R$  donne la luminosité totale de la galaxie :  $L_{tot} = 2\pi\Sigma_0 r_d^2$ .

Ces galaxies sont également riches en gaz, la profondeur optique dans les parties centrales étant de l'ordre de  $\tau = (1-2)$  dans la bande B. Le contenu en gaz (rapport  $M_{\text{gas}}/M_{\text{total}}$ ) augmentant au fur et à mesure que l'on considère des types morphologiques tardifs. Ceci implique que ces galaxies forment des étoiles activement. En particulier, les disques présentent souvent des structures sous forme de bras spiraux, représentant des sur-densités en brillance où se produit l'essentiel de la formation stellaire (d'où le nom de *galaxie spirale* qu'on leur accorde souvent). Certaines de ces galaxies présentent également des *barres* (Fig. 1.6) que l'on croit jouer un rôle important dans l'évolution des ces objets. Leur lumière est ainsi dominée par celle provenant des étoiles jeunes et leurs couleurs sont donc plus bleues ( $B - V \sim 0.6$ ) que celles des galaxies elliptiques. Du point de vue de la dynamique, les étoiles dans la galaxie sont animées d'un mouvement de rotation circulaire rapide qui assure l'équilibre de la galaxie à grande échelle. L'étude des vitesses de rotation en fonction de la distance radiale a permis entre autres, comme décrit auparavant, de mettre en évidence la présence d'un halo de matière noire. Il existe, de plus, des relations simples entre la luminosité de la galaxie et la vitesse maximale de rotation. En effet, en partant simplement de la relation entre la masse contenue à l'intérieur d'un rayon  $R$  et la vitesse de rotation :  $M \propto Rv_{max}^2/G$  et en supposant que le rapport entre masse et luminosité ( $M/L$ ) est constant en première approximation on en déduit que  $L \propto v_{max}^2 R$ . Si on suppose finalement que toutes les galaxies spirales ont approximativement la même brillance de surface centrale, alors le rapport  $L/R^2$  doit rester constant. On en déduit donc une relation entre la luminosité et la vitesse maximale de rotation :  $L \propto v_{max}^4$  qui peut s'écrire comme une relation entre la magnitude absolue et la vitesse de rotation :

$$M_B = -\alpha \log(v_{max}) + \beta \quad (1.4)$$

Cette relation appelée relation de Tully-Fischer (Tully & Fisher, 1977) est une caractéristique fondamentale des galaxies disque qui a été largement vérifiée dans l'Univers local.

### 1.3.1.3 Bimodalité

L'Univers local présente donc une dichotomie dans la distribution des galaxies : d'un côté des galaxies elliptiques, avec des étoiles vieilles, pauvres en gaz et soutenues par la dispersion des vitesses de leurs étoiles ; de l'autre, des galaxies spirales avec des étoiles jeunes, riches en gaz, supportées par la révolution quasi circulaire des étoiles autour de la galaxie. En nombre les galaxies spirales représentent une majorité des galaxies de l'Univers local :  $\sim 75\%$  contre  $\sim 25\%$  de galaxies elliptiques. Cependant, les bulbes (galaxies elliptiques plus bulbes de galaxies spirales) contiennent la majorité de la masse baryonique ( $\sim 63\%$ ).

TAB. 1.1 – Résumé des paramètres des galaxies de l'Univers local pour les deux « extrêmes » de la séquence de Hubble.

Paramètre	E	Sd	Variation
Luminosité B : $M_B$	-15 – -23	-15 – -20	\
Importance du bulbe : $L_B/T_{tot}$	$\sim 1$	$\sim 0$	\
Diamètre : $D_{25}$ (kpc)	1-200	0.5-50	\
Couleur : (B-V)	$\sim 0.5$	$\sim 0.8$	rouge $\rightarrow$ bleu
Gaz : $M_{gaz}/M_{total}$	–	0.25	/
Dynamique : $V_{rot}/\sigma$	$\sim 0.01$	$> 1$	/

Au vu de l'Univers quelques centaines de milliers d'années après le Big-Bang (tel que révèle le CMB) il est clair que cette dichotomie n'a pas toujours été présente et qu'elle s'est construite au cours du temps. Ces deux familles d'objets ont dû suivre des chemins évolutifs différents. Un des enjeux majeurs de la cosmologie observationnelle est de répondre aux questions : *Quand et comment cette bimodalité est-elle apparue ? Les baryons suivent-ils une croissance hiérarchique comme la matière noire ? L'effondrement des baryons dans les halos de matière sombre peut produire un disque sous l'effet de la rotation puis celui-ci peut croître par fusions successives. Qu'en est-il des galaxies elliptiques ?* Leur formation et évolution demeure toujours une question ouverte. L'homogénéité dans la distribution de leurs propriétés a pendant longtemps fait penser qu'elles s'étaient toutes formées pendant une courte période de temps suite à une formation stellaire intense. Est-il possible de vérifier ces hypothèses par les observations ?

Les réponses nécessitent de tenir compte d'un nombre élevé de variables toutes reliées entre elles. Il s'agit de faire le lien entre le cadre dans lequel ces galaxies évoluent (le modèle  $\Lambda$ CDM), leurs propriétés internes (morphologie, luminosité, histoire de formation stellaire) et leurs propriétés statistiques (densité, luminosité globale) et de suivre toutes ces relations au cours du temps afin de réalimenter les modèles pour qu'ils puissent reproduire les observations.

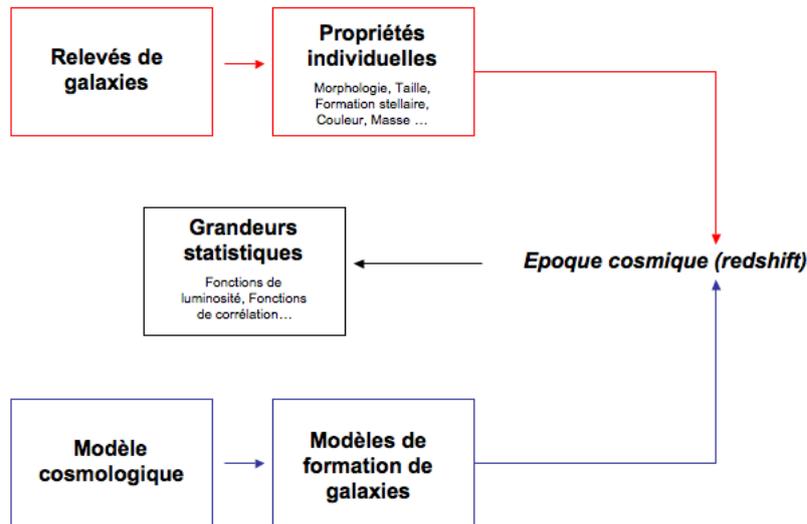


FIG. 1.7 – Schéma du fonctionnement d'un relevé de galaxies : les données observationnelles (rouge) sont confrontées aux prédictions des modèles (bleu) à travers des grandeurs statistiques (noir).

### 1.3.2 Archéologie galactique : l'époque des grands relevés

L'étude de la construction de la séquence de Hubble d'un point de vue observationnel relève d'un travail archéologique : remonter le temps à la recherche des précurseurs des galaxies actuelles afin de tracer leur chemin évolutif. Par chance, la vitesse finie de la lumière nous permet de figer le temps. Il suffit pour cela de regarder des objets éloignés de nous. On est capable ainsi de réaliser des « coupes » de l'Univers à différentes époques (i.e. différents *redshifts*) et d'analyser le spectre lumineux provenant d'une époque donnée. C'est ce que l'on appelle *les relevés de galaxies*. On remonte ainsi aux propriétés physiques des objets peuplant l'Univers à cette époque et, en mesurant des grandeurs statistiques, il est possible de faire le lien entre les observations et les prédictions des modèles. Le schéma 1.7 illustre ce procédé.

Un procédé facile, semblerait-il, et pourtant les relevés de galaxies représentent l'un des plus grands défis de l'astronomie moderne. L'objectif final est en effet d'obtenir des indicateurs statistiques qui puissent être prédits par les modèles. Qui dit « statistique » dit « grand nombre d'objets » où ce qui revient au même, « une grande aire ». On doit donc pouvoir observer des objets très lointains mais aussi sur des grandes surfaces pour ne pas être soumis à ce qui est appelé la *variance cosmique*. Cela n'est pas tout : pour arriver à des statistiques globales, on doit d'abord mesurer les propriétés individuelles des objets. Ceux-ci, très lointains, sont peu lumineux et petits (angulairement parlant) dans le ciel. Leur détection et leur analyse constituent également un défi pour les technologies actuelles. De plus, afin de comprendre la complexité de tous les phénomènes physiques ayant lieu, on doit

avoir une vision la plus large possible du spectre lumineux, c'est à dire observer à plusieurs longueurs d'onde et donc disposer d'un grand nombre d'instruments dédiés.

Un relevé se caractérise donc généralement par sa profondeur (quels sont les objets les plus faibles détectés), le volume sondé (époque cosmique et champ observé) et l'étendue du spectre lumineux analysé (nombre de filtres employés). Ils sont souvent la résultante d'une collaboration internationale mettant en jeu un grand nombre de personnes et d'instruments. Le tableau 1.3.2 cite quelques uns des plus récents.

TAB. 1.2 – Caractéristiques principales de quelques relevés de galaxies (photométriques)

Nom	Aire	Mag. Lim.	Filtres
2MASS	41252	$K < 16.1$	JHK
SDSS	8000	$I < 21.3$	ugriz
UKIDSS	400	$K < 20.2$	JHK
GOODS	0.05	$I < 26.8$	Bviz
COSMOS	2	$I < 27$	X→Radio
CFHTLS (deep)	4	$I < 26.5$	ugriz
HDF	0.001	$I < 28.0$	UBVIH
UDF	0.003	$I < 29.0$	BViz
ALHAMBRA	4	$I < 25.0$	3500Å–22000Å

*Qu'avons nous appris de ces relevés concernant la naissance et évolution des galaxies et en particulier de la séquence de Hubble ? Quels sont les enjeux pour l'avenir ?* Je passe en revue par la suite les principales avancées dans ce domaine au cours des dernières années ainsi que les implications quant aux modèles de formation. Il ne s'agit que d'une revue très brève des principaux points afin de situer le cadre dans lequel se déroule cette thèse. Pour une revue extensive, le lecteur intéressé pourra consulter les références indiquées.

### 1.3.3 Quand les galaxies se forment-elles ? Histoire de la formation stellaire

#### 1.3.3.1 Taux de formation stellaire

Une première approche pour identifier les principales époques de formation des galaxies est d'étudier quand le gaz se transforme en étoiles au cours de l'histoire de l'Univers. En définitive on souhaite répondre à la question : quand les galaxies ont-elles formé leur étoiles ? Cela peut être mesuré par le taux de formation stellaire ( $\rho_{SFR}$ ) qui donne une valeur moyenne de la masse de gaz transformée en étoiles par unité de temps et de volume.

L'estimation de cette grandeur se fait par des méthodes indirectes comme la mesure du continuum UV, directement lié à l'émission d'étoiles OB jeunes, la mesure du flux dans l'infrarouge lointain qui provient de la poussière autour d'étoiles en formation ou la raie en émission  $H_\alpha$  qui mesure le flux issu de la recombinaison du gaz ionisé par le flux UV.

Les premières mesures du taux de formation stellaire globale en fonction du *redshift* datent de 1996 (Lilly *et al.*, 1996; Ellis *et al.*, 1996; Madau *et al.*, 1996). Globalement on s'aperçoit que celui-ci est maximal entre  $z=1$  et  $z=2$  et décroît fortement jusqu'à nos jours. Hopkins (2004) et Hopkins & Beacom (2006) ont rassemblé l'ensemble des mesures réalisées par des méthodes différentes. La figure 1.8 résume les résultats. Même si à première vue elle peut sembler un peu confuse, on observe des tendances assez nettes : d'une part une croissance systématique jusqu'à  $z \sim 1$  ajustée par une loi de la forme :  $\rho_{SFR}(z) \propto (1+z)^{3.4}$ . Il existe ensuite un plateau quelque part entre  $z = 2$  et  $z = 4$  puis une diminution possible au delà de  $z \sim 4$  quoique il existe de fortes incertitudes sur les mesures à des *redshifts* si élevés.

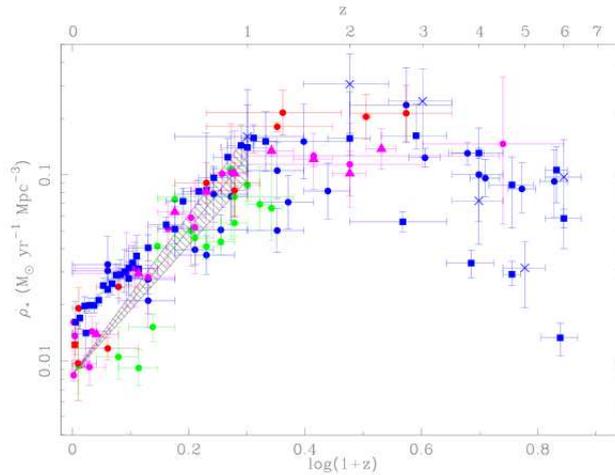


FIG. 1.8 – Bilan des dernières mesures de l'histoire de la formation stellaire dans l'Univers. Les cercles sont des données de Hopkins (2004), le code de couleurs indique la méthode employée : bleu : continuum UV, vert : [OII], rouge :  $H_\alpha$ , magenta : sub-mm et radio. Les nouvelles données de Hopkins & Beacom (2006) sont représentées par des triangles. Les droites montrent le meilleur ajustement pour  $z < 1$ .

L'étude de l'évolution du taux de formation stellaire moyen permet d'identifier les principales époques d'activité des galaxies. Le constat plus important au vu de la figure 1.8 est que la formation stellaire dans l'Univers est un processus étendu dans le temps. L'idée, longtemps soutenue par les observateurs, comme quoi la plupart des galaxies auraient formé leurs étoiles dans un court intervalle de temps doit être abandonnée. Les modèles semi-analytiques ont fait un travail de « rattrapage » au fur et à mesure que les données observationnelles se sont améliorées (Baugh *et al.*, 1998, 2005).

### 1.3.3.2 Fonction de luminosité

**Définition** Un deuxième indicateur souvent utilisé est la fonction de luminosité. Elle s'est avérée être un outil puissant et direct pour étudier la construction de la séquence de Hubble au cours du temps car elle peut être directement comparée aux modèles d'évolution. Elle compte tout simplement le nombre de galaxies dans un certain intervalle de luminosité par unité de volume et représente donc un des liens plus directs entre les observations et les modèles de formation. A première vue son estimation paraît simple, mais il faut cependant tenir compte d'un grand nombre de corrections en particulier à cause d'effets de sélection de l'échantillon en question. [Schechter \(1976\)](#) a montré que ces fonctions observées sont généralement bien décrites de façon empirique par une fonction analytique de la forme (*fonction de Schechter*) :

$$\Phi(L) \frac{dL}{L} = \Phi^* \left( \frac{L}{L^*} \right)^{-\alpha} \exp\left(-\frac{L}{L^*}\right) \frac{dL}{L^*} \quad (1.5)$$

Le paramètre  $\Phi^*$  est une normalisation globale,  $\alpha$  donne la pente de la fonction de luminosité aux faibles magnitudes et  $L^*$  est la luminosité caractéristique au delà de laquelle le nombre de galaxies décroît exponentiellement.

**Fonctions de luminosités locales** Les fonctions de luminosité de l'Univers d'aujourd'hui sont maintenant bien connues en plusieurs bandes photométriques grâce en particulier aux relevés SDSS ([Blanton et al., 2001](#)) et 2dF ([Norberg et al., 2002](#)).

La reproduction des fonctions de luminosité actuelles observées à l'aide de modèles semi-analytiques reste cependant un problème majeur ([Benson et al., 2003](#); [Croton et al., 2006](#); [Lucia et al., 2006](#)). En particulier, ils prédisent un excès de formation stellaire dans les galaxies plus massives ce qui impliquerait un nombre trop important de galaxies brillantes bleues et un excès de galaxies rouges faibles ([Bower et al., 2006](#)). Ces problèmes ont conduit les modélisateurs à réfléchir à des mécanismes de régulation de la formation stellaire permettant de déclencher ou d'arrêter la formation d'étoiles dans une galaxie :

- *feedback*<sup>3</sup> par réionisation : le flux provenant des étoiles naissantes ionise le milieu et retarde la formation de nouvelles étoiles ;
- *feedback* par supernovae : l'explosion des étoiles en supernovae chauffe le milieu interstellaire, réduisant le taux de formation stellaire ;
- *feedback* par l'activité du noyau actif : l'activité des noyaux actifs de la galaxie peut permettre un transfert d'énergie ayant une influence sur la formation stel-

---

<sup>3</sup>Le terme français « rétroaction » n'étant pas couramment employé nous avons décidé d'utiliser le terme anglais

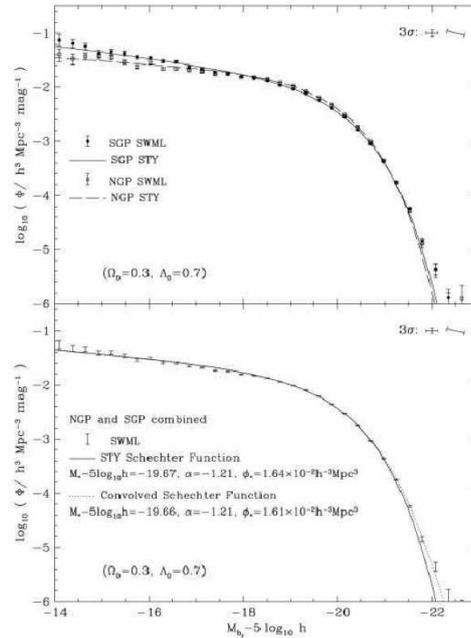


FIG. 1.9 – Fonction de luminosité en bande  $b_J$  obtenue à partir des données 2dF (Norberg *et al.*, 2002). (Haut) Comparaison des résultats entre les pôles galactiques nord et sud. (Bas) Combinaison des 2 hémisphères et ajustement par une fonction de Schechter.

laire. Ce mécanisme ayant été identifié plus récemment est par conséquent moins bien exploré mais il commence à prendre une place importante. (cf § 1.3.5)

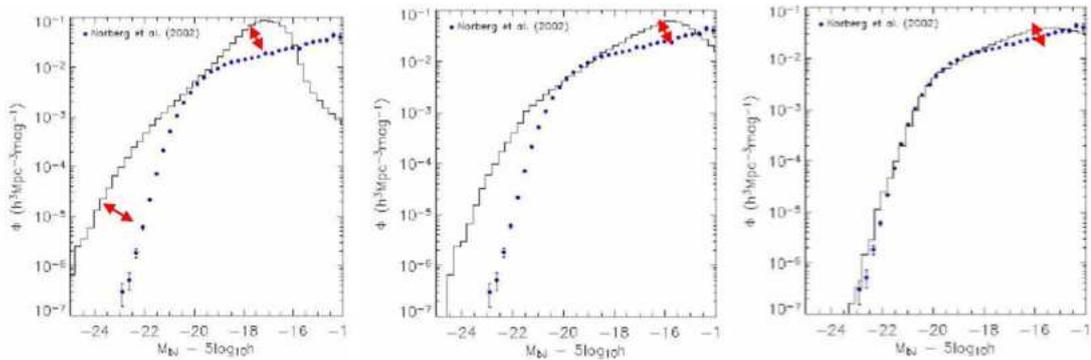


FIG. 1.10 – Effet de plusieurs formes de réalimentation sur la fonction de luminosité. De gauche à droite : pas de réalimentation, réalimentation par supernovae uniquement, réalimentation par supernovae, réionisation et AGN (Norberg *et al.*, 2002).

Les fonctions de luminosité peuvent également être étudiées par type morphologique, c'est à dire compter le nombre de galaxies de chaque type (spiral, elliptique)

par intervalle de luminosité. Ce travail requiert une étape supplémentaire : la classification morphologique de toutes les galaxies. Dans l'Univers local, ceci a pu être réalisé encore une fois grâce au relevé SDSS. On constate, entre autres, que les galaxies précoces dominent les parties brillantes des fonctions de luminosité alors que les types tardifs sont plus représentés aux luminosités plus faibles.

**Evolution** Plus récemment et grâce aux relevés de galaxies profonds, il est devenu possible d'étudier les fonctions de luminosité à différentes époques de l'Univers afin de suivre leur évolution. Il est particulièrement intéressant de les étudier pour les différents types morphologiques également. Il s'agit cependant d'un défi « technique » car il est nécessaire d'avoir un nombre important de galaxies, d'avoir leur *redshift* et de déterminer sans trop d'incertitude leur morphologie à grand  $z$ . *Ilbert et al. (2006)* ont réalisé ce travail dans le cadre du consortium VVDS jusqu'à  $z \sim 1.2$  en utilisant des données du télescope spatial. Ces études semblent indiquer que l'essentiel des galaxies elliptiques est en place à  $z \sim 1$  et qu'elles n'évoluent alors plus que passivement. Les galaxies plus tardives présentent quant à elles une évolution plus marquée témoignant d'un processus de formation plus continu et étendu dans le temps. En outre, ces études ont permis d'isoler une population de galaxies peu représentée dans l'Univers local, des galaxies présentant une morphologie de type elliptique mais avec des couleurs bleues témoignant d'une activité de formation stellaire importante. La nature de ces objets n'est toujours pas claire.

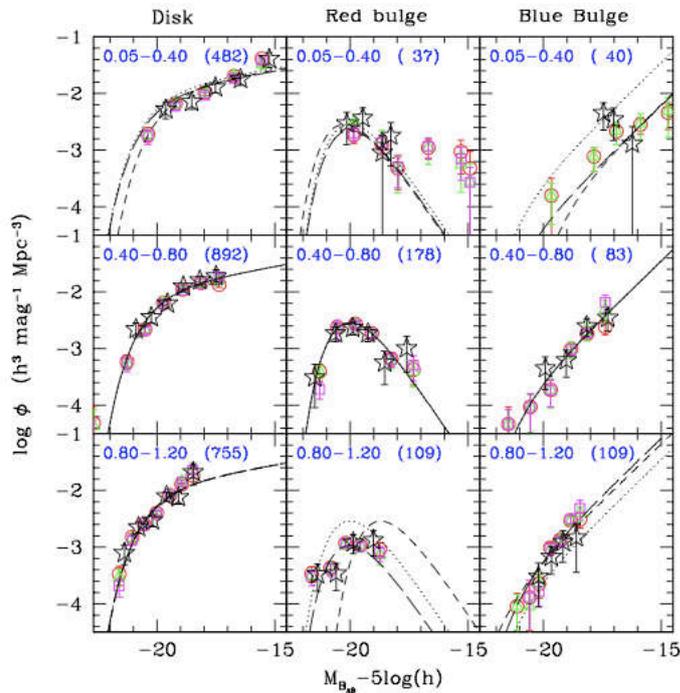


FIG. 1.11 – Evolution de la fonction de luminosité par type morphologique d'après *Ilbert et al. (2006)*.

## 1.3.3.3 Couleur-magnitude

Le diagramme couleur-magnitude est également un outil intéressant pour l'étude de l'évolution de la formation stellaire. Ces diagrammes représentent les galaxies dans un plan  $M_B$  vs.  $U - B$ . Dans l'Univers local les galaxies apparaissent alors bien séparées entre une *séquence rouge* et un *nuage bleu* (e.g. [Strateva et al., 2001](#); [Baldry et al., 2004](#), Fig. 1.12). La séquence rouge est essentiellement peuplée de galaxies de type elliptique ayant achevé leur formation stellaire tandis que le nuage bleu contient des galaxies de type tardif. Le diagramme traduit donc la bimodalité dans la distribution de galaxies dans l'Univers local. L'étude de celui-ci pour des différentes époques permet donc de déterminer les périodes d'apparition de cette bimodalité. [Conselice et al. \(2007\)](#) ont étudié l'évolution de la relation couleur-magnitude pour des galaxies massives ( $M > 10^{11} M_{sol}$ ) jusqu'à  $z \sim 1.2$ . On s'aperçoit que le nombre de galaxies dans la séquence rouge diminue quand le *redshift* augmente. Alors qu'à  $z \sim 0$  presque la totalité des galaxies massives se trouve dans la séquence, la fraction diminue fortement à grand  $z$ , témoignant d'une formation stellaire active (Fig. 1.13). Ceci suggère que la séquence rouge se forme à partir de galaxies bleues qui achèvent leur formation stellaire et est en accord avec la présence de cette population d'elliptiques bleues à grand  $z$  mise en évidence par l'étude des fonctions de luminosité.

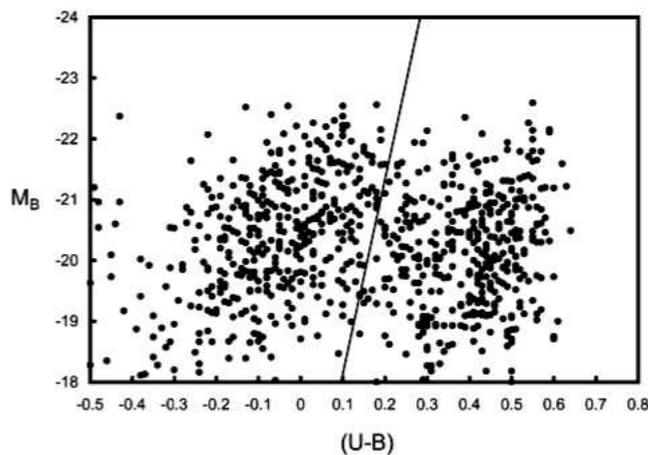


FIG. 1.12 – Diagramme couleur magnitude dans l'Univers local d'après [van den Bergh \(2007\)](#) : la distribution des couleurs est bimodale. La majorité des galaxies elliptiques se situent dans une *séquence rouge* alors que les galaxies spirales forment un *nuage bleu*. Les données sont issues du catalogue : *The Revised Shapley-Ames Catalog of Bright Galaxies* ([Sandage & Tammann, 1981](#))

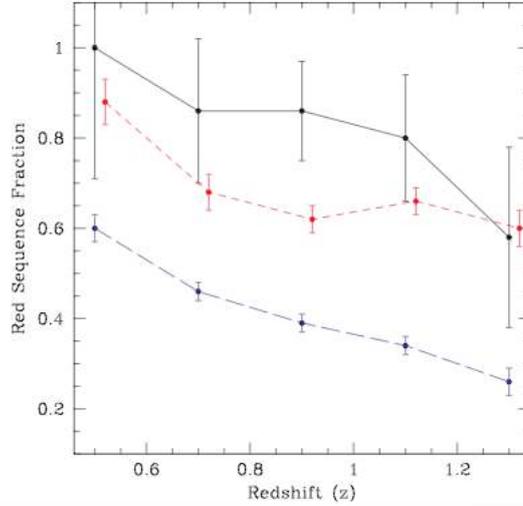


FIG. 1.13 – Evolution de la fraction de galaxies massives présentes dans la séquence rouge en fonction du décalage spectral (Conselice *et al.*, 2007). La courbe noire montre les galaxies ayant des masses  $M_* > 10^{11.5} M_\odot$ , la courbe rouge désigne celles ayant des masses  $10^{11} M_\odot < M_* < 10^{11.5} M_\odot$  et la courbe bleue celles ayant des masses  $10^{10.5} M_\odot < M_* < 10^{11} M_\odot$ .

### 1.3.4 Où les galaxies acquièrent-elles leur morphologie ? Le rôle de l’environnement

Dressler (1980) mesurèrent pour la première fois la relation entre le type morphologique des galaxies et l’environnement dans lequel elles résident en regardant la fraction de galaxies d’un type morphologique donné en fonction de la densité projetée (relation  $T - \Sigma$ ). Ils trouvèrent une forte dépendance du mélange morphologique avec l’environnement ; les galaxies de type elliptique étant plus abondantes dans les milieux denses. Depuis cette relation a été largement étudiée et confirmée, le cas extrême étant les amas de galaxies, dont la région centrale est essentiellement dominée par des galaxies elliptiques.

Ces observations donnèrent lieu à deux interprétations différentes du rôle de l’environnement dans la construction de la séquence de Hubble. Soit cette tendance n’est qu’un reflet des conditions initiales de formation (*nature*) ou alors il s’agit d’une manifestation des mécanismes d’évolution (*nurture*). Dans la première hypothèse, les galaxies se formant dans des environnements plus denses consommeraient leur gaz plus vite que celles se formant dans des environnements à plus faible densité. La ségrégation serait donc établie à la naissance. Dans la deuxième hypothèse, les galaxies seraient transformées de spirales en elliptiques par des processus induits par l’environnement.

Des travaux utilisant le télescope spatial Hubble à la fin des années 90, mirent en évidence cependant une évolution remarquable de la relation  $T - \Sigma$  entre les *redshifts*

$z=0$  et  $z=0.5$  (Couch *et al.*, 1998; Dressler *et al.*, 1997). Ces résultats soutiennent fortement l'hypothèse de type *nurture*. Des études plus récentes (Smith *et al.*, 2005; Postman *et al.*, 2005) ont révélé que la relation actuelle est essentiellement en place à  $z \sim 1$  (c'est-à-dire que l'on trouve effectivement plus d'elliptiques dans les régions denses) mais que la fraction de galaxies elliptiques dans les environnements denses a doublé depuis cette époque. Ceci suggère une transformation continue, dépendante de la densité, de galaxies spirales en elliptiques.

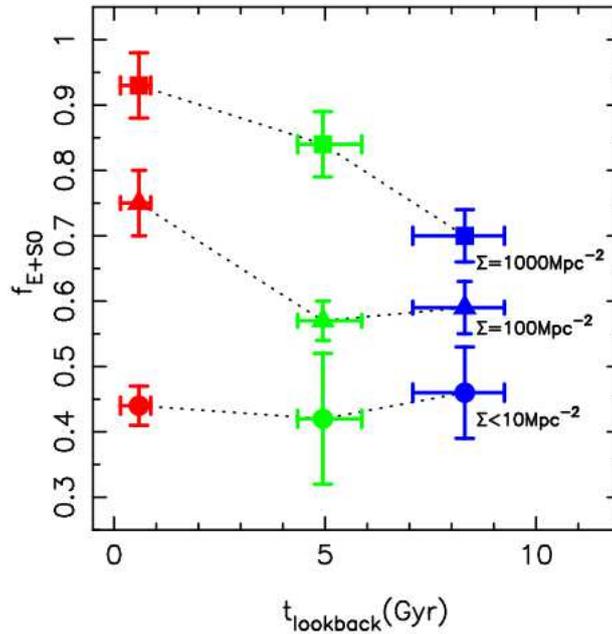


FIG. 1.14 – Evolution de la fraction de galaxies elliptiques en fonction de la densité projetée au cours du temps (Smith *et al.*, 2005).

### 1.3.5 Comment les galaxies acquièrent-elles leur morphologie ? Interactions, *Downsizing*, noyaux actifs

La forte dépendance avec la densité suggère que l'interaction de galaxies est un moteur majeur de la transformation morphologique. Les premières simulations N-corps visant à déterminer le rôle de la fusion des galaxies datent des années 70 (Toomre & Toomre, 1972). Il est maintenant bien admis que la fusion de deux disques produit un système de type sphéroïdal et que les étoiles acquièrent une dispersion de vitesse importante. D'un point de vue observationnel, la découverte de coeurs multiples dans les galaxies elliptiques locales (Davies *et al.*, 2001) révèle qu'elles ont subi des interactions récentes.

D'importants efforts ont donc été effectués afin d'étudier l'évolution du taux de fusion de galaxies au cours du temps. Les images HST suggèrent en effet une aug-

mentation de la fréquence de fusions avec le *redshift* (LeFèvre *et al.*, 2000), quoique l'estimation reste soumise à de fortes incertitudes (Bundy *et al.*, 2004). Déterminer si deux galaxies sont en interaction requiert une connaissance précise de leur *redshift*. De plus il est difficile d'identifier les états pré et post fusion. Des études sur la morphologie, ont révélé également que le nombre de galaxies présentant une morphologie irrégulière augmente significativement lorsque l'on augmente le *redshift* (Abraham *et al.*, 1996). Ces constats pointent vers une vision hiérarchique de l'évolution de la matière baryonique, les grandes structures se formant par fusions successives de plus petites entités.

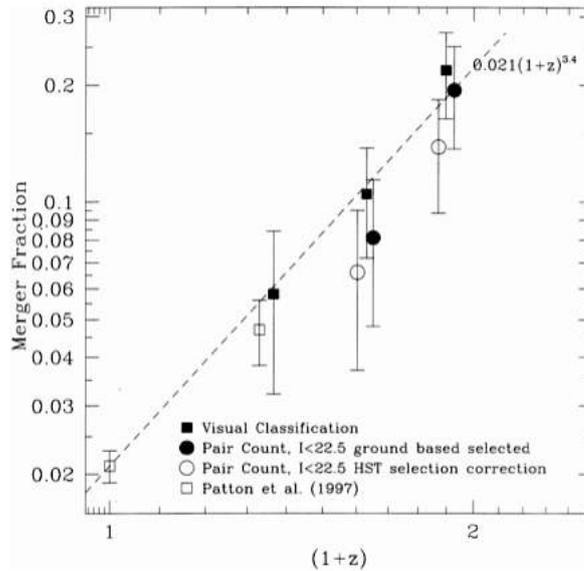


FIG. 1.15 – Evolution du taux de fusions, d'après LeFèvre *et al.* (2000).

En étudiant le taux de *formation stellaire spécifique* (rapport entre la masse stellaire de la galaxie  $M_{star}$  et le taux de formation stellaire  $SFR$ ), Cowie *et al.* (1996), Brinchmann & Ellis (2000) et plus récemment Juneau *et al.* (2005) ont trouvé une tendance surprenante au premier abord : entre les *redshifts*  $z = 0.5 - 1.5$ , les galaxies plus massives ne forment presque plus d'étoiles alors que les galaxies moins massives semblent être plus actives. Ceci implique que les galaxies massives forment plus rapidement leurs étoiles et que donc l'activité de formation stellaire se déplace au cours du temps des systèmes de plus grande masse aux systèmes moins massifs. Cette croissance de type *anti-hiérarchique* est connue sous le nom de *downsizing* en anglais et a été considérée pendant un temps un problème fondamental du modèle  $\Lambda$ CDM qui prédit une croissance hiérarchique des structures comme exposé précédemment. En réalité cela ne pose pas de véritables problèmes au scénario hiérarchique. Il suffit de trouver un mécanisme qui stoppe l'activité de formation stellaire plus rapidement dans les systèmes massifs. Les mécanismes de *feedback* interviennent à nouveau. En particulier l'activité des AGN semble jouer un rôle fondamental en chauffant le milieu interstellaire et évitant ainsi l'effondrement du

gaz et donc la formation stellaire. Les derniers modèles semi-analytiques (Granato *et al.*, 2000; Lucia *et al.*, 2006; Cimatti *et al.*, 2006), incorporant un système de rétroaction par l'activité du noyau actif arrivent effectivement à reproduire l'effet *downsizing* dans l'activité de formation stellaire.

Plus récemment cependant, (e.g. Bundy *et al.*, 2006) ont mis en évidence un processus de downsizing, non seulement dans la formation stellaire mais également dans l'assemblage de la masse proprement dit. C'est à dire que les galaxies moins massives assembleraient leur masse plus tard que les galaxies plus massives. Ce comportement purement anti-hiérarchique est plus difficile à expliquer par les modèles dans un contexte de formation hiérarchique  $\Lambda$ CDM.

## Bibliographie

- Abraham R.G., van den Bergh S., Glazebrook K., Ellis R.S., Santiago B.X., Surma P., & Griffiths R.E., 1996. The Morphologies of Distant Galaxies. II. Classifications from the Hubble Space Telescope Medium Deep Survey. *Astrophysical Journal Supplement* v.107, 107 : 1. 34
- Astier P., Guy J., Regnault N., Pain R., Aubourg E., Balam D., Basa S., Carlberg R.G., Fabbro S., Fouchez D., Hook I.M., Howell D.A., Lafoux H., Neill J.D., Palanque-Delabrouille N., Perrett K., Pritchet C.J., Rich J., Sullivan M., Taillet R., Aldering G., Antilogus P., Arsenijevic V., Balland C., Baumont S., Bronder J., Courtois H., Ellis R.S., Filiol M., Gonçalves A.C., Goobar A., Guide D., Hardin D., Lusser V., Lidman C., McMahon R., Mouchet M., Mourao A., Perlmutter S., Ripoche P., Tao C., & Walton N., 2006. The Supernova Legacy Survey : measurement of  $\Omega_M$ ,  $\Omega_{\Lambda}$  and  $w$  from the first year data set. *Astronomy and Astrophysics*, 447 : 31. 17
- Baldry I.K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R.H., Nichol R.C., & Szalay A.S., 2004. Quantifying the Bimodal Color-Magnitude Distribution of Galaxies. *ApJ*, 600 : 681. 31
- Baugh C.M., Cole S., Frenk C.S., & Lacey C.G., 1998. The Epoch of Galaxy Formation. *ApJ*, 498 : 504. 27
- Baugh C.M., Lacey C.G., Frenk C.S., Granato G.L., Silva L., Bressan A., Benson A.J., & Cole S., 2005. Can the faint submillimetre galaxies be explained in the  $\Lambda$  cold dark matter model? *MNRAS*, 356 : 1191. 27
- Benson A.J., Bower R.G., Frenk C.S., Lacey C.G., Baugh C.M., & Cole S., 2003. What Shapes the Luminosity Function of Galaxies? *The Astrophysical Journal*, 599 : 38. (c) 2003 : The American Astronomical Society. 28
- Blanton M.R., Dalcanton J., Eisenstein D., Loveday J., Strauss M.A., SubbaRao M., Weinberg D.H., Anderson J.E., Annis J., Bahcall N.A., Bernardi M., Brinkmann

- J., Brunner R.J., Burles S., Carey L., Castander F.J., Connolly A.J., Csabai I., Doi M., Finkbeiner D., Friedman S., Frieman J.A., Fukugita M., Gunn J.E., Hennessy G.S., Hindsley R.B., Hogg D.W., Ichikawa T., Ivezić Ž., Kent S., Knapp G.R., Lamb D.Q., Leger R.F., Long D.C., Lupton R.H., McKay T.A., Meiksin A., Merelli A., Munn J.A., Narayanan V., Newcomb M., Nichol R.C., Okamura S., Owen R., Pier J.R., Pope A., Postman M., Quinn T., Rockosi C.M., Schlegel D.J., Schneider D.P., Shimasaku K., Siegmund W.A., Smee S., Snir Y., Stoughton C., Stubbs C., Szalay A.S., Szokoly G.P., Thakar A.R., Tremonti C., Tucker D.L., Uomoto A., Berk D.V., Vogeley M.S., Waddell P., Yanny B., Yasuda N., & York D.G., 2001. The Luminosity Function of Galaxies in SDSS Commissioning Data. *The Astronomical Journal*, 121 : 2358. (c) 2001 : The American Astronomical Society. 28
- Bower R.G., Benson A.J., Malbon R., Helly J.C., Frenk C.S., Baugh C.M., Cole S., & Lacey C.G., 2006. Breaking the hierarchy of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 370 : 645. 28
- Brinchmann J. & Ellis R.S., 2000. The Mass Assembly and Star Formation Characteristics of Field Galaxies of Known Morphology. *ApJL*, 536 : L77. 34
- Bundy K., Ellis R.S., Conselice C.J., Taylor J.E., Cooper M.C., Willmer C.N.A., Weiner B.J., Coil A.L., Noeske K.G., & Eisenhardt P.R.M., 2006. The Mass Assembly History of Field Galaxies : Detection of an Evolving Mass Limit for Star-Forming Galaxies. *The Astrophysical Journal*, 651 : 120. (c) 2006 : The American Astronomical Society. 35
- Bundy K., Fukugita M., Ellis R.S., Kodama T., & Conselice C.J., 2004. A Slow Merger History of Field Galaxies since  $z = 1$ . *The Astrophysical Journal*, 601 : L123. (c) 2004 : The American Astronomical Society. 34
- Cimatti A., Daddi E., & Renzini A., 2006. Mass downsizing and “top-down” assembly of early-type galaxies. *Astronomy and Astrophysics*, 453 : L29. 35
- Cole S., Lacey C.G., Baugh C.M., & Frenk C.S., 2000. Hierarchical galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 319 : 168. (c) 2000 The Royal Astronomical Society. 21
- Conselice C.J., Bundy K., Trujillo I., Coil A., Eisenhardt P., Ellis R.S., Georgakakis A., Huang J., Lotz J., Nandra K., Newman J., Papovich C., Weiner B., & Willmer C., 2007. The properties and evolution of a K-band selected sample of massive galaxies at  $z \sim 0.4-2$  in the Palomar/DEEP2 survey. *MNRAS*, 381 : 962. 31, 32
- Couch W.J., Barger A.J., Smail I., Ellis R.S., & Sharples R.M., 1998. Morphological Studies of the Galaxy Populations in Distant “Butcher-Oemler” Clusters with the Hubble Space Telescope. II. AC 103, AC 118, and AC 114 at  $Z = 0.31$ . *Astrophysical Journal* v.497, 497 : 188. (c) 1998 : The American Astronomical Society. 33

- Cowie L.L., Songaila A., Hu E.M., & Cohen J.G., 1996. New Insight on Galaxy Formation and Evolution From Keck Spectroscopy of the Hawaii Deep Fields. *Astronomical Journal* v.112, 112 : 839. 34
- Croton D.J., Springel V., White S.D.M., Lucia G.D., Frenk C.S., Gao L., Jenkins A., Kauffmann G., Navarro J.F., & Yoshida N., 2006. The many lives of active galactic nuclei : cooling flows, black holes and the luminosities and colours of galaxies. *Monthly Notices of the Royal Astronomical Society*, 365 : 11. 21, 28
- Davies R.L., Kuntschner H., Emsellem E., Bacon R., Bureau M., Carollo C.M., Copin Y., Miller B.W., Monnet G., Peletier R.F., Verolme E.K., & de Zeeuw P.T., 2001. Galaxy Mapping with the SAURON Integral-Field Spectrograph : The Star Formation History of NGC 4365. *ApJL*, 548 : L33. 33
- de Vaucouleurs G., 1959. Classification and Morphology of External Galaxies. *Handbuch der Physik*, 53 : 275. 22
- Dressler A., 1980. Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *Astrophysical Journal*, 236 : 351. A&AA ID. AAA027.158.034. 32
- Dressler A., Oemler A., Couch W.J., Smail I., Ellis R.S., Barger A., Butcher H., Poggianti B.M., & Sharples R.M., 1997. Evolution since  $Z = 0.5$  of the Morphology-Density Relation for Clusters of Galaxies. *Astrophysical Journal* v.490, 490 : 577. (c) 1997 : The American Astronomical Society. 33
- Ellis R.S., 2007. Observations of the High Redshift Universe. arXiv, astro-ph.
- Ellis R.S., Colless M., Broadhurst T., Heyl J., & Glazebrook K., 1996. Autofib Redshift Survey - I. Evolution of the galaxy luminosity function. *MNRAS*, 280 : 235. 27
- Freedman W.L., Madore B.F., Gibson B.K., Ferrarese L., Kelson D.D., Sakai S., Mould J.R., Kennicutt R.C., Ford H.C., Graham J.A., Huchra J.P., Hughes S.M.G., Illingworth G.D., Macri L.M., & Stetson P.B., 2001. Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. *The Astrophysical Journal*, 553 : 47. (c) 2001 : The American Astronomical Society. 15
- Freeman K.C., 1970. On the Disks of Spiral and so Galaxies. *ApJ*, 160 : 811. 23
- Gerhard O., Kronawitter A., Saglia R.P., & Bender R., 2001. Dynamical Family Properties and Dark Halo Scaling Relations of Giant Elliptical Galaxies. *The Astronomical Journal*, 121 : 1936. (c) 2001 : The American Astronomical Society. 17
- Granato G.L., Lacey C.G., Silva L., Bressan A., Baugh C.M., Cole S., & Frenk C.S., 2000. The Infrared Side of Galaxy Formation. I. The Local Universe in the Semianalytical Framework. *ApJ*, 542 : 710. 35

- Hopkins A.M., 2004. On the Evolution of Star-forming Galaxies. *The Astrophysical Journal*, 615 : 209. (c) 2004 : The American Astronomical Society. 27
- Hopkins A.M. & Beacom J.F., 2006. On the Normalization of the Cosmic Star Formation History. *The Astrophysical Journal*, 651 : 142. (c) 2006 : The American Astronomical Society. 27
- Hubble E.P., 1926. Extragalactic nebulae. *Astrophys. J.*, 64 : 321. 22
- Ilbert O., Lauger S., Tresse L., Buat V., Arnouts S., Fèvre O.L., Burgarella D., Zucca E., Bardelli S., Zamorani G., Bottini D., Garilli B., Brun V.L., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Arnaboldi M., Bolzonella M., Cappi A., Charlot S., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Iovino A., McCracken H.J., Marano B., Marinoni C., Mathez G., Mazure A., Meneux B., Merighi R., Paltani S., Pello R., Pollo A., Pozzetti L., Radovich M., Bondi M., Bongiorno A., Busarello G., Ciliegi P., Mellier Y., Merluzzi P., Ripepi V., & Rizzo D., 2006. The VIMOS-VLT Deep Survey. Galaxy luminosity function per morphological type up to  $z = 1.2$ . *Astronomy and Astrophysics*, 453 : 809. 30
- Juneau S., Glazebrook K., Crampton D., McCarthy P.J., Savaglio S., Abraham R., Carlberg R.G., Chen H.W., Borgne D.L., Marzke R.O., Roth K., Jørgensen I., Hook I., & Murowinski R., 2005. Cosmic Star Formation History and Its Dependence on Galaxy Stellar Mass. *The Astrophysical Journal*, 619 : L135. (c) 2005 : The American Astronomical Society. 34
- Kauffmann G., White S.D.M., & Guiderdoni B., 1993. The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes. *R.A.S. MONTHLY NOTICES* V.264, 264 : 201. 21
- Koopmans L.V.E. & Treu T., 2003. The Structure and Dynamics of Luminous and Dark Matter in the Early-Type Lens Galaxy of 0047-281 at  $z = 0.485$ . *The Astrophysical Journal*, 583 : 606. (c) 2003 : The American Astronomical Society. 17
- LeFèvre O., Abraham R., Lilly S.J., Ellis R.S., Brinchmann J., Schade D., Tresse L., Colless M., Crampton D., Glazebrook K., Hammer F., & Broadhurst T., 2000. Hubble Space Telescope imaging of the CFRS and LDSS redshift surveys - IV. Influence of mergers in the evolution of faint field galaxies from  $z = 1$ . *Monthly Notices of the Royal Astronomical Society*, 311 : 565. 34
- Lilly S.J., Fèvre O.L., Hammer F., & Crampton D., 1996. The Canada-France Redshift Survey : The Luminosity Density and Star Formation History of the Universe to  $Z$  approximately 1. *Astrophysical Journal Letters* v.460, 460 : L1. (c) 1996 : The American Astronomical Society. 27

- Lucia G.D., Springel V., White S.D.M., Croton D., & Kauffmann G., 2006. The formation history of elliptical galaxies. *Monthly Notices of the Royal Astronomical Society*, 366 : 499. 28, 35
- Madau P., Ferguson H.C., Dickinson M.E., Giavalisco M., Steidel C.C., & Fruchter A., 1996. High-redshift galaxies in the Hubble Deep Field : colour selection and star formation history to  $z = 4$ . *Monthly Notices of the Royal Astronomical Society*, 283 : 1388. (c) 1996 The Royal Astronomical Society. 27
- Mather J.C., Cheng E.S., Eplee R.E., Isaacman R.B., Meyer S.S., Shafer R.A., Weiss R., Wright E.L., Bennett C.L., Bogges N.W., Dwek E., Gulkis S., Hauser M.G., Janssen M., Kelsall T., Lubin P.M., Moseley S.H., Murdock T.L., Silverberg R.F., Smoot G.F., & Wilkinson D.T., 1990. A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite. *Astrophysical Journal*, 354 : L37. 15
- Norberg P., Cole S., Baugh C.M., Frenk C.S., Baldry I., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Cross N.J.G., Dalton G., Propris R.D., Driver S.P., Efsthathiou G., Ellis R.S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Peacock J.A., Peterson B.A., Sutherland W., & Taylor K., 2002. The 2dF Galaxy Redshift Survey : the bJ-band galaxy luminosity function and survey selection function. *Monthly Notice of the Royal Astronomical Society*, 336 : 907. (c) RAS. 28, 29
- Padmanabhan T., 2002. *Theoretical Astrophysics, Volume III : Galaxies and Cosmology*. Cambridge University Press.
- Penzias A.A. & Wilson R.W., 1965. Measurement of the Flux Density of CAS a at 4080 Mc/s. *Astrophysical Journal*, 142 : 1149. 15
- Postman M., Franx M., Cross N.J.G., Holden B., Ford H.C., Illingworth G.D., Goto T., Demarco R., Rosati P., Blakeslee J.P., Tran K.V., Benítez N., Clampin M., Hartig G.F., Homeier N., Ardila D.R., Bartko F., Bouwens R.J., Bradley L.D., Broadhurst T.J., Brown R.A., Burrows C.J., Cheng E.S., Feldman P.D., Golimowski D.A., Gronwall C., Infante L., Kimble R.A., Krist J.E., Lesser M.P., Martel A.R., Mei S., Menanteau F., Meurer G.R., Miley G.K., Motta V., Sirianni M., Sparks W.B., Tran H.D., Tsvetanov Z.I., White R.L., & Zheng W., 2005. The Morphology-Density Relation in  $z = 1$  Clusters. *The Astrophysical Journal*, 623 : 721. (c) 2005 : The American Astronomical Society. 33
- Puech M., 2006. *Contraintes apportées par la spectroscopie intégrale de champ à la formation et à l'évolution des galaxies*. Thèse de doctorat, Université Paris VII - Denis Diderot.
- Refregier A., 2003. Weak Gravitational Lensing by Large-Scale Structure. *Annual Review of Astronomy & Astrophysics*, 41 : 645. 17

- Rubin V.C., 2000. One Hundred Years of Rotating Galaxies. The Publications of the Astronomical Society of the Pacific, 112 : 747. (c) 2000 : The Astronomical Society of the Pacific. 16
- Sandage A. & Tammann G.A., 1981. Book-Review - a Revised Shapley-Ames Catalog of Bright Galaxies. Journal of the British Astronomical Association, 91 : 522. 31
- Schechter P., 1976. An analytic expression for the luminosity function for galaxies. Astrophys. J., 203 : 297. A&AA ID. AAA017.158.009. 28
- Shostak G.S., 1973. Aperture Synthesis Study of Neutral Hydrogen in NGC 2403 and NGC 4236. II. Discussion. Astronomy and Astrophysics, 24 : 411. 17
- Smith G.P., Treu T., Ellis R.S., Moran S.M., & Dressler A., 2005. Evolution since  $z = 1$  of the Morphology-Density Relation for Galaxies. The Astrophysical Journal, 620 : 78. (c) 2005 : The American Astronomical Society. 33
- Smoot G.F., Bennett C.L., Kogut A., Wright E.L., Aymon J., Boggess N.W., Cheng E.S., de Amici G., Gulkis S., Hauser M.G., Hinshaw G., Jackson P.D., Janssen M., Kaita E., Kelsall T., Keegstra P., Lineweaver C., Loewenstein K., Lubin P., Mather J., Meyer S.S., Moseley S.H., Murdock T., Rokke L., Silverberg R.F., Tenorio L., Weiss R., & Wilkinson D.T., 1992. Structure in the COBE differential microwave radiometer first-year maps. Astrophysical Journal, 396 : L1. 16
- Spergel D., 2003. The Cosmological Implications of the WMAP Data : Parameters and Inflation. The Davis Meeting On Cosmic Inflation. 2003 March 22-25, p. 3. 15, 16
- Spergel D., 2006. The New Standard Cosmology. American Physical Society, p. 5002. 18
- Springel V., White S.D.M., Jenkins A., Frenk C.S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J.A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., & Pearce F., 2005. Simulations of the formation, evolution and clustering of galaxies and quasars. Nature, 435 : 629. (c) 2005 : Nature. 20
- Strateva I., Ivezić Ž., Knapp G.R., Narayanan V.K., Strauss M.A., Gunn J.E., Lupton R.H., Schlegel D., Bahcall N.A., Brinkmann J., Brunner R.J., Budavári T., Csabai I., Castander F.J., Doi M., Fukugita M., Györy Z., Hamabe M., Hennessy G., Ichikawa T., Kunszt P.Z., Lamb D.Q., McKay T.A., Okamura S., Racusin J., Sekiguchi M., Schneider D.P., Shimasaku K., & York D., 2001. Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data. AJ, 122 : 1861. 31
- Toomre A. & Toomre J., 1972. Model of the Encounter Between NGC 5194 and 5195. Bulletin of the American Astronomical Society, 4 : 214. A&AA ID. AAA007.158.132. 33

- Treu T., Koopmans L.V., Bolton A.S., Burles S., & Moustakas L.A., 2006. The Sloan Lens ACS Survey. II. Stellar Populations and Internal Structure of Early-Type Lens Galaxies. *The Astrophysical Journal*, 640 : 662. (c) 2006 : The American Astronomical Society. 17
- Tully R.B. & Fisher J.R., 1977. A new method of determining distances to galaxies. *Astronomy and Astrophysics*, 54 : 661. A&AA ID. AAA019.158.045. 23
- van den Bergh S., 2007. Some Musings on Galaxy Morphology, Galactic Colors, and the Environments of Galaxies. *AJ*, 134 : 1508. 31
- Zwicky F., 1933. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6 : 110. 16



# Morphologies de galaxies lointaines

---

## Sommaire

<b>2.1</b>	<b>D'une classification visuelle à une « science morphologie »</b>	<b>44</b>
2.1.1	Les premières classifications . . . . .	44
2.1.2	Morphologie ou traceur de la structure physique des galaxies .	45
<b>2.2</b>	<b>Morphologies de galaxies lointaines . . . . .</b>	<b>45</b>
2.2.1	Une automatisation des méthodes . . . . .	46
2.2.1.1	Classifications paramétriques . . . . .	46
2.2.1.2	Classifications non-paramétriques . . . . .	48
2.2.1.3	Couleurs et indices spectraux . . . . .	51
2.2.2	Le télescope spatial Hubble . . . . .	52
2.2.2.1	La mission . . . . .	52
2.2.2.2	« Hubble Deep Fields » . . . . .	52
2.2.3	Une difficulté d'interprétation : la k-correction morphologique .	55
<b>2.3</b>	<b>Le travail de cette thèse . . . . .</b>	<b>57</b>
	<b>Bibliographie . . . . .</b>	<b>58</b>

---

**Abstract :** This chapter describes major advances and issues encountered when trying to estimate morphologies of distant galaxies. The different methods (parametric and non-parametric) which are usually employed are presented. We discuss last advances in this field thanks in particular to HST imaging and show the need of NIR morphologies in the coming future because of morphological k-correction effects.

**Resumen :** El siguiente capítulo describe los avances y las dificultades encontradas en la estimación de las morfologías de galaxias lejanas. Se presentan primero los métodos automáticos generalmente empleados (paramétricos y no-paramétricos). Se abordan igualmente los últimos avances realizados, gracias en particular a los datos del telescopio espacial Hubble y se muestra el porque de la necesidad actual de realizar estudios morfológicos en el infrarrojo cercano.

---

L'étude des morphologies de galaxies est passée, depuis leur découverte dans les années 20, d'un travail purement descriptif à une *science de la morphologie* visant à relier les propriétés morphologiques à l'histoire de formation de celles-ci. En effet, il est maintenant bien connu que la morphologie est très fortement corrélée aux propriétés physiques de la galaxie (c.f. chapitre 1), et contient donc des informations importantes concernant son évolution. Il s'agit donc d'un traceur direct de la structure physique.

Remonter le temps à la recherche des formes des galaxies lointaines est devenu donc un axe central pour comprendre les processus physiques expliquant la formation de structures que l'on connaît actuellement dans l'Univers.

Ce chapitre a pour but de revoir l'évolution et les progrès dans ce domaine au cours des dernières années et de soulever les difficultés principales associées aux études morphologiques.

## 2.1 D'une classification visuelle à une « science morphologie »

### 2.1.1 Les premières classifications

Les premières tentatives de classification de galaxies furent visuelles et donc subjectives et commencèrent par les travaux de [Curtis \(1918\)](#), [Hubble \(1926\)](#) et [Sandage \(1961\)](#). Ces observations donnèrent lieu à la célèbre *classification de Hubble* qui divise les galaxies essentiellement en deux grandes catégories (spirales/elliptiques) à partir d'un critère de forme. Les galaxies elliptiques présentent une forme d'ellipsoïde sans trop de structures alors que les galaxies spirales sont sous forme de disque fin et présentent souvent des structures tels que des bras spiraux. (c.f. chapitre 1 pour une description détaillée). Probablement à cause de sa simplicité et grâce à son efficacité pour classer près de 97% des galaxies brillantes de l'Univers local, cette classification fut très vite adoptée par la communauté d'astronomes et a peu évoluée depuis mis à part quelques modifications mineures ([van den Bergh, 1960](#); [de Vaucouleurs, 1959](#)).

Il s'est avéré par la suite que la classification n'était pas purement descriptive mais cachait un grand nombre d'informations sur la structure physique de la galaxie. En effet, comme il a été expliqué dans le chapitre précédent, le simple fait de connaître la forme de la galaxie nous donne des informations sur l'âge de ses étoiles ([Kennicutt, 1992](#)), le mouvement de celles-ci au sein de la galaxie ([Tully & Fisher, 1977](#)), sa masse ([Kauffmann et al., 2003](#)), son contenu en gaz ([Roberts, 1969](#); [Kennicutt, 1992](#)). La morphologie est passée ainsi d'une simple description phénoménologique à un traceur de la structure physique.

### 2.1.2 Morphologie ou traceur de la structure physique des galaxies

Comprendre comment se sont formées les galaxies implique un travail archéologique ou ce qui revient au même, un regard sur l'Univers lointain. De ce point de vue, la morphologie, à cause de sa forte corrélation avec les propriétés physiques des galaxies, est un traceur idéal de l'évolution. C'est ainsi que l'étude des morphologies de galaxies lointaines est devenu un objectif majeur de la cosmologie moderne et l'un des buts principaux du télescope spatial Hubble comme on le verra par la suite.

En effet, la classification morphologique occupe généralement une place préliminaire dans les relevés de galaxies présentés dans le chapitre précédent. Comme il a été expliqué, la plupart des grandeurs statistiques mesurées, le sont par type morphologique, i.e. les fonctions de luminosité par type, les relations morphologie-densité ou l'évolution de la masse par type. Avant de pouvoir réaliser ces mesures il s'avère donc nécessaire de classer les galaxies à une époque donnée (*redshift*). Le schéma 2.1 illustre la place occupée par la morphologie dans les études statistiques à partir de relevés.

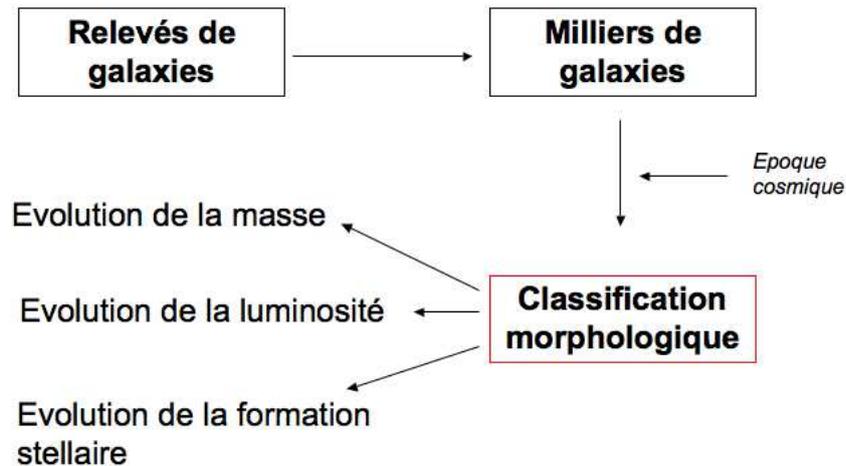


FIG. 2.1 – Place de l'étude morphologique dans les relevés cosmologiques. Celle-ci occupe généralement une place préliminaire.

## 2.2 Morphologies de galaxies lointaines

Tandis qu'au premier abord l'estimation de la morphologie des galaxies semble un problème relativement simple, ceci devient particulièrement difficile lorsqu'il s'agit d'objets lointains.

D'une part, la *résolution angulaire* devient critique. Comme nous le montrerons

dans le chapitre suivant, la résolution angulaire est une mesure de la taille minimale que l'on est capable de discerner avec les yeux ou avec un instrument d'optique. Evidemment, plus une galaxie est loin, plus elle va apparaître à l'observateur avec une taille apparente dans le ciel petite et plus cela va devenir difficile d'estimer précisément sa forme. En effet, une galaxie à  $z \sim 1$  (quand l'Univers avait approximativement la moitié de son âge actuel) peut apparaître jusqu'à 100 fois plus petite en nombre de pixels sur un détecteur qu'une galaxie proche. Ainsi, les traits caractéristiques permettant de réaliser une classification visuelle, comme par exemple les bras spiraux, sont moins visibles. Avoir des instruments avec la meilleure résolution angulaire possible est devenu donc un critère fondamental pour étudier la morphologie. Malheureusement cela est compliqué sur Terre à cause de l'effet de l'atmosphère sur la résolution comme je le montrerai dans le chapitre 3. Ce fut en effet une des raisons principales pour justifier le lancement du télescope spatial Hubble.

Le fait de regarder loin a également un effet dramatique sur la brillance de surface des galaxies. Alors que dans un Univers statique, la diminution en  $1/distance^2$  du flux reçu se compense exactement par l'augmentation de l'angle sous lequel on voit l'objet, ceci n'est plus vrai dans un Univers en expansion. Le flux reçu par un observateur décroît en effet en  $1/(1+z)^4$ . C'est ce que l'on appelle le *cosmological dimming factor* en anglais ou facteur d'atténuation cosmologique. Cela implique, que, très vite, les galaxies lointaines cessent d'être visibles. Des instruments avec une grande sensibilité deviennent donc indispensables.

Finalement, avec l'émergence de relevés de galaxies contenant un grand nombre d'objets lointains, les limites des classifications classiques commencent à être atteints. Avec des centaines de milliers d'objets il est devenu nécessaire de trouver des solutions automatiques afin de quantifier la morphologie avec des indicateurs reproductibles. Sans oublier que la subjectivité inhérente aux classifications visuelles peut induire des erreurs du fait que différents observateurs peuvent ne pas être d'accord sur la morphologie d'une même galaxie.

## 2.2.1 Une automatisation des méthodes

Pendant les dix dernières années, des progrès ont donc été réalisés dans la conception de méthodes permettant d'estimer la forme d'une galaxie. Ces travaux ont globalement convergé vers deux types de méthodes : des méthodes de type *paramétrique* qui visent à paramétriser analytiquement la galaxie et des méthodes dites *non-paramétriques* qui effectuent des mesures directes sur l'image de la galaxie.

### 2.2.1.1 Classifications paramétriques

Le paramètre principal décrivant la séquence de Hubble est la taille du bulbe par rapport au reste de la galaxie. Plus le bulbe est important, plus la galaxie est dite précoce. Ceci peut être quantifié par le rapport entre le flux du bulbe et le flux total de la galaxie ( $B/T$ ). Les classifications de galaxies ont donc tenté de quantifier

ce paramètre plus précisément et de façon objective afin de se rapprocher de la classification visuelle. Les approches de ce type consistent généralement en une modélisation des galaxies suivie d'un ajustement aux données par minimisation de moindres carrés après convolution par la FEP (Fonction d'étalement du point, c.f. chapitre 3 pour une définition détaillée).

En effet, comme il a été expliqué dans le chapitre d'introduction, les disques et les bulbes présentent des profils lumineux différents ; le disque étant bien décrit par un profil de type exponentiel (Freeman, 1970) alors que le bulbe s'ajuste plus à un profil de *de Vaucouleurs* (de Vaucouleurs, 1959). Une galaxie normale de la séquence de Hubble peut donc être vue comme une superposition de ces deux composantes, chacune d'elles contribuant au flux total de la galaxie plus ou moins en fonction du type morphologique :

$$\Sigma(r) = \frac{B}{T} \cdot \exp\left[-7.6697 \cdot \frac{r}{r_e}\right] + \left(1 - \frac{B}{T}\right) \cdot \exp\left[-\frac{r}{r_d}\right] \quad (2.1)$$

Ainsi, une galaxie elliptique pure aura une valeur de  $B/T = 1$  alors que pour une galaxie disque,  $B/T = 0$ .

L'ajustement a typiquement une dizaine de paramètres : le flux total de la galaxie, la fraction de bulbe ( $B/T$ ), le rayon effectif du bulbe ( $r_e$ ), l'ellipticité du bulbe ( $e = 1 - b/a$ ), le rayon caractéristique du disque ( $r_d$ ), l'inclinaison du disque ( $i$ ), l'estimation du fond ( $db$ ).

Ce modèle est bien entendu une approximation simpliste de la vraie structure d'une galaxie. Les galaxies réelles présentent plus de deux composantes uniques (barres, bras spiraux, régions HII). Cependant, il s'avère suffisant pour décrire la forme globale de la galaxie. Il s'agit en particulier d'un bon compromis entre une description réaliste de la galaxie et un nombre de paramètres d'ajustement suffisamment petit pour ne pas introduire trop de dégénérescences. D'autres techniques basées sur des décompositions en *ondelettes* (e.g. Khalil *et al.*, 2006), permettent de reproduire plus finement les détails de la galaxie, mais celles-ci sont rarement utilisées pour des études statistiques.

L'ajustement permet non seulement d'extraire la morphologie de la galaxie (rapport  $B/T$ ) mais aussi d'autres caractéristiques comme les tailles des bulbes et des disques, permettant ainsi des analyses plus poussées de la structure.

Un point fondamental dans ce type d'approche est l'estimation de la FEP des observations. En effet, une mauvaise estimation de celle-ci conduit à un modèle de galaxie qui n'est pas représentatif des données observées et peut induire des erreurs en particulier pour des galaxies lointaines ou mal résolues. Comme je le montrerai dans le chapitre 4 une attention particulière a été portée sur ce point dans l'analyse de nos données.

Il existe actuellement deux codes principaux, disponibles pour la communauté, fonctionnant sur ce principe : GALFIT (Peng *et al.*, 2002) et GIM2D (Marleau & Si-

mard, 1998; Simard *et al.*, 2002). Nous avons utilisé GIM2D pour notre étude du chapitre 4. Je détaillerai donc à ce stade les particularités de ce logiciel.

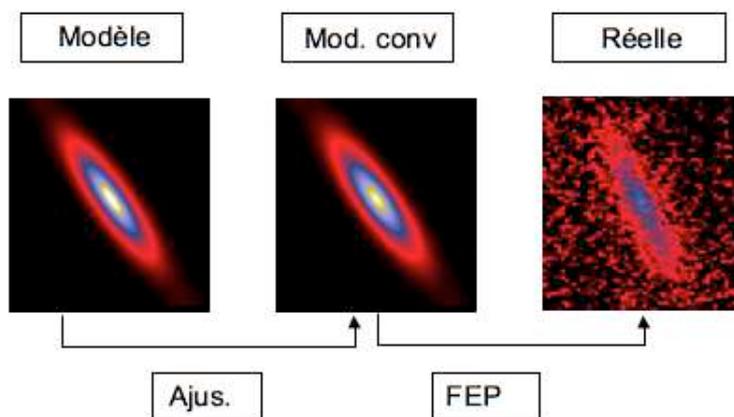


FIG. 2.2 – Illustration du principe des méthodes paramétriques. Le modèle de la galaxies (gauche) est convolué par la fonction d'étalement du point (milieu) est ajusté sur la galaxies réelle (droite).

### 2.2.1.2 Classifications non-paramétriques

Dans les classifications dites non paramétriques, l'approche est différente. On ne paramétrise pas la galaxie, mais on mesure sur la galaxie réelle des paramètres qui sont corrélés avec le type morphologique. L'avantage principal de ce type de méthode c'est qu'elle ne présuppose pas une expression analytique pour la galaxie et peut donc être employée pour classer des galaxies irrégulières également. La morphologie résultante sera cependant plus qualitative que pour les méthodes paramétriques dans le sens où on ne récupère pas d'information sur les rayons effectifs ni les ellipticités mais juste une mesure de la forme. Abraham *et al.* (1994, 1996) ont proposé pour la première fois ce type de méthode pour classer des galaxies à *redshift* élevé. Ils ont en effet défini deux paramètres : concentration et asymétrie (notés  $C$  et  $A$  respectivement) et montré, en utilisant des données du télescope spatial, qu'une représentation de ces paramètres dans un plan (*le plan C/A*) permet de bien séparer les trois types morphologiques principaux (spirale, elliptique et irrégulière). Ces 2 paramètres sont essentiellement une mesure de la concentration de lumière au centre de la galaxie, qui est donc corrélée avec la taille du bulbe (les galaxies elliptiques étant généralement plus « concentrées » que les galaxies disques), et du degré de symétrie de la galaxie, également corrélé avec le type morphologique, puisque les galaxies de type disque seront plus asymétriques à cause en particulier des structures telles que les bras spiraux par exemple.

Ainsi, dans sa première définition, le paramètre *concentration* est calculé comme le rapport entre le flux contenu dans une isophote de la galaxie et le flux total de

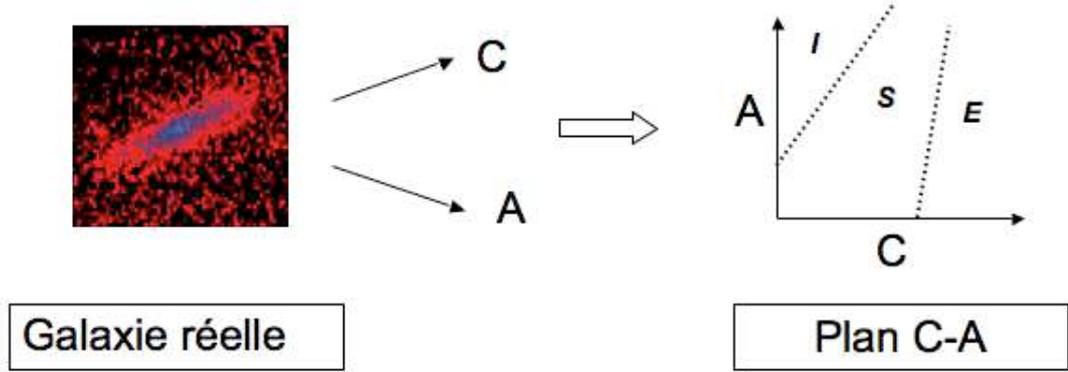


FIG. 2.3 – Illustration du principe des méthodes non-paramétriques. Plusieurs paramètres sont mesurés sur l'image de la galaxie et représentés dans un plan afin de définir des zones pour chaque type morphologique.

celle-ci :

$$C = \frac{\sum_{i,j \in E(\alpha)} I_{ij}}{\sum_{i,j \in E(1)} I_{ij}} \quad (2.2)$$

avec  $I(i, j)$  étant la valeur de l'intensité au pixel de coordonnées  $(i, j)$  et  $E(\alpha)$  des ellipses isophotales de la galaxie paramétrisées selon le rayon normalisé  $\alpha$  de telle sorte que pour  $\alpha = 1$  toute la galaxie est englobée. Il existe cependant dans la littérature de nombreuses définitions de la concentration toutes basées sur le même principe. Une définition souvent employée est celle de [Conselice et al. \(2000\)](#) et que j'utiliserai également dans ce manuscrit :

$$C = 5 \log(r_{80}/r_{20}) \quad (2.3)$$

avec  $r_{80}$  et  $r_{20}$  étant les rayons contenant 80% et 20% du flux respectivement.

Le paramètre *asymétrie* est une mesure du degré de symétrie de la galaxie et se calcule en obtenant la valeur absolue de la différence entre l'intensité de la galaxie et l'intensité de la galaxie tournée de  $180^\circ$ . La définition employée dans ce manuscrit est celle de [Conselice et al. \(2000\)](#) :

$$A = \frac{1}{2} \left( \frac{\sum |I(i, j) - I_{180}(i, j)|}{\sum I(i, j)} - \frac{\sum |B(i, j) - B_{180}(i, j)|}{\sum I(i, j)} \right) \quad (2.4)$$

$I(i, j)$  étant encore l'intensité de la galaxie et  $B(i, j)$  l'intensité du bruit de l'image dans une zone de taille égale à celle de la galaxie. Il est particulièrement important pour le calcul de l'asymétrie de bien déterminer le centre de la galaxie. [Conselice](#)

*et al.* (2000) emploient une méthode empirique qui définit la position du centre comme celle qui minimise la valeur de l'asymétrie. C'est cette méthode qui est utilisée tout au long du manuscrit.

En plus de ces deux paramètres originaux, d'autres auteurs ont introduit de nouveaux paramètres qui vont être également utilisés dans ce qui suit et que je trouve donc utile de définir.

*Conselice* (2003) ont introduit le paramètre de *smoothness* ( $S$ ) qui vise à mesurer le nombre de structures de petite taille à l'intérieur de la galaxie ou autrement dit les écarts à l'homogénéité dans le brillance de surface :

$$S = \frac{1}{2} \left( \frac{\sum |I(i, j) - I_S(i, j)|}{\sum I(i, j)} - \frac{\sum |B(i, j) - B_S(i, j)|}{\sum I(i, j)} \right), \quad (2.5)$$

Ce paramètre a donné nom au système de classification CAS (*Concentration-Asymmetry-Smoothness*).

*Abraham et al.* (2003) ont défini le coefficient *gini* ( $G$ ). La définition de celui-ci est basée sur un indicateur économique, la courbe de Lorentz, qui mesure la distribution de richesse dans une société ou dans notre cas la distribution d'intensité dans les pixels de la galaxie. Ainsi, le gini vaut 1 quand toutes les richesses (intensités) appartiennent à une seule personne (pixel) et 0 si elles sont distribuées uniformément. Ce paramètre est donc fortement corrélé avec la concentration mais a la propriété d'être indépendant de la façon dont est distribuée la lumière. L'expression couramment utilisée est :

$$G = \frac{1}{\bar{X}n(n-1)} \sum_{i=1}^n (2i - n - 1)X_i, \quad (n > 2) \quad (2.6)$$

avec  $X$  étant une liste des intensités des pixels ordonnées par ordre croissant.

Le dernier de ces paramètres est noté  $M_{20}$  et est défini dans *Lotz et al.* (2004). On définit d'abord le moment de deuxième ordre :

$$M_{tot} = \sum I_i [(x_i - x_c)^2 + (y_i - y_c)^2] \quad (2.7)$$

avec  $x_c$  et  $y_c$  les coordonnées du centre de la galaxie. Le coefficient  $M_{20}$  est alors défini comme le moment de deuxième ordre des 20% pixels les plus brillants.

Un des problèmes de ces définitions réside dans le fait que la valeur obtenue dépend fortement de l'estimation du fond d'une part, puisque l'estimation des pixels appartenant à la galaxie en dépend, mais aussi du *redshift*. Une même galaxie observée à différents *redshifts* peut en effet présenter des valeurs de concentration différentes du fait que dans la galaxie plus lointaine, une partie de celle-ci peut être noyée dans le bruit (*Brinchmann et al.*, 1998). Ceci pose un problème pour des études d'évolution en particulier où l'on souhaite justement comparer des populations à différents *redshifts*. Plusieurs améliorations ont donc été proposées. En particulier,

les derniers travaux utilisent le rayon *petrosien* (Petrosian, 1976; Blanton *et al.*, 2001) pour définir l'aire de la galaxie. Pour le définir on introduit la notion de *rapport pétrosien* :

$$R_p(r) \equiv \frac{\int_{0.8r}^{1.25r} dr' 2\pi r' I(r') / [\pi(1.25^2 - 0.8^2)r^2]}{\int_0^r dr' 2\pi r' I(r') / (\pi r^2)} \quad (2.8)$$

avec  $I(r)$  la brillance de surface de la galaxie. Le *rayon petrosien*  $r_p$  est alors défini comme le rayon pour lequel  $R_p(r) = 0.2$ . Ce rayon possède la propriété d'être plus robuste aux effets de variations de bruit de fond ou de *redshift*. Les calculs de concentration sont donc en général réalisés à l'intérieur d'une surface définie par un rayon  $2r_p$ .

Malgré cela, les valeurs calculées restent dépendantes des observations et donc il est particulièrement important de réaliser un étalonnage correct des paramètres avant d'effectuer la classification. La deuxième partie de cette thèse est dédiée à ce type de techniques. Une revue exhaustive des évolutions de ce type de méthodes et de l'état d'art de celles-ci est donc présentée au chapitre 7.

### 2.2.1.3 Couleurs et indices spectraux

L'estimation des morphologies de galaxies lointaines présente donc un grand nombre de difficultés techniques et requiert des observations de haute qualité tant d'un point de vue de résolution comme de sensibilité. Ce n'est évidemment pas toujours facile à obtenir et par conséquent des méthodes alternatives simplifiées sont souvent utilisées. Celles-ci exploitent les fortes corrélations existantes dans l'Univers local entre le type morphologique et d'autres propriétés telles que la couleur (c.f chapitre 1). Ainsi on peut par exemple définir les galaxies elliptiques comme celles ayant une couleur plus rouge qu'un certain seuil défini empiriquement d'après les observations et inversement les galaxies spirales comme celles étant plus bleues qu'un seuil donné. Il est possible également d'employer les couleurs de galaxies pour ajuster des modèles de spectres de galaxies locales et en déduire ainsi le type morphologique du meilleur ajustement (c.f. Annexe A). Ces types de classifications sont souvent appelés *classifications par couleurs* ou *classifications spectrales* pour les différencier des *classifications structurelles* présentées précédemment. Elles ont l'avantage d'être très simples et plus rapides mais possèdent également un certain nombre de problèmes majeurs. En particulier rien ne nous assure que les couleurs des galaxies n'évoluent pas et que ce seuil défini dans l'Univers d'aujourd'hui est valable à grand *redshift*. En particulier il a été noté par plusieurs auteurs (e.g. Ilbert *et al.*, 2006) la présence de galaxies elliptiques bleues dans le passé qui peuvent jouer un rôle important dans l'évolution. Celles-ci ne peuvent pas être identifiées avec une classification n'utilisant que les couleurs.

## 2.2.2 Le télescope spatial Hubble

Il est difficile de parler de morphologies de galaxies lointaines sans parler du télescope spatial Hubble (HST). Presque la totalité des études menées jusqu'à aujourd'hui ont été réalisées grâce à des données issues des instruments de celui-ci. Il a permis ainsi des études exhaustives de morphologies jusqu'à  $z \sim 1$ .

### 2.2.2.1 La mission

Les simulations réalisées avant son lancement suggéraient pourtant que le HST n'aurait pas apporter un gain fondamental pour l'étude de galaxies lointaines (Bahcall *et al.*, 1990). Dès la parution des premières images (Dressler *et al.*, 1994) il fut cependant évident que les galaxies lointaines avaient une brillance de surface plus importante que prévu. Ces premières observations apportèrent par conséquent des résultats inédits sur les morphologies de galaxies, l'évolution du taux de fusion de galaxies ou l'évolution de la densité (Mutz *et al.*, 1994; Schade *et al.*, 1995; Cowie *et al.*, 1995). Dans ce contexte, l'idée de réaliser un relevé plus ambitieux et plus profond prit forme.

### 2.2.2.2 « Hubble Deep Fields »

**Présentation** Les « Hubble Deep Fields » nord et sud (HDF-N et HDF-S) ont révolutionné l'étude de galaxies distantes mais aussi des étoiles du halo galactique ou de supernovae. Les champs furent choisis près du pôle galactique afin d'avoir une faible extinction galactique et pour ne pas être trop contaminé par la présence d'étoiles brillantes (cela s'avère un problème pour les études depuis le sol comme on le verra par la suite) (Williams *et al.*, 1996, 2000).

Une revue de tous les résultats issus de l'étude des ces champs est au delà des objectifs de ce manuscrit. Je décrirai cependant de manière assez succincte les résultats qui concernent directement le travail de cette thèse, c'est-à-dire les morphologies de galaxies lointaines.

**Morphologies de galaxies** L'étude de l'évolution des morphologies est devenu un axe central dans l'exploitation des données des HDF grâce à la qualité des images obtenues bien supérieure à celles obtenues avec n'importe quel télescope terrestre. Toutes les méthodes décrites ci-dessus ont pu être employées (e.g. van den Bergh *et al.*, 1996; Marleau & Simard, 1998). Les résultats obtenus avec les différentes méthodes ne sont pas toujours en accord; comme on l'a vu l'étude de galaxies lointaines présente un grand nombre de difficultés. Malgré cela, un certain nombre de résultats sont maintenant acceptés.

### Galaxies irrégulières

On constate une augmentation du nombre de galaxies de type irrégulier à grand *redshift*. Comme il a été dit dans le chapitre d'introduction, dans l'Univers local la grande majorité des galaxies peuvent être classées comme spirales ou elliptiques ; la fraction de galaxies irrégulières ne représentant que de l'ordre de 3% du nombre total de galaxies. En employant des méthodes non-paramétriques, Abraham *et al.* (1996) réalisèrent l'étude des morphologies visibles dans les HDF jusqu'à une magnitude de  $I=25$  et trouvèrent une fraction de l'ordre de 40% de galaxies irrégulières, fraction nettement supérieure à celle de l'Univers local. Cet effet peut être en partie dû à la k-correction morphologique (c.f. § 2.2.3) du fait que à grand *redshift* on sonde l'émission UV des galaxies plus sensible donc à des épisodes ponctuels de formation stellaire (Bohlin *et al.*, 1991; Giavalisco *et al.*, 1996). Cependant des quantifications plus précises de ces effets (Abraham *et al.*, 1999) suggèrent que la k-correction n'est pas l'effet dominant. De plus des observations avec la caméra infrarouge du HST (NICMOS) semblent confirmer que les perturbations des morphologies sont physiques et non pas un effet de longueur d'onde. Ces observations semblent être une conséquence d'une interaction de galaxies plus fréquente dans le passé (LeFèvre *et al.*, 2000) et donc d'une vision hiérarchique de l'Univers.

### Galaxies elliptiques

La formation des galaxies elliptiques massives a toujours été un thème de controverse dans les scénarii d'évolution. En se basant sur l'homogénéité de leurs propriétés photométriques et structurales dans l'Univers local Eggen *et al.* (1962) proposèrent un scénario de formation de type « monolithique » où les galaxies elliptiques se formeraient toutes pendant une période relativement courte de temps par un épisode violent de formation stellaire à des *redshifts* élevés ( $z > 5$ ) et évolueraient ensuite passivement jusqu'à nos jours. D'un autre côté, les simulations numériques menées par Toomre & Toomre (1972) montrant qu'une galaxie elliptique peut se former par fusion de deux galaxies spirales ainsi que l'émergence des modèles hiérarchiques de croissance de structures ( $\Lambda$ CDM) introduirent une vision « hiérarchique » où les galaxies elliptiques se formeraient par fusions successives (e.g. Kauffmann *et al.*, 1993). Etant donné que les conséquences de ces deux scénarii d'un point de vue observationnel sont très différentes (dans le premier cas, la densité de galaxies elliptiques devrait rester constante alors que dans le deuxième elle devrait diminuer avec le *redshift*), un grand nombre de travaux ont été dédiés à tenter d'élucider ce point. Dans ce contexte, le télescope spatial Hubble a rendu possible la sélection, à l'aide des techniques exposées dans le chapitre 2, d'échantillons de galaxies elliptiques lointaines afin d'étudier leurs propriétés (e.g. Driver *et al.*, 1995; Glazebrook *et al.*, 1995). Les résultats des différentes études ne sont pas toujours en accord. Tandis qu'un certain nombre de travaux trouvent peu ou pas d'évolution, ce qui serait compatible avec une évolution passive des galaxies elliptiques (Kodama *et al.*, 1999; Schade *et al.*, 1999), d'autres trouvent des traces d'une formation stellaire récente dans des galaxies elliptiques (Menanteau *et al.*, 1999; van Dokkum & van der Marel, 2007) indiquant donc une formation plus étendue dans le temps. La

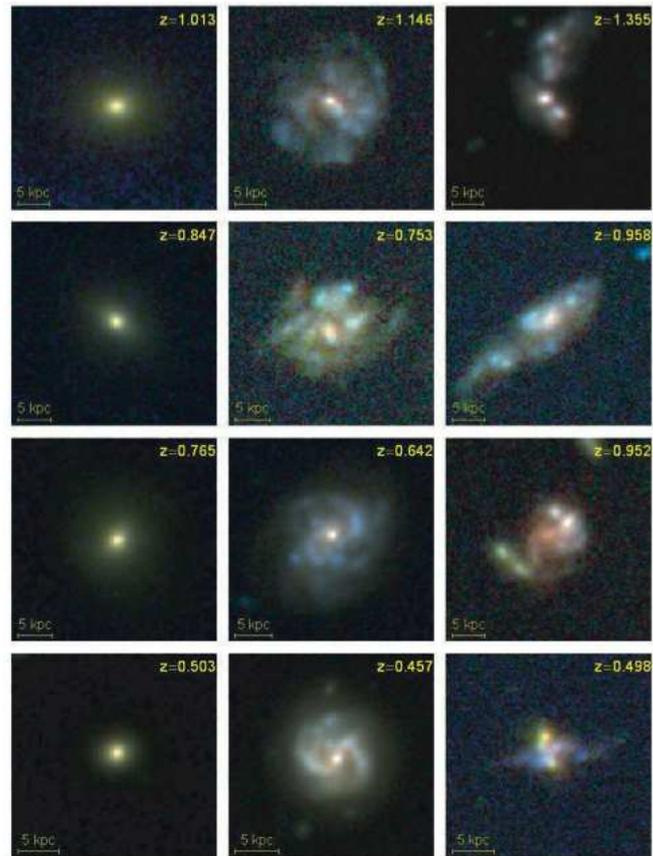


FIG. 2.4 – Exemple d’objets observés dans les HDF d’après [Abraham & van den Bergh \(2001\)](#). Les couleurs sont le résultat de la combinaison de 3 filtres. Gauche : galaxies elliptiques, milieu : galaxies spirales, droite : galaxies irrégulières.

question est restée et reste toujours ouverte. Aujourd'hui, il semblerait que l'histoire de ces objets dépend fortement de la masse (c.f § 1.3.5) : les galaxies plus massives formeraient leurs étoiles plus tôt et plus rapidement comme le prévoit la vision « monolithique » alors que les galaxies moins massives auraient une histoire de formation stellaire plus étendue dans le temps. Ces résultats commencent à être reproduits par les modèles semi-analytiques comme nous l'avons vu dans le chapitre d'introduction (Lucia *et al.*, 2006).

### 2.2.3 Une difficulté d'interprétation : la k-correction morphologique

Les résultats exposés précédemment sont basés essentiellement sur l'exploitation de données dans le domaine visible. Le fait d'utiliser une seule bande spectrale est susceptible d'introduire des biais dans l'interprétation des observations. En effet, lorsque l'on étudie l'évolution de la morphologie au cours du temps on sonde des époques différentes et donc des *redshifts* différents. Par conséquent, les observations réalisées dans un filtre unique sondent chaque galaxie dans une longueur d'onde différente en fonction du *redshift*. La définition même du *redshift* lie la longueur d'onde du filtre d'observation à la longueur d'onde réellement observée par :

$$\lambda_{obs} = \frac{\lambda_{filtre}}{1 + z} \quad (2.9)$$

En définitive, plus une galaxie est lointaine plus la lumière reçue pour un même filtre fixe est bleue (longueur d'onde plus petite). Ceci pose un problème pour l'interprétation de la morphologie puisque l'aspect d'une galaxie peut fortement varier selon la longueur d'onde d'observation. En effet, si l'on observe une galaxie dans l'UV, la lumière provient essentiellement d'étoiles jeunes et chaudes qui sont souvent situées dans des zones de formation stellaire pouvant donc donner lieu à une morphologie de type irrégulier. Dans le visible, la lumière dominante provient des étoiles de la séquence principale et la morphologie est donc proche de la classification originale de Hubble. Dans l'infrarouge, la plupart du flux provient d'étoiles vieilles, évoluées qui se distribuent en général d'une façon uniforme. La morphologie apparente devient donc plus symétrique (Fig. 2.6).

Par conséquent, des galaxies qui devraient être identiques peuvent présenter des morphologies apparentes différentes par le simple fait d'être à des *redshifts* différents. Cet effet reçoit le nom de *k correction morphologique* et représente un problème fondamental dans les études de l'évolution de la morphologie. On appelle *morphologie au repos* (*rest-frame* en anglais) la morphologie d'une galaxie dans une bande donnée fixe. L'analyse à plusieurs *redshift* requiert donc une analyse multi-longueur d'onde, en particulier il est nécessaire d'observer dans des longueurs d'onde de plus en plus longues au fur et à mesure que le *redshift* est élevé afin de rester toujours dans des longueurs d'onde visibles.



FIG. 2.5 – La morphologie d’une même galaxie peut apparaître différente en fonction de la longueur d’onde d’observation. La figure montre la galaxie du « sombrero » observée en rayons X par le satellite Chandra (en haut à droite), dans le visible avec le télescope spatial Hubble (milieu à droite) et dans l’infrarouge lointain avec Spitzer (en bas à droite). L’image centrale montre la combinaison des trois. [Source : NASA]

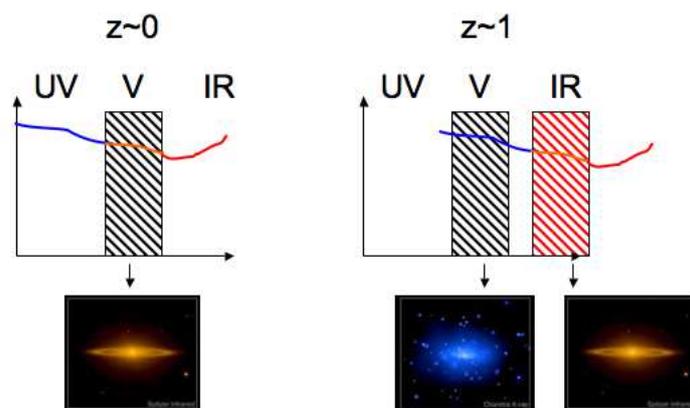


FIG. 2.6 – Illustration de l’effet de k-correction morphologique.

## 2.3 Le travail de cette thèse

La morphologie est donc un traceur direct de la structure de la galaxie. Son estimation qui au premier abord paraît simple est cependant un problème complexe lorsqu'il s'agit de galaxies lointaines. Non seulement il est nécessaire d'avoir des instruments capables de scruter des détails très petits, il faut être capable ensuite de la quantifier d'une manière objective et reproductible. Le télescope spatial a permis de faire des énormes avancées dans ce domaine grâce à la qualité et la profondeur des images obtenues. Les relevés actuels commencent désormais à sonder des populations de galaxies à des *redshifts* supérieurs à 1. À partir de ces distances, les caméras visibles du télescope spatial sondent l'émission UV des galaxies et la morphologie obtenue peut être donc fortement biaisée à cause des effets de *k-correction* exposés ci-dessus. Les études morphologiques à venir doivent se faire par conséquent dans ces bandes plus rouges. Le HST possède certes une caméra infrarouge, NICMOS, mais la résolution angulaire obtenue est nettement inférieure qu'avec les caméras visibles (c.f. chapitre 6).

Cette thèse tente donc de répondre à une question principale : **est-il possible aujourd'hui d'estimer les morphologies de galaxies à  $z \sim 1$  et au delà dans l'infrarouge avec des télescopes terrestres ? Si oui, quels sont alors les effets de la *k-correction* sur les morphologies des galaxies estimées ?** Deux voies principales sont explorées dans ce manuscrit :

- Les systèmes de correction des effets de l'atmosphère (optique adaptative) installés dans les grands télescopes actuels sont suffisants pour atteindre une qualité comparable à celle obtenue dans l'espace ? Nous analysons un échantillon de 79 galaxies observées avec le VLT. Une analyse approfondie de la fiabilité des classifications est présentée ainsi qu'une première estimation de l'évolution des morphologies en trois types. (Partie I)
- Peut-on améliorer les méthodes actuelles d'estimation morphologique afin d'extraire une information même à partir de données non corrigées ? Nous présentons une nouvelle méthode non-paramétrique pour la quantification de morphologies de galaxies lointaines basée sur des machines d'apprentissage. La méthode est employée pour analyser un échantillon de  $\sim 50\,000$  galaxies dans la bande K jusqu'à  $z \sim 2$ . Une comparaison détaillée avec les résultats obtenus dans le visible avec le télescope spatial est présentée afin de quantifier les effets de *k-correction*. (Partie II)

Dans les deux parties des études de performances ainsi que des résultats scientifiques seront présentés.

## Bibliographie

- Abraham R.G., Ellis R.S., Fabian A.C., Tanvir N.R., & Glazebrook K., 1999. The star formation history of the Hubble sequence : spatially resolved colour distributions of intermediate-redshift galaxies in the Hubble Deep Field. *Monthly Notices of the Royal Astronomical Society*, 303 : 641. (c) 1999 The Royal Astronomical Society. 53
- Abraham R.G., Valdes F., Yee H.K.C., & van den Bergh S., 1994. The morphologies of distant galaxies. 1 : an automated classification system. *Astrophysical Journal*, 432 : 75. 48
- Abraham R.G. & van den Bergh S., 2001. The Morphological Evolution of Galaxies. *Science*, 293 : 1273. (c) 2001 : Science. 54
- Abraham R.G., van den Bergh S., Glazebrook K., Ellis R.S., Santiago B.X., Surma P., & Griffiths R.E., 1996. The Morphologies of Distant Galaxies. II. Classifications from the Hubble Space Telescope Medium Deep Survey. *Astrophysical Journal Supplement v.107*, 107 : 1. 48, 53
- Abraham R.G., van den Bergh S., & Nair P., 2003. A New Approach to Galaxy Morphology. I. Analysis of the Sloan Digital Sky Survey Early Data Release. *The Astrophysical Journal*, 588 : 218. 50
- Bahcall J.N., Guhathakurta P., & Schneider D.P., 1990. What the longest exposures from the Hubble Space Telescope will reveal. *Science*, 248 : 178. 52
- Blanton M.R., Dalcanton J., Eisenstein D., Loveday J., Strauss M.A., SubbaRao M., Weinberg D.H., Anderson J.E., Annis J., Bahcall N.A., Bernardi M., Brinkmann J., Brunner R.J., Burles S., Carey L., Castander F.J., Connolly A.J., Csabai I., Doi M., Finkbeiner D., Friedman S., Frieman J.A., Fukugita M., Gunn J.E., Hennessy G.S., Hindsley R.B., Hogg D.W., Ichikawa T., Ivezić Ž., Kent S., Knapp G.R., Lamb D.Q., Leger R.F., Long D.C., Lupton R.H., McKay T.A., Meiksin A., Merelli A., Munn J.A., Narayanan V., Newcomb M., Nichol R.C., Okamura S., Owen R., Pier J.R., Pope A., Postman M., Quinn T., Rockosi C.M., Schlegel D.J., Schneider D.P., Shimasaku K., Siegmund W.A., Smee S., Snir Y., Stoughton C., Stubbs C., Szalay A.S., Szokoly G.P., Thakar A.R., Tremonti C., Tucker D.L., Uomoto A., Berk D.V., Vogeley M.S., Waddell P., Yanny B., Yasuda N., & York D.G., 2001. The Luminosity Function of Galaxies in SDSS Commissioning Data. *The Astronomical Journal*, 121 : 2358. (c) 2001 : The American Astronomical Society. 51
- Bohlin R.C., Cornett R.H., Hill J.K., Hill R.S., Landsman W.B., O'Connell R.W., Neff S.G., Smith A.M., & Stecher T.P., 1991. Simulated optical images of galaxies at Z about 1 using ultraviolet images of nearby galaxies. *Astrophysical Journal*, 368 : 12. 53

- Brinchmann J., Abraham R., Schade D., Tresse L., Ellis R.S., Lilly S., Fevre O.L., Glazebrook K., Hammer F., Colless M., Crampton D., & Broadhurst T., 1998. Hubble Space Telescope Imaging of the CFRS and LDSS Redshift Surveys. I. Morphological Properties. *Astrophysical Journal* v.499, 499 : 112. 50
- Conselice C.J., 2003. The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories. *The Astrophysical Journal Supplement Series*, 147 : 1. (c) 2003 : The American Astronomical Society. 50
- Conselice C.J., Bershadsky M.A., & Jangren A., 2000. The Asymmetry of Galaxies : Physical Morphology for Nearby and High-Redshift Galaxies. *The Astrophysical Journal*, 529 : 886. 49
- Cowie L.L., Hu E.M., & Songaila A., 1995. Detection of Massive Forming Galaxies at Redshifts  $Z > 1$ . *NATURE* V.377, 377 : 603. 52
- Curtis H.D., 1918. The planetary nebulae. *Publ. Lick Obs.*, 13 : 55. 44
- de Vaucouleurs G., 1959. Classification and Morphology of External Galaxies. *Handbuch der Physik*, 53 : 275. 44, 47
- Dressler A., Oemler A., Sparks W.B., & Lucas R.A., 1994. New images of the distant, rich cluster CL 0939+4713 with WFPC2. *Astrophysical Journal*, 435 : L23. 52
- Driver S.P., Windhorst R.A., & Griffiths R.E., 1995. The Contribution of Late-Type/Irregulars to the Faint Galaxy Counts from Hubble Space Telescope Medium-Deep Survey Images. *Astrophysical Journal* v.453, 453 : 48. 53
- Eggen O.J., Lynden-Bell D., & Sandage A.R., 1962. Evidence from the motions of old stars that the Galaxy collapsed. *Astrophysical Journal*, 136 : 748. 53
- Freeman K.C., 1970. On the Disks of Spiral and so Galaxies. *ApJ*, 160 : 811. 47
- Gialalisco M., Livio M., Bohlin R.C., Macchetto F.D., & Stecher T.P., 1996. On the Morphology of the HST Faint Galaxies. *Astronomical Journal* v.112, 112 : 369. 53
- Glazebrook K., Ellis R., Santiago B., & Griffiths R., 1995. The morphological identification of the rapidly evolving population of faint galaxies. *Monthly Notices of the Royal Astronomical Society*, 275 : L19. 53
- Hubble E.P., 1926. Extragalactic nebulae. *Astrophys. J.*, 64 : 321. 44
- Ilbert O., Lauger S., Tresse L., Buat V., Arnouts S., Fèvre O.L., Burgarella D., Zucca E., Bardelli S., Zamorani G., Bottini D., Garilli B., Brun V.L., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Arnaboldi M., Bolzonella M., Cappi A., Charlot S., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Iovino A., McCracken H.J., Marano B.,

- Marinoni C., Mathez G., Mazure A., Meneux B., Merighi R., Paltani S., Pello R., Pollo A., Pozzetti L., Radovich M., Bondi M., Bongiorno A., Busarello G., Ciliegi P., Mellier Y., Merluzzi P., Ripepi V., & Rizzo D., 2006. The VIMOS-VLT Deep Survey. Galaxy luminosity function per morphological type up to  $z = 1.2$ . *Astronomy and Astrophysics*, 453 : 809. 51
- Kauffmann G., Heckman T.M., White S.D.M., Charlot S., Tremonti C., Brinchmann J., Bruzual G., Peng E.W., Seibert M., Bernardi M., Blanton M., Brinkmann J., Castander F., Csábai I., Fukugita M., Ivezić Z., Munn J.A., Nichol R.C., Padmanabhan N., Thakar A.R., Weinberg D.H., & York D., 2003. Stellar masses and star formation histories for 105 galaxies from the Sloan Digital Sky Survey. *Monthly Notice of the Royal Astronomical Society*, 341 : 33. (c) 2003 RAS. 44
- Kauffmann G., White S.D.M., & Guiderdoni B., 1993. The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes. *R.A.S. MONTHLY NOTICES* V.264, 264 : 201. 53
- Kennicutt R.C., 1992. The integrated spectra of nearby galaxies - General properties and emission-line spectra. *Astrophysical Journal*, 388 : 310. 44
- Khalil A., Joncas G., Nekka F., Kestener P., & Arneodo A., 2006. Morphological Analysis of H I Features. II. Wavelet-based Multifractal Formalism. *ApJs*, 165 : 512. 47
- Kodama T., Bower R.G., & Bell E.F., 1999. The colour-magnitude relation of early-type galaxies in the Hubble Deep Field. *Monthly Notices of the Royal Astronomical Society*, 306 : 561. (c) 1999 RAS. 53
- LeFèvre O., Abraham R., Lilly S.J., Ellis R.S., Brinchmann J., Schade D., Tresse L., Colless M., Crampton D., Glazebrook K., Hammer F., & Broadhurst T., 2000. Hubble Space Telescope imaging of the CFRS and LDSS redshift surveys - IV. Influence of mergers in the evolution of faint field galaxies from  $z = 1$ . *Monthly Notices of the Royal Astronomical Society*, 311 : 565. 53
- Lotz J.M., Primack J., & Madau P., 2004. A New Nonparametric Approach to Galaxy Morphological Classification. *The Astronomical Journal*, 128 : 163. 50
- Lucia G.D., Springel V., White S.D.M., Croton D., & Kauffmann G., 2006. The formation history of elliptical galaxies. *Monthly Notices of the Royal Astronomical Society*, 366 : 499. 55
- Marleau F.R. & Simard L., 1998. Quantitative Morphology of Galaxies in the Hubble Deep Field. *The Astrophysical Journal*, 507 : 585. (c) 1998 : The American Astronomical Society. 47, 52
- Menanteau F., Ellis R.S., Abraham R.G., Barger A.J., & Cowie L.L., 1999. The optical-infrared colour distribution of a statistically complete sample of faint field

- spheroidal galaxies. *Monthly Notices of the Royal Astronomical Society*, 309 : 208. 53
- Mutz S.B., Windhorst R.A., Schmidtke P.C., Pascarelle S.M., Griffiths R.E., Ratnatunga K.U., Casertano S., Im M., Ellis R.S., Glazebrook K., Green R.F., & Sarajedini V.L., 1994. The Theta-z relation for HST bulges and disks out to Z approximately equal 0.8. *Astrophysical Journal*, 434 : L55. 52
- Peng C.Y., Ho L.C., Impey C.D., & Rix H.W., 2002. Detailed Structural Decomposition of Galaxy Images. *The Astronomical Journal*, 124 : 266. (c) 2002 : The American Astronomical Society. 47
- Petrosian V., 1976. Surface brightness and evolution of galaxies. *Astrophysical Journal*, 209 : L1. 51
- Roberts M.S., 1969. Integral Properties of Spiral and Irregular Galaxies. *Astronomical Journal*, 74 : 859. A&AA ID. AAA002.158.029. 44
- Sandage A., 1961. The Hubble atlas of galaxies. Washington : Carnegie Institution. 44
- Schade D., Lilly S.J., Crampton D., Ellis R.S., Fèvre O.L., Hammer F., Brinchmann J., Abraham R., Colless M., Glazebrook K., Tresse L., & Broadhurst T., 1999. Hubble Space Telescope Imaging of the CFRS and LDSS Redshift Surveys. III. Field Elliptical Galaxies at  $0.2 < z < 1.0$ . *The Astrophysical Journal*, 525 : 31. (c) 1999 : The American Astronomical Society. 53
- Schade D., Lilly S.J., Crampton D., Hammer F., Fèvre O.L., & Tresse L., 1995. Canada-France Redshift Survey : Hubble Space Telescope Imaging of High-Redshift Field Galaxies. *Astrophysical Journal Letters* v.451, 451 : L1. (c) 1995 : The American Astronomical Society. 52
- Simard L., Willmer C.N.A., Vogt N.P., Sarajedini V.L., Phillips A.C., Weiner B.J., Koo D.C., Im M., Illingworth G.D., & Faber S.M., 2002. The DEEP Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip. *The Astrophysical Journal Supplement Series*, 142 : 1. 48
- Toomre A. & Toomre J., 1972. Model of the Encounter Between NGC 5194 and 5195. *Bulletin of the American Astronomical Society*, 4 : 214. A&AA ID. AAA007.158.132. 53
- Tully R.B. & Fisher J.R., 1977. A new method of determining distances to galaxies. *Astronomy and Astrophysics*, 54 : 661. A&AA ID. AAA019.158.045. 44
- van den Bergh S., 1960. A Preliminary Luminosity Classification of Late-Type Galaxies. *Astrophysical Journal*, 131 : 215. 44

- van den Bergh S., Abraham R.G., Ellis R.S., Tanvir N.R., Santiago B.X., & Glazebrook K.G., 1996. A Morphological Catalog of Galaxies in the Hubble deep Field. *Astronomical Journal* v.112, 112 : 359. 52
- van Dokkum P.G. & van der Marel R.P., 2007. The Star Formation Epoch of the Most Massive Early-Type Galaxies. *The Astrophysical Journal*, 655 : 30. 53
- Williams R.E., Baum S., Bergeron L.E., Bernstein N., Blacker B.S., Boyle B.J., Brown T.M., Carollo C.M., Casertano S., Covarrubias R., de Mello D.F., Dickinson M.E., Espey B.R., Ferguson H.C., Fruchter A., Gardner J.P., Gonnella A., Hayes J., Hewett P.C., Heyer I., Hook R., Irwin M., Jones D., Kaiser M.E., Levay Z., Lubenow A., Lucas R.A., Mack J., MacKenty J.W., Madau P., Makidon R.B., Martin C.L., Mazzuca L., Mutchler M., Norris R.P., Perriello B., Phillips M.M., Postman M., Royle P., Sahu K., Savaglio S., Sherwin A., Smith T.E., Stiavelli M., Suntzeff N.B., Teplitz H.I., van der Marel R.P., Walker A.R., Weymann R.J., Wiggs M.S., Williger G.M., Wilson J., Zacharias N., & Zurek D.R., 2000. The Hubble Deep Field South : Formulation of the Observing Campaign. *The Astronomical Journal*, 120 : 2735. (c) 2000 : The American Astronomical Society. 52
- Williams R.E., Blacker B., Dickinson M., Dixon W.V.D., Ferguson H.C., Fruchter A.S., Giavalisco M., Gilliland R.L., Heyer I., Katsanis R., Levay Z., Lucas R.A., McElroy D.B., Petro L., Postman M., Adorf H.M., & Hook R., 1996. The Hubble Deep Field : Observations, Data Reduction, and Galaxy Photometry. *Astronomical Journal* v.112, 112 : 1335. 52

Première partie

Morphologies de galaxies à  
grand redshift avec optique  
adaptative



# Cosmologie et optique adaptative

---

## Sommaire

<b>3.1</b>	<b>Formation d'une image au foyer d'un télescope</b>	<b>66</b>
3.1.1	Cas idéal	66
3.1.2	L'effet de l'atmosphère	67
<b>3.2</b>	<b>L'optique adaptative</b>	<b>68</b>
3.2.1	Principe	68
3.2.2	FEP d'un système d'optique adaptative	70
3.2.3	Les limites de l'OA classique	71
<b>3.3</b>	<b>Relevés cosmologiques, morphologie et optique adaptative</b>	<b>72</b>
3.3.1	A la limite de la technique	72
3.3.2	Revue des travaux précédents	73
	<b>Bibliographie</b>	<b>74</b>

---

**Abstract :** We describe the basic principles of adaptive optics and show the major problems encountered when trying to use it for performing large cosmological surveys (limited field of view, variable PSF). We finally describe recent works which have been using adaptive optics for studying galactic morphology.

**Resumen :** Se describen a continuación los principios básicos de funcionamiento de la óptica adaptativa. Se subrayan especialmente los problemas encontrados a la hora de aplicar dicha técnica en el campo de la cosmología (variación de la respuesta al impulso, pequeñas áreas de observación). El capítulo concluye con una descripción de los trabajos anteriores basados en el uso de óptica adaptativa para el estudio de la morfología de galaxias.

Dans les chapitres précédents, nous avons vu que l'étude de l'évolution des galaxies requiert de remonter le temps en sondant des galaxies très lointaines et par conséquent très petites et faibles. Cela nécessite des instruments très sensibles et capables en même temps de scruter les plus petits détails du ciel. Le télescope spatial Hubble

(de 2.4m de diamètre) en est un et a permis effectivement de faire les plus grandes avancées dans la compréhension de l'Univers lointain durant les dernières décennies. Cependant nous avons maintenant des télescopes au sol qui font 8-10m de diamètre et bientôt 30m à 42m, une taille qu'on n'envisage pas pour des télescopes dans l'espace. La turbulence atmosphérique dégrade malheureusement la résolution angulaire des télescopes au sol. Corriger des effets de l'atmosphère pour ce type d'instruments permet d'obtenir des résolutions angulaires bien supérieures à celle de Hubble tout en bénéficiant de la surface collectrice. L'optique adaptative réalise cette correction mais avec certaines limites.

### 3.1 Formation d'une image au foyer d'un télescope

#### 3.1.1 Cas idéal

La lumière provenant des objets dans l'espace est une onde électromagnétique sphérique. Celle-ci se caractérise par un front d'onde, c'est-à-dire, une surface où le champ électrique est en phase. En absence de toute perturbation, l'amplitude dans un plan au niveau du front d'onde peut donc être décrite par :

$$\Psi(x) = |\Psi(x)| \exp(-i\phi(x)) \quad (3.1)$$

Avec  $\Psi(x)$  étant l'amplitude ou intensité et  $\phi(x)$  la phase qui dans ce cas reste constante.

La théorie de la diffraction nous dit qu'une onde lumineuse provenant d'une source ponctuelle (par exemple une étoile) qui pénètre dans un système optique tel un télescope va former sur un capteur situé au foyer de l'instrument non pas un point mais une « tâche ». Cette tâche est en fait la réponse impulsionnelle de l'instrument et reçoit le nom de *fonction d'étalement du point* (FEP). Plus précisément, la forme exacte de la FEP dépend de la pupille du télescope à travers une transformée de Fourier :

$$FEP \propto |TF[P]|^2 \quad (3.2)$$

Où P est la fonction décrivant la forme de la pupille. Pour une pupille circulaire, la FEP obtenue est appelée tâche d'Airy et la taille caractéristique de celle-ci, mesurée par la *largeur à mi-hauteur* (FWHM), en seconde d'angle<sup>1</sup>, dépend essentiellement du diamètre du télescope (D) et de la longueur d'onde d'observation ( $\lambda$ ) :

$$\theta = 0.25 \left( \frac{\lambda}{1\mu m} \right) \left( \frac{1m}{D} \right) \quad (3.3)$$

L'angle  $\theta$  est la *résolution angulaire* théorique de l'instrument et représente donc, en première approche, une mesure de la taille minimale des détails que l'on peut

<sup>1</sup>1 seconde =  $4.85 \times 10^{-6}$  radians =  $5.56 \times 10^{-4}$  degrés

distinguer à l'intérieur d'un objet astronomique avec un télescope donné. En effet, l'image d'un objet astronomique ayant une distribution d'intensité  $O$ , au foyer du télescope, est donnée par la convolution de celui-ci par la FEP (ou multiplication dans l'espace de Fourier) :

$$I = O(\mathbf{r}) * FEP(\mathbf{r}) \Leftrightarrow O(f) \times FTO(f) \quad (3.4)$$

$FTO$  étant la *fonction de transfert optique* qui n'est autre que la transformée de Fourier de la FEP. Celle-ci est bornée, c'est-à-dire que au delà d'une *certaine fréquence de coupure*, qui dépend du diamètre et de la longueur d'onde ( $\sim D/\lambda$ ), elle prend des valeurs nulles. L'équation 5.1 permet donc de voir clairement que l'effet du télescope est d'éliminer toute l'information de l'objet contenue au delà d'une certaine fréquence. Les détails trop petits ou les variations trop contrastées faisant intervenir des fréquences spatiales élevées sont perdus.

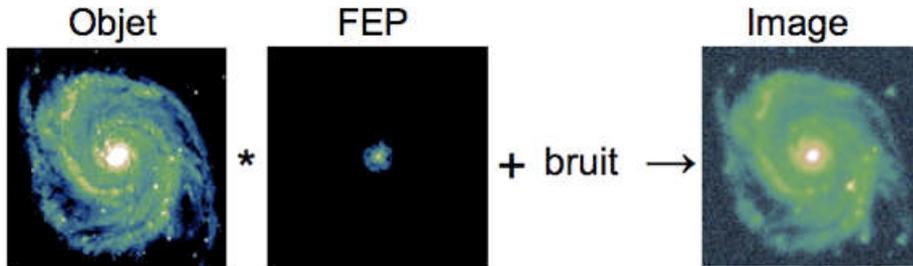


FIG. 3.1 – Illustration de la formation d'une image au foyer d'un télescope. [Source : E. Thiébaud]

D'après l'équation 3.3, en augmentant cependant le diamètre du télescope, on peut en théorie atteindre une résolution angulaire aussi haute que l'on veut et sonder alors les objets les plus lointains de l'Univers. Ainsi par exemple, une galaxie typique observée à  $z \sim 1$  a un diamètre angulaire dans le ciel de  $\sim 1''$ . Un télescope de 1m de diamètre suffirait pour résoudre sa structure.

### 3.1.2 L'effet de l'atmosphère

Malheureusement la nature n'est pas si bienveillante. Sur Terre, les rayons lumineux qui nous parviennent traversent l'atmosphère qui est un milieu turbulent et dégrade la planéité des fronts d'onde : elle agit en effet sur le front d'onde lumineux provenant de l'espace en introduisant des différences de chemins optiques. On dit alors qu'elle introduit une différence de phase : la phase de l'équation 3.1 n'est plus constante. L'échelle de ces perturbations peut être caractérisée par le paramètre de Fried ( $r_0$ , Fried, 1965) qui est en quelque sorte une mesure de la distance sur laquelle le front d'onde est suffisamment peu perturbé. L'expression de  $r_0$  dépend de la longueur d'onde ( $\sim \lambda^{6/5}$ ) ainsi que de paramètres décrivant la structure de l'atmosphère (masse d'air et profil de variation de l'indice de réfraction).

L'effet sur la qualité de l'image obtenue est accablant : la taille caractéristique de la FEP moyenne que l'on obtient n'est plus déterminée par le diamètre du télescope mais par le paramètre de Fried (si celui-ci est plus petit que le diamètre du télescope) :

$$\theta \sim \left(\frac{\lambda}{1\mu m}\right)\left(\frac{1m}{r_0}\right) \quad (3.5)$$

On assiste ainsi à une perte d'information forte au delà de la fréquence caractéristique dictée par l'atmosphère ( $\sim r_0/\lambda$ ) qui se traduit par une perte de contraste de l'image (Fig. 3.3). Cette nouvelle grandeur donnant la qualité de l'image obtenue en présence d'atmosphère est appelée *seeing* en anglais. Etant donné que le paramètre de Fried croît avec la longueur d'onde, il est intéressant de noter également que la qualité d'image obtenue sera meilleure à des longueurs d'onde plus grandes. L'analyse de cette expression nous dit d'autre part que pour un télescope terrestre, un diamètre plus grand n'apporte pas un gain en qualité d'image. Pourquoi alors construire des télescopes de plusieurs mètres de diamètre ? C'est une question de flux. Une ouverture plus grande permet de capter un plus grand nombre de photons et donc de détecter des objets de plus en plus faibles.

Cependant comme nous l'avons expliqué au chapitre précédent, l'étude des morphologies de galaxies requiert non seulement une bonne sensibilité mais aussi une très haute résolution angulaire.

Nous avons alors deux solutions : soit on se situe au dessus de l'atmosphère en construisant des télescopes spatiaux (e.g. télescope spatial Hubble, c.f. chapitre 2) ou alors on développe des techniques pour corriger des effets de l'atmosphère sur Terre (e.g. optique adaptative).

## 3.2 L'optique adaptative

### 3.2.1 Principe

De nombreuses techniques ont été proposées depuis les années 70 pour essayer d'améliorer la résolution angulaire des télescopes terrestres. [Labeyrie \(1970\)](#) a proposé par exemple la technique *d'interférométrie des tavelures* qui consiste en grandes lignes à figer la turbulence en prenant des images avec un temps de pose plus court que le temps caractéristique de variation de l'atmosphère. On peut ainsi remonter à des informations qui sont perdues lorsque l'image est moyennée comme le diamètre angulaire de certains objets ou la distance angulaire entre deux étoiles doubles.

L'idée de l'optique adaptative (OA) fut introduite par [Babcock \(1953\)](#). Les premières applications civiles pour l'astronomie datent cependant des années 90 ([Rodier, 1999](#)). Aujourd'hui, la grande majorité des télescopes terrestres de plus de 4m

de diamètre sont équipés avec un système d'OA.

L'idée derrière un tel système c'est d'essayer de redonner au front d'onde, déformé par l'effet de la turbulence atmosphérique, la forme originale qu'il avait avant de pénétrer dans l'atmosphère, de récupérer toute l'information contenue jusqu'à la fréquence de coupure théorique du télescope et de restaurer ainsi le contraste de l'image par rapport au cas turbulent. Pour cela on interpose sur le trajet lumineux de la lumière un miroir capable de se déformer pour corriger les déformations du front d'onde. Ce miroir dit *déformable* est réalisé en collant sous une surface réfléchissante souple, un certain nombre d'actionneurs qui permettent d'élever ou d'abaisser localement la surface réfléchissante pour que le miroir prenne des formes variées et corrige ainsi la forme du front d'onde. Afin de déterminer comment doit se déformer le miroir, des mesures du front d'onde sont réalisées à la sortie du miroir déformable. On analyse ainsi les aberrations résiduelles du front d'onde et on déforme le miroir en conséquence.

La figure 3.2 illustre le fonctionnement général d'un système d'optique adaptative : le miroir déformable est placé sur le trajet lumineux pour corriger les aberrations. Celui-ci est piloté par un système de contrôle qui reçoit les informations à son tour d'un système d'analyse de surface d'onde et les transforme en commandes pour le miroir. On dit que le système travaille en *boucle fermée* car l'analyse du front d'onde se fait à la sortie du miroir. Après fermeture de la boucle, l'analyseur ne mesure qu'un front d'onde résiduel et le miroir ne doit que réajuster légèrement sa forme.

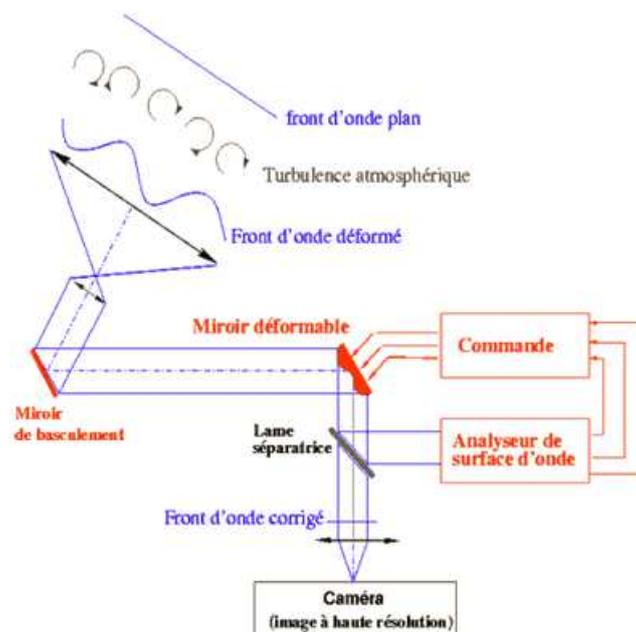


FIG. 3.2 – Principe de l'optique adaptative. [Source : T. Fusco]

### 3.2.2 FEP d'un système d'optique adaptative

L'intérêt donc de l'optique adaptative réside dans la possibilité d'obtenir une image corrigée des effets de la turbulence atmosphérique, stable pendant un temps très supérieur au temps de cohérence de l'atmosphère.

L'OA ne délivre cependant qu'une correction partielle car les possibilités de mesure (taille finie de l'analyseur de surface d'onde) et de reconstruction (nombre d'actionneurs du miroir fini) sont limitées. Ainsi la FEP obtenue est quelque part entre une FEP turbulente longue pose et une FEP théorique sans atmosphère. Dans la pratique, la FEP obtenue présente en général un pic ayant une largeur caractéristique donnée par le diamètre du télescope ( $\sim \lambda/D$ ) posé sur un halo ayant la largeur imposée par la turbulence ( $\sim \lambda/r_0$ ). Plus la correction est de bonne qualité, plus l'intensité du pic cohérent est élevée jusqu'à se rapprocher de la FEP idéale théorique (c.f. Fig. 3.3). La qualité de la correction dépend, entre autres, de la qualité de la mesure des aberrations du front d'onde qui dépend à son tour de la quantité de flux reçu. Ainsi, plus un objet est brillant, plus la correction sera de bonne qualité.

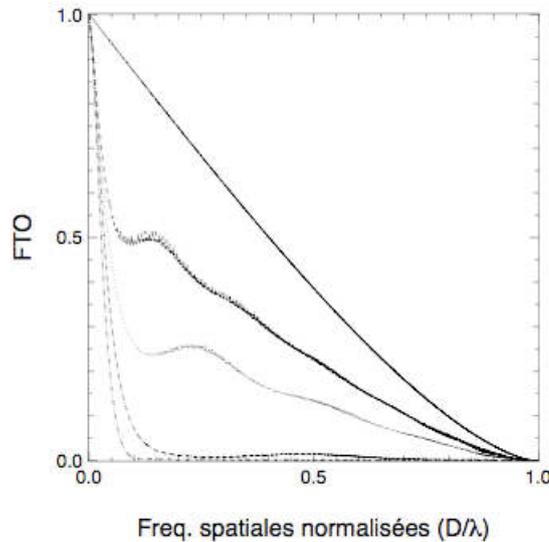


FIG. 3.3 – Exemples de coupes de la FTO longue pose pour différents niveaux de correction d'optique adaptative obtenues par simulation numérique. Trait plein : FTO théorique (limite de diffraction), pointillés : FTO après correction de 120 polynômes de Zernike, points : FTO après correction de 45 polynômes de Zernike, traits-points : FTO après correction de 10 polynômes de Zernike, traits-points-points : FTO turbulente, sans correction. [Source : F. Assémat]

Un paramètre permettant de quantifier plus précisément la qualité de la correction est le *rapport de Strehl* (SR). Celui-ci est défini justement comme étant le rapport entre l'intensité au centre de la FEP corrigée par OA et l'intensité maximale

théorique donnée par la limite de diffraction :

$$SR = \frac{I_{OA}(x=0)}{I_{diff}(x=0)} \quad (3.6)$$

Dans le cas d'une correction parfaite le rapport de Strehl vaut donc 1.

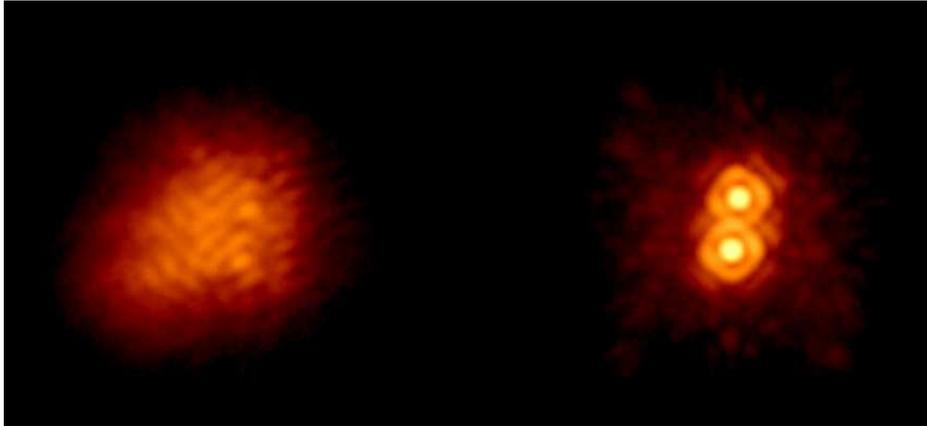


FIG. 3.4 – Démonstration du gain apporté par un système d'OA. A gauche image d'une étoile double sans correction. L'effet de l'atmosphère empêche de séparer les deux étoiles. A droite la même étoile observée avec un système d'optique adaptative. On aperçoit le coeur corrigé sur le halo non corrigé. [Source : Observatoire du Mont Palomar]

### 3.2.3 Les limites de l'OA classique

Comme on vient de le voir, pour qu'un système d'optique adaptative arrive à corriger correctement les effets de la turbulence, les mesures des aberrations par l'analyseur de surface d'onde doivent être réalisées sur des objets brillants. Si l'objet astrophysique que l'on veut observer est suffisamment brillant les mesures peuvent être réalisées directement sur celui-ci. Malheureusement ce n'est pas souvent le cas, en particulier pour des objets d'intérêt cosmologique qui se caractérisent par leur faible luminosité. La solution consiste alors à réaliser les mesures sur une étoile brillante proche de l'objet d'intérêt (*étoile guide ou de référence*).

Il existe une distance maximale de séparation entre l'étoile guide et l'objet d'intérêt. En effet si on observe deux sources séparées d'un angle  $\theta$  dans le ciel, la lumière qui nous parvient ne traverse pas les mêmes couches d'atmosphère et ne subit donc pas les mêmes aberrations. Si on essaye donc de corriger le deuxième objet à partir des mesures d'aberrations réalisées sur le premier on risque de ne pas corriger les bonnes déformations (Fig. 3.5). L'angle maximal sous lequel le front d'onde turbulent subit globalement les mêmes aberrations est appelé *angle d'isoplanétisme* et peut être

exprimé en fonction du paramètre de Fried (Fried, 1979) :

$$\theta_0 = 0.314 \frac{r_0}{\bar{h}} \quad (3.7)$$

où  $\bar{h}$  est la hauteur moyenne de la couche turbulente. Par conséquent si l'objet d'intérêt astrophysique se situe à une distance plus grande que l'angle isoplanétique de l'objet brillant servant à asservir la boucle, la gain de l'optique adaptative sera limité. Cela se traduira par une variation de la FEP dans le champ : à proximité de l'étoile de référence le rapport de Strehl sera maximal et diminuera au fur et à mesure que la distance à l'étoile augmente. On est donc contraints à observer des petits champs autour de l'objet de référence. Cela est spécialement critique pour des applications cosmologiques où l'on cherche justement à avoir les champs les plus grands possibles.

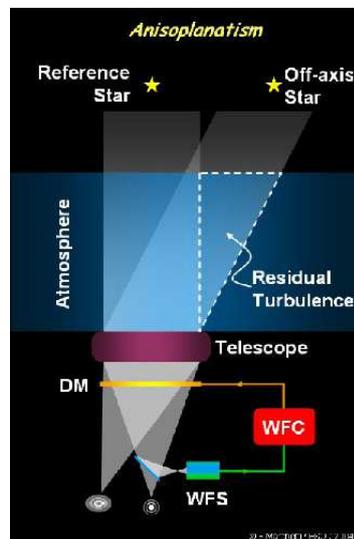


FIG. 3.5 – Illustration de la notion d'angle isoplanétique. Des objets sur l'axe et hors-axe ne voient pas la même couche turbulente [Source : ESO].

### 3.3 Relevés cosmologiques, morphologie et optique adaptative

#### 3.3.1 A la limite de la technique

Par la nature même des besoins en cosmologie observationnelle, l'utilisation de l'optique adaptative représente un défi technique remarquable. En effet, comme nous venons de le voir, l'OA classique a un fonctionnement optimal sur des objets brillants et pas très éloignés de l'objet d'asservissement. En cosmologie cependant, nous observons des galaxies lointaines et faibles. Il est indispensable alors d'utiliser

des étoiles suffisamment brillantes (typiquement  $m_v < 15$ ) pour réaliser les mesures d'aberration sur le front d'onde. Malheureusement, afin d'éviter l'extinction par la poussière de la Voie Lactée, les relevés de galaxies sont choisis à des latitudes proches des pôles galactiques. Ces régions sont par conséquent pauvres en étoiles assez brillantes pour l'OA et il devient particulièrement difficile de trouver des galaxies suffisamment proches d'une des ces étoiles. La *couverture du ciel*, c'est-à-dire le pourcentage du ciel que l'on peut atteindre, avec l'optique adaptative est donc très limitée. De plus, le but est de réaliser des mesures de grandeurs statistiques nous avons alors besoin d'observer un grand nombre d'objets, ce qui revient à observer des grands champs. On se situe par conséquent dans la zone de moindre performance de l'optique adaptative. Comme on le verra dans le chapitre 6, le défi de l'OA pour les années à venir est justement d'arriver à corriger des grands champs de la manière la plus uniforme possible.

### 3.3.2 Revue des travaux précédents

Malgré toutes ces difficultés, l'utilisation de l'optique adaptative pour des fins cosmologiques est une idée fortement séduisante. Pouvoir obtenir des données à la limite de diffraction avec des télescopes terrestres de 8m de diamètre ouvre les portes à l'étude des morphologies des toutes premières galaxies. De plus, son fonctionnement étant meilleur dans l'infrarouge que dans le visible, les données obtenues sont un complément parfait des données HST au delà de  $z \sim 1$  car elles sont moins affectées par des effets de k-corection morphologique. Certaines tentatives d'observations cosmologiques ont donc été réalisées durant les dernières années.

- Larkin *et al.* (2000) et Steinbring *et al.* (2004) réalisèrent pour la première fois des études de faisabilité en observant six galaxies avec le système d'optique adaptative du télescope Keck à Hawaï en employant des étoiles guides brillantes ( $V \sim 10$ ). Ils prouvèrent en réalisant des ajustements avec des modèles analytiques (c.f. chapitre 2) qu'il est possible d'étudier des morphologies sur des échelles proches de la diffraction.
- Minowa *et al.* (2005) ont été les premiers à réaliser un relevé cosmologique plus extensif avec le système d'optique adaptative du télescope Subaru à Hawaii. Ils observèrent un champ de  $1' \times 1'$  avec un temps d'intégration de 27h dans l'infrarouge proche (bande Ks). Ils ne réalisèrent pas d'étude morphologique mais démontrèrent cependant que l'utilisation de l'optique adaptative améliore nettement la détection de sources faibles en atteignant une limite de détection en bande Ks de 24.7.
- Les premiers à réaliser une analyse morphologique à valeur statistique sur des données d'OA sont Cresci *et al.* (2006). Ils observèrent, en utilisant le système d'optique adaptative du VLT, une aire de  $15' \times 15'$ , en bande  $K_s$  également, et obtinrent donc un échantillon de  $\sim 400$  galaxies. A l'aide des méthodes paramétriques exposées dans le chapitre précédent, ils classèrent les galaxies en deux grands types morphologiques et réalisèrent des comptages de galaxies par type.

- Cependant, l'absence des *redshifts* des galaxies ne leur permis pas d'aller plus loin dans l'interprétation. Ils soulèverent de plus la difficulté de l'estimation de la FEP dans ce type d'applications. En effet, étant donné qu'ils ont observé des champs largement plus grands que le champ isoplanétique, la correction au centre et sur les bords de l'image n'est pas la même et par conséquent la FEP varie dans le champ.
- Le projet CATS (*Center for Adaptive Optics Treasury Survey*, Koo *et al.*, 2007) : Ce projet tente de réaliser un relevé complet (avec *redshift*) de galaxies à partir de l'imagerie par optique adaptative. Ils ont réalisé en effet des observations de  $\sim 300$  galaxies avec le télescope Keck. Les galaxies ont été choisies dans des champs déjà observés par d'autres relevés afin de disposer des données sur plusieurs longueurs d'onde. Les données ont été mises en libre accès pour la communauté<sup>2</sup>. Cependant très peu de résultats ont été publiés à présent. En particulier, il n'existe pas d'étude quantitative de l'évolution des morphologies au cours du temps.

## Bibliographie

- Assémat F., 2004. *Apport de la haute résolution angulaire sur l'étude des galaxies lointaines : imagerie, optique adaptative et spectroscopie 3D*. Thèse de doctorat, Université Paris VII.
- Babcock H.D., 1953. What's in the Air. *Astronomical Society of the Pacific Leaflets*, 6 : 322. 68
- Cresci G., Davies R.I., Baker A.J., Mannucci F., Lehnert M.D., Totani T., & Minowa Y., 2006. Galaxy morphology and evolution from SWAN adaptive optics imaging. *Astronomy and Astrophysics*, 458 : 385. 73
- Fried D.L., 1965. Statistics of a Geometric Representation of Wavefront Distortion. *Journal of the Optical Society of America (1917-1983)*, 55 : 1427. 67
- Fried D.L., 1979. The nature of atmospheric turbulence effects on imaging and pseudo-imaging systems, and its quantification. Dans *IAU Colloq. 50 : High Angular Resolution Stellar Interferometry*, édité par J. Davis & W.J. Tango, p. 4-1. 72
- Fusco T., 2000. *Correction partielle et anisoplanétisme en optique adaptative : Traitements a posteriori et optique adaptative multiconjuguée*. Thèse de doctorat, Université de Nice Sophia-Antipolis.
- Gratadour D., 2005. *Optique adaptative, traitement d'image et étude des noyaux actifs de galaxie*. Thèse de doctorat, Université Paris VII.

<sup>2</sup><http://www.astro.ucla.edu/~irlab/cats/index.shtml>

- Koo D.C., Melbourne J., Max C., Metevier A., Ammons M., Larkin J.E., Barczys M., Wright S.A., & Steinbring E., 2007. CATS : CfAO Treasury Survey of distant galaxies, supernovae, and AGN's. Dans *IAU Symposium*, édité par F. Combes & J. Palous, volume 235 de *IAU Symposium*, p. 355–357. 74
- Labeyrie A., 1970. Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images. *Astronomy and Astrophysics*, 6 : 85. A&AA ID. AAA003.031.022. 68
- Larkin J.E., Glassman T.M., Wizinowich P., Acton D.S., Lai O., Filippenko A.V., Coil A.L., & Matheson T., 2000. Exploring the Structure of Distant Galaxies with Adaptive Optics on the Keck II Telescope. *PASP*, 112 : 1526. 73
- Minowa Y., Kobayashi N., Yoshii Y., Totani T., Maihara T., Iwamuro F., Takami H., Takato N., Hayano Y., Terada H., Oya S., Iye M., & Tokunaga A.T., 2005. Subaru Super Deep Field with Adaptive Optics I. Observations and First Implications. arXiv, astro-ph. 73
- Roddier F., 1999. Adaptive optics in astronomy. *Adaptive optics in astronomy / edited by Francois Roddier*. Cambridge ; New York. ISBN : 052155375X. 68
- Steinbring E., Metevier A.J., Norton S.A., Raschke L.M., Koo D.C., Faber S.M., Willmer C.N.A., Larkin J.E., & Glassman T.M., 2004. Keck Adaptive Optics Imaging of  $0.5 < z < 1$  Field Galaxies from the Hubble Space Telescope Archive. *ApJs*, 155 : 15. 73



# Morphologies de galaxies à $z \sim 1$ avec optique adaptative

---

## Sommaire

---

<b>4.1</b>	<b>Description des données</b>	<b>79</b>
4.1.1	Quelques mots sur l'instrument NACO	79
4.1.2	Observations	80
4.1.3	Réduction	82
<b>4.2</b>	<b>Construction du catalogue</b>	<b>84</b>
4.2.1	Détection	84
4.2.2	Nettoyage	84
4.2.3	Complétude	86
4.2.4	Comptages	86
4.2.5	<i>Redshifts</i> photométriques	88
4.2.6	Magnitudes absolues au repos	88
<b>4.3</b>	<b>Morphologies automatiques</b>	<b>90</b>
4.3.1	Considérations préliminaires	90
4.3.2	Détection des galaxies irrégulières	91
4.3.2.1	Principe	91
4.3.2.2	Remarques sur le calcul de C et A	93
4.3.2.3	Résultats	94
4.3.3	Analyse des galaxies régulières	94
4.3.3.1	Morphologies non-paramétriques	94
4.3.3.2	Morphologies paramétriques	96
<b>4.4</b>	<b>Résultats</b>	<b>105</b>
4.4.1	Statistiques globales	105
4.4.2	Evolution morphologique	106
<b>4.5</b>	<b>Quantification des effets de k-correction morphologique</b>	<b>109</b>
4.5.1	Comparaison globale	109
4.5.2	Evolution morphologique	111
<b>4.6</b>	<b>Conclusions de l'étude</b>	<b>112</b>
	<b>Bibliographie</b>	<b>114</b>

---

**Abstract** : We present the results of an imaging program of distant galaxies

( $z \sim 0.8$ ) at high spatial resolution ( $\sim 0.1''$ ) aiming at studying their morphological evolution. We observed 7 fields of  $1' \times 1'$  with the NACO Adaptive Optics system (VLT) in Ks ( $2.16\mu m$ ) band with typical  $V \sim 14$  guide stars and 3h integration time per field. Observed fields are selected within the COSMOS survey area, in which multi-wavelength photometric and spectroscopic observations are ongoing.

We show that reliable results can be obtained and establish a first basis for larger observing programs. We analyze the morphologies by means of B/D (bulge/disk) decomposition with GIM2D and C-A (concentration-asymmetry) estimators for 79 galaxies with magnitudes between  $K_s = 17 - 23$  and classify them into three main morphological types (late type, early type and irregulars). Automated and objective classification allows precise error estimation. Simulations and comparisons with seeing-limited (CFHT/Megacam) and space (HST/ACS) data are carried out to evaluate the accuracy of adaptive optics-based observations for morphological purposes.

We show that galaxy parameters (disk scale length, bulge effective radius, and bulge fraction) can be estimated with a random error lower than 20% for the bulge fraction up to  $K_s = 19$  ( $AB = 21$ ) and that classification into the three main morphological types can be done up to  $K_s = 20$  ( $AB = 22$ ) with at least 70% of the identifications correct. We used the known photometric redshifts to obtain a redshift distribution over 2 redshift bins ( $z < 0.8$ ,  $0.8 < z < 1.5$ ) for each morphological type and compare to the distributions obtained from the luminosity functions per spectral type. The classification is compared to the one obtained with HST/ACS data in the I band on the same objects. We find indeed some differences : the I band data tend to find more late-type systems, however the small number of objects does not enable to conclude safely about the nature of these differences. Most of the results are described in a paper published in A&A : Huertas-Company *et al.* (2007) that can be found in appendix E.

---

**Resumen :** Se presentan los resultados de un programa de observaciones de galaxias lejanas ( $z \sim 0.8$ ) con el sistema de óptica adaptativa NACO del VLT. Se han observado 7 campos de  $1' \times 1'$  en la banda Ks con el objetivo principal de realizar un estudio morfológico. Los campos han sido seleccionados en el área del *survey* COSMOS con lo que se dispone de los *redshifts* fotométricos.

Se muestra que es posible obtener una morfología fiable y se establecen las bases para programas de observación más amplios.

Se obtienen las morfologías paramétricas (GIM2D) y no paramétricas (concentración, asimetría) de 79 galaxias con magnitudes comprendidas entre  $K_s = 17 - 23$  en tres grandes tipos (espiral, elíptico, irregular). Se llevan a cabo simulaciones así como comparaciones con datos no corregidos (CFHT) y espaciales (HST) para evaluar la robustez de la clasificación.

Los *redshifts* fotométricos se emplean para obtener una distribución por tipos en función de  $z$  ( $z < 0.8$ ,  $0.8 < z < 1.5$ ). La comparación con una clasificación basada en datos del telescopio espacial en la banda I revela ciertas diferencias más importantes a *redshifts* elevados. El tamaño reducido de la muestra no permite sin

---

embargo extraer conclusiones fiables sobre la naturaleza de dichas diferencias. Los resultados han sido publicados en A&A : Huertas-Company *et al.* (2007). El lector podrá consultar dicha publicación en el anexo E.

---

Les travaux précédents (c.f. chapitre 3) ont tenté de montrer que malgré les difficultés techniques, des images de très haute qualité peuvent être obtenues grâce à l'utilisation d'optique adaptative. Il n'existe pas cependant aujourd'hui des études quantitatives de l'évolution des morphologies au cours du temps à partir de l'imagerie par optique adaptative.

Nous avons donc lancé un *programme pilote* d'observations au VLT avec l'instrument NACO afin de réaliser un premier pas dans cette direction. Ce chapitre détaille les résultats obtenus à l'issue de ce programme qui ont donné lieu à une publication dans A&A : Huertas-Company *et al.* (2007). Tout le long du chapitre les magnitudes seront exprimées dans le système Vega sauf indications explicites.

## 4.1 Description des données

### 4.1.1 Quelques mots sur l'instrument NACO

L'instrument NAOS-CONICA (NACO) à été installé sur l'un des foyers Nasmyth du quatrième VLT en décembre 2002. Il est composé d'une caméra infrarouge, équipée de modes spectroscopiques, polarimétriques, coronographiques et d'un certain nombre de filtres larges et étroits (CONICA, Lenzen *et al.*, 1998) ainsi que d'un système d'OA capable d'analyser la lumière dans le visible et l'IR suivant de nombreuses configurations, permettant de corriger les effets de la turbulence en temps réel (NAOS, Rousset *et al.*, 2000).

La partie OA de l'instrument est le fruit d'une collaboration entre l'Observatoire de Paris, l'ONERA et le Laboratoire d'Astrophysique de Grenoble (LAOG). Elle représente un effort de 65 personnes sur 5 ans pour un coût de 11.5 millions d'euros. Le retour pour les instituts ayant participé à sa conception est de 50 nuits de temps garanti pour réaliser des programmes scientifiques. Le programme que je présente para la suite a en effet débuté dans le cadre de ce temps garanti.

NAOS a une épaisseur de 70 cm, un poids de 2.3 tonnes et un diamètre de 2 m pour qu'il puisse être installé sur le foyer Nasmyth. Il est équipé d'un miroir déformable de 185 actionneurs possédant une bande passante de 650 Hz, permettant de corriger les effets de la turbulence atmosphérique en temps réel.

L'analyse du front d'onde se réalise grâce à un analyseur de type Shack-Hartmann fonctionnant dans le visible ( $0.45\text{--}1\mu\text{m}$ ) et dans l'infrarouge ( $0.8\text{--}2.4\mu\text{m}$ ). Les analyseurs peuvent adopter deux configurations différentes :  $7 \times 7$  et  $14 \times 14$  sous pupilles. Le premier est équipé d'une caméra CCD de  $128 \times 128$  pixels, et permet d'utiliser des sources de référence visibles jusqu'à une magnitude de 17. Le second est équipé

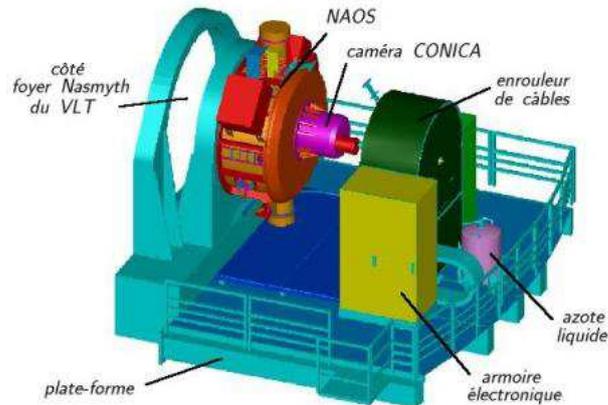


FIG. 4.1 – L'instrument NAOS-CONICA installé au VLT [Source : INSU].

d'un détecteur de  $1024 \times 1024$  pixels sensible jusqu'à la magnitude  $K = 12$ . NAOS est capable de délivrer un taux de correction supérieur à 40% (rapport de Strehl) dans la bande Ks avec une source brillante dans un cas de turbulence classique. Son angle isoplanétique est de l'ordre de  $20''$ .

#### 4.1.2 Observations

Nous avons observé sept champs de  $1' \times 1'$  dans la bande  $K_s$  ( $2.16\mu m$ ) sur une période de deux semestres<sup>1</sup>. Les champs ont été sélectionnés dans la région du relevé COSMOS (c.f. chapitres 1 et 10) permettant ainsi d'avoir de l'information multi-longueur d'onde de tous les objets. Afin d'assurer une correction par OA raisonnable, les champs ont été choisis autour d'étoiles brillantes ( $V \sim 14$ ). Un critère de couleur a été également rajouté ( $B - R < 1$ ) pour avoir une forte atténuation du flux stellaire dans l'infrarouge et réduire ainsi la région éblouie par l'étoile guide. La taille du pixel ( $0.054''$ ) a été choisie dans le but de maximiser le champ observé mais du coup les données sont sous-échantillonnées d'un facteur 2 par rapport au critère de Shannon-Nyquist. En effet, le théorème dit essentiellement que pour qu'un signal soit échantillonné correctement il faut au moins deux éléments de résolution (pixels) dans la FEP. Avec une correction parfaite, la largeur de la FEP des observations à  $2.16\mu m$  au VLT (diamètre  $\sim 8m$ ) vaut  $\lambda/D \sim 0.06''$ , on aurait donc moins de deux pixels. Comme on le montrera plus tard cela implique un certain nombre de difficultés de réduction, en particulier pour la reconstruction de la FEP, mais surmontables grâce au fait que la correction parfaite n'est jamais atteinte. La figure 4.2 montre les profils radiaux de 5 étoiles détectées sur l'ensemble des sept champs : on peut observer en effet que la largeur à mi-hauteur (FWHM) moyenne obtenue est de  $\sim 0.1''$ , plus large que la limite de diffraction du télescope par pratiquement un facteur 2.

<sup>1</sup>Programmes P73.A-0814A et P75.A-0569A

Champ	$\alpha$	$\delta$	temps. exp. (s)	FWHM (secondes d'angle)
STAR1	10 :00 :16	+02 :16 :22	10350	0.09
STAR2	10 :00 :52	+02 :19 :52	7650	–
STAR3	10 :00 :10	+02 :06 :08	7650	0.12
STAR4	09 :59 :52	+02 :05 :00	7170	0.08
STAR5	10 :00 :14	+02 :09 :09	10200	0.09
STAR6	10 :00 :02	+02 :06 :57	7650	0.13
STAR10	09 :59 :56	+02 :04 :07	10000	–

TAB. 4.1 – Résumé des observations pour les sept champs analysés. La résolution est calculée lorsque des étoiles non saturées ont été détectées dans le champ.

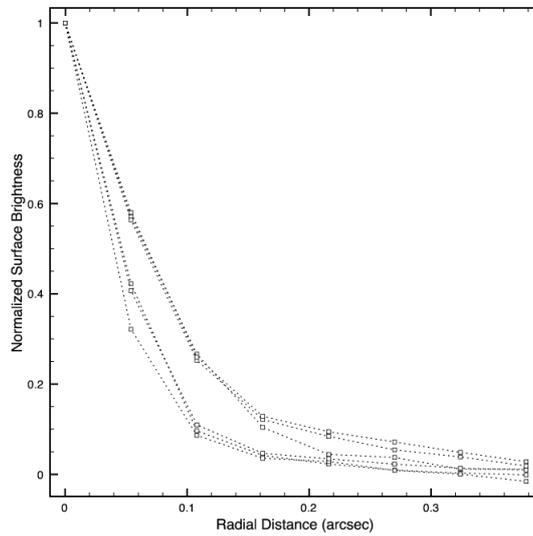


FIG. 4.2 – Profils radiaux de toutes les étoiles détectées plus brillantes que  $K_s \sim 19$ . La résolution moyenne est de  $\sim 0.1$  secondes d'angle. La limite de diffraction n'est pas atteinte à cause en particulier du long temps d'intégration.

### 4.1.3 Réduction

Le première partie du travail a été focalisée sur la réduction de données. Celle-ci a été réalisée de façon relativement standard. Les poses ont été acquises avec le mode *auto-jitter*, ce qui implique que le pointage du télescope est déplacé aléatoirement dans une boîte de 7" afin d'améliorer l'estimation du champ plat et de réduire l'effet des pixels morts. La valeur du ciel sur chaque pixel est obtenue à partir d'une moyenne tronquée en éliminant les 10% valeurs plus faibles et plus brillantes. Le recentrage est fait avec une précision sub-pixelique par maximum de corrélation entre les images. Il faut noter cependant que cette étape de réduction a été particulièrement difficile étant donné la nature des données. En effet, les galaxies observées étant très faibles, elles ne sont pas détectées sur les poses individuelles, ce qui empêche leur utilisation pour effectuer le recentrage. De plus, l'étoile guide ayant servi pour asservir la boucle d'optique adaptative était saturée sur une grande majorité des poses donc difficilement exploitable également. Finalement nous avons décidé d'utiliser les « ailes » de l'étoile guide après masquage du centre. Après recombinaison de toutes les poses individuelles, une estimation globale du fond de ciel à partir de la médiane spatiale de l'image est calculée. L'image finale est donc obtenue après soustraction de cette valeur.

Les points zéro photométriques (la magnitude correspondante à une ADU<sup>2</sup> dans le détecteur) ont été calculés en réalisant une photométrie d'ouverture sur des poses courtes des étoiles guides puis en comparant aux données disponibles sur le catalogue 2MASS (Kleinmann, 1992). Etant donné qu'il y a eu un changement de détecteur durant les observations, le point zéro varie d'une époque à l'autre : il vaut  $22.82 \pm 0.06$  pour le premier jeu de données et de  $23.29 \pm 0.06$  pour le deuxième.

---

<sup>2</sup>*Analog to Digital Units* en anglais. Unité de mesure du signal dans les images

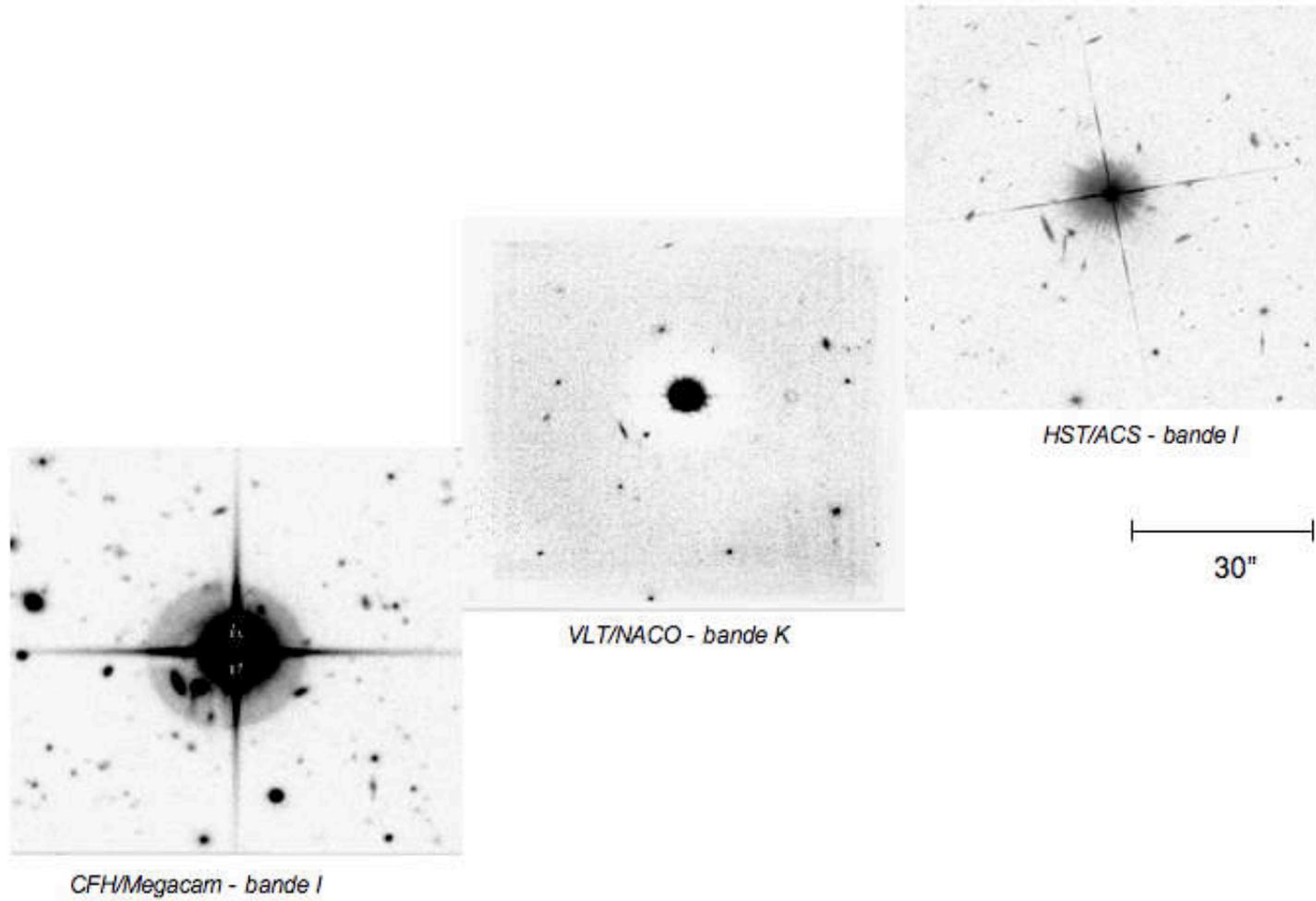


FIG. 4.3 – Même champ observé depuis l'espace avec le télescope spatial Hubble en bande I (en haut à droite), avec optique adaptative en bande Ks (milieu) et avec un télescope terrestre sans correction (CFHT, en bas à gauche).

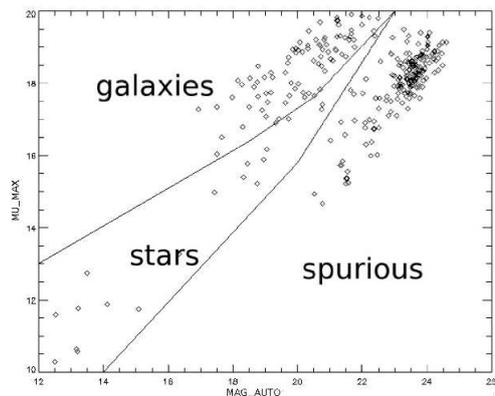


FIG. 4.4 – Classification des objets dans le plan MAG\_AUTO–MU\_MAX.

## 4.2 Construction du catalogue

### 4.2.1 Détection

Nous avons employé SExtractor (Bertin & Arnouts, 1996) pour la détection d’objets. Le seuil de détection a été fixé à  $3\sigma$  au dessus du fond de ciel sur au moins 4 pixels juxtaposés. Cette valeur peut paraître un peu petite, surtout pour une analyse morphologique, mais plusieurs raisons ont motivé ce choix. Tout d’abord nous voulions tester les performances de l’OA pour étudier les morphologies. Pour cela il est donc nécessaire d’atteindre les limites. Deuxièmement, nous voulions nous assurer que tous les objets sont détectés afin de comparer les comptages avec ceux obtenus sur d’autres observations infrarouges de la littérature (c.f. § 4.2.3).

### 4.2.2 Nettoyage

Deux-cents quatre-vingt cinq objets sur l’ensemble des 7 champs ont été détectés. Nous avons ensuite procédé à une étape de nettoyage afin de séparer les galaxies des étoiles et des fausses détections dues au bruit. Cela a été réalisé à l’aide de deux paramètres de SExtractor : MU\_MAX qui donne la valeur du pixel le plus brillant de l’objet, et MAG\_AUTO qui donne la magnitude apparente de la galaxie. Dans le plan défini par ces 2 grandeurs, les objets se situent clairement dans trois zones distinctes en fonction de leur nature comme on peut l’apprécier dans la figure 4.4. En effet, les objets ayant un pixel très brillant mais une magnitude totale élevée, sont considérés comme des fausses détections. Les frontières de séparation ont été tracées à la main puis une vérification visuelle postérieure a permis de confirmer que les étoiles connues dans le champ sont effectivement identifiées en tant que telles. Nous avons donc identifié finalement 79 galaxies, 19 étoiles (ou objets non résolus) et 187 fausses détections.

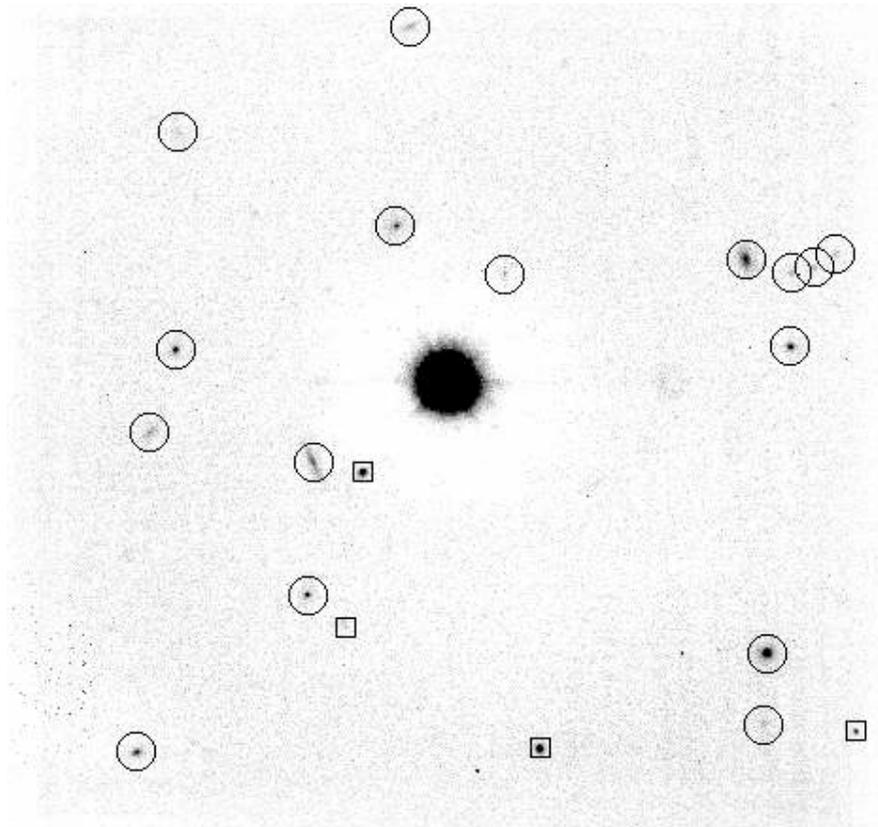


FIG. 4.5 – Image du champ centré sur  $\alpha = 10 :00 :16$ ,  $\delta = +02 :16 :22$ . Le temps d'intégration total est de 10350s. La taille du champ est 1 arcmin $\times$ 1 arcmin avec un pixel de 54mas. Les cercles marquent les galaxies et les carrés les étoiles. La largeur à mi-hauteur moyenne est de 0.1arcsec. L'étoile au centre est l'étoile utilisée pour asservir la boucle d'OA.

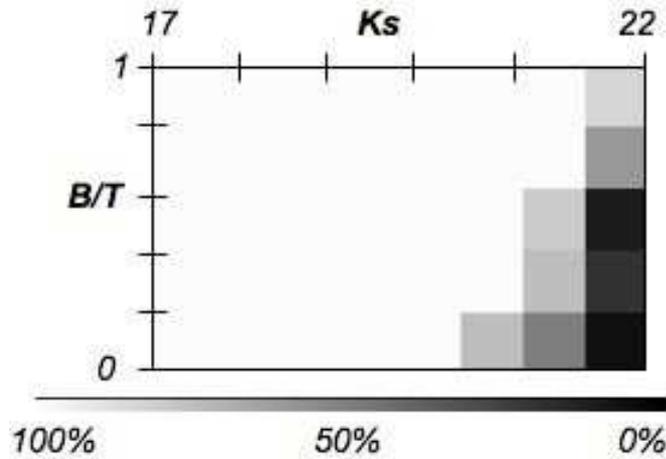


FIG. 4.6 – Complétude pour les sources étendues en fonction de la morphologie ( $B/T$ )

### 4.2.3 Complétude

La complétude d'un échantillon de galaxies mesure la limite de détection ou profondeur de l'échantillon, c'est-à-dire la magnitude à partir de laquelle on ne détecte plus les objets de façon significative. Cette grandeur permet ainsi de quantifier les effets de sélection et représente donc une étape indispensable préliminaire, à toute étude statistique. En général on mesure deux grandeurs : la complétude pour les sources ponctuelles (étoiles) et celle pour les sources étendues (galaxies).

- Afin d'estimer la complétude pour les sources ponctuelles, nous avons simulé des sources artificielles avec des magnitudes comprises entre  $K_s = 18$  et  $K_s = 24$  que nous avons plongé dans les champs réels. Celles-ci ont été simulées à partir de la FEP reconstruite (c.f. § 4.3.3.2). Nous avons ensuite fait tourner SExtractor avec les mêmes paramètres que pour les objets réels et regardé la fraction d'objets détectés. L'échantillon est complet à 50% à  $K_s = 22.5$  pour les sources ponctuelles.
- La complétude pour les sources étendues a été estimée d'une façon similaire. Nous avons généré 1000 galaxies synthétiques de différents types morphologiques et avec des paramètres distribués uniformément (c.f. § 4.3.1 pour plus de détails). On en déduit également que l'échantillon est complet à 50% à  $K_s = 21.5$  (Fig. 4.6).

### 4.2.4 Comptages

La connaissance précise de la complétude permet de réaliser des comptages de galaxies en fonction de la magnitude apparente. Les comptages constituent un test classique des modèles d'évolution de l'Univers. Il est cependant bien connu que ceux-ci représentent une information intégrée des galaxies soumise souvent à de fortes dégénérescences. Nous avons néanmoins voulu établir les comptages dans notre

échantillon afin de les comparer à d'autres observations profondes dans la bande Ks. Afin de pouvoir réaliser une comparaison directe avec les modèles, il est nécessaire cependant d'appliquer un certain nombre de corrections aux comptages bruts :

1. Tout d'abord il faut corriger de la complétude de l'échantillon pour les sources étendues. Cette fonction de complétude varie en fonction du type morphologique, les disques ayant généralement une brillance de surface plus faible que les bulbes, ils auront tendance à disparaître à des magnitudes plus élevées. Nous avons ainsi utilisé la fonction de complétude estimée dans la section 4.2.3 pour obtenir une probabilité de détection  $P_{det}$  en fonction de la fraction de bulbe estimée par GIM2D (c.f. chapitre 2, Simard *et al.*, 2002, § 4.3.3.2) pour chaque galaxie. De cette façon, une galaxie ayant une probabilité de détection  $P_{det}$  compte comme  $1/P_{det}$ .
2. Dans un deuxième temps, les comptages peuvent contenir des fausses détections qui peuvent devenir importantes pour les magnitudes plus faibles où le bruit devient plus important. Ces fausses détections ont été supprimées de l'échantillon en utilisant le plan *MAG\_AUTO/MU\_MAX* comme expliqué dans le paragraphe 4.2.2. Les étoiles sont également enlevées de l'échantillon à ce stade.
3. Finalement, les erreurs dans la photométrie peuvent induire des fluctuations dans les comptages à cause de galaxies qui passeraient dans d'autres intervalles de magnitude. Afin d'estimer les erreurs dues à cet effet nous avons généré une matrice de probabilités  $P_{ij}$  qui donne la probabilité pour qu'une galaxie de magnitude  $m_i$  soit détectée avec une magnitude  $m_j$ . Nous avons employé pour cela les simulations de galaxies utilisées pour estimer la complétude. En effet pour chaque galaxie simulée avec une magnitude donnée, nous avons obtenu sa magnitude mesurée par SExtractor. Les comptages ont été corrigés par la suite à partir de ces probabilités.

La figure 4.7 présente les comptages jusqu'à la magnitude  $K_s = 22$ . Au delà de cette magnitude, le facteur de correction que l'on doit appliquer aux galaxies détectées est tellement important que l'incertitude sur le nombre devient trop élevée. Nous avons réalisé par la suite un ajustement par une loi de puissance ainsi qu'une comparaison avec d'autres comptages en bande K de la littérature (Cresci *et al.*, 2006; Minowa *et al.*, 2005; Baker *et al.*, 2003; McCracken *et al.*, 2000; Ber-shady *et al.*, 1998; McLeod *et al.*, 1995). Le meilleur ajustement est une pente de  $d\log(N)/dm = 0.42 \pm 0.05$  qui est globalement en bon accord avec les valeurs mesurées par d'autres auteurs. Cette valeur est cependant significativement plus grande que celle obtenue dans le relevé SWAN (Cresci *et al.*, 2006), qui a été réalisé également avec de l'optique adaptative. En effet, ces auteurs trouvent une valeur de  $d\log(N)/dm = 0.26 \pm 0.01$  dans l'intervalle  $16 < K_s < 22$ . Cependant, leur sélection de champs présente un excès de galaxies brillantes, ce qui pourrait être à l'origine des différences.

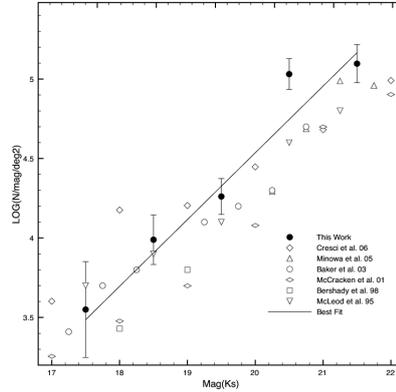


FIG. 4.7 – Comptages de galaxies comparés à d’autres relevés. La ligne droite représente le meilleur ajustement par une loi de puissance entre les magnitudes  $17 < K_s < 22$  avec une pente de  $d\log(N)/dm = 0.42 \pm 0.05$ . Les barres d’erreur sont des erreurs poissonniennes.

#### 4.2.5 *Redshifts* photométriques

Les comptages de galaxies apportent des informations intéressantes concernant la formation et l’évolution de celles-ci. Il existe cependant de fortes dégénérescences lorsque on essaye d’extraire des informations évolutives. Ces dégénérescences peuvent être brisées en introduisant l’information apportée par les *redshifts* photométriques qui permettent en effet d’isoler les populations à des époques différentes. C’est une des raisons pour laquelle les champs observés ont été choisis dans le champ COSMOS où des données multi-longueur d’onde sont disponibles.

L’estimation de *redshifts* photométriques a été réalisée à l’aide des données multi-longueur d’onde disponibles dans le relevé COSMOS (c.f. chapitres 1 et 10) en employant le code Le Phare (Ilbert *et al.*, 2006). Pour plus de détails sur la technique d’estimation de *redshifts* photométriques le lecteur peut consulter l’annexe A. La figure 4.8 montre l’histogramme de *redshifts* photométriques des 79 galaxies analysées. Le pic de la distribution est autour de  $z \sim 0.8$ .

#### 4.2.6 Magnitudes absolues au repos

Nous avons également estimé les magnitudes absolues des galaxies de l’échantillon dans la bande Ks au repos. Par définition, la magnitude absolue d’un objet est la luminosité qu’il aurait s’il était situé à 10pc. Il est donc nécessaire de connaître la distance de l’objet et appliquer un facteur de correction pour l’estimer. Lorsqu’il s’agit de distances cosmologiques, on parle de distance lumineuse et elle est calculée à l’aide du *redshift* en supposant un modèle cosmologique donné. A cela il faut ajouter un facteur de *k-correction* dû également au fait que les objets sont situés à différents *redshifts*. En effet, comme on l’a vu au chapitre 2, le flux reçu dans

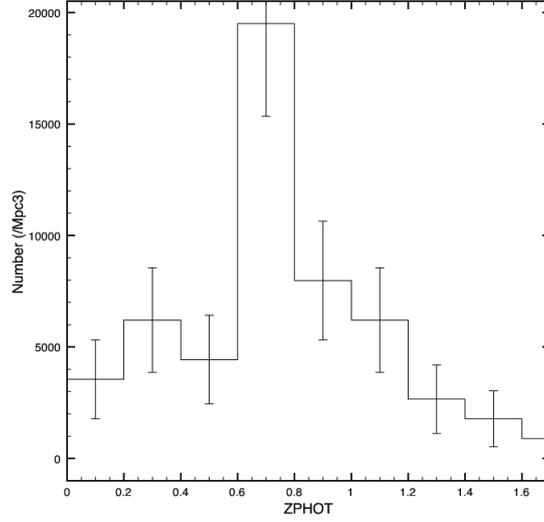


FIG. 4.8 – *Redshifts* photométriques estimés avec Le Phare (Ilbert *et al.*, 2006) pour l’ensemble des 60 objets détectés sur les champs COSMOS. Le *redshift* moyen est autour de  $z \sim 0.8$ . Les barres d’erreurs sont poissoniennes.

un filtre donné dépend du *redshift*. Il est donc nécessaire de ramener tous les flux reçus à une bande de référence appelée *bande au repos*. Ainsi, la relation entre la magnitude apparente et la magnitude absolue est :

$$m = M + DM(z) + K(z) \quad (4.1)$$

où  $m$  est la magnitude apparente,  $DM(z)$  est le module de distance et  $K(z)$  le facteur de k-correction. Le calcul de ce dernier fait nécessairement intervenir un modèle du spectre de la galaxie pour estimer le flux reçu lorsque le spectre est décalé vers le rouge.

Etant donné que le *redshift* est relativement bien connu (pour 60 objets) grâce à l’estimation du code Le Phare (c.f. § 4.2.5) nous avons employé le code HyperZ (Bolzonella *et al.*, 2000) pour estimer les magnitudes absolues. Ce code permet également d’estimer le *redshift* photométrique par ajustement de spectres (c.f. annexe A). Nous avons cependant ici fixé le *redshift* à la valeur estimée par Le Phare et obtenu la magnitude absolue par ajustement d’un spectre de galaxie théorique. Nous avons pour cela employé cinq bandes photométriques (u,g,r,i et z) en plus de la bande Ks issues du relevé CFHTLS. La figure 4.9 montre les magnitudes calculées en fonction du *redshift*. Cette représentation permet d’estimer la complétude de l’échantillon en fonction de  $z$ . On constate en effet que l’on détecte des objets plus brillants que  $K_s(AB) \sim -20$  jusqu’à  $z \sim 1.2$ .

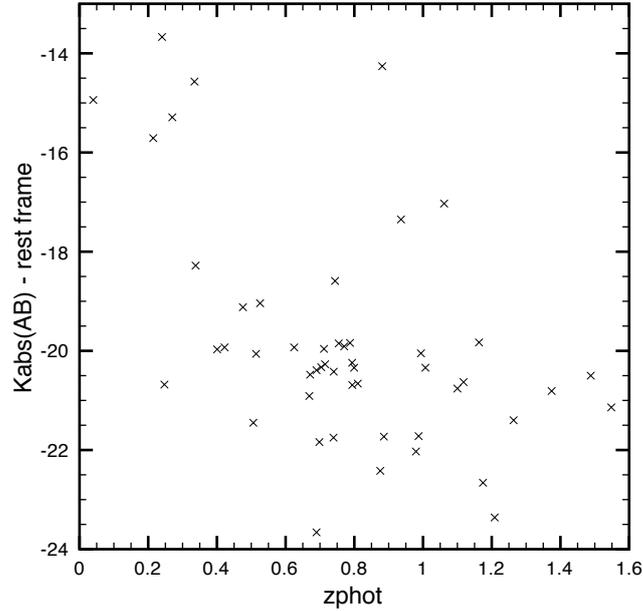


FIG. 4.9 – Magnitudes absolues dans la bande Ks au repos en fonction du *redshift* photométrique. Le *redshift* est estimé avec le code Le Phare et les magnitudes absolues avec HyperZ. L'échantillon est complet jusqu'à  $Ks(AB) \sim -20$  à  $z \sim 1.2$ .

### 4.3 Morphologies automatiques

Les 79 galaxies détectées dans l'ensemble des 7 champs ont été classées par type morphologique à l'aide des deux techniques automatiques présentées dans le chapitre 2. La première, paramétrique, est basée sur l'ajustement d'un modèle à la galaxie. Nous avons employé en particulier le logiciel GIM2D<sup>3</sup>. (Marleau & Simard, 1998; Simard *et al.*, 2002). La deuxième, non-paramétrique, consiste à mesurer des grandeurs qui gardent une corrélation avec la morphologie : les paramètres concentration et asymétrie tels que définis par Abraham *et al.* (1994, 1996). Pour plus de détails consulter le chapitre 2. Nous avons décidé de réaliser des classifications automatiques car celles-ci possèdent l'avantage d'être objectives et reproductibles, permettant ainsi une estimation précise des erreurs, qui est un des objectifs de cette étude.

#### 4.3.1 Considérations préliminaires

Avant de décrire les résultats des classifications proprement dits, je décris les caractéristiques principales des simulations qui ont été réalisées dans le but d'étalonner les classifications et d'estimer proprement les erreurs.

<sup>3</sup><http://www.hia-ihp.nrc-cnrc.gc.ca/STAFF/lst/gim2d/>

Nous avons donc considéré que le profil photométrique d'une galaxie est formé essentiellement de deux composantes : un bulbe décrit par un profil de De Vaucouleurs (de Vaucouleurs, 1959) et un disque ayant un profil exponentiel (Freeman, 1970) (c.f. chapitres 1 et 2). De cette façon, le type morphologique de la galaxie simulée est essentiellement déterminé par le rapport entre le flux contenu dans le bulbe (B) et le flux total (T). Ainsi, une galaxie est dominée par un bulbe (*bulge dominated*) lorsque  $B/T > 0.5$  et dominée par un disque (*disk dominated*) si  $B/T < 0.5$ . Les tailles caractéristiques du bulbe et du disque sont déterminées par les rayons caractéristiques  $r_e$  et  $r_d$  respectivement. Je rappelle l'expression employée :

$$\Sigma(r) = \frac{B}{T} \cdot \exp[-7.6697 \cdot \frac{r}{r_e}] + (1 - \frac{B}{T}) \cdot \exp[-\frac{r}{r_d}] \quad (4.2)$$

Nous avons donc généré des galaxies ayant des paramètres distribués uniformément dans l'espace de paramètres suivant :  $0 < B/T < 1$ ,  $0'' < r_d < 0.7''$ ,  $0'' < r_e < 0.7''$ ,  $0^\circ < i < 70^\circ$ , et  $0 < e < 0.7$ . L'objectif principal de ces simulations est de caractériser les erreurs de classification. Elles n'ont donc pas de caractère physique, d'où la distribution uniforme des paramètres. Pour la même raison on ne tient pas en compte des effets liés au *redshift* de la galaxie.

Afin de reproduire correctement les effets instrumentaux, les images sont convoluées par la FEP reconstruite des observations (c.f. § 4.3.3.2) et plongées dans un vrai fond. On obtient ainsi un échantillon de galaxies simulées analytiquement qui sera utilisé tout le long du processus de classification.

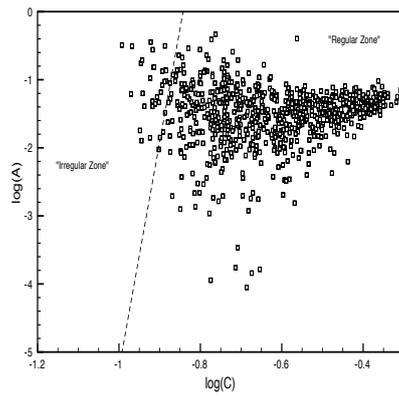
### 4.3.2 Détection des galaxies irrégulières

#### 4.3.2.1 Principe

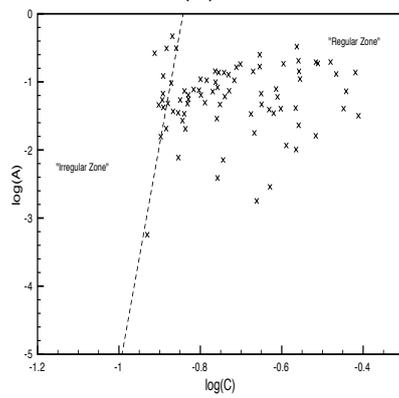
Tout d'abord, nous avons séparé les galaxies en deux grandes classes : irrégulières et régulières. Pour cela nous avons adopté une approche non-paramétrique en utilisant les paramètres concentration et asymétrie de la façon suivante : nous avons simulé un échantillon de galaxies synthétiques avec des magnitudes  $17 < K_s < 23$  et tracé les paramètres de concentration et d'asymétrie (Fig. 4.10). Ces galaxies synthétiques sont par définition symétriques et donc régulières. Nous avons donc raisonné par extrapolation afin d'établir la zone correspondant aux galaxies irrégulières :

- D'une part, les galaxies irrégulières doivent en principe avoir un profil photométrique plus plat (donc être moins concentrées) et doivent être plus asymétriques.
- D'autre part, la zone irrégulière doit être dépeuplée de galaxies simulées puisque ces galaxies sont par définition régulières.

La « zone irrégulière » est donc définie comme la zone « en haut à gauche » du plan C–A où on ne trouve pas ou peu de galaxies simulées.



(a)



(b)

FIG. 4.10 – Séparation des galaxies régulières des galaxies irrégulières. (a) Simulations. (b) Galaxies réelles.

## 4.3.2.2 Remarques sur le calcul de C et A

Dans la section précédente nous avons employé les paramètres concentration et asymétrie. L'expression exacte de ces paramètres ainsi qu'une revue détaillée des méthodes paramétriques ont été présentées dans le chapitre 2. La concentration est calculée à l'aide de l'expression 2.2 comme le rapport entre les flux contenus dans les rayons normalisés 0.3 et 1 dans une aire isophotale contenant les pixels  $3\sigma$  au dessus du niveau du ciel. Le centre de rotation est la position du maximum de l'image filtrée par un filtre gaussien de  $\sigma = 1$ . La brillance de surface limite varie entre  $\mu = 18.76 - 20.39 \text{ mag.arcsec}^{-2}$  à cause des variations du temps d'exposition entre les champs et de la variation intrinsèque du fond de ciel qui est importante pour des observations infrarouges au sol. La valeur de concentration calculée est très sensible à l'estimation du fond puisqu'une erreur sur cette valeur implique une variation dans les isophotes limites et donc une fraction du flux de la galaxie peut être perdue. Afin d'estimer l'erreur introduite dans le calcul de C à cause de ces variations, nous avons calculé les variations sur la valeur estimée pour des variations de la brillance de surface de  $\Delta\mu = 1$ . La variation moyenne sur C étant petite ( $\Delta C \sim 0.06$ ) nous avons décidé de ne pas appliquer de corrections particulières. L'asymétrie de la galaxie est calculée en prenant la valeur absolue de la différence entre l'image de la galaxie et l'image tournée de  $180^\circ$  tout en appliquant une correction pour le bruit (Eq. 2.4). Le calcul de l'asymétrie est, malgré tout, affecté par le bruit de l'image ce qui fait que des galaxies faibles peuvent présenter une asymétrie élevée même après correction. Afin de quantifier cet effet nous avons tracé l'asymétrie en fonction de la magnitude pour un ensemble de 1000 galaxies simulées (Fig. 4.11). On observe que la valeur calculée ne commence à être affectée par le bruit que pour des magnitudes plus faibles que 22.2 qui est au dessus de la limite de complétude de l'échantillon étudié.

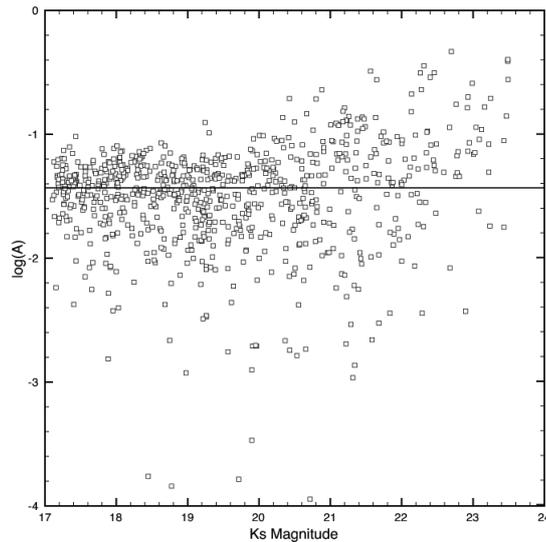


FIG. 4.11 – Asymétrie en fonction de la magnitude pour 1000 galaxies simulées.

### 4.3.2.3 Résultats

On compte 10 galaxies réelles dans la zone définie comme irrégulière, c'est-à-dire,  $\sim 12\%$  de l'échantillon. L'erreur sur cette valeur peut être estimée en regardant le nombre d'objets simulés qui se situent dans cette zone irrégulière. On trouve en effet 27 sur 1000 galaxies simulées ce qui nous permet de conclure que  $12\% \pm 2.7$  des galaxies correspond à des galaxies irrégulières.

## 4.3.3 Analyse des galaxies régulières

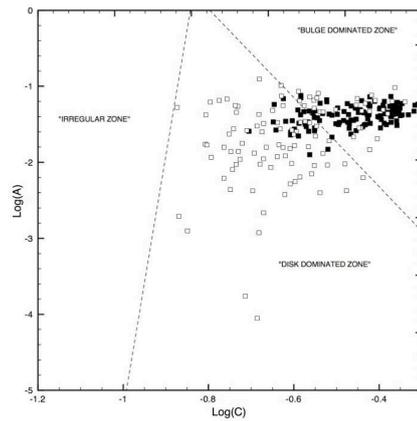
### 4.3.3.1 Morphologies non-paramétriques

**Etalonnage** Comme pour le cas des galaxies irrégulières, il est indispensable d'étalonner le plan C/A avant de procéder à la classification non-paramétrique. Il s'agit donc ici de trouver la frontière optimale entre les galaxies à bulbe dominant et celles dominées par un disque. L'avantage est que nous considérons maintenant des galaxies en principe symétriques et qui peuvent donc être simulées plus facilement. Nous avons donc modélisé des galaxies avec des rapports B/T allant de 0 à 1 pour lesquelles la morphologie est donc connue et avons représenté les paramètres de concentration et asymétrie dans le plan (Fig. 4.12). Nous avons par la suite tracé la frontière linéaire séparant les deux populations de galaxies. L'approche adoptée ici diffère légèrement de celle employée dans la littérature où le tracé de la frontière est généralement réalisé à la main ce qui comporte une certaine part d'arbitraire et de subjectif. Elle a été tracée en effet de manière automatique à l'aide d'un type particulier de « machines » d'apprentissage appelé *machines à vecteurs support* (SVM)<sup>4</sup>. Ce type de machine cherche la frontière optimale entre deux ou plusieurs familles de points. Le lecteur pourra trouver une description détaillée du mode de fonctionnement de celles-ci dans la deuxième partie de cette thèse où une nouvelle méthode de classification non-paramétrique basée sur les SVM est présentée. Pour une compréhension correcte de ce chapitre il suffit cependant de retenir que l'algorithme retrouve la frontière optimale (linéaire ou non) entre les populations en se basant essentiellement sur des critères de distance. Pour cette application en particulier nous avons forcé une frontière linéaire dans le but de pouvoir comparer plus facilement avec les travaux précédents. Cette approche a été choisie, par rapport à une approche classique car elle permet de réduire le nombre d'interventions subjectives ; le tracé manuel peut en effet varier d'une personne à l'autre alors que cette approche est complètement reproductible.

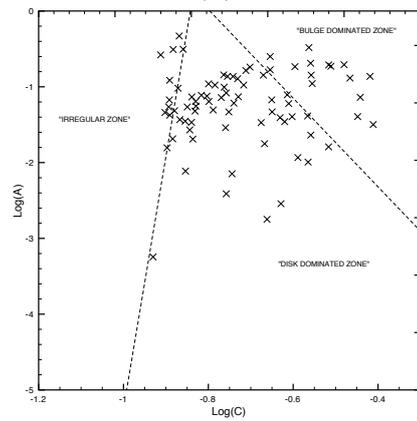
**Quantification des erreurs** La méthode employée permet également de quantifier les erreurs d'une manière précise et objective. Une fois la frontière tracée un

---

<sup>4</sup>SVM est la notation souvent employée pour ce type de machines, de l'anglais *Support Vector Machines*



(a)



(b)

FIG. 4.12 – Etalonnage du plan  $C/A$ . (a) Simulations, carrés vides : objets avec  $B/T < 0.2$ , carrés remplis : objets avec  $B/T > 0.8$ . (b) Objets réels. La frontière entre les galaxies à bulbe dominant et celles à disque dominant a été tracée à l'aide des SVM.

deuxième échantillon de 500 galaxies simulées pour lesquelles la morphologie est donc connue a été simulé et classée avec la frontière obtenue grâce aux premières simulations. Nous avons ensuite comparé les morphologies simulées et les morphologies estimées. Le tableau 4.2 illustre les résultats par intervalles de magnitude. On atteint 70% d'identifications correctes jusqu'à une magnitude  $K_s = 21$ .

Magnitude	Pourcentage de classifications correctes
$K_s < 19$	80%
$K_s < 20$	73%
$K_s < 21$	70%
$K_s < 22$	66%

TAB. 4.2 – Quantification des erreurs de la classification C-A.

#### 4.3.3.2 Morphologies paramétriques

L'estimation paramétrique des morphologies a été réalisée avec le logiciel GIM2D. Le principe de fonctionnement a été présenté au chapitre 2. Il est basé sur l'ajustement d'un modèle à la galaxie après convolution par la FEP des observations. Une description détaillée du logiciel est présentée en particulier dans [Marleau & Simard \(1998\)](#); [Simard \*et al.\* \(2002\)](#)

**Reconstruction de la FEP** Une utilisation propre de GIM2D requiert une bonne estimation de la FEP non bruitée et bien échantillonnée. Il existe des techniques classiques pour reconstruire une FEP à partir d'étoiles du champ (Tiny Tim, DAOPHOT). Cependant celles-ci ne peuvent pas être utilisées dans notre cas pour deux raisons principalement. Tout d'abord, la forme de la FEP d'optique adaptative est très spécifique. En effet, étant donné que le système ne réalise qu'une correction partielle des effets de la turbulence atmosphérique, la FEP est généralement formée par un noyau de largeur fixée par la diffraction superposé sur un halo limité par le *seeing* (c.f. chapitre 3). Deuxièmement, dans notre cas particulier, elle est sous-échantillonnée d'un facteur 2 (0.054" par pixel alors que 0.02" sont nécessaires pour respecter les critères de Nyquist).

Nous avons ainsi développé un algorithme simple qui permet d'obtenir une approximation de premier ordre de la FEP en utilisant des étoiles du champ. Pour cela nous générons d'abord un modèle de FEP constitué d'un noyau à la diffraction et d'un halo gaussien :

$$\text{FEP}_{art}(x, y) = k \left[ \text{SR} \times A(x, y) + (1 - \text{SR}) \times \exp\left(-\frac{x^2 + y^2}{\sigma^2}\right) \right]$$

SR représente le rapport de Strehl,  $A(x, y)$  la fonction d'Airy bi-dimensionnelle,  $\sigma$  la dispersion de la gaussienne qui est corrélée avec le seeing et  $k$  un facteur de

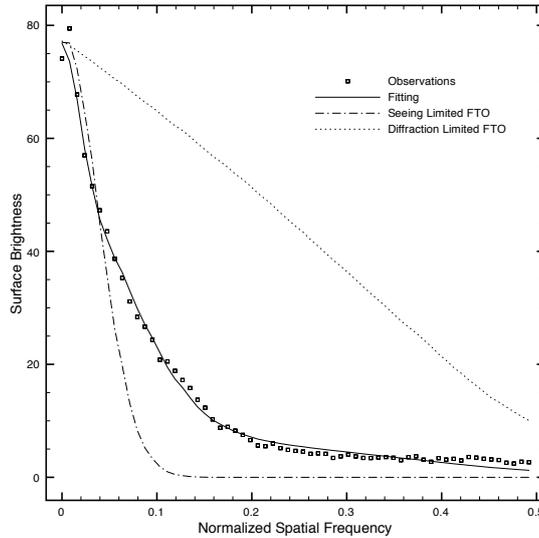
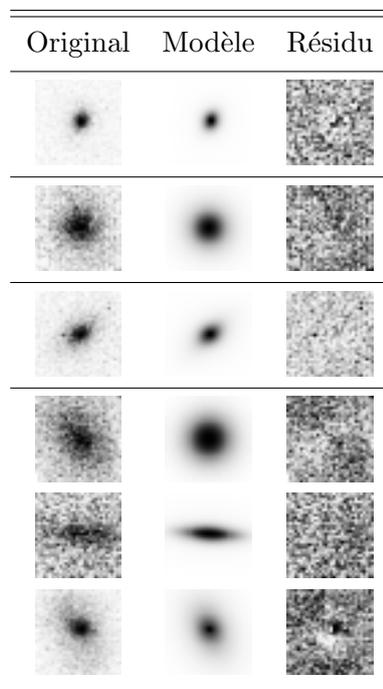


FIG. 4.13 – Exemple d’ajustement dans l’espace de Fourier. Carrés : Observations, ligne continue : ajustement, ligne discontinue : FTO d’observations limitées par le seeing, ligne en pointillés : FTO d’observations limitées par la diffraction. Les observations d’optique adaptative se situent donc entre les deux.

normalisation. On construit donc cette FEP avec un échantillonnage de Shannon ( $0.02''/\text{pixel}$ ), elle est ensuite rééchantillonnée à la taille de pixel des observations puis finalement on réalise une transformation de Fourier afin d’obtenir la fonction de transfert optique (FTO). D’un autre côté, pour chaque étoile détectée dans un champ donné on réalise également une transformation de Fourier. Finalement, on fait un ajustement à trois paramètres libres ( $SR$ ,  $\sigma$ ,  $k$ ) entre le modèle et les étoiles dans l’espace de Fourier. Les paramètres correspondant au meilleur ajustement sont ensuite utilisés pour obtenir une estimation de la FEP avec un échantillonnage à Shannon. Travailler dans l’espace de Fourier permet d’éliminer deux paramètres d’ajustement qui interviennent dans l’espace direct : l’estimation du niveau de fond ainsi que les coordonnées du centre de la FEP. Ces paramètres sont en effet particulièrement difficiles à estimer dans notre cas étant donné que la FWHM vaut moins de 2 pixels. La figure 4.13 montre le résultat de l’ajustement pour une étoile dans l’espace de Fourier. On peut observer que cette description simple à 3 paramètres suffit pour récupérer la forme générale de la FEP. Pour quelques étoiles cependant, la procédure d’ajustement n’arrivait pas à converger correctement : nous avons donc introduit pour ces cas, un deuxième halo gaussien.

**Analyse des erreurs** Nous avons fait tourner GIM2D sur l’ensemble de 79 objets détectés avec des magnitudes comprises entre  $K_s = 17$  et  $K_s = 22$  en utilisant la

FEP reconstruite avec la méthode exposée précédemment. L'ajustement a convergé pour toutes les galaxies. Une inspection visuelle montre un bon accord entre les modèles et les galaxies réelles ainsi que des résidus raisonnables (Tab. 4.3). Pour les objets les plus faibles cependant, il devient beaucoup plus difficile d'estimer la qualité de l'ajustement. En effet, les résidus peuvent être très bons même en présence de fortes dégénérescences qui peuvent produire un modèle faux. Nous avons donc décidé de réaliser une estimation systématique et objective des erreurs. Nous avons pour cela effectué des simulations de 1000 galaxies synthétiques (c.f. § 4.3.1) avec des paramètres galactiques connus et plongé ces galaxies dans un fond réel afin de reproduire au mieux les bruits de fond et de détection. Nous avons par la suite fait tourner GIM2D sur ces galaxies simulées et examiné les erreurs sur les paramètres récupérés en fonction des paramètres simulés. Etant donné que la même FEP est employée pour produire les simulations, les erreurs calculées ne contiennent pas des erreurs éventuelles dues à une mauvaise estimation de celle-ci.



TAB. 4.3 – Exemple de résultats d'ajustement avec GIM2D. Gauche : images originales, Milieu : modèles GIM2D, Droite : Résidus. La taille des images est de  $1.7' \times 1.7'$ .

Nous avons adopté, pour la représentation des erreurs, la méthode de [Simard \*et al.\* \(2002\)](#) qui consiste à représenter des plans 2-D donnant les biais et les dispersions pour un paramètre donné en fonction de deux autres paramètres. En effet, étant donné que l'ajustement se réalise dans un espace complexe de 11 dimensions il est impossible d'avoir une représentation des erreurs en fonction de tous les paramètres simultanément. Ces plans 2-D ne donnent par conséquent qu'une vision partielle de l'espace de paramètres. Puisque l'information principale que l'on souhaite récupérer

est le type morphologique, nous avons centré nos analyses sur le paramètre donnant la fraction de bulbe ( $B/T$ ) en fonction de deux paramètres fondamentaux de la galaxie : la luminosité et la taille. La luminosité est représentée par la magnitude apparente et la taille par le rayon contenant la moitié du flux de la galaxie ( $r_{hl}$ ). Pour chaque intervalle de « magnitude- $r_{hl}$  » on calcule donc la différence moyenne entre la fraction de bulbe simulée et celle récupérée :  $\overline{\Delta B/T} = \frac{\sum(B/T_i - B/T_r)}{N}$  qui donne le biais et la racine carrée de la variance de la différence qui donne la dispersion :  $\sigma \Delta B/T = \sqrt{\frac{\sum(\Delta B/T - \overline{\Delta B/T})^2}{N-1}}$ . Le tableau 4.4 montre les résultats obtenus.

Le constat principal, après une première inspection des valeurs obtenues, est que pour des magnitudes plus brillantes que  $K_s \sim 19$  ( $AB \sim 21$ ), la fraction de bulbe est récupérée avec un biais presque nul ( $\sim 0.02$ ) et une dispersion de l'ordre de 20%. Cela est vrai même pour les objets plus petits ( $-1 < \log(r_{hl}) < -0.75$ ) et est comparable aux résultats obtenus pour les objets les plus brillants avec le télescope spatial (Simard *et al.*, 2002).

Pour les objets plus faibles, nous constatons deux effets principalement :

- Un accroissement progressif du biais systématique qui passe à un niveau de l'ordre de 20% spécialement pour les sources plus étendues ( $\log(r_{hl}) > -0.75$ ). Ceci est un effet bien connu de GIM2D (Simard *et al.*, 2002). Les ailes d'un profil classique de *de Vaucouleurs* se trouvent noyées dans le bruit ce qui provoque une sous estimation systématique de la fraction de bulbe récupérée. A ceci peuvent s'ajouter également des erreurs dans l'estimation du fond de ciel qui font également qu'une partie du flux soit perdue.
- Deuxièmement, une légère augmentation de la dispersion qui atteint une valeur de 30%. Ceci est simplement une conséquence d'une accentuation de la dégénérescence lorsque le rapport S/B diminue.

**Comparaison avec des données sol** Afin de définir plus précisément les performances de l'optique adaptative pour l'analyse morphologique de galaxies lointaines, nous comparons nos observations avec des données sol non corrigées.

Nous avons donc comparé nos images avec des données en bande  $i$  limitées par le seeing appartenant au CFHTLS (Canada France Hawai Telescope Legacy Survey, c.f. chapitre 1). Un des champs profonds de ce relevé couvre en effet une partie l'aire observée par COSMOS. Nous avons utilisé ici le « release » T0003 de mars 2006. L'image  $i$  compte 20h d'intégration et la résolution moyenne est de 0.7".

Nous avons donc sélectionné une galaxie classée disque par les deux méthodes de classification avec une magnitude  $K_{AB} = 20.5$  et  $i = 21.3$ , afin d'être le plus sûr possible de son type morphologique et regardé cette même galaxie sur les données CFHTLS. Le profil radial moyen est obtenu sur une surface isophotale contenant tous les pixels ayant un signal  $3\sigma$  au dessus du bruit. Ce seuil correspond à une

Magnitude	$-1 < \log(r_{hl}) < -0.75$			$-0.75 < \log(r_{hl}) < -0.5$		
	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	N	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	N
[17 – 17.5]	-0.110	0.247	(19)	-0.034	0.140	(22)
[17.5 – 18]	0.165	0.088	(4)	0.037	0.277	(18)
[18 – 18.5]	-0.002	0.141	(19)	0.020	0.262	(58)
[18.5 – 19]	-0.114	0.324	(14)	0.021	0.309	(71)
[19 – 19.5]	-0.064	0.221	(20)	0.212	0.266	(90)
[19.5 – 20]	0.100	0.259	(24)	0.270	0.298	(110)
[20 – 20.5]	0.105	0.430	(36)	0.163	0.332	(107)
[20.5 – 21]	0.050	0.476	(56)	0.148	0.351	(81)

Magnitude	$-0.5 < \log(r_{hl}) < -0.25$			$-0.25 < \log(r_{hl}) < 0$		
	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	N	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	N
[17 – 17.5]	-0.002	0.056	(4)	-0.030	0.060	(8)
[17.5 – 18]	0.027	0.157	(23)	0.010	0.110	(11)
[18 – 18.5]	0.162	0.174	(72)	0.226	0.155	(28)
[18.5 – 19]	0.213	0.196	(83)	0.170	0.159	(30)
[19 – 19.5]	0.192	0.252	(51)	0.145	0.027	(3)
[19.5 – 20]	0.181	0.258	(34)	0.140	0.100	(6)
[20 – 20.5]	0.113	0.281	(13)	N/A	N/A	(0)
[20.5 – 21]	0.194	0.054	(3)	N/A	N/A	(0)

TAB. 4.4 – Analyse des erreurs dans l’estimation du rapport  $B/T$  en fonction de la magnitude et la taille des galaxies. La taille de la galaxie est représentée par le rayon contenant la moitié du flux ( $r_{hl}$ ). Les objets petits et brillants se trouvent en haut à gauche et les faibles et grands en bas à droite.  $\overline{\Delta B/T}$  c’est la différence moyenne entre la valeur de  $B/T$  simulée et la la valeur récupérée et  $\sigma\Delta B/T$  représente la dispersion moyenne

brillance de surface limite de  $\mu = 20 \text{ mag.arcsec}^{-2}$  pour les images NACO et de  $\mu = 25 \text{ mag.arcsec}^{-2}$  pour les images MegaCam. Le niveau du fond de ciel est estimé avec SExtractor.

Nous avons réalisé par la suite un ajustement par un profil de type de Vaucouleurs et de type exponentiel (après convolution par la FEP) afin d'estimer les capacités des deux types d'observations pour distinguer les différents profils. La figure 4.14 montre les résultats des ajustements. Nous constatons que pour les observations non corrigées il est plus difficile de faire la différence entre les deux types de profils pour des distances proches du centre de la galaxie ( $\sim 0.2''$ ) : le gain apporté par la correction d'optique adaptative est donc clair.

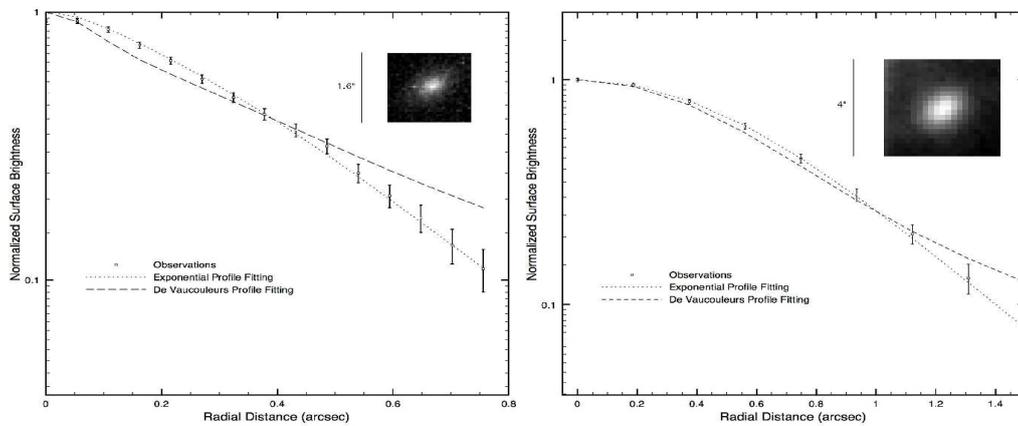


FIG. 4.14 – Comparaison du profil radial d'une galaxie spirale observée depuis le sol ; ajustement d'un profil de Vaucouleurs et d'un profil exponentiel. Gauche : NACO, droite : MegaCam. La galaxie analysée est située à  $\sim 20''$  de l'étoile guide.

**Remarques sur les variations de la FEP** Le modèle de FEP employé précédemment permet d'obtenir une estimation simple de la réponse impulsionnelle. Il ne tient pas en compte cependant d'importants effets liés à l'utilisation de l'optique adaptative. Comme il a été expliqué dans le chapitre précédent, la correction apportée par cette technique n'est pas uniforme dans le champ, en particulier la qualité décroît lorsque la distance à l'étoile de référence augmente. On assiste en effet à un élargissement de la largeur à mi-hauteur de la réponse impulsionnelle accompagnée d'une élongation en direction de la source d'asservissement de la boucle. Une analyse plus rigoureuse de la FEP devrait donc tenir compte de ces variations : dans le cas idéal on devrait estimer une FEP différente pour chaque galaxie du champ en fonction de la distance de cette dernière à l'étoile guide et de sa position par rapport à celle-ci.

Cresci *et al.* (2005); Cresci *et al.* (2006) ont proposé une approche qui consiste à paramétrer les variations de la FEP dans le champ à l'aide de deux paramètres : l'élongation en direction de l'étoile guide et la largeur à mi-hauteur. L'idée est de chercher une loi permettant de retrouver la forme de la FEP en fonction de la dis-

tance à l'étoile de référence et de la position de la galaxie dans le champ. Pour cela, il est nécessaire de disposer d'un échantillon d'étalonnage, c'est-à-dire des étoiles situées à des positions différentes pour estimer les variations. Il peut s'agir par exemple d'un amas d'étoiles observé dans les mêmes conditions que le champ profond de galaxies.

Malheureusement nous ne disposons pas d'un échantillon de ce type. La paramétrisation de la FEP devient donc particulièrement difficile dans notre cas car très peu d'étoiles suffisamment brillantes pour permettre un étalonnage sont disponibles dans les champs observés. En effet nous disposons au maximum de trois étoiles en plus de l'étoile de référence dans certains champs mais dans d'autres aucune étoile n'est disponible. En notre faveur nous avons le fait que le temps d'exposition est important ( $\sim 3h$ ), par conséquent la qualité de la correction, comme on a pu le voir, n'est pas maximale même à proximité de l'étoile guide. Les variations de la FEP devraient donc être moins importantes que pour des poses plus courtes où la correction est de très haute qualité au centre et la dégradation est donc plus marquée. Nous avons malgré cela réfléchi à des solutions pour pallier le problème de la variation. En particulier nous avons envisagé la possibilité d'estimer la FEP en même temps que les paramètres galactiques. Autrement dit, rajouter des paramètres libres appartenant à la FEP dans le processus d'ajustement ; de cette manière non seulement on crée un modèle de galaxie mais aussi un modèle de FEP. Le problème majeur de cette approche est que l'espace de paramètres est rendu encore plus complexe : en plus des 11 paramètres décrivant la galaxie il y aurait au moins 2 paramètres pour décrire la FEP. La question qui vient à l'esprit est donc : le gain apporté par une meilleure estimation de la FEP est-il suffisamment important pour compenser l'éventuelle dégénérescence introduite par l'augmentation du nombre de paramètres ?

Avant donc de se lancer dans le processus de codage de cette nouvelle procédure nous avons voulu estimer au première ordre le gain apporté. Nous avons pour cela choisi un champ dans lequel on dispose d'une étoile brillante un peu éloignée du centre et nous avons construit un modèle de FEP sur le même principe que celui exposé dans le paragraphe 4.3.3.2 mais en rajoutant deux nouveaux paramètres au modèle pour prendre en compte l'éventuelle asymétrie de la FEP. Ainsi, du modèle à deux paramètres présenté précédemment (rapport de Strehl, *seeing*) on passe à un modèle à quatre paramètres :

- Le rapport de Strehl
- Le *seeing* ou largeur du halo gaussien dans la direction de l'axe x
- Le *seeing* ou largeur du halo gaussien dans la direction de l'axe y
- Un angle de rotation permettant d'orienter la FEP estimée

Nous avons donc réalisé un ajustement avec ce modèle asymétrique sur une étoile du champ STAR1. La figure 4.15 illustre le résultat. On constate que la valeur du rapport de Strehl estimée est cohérente avec celle obtenue avec le modèle symétrique et que l'on obtient comme espéré un léger allongement en direction de l'étoile de référence.

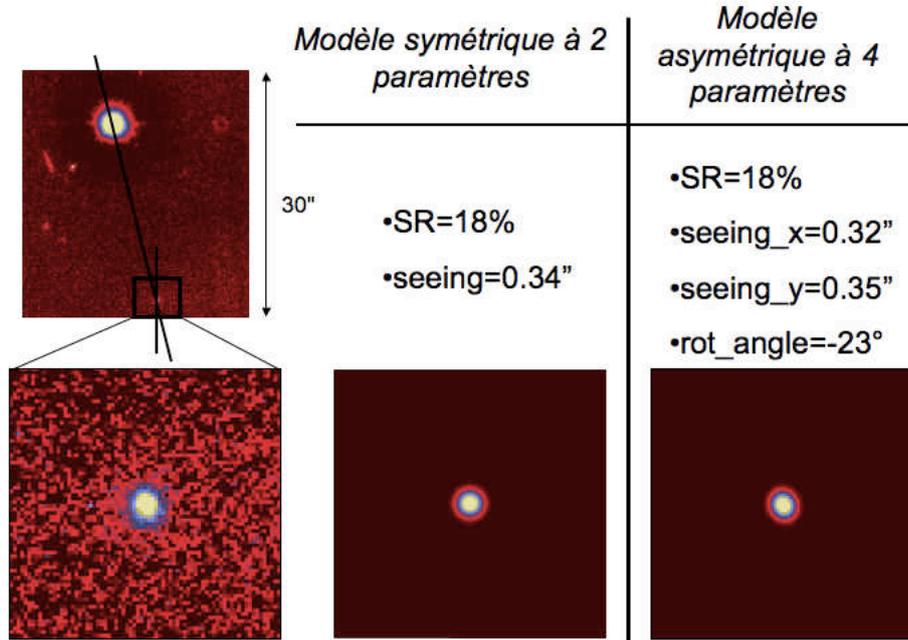


FIG. 4.15 – Résultat de l’ajustement d’un modèle asymétrique de FEP à une étoile réelle.

Il est maintenant question de déterminer si ce changement de FEP implique une forte différence dans l’estimation des paramètres galactiques. Nous avons donc exécuté GIM2D une deuxième fois sur l’ensemble des galaxies de ce champ avec cette FEP asymétrique et comparé les paramètres estimés avec ceux obtenus en employant le modèle symétrique. Les résultats sont illustrés dans la figure 4.16.

Nous constatons que les paramètres obtenus avec les deux FEP sont de façon générale en bon accord, en particulier pour le paramètre B/T qui est celui présentant le plus fort intérêt pour notre étude. Il existe cependant 5 objets pour lesquels la différence estimée est plus importante. Une inspection visuelle de ces galaxies montre qu’il s’agit d’objets faibles, se situant donc dans la zone de moins bon fonctionnement de GIM2D d’après les simulations. Les objets brillants présentent une cohérence dans les paramètres estimés malgré la modification de la FEP.

Au vu de ces résultats nous avons décidé pour cette étude de ne pas continuer dans l’élaboration d’un nouveau logiciel d’ajustement qui estimerait la FEP en même temps que les paramètres de la galaxie. Pour les observations futures on envisage cependant d’observer un amas d’étoiles pour étalonner proprement les variations. Ce thème sera à nouveau abordé dans le chapitre 5 à travers une discussion sur la déconvolution.

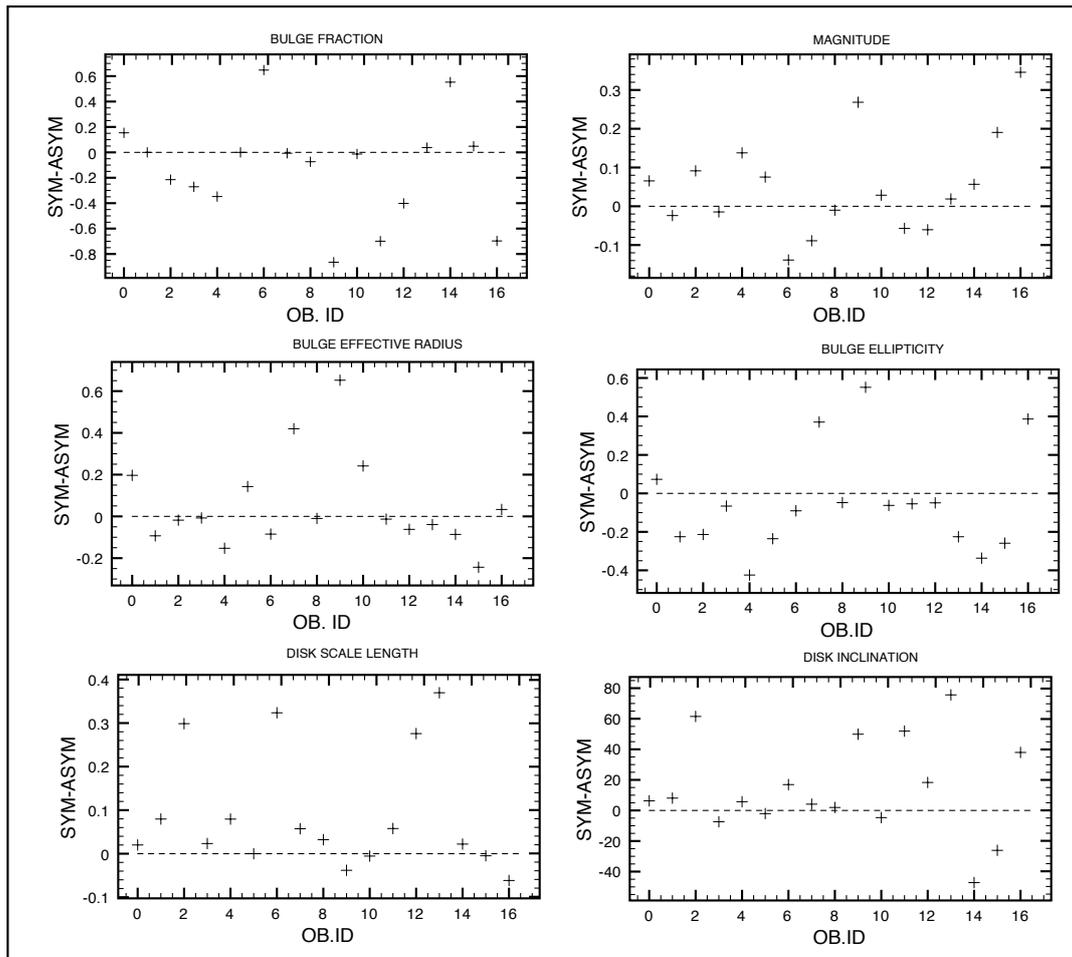


FIG. 4.16 – Différences dans l'estimation de quelques paramètres galactiques lorsque un modèle asymétrique de FEP est employé.

## 4.4 Résultats

Les galaxies ont donc été classées en trois types morphologiques, i.e. galaxies à disque dominant, galaxies à bulbe dominant et galaxies irrégulières.

### 4.4.1 Statistiques globales

Un des résultats principaux est que  $12 \pm 2.7\%$  des galaxies analysées présentent une morphologie de type irrégulier (10 sur 79). Pour le reste de l'échantillon, l'analyse paramétrique donne 12 ( $\sim 26\%$ ) galaxies à bulbe dominant ( $B/T > 0.5$ ) et 48 ( $\sim 60\%$ ) galaxies à disque dominant. D'un autre côté, l'analyse non paramétrique trouve 15 ( $\sim 19\%$ ) galaxies à bulbe dominant et 54 ( $\sim 67\%$ ) galaxies à disque dominant.

Afin de mieux comprendre les différences entre les deux classifications, nous avons réalisé une comparaison galaxie par galaxie du type morphologique obtenu avec GIM2D et avec la méthode *concentration et asymétrie*. La figure 4.17 montre ainsi la probabilité pour qu'une galaxie soit classée de la même façon par les deux méthodes : pour chaque type morphologique de la classification C/A on regarde la fraction de galaxies de chaque type GIM2D.

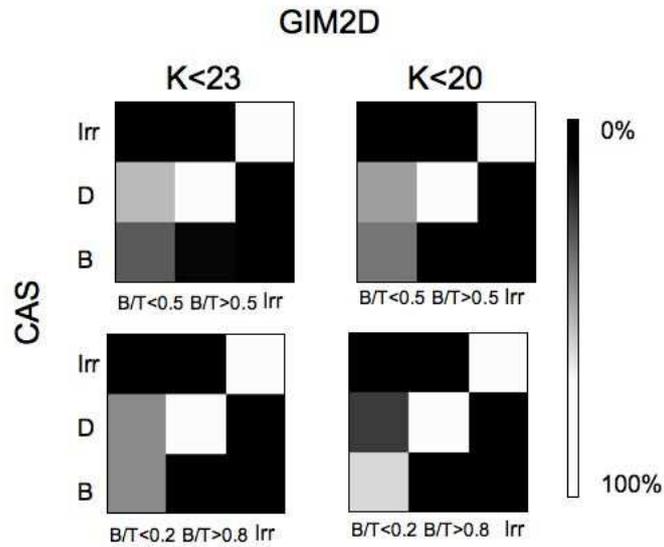


FIG. 4.17 – Comparaison des deux méthodes de classification : probabilité qu'une galaxie soit classée pareil par GIM2D et C/A pour différents intervalles de magnitude (c.f texte pour plus de détails).

La probabilité pour qu'une galaxie disque C/A soit aussi un disque dans la classification paramétrique est de  $p = 0.81$  par contre elle n'est que de  $p = 0.30$  pour les

galaxies bulbes. Pour les galaxies irrégulières la probabilité est évidemment égale à un puisqu'elles sont détectées avec la même méthode.

Il peut y avoir plusieurs raisons à l'origine de ces différences. Premièrement il peut s'agir d'une conséquence du bas rapport S/B. En effet, comme il a été dit dans la section 4.3.3.2, lorsque celui-ci diminue, GIM2D tend à sous estimer le flux provenant du bulbe. Ceci pourrait expliquer qu'une fraction si importante de galaxies identifiées comme bulbes par la méthode C/A soit considérées des disques par GIM2D. Afin de vérifier cette hypothèse, nous avons regardé cette même correspondance pour les galaxies plus brillantes de l'échantillon ( $K_s < 20$ ). La fraction de galaxies de type bulbe classés de même façon par les deux méthodes augmente alors jusqu'à  $p = 0.7$ .

Deuxièmement, les différences peuvent provenir d'une différence dans la définition des types entre les deux méthodes. En particulier, les galaxies avec une fraction B/T proche de 0.5 peuvent causer des différences. Si l'on répète l'analyse en enlevant ces objets (Fig. 4.17), 80% des bulbes C/A sont aussi classés en tant que bulbes par GIM2D et 95% des disques.

La comparaison entre les deux méthodes permet en quelque sorte de quantifier les erreurs dans la classification dans le sens où il paraît logique de penser que la « vraie » valeur doit se trouver quelque part entre les deux estimations. On peut donc estimer que le « mélange » morphologique dans notre échantillon est :  $24\% \pm 4\%$  de galaxies à bulbe dominant,  $64\% \pm 4\%$  de galaxies de type disque, et  $12\% \pm 2.7\%$  de galaxies irrégulières.

Ces résultats constituent la première mesure de la contribution de chaque type morphologique autour de  $z \sim 1$  à partir de données dans l'infrarouge proche à haute résolution angulaire.

#### 4.4.2 Evolution morphologique

Un des intérêts de cette étude réside dans le fait que l'on dispose d'une estimation des *redshifts* photométriques d'un certain nombre d'objets grâce aux données multi-longueur d'onde du relevé COSMOS. Plus précisément 60 galaxies parmi les 79 détectées ont une estimation du *redshift*. L'échantillon est petit et il est par conséquent difficile de tirer de conclusions définitives, nous avons tout de même voulu réaliser une première étude de l'évolution morphologique en fonction du *redshift*. Etant donné le nombre limité d'objets, nous avons limité l'étude à deux intervalles de *redshift* ( $z < 0.8$  et  $z > 0.8$ ) et avons tracé les fractions de chaque type morphologique pour chaque intervalle. Nous avons utilisé ici les résultats de la classification non-paramétrique afin de rester cohérents avec les comparaisons qui seront effectuées dans la section 4.5 avec des données spatiales d'une part et parce qu'elle est plus robuste jusqu'à des magnitudes plus faibles. Les résultats sont montrés dans la figure 4.19.

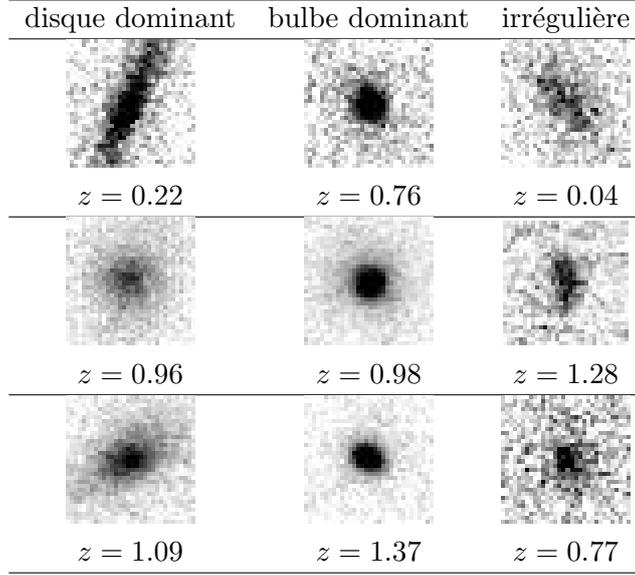


FIG. 4.18 – Exemples de classifications. Les tailles de l’images sont de  $1.7' \times 1.7'$

Nous avons tracé de même, les estimations d’après des mesures de fonctions de luminosités par type spectral réalisées par [Bolzonella et al. \(2008\)](#) dans le cadre du consortium VVDS. Les auteurs ont en effet établi les fonctions de luminosité dans la bande  $K_s$  au repos, jusqu’à  $z \sim 1.5$ , à plusieurs intervalles de *redshift* et pour quatre types spectraux définis par ajustement de spectres. Le premier type correspondant à des spectres de galaxies précoces, les types deux et trois à des galaxies spirales et le quatrième type à des galaxies irrégulières. En intégrant donc ces fonctions de luminosités il est possible d’obtenir une estimation des densités de chaque type en fonction du *redshift* comparable à nos mesures. Les types morphologiques ayant été estimés différemment, il est intéressant d’analyser si les résultats sont compatibles. Nous avons donc intégré les fonctions de Schechter en prenant comme limites d’intégration les limites de complétude approximatives estimées dans le paragraphe 4.2.6 pour chaque intervalle de *redshift*. Nous avons de même représenté comme référence un cas de *non-évolution* depuis  $z \sim 0.2$ , c’est-à-dire, nous avons calculé les fractions de chaque type à partir de la fonction de luminosité mesurée à  $z \sim 0.2$ .

On constate que les fractions estimées sont globalement compatibles, ce qui semblerait indiquer que l’estimation du type morphologique de manière structurelle ou à partir d’un spectre donnent des résultats cohérents. Cependant, nos deux points de mesure sont compatibles de la même manière avec les prévisions du scénario non-évolutif. Il est donc difficile de tirer des conclusions sûres concernant l’évolution des types morphologiques juste à partir de ces mesures. Notons cependant, que la fraction petite de galaxies irrégulières à grand  $z$  est obtenue également à partir de l’intégration des fonctions de luminosité.

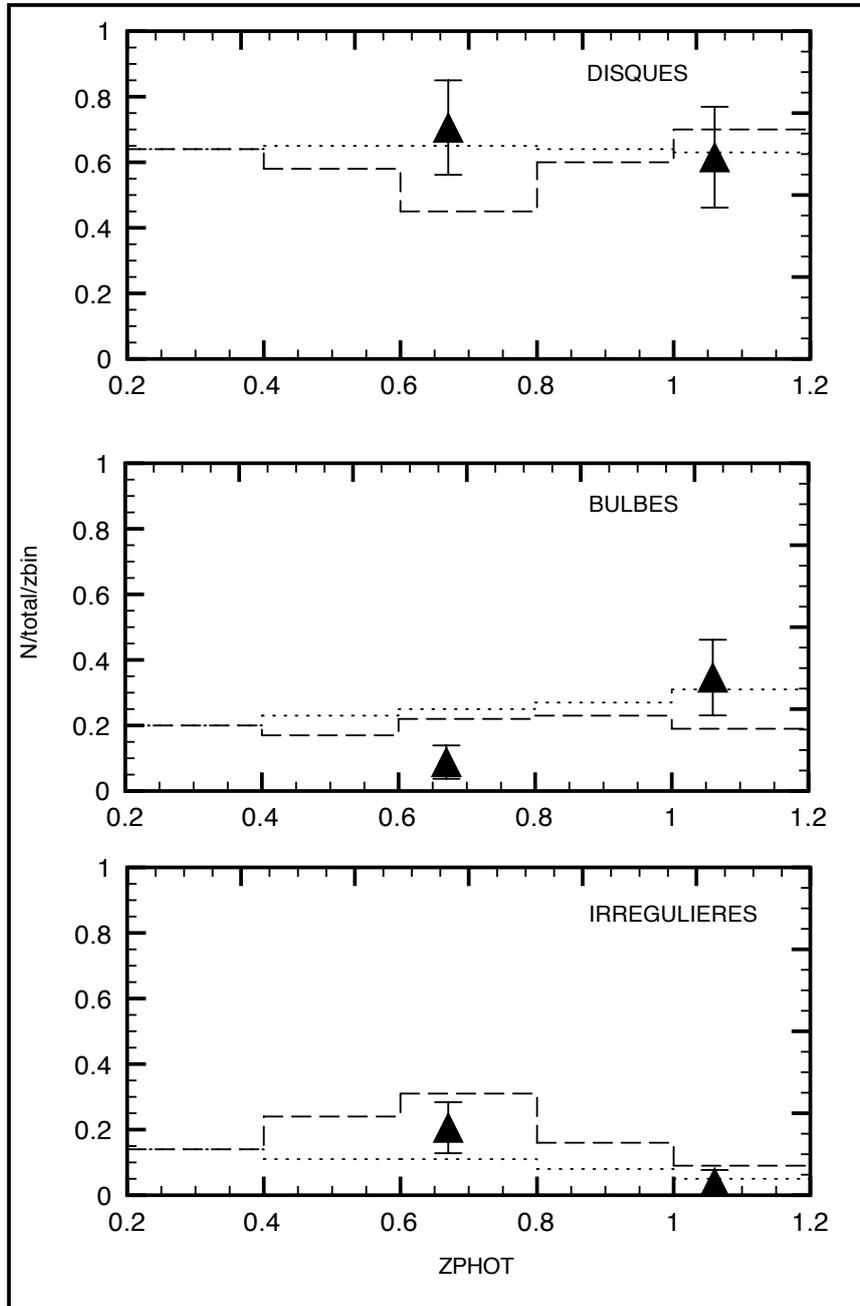


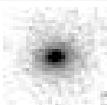
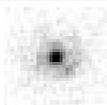
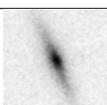
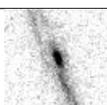
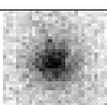
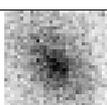
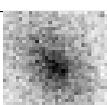
FIG. 4.19 – Distribution en *redshift* des trois types morphologiques d'après la classification non-paramétrique. L'échantillon a été divisé en deux intervalles de *redshifts* ( $z < 0.8$  et  $z > 0.8$ ). Le *redshift* représenté est le *redshift* médian de chaque intervalle. Les histogrammes montrent les estimations d'après les fonctions de luminosité établies par Bolzonella *et al.* (2008) : en pointillés : pas d'évolution depuis  $z \sim 0.2$  ; en trait interrompu : évolution suivant les mesures de Bolzonella *et al.* (2008). Noter que l'échelle des *redshifts* commence à  $z = 0.2$  car c'est à partir de ce *redshift* que l'on dispose des données sur les fonctions de luminosité. Consulter le texte pour plus de détails.

## 4.5 Quantification des effets de k-correction morphologique

### 4.5.1 Comparaison globale

Nous avons également comparé nos résultats avec des données du télescope spatial (HST) en bande i du relevé COSMOS. Soixante objets parmi les 79 analysés ont une contre partie dans les images HST. Cette comparaison a comme but principal l'estimation des effets de changement de bande dans la quantification de la morphologie (k-correction morphologique, c.f. chapitre 2). En effet, lorsque l'on observe une galaxie à  $z \sim 1$  en bande Ks, la bande équivalente au repos est autour de la bande z alors que si elle est observée en bande I, la bande au repos est centrée dans la bande B. Etant donné que les deux échantillons ont une très haute résolution angulaire et que l'on peut connaître la morphologie pour les deux, ce sont d'excellents candidats pour quantifier la k-correction. Nous avons donc classé ces objets en utilisant la méthode de *concentration et asymétrie* (c.f. §. 4.3.3.1). Le plan C/A a été étalonné de la même façon que pour les données NACO. Le résultat de la classification donne finalement :  $32\% \pm 1.6\%$  de galaxies irrégulières,  $47\% \pm 1.5\%$  de galaxies de type disque et  $21\% \pm 2.5\%$  de galaxies de type bulbe. On observe un certain nombre de différences dans les proportions de chaque type morphologique dans la bande I et la bande Ks. En particulier, dans la bande I on trouve plus de galaxies présentant une morphologie irrégulière. Si on regarde par ailleurs chaque objet individuellement (Fig. 4.20) cette tendance est largement confirmée. Il existe de grandes incertitudes de classement entre les irrégulières en bande K et les disques en bande I ainsi qu'entre les disques en bande I et les bulbes en bande K. La moitié des galaxies dominées par un bulbe en bande K se retrouvent dominées par un disque en bande I et plus de la moitié des disques en bande K sont classés comme irrégulières en bande I comme s'il existait une tendance à voir des types plus tardifs lorsqu'on les observe vers des longueurs d'onde plus petites. Il faut tenir compte cependant, avant de tirer des conclusions définitives, du fait que l'analyse est faite sur un petit nombre d'objets et que par conséquent des petites différences se traduisent rapidement sur des grandes différences dans la figure 4.20.

Malgré tout, cette tendance est cohérente avec l'effet attendu : la bande I sonde en effet des populations stellaires plus jeunes et par conséquent les régions de formation stellaire peuvent faire apparaître des inhomogénéités dans la galaxie qui la font apparaître sous une forme plus irrégulière. Une inspection visuelle révèle en effet qu'une partie de ces galaxies classées comme irrégulières dans la bande I sont en fait des galaxies spirales bien résolues mais pas tout à fait symétriques (Tab. 4.5).

NACO		<i>redshift</i>	ACS	
type morph.	image		image	type morph.
Disk dom.		1.090		Irr.
Irr.		0.4223		Irr.
Disk dom.		0.6689		Irr.
Bulge dom.		1.17460		Disk dom.
Disk dom.		0.6689		Irr.
Disk dom.		0.8861		Irr.
Disk dom.		0.7394		Irr.

TAB. 4.5 – *K correction morphologique* : différences morphologiques entre les bandes K (NACO) et I (HST). La taille des images est  $1.7' \times 1.7'$

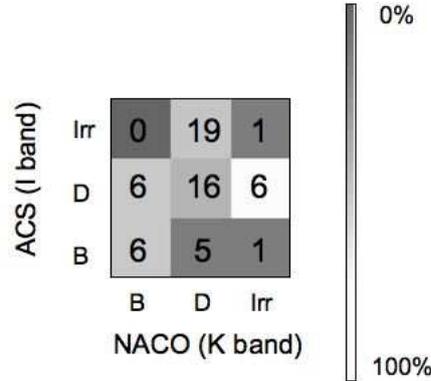


FIG. 4.20 – Comparaison entre la bande K et la bande I : la figure montre la probabilité pour qu’une galaxie soit classée de façon identique dans les deux bandes. Les chiffres indiquent le nombre de galaxies.

#### 4.5.2 Evolution morphologique

Ces différences sont-elles une manifestation des effets de k-correction ou s’agit-il d’erreurs de classification des méthodes automatiques ? Afin de mieux comprendre cet effet, nous avons essayé de corriger des effets de changement de bande par une méthode statistique. Si on arrivait en effet à « ramener » les morphologies des deux échantillons dans un filtre commun (*k-corriger*), les deux résultats devraient être identiques si les différences proviennent uniquement des effets de k-correction. Brinchmann *et al.* (1998) ont proposé une méthode qui consiste à simuler des galaxies en utilisant des spectres réalistes, les *redshifter* et regarder le nombre d’objets qui changent de classe morphologique. On peut alors définir un coefficient donnant la fraction d’objets qui passent d’une classe X à une classe Y :

$$D_{XY} = \frac{N_{X \rightarrow Y}}{N_X} \quad (4.3)$$

Il existe alors une relation entre le nombre d’objets observés dans une classe donnée et le nombre d’objets appartenant réellement à cette classe :

$$N_X^{obs} = N_X + \sum N_Y D_{YX} - N_X \sum D_{XY} \quad (4.4)$$

Brinchmann *et al.* (1998) ont obtenu les coefficients permettant de passer de la bande I observée à la bande R au repos pour des données HST. Etant donné que le filtre utilisé pour leurs observations (F814W) est le même que le notre, on peut employer leurs coefficients pour obtenir les morphologies R au repos de nos données spatiales pour pouvoir les comparer aux morphologies obtenues en bande Ks. La bande R n’est pas tout à fait la bande au repos à  $z \sim 1$  correspondant aux observations K mais devrait s’en rapprocher en première approximation. Ainsi, une

fois le facteur de correction appliqué, il est possible de comparer les distributions morphologiques corrigées avec celles observées avec NACO non corrigées.

Nous avons donc tracé à nouveau l'évolution morphologique sur deux intervalles de *redshift* ( $z < 0.8$  et  $z > 0.8$ ) en incluant cette fois la classification en bande I « k-corrigée ». La figure 4.21 montre les résultats :

- A bas *redshift* ( $z < 0.8$ ) l'accord entre les distributions en bande I corrigées et en bande K non corrigées est raisonnable. Cela semble indiquer qu'à ces *redshifts*, les effets de k-correction ne sont pas très importants.
- A plus grand *redshift* la distribution en bande K estime une fraction plus importante de galaxies de type elliptique et nettement moins de galaxies irrégulières. La fraction de galaxies irrégulières obtenues avec la bande Ks semble être plus en accord avec les mesures des fonctions de luminosités quoique l'incertitude est grande. Cela peut être dû au fait que les coefficients calculés par Brinchmann *et al.* (1998) sont plus susceptibles de contenir des erreurs à grand  $z$ . En effet, étant donné que le calcul est basé sur un ajustement de spectres il est assez probable que les incertitudes deviennent plus importantes. Il se peut aussi que ce soit une conséquence de la taille réduite de l'échantillon : des erreurs sur une petite fraction peuvent en effet induire des différences importantes sur les distributions globales. La résolution de ce point nécessiterait d'une augmentation significative de l'échantillon.

## 4.6 Conclusions de l'étude

J'ai présenté l'analyse morphologique de 79 galaxies dans l'infrarouge proche observées avec optique adaptative avec une résolution angulaire moyenne de  $0.1''$ . Grâce à des simulations intensives, nous avons pu montrer que malgré les difficultés techniques, en particulier pour l'estimation de la FEP, l'optique adaptative permet d'obtenir des morphologies de galaxies à  $z \sim 1$  d'une manière automatique, avec un bon degré de confiance, bien adaptée donc à des grands relevés de galaxies :

- Les paramètres galactiques peuvent être estimés par ajustement d'un modèle analytique avec des erreurs plus petites que 20% jusqu'à une magnitude de  $K_s = 19$ . Cela est comparable aux résultats obtenus sur des données spatiales.
- Une morphologie plus qualitative à l'aide de méthodes non paramétriques classiques peut être obtenue jusqu'à  $K_s = 21$  avec au moins 70% d'identifications correctes.

Cela a permis donc d'obtenir, pour la première fois, une estimation de la distribution morphologique entre  $z = 0.6$  et  $z = 1.2$  à partir de données dans l'infrarouge proche à haute résolution angulaire. Le mélange morphologique global est :  $64\% \pm 4\%$  de galaxies tardives,  $24\% \pm 4\%$  de galaxies précoces et  $12\% \pm 2.7\%$  de galaxies irrégulières. La fraction élevée de galaxies irrégulières ( $12\% \pm 2.7\%$ ) bien supérieure à celle

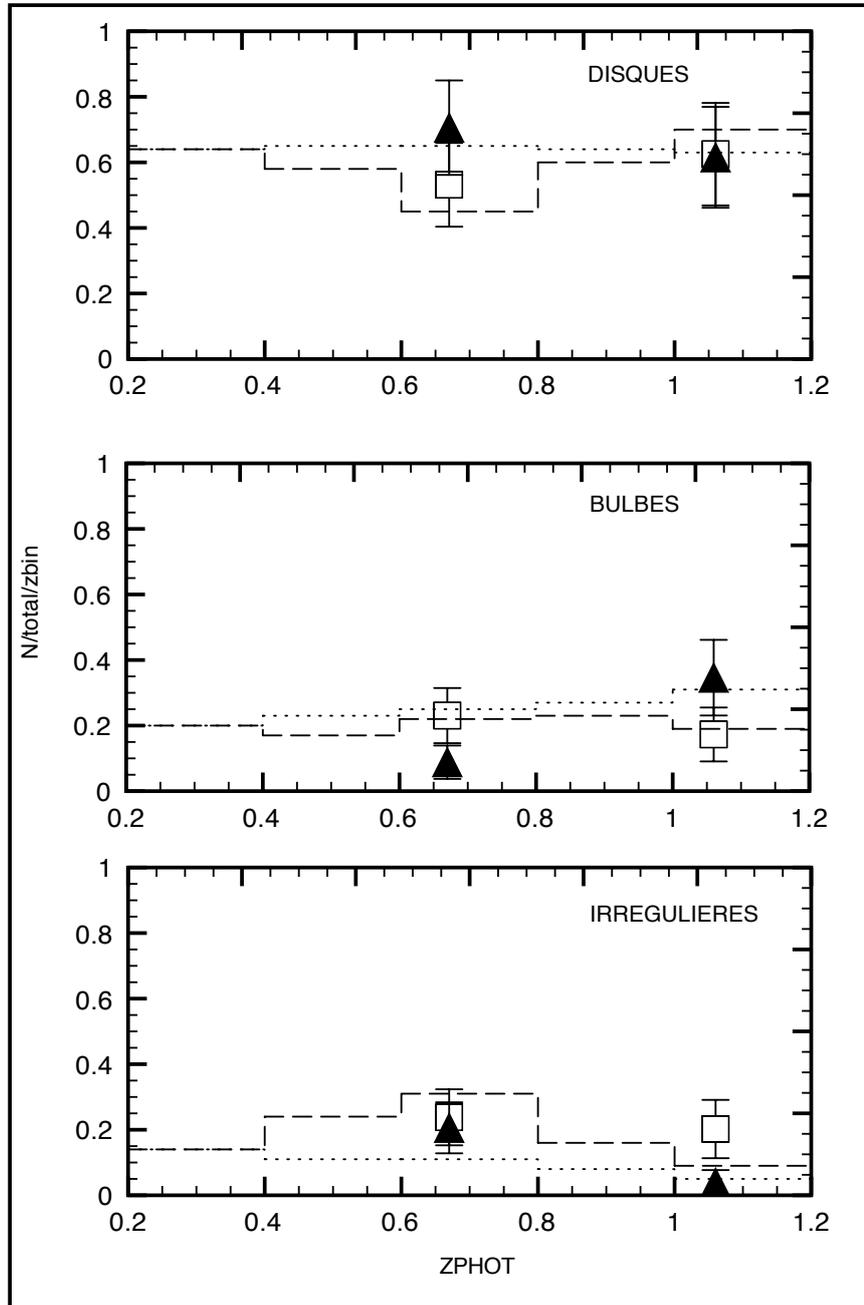


FIG. 4.21 – Distribution en *redshift* des trois types morphologiques. On représente également les données notre échantillon observé avec ACS (carrés). Les données ACS sont corrigées pour être vues dans la bande R au repos. Les données NACO sont dans la bande Ks sans correction. Le *redshift* représenté est le *redshift* médian de chaque intervalle.

que l'on trouve dans l'Univers local confirme que le nombre de galaxies irrégulières augmente significativement avec le *redshift*. Nous avons abordé l'étude de l'évolution morphologique en comparant aux mesures à partir des fonctions de luminosité. Nos mesures sont globalement compatibles avec celles obtenues à partir des fonctions de luminosité par type spectral, cependant, l'échantillon est encore trop petit pour pouvoir tirer des conclusions robustes sur l'évolution des populations de galaxies elliptiques et spirales.

Un point crucial de cette étude résidait dans la quantification des effets de k-correction. Nous avons pour cela comparé les résultats des classifications aux classifications obtenues sur les mêmes objets dans le visible (HST). Encore une fois, le nombre limité d'objets nous empêche de tirer des conclusions définitives. Il semblerait que les distributions visibles et infrarouges sont cohérentes pour l'échantillon à  $z < 0.8$ . A plus grand  $z$ , les différences persistent même après application des coefficients de correction, en particulier pour la population de galaxies irrégulières, moins représentée dans la classification infrarouge. S'agit-il d'erreurs de classification ou d'erreurs dans les facteurs de correction appliqués qui reflètent donc des effets de k-correction ?

La réponse à ces questions ouvertes passe donc nécessairement par l'observation d'un échantillon statistiquement plus représentatif. Nous avons ainsi proposé à l'ESO un programme d'observations plus systématique du champ COSMOS. Les caractéristiques principales de ce programme sont présentées dans le chapitre 6.

## Bibliographie

- Abraham R.G., Valdes F., Yee H.K.C., & van den Bergh S., 1994. The morphologies of distant galaxies. 1 : an automated classification system. *Astrophysical Journal*, 432 : 75. 90
- Abraham R.G., van den Bergh S., Glazebrook K., Ellis R.S., Santiago B.X., Surma P., & Griffiths R.E., 1996. The Morphologies of Distant Galaxies. II. Classifications from the Hubble Space Telescope Medium Deep Survey. *Astrophysical Journal Supplement* v.107, 107 : 1. 90
- Baker A.J., Davies R.I., Lehnert M.D., Thatte N.A., Vacca W.D., Hainaut O.R., Jarvis M.J., Miley G.K., & Röttgering H.J.A., 2003. Galaxies in southern bright star fields. I. Near-infrared imaging. *A&A*, 406 : 593. 87
- Bershady M.A., Lowenthal J.D., & Koo D.C., 1998. Near-Infrared Galaxy Counts to J and K  $\sim 24$  as a Function of Image Size. *ApJ*, 505 : 50. 87
- Bertin E. & Arnouts S., 1996. SExtractor : Software for source extraction. *A&A supplement*, 117 : 393. 84
- Bolzonella M., Miralles J.M., & Pelló R., 2000. Photometric redshifts based on standard SED fitting procedures. *A&A*, 363 : 476. 89

- Bolzonella N., Zucca E., & VVDS consortia, 2008. in preparation. 107, 108
- Brinchmann J., Abraham R., Schade D., Tresse L., Ellis R.S., Lilly S., Le Fevre O., Glazebrook K., Hammer F., Colless M., Crampton D., & Broadhurst T., 1998. Hubble Space Telescope Imaging of the CFRS and LDSS Redshift Surveys. I. Morphological Properties. *ApJ*, 499 : 112. 111, 112
- Cresci G., Davies R., Baker F., Mannucci M., Lehnert T., Totani T., & Minowa Y., 2006. Galaxy morphology and evolution from the SWAN Adaptive Optics imaging. astro-ph/0607221. 87, 101
- Cresci G., Davies R.I., Baker A.J., & Lehnert M.D., 2005. Accounting for the anisoplanatic point spread function in deep wide-field adaptive optics images. *A&A*, 438 : 757. 101
- de Vaucouleurs G., 1959. Classification and Morphology of External Galaxies. *Handbuch der Physik*, 53 : 275. 91
- Freeman K.C., 1970. On the Disks of Spiral and so Galaxies. *ApJ*, 160 : 811. 91
- Huertas-Company M., Rouan D., Soucail G., Le Fèvre O., Tasca L., & Contini T., 2007. Morphological evolution of  $z \sim 1$  galaxies from deep K-band AO imaging in the COSMOS deep field. *A&A*, 468 : 937. 78, 79
- Ilbert O., Arnouts S., McCracken H.J., Bolzonella M., Bertin E., Le Fevre O., Mellier Y., Zamorani G., Pello R., Iovino A., Tresse L., Bottini D., Garilli B., Le Brun V., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Bardelli S., Cappi A., Charlot S., Ciliegi P., Contini T., Cucciati O., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Marano B., Marinoni C., Mazure A., Meneux B., Merighi R., Paltani S., Pollo A., Pozzetti L., Radovich M., Zucca E., Bondi M., Bongiorno A., Busarello G., De La Torre S., Gregorini L., Lamareille F., Mathez G., Merluzzi P., Ripepi V., Rizzo D., & Vergani D., 2006. Accurate photometric redshifts for the CFHT Legacy Survey calibrated using the VIMOS VLT Deep Survey. *ArXiv Astrophysics e-prints*. 88, 89
- Kleinmann S.G., 1992. 2MASS - The 2 Micron All Sky Survey. Dans *ASP Conf. Ser. 34 : Robotic Telescopes in the 1990s*, édité par A.V. Filippenko, p. 203–212. 82
- Lenzen R., Hofmann R., Bizenberger P., & Tusche A., 1998. CONICA : the high-resolution near-infrared camera for the ESO VLT. Dans *Proc. SPIE Vol. 3354, p. 606-614, Infrared Astronomical Instrumentation, Albert M. Fowler ; Ed.*, édité par A.M. Fowler, volume 3354 de *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, p. 606–614. 79
- Marleau F.R. & Simard L., 1998. Quantitative Morphology of Galaxies in the Hubble Deep Field. *The Astrophysical Journal*, 507 : 585. (c) 1998 : The American Astronomical Society. 90, 96

- McCracken H.J., Metcalfe N., Shanks T., Campos A., Gardner J.P., & Fong R., 2000. Galaxy number counts - IV. Surveying the Herschel Deep Field in the near-infrared. *MNRAS*, 311 : 707. 87
- McLeod B.A., Bernstein G.M., Rieke M.J., Tollestrup E.V., & Fazio G.G., 1995. K-band galaxy counts. *ApJs*, 96 : 117. 87
- Minowa Y., Kobayashi N., Yoshii Y., Totani T., Maihara T., Iwamuro F., Takami H., Takato N., Hayano Y., Terada H., Oya S., Iye M., & Tokunaga A.T., 2005. Subaru Super Deep Field with Adaptive Optics. I. Observations and First Implications. *ApJ*, 629 : 29. 87
- Rousset G., Lacombe F., Puget P., Gendron E., Arsenault R., Kern P.Y., Rabaud D., Madec P.Y., Hubin N.N., Zins G., Stadler E., Charton J., Gigan P., & Feautrier P., 2000. Status of the VLT Nasmyth adaptive optics system (NAOS). Dans *Proc. SPIE Vol. 4007, p. 72-81, Adaptive Optical Systems Technology, Peter L. Wizinowich; Ed.,* édité par P.L. Wizinowich, volume 4007 de *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, p. 72–81. 79
- Simard L., Willmer C.N.A., Vogt N.P., Sarajedini V.L., Phillips A.C., Weiner B.J., Koo D.C., Im M., Illingworth G.D., & Faber S.M., 2002. The DEEP Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip. *ApJs*, 142 : 1. 87, 90, 96, 98, 99

# Déconvolution

---

## Sommaire

---

<b>5.1 Principe</b> . . . . .	<b>118</b>
<b>5.2 Méthodes de déconvolution</b> . . . . .	<b>118</b>
5.2.1 Maximum de vraisemblance . . . . .	119
5.2.2 Maximum à posteriori . . . . .	119
5.2.3 Déconvolution myope . . . . .	121
<b>5.3 Application au champ NACO</b> . . . . .	<b>121</b>
<b>5.4 Sommaire et conclusions</b> . . . . .	<b>123</b>
<b>Bibliographie</b> . . . . .	<b>124</b>

---



---

**Abstract :** Adaptive optics enables a better estimation of the morphological parameters of distant galaxies compared to seeing limited observations as shown in the previous chapter. In this chapter we explore the possibility of increasing the image quality by means of deconvolution. I describe the basic principles of deconvolution followed by some preliminary results on the NACO fields.

---

**Resumen :** Hemos tratado de demostrar en el capítulo anterior, que el uso de la óptica adaptativa permite una mejora significativa de la calidad de la estimación de los parámetros morfológicos de las galaxias distantes. En el siguiente capítulo se investiga la posibilidad del empleo de técnicas de deconvolución para tratar de mejorar la calidad de la imagen. Se describen en un primer lugar los principios básicos de la deconvolución y se muestran algunos resultados preliminares sobre las galaxias observadas.

---

L'optique adaptative permet d'améliorer la qualité de l'estimation de la morphologie comme il a été présenté dans le chapitre précédent. La résolution angulaire obtenue est, en effet, nettement supérieure à celle obtenue sans correction et la qualité de l'estimation des paramètres galactiques est comparable, dans certains domaines de brillance, à celle obtenue sur des données spatiales. Dans ce chapitre on explore la technique de déconvolution pour essayer d'améliorer un peu plus la qualité de l'image. Je décris donc dans un premier temps les principes de base de la déconvolution puis les résultats obtenus pour nos images de galaxies lointaines.

## 5.1 Principe

Rappelons l'expression de l'image au foyer d'un télescope :

$$I(\mathbf{r}) = (O * H)(\mathbf{r}) + N(\mathbf{r}) \quad (5.1)$$

où  $O$  représente l'objet,  $H$  la FEP de l'instrument et  $N$  le bruit de mesure. Comme nous l'avons vu dans les chapitres précédents, en présence d'atmosphère les fréquences plus grandes que  $\lambda/r_0$  se trouvent fortement atténuées. L'optique adaptative permet de récupérer une partie de l'information en se rapprochant de la FEP théorique du télescope qui a une fréquence de coupure à  $\lambda/D$ . La déconvolution quant à elle, essaye de faire un pas de plus en accédant à l'objet avant convolution à travers l'inversion de l'équation 5.1. Ce processus n'a pas pour but de récupérer des informations au delà de la fréquence de coupure du télescope puisque celles-ci sont complètement absentes de l'image mais plutôt de rétablir l'information de manière uniforme dans tout le domaine spectral jusqu'à la fréquence de coupure, comme si la FTO était une fonction porte uniforme. En effet, même avec une correction parfaite, la FTO du télescope n'est pas distribuée uniformément dans tout le domaine spectral, les hautes fréquences étant plus atténuées que les basses fréquences. La déconvolution redistribue l'information dans toutes les fréquences permettant ainsi une augmentation du contraste de l'image.

Avant de poursuivre avec l'exposé, je réponds à une question que le lecteur pourra se poser à ce stade : s'il est possible de remonter à l'objet en inversant l'équation 5.1 pourquoi utiliser alors l'optique adaptative ? En fait, comme je le montrerai par la suite, le problème d'inversion est complexe, en particulier à cause de la présence du bruit inconnu. Il appartient à la famille des problèmes « mal posés » qui n'a pas de solution unique ou qui demande des hypothèses supplémentaires. La restitution des fréquences intermédiaires est d'autant plus facile que la perte d'information est petite.

La déconvolution se présente donc ici comme un complément de l'optique adaptative : la correction apportée par la deuxième permet d'améliorer les performances de la première. Nous nous interrogeons donc sur le fait de si l'application de la déconvolution à nos images de galaxies lointaines présentées dans le chapitre précédent peut apporter un gain dans la détermination de la morphologie.

## 5.2 Méthodes de déconvolution

L'objectif principal est donc d'inverser l'équation 5.1. Dans l'espace de Fourier, la convolution devient une multiplication. Le problème d'inversion semble donc simple : il suffirait de diviser la transformée de Fourier de l'image par celle de la

FEP et en revenant dans l'espace réel, on obtiendrait la FEP de l'objet. Cela produit malheureusement une image extrêmement bruitée. En effet, étant donné que la FEP prend des valeurs très petites aux hautes fréquences, la division implique une très forte amplification du bruit.

### 5.2.1 Maximum de vraisemblance

Il est donc indispensable de prendre en compte la statistique du bruit dans le processus de minimisation. Les méthodes classiques employées en astronomie sont de type probabiliste, c'est à dire qu'on essaye de trouver l'objet qui maximise la probabilité d'avoir observé cette image ou ce qui revient au même on maximise la probabilité de l'image connaissant l'objet (*maximum de vraisemblance*) :

$$O_{MV} = \arg \max_O Pr(I/O) \quad (5.2)$$

ce qui est équivalent à minimiser l'anti-logarithme de cette probabilité :

$$E_{MV} = -\ln[Pr(I/O)] \quad (5.3)$$

Pour résoudre le problème, il faut donc supposer une statistique de bruit. Pour un bruit gaussien par exemple, la solution revient à une simple inversion. Dans le cas d'un bruit de Poisson, une solution fut proposée par [Lucy \(1974\)](#). Cette solution a cependant tendance également à amplifier le bruit. Afin de réduire ces effets, la minimisation est arrêtée avant convergence par un critère plus ou moins arbitraire.

### 5.2.2 Maximum à posteriori

Une autre façon de pallier le problème est une approche bayésienne proposée par [Ridolfi & Idier \(2001\)](#). Dans ce cadre, on essaye de maximiser la probabilité de l'objet connaissant l'image observée (*maximum à posteriori*, MAP) :

$$O_{MAP} = \arg \max_O Pr(O/I) \quad (5.4)$$

La formule de Bayes permet d'estimer cette probabilité en introduisant un à priori sur l'objet :

$$Pr(O/I) = \frac{Pr(I/O) \times Pr(O)}{Pr(I)} \quad (5.5)$$

Etant donné que la probabilité de l'image vis-à-vis de l'objet est constante, le critère à minimiser devient :

$$E_{MAP} = E_{MV} + E_{prior} = -\ln[Pr(I/O) - \ln[Pr(O)]] \quad (5.6)$$

Le terme  $E_{prior}$  est appelé terme de régularisation et donne en fait une ou plusieurs contraintes à l'objet, empêchant ainsi l'amplification du bruit aux hautes fréquences. Il existe plusieurs types de régularisations que l'on peut appliquer. La régularisation quadratique considère que l'objet est gaussien est que la différence entre ce dernier et sa moyenne est stationnaire :

$$E_{L2} = \frac{1}{2} \sum_f \frac{\|TF(O) - TF(O_m)\|}{S_O} \quad (5.7)$$

où  $S_O$  est la densité spectrale de l'objet. Celle-ci peut être modélisée avec peu de paramètres (Conan *et al.*, 1998) :

$$S_O = \frac{K}{1 + (f/f_0)^p} \quad (5.8)$$

où  $f_0$  est inversement proportionnelle à l'étendue spatiale de l'objet et  $p$  caractérise la douceur de l'objet. Finalement on ajoute souvent un paramètre ( $\mu$ ) permettant d'ajuster le poids accordé à la régularisation :

$$E_{MAP} = E_{MV} + \mu \times E_{L2} \quad (5.9)$$

Ainsi, si  $\mu$  est trop grand l'a priori sur l'objet domine et on reconstruit un objet trop lisse et au contraire si  $\mu$  est trop petit, on retrouve le cas de l'estimateur par maximum de vraisemblance qui peut fournir un objet trop bruité. Avec une telle modélisation les sauts brusques qui pourraient se trouver dans l'objet ne sont pas bien reproduits cependant.

Pour faire face à ce type d'objets ayant de forts sauts en densité, on introduit un autre type de régularisation. Un objet peut être en effet doux localement mais avoir des sursauts d'intensité aux bords ou à l'inverse. La méthode de régularisation L1-L2 (Green, 1990; Mugnier *et al.*, 2001, 2004) permet justement de pénaliser les gradients faibles d'intensité entre les pixels (dus au bruit probablement) tout en restaurant les forts gradients :

$$E_{L1L2} = \mu_x \left[ \frac{\Delta O}{\delta} - \ln \left( 1 + \frac{\Delta O}{\delta} \right) \right] \quad (5.10)$$

où  $\Delta O$  est le gradient de l'objet par rapport à la dimension spatiale.

Dans cette expression, si  $\Delta O$  est petit (variations faibles dans l'image), le second terme domine et pénalise fortement les gradients d'intensité (après un développement à l'ordre 2, il devient un critère quadratique). L'objet tend donc à devenir lisse. Si au contraire le gradient est grand, l'expression se rapproche d'une forme linéaire permettant de restaurer les forts contrastes.

La difficulté majeure de ce type de méthode est l'ajustement des deux paramètres  $\delta$  et  $\mu$ . Le paramètre  $\delta$  permet justement d'ajuster le passage entre les 2 régimes (linéaire et quadratique) et  $\mu$  mesure le poids accordé à la régularisation comme pour

la régularisation quadratique simple. Un mauvais ajustement de ces paramètres peut donc mener à un objet trop lissé (si  $\mu$  est trop grand par exemple) ou trop bruité (si  $\delta$  est trop petit).

Une contrainte additionnelle, appelée *contrainte de positivité* peut être rajoutée. En effet, l'objet est caractérisé tout simplement par une distribution d'intensité; toutes les valeurs doivent donc être positives. L'introduction de cette information complémentaire facilite en effet la minimisation.

### 5.2.3 Déconvolution myope

En pratique, la FEP des observations n'est pas connue avec exactitude (c'est spécialement le cas pour nos observations). Une erreur dans l'estimation de celle-ci peut malheureusement induire des erreurs importantes dans l'estimation de l'objet. Une solution passe par estimer conjointement la galaxie et les paramètres de la FEP. C'est le principe de la *déconvolution aveugle* proposé par un certain nombre d'auteurs (Ayers & Dainty, 1988; Lane, 1996; Thiebaut & Conan, 1995). L'inconvénient de cette approche est l'augmentation non négligeable du nombre d'inconnues.

Une approche intermédiaire a été proposée par (e.g. Conan *et al.*, 1998) qui consiste à rajouter un troisième terme de régularisation portant sur la FEP dans l'approche Bayésienne :

$$Pr(O/H, I) = \frac{Pr(I/O, H) \times Pr(O) \times Pr(H)}{Pr(I)} \quad (5.11)$$

où H désigne la FEP. Cette façon d'aborder le problème est appelée *déconvolution myope*.

## 5.3 Application au champ NACO

Nous avons réalisé un travail préliminaire afin de déterminer les gains éventuels d'une déconvolution sur les images NACO. Tout d'abord insistons sur le fait que la déconvolution sur nos images est un travail particulièrement difficile à cause de plusieurs aspects :

- D'une part, il s'agit d'objets petits avec un rapport S/B faible. L'identification des bords de l'objet n'est donc pas évidente.
- De plus, la déconvolution requiert une bonne estimation de la FEP pour inverser proprement l'équation 5.1. Ce point est spécialement délicat dans notre cas, comme il a été expliqué dans le chapitre précédent, où la FEP est variable dans le champ.
- Finalement, nos données sont sous-échantillonnées de presque un facteur deux (c.f. chapitre 4). Des effets *d'aliasing* peuvent donc poser des problèmes importants lors de l'inversion de l'équation 5.1

Nous avons voulu tout de même réaliser une étude préliminaire afin de quantifier ces difficultés plus proprement. Nous avons pour cela employé le logiciel MAAD

disponible en ligne<sup>1</sup> développé par Eric Thiébaud.

En utilisant comme estimation de la FEP celle reconstruite par la méthode exposée dans le chapitre 4, nous avons réalisé une déconvolution d'un des champs observés (STAR1). Celle-ci a été effectuée par la méthode de maximum à posteriori avec une régularisation L1-L2 et une contrainte de positivité.

Les résultats pour trois objets sont montrés dans la figure 5.1. J'aimerais insister à ce stade sur le fait qu'il ne s'agit ici que de résultats *au premier ordre*. Les images montrent effectivement un léger gain en résolution, les contours apparaissant plus contrastés de manière générale. Dans la première galaxie on devine même un structure de type bras spiral qui n'apparaît pas dans l'image avant déconvolution.

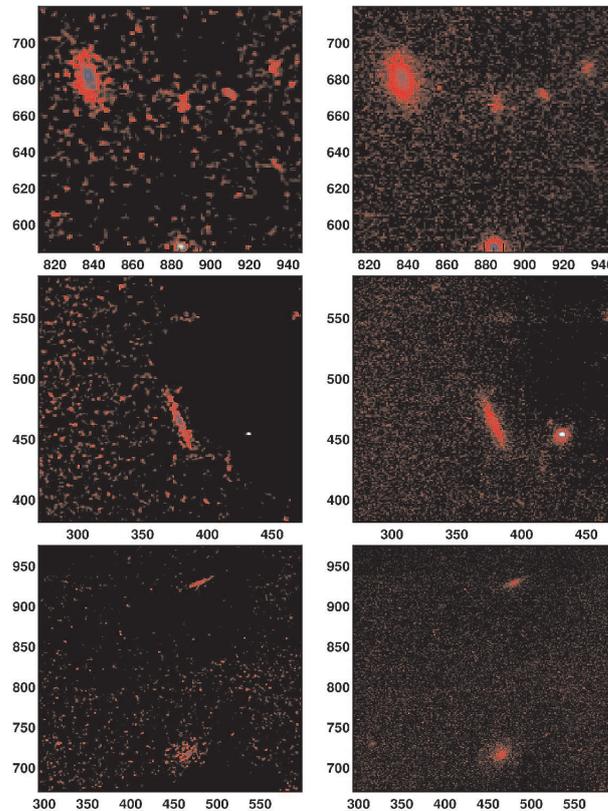


FIG. 5.1 – Exemple de déconvolution par maximum à postérieur de trois objets du champ. Colonne de droite : Galaxie avant déconvolution. Colonne de gauche : Galaxie après déconvolution.

Il peut s'agir cependant *d'artefacts de déconvolution*, c'est-à-dire de structures qui sont créés artificiellement à cause des à priori introduits sur l'objet. Afin d'enlever cette ambiguïté, une étude plus poussée est requise. En particulier une estimation plus fine de la FEP serait nécessaire. Ce point est extrêmement compliqué pour

<sup>1</sup><http://isis.iap.fr:8080/MAAD/>

nos observations où nous ne disposons que de très peu d'étoiles dans le champ (c.f. chapitre 4).

Nous avons malgré tout envisagé une telle possibilité, en deux temps :

1. Paramétrer les variations de la FEP par une loi simple à partir d'un ajustement d'un modèle asymétrique comme celui présenté dans le chapitre précédent. L'idée est relativement simple : il suffit d'ajuster ce modèle asymétrique sur l'ensemble des étoiles du champ afin d'essayer de trouver une loi simple qui donne l'élargissement de la largeur à mi hauteur en fonction de la distance. L'application de cette loi par la suite, permet donc d'obtenir une estimation de la FEP dans tout le champ. Pour les champs où il n'y aurait pas d'étoiles on appliquerait la loi obtenue à partir des autres champs quitte à l'ajuster finement dans la deuxième étape.
2. Réaliser une déconvolution myope, c'est-à-dire estimer simultanément la FEP et l'objet par une approche Bayésienne avec un critère de régularisation.

Après réflexion et suite à un certain nombre de discussions avec des spécialistes de la déconvolution, nous avons décidé cependant d'arrêter le travail à ce stade. En effet la raison principale est le rapport entre le gain que la déconvolution pourra apporter au travail morphologique et l'effort que celle-ci requiert pour ce type d'observations en particulier. Tout d'abord interrogeons nous sur le but scientifique de ces données. Comme il a été exposé auparavant, il s'agit d'effectuer une étude statistique sur la morphologie. Ce qui nous intéresse est finalement une vision globale et non pas un objet en particulier. La déconvolution peut en effet permettre de faire apparaître certains détails (comme les bras spiraux dans l'exemple ci-dessus) et peut être très utile pour une étude détaillée d'une galaxie unique. Notre objectif est cependant de réaliser une étude automatique et statistique. Dans l'idéal on voudrait refaire toute l'analyse morphologique présentée dans le chapitre 4 sur les images déconvoluées. Le risque d'introduire des artefacts de déconvolution (vue la difficulté du problème) qui peuvent biaiser l'estimation morphologique est important. De plus, étant donné le bas rapport S/B des objets, le gain attendu de la déconvolution (d'après l'avis des spécialistes consultés) ne devrait pas être très important.

## 5.4 Sommaire et conclusions

Nous avons exploré dans ce chapitre la possibilité de déconvoluer les images de galaxies lointaines observées avec optique adaptative. Le but étant de gagner en contraste. La déconvolution est particulièrement difficile ici puisque les objets sont faibles, la FEP est mal connue et les données sont sous-échantillonnées. Le problème évoqué récemment à la réunion scientifique SPIE de Marseille (2008) a été reconnu par tous les participants comme très difficile et demandant un gros effort à la fois théorique et sur l'obtention de données secondaires facilitant la reconstruction de la FEP (e.g. profils de  $Cn^2$ , covariance de l'analyseur de surface d'onde).

Nous avons malgré tout réalisé une déconvolution sur l'un des champs par la méthode de maximum à posteriori avec une régularisation L1-L2 afin d'obtenir un premier retour. Nous obtenons un léger gain sur quelques objets après une inspection visuelle. Artefacts? Gain réel? Un travail plus poussé requiert une estimation propre de la FEP et une quantification des effets de sous-échantillonnage. Nous avons réfléchi à plusieurs façons d'aborder le problème. Cependant, finalement nous avons décidé d'arrêter le travail à ce stade à cause de plusieurs raisons :

- On ne s'attend pas à avoir un gain très important étant donné la difficulté du problème ;
- Il existe un risque d'introduire des artefacts susceptibles de nuire l'estimation morphologique.

## Bibliographie

- Ayers G.R. & Dainty J.C., 1988. Iterative blind deconvolution method and its applications. *Opt. Lett.*, 13(7) : 547. [121](#)
- Conan J.M., Mugnier L.M., Fusco T., Michau V., & Rousset G., 1998. Myopic deconvolution of adaptive optics images by use of object and point-spread function power spectra. *Appl. Opt.*, 37(21) : 4614. [120](#), [121](#)
- Green P., 1990. *IEEE Trans. Med. Imag.*, 9 : 84. [120](#)
- Lane R.G., 1996. Methods for maximum-likelihood deconvolution. *J. Opt. Soc. Am. A*, 13(10) : 1992. [121](#)
- Lucy L.B., 1974. An iterative technique for the rectification of observed distributions. *AJ*, 79 : 745. [119](#)
- Mugnier L.M., Fusco T., & Conan J.M., 2004. MISTRAL : a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images. *J. Opt. Soc. Am. A*, 21(10) : 1841. [120](#)
- Mugnier L.M., Robert C., Conan J.M., Michau V., & Salem S., 2001. Myopic deconvolution from wave-front sensing. *J. Opt. Soc. Am. A*, 18(4) : 862. [120](#)
- Ridolfi A. & Idier J., 2001. Penalized Maximum Likelihood Estimation for univariate normal mixture distributions. Dans *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, édité par A. Mohammad-Djafari, volume 568 de *American Institute of Physics Conference Series*, p. 229–237. [119](#)
- Thiebaut E. & Conan J.M., 1995. Strict a priori constraints for maximum-likelihood blind deconvolution. *J. Opt. Soc. Am. A*, 12(3) : 485. [121](#)

# Travaux en cours et perspectives

---

## Sommaire

<b>6.1 Construction de la séquence de Hubble</b>	<b>126</b>
6.1.1 Caractéristiques principales du programme	126
6.1.2 Objectifs scientifiques	127
6.1.3 Des outils prêts	127
6.1.4 Du neuf par rapport à NICMOS?	128
6.1.5 Une nouvelle approche, l'étoile laser	128
6.1.6 Réactions aux premières soumissions	130
<b>6.2 Nouveaux enjeux de l'optique adaptative</b>	<b>130</b>
<b>Bibliographie</b>	<b>132</b>

---

**Abstract :** We describe a *Large Programme* proposal which has been submitted to ESO. The main goal is to study the build-up of the Hubble sequence using the morphological information offered by AO imaging in the K-band. The proposal aims at observing  $\sim 800$  objects in a  $\sim 70(')^2$  area in the COSMOS field. At the moment of writing this report, the proposal is not yet accepted despite the optimistic comments from the OPC. The proposal can be found in appendix C

**Resumen :** El presente capítulo describe un programa de observaciones propuesto a la ESO durante el transcurso de la tesis doctoral. El objetivo principal es el estudio de la construcción de la secuencia de Hubble mediante el análisis morfológico de  $\sim 800$  galaxias contenidas en un área de  $\sim 70(')^2$  observadas en la banda K con óptica adaptativa.

L'exploitation des données du *programme pilote* nous a permis d'étalonner les performances de l'optique adaptative pour étudier la morphologie de galaxies à grand *redshift*. Nous avons vu en effet que malgré les difficultés techniques, on atteint une résolution nettement supérieure qu'avec des télescopes terrestres sans correction et que la qualité de la morphologie obtenue est comparable (jusqu'à  $K_{AB=21}$ ) à celle

obtenue avec le télescope spatial. Dans ce contexte, nous avons décidé de proposer un programme d'observation plus large, visant à étudier la construction de la séquence de Hubble au cours du temps. Le programme s'inscrit dans le cadre d'un *Large Programme* ESO dont je suis le « chercheur principal » et est le résultat de la collaboration entre 10 instituts de recherche de 5 pays européens (c.f. annexe C). La proposition a été envoyée à l'ESO mais malgré les commentaires très positifs du comité de sélection elle n'a pas été acceptée. Il est en particulier demandé de vérifier les performances obtenues avec une étoile laser.

Etant donné cependant que la préparation de ce programme d'observations a occupé une partie non négligeable de mon travail et qu'il va être probablement soumis une nouvelle fois, il m'a semblé important de décrire brièvement les caractéristiques principales de celui-ci. Pour plus de détails, consulter l'annexe C contenant une copie de la proposition envoyée.

## 6.1 Construction de la séquence de Hubble depuis $z \sim 1$ à l'aide d'observations d'OA avec étoile laser

### 6.1.1 Caractéristiques principales du programme

L'objectif principal du programme est d'observer une aire de  $\sim 70''^2$  dans les champs COSMOS avec le système d'optique adaptative NAOS/CONICA du VLT asservi par une étoile laser dans la bande  $K_s$  ( $2.12\mu m$ ) afin d'obtenir des données sur  $\sim 800$  objets. Comme il a été dit dans les chapitres précédents, ce programme présente un double intérêt. D'une part, les observations dans le proche infrarouge permettent de sonder des populations stellaires vieilles autour de  $z \sim 1$  et on est donc moins sensibles aux effets de k-correction morphologique. D'autre part, la très haute résolution angulaire obtenue ( $\sim 0.1''$ ) permet une étude de la morphologie avec une qualité comparable avec celle du télescope spatial dans le visible comme il l'a été prouvé à l'aide du programme pilote. Le fait d'utiliser une étoile laser est une nouveauté importante par rapport au programme pilote et comme je le montrerai par la suite (§ 6.1.5), elle présente un certain nombre d'intérêts. Etant donné que les champs observés appartiennent au relevé COSMOS nous aurons accès aux *redshifts* photométriques de tous nos objets mais aussi à un certain nombre de *redshifts* spectroscopiques (zCOSMOS). Ceci permettra donc d'analyser pour la première fois, l'évolution des comptages, des luminosités, des fonctions de corrélation ainsi que d'autres estimateurs d'évolution pour les 3 types morphologiques principaux (elliptique, spiral, irrégulier) sur au moins deux intervalles de *redshift* à partir de données infrarouges à très haute résolution angulaire.

### 6.1.2 Objectifs scientifiques

On espère donc obtenir de l'ordre de 800 objets avec une résolution moyenne de 0.1'' allant jusqu'à un *redshift* de  $z \sim 1.5$  et une profondeur de  $K_s < 22$ . Il s'agit donc d'un gain d'un facteur 10 en nombre d'objets. En combinant les morphologies obtenues sur ces objets avec les *redshifts* du relevé COSMOS, on essayera de répondre à quelques unes des questions soulevées lors du chapitre introductif.

- *Quand les galaxies ont-elles formé leurs étoiles ?* Comme on l'a vu, l'histoire de la formation des galaxies elliptiques en particulier pose toujours des problèmes. Ce sont-elles formées après un épisode violent de formation stellaire à des *redshifts* très élevés ou alors évoluent-elles par fusion de plus petites entités ? L'étude de l'évolution de la fonction de luminosité par type morphologique (e.g. [Ilbert et al., 2006](#)) est un outil puissant pour essayer de répondre à ces questions puisqu'elle permet de suivre l'évolution de chaque population. Aujourd'hui, il est admis que l'activité de formation stellaire dépend de la masse de la galaxie ; les galaxies plus massives formant des étoiles plus rapidement (*downsizing*). Cependant, la plupart de ces travaux sont basés sur l'étude de morphologies dans le visible (avec HST) ou en employant des couleurs. Les résultats pourraient être affectés par des effets de k-correction morphologique que l'on peut limiter en obtenant les morphologies dans le proche infrarouge.
- *Comment évolue le taux de fusion ?* Dans un scénario de formation de structures hiérarchique (c.f. chapitre 1), les fusions de galaxies sont censées jouer un rôle clef dans l'évolution de celles-ci. Cependant, d'un point de vue observationnel, les effets des fusions sur l'assemblage de la masse et la formation stellaire n'est toujours pas bien éclairci. Les études précédentes, encore une fois basées pour la plupart sur des morphologies HST, obtiennent des résultats incertains ([Conselice, 2003](#); [LeFèvre et al., 2000](#)). Des données en infrarouge à très haute résolution vont donc permettre de sélectionner les paires de galaxies en fusion sans être biaisé par la présence de zones de formation stellaire, plus visibles dans les bandes bleues.
- *Où les galaxies acquièrent-elles leur morphologie ?* L'environnement, comme on l'a dit à plusieurs reprises dans ce manuscrit, a une influence importante sur l'évolution des galaxies. La présence de plus de galaxies elliptiques dans les zones plus denses de l'Univers semble indiquer que l'évolution se fait plus rapidement dans ces zones (e.g. [Meneux et al., 2006](#)). Encore une fois ces études sont basées sur une sélection morphologique utilisant des couleurs ce qui peut introduire des biais.

### 6.1.3 Des outils prêts

Le travail réalisé pendant la thèse m'a permis de mettre en place les outils nécessaires pour une exploitation rapide des données. Ainsi, le plan de travail s'organise en deux étapes :

1. **Analyse morphologique** : On analysera les objets plus brillants ( $K < 19$ ) avec une décomposition bulbe/disque puisque le travail sur le programme pilote a permis de prouver que, pour ces objets, l'erreur dans la détermination du rapport  $B/T$  est de l'ordre de 20% (c.f. chapitre 4). Pour les objets plus faibles on réalisera une étude non-paramétrique en employant une méthode améliorée développée pendant ma thèse (GalSVM) que je présenterai dans la deuxième partie de ce manuscrit. Cela permettra d'obtenir le type morphologique avec au moins un taux de réussite de l'ordre de  $\sim 80\%$ .
2. **Grandeurs statistiques** : Les galaxies vont être ensuite séparées en deux intervalles de *redshift*. Pour chaque intervalle de « *redshift* – type morphologique » on mesurera la fonction de luminosité en utilisant les codes développés dans le cadre du consortium VVDS. D'après *Zucca et al. (2006)* il faut au moins 30 objets par intervalle pour une estimation correcte des fonctions de luminosité. Cela contraint donc le temps d'exposition nécessaire pour mener à bout ce programme (c.f. annexe C).

#### 6.1.4 Du neuf par rapport à NICMOS ?

L'objectif principal du programme est d'obtenir des morphologies en infrarouge proche à très haute résolution. Est-il possible de réaliser ce travail depuis l'espace actuellement ? Le télescope spatial Hubble est en effet équipé d'une caméra infrarouge (NICMOS) et un relevé est en train d'être réalisé dans le cadre du consortium COSMOS en bande H ( $1.6\mu m$ ). On peut donc se demander si le programme présenté ici apportera un gain significatif par rapport à ce relevé.

Les images NICMOS ont une taille de pixel de  $0.1''$  et une résolution moyenne de  $0.28''$ . Les données du programme pilote ont un pixel deux fois plus petit ( $0.054''$ ) et une résolution moyenne deux fois meilleure ( $\sim 0.1''$ ). De plus, la procédure de reconstruction de la FEP présentée dans le chapitre 4 permet d'obtenir une FEP avec une taille de pixel de  $0.027''$ . Comme on peut l'apprécier dans la figure 6.1 ceci représente un gain crucial puisque d'après les données du programme pilote, la plupart des objets ont un rayon caractéristique plus petit que  $0.3''$ .

#### 6.1.5 Une nouvelle approche, l'étoile laser

La stratégie d'observation du programme incorpore une nouveauté importante par rapport au programme pilote : l'utilisation d'une étoile laser pour asservir la boucle d'optique adaptative. Dans le chapitre 3 on a vu que la correction par optique adaptative nécessite d'une source brillante de référence afin de mesurer les aberrations du front d'onde. Ce dernier point pose un problème en particulier pour des applications cosmologiques car les champs observés sont souvent pauvres en étoiles brillantes, ce qui limite donc fortement le choix des objets observables. Afin de pallier ce problème Foy et Labeyrie ont proposé en 1985 l'utilisation d'une étoile

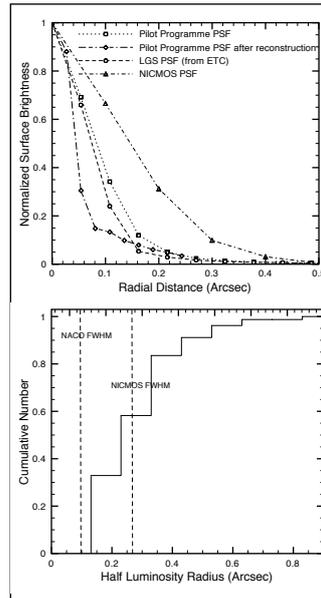


FIG. 6.1 – Gauche : Distribution des rayons contenant la moitié du flux pour les 79 objets analysés dans le programme pilote. Droite : Comparaison des profils radiaux des FEP de NICMOS et de NACO.

laser. Le principe consiste à exciter les atomes de la couche de sodium située à une altitude de  $\sim 90km$  dans l'atmosphère à l'aide d'un laser afin de créer une source d'asservissement artificielle. Cela permet de s'affranchir du besoin d'avoir une étoile assez brillante dans le champ. Il faut malgré tout la présence d'une étoile naturelle, même si plus faible, pour mesurer un mode d'aberration (le *tip-tilt*) qui n'est pas mesurable avec une source artificielle.

Un système d'étoile laser est disponible depuis 2006 au VLT. Nous avons donc décidé de l'utiliser pour nos observations pour quatre raisons fondamentalement :

- L'étoile laser émettant dans le visible uniquement, elle n'est pas visible sur les images scientifiques. Il n'y a donc pas d'aire perdue à cause de l'étoile guide saturée comme dans les premières observations (c.f. Fig. 4.5). Cela diminue donc le temps d'exposition nécessaire pour obtenir le même nombre d'objets.
- Le fait de ne pas sélectionner les champs autour d'étoiles très brillantes permettra sûrement d'obtenir plus de *redshifts* spectroscopiques étant donné que les zones proches de ces étoiles sont souvent masquées dans un relevé spectroscopique.
- La FEP obtenue ainsi que la correction sera plus homogène dans l'ensemble des champs puisque la source d'asservissement reste inchangée.
- Finalement, la résolution obtenue, d'après les simulations, devrait être très similaire ou même meilleure qu'en employant une étoile guide naturelle (Fig. 6.1).

### 6.1.6 Réactions aux premières soumissions

Le programme a été soumis trois fois. Dans tous les cas les réactions du comité ont été positives :

*”There is a clear need for morphological information for galaxies well above  $z=1$ , and the panel recognises this and feel that the goals of the project are valuable and the COSMOS field is clearly the right field to study in view of the supporting information. This proposal also aims to make good use of the adaptive optics imaging facilities at ESO.”*

Les questions posées concernent essentiellement le gain par rapport à NICMOS. Une comparaison plus exhaustive sera donc menée avant la prochaine soumission afin de comparer correctement les FEP obtenues avec les deux instruments.

## 6.2 Nouveaux enjeux de l’optique adaptative

Comme nous avons pu le voir, l’utilisation de l’optique adaptative classique pour l’observation de grands champs cosmologiques est une application extrême de la technique. La plus grande limitation est sans doute l’étroitesse du champ isolplanétique et donc l’impossibilité de pouvoir observer des grands champs avec une correction uniforme. L’utilisation de télescopes au sol pour l’observation de galaxies lointaines est, malgré cela, une priorité et un des objectifs principaux de la nouvelle génération de télescopes (*Extremely Large Telescopes* ou ELTs) ayant plusieurs dizaines de mètres de diamètre.

De nombreux travaux visant à développer de nouvelles techniques d’OA grand champ sont donc en cours. Globalement deux nouveaux concepts ont émergé :

- L’optique adaptative multiconjuguée (MCAO, *Multi Conjugate Adaptive Optics*). Ce concept a été proposé pour la première fois par Dicke (1975) et repris par Beckers (1988). L’idée principale est d’utiliser plusieurs étoiles d’analyse dans le champ afin d’accéder à une connaissance du volume de turbulence plus complète. Plusieurs miroirs déformables sont alors utilisés pour corriger la turbulence (Tallon *et al.*, 1992; Fusco *et al.*, 2000). Le principe est illustré dans la figure 6.2. Les systèmes actuellement en cours d’élaboration visent à obtenir une correction uniforme et proche de la limite de diffraction sur des champs de 1 à 2 minutes d’angle en utilisant de 2 à 3 miroirs déformables. Le problème majeur de cette technique est probablement la nécessité d’un nombre important d’étoiles brillantes ce qui est, encore une fois, rare dans les champs d’intérêt cosmologique. L’utilisation d’une constellation d’étoiles laser est une solution envisageable qui commence à fournir des résultats prometteurs (Lloyd *et al.*, 2006). Récemment un démonstrateur a été installé au VLT (MAD, Marchetti *et al.*, 2003) et proposé à la communauté pour réaliser des tests de performances.
- L’optique adaptative multi-objet (MOAO, *Mult Object Adaptive Optics*). Cette technique a été conçue spécifiquement pour des applications cosmologiques (Puech

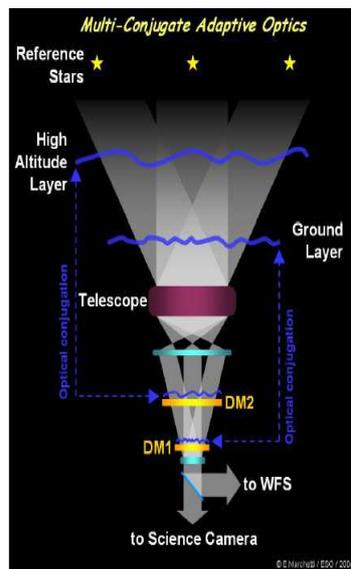


FIG. 6.2 – Illustration du concept d'optique adaptative multi conjuguée (MCAO).  
[Source : ESO]

*et al.*, 2006). L'idée de base est que dans les relevés de galaxies il est nécessaire, certes, d'observer des grands champs, cependant on ne s'intéresse qu'aux objets du champ et non pas au champ en entier. Il suffit donc de corriger les zones autour des objets d'intérêt scientifique. Pour cela il faut qu'il y ait des étoiles brillantes relativement proches de chaque galaxie à analyser ce qui pose encore une fois un problème mais l'utilisation d'étoiles lasers est envisageable. La figure 6.3 illustre cette technique.

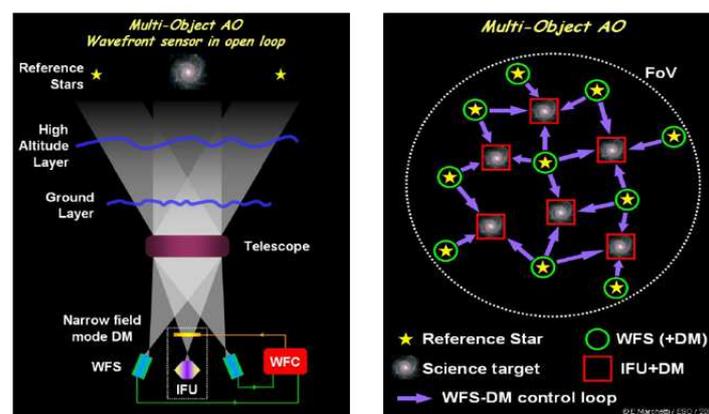


FIG. 6.3 – Illustration du concept d'optique adaptative multi objet (MOAO).  
[Source : ESO]

## Bibliographie

- Beckers J.M., 1988. Increasing the size of the isoplanatic patch with multiconjugate adaptive optics. Dans *ESO Conference on Very Large Telescopes and their Instrumentation, Vol. 2, p. 693 - 703*, volume 2, p. 693–703. 130
- Conselice C.J., 2003. The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories. *The Astrophysical Journal Supplement Series*, 147 : 1. (c) 2003 : The American Astronomical Society. 127
- Dicke R.H., 1975. Phase-contrast detection of telescope seeing errors and their correction. *ApJ*, 198 : 605. 130
- Fusco T., Conan J.M., Mugnier L.M., Michau V., & Rousset G., 2000. Characterization of adaptive optics point spread function for anisoplanatic imaging. Application to stellar field deconvolution. *A&A*, 142 : 149. 130
- Ilbert O., Lauger S., Tresse L., Buat V., Arnouts S., Fèvre O.L., Burgarella D., Zucca E., Bardelli S., Zamorani G., Bottini D., Garilli B., Brun V.L., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Arnaboldi M., Bolzonella M., Cappi A., Charlot S., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Iovino A., McCracken H.J., Marano B., Marinoni C., Mathez G., Mazure A., Meneux B., Merighi R., Paltani S., Pello R., Pollo A., Pozzetti L., Radovich M., Bondi M., Bongiorno A., Busarello G., Ciliegi P., Mellier Y., Merluzzi P., Ripepi V., & Rizzo D., 2006. The VIMOS-VLT Deep Survey. Galaxy luminosity function per morphological type up to  $z = 1.2$ . *Astronomy and Astrophysics*, 453 : 809. 127
- LeFèvre O., Abraham R., Lilly S.J., Ellis R.S., Brinchmann J., Schade D., Tresse L., Colless M., Crampton D., Glazebrook K., Hammer F., & Broadhurst T., 2000. Hubble Space Telescope imaging of the CFRS and LDSS redshift surveys - IV. Influence of mergers in the evolution of faint field galaxies from  $z = 1$ . *Monthly Notices of the Royal Astronomical Society*, 311 : 565. 127
- Lloyd M., Baranec C., Milton N., Snyder M., Stalcup T., & Angel J., 2006. Experimental results of ground-layer and tomographic wavefront reconstruction from multiple laser guide stars. *Opt. Exp.*, 14 : 7541. 130
- Marchetti E., Hubin N.N., Fedrigo E., Brynnel J., Delabre B., Donaldson R., Franza F., Conan R., Le Louarn M., Cavadore C., Balestra A., Baade D., Lizon J.L., Gilmozzi R., Monnet G.J., Ragazzoni R., Arcidiacono C., Baruffolo A., Diolaiti E., Farinato J., Vernet-Viard E., Butler D.J., Hippler S., & Amorin A., 2003. MAD the ESO multi-conjugate adaptive optics demonstrator. Dans *Adaptive Optical System Technologies II. Edited by Wizinowich, Peter L.; Bonaccini, Domenico. Proceedings of the SPIE, Volume 4839, pp. 317-328 (2003).*, édité par P.L. Wizinowich & D. Bonaccini, volume 4839 de *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, p. 317–328. 130

- Meneux B., Le Fèvre O., Guzzo L., Pollo A., Cappi A., Ilbert O., Iovino A., Marinoni C., McCracken H.J., Bottini D., Garilli B., Le Brun V., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Tresse L., Vettolani G., Zanichelli A., Adami C., Arnouts S., Arnaboldi M., Bardelli S., Bolzonella M., Charlot S., Ciliegi P., Contini T., Foucaud S., Franzetti P., Gavignaud I., Marano B., Mazure A., Merighi R., Paltani S., Pellò R., Pozzetti L., Radovich M., Zamorani G., Zucca E., Bondi M., Bongiorno A., Busarello G., Cucciati O., Gregorini L., Lamareille F., Mathez G., Mellier Y., Merluzzi P., Ripepi V., & Rizzo D., 2006. The VIMOS-VLT Deep Survey. The evolution of galaxy clustering per spectral type to  $z = 1.5$ . *A&A*, 452 : 387. 127
- Nicolle M., 2006. *Analyse de front d'onde pour les optiques adaptatives de nouvelle génération*. Thèse de doctorat, Université Paris XI.
- Puech M., Hammer F., Jagourel P., Gendron E., Assémat F., Chemla F., Flores H., Laporte P., Conan J.M., Fusco T., Liotard A., & Zamkotsian F., 2006. FALCON : Extending adaptive corrections to cosmological fields. *New Astronomy Review*, 50 : 382. 130
- Tallon M., Foy R., & Vermin J., 1992. 3-D Wavefront Sensing for Multiconjugate Adaptive Optics. Dans *Progress in Telescope and Instrumentation Technologies*, édité par M.H. Ulrich, p. 517–+. 130
- Zucca E., Ilbert O., Bardelli S., Tresse L., Zamorani G., Arnouts S., Pozzetti L., Bolzonella M., McCracken H.J., Bottini D., Garilli B., Le Brun V., Le Fèvre O., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Arnaboldi M., Cappi A., Charlot S., Ciliegi P., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Iovino A., Marano B., Marinoni C., Mazure A., Meneux B., Merighi R., Paltani S., Pellò R., Pollo A., Radovich M., Bondi M., Bongiorno A., Busarello G., Cucciati O., Gregorini L., Lamareille F., Mathez G., Mellier Y., Merluzzi P., Ripepi V., & Rizzo D., 2006. The VIMOS VLT Deep Survey. Evolution of the luminosity functions by galaxy type up to  $z = 1.5$  from first epoch data. *A&A*, 455 : 879. 128



# Conclusions de la première partie

J'ai voulu montrer dans la première partie de cette thèse que l'optique adaptative peut être employée pour analyser des morphologies de galaxies lointaines dans l'infrarouge proche. Grâce à l'analyse sur un échantillon de 79 galaxies réelles obtenu à partir de mesures profondes avec le système NACO du VLT, j'ai pu montrer que des performances similaires à des données spatiales dans le visible (HST/ACS) peuvent être obtenues au moins pour les objets plus brillants que  $K_{AB} = 21$ . J'ai obtenu ainsi la première estimation quantitative de l'évolution morphologique en trois types à partir de l'imagerie infrarouge. Cependant, il existe encore de nombreuses limitations, la principale étant probablement la taille des champs observables. De nos jours, réaliser une étude sur un échantillon représentatif en utilisant des données d'optique adaptative requiert donc un nombre important d'heures d'observations et un travail complexe de réduction. Les nouvelles techniques d'optique adaptative en cours de développement (en particulier MOAO) et l'avènement des télescopes extrêmement grands, vont probablement résoudre le problème. Aujourd'hui, les ressources disponibles à l'échelle planétaire étant assez limitées, un tel programme se heurte à une concurrence très forte. J'ai voulu, tout de même, réaliser une étude plus complète pendant la durée de ma thèse sous forme d'un *ESO Large Programme* mais ceci n'a encore été malheureusement pas possible.

L'obtention de morphologies de galaxies à grand *redshift* dans l'infrarouge reste malgré tout un besoin important pour les études d'évolution de galaxies. Les difficultés rencontrées à l'heure d'obtenir du temps d'observation pour obtenir un échantillon à haute résolution angulaire m'ont mené à me poser la question : est-il possible d'obtenir une information morphologique à partir de données limitées par le *seeing* ? En effet, des relevés profonds en infrarouge proche limités par le *seeing* mais dans de très bons sites et couvrant une aire du ciel importante sont disponibles actuellement. J'ai donc commencé une réflexion sur des améliorations éventuelles aux techniques d'estimation morphologique afin de les appliquer à des données à plus basse résolution angulaire. Les résultats de ces recherches sont présentés dans la deuxième partie de cette thèse.



## Deuxième partie

# Morphologies automatiques non paramétriques sur des données au sol non corrigées



# Morphologies non paramétriques : état de l'art et limitations

---

## Sommaire

---

7.1	Survol historique . . . . .	140
7.2	Plus de deux paramètres ? . . . . .	140
7.3	Etalonnage de l'espace de paramètres ? . . . . .	141
7.4	Frontières optimales et estimation des erreurs ? . . . . .	141
	Bibliographie . . . . .	142

---

**Abstract :** This chapter is a complement for chapter 2 in which non-parametric methods are described. I focus here more precisely on the main limitations of these methods in order to better understand the new elements that will be presented in the following chapters.

---

**Resumen :** Se trata de un capítulo introductorio que ha de ser visto como un complemento al capítulo 2 en el que se presentan las características principales de los métodos no paramétricos de estimación de morfología. Se presentan pues aquí las limitaciones principales de este tipo de métodos con el fin de mejorar la comprensión de las mejoras aportadas, que serán presentadas en los capítulos siguientes.

---

Dans le deuxième chapitre j'ai décrit les méthodes non paramétriques pour estimer les morphologies de galaxies lointaines. Celles-ci se basent sur la mesure de paramètres corrélés avec le type morphologique de la galaxie. J'ai voulu apporter un certain nombre d'améliorations à ces méthodes afin de les rendre plus générales et de pouvoir les appliquer à des données limitées par le *seeing*. Elles seront présentées dans les chapitres qui suivent. Il me semble cependant nécessaire, pour mieux les comprendre, de revoir d'abord précisément l'état d'art de ces méthodes et leurs principales limitations.

## 7.1 Survol historique

Comme il a été dit, [Abraham et al. \(1994\)](#) et [Abraham et al. \(1996\)](#) définirent deux paramètres, concentration et asymétrie, qui présentent une forte corrélation avec le type morphologique de la séquence de Hubble. Ils montrèrent qu’une représentation sur un plan (*le plan C/A*) de ces deux paramètres permet de définir différentes zones peuplées par des types morphologiques différents. Les galaxies elliptiques sont concentrées et symétriques et se retrouvent donc *en bas à droite* du plan C/A alors que les galaxies plus tardives sont moins concentrées et plus asymétriques et tombent donc *en haut à gauche*.

Le problème principal de ce type de paramètres est leur dépendance avec les conditions d’observation et avec le *redshift* (c.f. chapitre 2). De nombreux auteurs ont donc apporté des modifications afin de les rendre plus robustes (e.g. [Brinchmann et al., 1998](#); [Wu, 1999](#); [Bershady et al., 2000](#); [Conselice et al., 2000](#)).

De plus, d’autres paramètres ont également été proposés, chacun apportant des informations légèrement différentes sur la galaxie : le paramètre de *smoothness* (S) introduit par [Conselice et al. \(2003\)](#) et qui a donné lieu à une classification à trois paramètres appelée CAS, le coefficient de *gini* ([Abraham et al., 2003](#)) et le  $M_{20}$  ([Lotz et al., 2004](#)). Tous ces paramètres ont été définis dans le chapitre 2.

## 7.2 Plus de deux paramètres ?

Chacun de ces paramètres apporte une information différente concernant la galaxie. Une question intéressante est de savoir si l’utilisation simultanée de tous ces paramètres peut, dans certaines conditions, améliorer les performances de la classification. Répondre à cette question n’est malheureusement pas possible avec les techniques actuelles. En effet dans les approches classiques, les zones dans l’espace de paramètres sont définies en traçant des frontières linéaires manuellement. Cette tâche est simple sur le plan C/A original à deux dimensions, devient un peu plus complexe dans l’espace CAS à 3 dimensions et impossible si on augmente encore le nombre de paramètres. [Bershady et al. \(2000\)](#) ont réalisé la première tentative d’analyse multi-dimensionnelle dans un espace à 4 dimensions contenant la concentration, l’asymétrie ainsi que des informations de couleur et de luminosité. Ils trouvèrent des corrélations entre ces paramètres et définirent six plans bi-dimensionnels sur lesquels ils réalisèrent la classification finale. Leur approche utilise donc l’information des 4 paramètres mais la classification finale continue à se faire en 2D. Plus récemment dans le cadre du consortium COSMOS, [Scarlata et al. \(2006\)](#) ont proposé un système de classification multi-dimensionnel en réalisant une analyse par composantes principales dans un espace à 5 dimensions. Le système est cependant construit spécifiquement pour être appliqué à des données du télescope spatial.

### 7.3 Etalonnage de l'espace de paramètres ?

Un autre point fondamental est l'étalonnage de l'espace de paramètres. On a vu en effet que les valeurs des paramètres dépendent, entre autres, des conditions d'observation. Autrement dit, il n'existe pas des *seuils absolus* qui puissent nous permettre d'affirmer : « si la concentration est plus grande que ce seuil, alors il s'agit toujours d'une galaxie elliptique ». Il est donc très important, avant de réaliser une classification, d'étalonner l'espace pour cette application en particulier.

L'approche classique consiste à réaliser l'étalonnage en utilisant un échantillon local pour lequel on connaît bien le type morphologique et garder les mêmes frontières pour classer l'échantillon lointain (Abraham *et al.*, 1996). Cette approche ne tient évidemment pas compte ni des variations à cause de la longueur d'onde (k-correction) ni des conditions spécifiques des observations à analyser.

Une autre approche consiste à sélectionner directement une fraction de l'échantillon à classer et réaliser un classement à l'oeil (Menanteau *et al.*, 2006; Scarlata *et al.*, 2006). Cette approche prend en compte les conditions d'observation mais requiert cependant une résolution angulaire élevée ainsi qu'un bon rapport S/B. Cela est donc envisageable pour des observations spatiales mais devient plus compliqué pour des observations à plus basse résolution. De plus, même pour les données spatiales, l'étalonnage se fait le plus souvent sur les objets plus brillants ce qui n'est pas tout à fait adapté à l'échantillon total et peut donc introduire un biais.

### 7.4 Frontières optimales et estimation des erreurs ?

Dans toutes les approches classiques énoncées ci-dessus, les frontières sont généralement tracées manuellement et de façon linéaire (droite ou plan). Le tracé manuel implique un côté subjectif non reproductible, deux personnes différentes pouvant tracer des frontières différentes menant à des classifications différentes également. On peut alors se demander s'il n'y aurait pas un moyen de s'assurer que les frontières tracées soient optimales. De plus, en quelle mesure pouvons-nous être certains que la frontière linéaire est la mieux adaptée ? Est-il possible d'améliorer les performances en employant des frontières non linéaires ?

Dans les chapitres qui suivent, je décris les améliorations apportées aux méthodes classiques non-paramétriques dans le but de réduire les limitations exposées. En particulier, j'ai utilisé un type de machines d'apprentissage, *les machines à vecteurs support* pour tracer les frontières et un échantillon local de galaxies adapté aux conditions d'observations pour réaliser l'étalonnage. Le chapitre suivant est dédié à la description de quelques propriétés principales de ces machines. Par la suite, je présente le fonctionnement global de la méthode de classification. Les deux derniers

chapitres illustrent quelques applications astrophysiques sur des galaxies de champ et dans des amas.

## Bibliographie

- Abraham R.G., Valdes F., Yee H.K.C., & van den Bergh S., 1994. The morphologies of distant galaxies. 1 : an automated classification system. *ApJ*, 432 : 75. [140](#)
- Abraham R.G., van den Bergh S., Glazebrook K., Ellis R.S., Santiago B.X., Surma P., & Griffiths R.E., 1996. The Morphologies of Distant Galaxies. II. Classifications from the Hubble Space Telescope Medium Deep Survey. *ApJs*, 107 : 1. [140](#), [141](#)
- Abraham R.G., van den Bergh S., & Nair P., 2003. A New Approach to Galaxy Morphology. I. Analysis of the Sloan Digital Sky Survey Early Data Release. *ApJ*, 588 : 218. [140](#)
- Bershady M.A., Jangren A., & Conselice C.J., 2000. Structural and Photometric Classification of Galaxies. I. Calibration Based on a Nearby Galaxy Sample. *AJ*, 119 : 2645. [140](#)
- Brinchmann J., Abraham R., Schade D., Tresse L., Ellis R.S., Lilly S., Le Fevre O., Glazebrook K., Hammer F., Colless M., Crampton D., & Broadhurst T., 1998. Hubble Space Telescope Imaging of the CFRS and LDSS Redshift Surveys. I. Morphological Properties. *ApJ*, 499 : 112. [140](#)
- Conselice C.J., Bershady M.A., Dickinson M., & Papovich C., 2003. A Direct Measurement of Major Galaxy Mergers at  $z < 3$ . *AJ*, 126 : 1183. [140](#)
- Conselice C.J., Bershady M.A., & Jangren A., 2000. The Asymmetry of Galaxies : Physical Morphology for Nearby and High-Redshift Galaxies. *ApJ*, 529 : 886. [140](#)
- Lotz J.M., Primack J., & Madau P., 2004. A New Nonparametric Approach to Galaxy Morphological Classification. *AJ*, 128 : 163. [140](#)
- Menanteau F., Ford H.C., Motta V., Benítez N., Martel A.R., Blakeslee J.P., & Infante L., 2006. The Morphological Demographics of Galaxies in the Advanced Camera for Surveys Hubble Ultra Deep Parallel Fields. *AJ*, 131 : 208. [141](#)
- Scarlata C., Carollo C.M., Lilly S.J., Sargent M.T., Feldmann R., Kampezyk P., Porciani C., Koekemoer A., Scoville N., Kneib J., Leauthaud A., Massey R., Rhodes J., Tasca L., Capak P., Maier C., McCracken H.J., Mobasher B., Renzini A., Taniguchi Y., Thompson D., Sheth K., Ajiki M., Aussel H., Murayama T., Sanders D.B., Sasaki S., Shioya Y., & Takahashi M., 2006. COSMOS morphological classification with ZEST (the Zurich Estimator of Structural Types) and the evolution since  $z=1$  of the Luminosity Function of early-, disk-, and irregular galaxies. *ArXiv Astrophysics e-prints*. [140](#), [141](#)

- Wu K.L., 1999. A Restframe Comparison of Morphological Peculiarities in the Hubble Deep Field - North. Dans *Photometric Redshifts and the Detection of High Redshift Galaxies*, édité par R. Weymann, L. Storrie-Lombardi, M. Sawicki, & R. Brunner, volume 191 de *Astronomical Society of the Pacific Conference Series*, p. 332–+. 140



# Machines à vecteurs support

---

## Sommaire

<b>8.1</b>	<b>Machines d'apprentissage : définition</b>	<b>146</b>
<b>8.2</b>	<b>Les machines à vecteurs support</b>	<b>146</b>
8.2.1	Machines linéaires, données séparables	147
8.2.1.1	Entraînement	147
8.2.1.2	Classification	148
8.2.2	Machines linéaires, données non-séparables	149
8.2.3	Machines non linéaires, données non-séparables	149
<b>8.3</b>	<b>Remarques sur l'entraînement et le réglage des paramètres</b>	<b>152</b>
<b>8.4</b>	<b>Probabilités</b>	<b>152</b>
<b>8.5</b>	<b>Bilan</b>	<b>153</b>
	<b>Bibliographie</b>	<b>153</b>

---

**Abstract :** I present here some properties of support vector machines that will be employed in the following chapters. They are basically a particular kind of learning machine which find the optimal boundaries to separate a number of distributions.

**Resumen :** Se presentan a continuación las propiedades básicas de las *support vector machines* que serán empleadas en los capítulos siguientes. Se trata básicamente de un tipo particular de máquinas de aprendizaje que busca las fronteras óptimas entre varias distribuciones de puntos.

Le propos de ce chapitre est d'introduire les notions de base concernant une famille particulière de machines d'apprentissage appelées machines à vecteurs support (SVM de l'anglais *Support Vector Machines*) qui seront utilisées par la suite. Elles furent introduites pour la première fois dans les années 70 (Vapnik, 1979) mais ce n'est que récemment qu'elles ont commencé à prendre de l'ampleur dans de nombreux domaines d'application. Il ne s'agit ici que d'un court exposé afin de permettre au lecteur de comprendre la suite du manuscrit. Pour une revue formelle, le lecteur intéressé pourra consulter les références Vapnik (1995); Burges (1998); Gunn (1998) ainsi que celles qui seront indiquées au fur et à mesure.

Tout le long de l'exposé, les variables en gras représenteront des grandeurs vectorielles.

## 8.1 Machines d'apprentissage : définition

De manière générale, l'objectif des machines d'apprentissage est d'apprendre à classer des objets à partir d'une série de mesures sur ceux-ci. L'apprentissage se fait en employant un *échantillon d'apprentissage* dont on connaît la classe et les mesures. La machine doit être ensuite capable de classer un échantillon dont on ne connaît que les mesures.

Toute classification effectuée à l'aide des machines d'apprentissage se structure donc en deux étapes, une première étape dite d'apprentissage où la machine est entraînée pour résoudre un problème donné, et une deuxième étape de classification proprement dite où la machine classe un échantillon inconnu.

Un échantillon d'apprentissage doit donc être composé de :

1. Un vecteur de mesures :  $\mathbf{x}_i \in \mathbb{R}^n, i = 1, \dots, l$ ;
2. Une classe associée à chaque mesure  $y_i$  que l'on connaît avec certitude

Celui-ci peut correspondre par exemple à un échantillon de galaxies dont on connaît de manière fiable le type morphologique ( $y_i = [\text{spirale, elliptique, irrégulière}]$ ) et dont on a mesuré les paramètres morphologiques présentés dans le chapitre 2 ( $\mathbf{x}_i = [C, A, S, G]$ ). On suppose alors qu'il existe une distribution de probabilités  $P(\mathbf{x}, \mathbf{y})$  qui associe une classe probable  $y$  à partir des mesures  $\mathbf{x}$ . Une machine d'apprentissage est donc définie comme une famille de fonctions  $f(\mathbf{x}, \alpha)$ ,  $\mathbf{x}_i \mapsto y_i$  permettant le passage de l'espace de mesures à l'espace des classes. Le choix d'une valeur de  $\alpha$  pour un cas particulier donne lieu à ce que l'on appelle une *machine entraînée*.

## 8.2 Les machines à vecteurs support

Les machines à vecteurs support sont une famille particulière de machines d'apprentissage dont l'approche consiste à essayer de trouver les frontières optimales entre plusieurs nuages de points dans l'espace des mesures.

Afin de simplifier la présentation des caractéristiques principales, on se limitera au cas où il n'y a que deux classes :  $\{\mathbf{x}_i, y_i\}, i = 1, \dots, l$   $y_i \in \{-1, 1\}, \mathbf{x}_i \in \mathbb{R}^d$  sans que cela implique une perte de généralité. On appelle alors *classe positive* le cas  $y = 1$  et *classe négative* le cas  $y = -1$ .

### 8.2.1 Machines linéaires, données séparables

#### 8.2.1.1 Entraînement

Commençons d'abord par le cas le plus simple, c'est-à-dire des machines linéaires entraînées sur des données séparables (Fig. 8.1). Cela signifie simplement qu'il existe un hyperplan séparant les objets appartenant à l'une ou à l'autre classe que l'on appelle *hyperplan séparateur*. Le but de l'algorithme est de trouver cet hyperplan. Ainsi, les points  $\mathbf{x}$  situés sur ce plan vérifient l'équation :

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (8.1)$$

$\mathbf{w}$  étant le vecteur normal à l'hyperplan,  $|b|/\|\mathbf{w}\|$  la distance perpendiculaire de l'hyperplan à l'origine et  $\|\mathbf{w}\|$  la norme Euclidienne de  $\mathbf{w}$ . On note alors  $d_+$  ( $d_-$ ) les distances entre l'hyperplan et le point de la classe positive (négative) plus proche de celui-ci et on définit alors la marge comme  $d_+ + d_-$ . Pour ce cas où les données sont séparables, l'algorithme cherche simplement l'hyperplan qui maximise la marge. On peut l'écrire de la façon suivante :

$$\begin{aligned} \mathbf{x}_i \cdot \mathbf{w} + b &\geq +1 & \text{si } y_i = +1 \\ \mathbf{x}_i \cdot \mathbf{w} + b &\leq -1 & \text{si } y_i = -1 \end{aligned} \quad (8.2)$$

Le système d'équations peut être combiné en une seule famille d'inégalités :

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall i \quad (8.3)$$

Considérons maintenant les points pour lesquels l'égalité dans l'équation 8.2 est vérifiée (pour que ces points existent, il suffit de choisir le bon facteur de normalisation pour  $\mathbf{w}$  et  $b$ ). Ces points sont donc sur l'hyperplan  $H_1 : \mathbf{x}_i \cdot \mathbf{w} + b = 1$  ou  $H_2 : \mathbf{x}_i \cdot \mathbf{w} + b = -1$ . Par conséquent,  $d_+ = d_- = 1/\|\mathbf{w}\|$  et la marge vaut tout simplement :  $2/\|\mathbf{w}\|$ .

Pour trouver l'hyperplan séparateur, il suffit donc de minimiser  $\|\mathbf{w}\|^2$  soumis aux contraintes 8.3. Dans la pratique, cette minimisation peut se faire en employant des multiplicateurs de Lagrange :  $\alpha_i$ ,  $i = 1, \dots, l$ . Les détails de son développement ne sont pas indispensables à la compréhension des principes de fonctionnement des machines et ne sont donc pas exposés dans ce chapitre de présentation. On se limitera à rappeler l'expression du Lagrangien qu'il faut minimiser par rapport à  $\mathbf{w}$  et  $b$ , en imposant que toutes les dérivées par rapport aux coefficients  $\alpha_i$  s'annulent et que tous les coefficients restent positifs :

$$\begin{aligned} \min_{\mathbf{w}, b} [L_P &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^l \alpha_i] \\ \frac{\delta L_P}{\delta \alpha_i} &= 0 \\ \alpha_i &\geq 0 \end{aligned} \quad (8.4)$$

Il est possible de prouver que cela revient à résoudre ce que l'on appelle le problème dual (Fletcher, 1987), c'est-à-dire maximiser le Lagrangien par rapport aux coefficients  $\alpha_i$  en imposant que les dérivées par rapport à  $\mathbf{w}$  et  $b$  s'annulent. Cela implique alors que :

$$\begin{aligned} \mathbf{w} &= \sum_i \alpha_i y_i \mathbf{x}_i \\ \sum_i \alpha_i y_i &= 0 \end{aligned} \quad (8.5)$$

En remplaçant ces conditions dans l'expression du Lagrangien (Eq. 8.4), on obtient l'expression du Lagrangien dual.

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (8.6)$$

Il est important de noter pour la suite, que le Lagrangien dual ne dépend des vecteurs de mesures qu'à travers des produits scalaires :  $\mathbf{x}_i \cdot \mathbf{x}_j$ . Ceci sera utilisé pour introduire des séparateurs non linéaires dans le paragraphe 8.2.3.

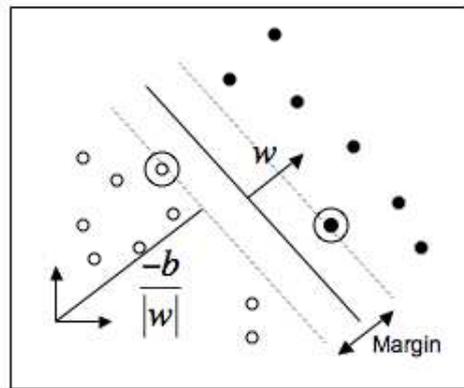


FIG. 8.1 – Illustration du cas linéaire avec des données séparables en deux dimensions. Les points encadrés sont les vecteurs de support.

### 8.2.1.2 Classification

Une fois la machine entraînée, c'est-à-dire une fois l'hyperplan optimal pour un problème donné connu, l'utilisation de celle-ci pour classer un nouvel échantillon est simple. Il suffit d'identifier de quel côté de la frontière est le point que l'on veut classer : la classe d'un vecteur  $\mathbf{x}$  est simplement :  $f(x) = \text{signe}(\mathbf{w} \cdot \mathbf{x} + b)$ . Le processus de classement est donc extrêmement rapide.

### 8.2.2 Machines linéaires, données non-séparables

Si les données ne sont pas séparables par une surface linéaire (Fig. 8.2), ce qui est souvent le cas pour des données réelles, les inéquations 8.2 n'ont pas de solution. Comment peut-on alors étendre la formulation du paragraphe précédent à ce cas plus général? Cortes & Vapnik (1995) ont proposé d'introduire des variables positives  $\xi_i$ ,  $i = 1, \dots, l$  afin d'autoriser des erreurs de classification dans l'échantillon d'entraînement. Ainsi, les équations définissant l'hyperplan optimal deviennent :

$$\begin{aligned} \mathbf{x}_i \cdot \mathbf{w} + b &\geq +1 - \xi_i & \text{si } y_i = +1 \\ \mathbf{x}_i \cdot \mathbf{w} + b &\leq -1 + \xi_i & \text{si } y_i = -1 \\ \xi_i &\geq 0 \forall i \end{aligned} \quad (8.7)$$

Pour qu'une erreur intervienne il faut alors que  $\xi_i$  soit supérieure à 1, donc  $\sum \xi_i$  est une borne supérieure du nombre d'erreurs dans l'échantillon d'entraînement. Le résultat est que la fonction à minimiser est maintenant :  $\|\mathbf{w}\|^2/2 + T(\sum \xi_i)$ , avec T représentant un paramètre choisi par l'utilisateur. Plus T est grand, plus le poids donné aux erreurs est important.

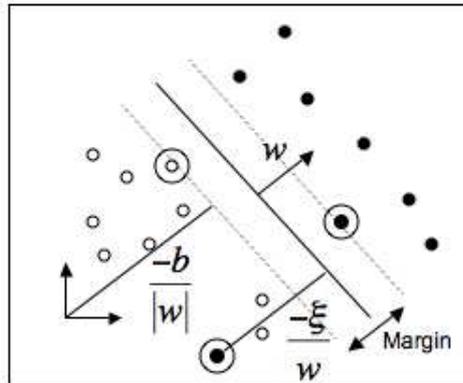


FIG. 8.2 – Illustration du cas linéaire avec des données non séparables.

### 8.2.3 Machines non linéaires, données non-séparables

On est souvent confrontés, lorsqu'il s'agit de classer un échantillon réel, à des problèmes qui sont loin d'être séparables par des frontières linéaires. Le fait d'autoriser des erreurs, comme il a été exposé au paragraphe précédent, confère une solution aux problèmes mais ne veut pas dire que la qualité de la classification soit acceptable. L'une des caractéristiques majeures des machines à vecteur support est qu'il est possible de généraliser davantage le problème au cas où la surface séparatrice n'est plus linéaire. Il s'agit d'un problème plus complexe puisqu'on ne cherche plus à tracer un hyperplan dont l'expression mathématique est simple mais une hypersurface quelconque.

Boser *et al.* (1992) ont montré cependant qu'en utilisant une vieille approche présentée dans Aizerman *et al.* (1964), il est possible de contourner le problème d'une manière élégante et simple. L'idée consiste à projeter les mesures dans un nouvel espace Euclidien de plus grande dimension. Un hyperplan dans ce nouvel espace peut devenir une hypersurface dans l'espace d'origine.

Pour cela il suffit de remarquer que dans le processus de minimisation avec des multiplicateurs de Lagrange, les vecteurs de données n'interviennent qu'à travers des produits scalaires de la forme :  $\mathbf{x}_i \cdot \mathbf{x}_j$  (équation 8.6). Ainsi, il est possible de définir une fonction  $\Phi$  effectuant le passage de l'espace des mesures à un autre espace  $\mathbb{H}$  :

$$\Phi : \mathbb{R}^d \mapsto \mathbb{H} \quad (8.8)$$

L'algorithme d'entraînement ne dépend alors des mesures que par l'intermédiaire de produits scalaires dans  $\mathbb{H}$  de la forme :  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ . S'il existait alors une fonction *noyau*  $K$  telle que :

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \quad (8.9)$$

il suffirait d'utiliser cette fonction pour résoudre le problème de minimisation sans nécessairement connaître la fonction  $\Phi$ . En remplaçant tout simplement  $\mathbf{x}_i \cdot \mathbf{x}_j$  par  $K(\mathbf{x}_i, \mathbf{x}_j)$  dans l'équation 8.6, on produit une machine vivant dans un espace de plus grande dimension (même infini), et ceci sans rendre plus complexe l'algorithme d'entraînement.

Mais cette machine est-elle utilisable facilement une fois entraînée ? Autrement dit, durant la phase de classification, comment peut-on déterminer de quel côté de l'hypersurface se situent les points ? En effet, il semblerait qu'il faut fatalement connaître l'expression du vecteur normal  $\mathbf{w}$  dans l'espace  $\mathbb{H}$ . Par chance, encore une fois, la classification peut se faire en ne faisant intervenir que des produits scalaires entre les vecteurs de support et le vecteur de mesures :

$$f(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i y_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{x}) + b = \sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}_j) + b \quad (8.10)$$

où  $\mathbf{s}_i$  sont les vecteurs de support. Par conséquent, il n'est pas nécessaire de connaître explicitement la fonction  $\Phi$ .

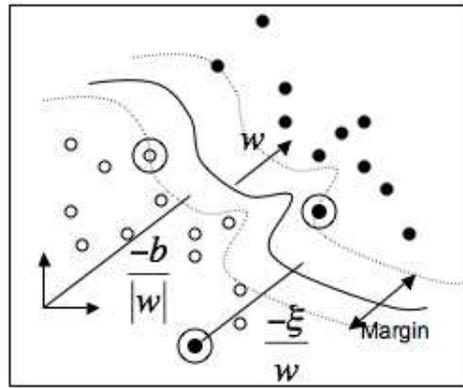


FIG. 8.3 – Illustration du cas non linéaire avec des données non séparables en deux dimensions.

A la vue de ces explications, il peut paraître compliqué de trouver des fonctions satisfaisant les critères nécessaires. Pour mieux l'illustrer, posons donc un exemple simple. Supposons que les données vivent dans  $\mathbb{R}^2$  et que l'on choisit  $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^2$ . Il est alors facile de trouver un espace  $\mathbb{H}$  et une fonction  $\Phi$  allant de  $\mathbb{R}^2$  vers  $\mathbb{H}$  telle que  $(\mathbf{x}_i \cdot \mathbf{x}_j)^2 = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})$ . On peut choisir par exemple  $\mathbb{H} = \mathbb{R}^3$  et :

$$\Phi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix} \quad (8.11)$$

Les questions qui viennent à l'esprit sont donc : existe-t-il toujours une paire  $\{\mathbb{H}, \Phi\}$  ayant les propriétés nécessaires ? Quelles sont les fonctions utilisables ?

La réponse à la première question est donnée par la condition de Mercer (Vapnik, 1995; Courant & D., 1953). Par la suite, on admet que ces fonctions existent.

Quant aux fonctions utilisables, il existe en particulier trois types de fonctions employées classiquement :

Noyau polynomial :	$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^p$
Noyau sigmoïde :	$K(\mathbf{x}, \mathbf{y}) = \tanh(\kappa \mathbf{x} \cdot \mathbf{y} - \delta)$
Noyau RBF (/emphRadial Basis Function) :	$K(\mathbf{x}, \mathbf{y}) = e^{-\ \mathbf{x} - \mathbf{y}\ ^2 / 2\sigma^2}$

Lequel de ces noyaux est-il recommandable d'employer ? La réponse dépend des mesures et de la relation existant entre les mesures et les classes associées. Malheureusement, cette relation est rarement connue. Il est difficile de connaître en avance la forme optimale que devrait avoir la frontière. La grande majorité des auteurs recommandent l'utilisation des noyaux RBF comme la cas le plus général. Ce choix est basé sur plusieurs raisons.

D'une part, Keerthi & Lin (2003) ont prouvé que le cas linéaire est un cas particulier de celui-ci, impliquant que si les données sont mieux séparées par une frontière

linéaire, on devrait la retrouver en employant un noyau RBF. De même, le noyau sigmoïde peut finalement se comporter comme un noyau RBF pour un certain choix de paramètres (Lin & C.J., 2003). D'autre part, le noyau RBF est complètement décrit par un seul paramètre  $\sigma$ , ce qui simplifie le choix en comparaison avec le noyau sigmoïde qui en emploie deux. Finalement, d'un point de vue purement numérique, l'utilisation de celui-ci pose moins de problèmes comparé aux noyaux polynomiaux qui peuvent prendre des valeurs non bornées pour certaines valeurs de  $p$ .

### 8.3 Remarques sur l'entraînement et le réglage des paramètres

Comme il a été expliqué ci-dessus, les SVM peuvent donc être généralisés pour résoudre des problèmes non-séparables et non-linéaires. Pour ce faire, nous avons dû cependant rajouter plusieurs paramètres par rapport au cas de base linéairement séparable. En effet, pour faire face à des cas séparables, il a fallu introduire le paramètre de tolérance  $T$  (Eq. 8.7), et pour autoriser des frontières non-linéaires nous avons introduit le concept de noyau qui incorpore également des paramètres. Le noyau RBF par exemple est paramétrisé par  $\sigma$ . Donc, avant de réaliser le travail d'entraînement proprement dit, il est nécessaire de fixer ces paramètres. Comment cela peut être fait ? Il existe plusieurs manières d'aborder le problème. La plus simple, bien qu'elle ne soit pas optimale du point de vue du temps de calcul, est appelée méthode par *validation croisée*. Il s'agit simplement d'une recherche systématique ; on sépare ainsi l'échantillon d'entraînement en plusieurs sous-échantillons puis chacun de ces sous-échantillons est utilisé pour valider un entraînement réalisé avec le reste des points. La qualité de la *validation croisée* est obtenue comme la moyenne de toutes les prédictions.

Afin de fixer les paramètres  $(T, \sigma)$  on réalise donc des *validation croisées* avec toutes les combinaisons possibles de valeurs dans une grille fixée par l'utilisateur. On garde alors la combinaison donnant la qualité de la *validation croisée* la plus élevée. C'est ensuite ce choix de paramètres qui est utilisé pour entraîner l'échantillon total.

### 8.4 Probabilités

Une machine entraînée pour un problème particulier offre donc en sortie une prédiction strictement binaire : l'objet appartient ou non à une classe donnée. Cependant, il peut parfois s'avérer plus intéressant d'avoir une information plus *nuancée*. Autrement dit, une estimation d'une probabilité d'appartenance à une classe en fonction de la position de l'objet dans l'espace de paramètres :

$$p_i = p(y = i | \mathbf{x}) \quad (8.12)$$

De nombreux auteurs ont proposé des méthodes pour résoudre ce problème. Probablement, la technique la plus répandue est celle imaginée par Platt (1999) et améliorée par la suite par Wu *et al.* (2004). Ils démontrent en effet que la probabilité peut être paramétrée par une loi de la forme :

$$P(y = 1|f) = \frac{1}{1 + \exp(Af + B)} \quad (8.13)$$

où  $f$  est défini dans 8.10. Les paramètres  $A$  et  $B$  sont déterminés par un ajustement sur l'échantillon d'entraînement.

De cette manière, une fois les paramètres connus, il est possible d'appliquer cette loi sur l'échantillon à classer et d'associer ainsi non seulement une classe d'appartenance, mais aussi une probabilité.

## 8.5 Bilan

Les machines à vecteur support sont une classe particulière de machines d'apprentissage offrant une interprétation géométrique directe : elles trouvent la frontière optimale entre plusieurs nuages de points dans un espace  $n$ -dimensionnel. Un processus de classification standard peut se décomposer en quatre étapes principales :

1. Choix du noyau (dans ce qui suit le noyau c'est le noyau RBF qui sera employé) ;
2. Choix des meilleurs paramètres  $(T, \sigma)$  (par validation croisée) ;
3. Entraînement de la machine avec les paramètres sélectionnés ;
4. Test.

A la sortie de la classification, nous obtenons pour chaque objet, une classe estimée et une probabilité d'appartenance à cette classe.

## Bibliographie

- Aizerman M., Braverman E., & Rozoner L., 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control*, 25 : 821. 150
- Boser B.E., Guyon I., & Vapnik, 1992. A training algorithm for optical margin classifiers. *Fifth Annual Workshop on Computational Learning Theory*. 150
- Burges C., 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2 : 121. 145
- Cortes C. & Vapnik V., 1995. Support Vector Networks. *Machine Learning*, 20 : 273. 149

- Courant R. & D. H., 1953. *Methods of Mathematical Physics*. Interscience. 151
- Fletcher R., 1987. *Practical Methods of Optimization*. John Wiley and Sons. 148
- Gunn S., 1998. *Support Vector Machines for Classification and Regression*. 145
- Hsu C.W., Chang C.C., & Lin C.J., 2001. *A practical guide to Support Vector Classification*.
- Keerthi S. & Lin C., 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15 : 1667. 151
- Lin H.T. & C.J. L., 2003. *A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods*. . 152
- Platt J., 1999. *Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods*. 153
- Vapnik V., 1979. Estimation of dependences based on empirical data [en russe]. Nauka. Traduction en anglais Springer-Verlag, 1982. 145
- Vapnik V., 1995. *The nature of statistical learning theory*, volume 536. Springer-Verlag. 145, 151
- Wu T.F., J. L.C., & C. W.R., 2004. Probability estimates for multi-class classification by pairwise coupling. *Journal of machine learning research*, 5 : 975. 153



# GALSVM: méthode non-paramétrique pour l'estimation de morphologies de galaxies lointaines

---

## Sommaire

<b>9.1</b>	<b>Validation sur un échantillon bien résolu</b>	<b>157</b>
9.1.1	C/A classique et SVM 2-D	157
9.1.2	C/A classique et SVM 4-D	158
<b>9.2</b>	<b>Application à des galaxies lointaines</b>	<b>159</b>
9.2.1	Description de la méthode	161
9.2.2	L'échantillon réel	162
9.2.3	L'étalonnage	163
9.2.4	Mesure des paramètres de classification	164
9.2.5	Résultats	165
9.2.5.1	C/A classique vs. 2-D SVM	165
9.2.5.2	n-D vs. 2-D	165
9.2.5.3	Comment fixer le nombre de paramètres?	166
9.2.5.4	Influence de l'échantillon d'entraînement	168
9.2.6	Quelques mots sur le mélange morphologique	169
<b>9.3</b>	<b>Mise à disposition du code</b>	<b>170</b>
<b>9.4</b>	<b>Bilans et conclusions</b>	<b>170</b>
	<b>Bibliographie</b>	<b>171</b>

---

**Abstract :** We present a new non-parametric method to quantify morphologies of galaxies based on support vector machines. The method which can be seen as a generalization of the classical C/A classification but with an unlimited number of dimensions and non-linear boundaries between decision regions, is fully automated and thus particularly well adapted to large cosmological surveys. The source code is available for download at : <http://www.lesia.obspm.fr/~huertas/galsvm.html>. To test the method, we use a seeing limited near-infrared ( $K_s$  band,  $2,16\mu\text{m}$ ) sample

observed with WIRCam at CFHT at a median redshift of  $z \sim 0.8$ . The machine is trained with a simulated sample built from a local visually classified sample from the SDSS, chosen in the high-redshift sample's rest-frame (i band,  $0.77\mu m$ ) and artificially redshifted to match the observing conditions. We use a 12-dimensional volume, including 5 morphological parameters, and other characteristics of galaxies such as luminosity and redshift. A fraction of the simulated sample is used to test the machine and assess its accuracy.

We show that a qualitative separation in two main morphological types (late type and early type) can be obtained with an error lower than 20% up to the completeness limit of the sample ( $KAB \sim 22$ ), which is more than 2 times better than what would be obtained with a classical C/A classification on the same sample and indeed comparable to space data. The method is optimized to solve a specific problem, offering an objective and automated estimate of errors that enables a straightforward comparison with other surveys. Selecting the training sample in the high-redshift sample rest-frame makes the results free from wavelength dependent effects and hence its interpretation in terms of evolution easier.

The results shown here have been published in A&A (Huertas-Company *et al.*, 2008). The paper is presented in appendix E.

---

**Resumen :** Se presenta un método nuevo no paramétrico de estimación morfológica basado en máquinas de vector soporte. Dicho método puede ser analizado como una generalización de los métodos clásicos basados en la concentración y la asimetría a  $n$  dimensiones y con fronteras no lineales. El código fuente es público y puede ser descargado en la dirección : <http://www.lesia.obspm.fr/~huertas/galsvm.html>.

Para probar el método se he empleado una muestra de galaxias en la banda Ks observadas con el CFHT situada a un *redshift* medio de  $z \sim 0.8$ . Se entrena la máquina mediante una muestra local del SDSS adaptada para reproducir las condiciones observacionales e instrumentales de la muestra que se desea analizar. La clasificación se realiza en un espacio de 12 dimensiones que incluye parámetros morfológicos así como otras características de las galaxias. Una parte de la muestra simulada se emplea para el análisis de errores.

Se logra una separación en dos tipos con un error inferior al 20% hasta  $KAB \sim 22$ . Los resultados han dado lugar a una publicación en A&A (Huertas-Company *et al.*, 2008). El contenido íntegro del artículo puede ser consultado en el anexo E.

---

Lorsqu' Abraham *et al.* (1994); Abraham *et al.* (1996) ont proposé leur méthode de classification basée sur les positions des galaxies dans le plan C/A, ils ont montré, en utilisant un échantillon d'étalonnage, qu'il existe une bonne séparation entre les trois types morphologiques grâce à des séparateurs linéaires. Autrement dit, ils ont essayé de maximiser les marges entre trois populations dans un espace à deux dimensions en utilisant des noyaux linéaires. C'est exactement le travail réalisé par

les SVM dans le cas linéaire séparable (c.f. § 8.2.1). De ce point de vue, les machines à vecteur support, représentent une généralisation directe vers un classement non-paramétrique avec plus de dimensions et des frontières non-linéaires. Ce chapitre présente le résultat de ces recherches qui ont donné lieu à une publication dans A&A : (Huertas-Company *et al.*, 2008)

D'autres auteurs (e.g. Brinchmann *et al.*, 1998) ont prouvé que le problème de classifications de galaxies par des méthodes non-paramétriques est loin d'être un problème linéairement séparable du fait des importantes contaminations dans le plan C/A. Nous avons donc décidé d'employer, dans les paragraphes qui suivent, les machines à vecteurs support les plus générales, c'est-à-dire non-linéaires, avec un noyau RBF.

Pour l'utilisation des SVM, nous avons employé la librairie libSVM (Chang & Lin, 2001) disponible gratuitement.

## 9.1 Validation sur un échantillon bien résolu

Dans un premier temps, afin de vérifier que les SVM marchent correctement pour la classification de galaxies, nous avons réalisé un premier test sur un échantillon bien résolu. Nous avons pour cela utilisé un échantillon de  $\sim 500$  galaxies du *Sloan Digital Sky Survey* (SDSS, chapitre 1) en bande i dont la morphologie est bien connue car les galaxies ont été classées à l'oeil par plusieurs personnes (Tasca & White, 2005). Les galaxies sont donc des galaxies proches, bien résolues et avec un bon rapport S/B. Dans ces conditions, les méthodes C/A classiques ont des bonnes performances. Le but est donc de vérifier que l'on peut obtenir au moins les mêmes performances en employant des machines à vecteurs support.

### 9.1.1 C/A classique et SVM 2-D

Nous avons donc mesuré les paramètres de concentration et asymétrie sur l'ensemble de l'échantillon (c.f. § 2.2.1.2) puis représenté les valeurs dans un plan afin de réaliser une classification en deux types (elliptique/spirale). D'un côté, nous avons tracé une frontière linéaire à l'oeil comme dans les travaux précédents, puis d'un autre côté, nous avons employé une machine SVM à deux dimensions. Les résultats sont présentés dans la figure 9.1.

On constate tout d'abord que les frontières obtenues sont différentes étant donné que les SVM ne produisent pas de séparateurs linéaires. Cependant, les performances obtenues sont tout à fait consistantes. En effet, la complétude (fraction de galaxies pour lesquelles la morphologie est estimée correctement) et la contamination (fraction de galaxies mal classées) sont essentiellement les mêmes pour les deux méthodes (c.f. Tab. 9.1).

Afin de confirmer ce raisonnement, mais également pour vérifier qu'il n'y a pas

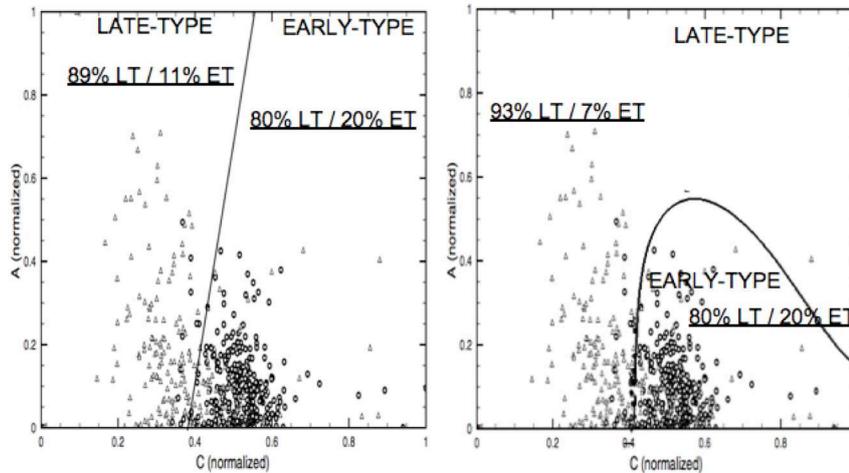


FIG. 9.1 – Comparaison d’une classification C/A classique (gauche) et d’une classification par SVM (droite) sur le même échantillon. Les triangles représentent des galaxies classées à l’œil comme tardives (*late-type*) et les cercles des galaxies précoces (*early-type*). Les chiffres montrent les probabilités qu’une galaxies classée automatiquement ait le même type morphologique qu’à l’œil.

de biais additionnels, nous avons réalisé également une comparaison entre les deux méthodes galaxie par galaxie. On trouve que 98% (94%) des galaxies classées elliptiques (spirales) par une méthode le sont aussi par la deuxième.

Nous pouvons donc conclure que pour un échantillon bien résolu, tracer les frontières avec des SVM ou à la main est équivalent. L’avantage principal de l’emploi des SVM étant cependant que le tracé se fait de manière automatique et optimale pour réduire les contaminations. Qu’en est-il si on augmente le nombre de paramètres ? Y a-t-il un gain significatif pour cet échantillon ?

### 9.1.2 C/A classique et SVM 4-D

En effet, un des avantages majeurs de l’utilisation de SVM est qu’un nombre de dimensions théoriquement illimité peut être employé. Il est donc intéressant de regarder si cela apporte un gain lorsque l’échantillon est bien résolu et que les méthodes à deux dimensions marchent déjà correctement. Pour ce faire, nous avons donc entraîné une machine à quatre paramètres : concentration, asymétrie, *smoothness* et *gini*. Les résultats sont résumés dans le tableau 9.1. On constate qu’il n’y a pas un gain significatif. Nous avons confirmation que, lorsqu’il s’agit d’un échantillon bien résolu et ayant un bon rapport S/B, les méthodes classiques à deux dimensions marchent correctement et l’apport de dimensions additionnelles ne permet pas d’améliorer la qualité.

## 9.2 Application à des galaxies lointaines observées avec des télescopes terrestres

Si l'on observe des objets à grand *redshift* avec un télescope au sol, le rapport S/B décroît, ils deviennent mal résolus et, par conséquent, plus symétriques et moins concentrés (e.g. [Conselice et al., 2000](#)). La séparation dans le plan C/A devient moins claire et le tracé des frontières difficile. C'est pour cela que des données spatiales sont généralement employées (c.f. première partie). Quand il s'agit de données prises au sol, les classifications se font généralement en employant des couleurs (e.g. [Zucca et al., 2006](#)). Il est bien connu cependant (c.f. § 2.2.1.3) que les classifications basées exclusivement sur des couleurs peuvent présenter d'importantes contaminations à cause de la présence par exemple de galaxies elliptiques *bleues* qui n'existent pratiquement pas dans l'Univers local. Une classification morphologique structurale est donc souhaitable d'autant plus qu'un nombre croissant de données sol est disponible à la communauté.

Dans les paragraphes qui suivent, nous nous interrogerons sur le fait suivant : les possibilités offertes par les machines à vecteurs support, telles que l'utilisation d'un grand nombre de dimensions ou l'emploi de frontières non linéaires, peuvent-elles permettre d'obtenir un gain dans la classification de galaxies lointaines observées depuis le sol ?

TAB. 9.1 – Comparaison des performances de trois méthodes (C/A classique, SVM à deux dimensions et SVM à 4 dimensions) sur un échantillon local bien résolu. Le tableau montre pour chaque méthode les relations entre le type visuel et le type automatique.

	C/A classique		SVM C/A		SVM 4-D	
	Early-Type	Late-Type	Early-Type	Late-Type	Early-Type	Late-Type
Visual Early-Type	0.80 (254)	0.09 (17)	0.79 (256)	0.08 (15)	0.79 (251)	0.10 (20)
Visual Late-Type	0.20 (65)	0.91 (172)	0.21 (72)	0.92 (166)	0.21 (67)	0.90 (171)

TAB. 9.2 – Comparaison des performances de classification sur l'échantillon WIRCam avec trois méthodes différentes : C/A classique, C/A SVM et SVM à 12 dimensions. Le nombre d'objets dans chaque intervalle est indiqué entre parenthèses.

	C/A classique		SVM C/A		SVM 12-D	
	Early-Type	Late-Type	Early-Type	Late-Type	Early-Type	Late-Type
Visual Early-Type	0.59 (96)	0.51 (321)	0.57 (304)	0.45 (113)	0.75 (365)	0.18 (52)
Visual Late-Type	0.41 (65)	0.49 (309)	0.43 (236)	0.55 (138)	0.25 (149)	0.82 (225)

### 9.2.1 Description de la méthode

La méthode proposée pour répondre à cette question peut être décomposée en quatre étapes principales (Fig. 9.2) :

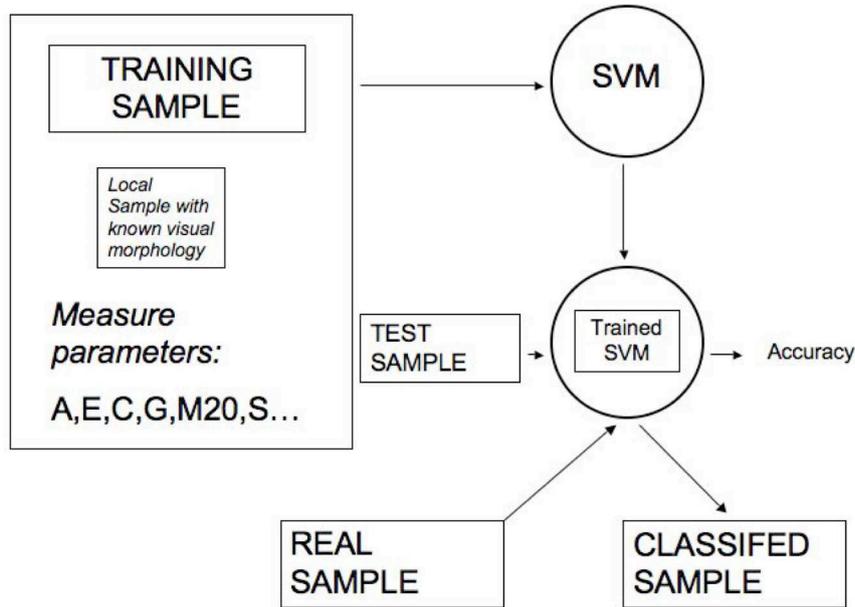


FIG. 9.2 – Etapes de la méthode de classification morphologique proposée.

1. **Construction de l'échantillon d'entraînement** : l'étalonnage est réalisé à partir d'un échantillon local dont on connaît la morphologie qui est ensuite adapté pour reproduire les conditions d'observation (bruit de fond et résolution), et les propriétés physiques (magnitude, *redshift*) de l'échantillon à grand *redshift* à analyser. Ceci est décrit en détail dans le paragraphe 9.2.3.
2. **Mesure des paramètres morphologiques sur l'échantillon d'entraînement** : on obtient les mesures de ces paramètres en utilisant les expressions décrites dans le chapitre 2. Comme on le verra par la suite, d'autres paramètres non-morphologiques pouvant être corrélés avec le type de la galaxie peuvent être également inclus.
3. **Entraînement d'une machine à vecteurs support** sur une fraction de l'échantillon d'entraînement et utilisation de la fraction restante pour estimer les erreurs de classification.
4. **Classification de l'échantillon réel** en utilisant la machine entraînée précédemment.

Dans les paragraphes qui suivent, on décrit chacune de ces étapes plus précisément.

### 9.2.2 L'échantillon réel

Afin d'illustrer la méthode concrètement, nous avons utilisé des données dans l'infrarouge proche (bande Ks,  $2.12\mu\text{m}$ ) obtenues avec la caméra WIRCam (Thibault *et al.*, 2003) installée au CFHT. Le champ observé fait partie du *Canada France Hawaii Legacy Survey* (CFHTLS) et est centré dans le champ COSMOS (Scoville & COSMOS Team, 2005). Une description détaillée de ces données sera faite dans le chapitre 10 qui décrit l'exploitation scientifique. Dans ce chapitre, nous avons utilisé un région de  $10' \times 10'$  dans le seul but d'illustrer la méthode. Il est utile de noter cependant, pour une compréhension correcte de ce qui suit, que les galaxies sont situées à un *redshift* photométrique moyen de  $\sim 0.8$  (Ilbert *et al.*, 2006), que l'échantillon est complet jusqu'à  $K_{AB} \sim 23$  et que la résolution moyenne des images est de  $FWHM \sim 0.7$  (Fig. 9.3).

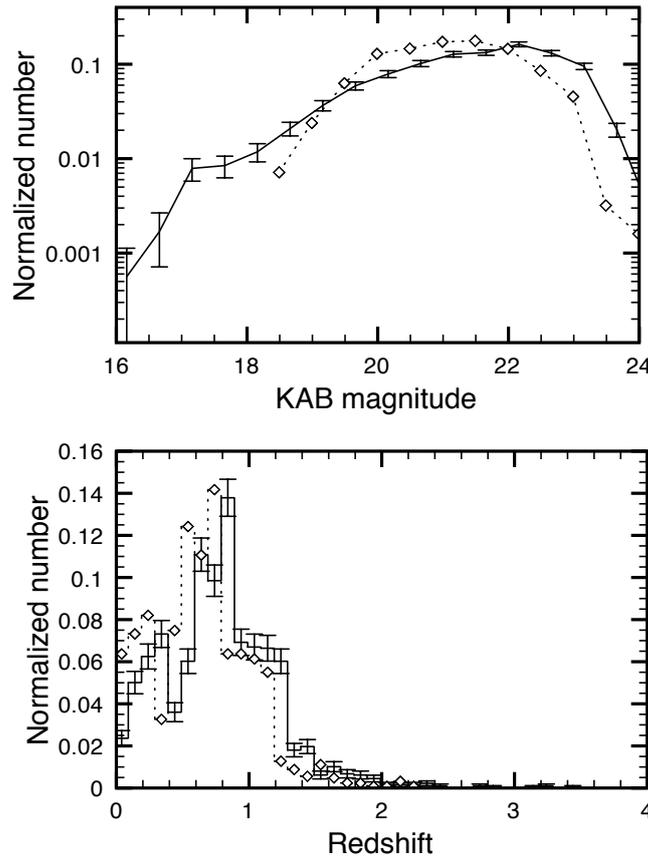


FIG. 9.3 – Distributions de magnitude réelles et simulées. Ligne continue : distribution réelle, ligne en pointillés : distribution simulée.

### 9.2.3 L'étalonnage

L'étalonnage de l'espace de paramètres est donc fait à l'aide d'un échantillon local. Concrètement, nous avons employé le même échantillon que dans le paragraphe 9.1 : 1472 objets du SDSS en bande *i*. L'objectif est de construire un échantillon d'étalonnage le plus proche possible de l'échantillon réel, mais pour lequel la morphologie est connue sans ambiguïté. Afin de réduire les effets de *k*-correction, l'échantillon est choisi dans la bande *i* qui correspond en première approximation à la bande au repos pour un échantillon à  $z \sim 1$  en bande *K<sub>s</sub>*. Il est possible également de raffiner ceci en utilisant plusieurs bandes en fonction du *redshift*, comme je le montrerai par la suite (c.f. § 10.4.1).

On génère ensuite une paire aléatoire de (magnitude, *redshift*) suivant une distribution de probabilités identique aux distributions de magnitude et *redshift* réelles. Ceci a pour but de générer un échantillon synthétique qui ait les mêmes distributions en *redshift* et magnitude que l'échantillon réel (Fig. 9.3). Dans un deuxième temps, une fois que chaque galaxie de l'échantillon local a une paire de valeurs associées, on procède en quatre étapes (Fig. 9.4) :

1. Tout d'abord, les étoiles d'avant plan sont supprimées en employant la carte de segmentation de SExtractor. Les zones où des étoiles sont détectées sont remplacées par un bruit aléatoire ayant la même moyenne et variance que le bruit de l'image.
2. La résolution est dégradée. On ramène le pixel à la taille angulaire qu'il aurait à grand *redshift* et on convolue par une gaussienne de largeur angulaire égale au *seeing* des observations.
3. Troisièmement, l'image de la galaxie est rééchantillonnée pour que la galaxie acquière la taille angulaire attendue avec la taille de pixel des observations à grand *redshift* (dans ce cas 0.15"). Dans cette étape, l'intensité est également mise à l'échelle pour que la galaxie ait la magnitude qui lui a été assignée. On introduit une contrainte supplémentaire : la valeur moyenne du fond de l'image créée doit être au moins un facteur 3 inférieure à la valeur du fond réel afin d'éviter que le bruit de l'image original domine. Cela a pour conséquence que les objets trop brillants (typiquement  $K_s < 17$ ) ne peuvent pas être simulés car le facteur d'échelle à appliquer s'avère être trop petit pour arriver à réduire le niveau du fond (c.f. Fig. 9.3).
4. L'image ainsi simulée est plongée dans le vrai fond de l'image.

En résumé, l'échantillon simulé reproduit les conditions physiques (magnitude, *redshift*) ainsi que les conditions d'observation (bruit de fond, résolution). De plus, il est choisi dans la bande correspondant à la bande au repos de l'échantillon à grand *redshift* réduisant ainsi les effets de *k*-correction.

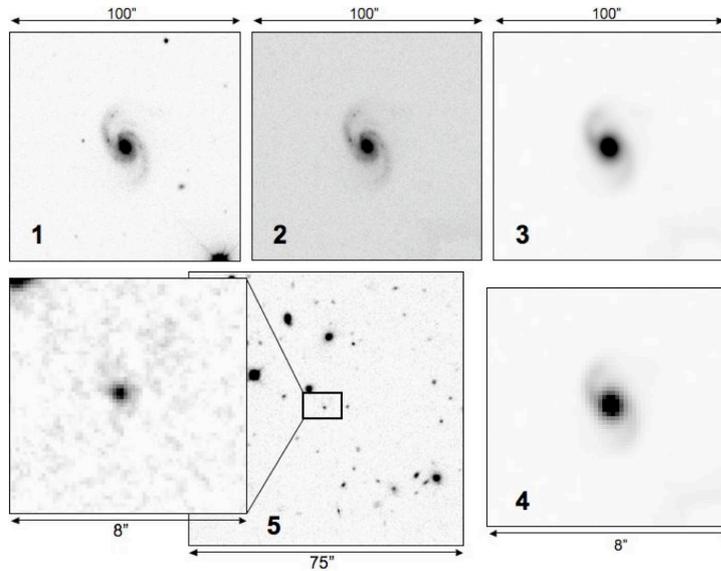


FIG. 9.4 – Illustration des différentes étapes de la construction de l'échantillon d'entraînement. 1 : Image originale, 2 : Suppression des étoiles d'avant plan, 3 : Image après convolution, 4 : Image rééchantillonnée, 5 : Image finale « plongée » dans le champ réel.

#### 9.2.4 Mesure des paramètres de classification

Une fois l'échantillon d'entraînement construit, on mesure des paramètres morphologiques sur celui-ci. On mesure en particulier les 5 paramètres décrits dans le chapitre 2, à savoir : la concentration, l'asymétrie, le gini, la smoothness et le  $M_{20}$ . De plus, comme il a été décrit dans les chapitres précédents, les valeurs obtenues pour ces paramètres peuvent dépendre de la taille, de la luminosité ou du *redshift* de la galaxie (Brinchmann *et al.*, 1998; Bershadsky *et al.*, 2000). Par conséquent, l'inclusion de ces paramètres dans l'espace de classification devrait aider la machine à identifier des biais éventuels. Nous avons donc mesuré 7 paramètres additionnels que l'on a classés en quatre catégories :

- **Forme** : l'ellipticité de la galaxie mesurée par SExtractor (rapport entre le petit et le grand axe des ellipses isophotales de la galaxie) et le paramètre CLASS\_STAR de SExtractor également qui permet de séparer les étoiles des galaxies. Etant donné qu'il peut varier entre 0 et 1, il peut être vu comme une mesure de la compacité de la galaxie.
- **Taille** : dans la taille, on inclut l'aire isophotale de la galaxie ainsi que le rayon pétrosien (c.f. chapitre 2).
- **Luminosité** : la magnitude et la brillance de surface sont utilisées en tant qu'indicateurs de luminosité.
- **Distance** : la mesure de distance est donnée par le *redshift* photométrique.

### 9.2.5 Résultats

Afin de tester la méthode, nous avons réalisé un certain nombre de tests décrits dans les paragraphes qui suivent. Dans tous les cas, la même procédure est adoptée : l'entraînement de la machine se fait avec un échantillon de 500 galaxies de catalogue simulé, et le reste de l'échantillon (*échantillon de test*, i.e.  $\sim 1000$  galaxies) est employé pour mesurer la qualité des classifications. La classification est limitée à deux classes (tardives et précoces) ; d'une part, parce que l'échantillon local utilisé présente très peu de galaxies irrégulières, et d'autre part, parce que la qualité des images permettrait difficilement de réaliser un classement plus fin. De plus, dans la plupart des études statistiques, il est largement suffisant de séparer les galaxies en deux grandes classes. Cela n'implique cependant pas une perte de généralité de la méthode. Le même type d'analyse peut être réalisé avec un nombre de classes plus important.

#### 9.2.5.1 C/A classique vs. 2-D SVM

Est-il possible de classer cet échantillon avec une approche C/A classique ?

Nous avons représenté dans le plan C/A les objets les plus brillants de l'échantillon d'entraînement ( $K_s < 20$ ) pour lesquels le type morphologique est connu, et nous avons essayé de tracer une frontière linéaire afin de séparer les deux populations. Comme prévu, les deux distributions sont maintenant difficilement séparables et le tracé de la ligne droite devient extrêmement difficile. Cela est confirmé par les performances obtenues lorsque tout l'échantillon est classé (Tab. 9.2) : les complétudes et les contaminations sont essentiellement les mêmes que celles obtenues avec un classement aléatoire (autour de  $\sim 50\%$ ). L'approche classique ne peut donc pas être utilisée pour classer cet échantillon.

Qu'en est-il d'une classification avec une machine à vecteurs support à 2 dimensions (C et A) ? Le fait d'utiliser des frontières non-linéaires apporte-t-il un gain ? Les résultats sont montrés également dans le tableau 9.2. On constate un léger gain, mais les performances sont toujours compatibles avec une classification aléatoire.

#### 9.2.5.2 n-D vs. 2-D

Peut-on améliorer la qualité de la classification en augmentant le nombre de paramètres ? Pour répondre à cette question, nous avons entraîné une machine à deux dimensions (C et A) qui doit globalement donner les mêmes performances qu'une classification C/A (d'après ce qui a été vu dans le paragraphe 9.1) puis une machine à 12 dimensions avec tous les paramètres décrits dans le paragraphe 9.2.4. Les résultats de la comparaison sont représentés dans le tableau 9.2. On constate que l'inclusion de tous les paramètres résulte en un gain très significatif. En effet, alors qu'une classification à deux paramètres est presque aléatoire, celle à 12 dimensions

a un taux de réussite de l'ordre de 80%.

Afin de vérifier la robustesse de cet effet, nous avons étudié la qualité des classifications en fonction de trois propriétés principales des galaxies : la luminosité, la distance et l'aire isophotale (Fig. 9.5). Pour ce faire, des objets sont rajoutés au fur et à mesure, et on calcule à chaque fois la fraction globale de galaxies classées correctement puis celle par type morphologique, c'est à dire le nombre d'elliptiques classées en tant qu'elliptiques et le nombre de spirales identifiées comme spirales ( $N_{E \rightarrow E}$ ,  $N_{S \rightarrow S}$ ).

Plusieurs conclusions peuvent être tirées de cette comparaison :

- L'utilisation de plus de deux paramètres fait augmenter la qualité globale de la classification dans tout le domaine de variation des magnitudes, *redshift* et des surfaces. En effet, dans le cas à deux dimensions, le taux de réussite moyen est autour de  $\sim 60\%$  et décroît jusqu'à  $\sim 50\%$ , ce qui implique une forte contamination dans le plan C/A, en accord avec les résultats du paragraphe 9.2.5.1. Dans le cas à 12 dimensions cependant, le taux de réussite se situe autour de 80%, ce qui est comparable à celui obtenu avec des données spatiales (Menanteau *et al.*, 2006).
- Le gain se traduit également dans les performances propres à chaque type morphologique. Dans le cas à deux dimensions, la réponse de la machine est asymétrique : les galaxies elliptiques sont classées avec un taux de réussite de  $\sim 65\%$  alors que pour les galaxies spirales celui-ci descend à  $\sim 50\%$ . Ceci implique une surestimation des galaxies elliptiques ; une fraction importante des galaxies spirales est classée comme elliptique. Lorsque 12 dimensions sont employées, la réponse devient symétrique.

### 9.2.5.3 Comment fixer le nombre de paramètres ?

A la vue des ces résultats on peut se demander s'il existe un nombre de paramètres optimal ? Autrement dit, tous ces paramètres sont-ils nécessaires ? Ou alors, y a-t-il des paramètres susceptibles d'introduire des biais dans la classification voire de dégrader la qualité de celle-ci ?

Afin d'y répondre, nous avons réalisé un test simple qui consiste à entraîner plusieurs machines avec un nombre de paramètres croissant. On commence par une machine à deux dimensions (C et A) et on rajoute successivement des paramètres. Pour chaque machine, on étudie les performances sur l'échantillon de test. Le résultat est montré dans la figure 9.6.

On constate d'une part que tous les paramètres n'apportent pas la même quantité d'informations. Les paramètres de type morphologique et de forme suffisent pratiquement pour obtenir la qualité maximale. En particulier, le coefficient gini ainsi que l'ellipticité semblent avoir un poids important. D'autre part, le taux de réussite est une fonction monotone croissante du nombre de paramètres. Information

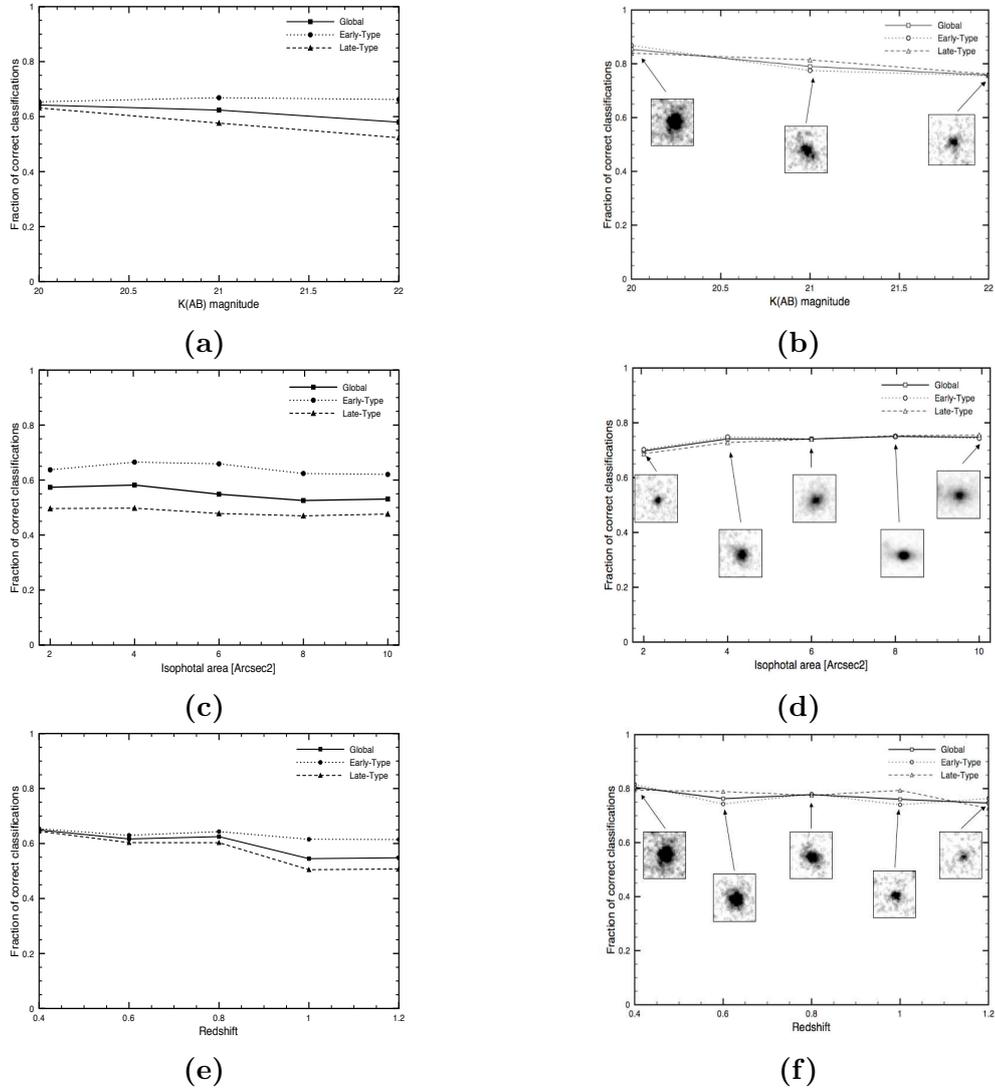


FIG. 9.5 – Performances cumulées des classifications pour une machine à 2D (colonne de gauche) et à 12D (colonne de droite) en fonction de la magnitude (a et b), de l’aire (c et d) et du *redshift* (e et f). Un échantillon de  $\sim 1000$  galaxies est employé pour réaliser les tests. Les lignes continues représentent les performances globales (nombre de galaxies classées correctement) et les lignes discontinues les fractions d’elliptiques et de spirales classées correctement. Les images montrent une galaxie typique dans un intervalle de magnitude, *redshift* ou aire donné.

particulièrement importante puisqu'elle implique que l'inclusion d'un nombre trop important de paramètres ne résulte pas en une dégénérescence. De plus, cela n'implique pas non plus une augmentation significative du temps de calcul.

On verra cependant dans le chapitre 10 que lorsque l'on classe des objets réels et non pas des simulés, certains paramètres sont susceptibles d'introduire des biais s'ils ne sont pas représentatifs de la distribution réelle.

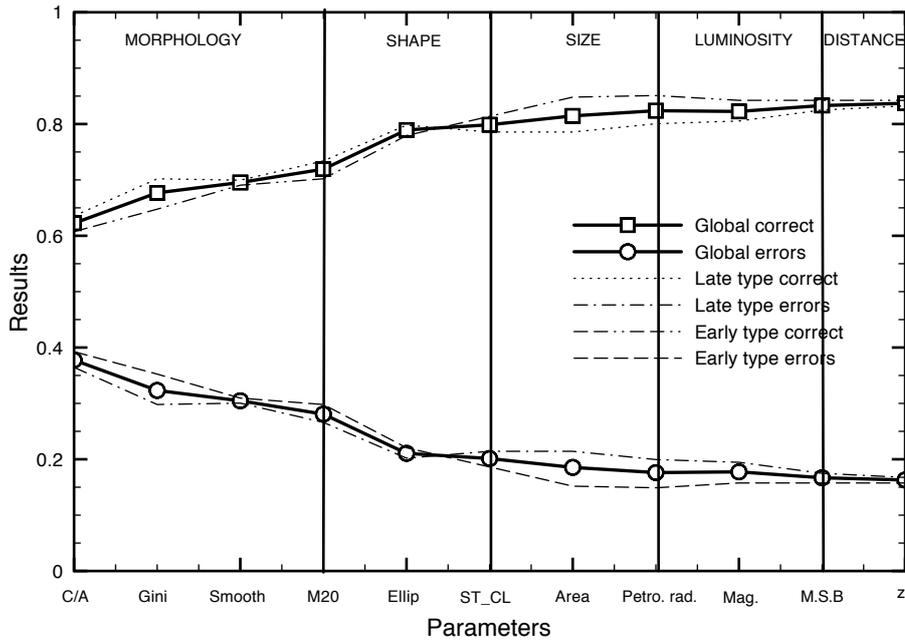


FIG. 9.6 – Résultats des classifications en fonction du nombre de paramètres employés. Le premier point correspond à une classification C/A et chaque point correspond à un nouveau paramètre ajouté.

#### 9.2.5.4 Influence de l'échantillon d'entraînement

Un autre point important est l'influence de l'échantillon d'entraînement. En effet, la méthode décrite ici est basée sur un échantillon qui essaye de reproduire le plus fidèlement possible les propriétés physiques et instrumentales de l'échantillon à analyser. Jusqu'à quel point est-ce important ? Peut-on employer un échantillon d'entraînement même s'il n'est pas tout à fait adapté au problème à résoudre ?

Nous avons simulé un échantillon reproduisant les conditions d'observation des données NACO de la première partie puis entraîné une machine qui a été utilisée pour classer l'échantillon WIRCam. Les résultats sont montrés dans le tableau 9.3. La qualité globale de la classification décroît jusqu'à  $\sim 62\%$  si le modèle NACO est employé. On en conclut donc que l'adaptation de l'échantillon d'entraînement aux données est fondamentale. Ce point sera à nouveau abordé lors de l'exploitation scientifique des données WIRCam dans le chapitre 10.

TAB. 9.3 – Effet sur les classifications après une modification de l'échantillon d'entraînement.

	Modèle WIRCam	Modèle NACO
Global	0.83	0.62
$N_{E \rightarrow E}$	0.81	0.96
$N_{S \rightarrow S}$	0.84	0.24

### 9.2.6 Quelques mots sur le mélange morphologique

Il est intéressant également de se demander quelle est l'influence de la distribution morphologique dans l'échantillon d'entraînement. Il est bien connu en effet, que le nombre de galaxies spirales est supérieur à celui des galaxies elliptiques dans l'Univers local. Ceci veut-il dire que l'échantillon d'entraînement doit refléter cet aspect ? Y-a-t-il une distribution optimale ? De manière générale, les SVM essaient de minimiser les erreurs, c'est-à-dire réduire au maximum le nombre d'identifications incorrectes dans l'échantillon d'entraînement. Cela peut poser des problèmes si les objets de chaque classe ne sont pas représentés équitablement en nombre. En effet, supposons un cas extrême dans lequel l'échantillon d'entraînement contiendrait 95% de galaxies spirales contre seulement 5% de galaxies elliptiques. Dans ce cas, le processus d'optimisation d'erreurs peut mener à prendre la décision suivante : *si je dis que toutes les galaxies sont systématiquement spirales, alors je ne me trompe que dans 5% des cas, obtenant ainsi un taux de réussite de 95%*. La classification d'un échantillon réel avec une machine entraînée de telle façon conduira inévitablement à un échantillon ne contenant que des galaxies spirales.

Ce mécanisme est un problème bien connu des SVM lorsque l'on essaye de classer des données *non équilibrées* (*unbalanced data* en anglais). Certains auteurs (e.g. [Osuna et al., 1997](#)) ont alors proposé d'utiliser des poids ou pénalités différents pour chacune des classes afin qu'une erreur dans la classe moins représentée soit plus pénalisée. La fonction à minimiser devenant alors :  $\|w\|^2/2 + T_1(\sum_{y_i=1} \xi_i) +$

$T_2(\sum_{y_i=-1} \xi_i)$ . La technique de minimisation reste essentiellement la même, mais il

est nécessaire maintenant de choisir des pénalités différentes pour chaque classe, ce qui complexifie inévitablement le problème. C'est pour cette raison que nous avons décidé d'employer systématiquement un échantillon d'entraînement contenant le même nombre d'objets de chaque classe. Il peut cependant s'avérer intéressant d'explorer cette voie en particulier pour classer des galaxies irrégulières. Celles-ci étant généralement beaucoup moins représentées que les galaxies régulières, il est difficile de trouver un échantillon d'entraînement qui soit équilibré. Cela fait partie des travaux en cours au moment de la rédaction de ce manuscrit.

### 9.3 Mise à disposition du code

Le code permettant d'appliquer les différentes étapes de cette méthode à un échantillon de galaxies lointaines a été mis à la disposition de la communauté scientifique sous le nom de *GalSVM*. Un site internet dédié à ce propos a été développé<sup>1</sup>, depuis lequel le code peut être téléchargé. Un manuel d'utilisation ainsi qu'une documentation de chacune des fonctions composant le code ont été également développés (c.f. annexe D).

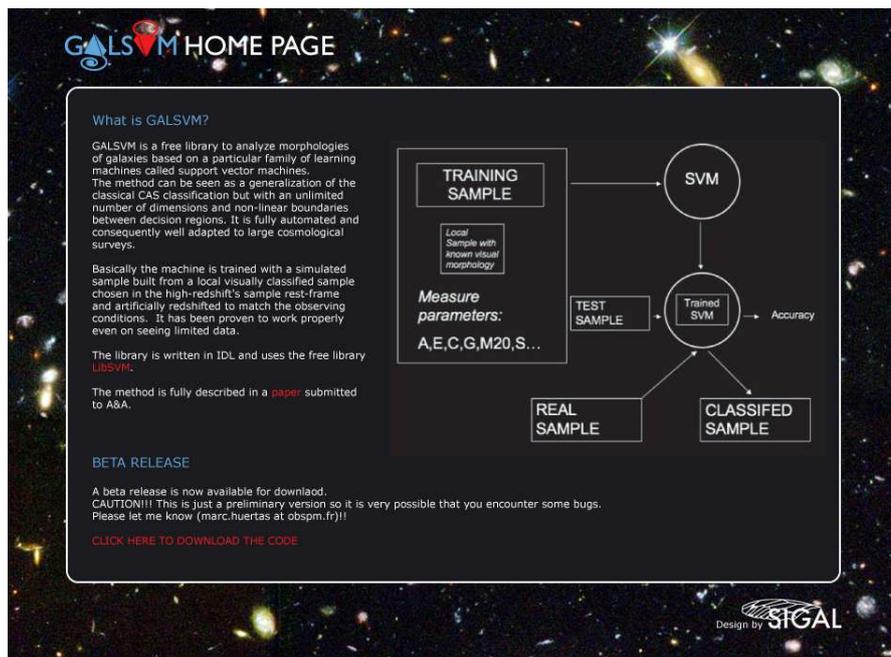


FIG. 9.7 – Interface du site internet permettant de télécharger le code source de GalSVM.

### 9.4 Bilans et conclusions

J'ai présenté une nouvelle méthode pour l'analyse morphologique d'échantillons cosmologiques basée sur l'utilisation de machines à vecteur support. Elle peut être vue comme une généralisation des méthodes non-paramétriques classiques mais avec un nombre illimité de dimensions et des frontières non-linéaires. La méthode est spécialement conçue pour être appliquée à des relevés de galaxies puisqu'elle est complètement automatique et permet une estimation objective et reproductible des erreurs de classement. Le fait d'employer un échantillon d'entraînement dans la bande au repos de l'échantillon à grand *redshift* à analyser rend le résultat plus robuste face aux effets de k-correction permettant une interprétation plus directe

<sup>1</sup><http://www.lesia.obspm.fr/~huertas/galsvm.html>

en termes évolutifs.

## Bibliographie

- Abraham R., van den Bergh S., Glazebrook K., Ellis R., Santiago B., Surma P., & Griffiths R., 1996. The morphologies of distant galaxies. *ApJ Supplement*, 107 : 1. 156
- Abraham R.G., Valdes F., Yee H.K.C., & van den Bergh S., 1994. The morphologies of distant galaxies. 1 : an automated classification system. *ApJ*, 432 : 75. 156
- Bershady M.A., Jangren A., & Conselice C.J., 2000. Structural and Photometric Classification of Galaxies. I. Calibration Based on a Nearby Galaxy Sample. *AJ*, 119 : 2645. 164
- Brinchmann J., Abraham R., Schade D., Tresse L., Ellis R.S., Lilly S., Le Fevre O., Glazebrook K., Hammer F., Colless M., Crampton D., & Broadhurst T., 1998. Hubble Space Telescope Imaging of the CFRS and LDSS Redshift Surveys. I. Morphological Properties. *ApJ*, 499 : 112. 157, 164
- Chang C.C. & Lin C.J., 2001. *LIBSVM : a library for support vector machines*. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. 157
- Conselice C.J., Bershady M.A., & Jangren A., 2000. The Asymmetry of Galaxies : Physical Morphology for Nearby and High-Redshift Galaxies. *ApJ*, 529 : 886. 159
- Huertas-Company M., Rouan D., Tasca L., Soucail G., & Le Fèvre O., 2008. A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images. I. Method description. *A&A*, 478 : 971. 156, 157
- Illbert O., Arnouts S., McCracken H.J., Bolzonella M., Bertin E., Le Fèvre O., Mellier Y., Zamorani G., Pellò R., Iovino A., Tresse L., Le Brun V., Bottini D., Garilli B., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Bardelli S., Cappi A., Charlot S., Ciliegi P., Contini T., Cucciati O., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Marano B., Marinoni C., Mazure A., Meneux B., Merighi R., Paltani S., Pollo A., Pozzetti L., Radovich M., Zucca E., Bondi M., Bongiorno A., Busarello G., de La Torre S., Gregorini L., Lamareille F., Mathez G., Merluzzi P., Ripepi V., Rizzo D., & Vergani D., 2006. Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *AAP*, 457 : 841. 162
- Menanteau F., Ford H.C., Motta V., Benítez N., Martel A.R., Blakeslee J.P., & Infante L., 2006. The Morphological Demographics of Galaxies in the Advanced Camera for Surveys Hubble Ultra Deep Parallel Fields. *AJ*, 131 : 208. 166

- Osuna E., Freund R., & Girosi F., 1997. Improved training algorithm for support vector machines. 169
- Scoville N.Z. & COSMOS Team, 2005. COSMOS. Dans *Bulletin of the American Astronomical Society*, p. 1309–+. 162
- Tasca L. & White S., 2005. Quantitative morphology of Galaxies from the SDSS I : Luminosity in Bulges and Disks. astro-ph/0507249. 157
- Thibault S., Cui Q., Poirier M., Vallee P., Doyon R., Rabou P., & Salmon D.A., 2003. Optical design of WIRCAM, the CFHT wide-field infrared camera. Dans *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. Edited by Iye, Masanori; Moorwood, Alan F. M. Proceedings of the SPIE, Volume 4841, pp. 932-943 (2003).*, édité par M. Iye & A.F.M. Moorwood, volume 4841 de *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, p. 932–943. 162
- Zucca E., Ilbert O., Bardelli S., Tresse L., Zamorani G., Arnouts S., Pozzetti L., Bolzonella M., McCracken H.J., Bottini D., Garilli B., Le Brun V., Le Fèvre O., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Arnaboldi M., Cappi A., Charlot S., Ciliegi P., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Iovino A., Marano B., Marinoni C., Mazure A., Meneux B., Merighi R., Paltani S., Pellò R., Pollo A., Radovich M., Bondi M., Bongiorno A., Busarello G., Cucciati O., Gregorini L., Lamareille F., Mathez G., Mellier Y., Merluzzi P., Ripepi V., & Rizzo D., 2006. The VIMOS VLT Deep Survey. Evolution of the luminosity functions by galaxy type up to  $z = 1.5$  from first epoch data. *AAP*, 455 : 879. 159

# Evolution morphologique de galaxies depuis $z \sim 2$ dans le champ COSMOS

---

## Sommaire

<b>10.1 Le relevé COSMOS</b> . . . . .	<b>175</b>
10.1.1 Description générale . . . . .	175
10.1.2 Données multi-longueur d'onde . . . . .	176
10.1.3 Principaux objectifs scientifiques . . . . .	176
<b>10.2 Motivations du travail</b> . . . . .	<b>177</b>
<b>10.3 Les données</b> . . . . .	<b>178</b>
10.3.1 Description . . . . .	178
10.3.2 Nettoyage . . . . .	179
10.3.3 Séparation étoiles/galaxies . . . . .	179
10.3.4 <i>Redshifts</i> photométriques . . . . .	179
10.3.5 Le catalogue morphologique final . . . . .	180
<b>10.4 Analyse morphologique</b> . . . . .	<b>181</b>
10.4.1 L'échantillon d'entraînement . . . . .	182
10.4.2 Classification . . . . .	182
10.4.3 Remarques sur l'influence du nombre de paramètres . . . . .	183
10.4.4 Catalogue de sortie et qualité des classifications . . . . .	183
<b>10.5 Résultats et discussions</b> . . . . .	<b>186</b>
10.5.1 Estimations des statistiques globales . . . . .	186
10.5.2 Comptages par type morphologique . . . . .	190
10.5.3 Evolution morphologique . . . . .	191
<b>10.6 Autour des effets de k-correction morphologique</b> . . . . .	<b>191</b>
10.6.1 Comparaison galaxie par galaxie . . . . .	193
10.6.2 Evolutions morphologiques comparées . . . . .	193
<b>10.7 Conclusions et perspectives</b> . . . . .	<b>195</b>
<b>Bibliographie</b> . . . . .	<b>198</b>

---

**Abstract** : We use the method presented in the previous chapter to classify  $\sim 50000$  galaxies in the COSMOS area. Galaxies are observed in the near-infrared

at a median *redshift* of  $z \sim 0.8$  with a median FWHM of  $0.8''$  using WIRCam at CFHT. These data are particularly important because K-band data have the advantage of probing old stellar populations in the rest-frame for  $z < 2$  enabling a determination of galaxy morphological types unaffected by recent star formation. We use a 10-dimensional volume, including 5 morphological parameters, and other characteristics of galaxies such as luminosity and redshift. The obtained classification is used to investigate the redshift distributions and number counts per morphological type up to  $z \sim 2$  and to compare to the results obtained with HST/ACS in the I-band on the same objects from other works in order to quantify the *morphological k-correction effects*.

We associate to every galaxy with  $Ks < 21.5$  and  $z < 2$  a probability between 0 and 1 of being late-type or early-type. We use this value to assess the accuracy of our classification as a function of physical parameters of the galaxy and to correct from classification errors. The classification is found to be reliable up to  $z \sim 2$ . The mean probability is  $p \sim 0.8$ . It decreases with redshift and with size, especially for the early-type population but remains above  $p \sim 0.7$ .

We observe an increase of the early-type population from  $z \sim 2$  ( $\sim 20\%$ ) to the present ( $\sim 30\%$ ) probably reflecting a progressive building up of the red sequence from late-type objects. The classification is globally in good agreement with the one obtained using HST/ACS for  $z < 1$ . Above  $z \sim 1$ , the I-band classification tends to find less early-type galaxies than the Ks-band one by a factor  $\sim 1.5$  which might be a consequence of morphological k-correction effects. We argue therefore that studies based on I-band HST/ACS classifications at  $z > 1$  could be underestimating the elliptical population.

These results are described in a paper that will be soon submitted to A&A and which is included in appendix E.

---

**Resumen :** A continuación se usa el método expuesto en el capítulo anterior para clasificar  $\sim 50000$  galaxias en el área de COSMOS. Las galaxias han sido observadas en el infrarrojo cercano con WIRCam (CFHT) y se sitúan a un *redshift* medio de  $\sim 0.8$ . Dichos datos son de particular interés puesto que la banda K permite captar la emisión proveniente de estrellas viejas y ofrece pues una morfología galáctica poco afectada por episodios recientes de formación estelar. Se emplea un espacio de 10 dimensiones para el análisis morfológico. La clasificación obtenida se emplea a continuación para analizar las cuentas por tipo morfológico así como las distribuciones en función del desplazamiento al rojo de cada tipo hasta  $z \sim 2$ . Los resultados son comparados con los obtenidos mediante el telescopio espacial en el óptico con el fin de estimar los efectos de la k-corrección morfológica. Asociamos a cada galaxia con  $Ks < 21.5$  y  $z < 2$  una probabilidad comprendida entre 0 y 1 de ser de tipo temprana o tardía. Se usa dicha medida para establecer la fiabilidad de la clasificación en función de parámetros físicos de las galaxias. La probabilidad media hasta  $z \sim 2$  es de  $p \sim 0.8$ . Se observa un decrecimiento a medida que el *redshift* aumenta pero la probabilidad se mantiene en torno a  $p \sim 0.7$ .

Se observa un aumento de la población de galaxias elípticas desde  $z \sim 2$  hasta nues-

tros días, reflejo probable de la construcción de la secuencia roja a partir de galaxias tardías. La clasificación presenta globalmente un buen acuerdo con la obtenida con el HST en la banda I hasta  $z \sim 1$ . Por encima de  $z \sim 1$ , cuando las cámaras ópticas del telescopio espacial empiezan a captar la emisión ultravioleta, la clasificación en la banda K tiende a encontrar 1.5 veces más de galaxias elípticas.

Los resultados aquí expuestos han dado lugar a una publicación que será enviada a A&A y que puede ser consultada en el anexo E.

---

La méthode présentée dans les chapitres précédents a été employée pour étudier l'évolution des morphologies des galaxies dans le champ COSMOS (*Cosmic Evolution Survey*, Scoville & COSMOS Team, 2005). Ce chapitre décrit les résultats de ces recherches. Dans un premier temps je présente les caractéristiques principales du relevé puis dans un deuxième temps le travail réalisé sur les données WIRCam en particulier. Les résultats ont donné lieu à un article qui a été soumis à A&A et qui peut être consulté en annexe. Les magnitudes seront exprimées dans le système AB tout le long du chapitre.

## 10.1 Le relevé COSMOS

### 10.1.1 Description générale

COSMOS est un relevé cosmologique pensé spécifiquement pour étudier l'évolution couplée entre les structures à grande échelle et les baryons dans l'Univers (e.g. galaxies, formation stellaire, rôle des noyaux actifs). Il s'agit principalement de l'aire la plus grande jamais observée par le télescope spatial Hubble ( $\sim 2 \text{deg}^2$  avec une profondeur de  $I_{AB} = 28$  mag pour les sources ponctuelles). L'observation d'une si grande surface permet de sonder des volumes suffisamment grands dans l'Univers primordial pour réduire de façon significative les effets de variance cosmique.

Le champ étant dans le plan équatorial (AD=10 :00 :28.6, DEC=+02 :12 :21.0), il est accessible à pratiquement tous les instruments au sol. Par conséquent, un suivi extensif (spectroscopie et imagerie) en plusieurs longueurs d'onde (des rayons X à l'infrarouge lointain) est également réalisé (c.f. § 10.1.2 pour plus de détails).

Comme il a été exposé dans les chapitres d'introduction, l'approche multi-longueur d'onde est le seul moyen de s'affranchir des effets dus au *redshift* lorsque l'on sonde des époques cosmiques différentes mais aussi d'étudier simultanément tous les objets de nature très diversifiée mais dont l'évolution semble être fortement couplée (populations stellaires jeunes et vieilles, *starbursts*, AGN...). C'est indispensable également pour obtenir les *redshifts* des objets et par conséquent l'époque dans laquelle il se situent.

Certaines données sont toujours entrain d'être acquises. L'avancement des observa-

tions peut être consulté sur le site officiel du relevé<sup>1</sup>.

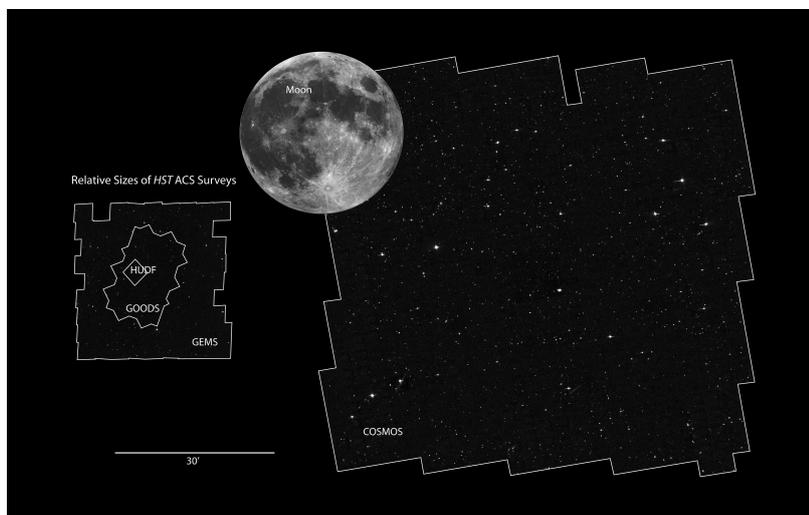


FIG. 10.1 – Le champ COSMOS comparé à la taille de la lune dans le ciel ainsi qu'à d'autres relevés récents. [Source : <http://hubblesite.org>]

### 10.1.2 Données multi-longueur d'onde

En plus de l'imagerie avec le télescope spatial Hubble, le champ COSMOS est suivi par un grand nombre d'instruments :

- **Imagerie sol** : Les télescopes Subaru, CFHT, UKIRT, NOAO ont permis d'obtenir des données photométriques en 12 bandes, de la bande U à la bande Ks, de 2 millions d'objets et par conséquent les *redshifts* photométriques de 860 000 objets (Mobasher *et al.*, 2007). Il s'agit justement de données issues du CFHT qui seront étudiées tout le long de ce chapitre.
- **Spectroscopie sol** : Un programme spectroscopique au VLT (z-COSMOS) vise à obtenir les *redshifts* de  $\sim 30\,000$  objets jusqu'à  $z \sim 3$  (Lilly *et al.*, 2007). De plus, un deuxième programme centré sur l'étude des noyaux actifs est entrain d'être conduit avec Magellan/IMACS (Impey *et al.*, 2007).
- **Espace** : Le satellite XMM a réalisé un relevé en rayons X de tout le champ (Hasinger *et al.*, 2007) et le satellite GALEX dans l'ultraviolet. Spitzer a réalisé également des observations dans l'infrarouge moyen et lointain (Sanders *et al.*, 2007).

### 10.1.3 Principaux objectifs scientifiques

Le relevé trace l'évolution des structures tout le long d'une période couvrant 75% de l'âge de l'Univers. La figure 10.2 montre le volume comobile sondé en fonction du

<sup>1</sup><http://www.astro.caltech.edu/~cosmos/>

*redshift*. Il est intéressant de constater que le relevé plus large réalisé dans l'Univers local, le SDSS (c.f. chapitre 1), sonde un volume de l'ordre de  $3 \times 10^7 h^{-3} \text{Mpc}^3$  à  $z < 0.1$ . Le relevé COSMOS sonde donc des volumes équivalents dans l'Univers primordial.

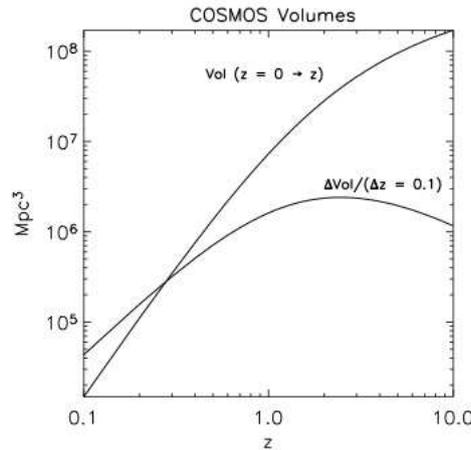


FIG. 10.2 – Volumes sondés par le relevé COSMOS entre les *redshifts* 0 et  $z$  et dans une « coquille » d'épaisseur  $\Delta z = 0.1$ . [Source : Scoville & COSMOS Team (2005)]

En combinant toutes les données disponibles, le relevé devrait donc pouvoir aborder certains des problèmes soulevés lors des chapitres d'introduction, à savoir :

- Distributions des masses et des luminosités des premières galaxies entre  $z = 3$  et  $z = 6$
- Evolution des noyaux actifs et leur rôle dans l'évolution des galaxies
- Reconstruction de la distribution de matière noire dans l'Univers jusqu'à  $z < 1.5$
- **Evolution morphologique des galaxies, des taux de fusion et de formation stellaire jusqu'à  $z = 6$**

## 10.2 Motivations du travail

Les résultats présentés dans ce chapitre sont centrés sur le dernier point et en particulier sur l'évolution de la morphologie des galaxies. En effet, les travaux précédents, utilisent des morphologies estimées dans le visible. Comme il a été exposé dans les chapitres précédents cela peut provoquer des biais à des *redshifts* élevés à cause des effets de k-correction. Grâce aux données multi-longueur d'onde disponibles, le champ COSMOS est donc un champ idéal pour mesurer ces effets. La morphologie visible est disponible grâce aux données spatiales. Des données au sol, obtenues avec l'instrument WIRCcam (Thibault *et al.*, 2003) au CFHT sur les mêmes objets dans

---

<sup>2</sup> $h=0.7$

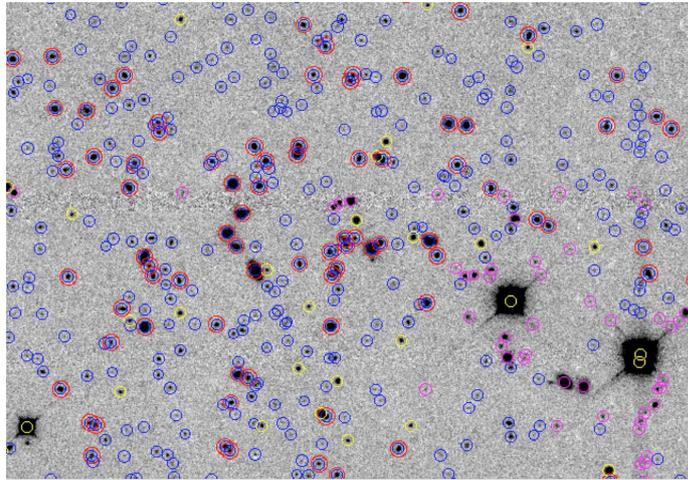


FIG. 10.3 – Région de  $3' \times 2'$  de l'image réduite. Les cercles jaunes indiquent les étoiles, les roses les objets masqués, les cercles bleus sont les galaxies et les cercles rouges les galaxies pour lesquelles la morphologie a été estimée (c.f. texte pour plus de détails.)

l'infrarouge proche sont disponibles quoiqu'avec une résolution angulaire nettement inférieure.

L'application de la méthode présentée dans le chapitre précédent, permet d'estimer, pour la première fois de façon quantitative, la morphologie de galaxies de champ au delà de  $z \sim 1$  dans l'infrarouge et de la comparer directement aux morphologies HST afin de quantifier les différences dues au changement de longueur d'onde.

## 10.3 Les données

### 10.3.1 Description

Les données analysées ont donc été observées au CFHT avec la caméra WIRCam en bande  $K_s$  ( $2.12\mu\text{m}$ ). Elles couvrent tout le champ COSMOS, i.e.  $1.4^{(\circ)^2}$  et correspondent également au champ D2 du relevé CFHTLS. La réduction des images a été réalisée avec la *pipeline* Térapix<sup>3</sup>. La taille du pixel sur l'image finale est de  $0.15''$  et la largeur à mi-hauteur moyenne des étoiles est de  $0.7''$ . La figure 10.3 montre une section de l'image finale. Une description plus détaillée des données ainsi que des techniques de réduction employées peut être trouvée dans [McCracken & The COSMOS Team \(2008\)](#).

<sup>3</sup><http://terapix.iap.fr>

### 10.3.2 Nettoyage

Tous les objets ayant un signal  $1.5\sigma$  aux dessus du niveau du ciel sur 3 pixels contigus ont été détectés en employant SExtractor (Bertin & Arnouts, 1996). Une opération de nettoyage afin d'éliminer les fausses détections a été réalisée par la suite en suivant une méthode analogue à celle décrite dans le chapitre 4, c'est-à-dire en utilisant le plan MU\_MAX-MAG\_AUTO. En plus de ce travail de nettoyage, certains objets ont dû être masqués car ils étaient sur des zones de l'image où le bruit est considérablement plus élevé que dans le reste. Les objets étant près d'étoiles brillantes ( $K_s < 15$ ) ont été également masqués.

Finalement 282 122 sources ont été retenues sur l'ensemble du champ.

### 10.3.3 Séparation étoiles/galaxies

Parmi les sources retenues dans l'étape de nettoyage, on trouve des étoiles ainsi que des galaxies. Afin de les séparer nous avons employé un certain nombre de paramètres photométriques :

- le paramètre CLASS\_STAR de SExtractor qui est une mesure de la compacité de l'objet. Plus ce paramètre est proche de 1, plus la probabilité pour que ce soit une étoile est élevée.
- le paramètre MU\_CLASS, c'est-à-dire la position dans le plan MU\_MAX-MAG\_AUTO. Comme il a été exposé dans le chapitre 4, les positions des objets dans le plan permettent également de séparer les étoiles des galaxies. La séparation est en générale correcte pour les objets les plus brillants et devient plus difficile lorsque la magnitude augmente.
- le type spectral résultant de l'ajustement de spectres lors de l'estimation des *redshifts* photométriques (c.f. annexe A).
- le *redshift* photométrique proprement dit. Sauf erreur dans l'estimation, les étoiles doivent avoir un *redshift* nul ou proche de zéro.

Nous avons donc combiné tous ces paramètres afin de séparer les étoiles des galaxies : Pour les objets plus brillants que  $K_s = 20$ , en plus du paramètre MU\_CLASS, nous avons considéré que ces objets étaient des étoiles lorsque  $CLASS\_STAR > 0.95$  avec un *redshift* photométrique plus petit que 0.05 ou lorsque le type spectral estimé correspondait à celui d'une étoile. Pour les objets plus faibles, le paramètre MU\_CLASS n'a pas été utilisé.

L'échantillon final contient 27 343 étoiles ou sources ponctuelles et 254 779 galaxies.

### 10.3.4 *Redshifts* photométriques

Les *redshifts* photométriques de 198 684 objets ont été estimés à l'aide du code *Le Phare* (Ilbert *et al.*, 2006), dans le cadre du consortium COSMOS, en employant un certain nombre de mesures photométriques issues des filtres disponibles dans

le champ COSMOS. Plus de détails peuvent être trouvés dans l'annexe A. La figure 10.4 illustre la distribution jusqu'à  $z = 2$ . Comme on peut l'apprécier, le pic de la distribution se situe autour de  $z \sim 1$ . Etant donné que l'objectif principal de ce travail est d'obtenir l'évolution des morphologies, l'estimation du *redshift* pour chaque objet est fondamentale. Nous avons donc restreint l'analyse morphologique aux objets possédant une estimation de *redshift*.

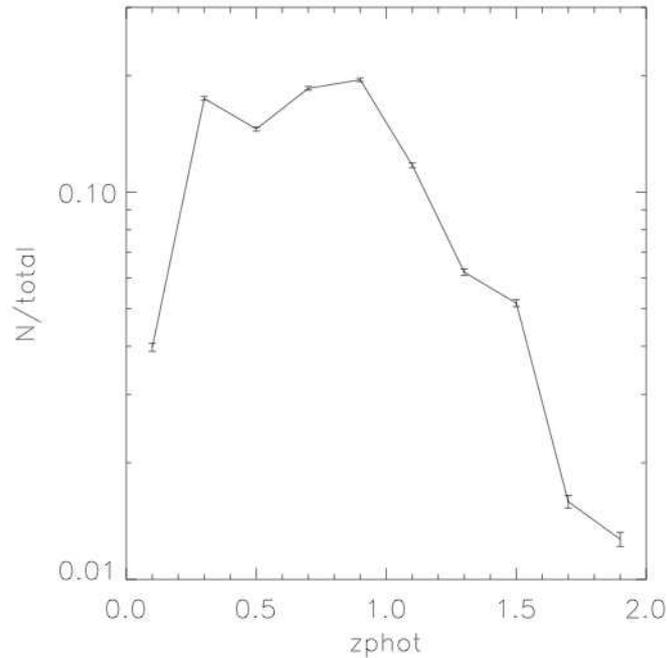


FIG. 10.4 – Distribution des *redshift* photométriques. Les barres d'erreurs sont poissoniennes.

### 10.3.5 Le catalogue morphologique final

En plus de la contrainte des *redshifts* photométriques nous avons rajouté deux autres contraintes pour sélectionner le catalogue final sur lequel réaliser l'étude morphologique :

- Seul les objets ayant une magnitude plus brillante que  $K_s = 21.5$  ont été retenus. Cette décision est basée sur une inspection visuelle ainsi que sur les simulations réalisées dans le chapitre précédent. En effet, les galaxies plus faibles présentent un rapport S/B bas ( $S/B \sim 5$ ) et sont trop petits ( $\sim 1 \text{ arcsec}^2$ ) pour estimer la morphologie de manière fiable.
- Nous avons également sélectionné uniquement les objets ayant un *redshift* plus petit que  $z = 2$ . En effet, avec les bandes photométriques employées, la fiabilité des *redshifts* photométriques décroît fortement au dessus de  $z = 2$ . Nous avons

donc décidé d'enlever ces objets de l'échantillon d'analyse.

Cette sélection résulte dans un catalogue final de 44 089 galaxies. La limite de complétude de l'échantillon étant significativement plus loin que la limite imposée (McCracken & The COSMOS Team, 2008), l'échantillon coupé à  $K_s < 21.5$  est complet, c'est-à-dire, tous les objets plus brillants que cette magnitude sont détectés.

Afin de quantifier cependant les effets de sélection plus précisément, nous avons simulé 1000 galaxies ayant des rapports bulbe/disque distribués uniformément entre 0 et 1 (c.f. chapitre 4), plongé ceux-ci dans un fond réel et essayé de les détecter avec SExtractor. Les résultats sont représentés dans la figure 10.5. On confirme qu'à  $K_s = 21.5$  l'échantillon est complet pour tous les types morphologiques.

Nous avons également analysé la complétude en fonction de la taille des galaxies. Nous avons pour cela simulé des disques (Fig. 10.6 a) et des bulbes pures (Fig. 10.6 b) ayant des tailles caractéristiques allant jusqu'à  $1.5''$ . On constate une dépendance de la complétude avec la taille, les galaxies plus étendues étant détectées avec plus de difficultés. Malgré tout, la complétude pour les bulbes est pratiquement de 100% pour toutes les tailles considérées jusqu'à  $K_s = 21.5$  et les disques sont complets à 90% à  $K_s = 21.5$  et  $r_d = 1.4''$ .

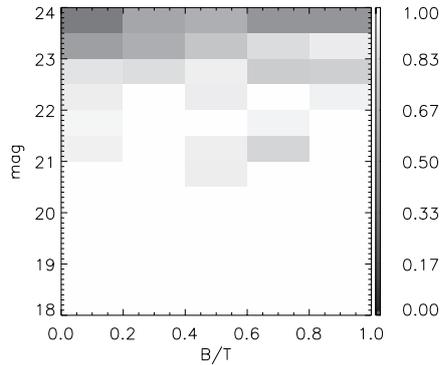


FIG. 10.5 – Complétude pour les sources étendues d'après les simulations analytiques de galaxies.

## 10.4 Analyse morphologique

L'estimation des morphologies (en deux types) a donc été réalisée avec le code GalSVM (Huertas-Company *et al.*, 2008) présenté dans le chapitre précédent. Le processus de classification se déroule en quatre étapes principales que l'on rappelle ici brièvement. Plus de détails peuvent être trouvés au chapitre 9.

1. Construction de l'échantillon d'entraînement à partir d'un échantillon local
2. Mesure des paramètres morphologiques

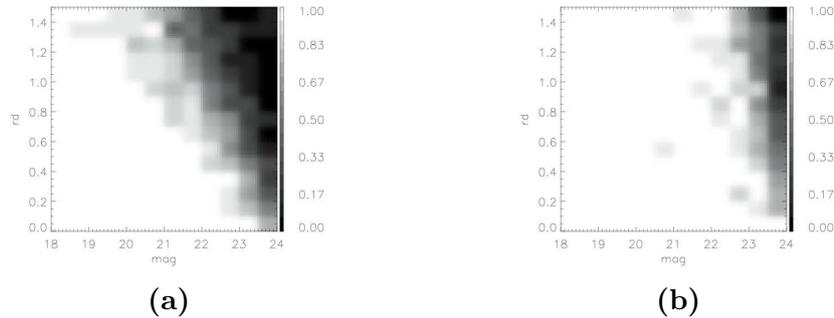


FIG. 10.6 – Complétude en fonction de la taille des galaxies disque (a) et bulbe (b) d’après les simulations de 5000 galaxies plongées dans les images réelles. La taille est représentée par l’échelle caractéristique des disques ( $r_d$ ) et le rayon effectif des bulbes ( $r_e$ ).

3. Entraînement d’une machine à vecteurs support
4. Classification de l’échantillon réel avec la machine entraînée

#### 10.4.1 L’échantillon d’entraînement

Dans le chapitre précédent, nous avons vu que l’adaptation de l’échantillon d’entraînement à l’échantillon réel est fondamentale pour obtenir une qualité de classification raisonnable. Nous avons vu en effet que celui-ci doit reproduire au mieux les conditions physiques et observationnelles. Le processus suivi ici pour générer l’échantillon est essentiellement le même que j’ai décrit précédemment. Nous avons donc employé le même échantillon local provenant du SDSS (1319 objets) mais au lieu d’employer une seule bande photométrique nous avons cette fois utilisé 2 bandes (z et i) en fonction du *redshift* associé à la galaxie : la bande la plus proche de la longueur d’onde au repos pour ce *redshift* donné est employée. Quatre cents cinquante objets ont été utilisés pour l’entraînement et les 869 restants pour les tests.

#### 10.4.2 Classification

Le rapport morphologique dans l’échantillon d’entraînement est fixé à 50/50, c’est-à-dire 50% de galaxies elliptiques et 50% de galaxies spirales. Comme il a été expliqué dans le chapitre précédent, ce mélange permet d’optimiser les performances sans se soucier des problèmes liés à des distributions non symétriques.

La classification finale est faite dans un espace à 10 paramètres : 6 paramètres morphologiques (asymétrie, concentration, gini,  $M_{20}$ , smoothness), un paramètre de distance (*redshift*), un paramètre de forme (ellipticité) et deux paramètres de luminosité (magnitude et brillance de surface).

### 10.4.3 Remarques sur l'influence du nombre de paramètres

Nous avons vu dans le chapitre précédent, que l'inclusion de paramètres additionnels ne provoque jamais une baisse de la qualité de la classification. En essayant de classer l'échantillon réel cependant, nous avons pu repérer un effet nouveau. Certains paramètres peuvent en effet introduire des biais s'ils ne sont pas représentatifs de la distribution réelle. C'est le cas en particulier de la magnitude. Il s'est avéré que l'inclusion de celle-ci peut provoquer un changement important des résultats obtenus. Alors que l'inclusion successive des autres paramètres semble faire converger les résultats, lorsque la magnitude est prise en compte la classification donne un fort excès de galaxies elliptiques faibles à grand *redshift*. Cependant, les tests de performances à partir de l'échantillon simulé donnent des performances meilleures lorsque la magnitude est intégrée. Cela veut-il dire que l'excès est réel? Afin de vérifier cela, nous avons regardé la distribution en magnitudes par type morphologique de l'échantillon classé avec 9 paramètres. On peut constater que les deux types de galaxies ont des distributions bien différentes surtout à des faibles magnitudes. Nous avons donc régénéré un nouvel échantillon d'entraînement mais cette fois en reproduisant ces différences dans la distribution des magnitudes. On s'aperçoit alors que la classification à 10 paramètres ne donne plus l'excès d'elliptiques. Cela semble indiquer par conséquent que si l'échantillon n'est pas représentatif de la réalité, la classification finale peut se voir affectée. Cet effet ne peut pas être repéré sur les simulations puisque, par construction, l'échantillon de test et d'entraînement sont identiques.

Etant donné que la simulation des tailles caractéristiques des objets peut ne pas être réaliste de même (nous avons juste reproduit les tailles de l'Univers local sans tenir compte de possibles effets évolutifs), nous avons décidé de ne pas inclure les 2 paramètres de taille qui ont été utilisés pour les simulations du chapitre précédent.

### 10.4.4 Catalogue de sortie et qualité des classifications

A chaque objet du catalogue final on associe donc un type morphologique (tardif ou précoce) ainsi qu'une probabilité d'appartenance à cette classe entre 0.5 et 1. Etant donné qu'il s'agit d'un problème à deux classes on a toujours  $p(\textit{galaxie} = \textit{elliptique}) = 1 - p(\textit{galaxie} = \textit{spirale})$ .

La qualité de la classification à l'aide de l'échantillon de test a été présentée en partie lors du chapitre précédent. Nous discutons ici un élément nouveau : la probabilité en tant qu'estimateur de fiabilité. Le tableau 10.1 montre l'évolution du taux de réussite en fonction de plusieurs seuils de probabilité. On s'aperçoit facilement que la valeur de la probabilité est fortement corrélée avec le nombre d'identifications correctes : le taux de réussite augmente nettement lorsque le seuil de probabilité augmente. En effet, si on ne garde que les galaxies ayant une probabilité entre 0.5 et 0.6 le taux de réussite n'est que de 58% alors que pour les objets ayant une probabilité

supérieure à  $p = 0.8$ , le classement est à 90% sûr. Les contaminations sont de l'ordre de 20% pour l'échantillon complet ( $p > 0.5$ ), en accord avec les résultats du chapitre précédent.

Ces relations entre la probabilité et le taux de réussite justifient l'emploi de la probabilité comme estimateur de la fiabilité des classifications dans l'échantillon réel. Nous avons, pour ce faire, représenté des cartes bi-dimensionnelles de la probabilité moyenne pour différents intervalles de magnitude, taille et *redshift*. Les résultats sont montrés dans la figure 10.7. L'étude de ces cartes permet d'extraire un certain nombre de conclusions pertinentes :

- Globalement, les galaxies tardives présentent des valeurs de probabilité plus élevées pour tous les intervalles de magnitude, taille ou *redshift* explorés. Cet effet peut probablement s'expliquer à cause de l'utilisation de l'ellipticité dans l'espace des paramètres. En effet, les galaxies présentant une forte ellipticité sont normalement classées comme tardives avec une forte probabilité, faisant ainsi augmenter la probabilité moyenne.
- La taille joue un rôle important comme attendu dans la détermination des deux types morphologiques. Les objets petits ( $r_{half} < 0.6''$ ) ont des probabilités de classement de l'ordre de  $p \sim 0.7$  alors que lorsque la taille augmente ( $r_{half} > 0.6''$ ) la probabilité croît au delà  $p \sim 0.8$  (Fig. 10.7 c,d,e,f).
- Il existe également une dépendance vis à vis du *redshift* de la galaxie, spécialement pour les galaxies précoces : au dessous de  $z \sim 1$  la probabilité est de  $p \sim 0.75$  avec un nombre important d'objets au dessus de  $p \sim 0.8$ . Les galaxies elliptiques à  $z > 1$  sont classées avec une probabilité moyenne de l'ordre de  $p \sim 0.7$  (Fig. 10.7 a,b,c,d).
- Finalement, le rôle de la magnitude est également important. Les galaxies plus faibles que  $K_s \sim 20$  la probabilité moyenne pour les galaxies elliptiques est de l'ordre de  $p \sim 0.7$  (Fig. 10.7 a,b,e,f). Étonnement cette tendance n'est pas présente pour les galaxies tardives. Cela est probablement un reflet de la manière de produire l'échantillon d'entraînement. En effet, comme il a été dit précédemment, les distributions de magnitude employées sont différentes en fonction du type morphologique concerné. Ainsi, aux faibles luminosités, les galaxies spirales sont dominantes et donc la machine a tendance à les classer avec une probabilité élevée.



Comment peut-on prendre en compte ces différentes tendances pour réaliser des études statistiques de morphologies? Intuitivement, le plus simple, semblerait de réaliser des seuils de probabilité et de ne garder ainsi que les objets ayant une probabilité minimale fixée en fonction de la fiabilité recherchée. Cependant, introduit-on des biais si on ne retient, pour une analyse scientifique, que les objets ayant une probabilité supérieure à un seuil donné? Afin de répondre à cette question, nous avons regardé la complétude de l'échantillon réel en fonction de la probabilité et de quatre paramètres : la magnitude, la taille, le *redshift* et le type morphologique (Fig. 10.8). Evidemment, plus le seuil de probabilité devient élevé, moins d'objets sont retenus. Cependant, ce qui est important à analyser est la tendance des courbes de complétude, c'est-à-dire, s'il n'y a pas de biais, on s'attend à ce que ces courbes soient « plates » en fonction de tous les paramètres : « tous les objets sont perdus de la même manière ». La figure 10.8 montre que ce n'est pas tout à fait le cas. En particulier il existe un biais important en fonction de la taille et du type morphologique ; les objets plus grands ont en effet, de façon générale, une probabilité plus élevée et sont donc retenus en priorité lorsque l'on applique un seuil de probabilité. Le biais plus important est cependant vis à vis des morphologies : comme on vient de le voir, les galaxies spirales ont des probabilités plus élevées et sont donc retenues en priorité. Un seuil en probabilité résulte donc en un échantillon biaisé envers des systèmes tardifs. On en déduit donc, que pour une analyse statistique globale de l'évolution des morphologies, la réalisation de coupes arbitraires en probabilité n'est pas adaptée. Des techniques pour l'estimation des erreurs doivent donc être développées pour ce type d'analyse comme nous le verrons par la suite. Des seuils en probabilité peuvent être cependant employés pour des applications ponctuelles où une très forte précision dans l'estimation morphologique est souhaitée.

## 10.5 Résultats et discussions

### 10.5.1 Estimations des statistiques globales

Peut-on utiliser l'information contenue dans les probabilités pour estimer plus précisément le nombre d'objets dans chaque classe ainsi que les erreurs associées? Considérons que nous avons  $N$  galaxies ayant une probabilité  $p_k$  d'appartenir à une classe A (elliptique ou spirale dans notre cas), nous avons défini deux fonctions pour estimer le nombre d'objets appartenant à cette classe donnée :

Le premier (*estimateur par comptages*), est l'approche la plus intuitive et consiste tout simplement à compter le nombre d'objets ayant une probabilité supérieure à 0.5. Cela peut s'exprimer mathématiquement en employant une fonction  $Z_k$  telle que :

$$Z_k = \begin{cases} 0 & \text{si } p_k < 0.5 \\ 1 & \text{si } p_k > 0.5 \end{cases}$$

Ainsi, le nombre d'objets dans la classe A est donné par  $N_A = \sum Z_k$ . Cet es-

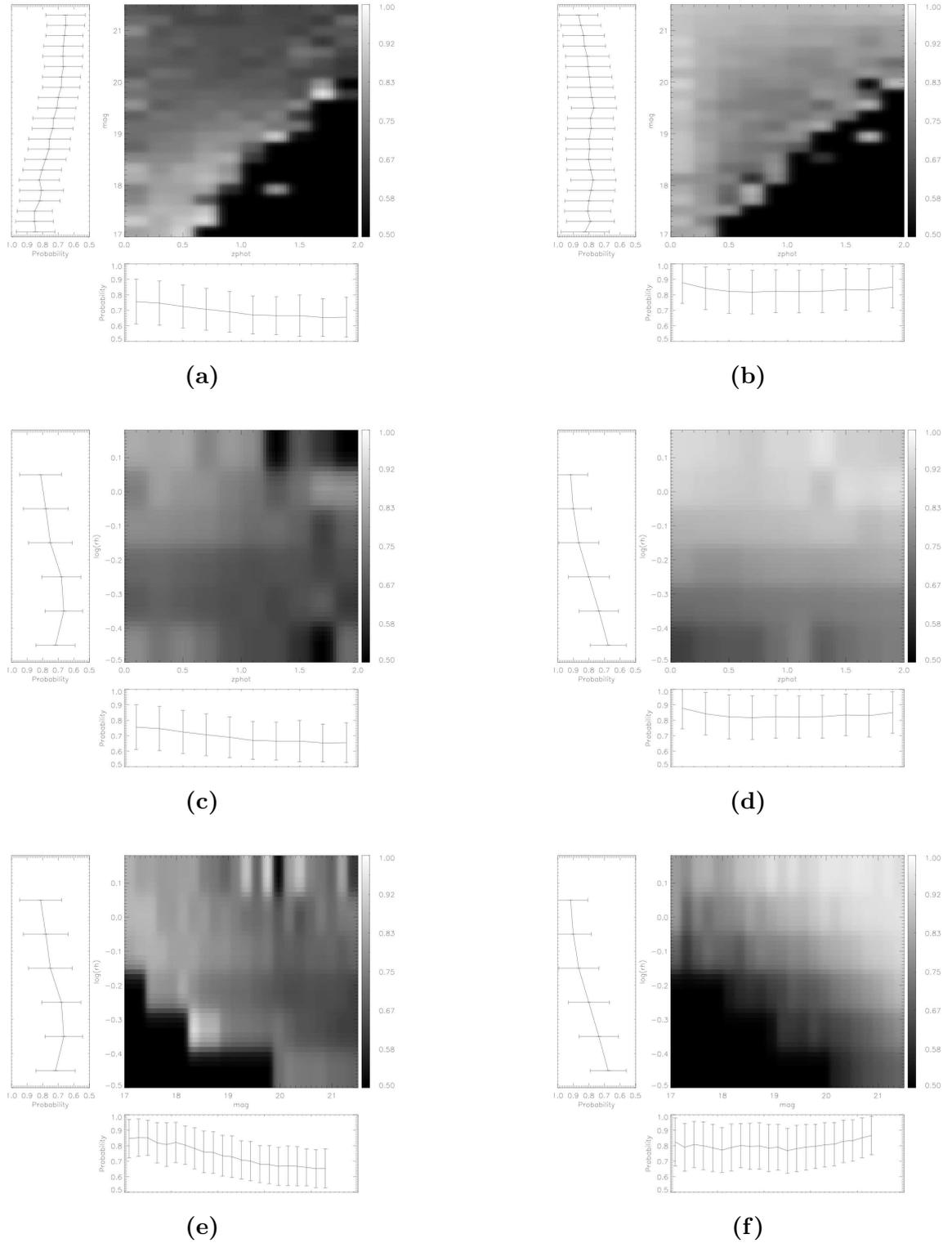


FIG. 10.7 – Cartes bi-dimensionnelles des probabilités moyennes dans l'échantillon réel en fonction de la magnitude, la taille et le redshift des galaxies. La taille est représentée par le rayon de moitié de luminosité ( $r_{half}$ ). Les courbes montrent les projections en fonction d'un seul paramètre. Les barres d'erreurs tracent la dispersion de la distribution de probabilités. Dans la colonne de gauche sont représentées les galaxies précoces et dans la colonne de droite les galaxies tardives.

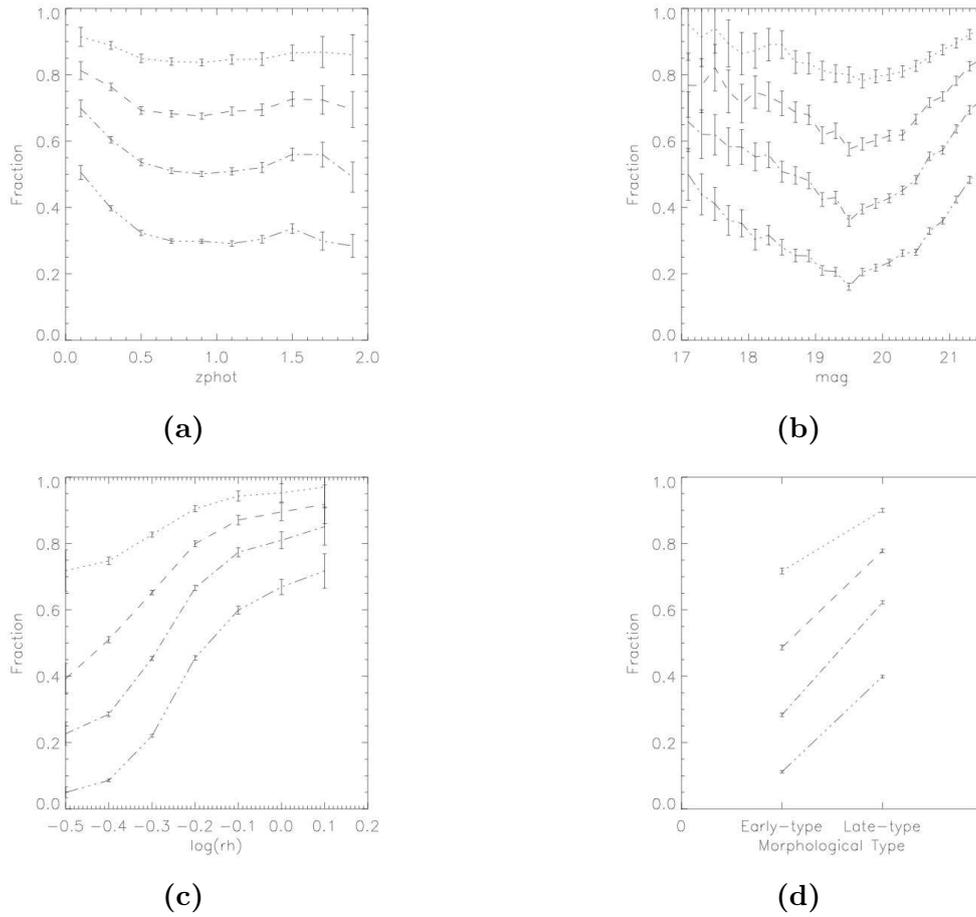


FIG. 10.8 – Complétude en fonction du *redshift* (a), de la magnitude (b), de la taille (c) et du type morphologique (d) pour différents seuils de probabilité. Ligne en pointillés :  $p > 0.6$ , ligne discontinue :  $p > 0.7$ , ligne discontinue avec un point :  $p > 0.8$ , ligne discontinue avec 3 points :  $p > 0.9$ . Les barres d'erreurs tracent la racine carrée du nombre d'objets.

timeur n'utilise donc pas l'information des probabilités. Une galaxie ayant une probabilité  $p = 0.51$  est traitée de la même manière qu'une autre ayant une probabilité  $p = 0.99$  : elle est tout simplement ajoutée à la classe correspondante. Dans le paragraphe précédent nous avons vu que la probabilité est cependant un bon indicateur de la qualité de la classification. Est-il possible de l'intégrer dans les comptages ? Nous avons pour cela défini la variable aléatoire suivante (*l'estimateur par probabilité*) :

$$Y_k = \begin{cases} 0 & \text{avec une probabilité } 1 - p_k \\ 1 & \text{avec une probabilité } p_k \end{cases}$$

Il devient possible alors d'estimer le nombre d'objets dans la classe A à partir de l'espérance mathématique de cette variable :  $E(Y) = \sum E(Y_k) = \sum p_k$ . De cette façon, une galaxie ayant une probabilité  $p = 0.51$  compte comme 0.51 dans la classe A mais aussi comme 0.49 dans la classe B. L'information probabilistique est ainsi intégrée. De plus, la définition de la variable aléatoire  $Y_k$  permet également d'estimer les erreurs à  $1-\sigma$  en calculant la racine carrée de la variance :  $Var(Y) = E(X^2) - E(X)^2 = \sum p_k(1 - p_k)$ .

Plus le nombre d'objets ayant une probabilité proche de 0.5 est important, plus les différences entre les deux estimateurs deviennent grandes. En effet, dans le cas idéal où tous les objets auraient une probabilité de 1, les deux estimateurs donneraient la même valeur. Au contraire, si tous les objets avaient une probabilité de 0.5, *l'estimateur par probabilité* donneraient la moitié d'objets que *l'estimateur par comptages*. La comparaison des deux méthodes est donc également une manière de mesurer la qualité de la classification à partir de l'échantillon réel dans le sens où s'il existe beaucoup de différences entre les deux c'est un signe que les objets sont classés avec une faible probabilité.

Nous avons donc utilisé ces deux estimateurs pour obtenir le nombre global de galaxies de chaque type morphologique dans l'échantillon de 44 089 objets. *L'estimateur par comptages* donne :

- 33 291 galaxies de type tardif soit  $\sim 75\%$  de l'échantillon.
- 10 798 galaxies de type précoce soit  $\sim 25\%$  de l'échantillon.

D'un autre côté, *l'estimateur par probabilités* donne :

- $30\,711.7 \pm 80$  galaxies de type tardif soit  $\sim 70\%$  de l'échantillon.
- $13\,376.3 \pm 80$  galaxies de type précoce soit  $\sim 30\%$  de l'échantillon.

On constate ainsi des différences entre les deux estimateurs comme prévu. Les tendances sont cependant les mêmes et globalement les deux estimations sont en bon accord indiquant que la classification est solide. Nous pouvons alors nous demander : où ces différences se concentrent-elles ? Sont-elles uniformément réparties pour toutes les galaxies ?

### 10.5.2 Comptages par type morphologique

La figure 10.9 montre les comptages de galaxies par type morphologique. Comme attendu, les galaxies elliptiques dominent parmi les galaxies plus brillantes alors que les galaxies spirales dominent à des luminosités plus faibles. L'effet de la prise en compte des probabilités est surtout appréciable dans la population d'elliptiques faibles ( $K_s > 20$ ). En particulier, le nombre de galaxies elliptiques faibles augmente avec *l'estimateur par probabilité*. Cela reflète qu'aux faibles luminosités, les galaxies elliptiques sont classées avec des probabilités plus faibles comme nous avons pu le voir dans les cartes de probabilité (Fig. 10.7). L'utilisation des probabilités permet ainsi de corriger les comptages en ajoutant des objets qui tombent du côté spiral mais présentant une probabilité non négligeable d'être elliptique.

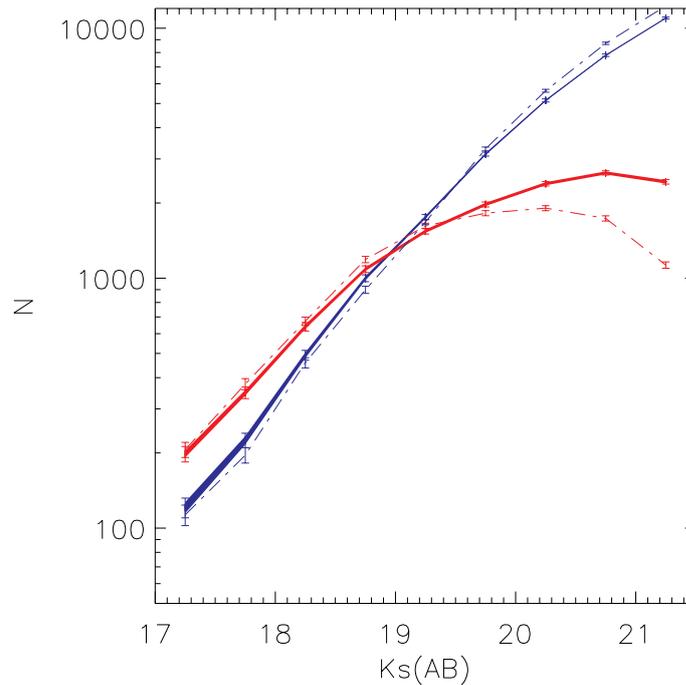


FIG. 10.9 – Comptages de galaxies par type morphologique pour les 44 089 galaxies analysées. Les courbes continues représentent les résultats obtenus avec *l'estimateur par probabilités* et les courbes en pointillés ceux estimés à partir de *l'estimateur par comptages*. L'épaisseur des courbes représente la dispersion calculée à partir de la distribution de probabilités. Les courbes rouges correspondent aux galaxies précoces et les courbes bleues aux galaxies tardives. Les barres d'erreurs montrent la racine carrée du nombre d'objets.

### 10.5.3 Evolution morphologique

La figure 10.10 illustre l'évolution du mélange morphologique depuis  $z \sim 2$ . Les deux estimateurs détectent une augmentation de la fraction de galaxies elliptiques, cependant l'effet est moins important lorsque l'information probabiliste est exploitée. En effet, avec *l'estimateur par probabilité*, nous comptons 20% de galaxies précoces à  $z \sim 2$  pour  $\sim 30\%$  dans l'Univers local. *L'estimateur par comptages* prédit une fraction de 15% à  $z \sim 1.5$ . L'inclusion de la probabilité semble donc permettre de corriger de l'incomplétude des galaxies elliptiques à  $z \sim 1.5$ .

Cette variation de la population de galaxies précoces est un effet relativement bien connu qui a pu être mis en évidence à travers l'analyse de morphologies grâce à des données du télescope spatial (e.g. Brinchmann *et al.*, 1998; Cassata *et al.*, 2005). On aurait pu penser cependant que cet effet pouvait être biaisé à cause justement des effets de k-correction qui font que, à grand *redshift*, on sur-estime le nombre de galaxies de type tardif en sondant l'émission UV des galaxies. Le fait que l'on continue à détecter cette tendance à partir de données dans l'infrarouge proche qui sondent donc des populations stellaires plus vieilles semble indiquer qu'il s'agit d'un effet réel et non pas d'une conséquence du changement de bande. Cependant, la fraction trouvée ici à  $z \sim 1.5$  est supérieure à celle mesurée par Cassata *et al.* (2005). Ces différences peuvent-elles être conséquence du changement de bande ?

C'est un point crucial de cette étude et la raison principale pour laquelle on réalise des analyses morphologiques dans l'infrarouge proche. Nous prêtons donc une attention particulière dans les sections qui suivent.

## 10.6 Autour des effets de k-correction morphologique

Dans les paragraphes précédents nous avons obtenu une classification morphologique mesurée dans l'infrarouge proche. L'intégration de la probabilité dans les comptages statistiques permet de corriger, dans la mesure du possible, des éventuelles erreurs de classification. Un point fondamental est maintenant de comprendre les différences existantes avec des morphologies estimées dans le visible. Pour cela, nous avons comparé le catalogue morphologique obtenu ci-dessus avec un autre catalogue obtenu d'une manière indépendante sur les mêmes objets observés avec le télescope spatial en bande I (caméra ACS). La méthode employée pour classer l'échantillon ACS est décrite dans Tasca & The COSMOS Team (2008). Le lecteur intéressé pourra consulter les détails dans la référence indiquée. Parmi les 44 089 objets de notre catalogue, 34 165 ont une contrepartie dans le catalogue HST. Les objets restants étaient trop faibles sur les images HST pour estimer une morphologie fiable.

Les proportions de chaque type morphologique sont globalement en bon accord :

- 10 458.8 galaxies précoces et 23 706.2 galaxies tardives avec WIRCam
- 9 163 galaxies précoces et 25 002 galaxies tardives avec ACS

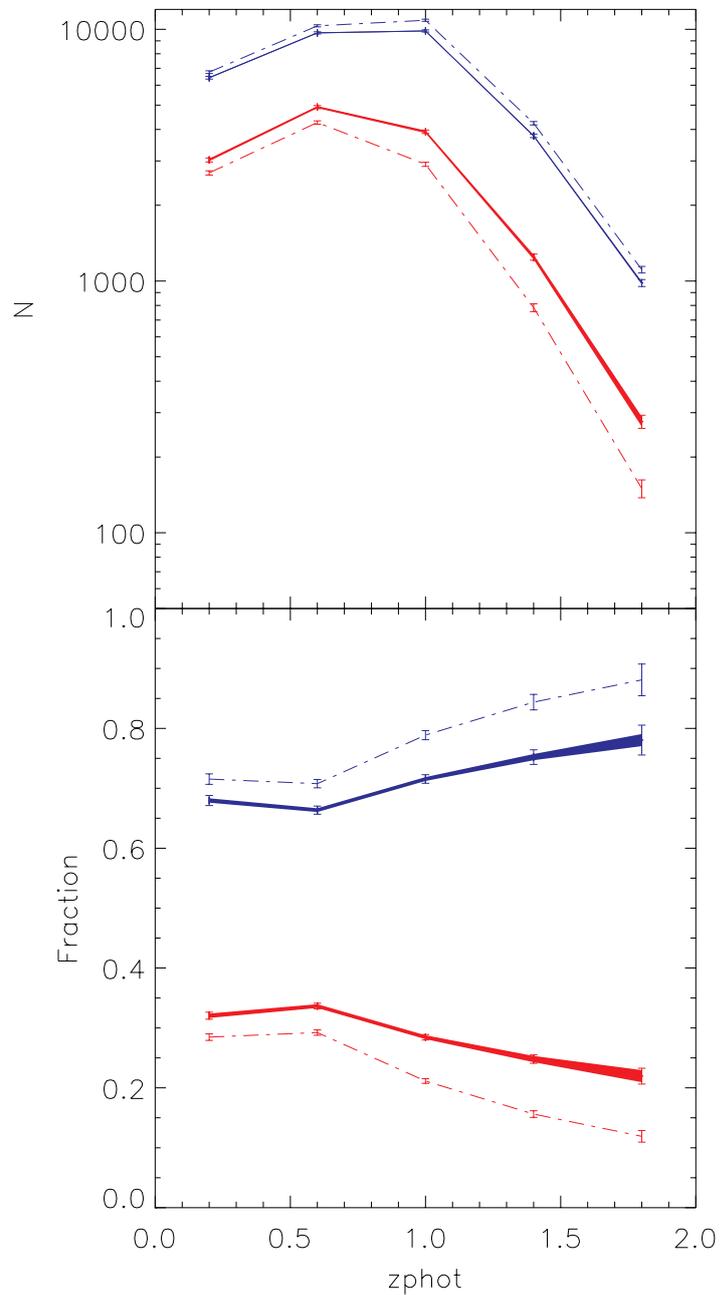


FIG. 10.10 – Distribution des morphologies en fonction du redshift pour les 44 089 galaxies analysées. Les courbes continues représentent les résultats obtenus avec *l'estimateur par probabilités* et les courbes en pointillés ceux estimés à partir de *l'estimateur par comptages*. L'épaisseur des courbes représente la dispersion calculée à partir de la distribution de probabilités. Les courbes rouges correspondent aux galaxies précoces et les courbes bleues aux galaxies tardives. Les barres d'erreurs montrent la racine carrée du nombre d'objets.

On s'intéresse par la suite à la localisation précise de ces différences en fonction du *redshift*.

### 10.6.1 Comparaison galaxie par galaxie

La figure 10.11 montre les probabilités pour qu'une galaxie classée précoce (tardive) dans la bande I soit classée précoce (tardive) également dans la bande Ks en fonction du *redshift*.

Nous constatons que globalement l'identification des galaxies tardives ne présente pas trop d'ambiguïté : une galaxie tardive dans la bande I est en effet tardive dans la bande Ks avec une probabilité moyenne de 0.9 jusqu'à  $z \sim 2$ .

Pour les galaxies précoces cependant, la correspondance dépend du *redshift*. En dessous de  $z \sim 1$  une galaxie précoce dans la bande I l'est aussi dans la bande K avec une probabilité moyenne  $p \sim 0.7$ . A  $z > 1$ , les différences entre les deux classements deviennent plus importantes et la probabilité diminue jusqu'à  $p \sim 0.4$  à  $z \sim 1.5$ . A ces *redshifts*, le filtre I commence à sonder le flux UV des galaxies et par conséquent la morphologie obtenue est essentiellement déterminée par la lumière provenant de populations stellaires jeunes. L'effet observé est donc probablement une k-correction morphologique : les galaxies elliptiques ont tendance à se déplacer vers des types morphologiques plus tardifs lorsqu'elles sont observées dans la bande I.

Quelles sont alors les différences entre une sélection morphologique dans la bande I et dans la bande Ks ?

### 10.6.2 Evolutions morphologiques comparées

La figure 10.12 montre l'évolution des types morphologiques en fonction de  $z$  estimés dans les deux filtres analysés. Les distributions des galaxies tardives présentent un bon accord comme attendu d'après l'analyse du paragraphe précédent. Un test de Kolmogorov-Smirnoff (*KS-test*) révèle en effet que les deux fonctions sont issues de la même statistique avec une probabilité de 96%. Les galaxies elliptiques présentent plus de différences en particulier à des *redshifts* élevés. L'accord entre les deux distributions est fort à  $z < 1$  (97%) mais décroît nettement au dessus de  $z \sim 1$  pour se situer à 67%. Nous constatons en particulier un excès de galaxies elliptiques d'un facteur  $\sim 1.5$  dans la bande Ks. Malgré le fait que, d'après les simulations effectuées, à ces *redshifts* la classification infrarouge est moins fiable, la différence est significative et persiste après correction des erreurs. Il semble donc que la k-correction morphologique aurait pour effet une sous estimation du nombre de galaxies elliptiques à grand  $z$  lorsque les mesures sont effectuées à partir de la bande I du télescope spatial.

Ces résultats doivent être cependant confirmés par une analyse détaillée de la population de galaxies elliptiques à  $z \sim 1.5$  en étudiant, entre autres, le taux de

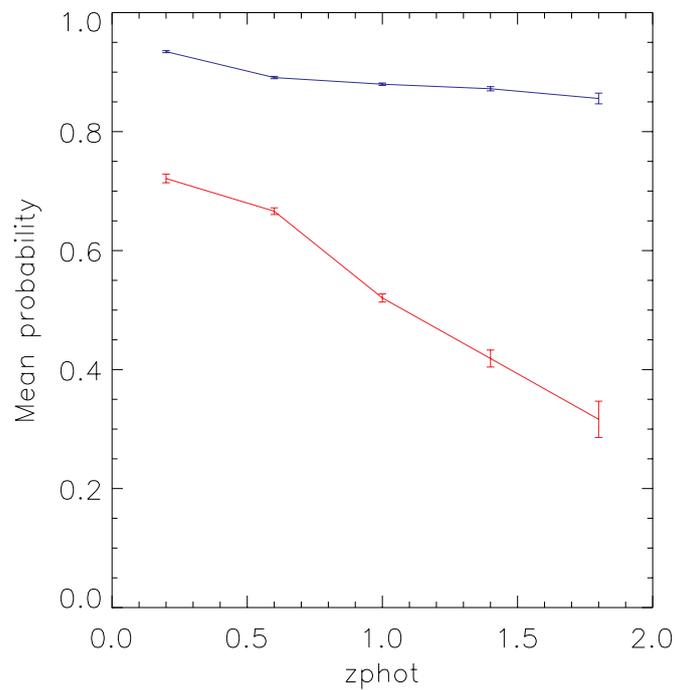


FIG. 10.11 – Comparaison galaxie par galaxie. Courbe rouge : Probabilité moyenne pour qu'une galaxie classée précoce dans la bande I soit classée pareil dans la bande Ks. Courbe bleue : Probabilité moyenne pour qu'une galaxie classée tardive dans la bande I soit classée pareil dans la bande Ks. Les barres d'erreurs donnent la dispersion de la distribution de probabilités.

formation stellaire et leurs masses. Ce travail est en cours de développement lors de la rédaction de ce manuscrit.

## 10.7 Conclusions et perspectives

Nous avons présenté l'analyse morphologique dans l'infrarouge proche de 44 089 galaxies dans le champ COSMOS. Les morphologies ont été estimées avec le code non-paramétrique présenté dans le chapitre précédent dans un volume à 10 dimensions avec des frontières non-linéaires.

Le catalogue final constitué offre pour chaque galaxie, un type morphologique et une probabilité d'appartenance à la classe donnée. Nous avons démontré que la probabilité est fortement corrélée au taux de réussite et représente donc un bon estimateur de la qualité des classifications.

La classification obtenue a été employée pour réaliser des comptages par type morphologique et étudier l'évolution de morphologies en fonction du *redshift* depuis  $z \sim 2$ . Les résultats ont été par la suite comparés à ceux obtenues à partir de mesures réalisées avec des données du télescope spatial dans la bande I afin de quantifier les effets de k-correction morphologique.

Les conclusions principales de cette étude sont exposées ci-dessous :

En ce qui concerne la fiabilité de la classification obtenue :

- (i) D'après l'échantillon de test simulé, le taux de réussite moyen est de l'ordre de  $\sim 80\%$  pour l'échantillon complet. Les contaminations sont par conséquent de l'ordre de  $\sim 20\%$  pour les deux types morphologiques.
- (ii) La probabilité résultante de la classification s'avère être un bon estimateur de la qualité de la classification à cause de sa forte corrélation avec le taux de réussite. Les objets ayant une probabilité supérieure à 0.8 sont en effet identifiés avec 90% de fiabilité.
- (iii) L'étude donc de la distribution des probabilités en fonction de la magnitude, la taille et le *redshift* des galaxies permet de conclure que les objets identifiés avec plus de difficultés sont les galaxies elliptiques faibles ( $K_s > 20$ ), petites ( $\log(r_{half}) < -0.2$ ) situées à  $z > 1$ . Cependant, même pour cette classe d'objets, la probabilité moyenne est autour de  $\sim 0.7$ .
- (iv) Nous avons montré de même que réaliser des seuils de probabilité afin d'essayer d'améliorer la fiabilité de l'échantillon résulte en un échantillon biaisé vers les types morphologiques tardifs. Nous avons cependant proposé un moyen de prendre en compte l'information probabiliste dans les analyses statistiques.

En ce qui concerne la distribution morphologique de l'échantillon étudié :

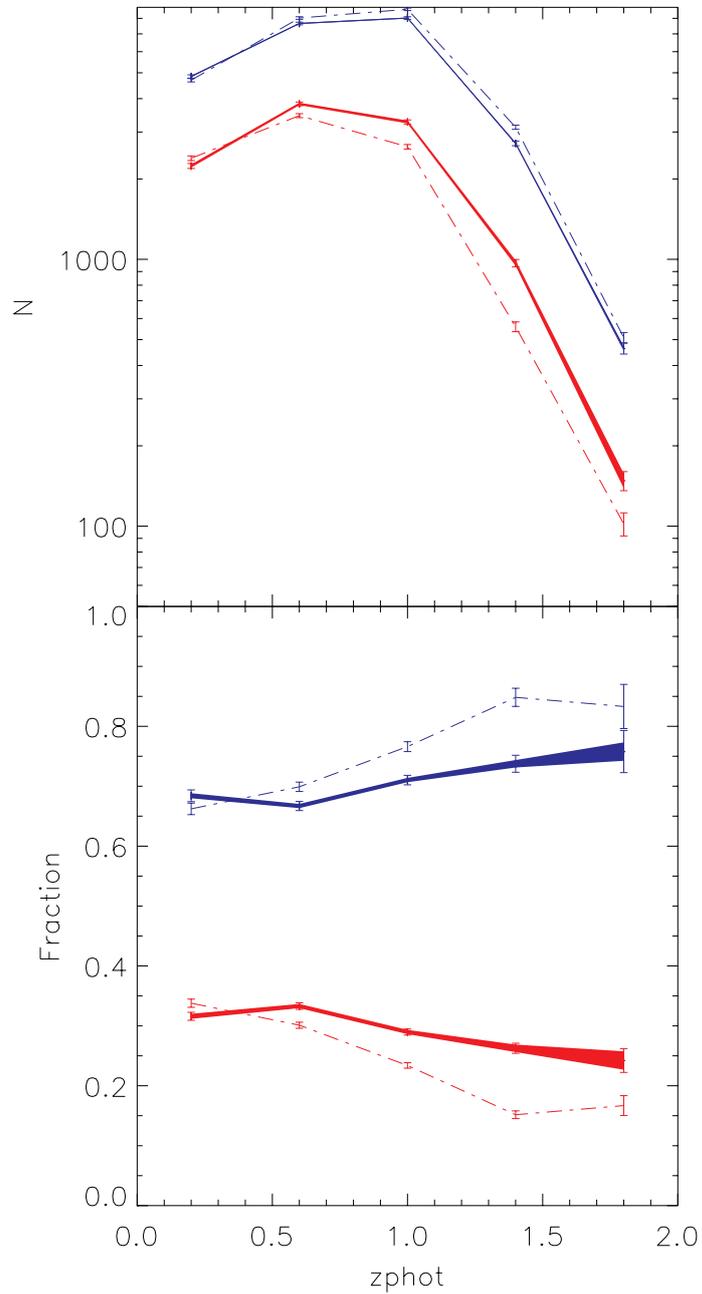


FIG. 10.12 – Distribution des morphologies en fonction du redshift pour les 34 165 galaxies analysées observées dans la bande Ks (courbe continue) et dans la bande I (courbe en pointillés). L'épaisseur des courbes représente la dispersion calculée à partir de la distribution de probabilités. Les courbes rouges correspondent aux galaxies précoces et les courbes bleues aux galaxies tardives. Les barres d'erreurs montrent la racine carrée du nombre d'objets.

- (i) La distribution morphologique globale est de  $30\,711.7 \pm 80$  ( $\sim 70\%$ ) de galaxies tardives et  $13\,376.3 \pm 80$  ( $\sim 30\%$ ) de galaxies précoces. Ces résultats prennent en compte les erreurs de classification à l'aide du paramètre de probabilité. La population de galaxies est donc dominée par les galaxies tardives depuis  $z \sim 2$ .
- (ii) Nos observons une augmentation du nombre de galaxies elliptiques depuis  $z \sim 2$ . Il y a, dans notre échantillon,  $\sim 30\%$  de galaxies elliptiques à  $z \sim 0.5$  alors que la fraction à  $z \sim 1.5$  est de  $\sim 20\%$ . Cet effet persiste même après l'application des corrections pour la possible perte de galaxies elliptiques à grand  $z$  et permet ainsi de confirmer un résultat obtenu dans des travaux précédents avec des données du télescope spatial. Ce résultat est probablement un reflet de la construction progressive de la séquence rouge à partir de galaxies tardives.

Finalement, la comparaison des morphologies obtenues avec celles estimées sur les mêmes objets observés avec HST dans la bande I, révèle un certain nombre de résultats intéressants :

- (i) Les distributions morphologiques globales sont similaires. Nous comptons 25 002 galaxies tardives et 9 163 galaxies précoces dans la bande I et 23 706 galaxies tardives pour 10 458 galaxies précoces dans la bande Ks.
- (ii) Une galaxie classée tardive dans la bande I a une probabilité moyenne de  $p \sim 0.9$  d'être classée tardive également dans la bande Ks pour tout le domaine de *redshift* analysé.
- (iii) La correspondance pour les galaxies elliptiques dépend du *redshift* : en dessous de  $z \sim 1$ , une galaxie classée elliptique dans la bande I, est classée elliptique dans la bande Ks avec une probabilité de  $p \sim 0.7$ . Au dessus de  $z \sim 1$ , lorsque le filtre I commence à sonder le spectre UV de la galaxie, la probabilité décroît jusqu'à  $p \sim 0.4$  à  $z \sim 1.5$ .
- (iv) La comparaison des distributions morphologiques en fonction du *redshift* révèle également une dépendance vis à vis de  $z$  : en dessous de  $z \sim 1$ , les deux distributions sont similaires. Le test KS estime à 97% la probabilité pour que les deux distributions soient issues de la même statistique. Au dessus de  $z \sim 1$ , la classification dans la bande I trouve un nombre de galaxies elliptiques inférieur d'un facteur  $\sim 1.5$  au nombre estimé dans la bande K. Cela est probablement un reflet des effets de k-correction morphologique. Les études basées sur des observations du télescope spatial dans la bande I à ces *redshifts* risquent par conséquent de sous estimer la fraction de galaxies précoces.

Les résultats exposés ci-dessus ouvrent un grand nombre de possibilités pour poursuivre l'analyse. Ces tendances doivent être en effet vérifiées par une étude approfondie de la population de galaxies elliptiques à  $z \sim 1.5$ . Les données multi-longueur d'onde du relevé COSMOS permettent d'étudier leurs masses et leurs histoires de

formation stellaire. Ce travail est en cours de développement au moment de la rédaction de ce manuscrit. De plus, la classification présentée peut être également utilisée pour analyser des indicateurs statistiques comparables aux modèles d'évolution comme les fonctions de luminosité par type, l'évolution des tailles de chaque type morphologique ou la relation entre la morphologie et la densité en fonction du *redshift*.

## Bibliographie

- Baker A.J., Davies R.I., Lehnert M.D., Thatte N.A., Vacca W.D., Hainaut O.R., Jarvis M.J., Miley G.K., & Röttgering H.J.A., 2003. Galaxies in southern bright star fields. I. Near-infrared imaging. *A&A*, 406 : 593.
- Bertin E. & Arnouts S., 1996. SExtractor : Software for source extraction. *A&A*, 117 : 393. 179
- Brinchmann J., Abraham R., Schade D., Tresse L., Ellis R.S., Lilly S., Le Fevre O., Glazebrook K., Hammer F., Colless M., Crampton D., & Broadhurst T., 1998. Hubble Space Telescope Imaging of the CFRS and LDSS Redshift Surveys. I. Morphological Properties. *ApJ*, 499 : 112. 191
- Cassata P., Cimatti A., Franceschini A., Daddi E., Pignatelli E., Fasano G., Rodighiero G., Pozzetti L., Mignoli M., & Renzini A., 2005. The evolution of the galaxy B-band rest-frame morphology to  $z \sim 2$  : new clues from the K20/GOODS sample. *MNRAS*, 357 : 903. 191
- Cresci G., Davies R.I., Baker A.J., Mannucci F., Lehnert M.D., Totani T., & Minowa Y., 2006. Galaxy morphology and evolution from SWAN adaptive optics imaging. *Astronomy and Astrophysics*, 458 : 385.
- Hasinger G., Cappelluti N., Brunner H., Brusa M., Comastri A., Elvis M., Finoguenov A., Fiore F., Franceschini A., Gilli R., Griffiths R.E., Lehmann I., Mainieri V., Matt G., Matute I., Miyaji T., Molendi S., Paltani S., Sanders D.B., Scoville N., Tresse L., Urry C.M., Vettolani P., & Zamorani G., 2007. The XMM-Newton Wide-Field Survey in the COSMOS Field. I. Survey Description. *ApJS*, 172 : 29. 176
- Huertas-Company M., Rouan D., Soucail G., Le Fèvre O., Tasca L., & Contini T., 2007. Morphological evolution of  $z \sim 1$  galaxies from deep K-band AO imaging in the COSMOS deep field. *A&A*, 468 : 937.
- Huertas-Company M., Rouan D., Tasca L., Soucail G., & Le Fèvre O., 2008. A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images. I. Method description. *A&A*, 478 : 971. 181

- Ilbert O., Lauger S., Tresse L., Buat V., Arnouts S., Fèvre O.L., Burgarella D., Zucca E., Bardelli S., Zamorani G., Bottini D., Garilli B., Brun V.L., Maccagni D., Picat J.P., Scaramella R., Scodreggio M., Vettolani G., Zanichelli A., Adami C., Arnaboldi M., Bolzonella M., Cappi A., Charlot S., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Iovino A., McCracken H.J., Marano B., Marinoni C., Mathez G., Mazure A., Meneux B., Merighi R., Paltani S., Pello R., Pollo A., Pozzetti L., Radovich M., Bondi M., Bongiorno A., Busarello G., Ciliegi P., Mellier Y., Merluzzi P., Ripepi V., & Rizzo D., 2006. The VIMOS-VLT Deep Survey. Galaxy luminosity function per morphological type up to  $z = 1.2$ . *Astronomy and Astrophysics*, 453 : 809. 179
- Impey C.D., Trump J.R., McCarthy P.J., Elvis M., Huchra J.P., Scoville N.Z., Lilly S.J., Brusa M., Hasinger G., Schinnerer E., Capak P., & Gabor J., 2007. A survey of AGN and supermassive black holes in the COSMOS Survey. Dans *IAU Symposium*, édité par V. Karas & G. Matt, volume 238 de *IAU Symposium*, p. 287–290. 176
- Lilly S.J., Le Fèvre O., Renzini A., Zamorani G., Scodreggio M., Contini T., Carollo C.M., Hasinger G., Kneib J.P., Iovino A., Le Brun V., Maier C., Mainieri V., Mignoli M., Silverman J., Tasca L.A.M., Bolzonella M., Bongiorno A., Bottini D., Capak P., Caputi K., Cimatti A., Cucciati O., Daddi E., Feldmann R., Franzetti P., Garilli B., Guzzo L., Ilbert O., Kampczyk P., Kovac K., Lamareille F., Leauthaud A., Borgne J.F.L., McCracken H.J., Marinoni C., Pello R., Ricciardelli E., Scarlata C., Vergani D., Sanders D.B., Schinnerer E., Scoville N., Taniguchi Y., Arnouts S., Aussel H., Bardelli S., Brusa M., Cappi A., Ciliegi P., Finoguenov A., Foucaud S., Franceschini R., Halliday C., Impey C., Knobel C., Koekemoer A., Kurk J., Maccagni D., Maddox S., Marano B., Marconi G., Meneux B., Mobasher B., Moreau C., Peacock J.A., Porciani C., Pozzetti L., Scaramella R., Schiminovich D., Shopbell P., Smail I., Thompson D., Tresse L., Vettolani G., Zanichelli A., & Zucca E., 2007. zCOSMOS : A Large VLT/VIMOS Redshift Survey Covering  $0 < z < 3$  in the COSMOS Field. *ApjS*, 172 : 70. 176
- McCracken H.J., Metcalfe N., Shanks T., Campos A., Gardner J.P., & Fong R., 2000. Galaxy number counts - IV. Surveying the Herschel Deep Field in the near-infrared. *MNRAS*, 311 : 707.
- McCracken H.J. & The COSMOS Team, 2008. in preparation. 178, 181
- Mobasher B., Capak P., Scoville N.Z., Dahlen T., Salvato M., Aussel H., Thompson D.J., Feldmann R., Tasca L., Lefevre O., Lilly S., Carollo C.M., Kartaltepe J.S., McCracken H., Mould J., Renzini A., Sanders D.B., Shopbell P.L., Taniguchi Y., Ajiki M., Shioya Y., Contini T., Giavalisco M., Ilbert O., Iovino A., Le Brun V., Mainieri V., Mignoli M., & Scodreggio M., 2007. Photometric Redshifts of Galaxies in COSMOS. *ApjS*, 172 : 117. 176
- Sanders D.B., Salvato M., Aussel H., Ilbert O., Scoville N., Surace J.A., Frayer D.T., Sheth K., Helou G., Brooke T., Bhattacharya B., Yan L., Kartaltepe J.S., Barnes

- J.E., Blain A.W., Calzetti D., Capak P., Carilli C., Carollo C.M., Comastri A., Daddi E., Ellis R.S., Elvis M., Fall S.M., Franceschini A., Giavalisco M., Hasinger G., Impey C., Koekemoer A., Le Fèvre O., Lilly S., Liu M.C., McCracken H.J., Mobasher B., Renzini A., Rich M., Schinnerer E., Shopbell P.L., Taniguchi Y., Thompson D.J., Urry C.M., & Williams J.P., 2007. S-COSMOS : The Spitzer Legacy Survey of the Hubble Space Telescope ACS  $2deg^2$  COSMOS Field I : Survey Strategy and First Analysis. *ApJS*, 172 : 86. 176
- Scoville N.Z. & COSMOS Team, 2005. COSMOS. Dans *Bulletin of the American Astronomical Society*, volume 37 de *Bulletin of the American Astronomical Society*, p. 1309–+. 175, 177
- Tasca L. & The COSMOS Team, 2008. in preparation. 191
- Thibault S., Cui Q., Poirier M., Vallee P., Doyon R., Rabou P., & Salmon D.A., 2003. Optical design of WIRCAM, the CFHT wide-field infrared camera. Dans *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. Edited by Iye, Masanori ; Moorwood, Alan F. M. Proceedings of the SPIE, Volume 4841, pp. 932-943 (2003).*, édité par M. Iye & A.F.M. Moorwood, volume 4841 de *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, p. 932–943. 177

# Travaux en cours : Autres applications de




---

## Sommaire

<b>11.1 Morphologies ALHAMBRA</b> . . . . .	<b>203</b>
11.1.1 Le relevé ALHAMBRA . . . . .	203
11.1.2 Objectifs scientifiques de l'étude . . . . .	203
11.1.3 Les données analysées . . . . .	204
11.1.4 Catalogue photométrique et <i>redshifts</i> photométriques . . . . .	204
11.1.5 Sélection de l'échantillon morphologique . . . . .	206
11.1.6 Analyse morphologique . . . . .	207
11.1.6.1 Principe . . . . .	207
11.1.6.2 Echantillon d'entraînement . . . . .	207
11.1.6.3 Analyse des erreurs . . . . .	207
11.1.6.4 Quelques résultats préliminaires . . . . .	209
11.1.7 Bilan et perspectives . . . . .	211
<b>11.2 Morphologies dans des amas à <i>redshift</i> intermédiaire (<math>z \sim 0.5</math>)</b>	<b>213</b>
11.2.1 Motivations du travail réalisé . . . . .	213
11.2.2 Description des données . . . . .	214
11.2.2.1 Contexte scientifique . . . . .	214
11.2.2.2 MS 1621.5+2640 . . . . .	214
11.2.3 Sélection des galaxies appartenant à l'amas . . . . .	215
11.2.4 Analyse morphologique . . . . .	215
11.2.5 Inspection visuelle . . . . .	216
11.2.6 Relation <i>couleur-magnitude</i> . . . . .	216
11.2.7 Relation <i>morphologie-distance radiale</i> . . . . .	219
11.2.8 Conclusions et travaux en cours . . . . .	219
<b>Bibliographie</b> . . . . .	<b>221</b>

---

**Abstract :** In this chapter I present some ongoing works at the moment of writing this report. First, I am using GalSVM in the framework of the ALHAMBRA consortium to analyze morphologies of field galaxies up to  $z \sim 1$ . The main goal is to use all the available filters (20) to obtain rest-frame morphologies. Once the survey will be complete, there will be 20 000 available galaxies. I present here a

preliminar analysis on 376 galaxies. Since the mean angular resolution is not very high (typically around  $\sim 1''$ ) the first task has been to find the limits of the morphological classification. We find that objects further than  $z \sim 1$  cannot be analyzed. A more precise quantification is on-going. We show however some preliminar scientific results : one of the major interests of these data is indeed that they enable a precise comparison of the morphological type with the spectral type. Our first results show that  $\sim 80\%$  of the early-type population has a spectral type of elliptical or Sa galaxies and that  $\sim 70\%$  of the late type galaxies are classified as Sab or Scd according to the spectral type. There are therefore some interesting contaminations which have to be analyzed.

Second, I am studying the morphologies in clusters at  $z \sim 0.5$  observed with MegaCam at CFHT and the relation between morphology and environment. I present here the analysis of the morphologies in MS 1621.5+2640 in the r band. We obtain the *color-magnitude* relation and the radius-morphology relation and compare to some published results obtained with the HST. The first results are in good agreement : we indeed find that the color-magnitude relation is in place at  $z \sim 0.5$  for this cluster and a clear correlation between the morphological mixing and the distance to the center.

Nevertheless, the reader must keep in mind that the results presented here are extremely preliminar.

---

**Resumen :** Este capítulo presenta algunos de los proyectos en curso en el momento de la redacción de la presente memoria. En un primer lugar, se presenta el trabajo iniciado en el marco del cartografiado cósmico ALHAMBRA que consiste principalmente en el análisis morfológico *rest-frame* de galaxias de campo hasta  $z \sim 1$  haciendo uso de la gran cantidad de filtros disponibles. Cuando las observaciones y la reducción de datos estén finalizados, se dispondrán de aproximadamente 20 000 galaxias para realizar el análisis. Se presenta a continuación un trabajo preliminar sobre 376 galaxias de campo. Dado que la resolución angular media no es excesivamente elevada ( $\sim 1''$ ) la primera parte del trabajo a consistido en establecer los límites del análisis morfológico. Se muestran de igual modo algunos análisis científicos preliminares como la comparación de los tipos morfológicos estructurales y espectrales. Se obtiene que el 80% de las galaxias clasificadas como tempranas desde un punto de vista morfológico tienen un tipo espectral temprano de igual modo mientras que el 70% de las galaxias tardías morfológicamente hablando son tardías desde un punto de vista espectral. Existen por consiguiente un cierto número de contaminaciones que están siendo analizadas.

En un segundo lugar, se abordan los primeros resultados del estudio de morfologías en cúmulos de galaxias en torno a  $z \sim 0.5$  observados con el CFHT así como de su relación con el entorno. Se obtienen las relaciones color-magnitud y morfología-distancia radial en el cúmulo MS 1621.5+2640 en el filtro r. Los primeros resultados presentan un buen acuerdo con los resultados publicados basados en datos del telescopio espacial.

El lector debe tener en cuenta que los resultados presentados están todavía en curso

de elaboración y son por lo tanto muy preliminares.

---

Je présente par la suite quelques uns des travaux en cours au moment de la rédaction de ce manuscrit. Il s'agit principalement de quelques applications de la méthode présentée dans le chapitre 9. Les résultats décrits sont par conséquent préliminaires et pourront donc subir des évolutions après la rédaction.

## 11.1 Morphologies ALHAMBRA

### 11.1.1 Le relevé ALHAMBRA

Le relevé ALHAMBRA (*Advanced Large Homogeneous Area Medium Band Redshift Astronomical survey*, Moles *et al.*, 2005) est un projet pensé pour sonder une fraction significative de l'Univers avec une précision suffisante pour suivre l'évolution du contenu et des propriétés de celui-ci en fonction du *redshift*. L'aire observée est de  $8^{(\circ)^2}$  et les observations sont réalisées sur 20 filtres contigus de même largeur spectrale ( $\sim 410\text{\AA}$ ) allant de  $3500\text{\AA}$  à  $9700\text{\AA}$  en plus de 3 filtres infrarouges (JHK). Les observations sont effectuées avec le télescope de 3.5 m de *Calar Alto* en Espagne. La limite de détection (pour les sources ponctuelles avec un rapport S/B=5) est fixée à AB=25 mag dans tous les filtres jusqu'à  $8300\text{\AA}$  et entre 24.7 et 23.4 pour les reste des filtres; la résolution moyenne est de l'ordre de  $\sim 1''$ .

La caractéristique principale de ce relevé réside dans l'échantillonnage continu du spectre lumineux. Il se place ainsi entre un relevé purement photométrique et un relevé spectroscopique. Il permet donc d'obtenir la distribution spectrale d'énergie de près de  $\sim 300\,000$  objets avec une précision suffisante pour obtenir des *redshifts* photométriques de grande qualité ( $\Delta z < 0.015(1+z)$ ).

Les données obtenues vont permettre d'étudier, entre autres, l'évolution des structures à grande échelle, l'évolution de galaxies (luminosités, morphologies), la sélection d'amas de galaxies ou l'identification de familles particulières d'objets sans nécessité de réaliser un suivi à posteriori avec d'autres instruments.

### 11.1.2 Objectifs scientifiques de l'étude

Le travail présenté par la suite est centré sur l'étude des morphologies. La haute résolution n'étant pas la priorité de ce relevé, les études morphologiques structurales sont particulièrement difficiles et comme on le verra par la suite, il est impossible d'obtenir une morphologie fiable d'objets très lointains (typiquement  $z \sim 1.5$ ) comme il a été fait pour COSMOS (c.f. chapitre 10). Par conséquent, la première question à laquelle nous essayons de répondre avec cette étude préliminaire est : est-il possible d'obtenir une estimation des morphologies des galaxies de champ ?

Si la réponse s'avérait être positive, une analyse morphologique dans ALHAMBRA présente plusieurs intérêts du fait du grand nombre de filtres disponibles et de l'uniformité de ceux-ci :

- (i) **Morphologies au repos** : L'emploi de tous les filtres simultanément permet d'obtenir en principe une morphologie au repos sur le même filtre de référence pour chaque *redshift* comme je le montrerai par la suite.
- (ii) **Comparaison SED - morphologie** : Comme il a été dit précédemment, le relevé devrait permettre l'obtention de distributions spectrales d'énergie des galaxies avec une grande précision. La comparaison de type morphologique estimé à partir de critères de forme uniquement avec le type spectrale peut permettre l'identification d'objets particuliers (e.g. elliptiques bleues).

### 11.1.3 Les données analysées

Le relevé couvre une aire de  $8^{(\circ)^2}$  distribuée en 8 champs (ALH01-ALH08) de  $1^{(\circ)^2}$  chacun, divisés à son tour en 2 pointages de  $30'' \times 30''$  et 4 CCDs de  $15'' \times 15''$ . Les résultats exposés ci-dessous portent sur le CCD 3 du pointage 1 du champ ALH08. Ce champ est l'un des premiers à avoir été observé et réduit et il existe une première estimation des *redshifts* photométriques.

L'aire analysée est donc de  $15'' \times 15''$  avec une taille de pixel de  $0.22''$ . Nous avons employé 10 filtres, 7 visibles (de  $7370\text{\AA}$  à  $9540\text{\AA}$ ) et deux infrarouges (J et H). Cette sélection est justifiée par le type d'étude morphologique que l'on souhaite faire, dans le paragraphe § 11.1.6.1 je détaille en effet le principe.

### 11.1.4 Catalogue photométrique et *redshifts* photométriques

Le catalogue photométrique employé pour la détection d'objets a été élaboré dans le cadre du consortium à l'aide de SExtractor et nettoyé des fausses détections. Il comprend 20 805 objets sur l'ensemble du champ. Les *redshifts* photométriques de 9 520 sources ont également été obtenus dans le cadre du consortium avec des codes propres. Il ne s'agit cependant que d'une version très préliminaire et c'est donc l'un des points qui est susceptible de faire évoluer ces premiers résultats.

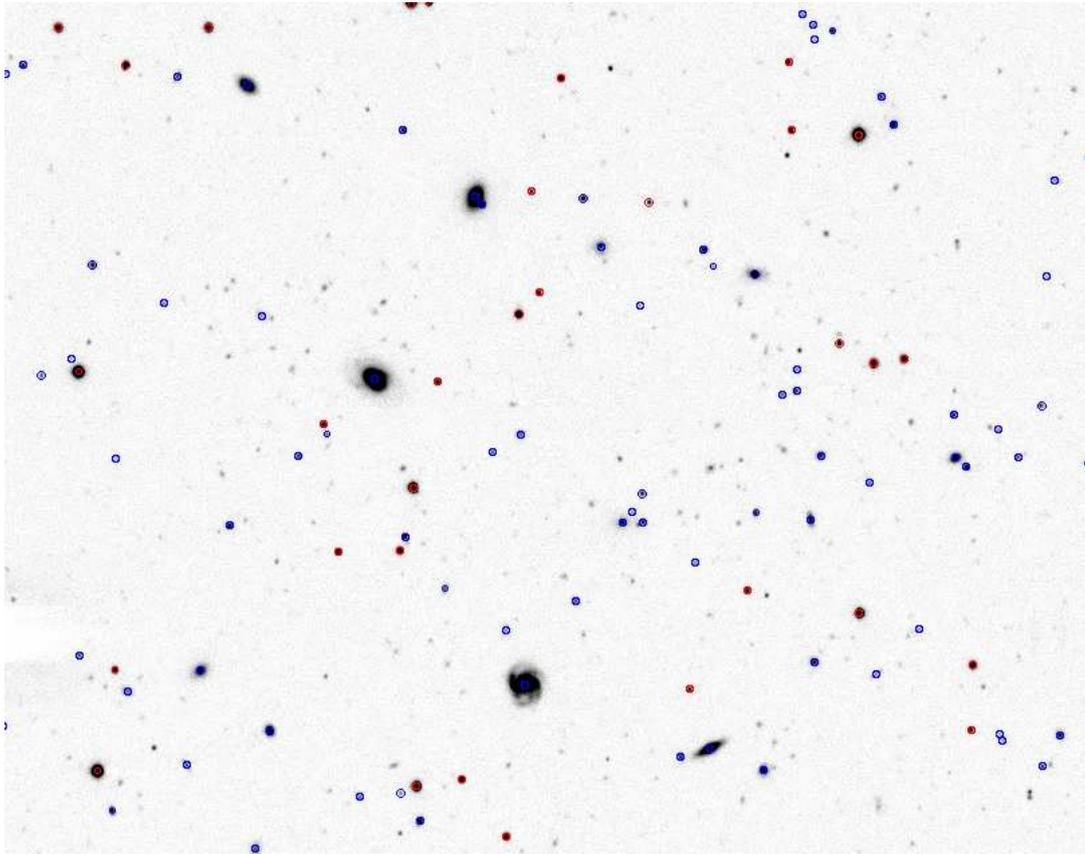


FIG. 11.1 – Section de  $5' \times 5'$  du champ analysé en bande I ( $7370\text{\AA}$ ). Les cercles rouges marquent les étoiles et les cercles bleus les galaxies sélectionnées pour l'analyse morphologique (c.f. § 11.1.5).

### 11.1.5 Sélection de l'échantillon morphologique

Comme pour le travail effectué dans le cadre de COSMOS, nous avons réduit l'échantillon avant de réaliser l'analyse morphologique à partir de deux critères basés sur des inspections visuelles et des simulations (ce point est certainement à améliorer pour raffiner la sélection) :

- Nous n'avons gardé que les objets ayant un *redshift* photométrique inférieur à 1.
- Concernant la luminosité, nous avons exigé que les objets sélectionnés aient une magnitude au moins 2 *mag* plus brillante que la position du pic de la fonction de magnitude dans chacun des filtres. Ainsi pour le filtre à  $7370\text{\AA}$ , (c.f Fig. 11.2) nous avons gardé tous les objets plus brillants que  $m_{7370} < 23.7$ . On compte ainsi 626 objets dans le catalogue final analysé.

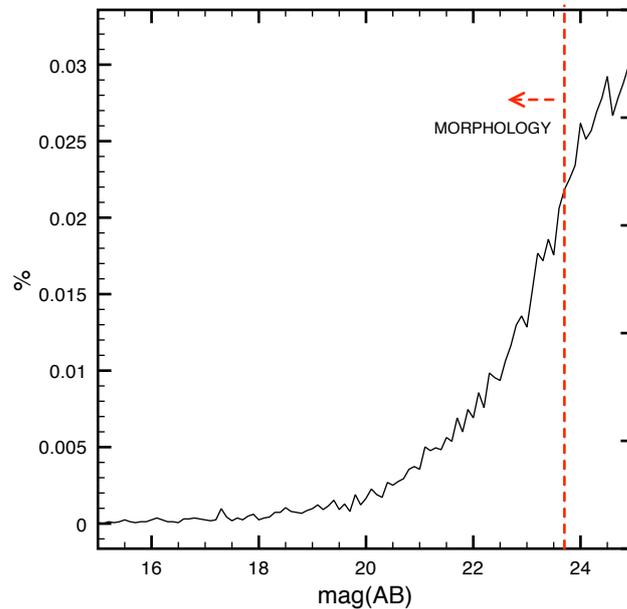


FIG. 11.2 – Comptages et sélection de l'échantillon pour le filtre à  $7370\text{\AA}$ . Ce critère a été appliqué à tous les filtres employés.

La séparation étoiles/galaxies a été réalisée d'une façon analogue à celle employée dans les chapitres précédents, à savoir en employant les paramètres MU\_CLASS et MAG\_AUTO de SExtractor. Nous obtenons 250 étoiles et 376 galaxies. Cette étape pourra être raffinée lors de l'étude définitive en incluant d'autres paramètres (CLASS\_STAR, type spectral) comme pour l'analyse dans le champ COSMOS. Même s'il peut avoir quelques contaminations, la séparation est suffisamment propre pour une première analyse. Ces valeurs nous permettent d'estimer le nombre de galaxies pour lesquelles nous sommes susceptibles d'avoir une estimation morphologique une fois que toutes les données seront disponibles. En effet, en considérant que  $\sim 400$  galaxies par champ de  $15' \times 15'$  pourront être employées, on obtient  $\sim 25\,000$  galaxies sur l'ensemble de l'aire observée.

### 11.1.6 Analyse morphologique

#### 11.1.6.1 Principe

Un des intérêts majeurs de ces données est sans doute le grand nombre de filtres disponibles et l’homogénéité de ceux-ci. Nous avons donc essayé d’exploiter cette propriété pour obtenir des morphologies au repos dans la bande I ( $\sim 7370\text{\AA}$ ) entre  $z=0$  et  $z=1$ . Pour cela, il a fallu changer légèrement l’approche suivie dans les chapitres précédents. Le but est maintenant de rester systématiquement centrés sur la bande I. Nous utilisons pour cela tous les filtres *plus rouges* que la bande I, et en fonction du *redshift*, chaque galaxie est analysée dans le filtre correspondant à la bande I au repos. Le choix du filtre est donc donné par l’expression :

$$\lambda_{\text{filtre}} = \lambda_I(1 + z) \quad (11.1)$$

Le choix de la bande I est justifié pour des raisons pratiques puisque l’échantillon SDSS que l’on possède pour réaliser l’entraînement est justement dans cette bande.

#### 11.1.6.2 Echantillon d’entraînement

La stratégie adoptée pour construire l’échantillon d’entraînement est légèrement différente à celle décrite dans le chapitre précédent. Nous n’avons utilisé que la bande I de notre échantillon SDSS (Tasca & White, 2005), soit  $\sim 500$  objets d’entraînement et  $\sim 500$  de test. Pour chaque galaxie :

- (i) On associe un *redshift* aléatoire qui suit la distribution réelle des *redshifts* photométriques.
- (ii) En fonction du *redshift* associé, le filtre plus proche de la bande I au repos est choisi d’après l’expression 11.1 et une magnitude aléatoire est tirée d’après la distribution de magnitudes réelle du filtre sélectionné.
- (iii) Une fois la magnitude et le *redshift* associés, la galaxie est simulée en suivant la procédure détaillée dans le chapitre 9.
- (iv) La galaxie simulée est finalement *plongée* dans un fond réel du filtre correspondant.

Ainsi, à la différence de l’échantillon employé pour analyser les données COSMOS, où l’on changeait l’échantillon local pour rester toujours dans la bande K à grand *redshift*, ici, l’échantillon d’entraînement se distribue dans les 10 filtres employés. La figure 11.3 illustre le processus schématiquement.

#### 11.1.6.3 Analyse des erreurs

Comme pour les données WIRCam, nous avons employé l’échantillon de test pour quantifier les erreurs de classification en fonction de la probabilité. Le tableau 11.1

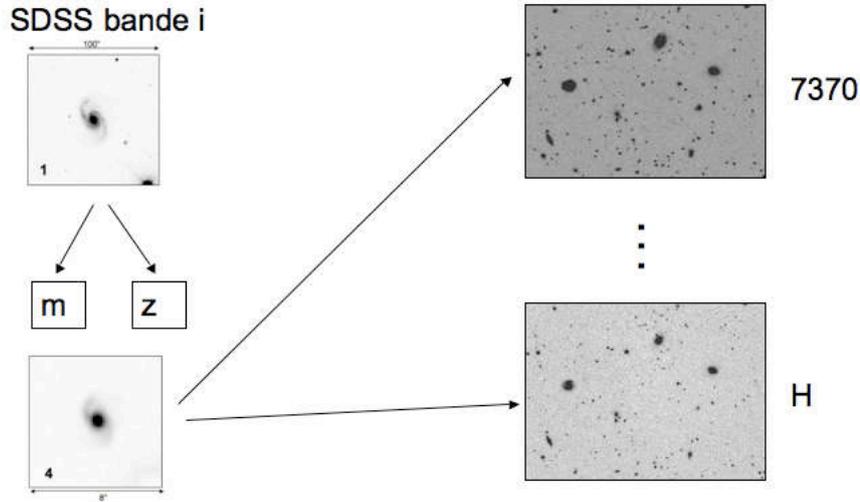


FIG. 11.3 – Construction de l'échantillon d'entraînement. En fonction du *redshift* associé, la galaxie simulée est plongée dans un filtre différent pour rester dans la même bande au repos.

résume les performances obtenues. Nous constatons tout d'abord que les contaminations sont plus importantes que pour l'échantillon WIRCam probablement à cause de la résolution plus basse. Il existe en particulier une tendance à sur-estimer le nombre de galaxies elliptiques, ce qui est logique si certains objets sont peu ou pas résolus. Les performances augmentent nettement avec la probabilité et nous atteignons un taux de réussite supérieur à 80%. Des recherches sont en cours pour améliorer la qualité, en particulier, l'inclusion de fonctions de magnitudes par type est susceptible d'apporter un gain. La figure 11.4 montre la distribution des objets réels en fonction de la probabilité. On constate qu'un nombre important d'objets (essentiellement des elliptiques) a une probabilité de classification proche de 0.5.

TAB. 11.1 – Taux de réussite de la classification en fonction de la probabilité

	$p > 0.5$		$p > 0.6$	
	Précoce	Tardive	Précoce	Tardive
Précoce visuelle	0.74	0.35	0.78	0.29
Tardive visuelle	0.25	0.65	0.22	0.71
	$p > 0.7$		$p > 0.8$	
	Précoce	Tardive	Précoce	Tardive
Précoce visuelle	0.81	0.20	0.80	0.14
Tardive visuelle	0.19	0.80	0.2	0.86

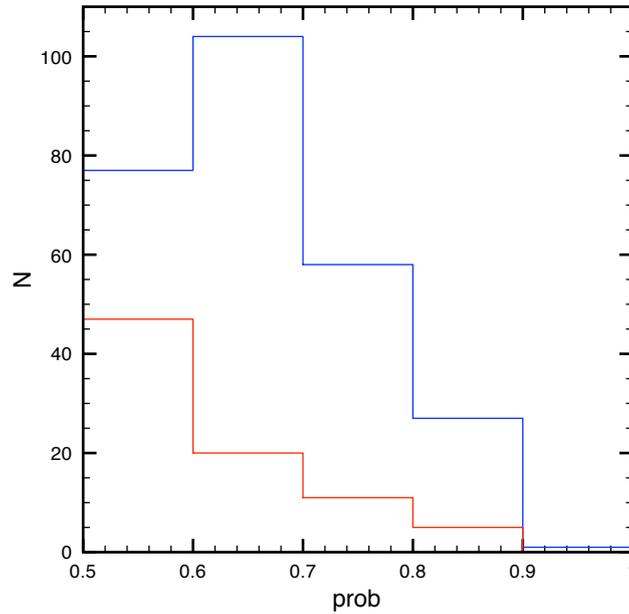


FIG. 11.4 – Histogramme des probabilités pour les galaxies réelles. Rouge : précoces, Bleu : tardives

#### 11.1.6.4 Quelques résultats préliminaires

Malgré le caractère encore peu fiable des classifications et des *redshifts* photométriques nous montrons par la suite quelques unes des exploitations scientifiques possibles afin d'illustrer ce que pourra apporter ce travail une fois abouti.

**Mélange morphologique global** L'analyse des classifications sur l'ensemble du champ donne (*estimateur par comptages*) :

- 84 galaxies précoces ( $\sim 25\%$ )
- 285 galaxies tardives ( $\sim 75\%$ )

Cette première estimation est cohérente avec celle obtenue dans le cadre de COSMOS. Une inspection visuelle révèle de plus que les galaxies plus brillantes et résolues pour lesquelles le type morphologique est facilement identifiable à l'oeil sont classées correctement.

**Evolution morphologique** La figure 11.5 montre l'évolution du mélange morphologique en fonction du *redshift*. Nous rappelons que ce résultat doit être pris avec énormément de précautions à cause du caractère provisoire des *redshifts*. Le nombre d'objets par intervalle est également petit et donc probablement affecté par des effets de variance cosmique.

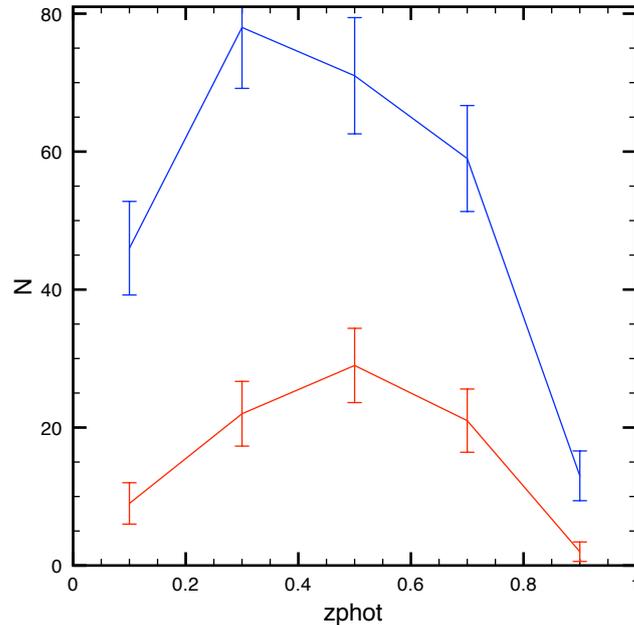


FIG. 11.5 – Evolution morphologique dans l'échantillon ALHAMBRA analysé. Rouge : précoces, Bleu : tardives. Les barres d'erreurs sont issues d'une statistique de Poisson.

**Relation morphologie-type spectral** Comme il a été dit auparavant, un des points intéressants de cette étude réside dans la comparaison des morphologies structurales avec les morphologies spectrales issues de l'estimation des *redshifts* photométriques. La figure 11.6 montre cette correspondance. Pour chaque type morphologique déterminé avec GalSVM nous avons regardé son type spectral d'après l'ajustement. Si la correspondance était *parfaite*, on s'attendrait à ce que toutes les galaxies classées comme elliptiques aient un type spectral de galaxie précoce et inversement que toutes les galaxies classées spirales soient associées à un spectre tardif. La figure 11.6 révèle en effet des tendances attendues mais aussi certains comportements relativement surprenants :

- Une fraction importante de galaxies classées comme tardives d'un point de vue structural est associée à un type spectral précoce. Une inspection visuelle de ces objets révèle, au moins pour les plus brillants, la présence d'une composante de type disque mais également d'un bulbe important. Une explication possible de cette divergence pourrait être donc que la photométrie effectuée pour l'estimation de *redshifts* n'ait été réalisée que sur la partie centrale de celle-ci, ne prenant donc en compte que la lumière provenant du bulbe. Ce point devra donc être vérifié lorsque les *redshifts* photométriques seront définitifs.
- Il existe également une population de galaxies classées comme elliptiques mais présentant un spectre tardif (19 objets). Peut-il s'agir de la population de galaxies elliptiques bleues ? Ou alors s'agit-il d'un reflet des erreurs de classification ? L'inspection visuelle montre que la plupart de ces objets sont faibles et ont

une probabilité associée proche de 0.5 ce qui pourrait donc aller dans le sens des erreurs. Malgré cela, ils semblent compacts et pour les plus brillants on arrive à discerner un bulbe. Une caractérisation plus précise de ces objets peut s'avérer donc un objectif intéressant de cette étude. Nous avons commencé à aborder ce point dans le paragraphe suivant.

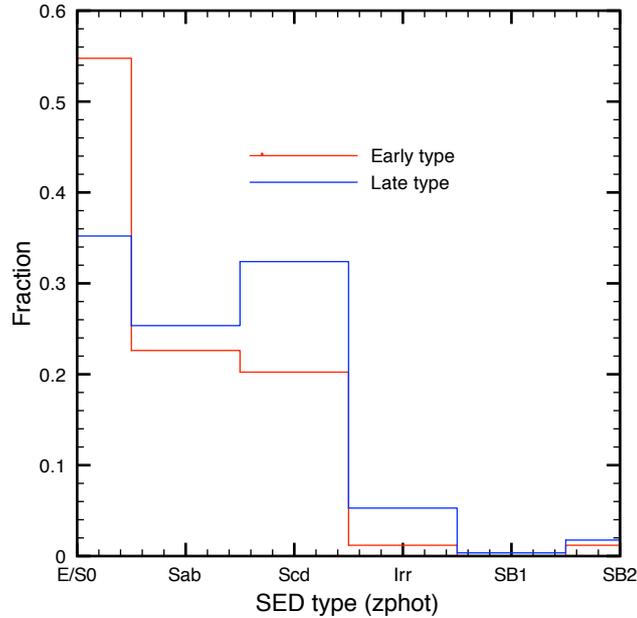


FIG. 11.6 – Comparaison du type morphologique dérivé avec GalSVM avec le type spectral issu de l'estimation des *redshifts* photométriques. Pour chaque type morphologique on montre la fraction d'objets de chaque type spectral.

**Galaxies elliptiques bleues** Nous avons donc regardé un peu plus dans le détail ces 19 objets qui présentent une morphologie elliptique et un spectre tardif. En particulier, nous avons regardé les magnitudes absolues dans la bande K qui doivent en principe être corrélées avec la masse. Il ne s'agit que d'une première approximation étant donné que nous n'avons appliqué aucune k-correction. La figure 11.7 compare la distribution de magnitudes absolues de toutes les galaxies elliptiques avec celle *elliptiques bleues*. On constate que ces dernières sont globalement moins brillantes (i.e. moins massives en première approximation).

### 11.1.7 Bilan et perspectives

Nous avons réalisé une première analyse très préliminaire des morphologies de 376 galaxies dans la bande I au repos à l'aide de GalSVM. L'échantillon d'étude pourra être augmenté de manière significative dès que la réduction des données sera finie et les *redshifts* photométriques seront stables. L'analyse de tous les champs du relevé

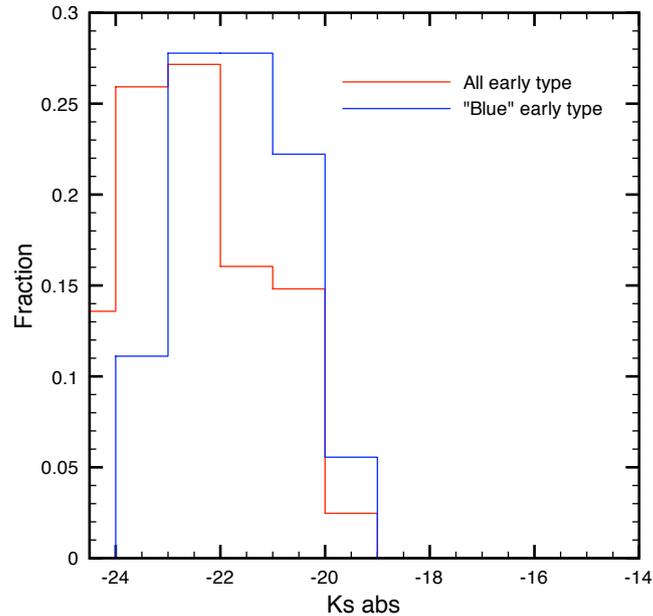


FIG. 11.7 – Distribution des magnitudes absolues en bande Ks pour toutes les galaxies elliptiques (rouge) et pour les galaxies elliptiques présentant des traces de formation stellaire (bleu).

devrait fournir un échantillon de plus de  $\sim 20\,000$  galaxies exploitables.

Les premiers tests donnent des performances inférieures à celles obtenues pour les données WIRCam probablement à cause de la résolution angulaire plus basse. Celles-ci vont peut être être améliorées en raffinant le processus de simulation de l'échantillon d'entraînement (e.g. distributions de magnitudes asymétriques). Une étude plus approfondie en termes de probabilité est en cours de réalisation.

Nous avons tout de même commencé une exploitation scientifique afin d'illustrer les principaux objectifs de cette étude. Nous avons ainsi pu repérer deux grands axes d'analyse qui feront l'objet d'une étude plus poussée lorsque les données seront définitives :

- *Comparaison type spectral - morphologie* : Cette première comparaison révèle des tendances attendues mais aussi des effets étranges. Une étude en termes de probabilité comme celle effectuée dans le chapitre précédent permettra peut-être d'éclaircir ce point. Rappelons cependant que ce résultat est extrêmement préliminaire à cause du caractère provisoire des *redshifts* photométriques.
- *Etude détaillée de populations particulières* : La comparaison de la morphologie avec le type spectral permet d'identifier des populations d'objets particuliers comme par exemple les elliptiques bleues. L'étude détaillée de ce type de galaxies est un point susceptible d'être abordé avec les données du relevé ALHAMBRA.

## 11.2 Morphologies dans des amas à *redshift* intermédiaire ( $z \sim 0.5$ )

Les amas de galaxies sont les structures plus grandes de l'Univers liées gravitationnellement. La densité élevée de l'environnement fait de ceux-ci un laboratoire particulièrement intéressant pour étudier, entre autres, les effets de l'environnement dans l'évolution des galaxies (Dressler, 1984; Martel *et al.*, 1998). Leurs coeurs représentent en effet les zones de plus forte densité galactique de l'Univers, par conséquent quelle qu'elle soit, la dépendance de la formation ou de l'évolution des galaxies avec l'environnement doit être plus marquée dans les amas. Le code développé est donc entrain d'être employé également pour étudier les morphologies dans des amas à *redshift* intermédiaire.

### 11.2.1 Motivations du travail réalisé

L'étude morphologique dans les amas peut donner des indications intéressantes concernant la formation et l'évolution des galaxies elliptiques. En effet, comme il a été exposé dans le chapitre d'introduction, les galaxies précoces dominent les régions centrales des amas (Hubble & Humason, 1931). Dans les amas locaux elles présentent de plus une relation étroite entre la couleur et la magnitude, avec une petite dispersion ce qui laisse entrevoir, soit une formation très tôt dans l'histoire de l'Univers soit une formation synchronisée (Bower *et al.*, 1992). Il semblerait que cette relation était déjà en place à  $z \sim 0.5$  (Ellis *et al.*, 1997) indiquant que ces galaxies on dû se former avant  $z \sim 3$ .

Un autre phénomène intéressant qui peut être également étudié dans les amas est la relation morphologie-densité ou la relation morphologie-distance radiale (Melnick & Sargent, 1977; Dressler, 1980; Whitmore *et al.*, 1993). Ces relations décrivent comment les nombres relatifs de chaque type morphologique évoluent en fonction de la densité locale ou de la distance au centre de l'amas. Les centres des amas sont dominés par des galaxies précoces (elliptiques ou S0) tandis que les régions plus extérieures présentent des rapports morphologiques plus proches de ceux que l'on retrouve pour les galaxies de champ. Dressler *et al.* (1997) ont étudié l'évolution de ces relations à partir de  $z \sim 0.5$  grâce à des données sur 10 amas. En prenant tous les amas sans distinction il s'avère que la relation morphologie-densité est pratiquement absente  $z \sim 0.5$  à l'exception des régions de très forte densité où les galaxies elliptiques dominent. En revanche, à bas *redshift* tous les amas présentent une relation nette entre la morphologie et la densité. En divisant cependant l'échantillon par type d'amas (très concentré, peu concentré et irrégulier) on s'aperçoit que seul les amas très concentrés présentent une relation entre la morphologie et la densité.

Les résultats énoncés ci-dessus emploient tous des morphologies obtenues grâce à

l'imagerie du télescope spatial. Les statistiques portent souvent sur peu d'objets (5 à 10 amas) et généralement les études sont effectuées dans les parties plus centrales des amas à cause du champ réduit du HST. L'imagerie MegaCam offre un très grand champ permettant d'aller sonder les périphéries des amas et d'augmenter les statistiques. L'inconvénient majeur est bien entendu la résolution angulaire obtenue, bien en dessous de celle délivrée par le télescope spatial. L'utilisation du code développé peu fournir des résultats fiables sur des amas à *redshift* intermédiaire? Nous nous interrogeons sur ce point dans les paragraphes qui suivent.

## 11.2.2 Description des données

### 11.2.2.1 Contexte scientifique

Les données employées ont été obtenues dans un contexte scientifique différent. Elles font en effet partie d'un suivi optique d'un échantillon d'amas X à *redshift* intermédiaire ( $\sim 0.5$ ) (Arnaud, 2004). Il est formé des 13 amas les plus brillants en X, donc à priori les plus massifs, pour en faire l'analyse *weak lensing*. Le but scientifique principal est d'établir les relations entre les grandeurs X et la masse déterminée indépendamment par *lensing*. Les données ont été obtenues avec MegaCam en quatre couleurs : 2h en r (imagerie profonde et de bonne qualité, avec un *seeing*  $< 0.8''$ , pour faire le *weak lensing*) et plus court en g, i et z.

### 11.2.2.2 MS 1621.5+2640

Afin de réaliser ce premier travail préliminaire de validation, nous avons choisi de travailler sur un des amas de l'échantillon : MS 1621.5+2640. Il existe des données spectroscopiques et photométriques sur cet amas (Ellingson *et al.*, 1997) obtenues avec le CFHT dans le cadre d'un relevé sur des amas mené par le CNOC (*Canadian Network for Observational Cosmology*) (Yee *et al.*, 1996). De plus, Saintonge *et al.* (2005) ont réalisé une étude détaillée des morphologies dans l'amas à partir de l'imagerie HST en deux bandes (F814W et F555W). Il s'agit donc de la cible idéale pour valider la qualité de la morphologie obtenue à partir des données sol. L'amas est situé à  $z = 0.4275$ , il a une masse de  $10^{15} M_{\odot}$  et une dispersion de vitesses de  $\sigma = 793 \text{ km.s}^{-1}$  (Carlberg *et al.*, 1996). Son flux en X (0.3 – 3.5 keV) est de  $L_X = 9.74 \times 10^{-13} \text{ ergs.cm}^{-2}.\text{s}^{-1}$  (Gioia *et al.*, 1990) et son rayon caractéristique (distance au centre où la densité de l'amas est 200 fois la densité critique de l'Univers) est  $r_{200} = 278''$  (Carlberg *et al.*, 1997).

### 11.2.3 Sélection des galaxies appartenant à l’amas

Le premier travail consiste à isoler les galaxies appartenant à l’amas. Il existe plusieurs méthodes pour ce faire, on peut par exemple utiliser la relation couleur-magnitude. Dans notre cas, nous disposons de données photométriques en quatre bandes (g,r,i et z) sur l’amas. Il est par conséquent possible d’estimer les *redshifts* photométriques des galaxies et d’isoler les galaxies grâce à ce critère. D’ autant plus, que pour cet amas en particulier, certains objets ont une mesure du *redshift* spectroscopique qui peut donc être employé pour étalonner les *redshifts* photométriques. Nous avons donc utilisé le code HyperZ (Bolzonella *et al.*, 2000). Au moment de la rédaction de ce manuscrit, cependant, certains problèmes sont encore en train d’être résolus. En effet, des problèmes de photométrie dans l’une des bandes semblent poser problème à l’heure d’estimer proprement les *redshifts*. Il a fallu donc introduire des facteurs de correction pour faire face à ce problème.

Pour cette première analyse, nous avons donc décidé d’employer la sélection de Saintonge *et al.* (2005). Nous avons donc sélectionné 167 galaxies par identification avec le catalogue publié. Les résultats exposés para la suite portent sur cet échantillon. Ce point reste donc à améliorer pour les travaux à venir.

### 11.2.4 Analyse morphologique

L’analyse morphologique a été réalisée avec GalSVM (c.f. chapitre 9, Huertas-Company *et al.*, 2008) dans la bande r (6250Å). Nous avons simulé un échantillon d’entraînement à partir de l’échantillon SDSS utilisé pour les études précédentes, mais cette fois-ci, nous avons employé la bande g qui correspond à la bande r au repos à  $z \sim 0.4$ . La simulation de l’échantillon est légèrement différente que pour les galaxies de champ car tous les objets sont situés à la même distance. Nous avons donc simulé une distribution en *redshift* qui pique au *redshift* de l’amas. Pour la distribution en magnitude, la distribution réelle mesurée sur les images r a été employée.

TAB. 11.2 – Résumé des propriétés de l’amas MS 1621.5+2640. Les références sont indiquées dans le texte

$\alpha$ (J2000)	$\delta$ (J2000)	$z$	$L_X$ ( $10^{-13} \text{ ergs.cm}^{-2}.s^{-1}$ )
16 25 38.36	26 27 42.59	0.4275	9.74

Masse ( $10^{15} M_\odot$ )	$r_{200}$ (arcsec)	$\sigma$ ( $km.s^{-1}$ )
1.52	278	793

Nous avons donc construit un échantillon d'entraînement de 500 galaxies dont la morphologie est connue. Un classement à 7 paramètres : concentration, asymétrie, ellipticité, gini,  $M_{20}$  et *smoothness* a été réalisé. Nous avons décidé dans cette première analyse de ne pas inclure des paramètres tels que la magnitude pour éviter des effets tels que ceux exposés dans le chapitre 10.

### 11.2.5 Inspection visuelle

Nous avons tout d'abord comparé les résultats de la classification à ceux obtenus par [Saintonge et al. \(2005\)](#) sur des images HST avec le filtre F555W (5407Å). Les classifications à partir des images spatiales ont été effectuées à l'aide d'un ajustement paramétrique d'un modèle analytique (c.f. chapitre 2). Le paramètre de classification obtenu est par conséquent la fraction de bulbe par rapport au total ( $B/T$ ). Nous avons donc séparé dans le catalogue publié les galaxies dominées par un bulbe ( $B/T > 0.5$ ) de celles dominées par un disque ( $B/T < 0.5$ ) et les avons comparé à notre classification en deux types à partir de l'image MegaCam en bande r. La figure 11.8 montre les deux images centrées sur la partie centrale de l'amas. Nous constatons tout d'abord que les galaxies précoces semblent dominer le centre de l'amas comme attendu, et ce pour les deux classifications. Il peut paraître cependant étrange que les deux galaxies les plus brillantes de l'amas (en particulier la galaxie centrale) soient classées comme tardives. En effet, il est bien connu que les galaxies plus brillantes sont généralement des galaxies elliptiques. En regardant l'image de ces galaxies plus en détail on s'aperçoit cependant que ce sont des galaxies doubles non séparées. Les procédures automatiques ont donc tendance à trouver un type tardif.

La comparaison entre les deux classifications est globalement en bon accord indiquant que la classification à partir des données limitées par le *seeing* semble robuste. Une quantification plus précise révèle que les deux classifications sont en accord pour 116 galaxies parmi les 167. Il existe donc un certain nombre de différences. Il peut y avoir plusieurs raisons à l'origine de celles-ci : (1) il peut s'agir d'un simple reflet des erreurs de classifications qui doivent être présentes dans les deux échantillons étant donné leur caractère automatique, (2) il se peut également qu'elles soient une conséquence des définitions différentes du type morphologique (nous définissons les galaxies elliptiques à partir de critères visuels sur l'échantillon local alors que [Saintonge et al. \(2005\)](#) le font à partir du rapport  $B/T$ ).

Nous réalisons par la suite d'autres tests afin de caractériser plus précisément ces différences.

### 11.2.6 Relation *couleur–magnitude*

Nous avons donc regardé également la relation *couleur–magnitude* dans l'amas. Comme nous l'avons vu auparavant et dans le chapitre introductif, il existe une

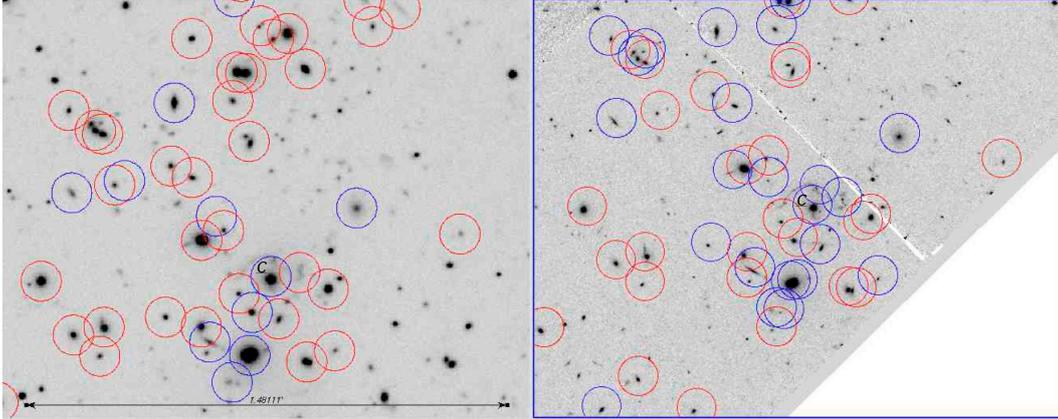


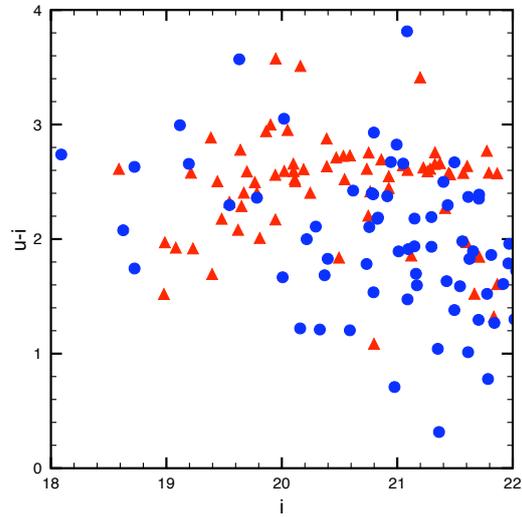
FIG. 11.8 – Comparaison des classifications dans la partie centrale de l’amas avec MegaCam en bande r (gauche) et avec le télescope spatial avec le filtre F555W (à droite). Les cercles bleus indiquent les galaxies tardives et les cercles rouges les galaxies précoces. La lettre « C » indique la galaxie centrale de l’amas.

bimodalité dans la distribution des couleurs et des types morphologiques. Cette dichotomie se traduit par une ségrégation dans les diagrammes couleur magnitude : les galaxies elliptiques étant plus rouges et avec des propriétés plus uniformes, se situent dans une *séquence rouge* peu dispersée alors que les galaxies tardives sont dans un *nuage bleu* plus étendu. L’analyse du diagramme *couleur–magnitude* s’avère donc être un bon test de la qualité de la classification.

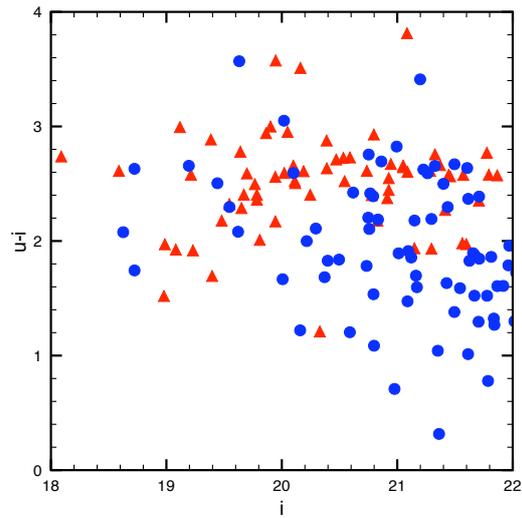
La figure 11.9 montre le diagramme  $i$  vs.  $u - i$  d’après notre classification et la classification de [Saintonge et al. \(2005\)](#). Puisque toutes les galaxies sont au même *redshift*, nous pouvons utiliser les magnitudes apparentes sans nécessité d’obtenir les magnitudes absolues.

Nous observons que les galaxies classées comme précoces sont globalement plus rouges que les galaxies tardives, comme attendu. De plus elles ont tendance à former une séquence bien définie alors que les galaxies tardives sont plus dispersées dans le diagramme. Cela semble encore être un signe en faveur de la robustesse de la classification obtenue.

Les deux diagrammes sont de plus en bon accord globalement. On constate malgré tout quelques « contaminations », c’est-à-dire des galaxies classées comme elliptiques mais se situant dans le nuage bleu qui semblent être légèrement plus abondantes dans la classification à partir des données MegaCam en particulier à des magnitudes faibles ( $i > 21$ ). S’agit-il d’erreurs de classification ? Ou au contraire ce sont réellement des galaxies elliptiques bleues ?



(a)



(b)

FIG. 11.9 – Comparaison des diagrammes *couleur-magnitude* obtenus à partir des classifications sur les images MegaCam (a) et sur les images HST (b). Les triangles rouges sont les galaxies classées comme précoces (E/S0 pour MegaCam et  $B/T > 0.5$  pour HST) et les cercles bleus celles classées comme tardives (S/Irr pour MegaCam et  $B/T < 0.5$  pour HST).

### 11.2.7 Relation *morphologie–distance radiale*

Nous avons par la suite étudié la relation entre la morphologie et la distance radiale pour les deux classifications. Pour cela, nous avons obtenu la distance au centre de l'amas de chacune des galaxies puis divisé l'échantillon en intervalles de distance de telle sorte que le nombre de galaxies soit le même dans chaque intervalle. Les résultats sont affichés dans la figure 11.10.

On constate tout d'abord qu'il existe effectivement une relation entre la distance au centre et la morphologie : les galaxies elliptiques dominent les parties centrales de l'amas ( $r \sim 50''$ ) et leur abondance diminue au fur et à mesure que la distance au centre augmente. L'interprétation est à prendre avec précaution car nous n'avons pas pris toutes les galaxies de l'amas mais seulement celles qui ont un équivalent dans le catalogue de [Saintonge et al. \(2005\)](#).

Le même travail a été réalisé à partir des classifications paramétriques sur les images HST. Les résultats sont affichés également dans la figure 11.10. Les tendances sont les mêmes mais on observe un certain nombre de différences. Celles-ci peuvent paraître importantes mais elles portent sur 3 objets maximum par intervalle de distance. Elles peuvent donc être une conséquence d'erreurs de classification.

### 11.2.8 Conclusions et travaux en cours

Nous avons employé le code développé dans le cadre de cette thèse pour réaliser une étude des morphologies dans un amas (MS 1621.5+2640) à *redshift* intermédiaire ( $z \sim 0.4275$ ) à partir de l'imagerie MegaCam en bande r. Il fait partie d'un échantillon de 13 amas X. L'amas ayant été analysé à partir de l'imagerie HST ([Saintonge et al., 2005](#)), il s'agit d'une cible idéale pour valider dans un premier temps l'information morphologique obtenue à partir des données sol.

La sélection des galaxies appartenant à l'amas a été réalisée dans un premier temps à partir du catalogue publié afin de faciliter les comparaisons. Cependant, tous les amas du catalogue ont été observés en quatre bandes (g, r, i et z). Il est donc possible d'obtenir des *redshifts* photométriques et de les employer pour sélectionner les objets. Cela fait partie des travaux en cours.

Nous avons obtenu les morphologies en deux types (précoce, tardif) de 167 galaxies de l'amas. Afin de vérifier la fiabilité de la morphologie obtenue, nous avons tracé les diagrammes *couleur–magnitude* et la relation *morphologie–distance radiale*. Pour le premier, nous constatons que les galaxies elliptiques se situent globalement dans une séquence rouge, comme attendu. Pour le deuxième, on identifie une relation entre la morphologie et la distance, les galaxies précoces dominant les parties centrales. Ces mêmes diagrammes, tracés en employant des classifications à partir des données du télescope spatial offrent des résultats très similaires.

Nous allons donc étendre les analyses présentées ci-dessus à l'ensemble de l'échantillon de 13 amas, avec une sélection des membres basée sur des *redshifts* photométriques.

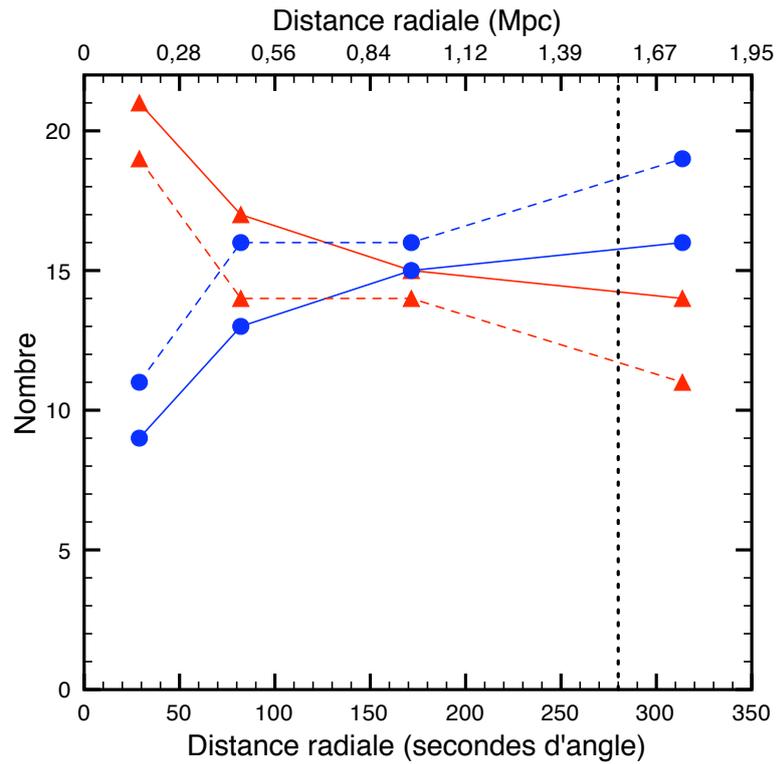


FIG. 11.10 – Relations *morphologie–distance radiale* obtenues à partir de la classification morphologique à partir des images MegaCam (lignes continues) et HST (ligne discontinue). Les lignes rouges sont les galaxies précoces et les bleues les tardives. La droite verticale indique le rayon caractéristique. On donne également la distance radiale en Mpc à  $z = 0.4275$  avec une cosmologie standard.

## Bibliographie

- Arnaud M., 2004. XMM-Newton Proposal 03025819. Dans *XMM-Newton Proposal ID #03025819*, p. 96–+. 214
- Bolzonella M., Miralles J.M., & Pelló R., 2000. Photometric redshifts based on standard SED fitting procedures. *A&A*, 363 : 476. 215
- Bower R.G., Lucey J.R., & Ellis R.S., 1992. Precision Photometry of Early Type Galaxies in the Coma and Virgo Clusters - a Test of the Universality of the Colour / Magnitude Relation - Part Two - Analysis. *MNRAS*, 254 : 601. 213
- Carlberg R.G., Yee H.K.C., & Ellingson E., 1997. The Average Mass and Light Profiles of Galaxy Clusters. *ApJ*, 478 : 462. 214
- Carlberg R.G., Yee H.K.C., Ellingson E., Abraham R., Gravel P., Morris S., & Pritchett C.J., 1996. Galaxy Cluster Virial Masses and Omega. *ApJ*, 462 : 32. 214
- Dressler A., 1980. Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *ApJ*, 236 : 351. 213
- Dressler A., 1984. The Evolution of Galaxies in Clusters. *ARA&A*, 22 : 185. 213
- Dressler A., Oemler A.J., Couch W.J., Smail I., Ellis R.S., Barger A., Butcher H., Poggianti B.M., & Sharples R.M., 1997. Evolution since  $Z = 0.5$  of the Morphology-Density Relation for Clusters of Galaxies. *ApJ*, 490 : 577. 213
- Ellingson E., Yee H.K.C., Abraham R.G., Morris S.L., Carlberg R.G., & Smecker-Hane T.A., 1997. The CNOC Cluster Redshift Survey Catalogs. III. MS 1621.5+2640 and MS 0302.7+1658. *ApJs*, 113 : 1. 214
- Ellis R.S., Smail I., Dressler A., Couch W.J., Oemler A.J., Butcher H., & Sharples R.M., 1997. The Homogeneity of Spheroidal Populations in Distant Clusters. *ApJ*, 483 : 582. 213
- Gioia I.M., Maccacaro T., Schild R.E., Wolter A., Stocke J.T., Morris S.L., & Henry J.P., 1990. The Einstein Observatory Extended Medium-Sensitivity Survey. I - X-ray data and analysis. *ApJs*, 72 : 567. 214
- Hubble E. & Humason M.L., 1931. The Velocity-Distance Relation among Extra-Galactic Nebulae. *ApJ*, 74 : 43. 213
- Huertas-Company M., Rouan D., Tasca L., Soucail G., & Le Fèvre O., 2008. A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images. I. Method description. *A&A*, 478 : 971. 215
- Martel H., Premadi P., & Matzner R., 1998. Morphological Evolution of Galaxies. *ApJ*, 497 : 512. 213

- Melnick J. & Sargent W.L.W., 1977. The radial distribution of morphological types of galaxies in X-ray clusters. *ApJ*, 215 : 401. 213
- Moles M., Alfaro E., Benítez N., Broadhurst T., Castander F.J., Cepa J., Cerviño M., Fernández-Soto A., González Delgado R.M., Infante L., López Aguerri A., Márquez I., Martínez V.J., Masegosa J., del Olmo A., Perea J., Prada F., Quintana J.M., & Sánchez S., 2005. The ALHAMBRA Survey : For a systematic Study of Cosmic Evolution. *ArXiv Astrophysics e-prints*. 203
- Saintonge A., Schade D., Ellingson E., Yee H.K.C., & Carlberg R.G., 2005. Catalog of Galaxy Morphology in Four Rich Clusters : Luminosity Evolution of Disk Galaxies at  $0.33 < z < 0.83$ . *ApJs*, 157 : 228. 214, 215, 216, 217, 219
- Tasca L. & White S., 2005. Quantitative morphology of Galaxies from the SDSS I : Luminosity in Bulges and Disks. *astro-ph/0507249*. 207
- Whitmore B.C., Gilmore D.M., & Jones C., 1993. What determines the morphological fractions in clusters of galaxies? *ApJ*, 407 : 489. 213
- Yee H.K.C., Ellingson E., & Carlberg R.G., 1996. The CNOC Cluster Redshift Survey Catalogs. I. Observational Strategy and Data Reduction Techniques. *ApJs*, 102 : 269. 214

# Conclusions de la deuxième partie

Dans la deuxième partie de ce manuscrit, j'ai exploré des moyens pour estimer les morphologies de galaxies lointaines ( $z < 2$ ) observées avec des télescopes terrestres sans correction de la turbulence atmosphérique.

J'ai, pour ce faire, apporté un certain nombre d'améliorations aux méthodes non-paramétriques. D'un côté j'ai résolu le problème de l'étalonnage de l'espace de paramètres grâce à l'utilisation d'un échantillon local, bien résolu et avec un rapport S/B élevé, qui est adapté pour reproduire les caractéristiques physiques et instrumentales de l'échantillon à analyser. Deuxièmement, j'ai montré que l'utilisation de machines à vecteur support permet de faire une analyse avec un grand nombre de paramètres simultanément. Ceci n'a pas d'effet significatif si l'échantillon est de bonne qualité mais produit un gain significatif (80% comparé à 60% de classifications correctes) pour un échantillon à  $z \sim 1$ . Le code permettant de réaliser ces analyses a été mis à disposition de la communauté : <http://www.lesia.obspm.fr/~huertas/galsvm.html>.

Dans un deuxième temps, j'ai employé la méthode développée pour analyser les morphologies de  $\sim 50\,000$  galaxies contenues dans le champ COSMOS dans l'infrarouge proche (bande Ks). Ceci représente la première étude quantitative de l'évolution morphologique des galaxies de champ entre  $z = 0$  et  $z = 2$  à partir de morphologies estimées dans l'infrarouge. On mesure une augmentation de la fraction de galaxies elliptiques depuis  $z \sim 2$  jusqu'à nos jours. C'est une étude particulièrement importante, car, au delà de  $z = 1$ , la lumière visible analysée avec le télescope spatial sonde l'émission UV de la galaxie. Un effort particulier a donc été fait pour quantifier proprement les effets de k-correction. J'ai ainsi comparé la classification obtenue avec une classification sur les mêmes objets basée sur des données visibles du HST. Alors que les classifications sont globalement en bon accord, il existe des différences à grand *redshift* ( $z > 1$ ), en particulier pour la population de galaxies elliptiques.

Finalement, j'ai présenté un certain nombre de travaux en cours de développement au moment de la rédaction du manuscrit. Il s'agit principalement d'autres applications du code développé. La première de ces études s'inscrit dans le cadre du consortium ALHAMBRA et vise à étudier les morphologies des galaxies de champ jusqu'à  $z \sim 1$  en restant toujours dans la même bande photométrique au repos. Les résultats sont pour l'instant préliminaires mais plusieurs voies d'étude ont été envisagées. En particulier, cette analyse devrait permettre une comparaison robuste du type morphologique structurel avec le type spectral issu de l'estimation des *redshifts*

photométriques jusqu'à  $z \sim 1$ .

La dernière étude présentée porte sur l'analyse de morphologies dans des amas situés autour de  $z \sim 0.5$ . J'ai montré en particulier les résultats de l'analyse sur l'amas MS 1621.5+2640 dans la bande r. Les morphologies obtenues sont en bon accord avec celles obtenues avec des données du télescope spatial. On constate en particulier que la plupart des galaxies elliptiques se situent dans une séquence rouge dans le plan couleur-magnitude et que la relation morphologie-distance radiale est en place pour cet amas.

# Conclusions

J'ai voulu, à travers le travail de cette thèse, étudier les possibilités actuelles de quantifier des morphologies de galaxies à grand redshift ( $z \sim 1$ ) dans l'infrarouge proche en employant des télescopes au sol. Jusqu'à présent, la grande majorité des études faisant intervenir des morphologies sont basées sur des données visibles du télescope spatial Hubble. Les nouveaux relevés cosmologiques sondant des objets au delà de  $z \sim 1$  sont demandants d'une morphologie estimée dans l'infrarouge à cause des effets de k-correction, cependant, les résultats employant la caméra infrarouge du HST (NICMOS) ne portent que sur un nombre réduit d'objets.

L'utilisation de télescopes terrestres pour ce faire est une voie peu explorée essentiellement à cause des problèmes de résolution angulaire et de sensibilité. L'estimation morphologique est en effet un travail complexe lorsqu'il s'agit d'objets éloignés qui requiert une qualité d'image très élevée. Cette thèse explore donc deux voies distinctes d'aborder le problème et montre les résultats scientifiques obtenus avec chacun des deux chemins.

J'ai dans un premier temps, exploré une solution *in vivo* sur le télescope. La génération des télescopes de 8-10m est en effet équipée d'un système d'optique adaptative permettant de corriger les effets de la turbulence atmosphérique sur l'image. L'utilisation de celui-ci pour des applications cosmologiques et en particulier pour des études morphologiques est donc extrêmement séduisante. Nous avons donc analysé un échantillon de 79 galaxies observées avec le système NAOS/CONICA du VLT. Plusieurs éléments peuvent être retenus à l'issue de ce travail. Premièrement, je dirais que les résultats sont prometteurs ; j'ai pu montrer en effet que la qualité sur la morphologie est significativement supérieure à celle obtenue avec les mêmes techniques sur des données non corrigées. J'ai obtenu ainsi une première estimation quantitative de l'évolution morphologique des galaxies (en trois types) jusqu'à  $z \sim 1.2$  à partir de morphologies déterminées directement dans l'infrarouge. Il existe un bon accord avec les nombres estimés à partir des fonctions de luminosité. J'ai quantifié également les effets de k-correction en comparant la morphologie obtenue avec celle estimée sur les mêmes objets avec le télescope spatial. On observe un certain nombre de différences à grand  $z$ , mais une conclusion définitive sur la nature de celle-ci requiert un nombre d'objets plus important. Cependant, il est évident ; à la vue des difficultés rencontrées que ces applications poussent la technique à ses limites. Mais n'est-ce pas de cela dont il s'agit ? Est-il envisageable de faire des progrès scientifiques sans atteindre les limites ? J'aurais bien voulu approfondir cette étude en observant donc un échantillon plus important grâce à une étoile laser en particulier. J'ai été malheureusement confronté à un problème bien connu de toute la communauté : les énormes difficultés pour obtenir du temps d'observation. C'est

d'ailleurs une autre conclusion que j'aimerais faire ressortir. Pour que l'optique adaptative puisse être compétitive dans le champ de la cosmologie observationnelle, il est indispensable de pouvoir observer un échantillon significatif, sachant que les relevés actuels comportent quelques centaines de milliers d'objets. Les ressources planétaires étant extrêmement limitées cela devient un problème fondamental. De ce point de vue, les nouvelles techniques d'optique adaptative (MOAO, MCAO) pourront probablement apporter des solutions en offrant, entre autres, la possibilité d'observer des champs plus grands.

Ces difficultés pour continuer le travail initialement prévu m'ont poussé à réfléchir à une deuxième voie pour obtenir une estimation morphologique fiable dans le proche infrarouge. Il s'agit d'une voie *à posteriori* qui essaye en effet d'améliorer les techniques actuelles d'estimation morphologique afin d'extraire un maximum d'information de données non corrigées. Jusqu'à présent, malgré un effort pour rendre les analyses morphologiques plus objectives et automatiques, toutes les approches existantes requièrent une intervention humaine importante. J'ai ainsi apporté deux nouveaux éléments aux méthodes non-paramétriques : d'une part j'ai résolu le problème de l'étalonnage de l'espace de paramètres en employant un échantillon local adapté aux conditions d'observations et d'autre part j'ai développé une méthode permettant d'employer un nombre illimité de paramètres simultanément. J'ai pu prouver que l'utilisation de ces deux éléments permet d'améliorer significativement la qualité de la morphologie (en deux grands types) sur des galaxies lointaines observées avec une résolution dictée par les conditions atmosphériques. Un effort particulier a été réalisé pour rendre cette approche disponible à toute la communauté. Ainsi, le code complet (galSVM) a été rendu publique afin qu'il puisse être utilisé pour d'autres applications. (<http://www.lesia.obspm.fr/~huertas/galsvm.html>).

Dans un deuxième temps j'ai pu employer le code développé pour étudier l'évolution morphologique de  $\sim 50\,000$  galaxies dans le champ COSMOS à partir d'imagerie infrarouge (WIRCam). La classification a été réalisée dans un espace à 10 paramètres. Nous avons associé à chaque galaxie ayant  $K_s < 21.5$  et  $z < 2$ , une probabilité comprise entre 0 et 1 d'appartenir à un type morphologique donné. Cette valeur de probabilité a été employé pour quantifier la fiabilité des classifications. La probabilité moyenne des objets de l'échantillon est de  $p \sim 0.8$ . Elle décroît avec le *redshift* et la taille (en particulier pour les types précoces) mais reste néanmoins autour de  $p \sim 0.7$  même à  $z \sim 2$ . La classification obtenue a été donc employée pour analyser les comptages par type morphologique ainsi que les distribution en *redshift* de chaque type. On observe une augmentation de la fraction de galaxies elliptiques depuis  $z \sim 2$  ( $\sim 20\%$ ) jusqu'à nos jours ( $\sim 30\%$ ) ce qui traduit probablement une construction progressive de la séquence rouge à partir de galaxies plus tardives. La classification est globalement en bon accord avec celle obtenue sur les mêmes objets à partir de données HST/ACS pour des galaxies situées à  $z < 1$ . Au delà de ce *redshift*, la classification dans la bande I a tendance à trouver moins de

galaxies elliptiques d'un facteur  $\sim 1.5$  ce qui peut être une conséquence des effets de *k - correction*.

Les résultats exposés ci-dessus ouvrent un grand nombre de possibilités pour poursuivre l'analyse. Ces tendances doivent être en effet vérifiées par une étude approfondie de la population de galaxies elliptiques à  $z \sim 1.5$ . Les données multi-longueur d'onde du relevé COSMOS permettent d'étudier leurs masses et leurs histoires de formation stellaire. Ce travail est en cours de développement au moment de la rédaction de ce manuscrit. De plus, la classification présentée peut être également utilisée pour analyser des indicateurs statistiques comparables aux modèles d'évolution comme les fonctions de luminosité par type, l'évolution des tailles de chaque type morphologique, la relation entre la morphologie et la densité en fonction du *redshift*.

D'autre part, la méthode développée est entrain d'être employée pour d'autres applications ce qui ouvre encore de nouvelles voies pour continuer le travail de cette thèse. Dans le cadre du consortium ALHAMBRA j'essaye d'exploiter le grand nombre de bandes disponibles pour obtenir une morphologie au repos jusqu'à  $z \sim 1$ . De même, je l'ai appliquée à un amas de galaxies situé à  $z \sim 0.5$  afin d'étudier les relations densité-morphologie ainsi que d'autres indicateurs statistiques. J'ai pu prouver que la morphologie obtenue est comparable à celle délivrée par le télescope spatial et un travail sur un échantillon plus important est donc en cours.

Tout ceci pour dire qu'il existe encore un grand nombre de domaines d'application et que malgré le fait que je me vois obligé à mettre un point final au travail de thèse proprement dit, de nombreuses portes restent ouvertes à court et moyen terme .



# Conclusiones (español)

El objetivo de esta tesis doctoral es estudiar las posibilidades actuales para cuantificar las morfologías de galaxias distantes ( $z \sim 1$ ) en el infrarrojo cercano mediante el uso de telescopios terrestres. La gran mayoría de los estudios morfológicos existentes en la actualidad se basan en datos obtenidos con el telescopio espacial Hubble en el óptico. Los *surveys* cosmológicos recientes empiezan sin embargo a estudiar objetos mas allá de  $z \sim 1$  y requieren por consiguiente una morfología estimada en el infrarrojo par evitar sesgos ligados a la k-corrección morfológica.

El uso de telescopios terrestres para dicho trabajo es una vía poco explorada hoy en día a causa esencialmente de la perdida de resolución angular provocada por la atmósfera terrestre. La estimación de la morfología de galaxias lejanas requiere en efecto una calidad de imagen importante tanto desde el punto de vista de la resolución angular como de la sensibilidad. En esta tesis doctoral se han explorado dos vías distintas para abordar el problema y se han mostrado los resultados científicos obtenidos con cada una de ellas.

En un primer lugar, hemos explorado una solución *in vivo* en el telescopio. La generación de telescopios de 8-10m está ya equipada con sistemas de óptica adaptativa que permiten corregir los efectos de la turbulencia atmosférica sobre la imagen creada. Se han obtenido datos de 79 objetos en la banda Ks con el sistema NAOS/CONICA del VLT. Los primeros resultados son esperanzadores. Mediante simulaciones exhaustivas hemos demostrado que la calidad de la morfología obtenida es significativamente superior a la obtenida con datos no corregidos y comparable en ciertos rangos de luminosidad a la obtenida con el telescopio espacial. La información morfológica obtenida se ha empleado para obtener una primera estimación de la evolución de las morfologías de los tres tipos principales desde  $z \sim 1.2$  y para obtener una primera medida de los efectos de la k-corrección morfológica gracias a la comparación con datos del telescopio espacial. Se observan ciertas diferencias, en particular una tendencia de los datos infrarrojos a encontrar morfologías más tempranas, sin embargo para establecer una conclusión definitiva sobre la naturaleza exacta de dichas diferencias requeriría un número más importante de objetos. Aumentar la muestra de forma significativa, es una tarea difícil con las técnicas actuales de óptica adaptativa. Como hemos mostrado, la variación espacial de la respuesta al impulso así como la imposibilidad de observar grandes áreas, limitan mucho el campo de acción. Pese a ello, se ha intentado emplear una de las nuevas posibilidades ofrecidas por el VLT, como es el empleo de una estrella láser como fuente de referencia, par obtener cerca de 800 objetos. Nos hemos visto confrontados, sin embargo, a uno de los problemas importantes de la comunidad astronómica : las enormes dificultades existentes para obtener tiempo de observación. Se trata de

hecho, de otra conclusión que me gustaría recalcar ; para que la óptica adaptativa llegue a ser competitiva en el campo de la cosmología, es indispensable poder obtener una muestra comparable a las obtenidas en los *surveys* recientes (i.e. cientos de miles de objetos). Con los recursos actuales, el tiempo necesario es demasiado importante. Desde este punto de vista, las nuevas generaciones de óptica adaptativa (MOAO, MCAO) podrán probablemente aportar soluciones como por ejemplo la posibilidad de observar grandes áreas con una corrección uniforme.

Dichas dificultades para continuar con el trabajo inicialmente previsto, nos llevaron a reflexionar sobre una segunda vía para obtener morfologías robustas en el infrarrojo cercano. Se trata de un camino *a posteriori* que busca mejorar las técnicas actuales de estimación no paramétrica con el objetivo de extraer la máxima cantidad de información de datos no corregidos. Hasta ahora, pese a los esfuerzos realizados por la comunidad científica para dotar de un carácter automático y objetivo a los estudios morfológicos, la mayoría de los métodos existentes requieren una intervención humana importante. Se han aportado dos modificaciones significativas a los métodos no-paramétricos. En primer lugar, se ha resuelto el problema de la calibración del espacio de parámetros mediante el uso de una muestra local que se adapta de tal forma que reproduzca las condiciones físicas y observacionales de la muestra que se desea analizar. En segundo lugar, se han generalizado dichos métodos para que puedan emplear un número ilimitado de dimensiones así como fronteras no lineales para la separación de regiones. Se ha demostrado a continuación que la inclusión de dichas innovaciones permite mejorar la fiabilidad de la morfología obtenida (en dos grandes tipos) sobre galaxias lejanas observadas a través de la atmósfera terrestre. Una parte del tiempo de trabajo ha sido consagrada a poner el código fuente (GalSVM) a disposición de la comunidad científica (<http://www.lesia.obspm.fr/~huertas/galsvm.html>).

El método propuesto ha sido empleado para el estudio de la evolución morfológica de 50 000 galaxias en el campo del *survey* COSMOS observadas en el infrarrojo cercano con la cámara WIRCam del *Canada France Hawaii Telescope*. Se ha empleado un espacio de 10 dimensiones que incluye 5 parámetros morfológicos para el análisis. A cada galaxia con  $K_s < 21.5$  y  $z < 2$  se le ha asociado una probabilidad entre 0 y 1 de ser tardía o precoz. Dicho valor ha sido empleado para medir la fiabilidad de la clasificación obtenida en función de parámetros físicos de la galaxia como la magnitud, el tamaño o el *redshift*. La clasificación es fiable hasta  $z \sim 2$  : la probabilidad media de los objetos es de  $p \sim 0.8$ . Se observa un decrecimiento en función del desplazamiento al rojo, más importante en las galaxias precoces, pero la probabilidad media se mantiene en torno a  $p \sim 0.7$ . Se observa pues un aumento de la fracción de galaxias elípticas desde  $z \sim 2$  ( $\sim 20\%$ ) hasta hoy en día ( $\sim 30\%$ ) que refleja probablemente una construcción progresiva de la secuencia roja a partir de galaxias tardías. Cuando se compara la clasificación con la obtenida a partir de datos del telescopio espacial en la banda I, se observa que las dos clasificaciones están

globalmente de acuerdo hasta  $z \sim 1$ . Más allá, cuando la banda I empieza a sondear la emisión ultravioleta de la galaxia, la estimación basada en la banda Ks tiende a encontrar  $\sim 1.5$  veces más galaxias precoces. Esta diferencia es probablemente un reflejo de los efectos de la k-corrección morfológica.

Estos últimos resultados abren un cierto número de puertas para proseguir con el trabajo. En particular, la gran cantidad de datos disponibles en el campo de COSMOS permite un estudio detallado de la población de galaxias elípticas identificada a  $z \sim 1.5$  mediante el análisis de las masas y de la actividad de formación estelar. Dicho trabajo está en curso de desarrollo. Además, la clasificación obtenida puede ser empleada para obtener otros indicadores estadísticos directamente comparables a los modelos de evolución, como pueden ser funciones de luminosidad por tipo o la relación morfología–densidad en función del desplazamiento al rojo.

En el momento de redactar la presente tesis doctoral, siguen abiertos además una serie de trabajos. En particular, se está empleando el método desarrollado para obtener las morfologías de galaxias en reposo hasta  $z \sim 1$  en el marco del *survey* ALHAMBRA, aprovechando así la gran cantidad de filtros disponibles y su alto grado de homogeneidad. De igual modo, se ha iniciado un estudio en cúmulos situados en torno a  $z \sim 0.5$ . El estudio de MS 1621.5+2640 ha permitido mostrar que la calidad de la morfología obtenida en estos objetos es similar a la obtenida con el telescopio espacial, lo que abre las puertas al estudio de una muestra más importante.

Con todo ello, me gustaría resaltar, que pese a que me veo obligado a poner punto final a la presente tesis doctoral, siguen abiertas un número importante de vías para continuar el trabajo a corto y medio plazo.



# Conclusions (english)

I have tried to explore different possibilities for quantifying morphologies of distant galaxies ( $z \sim 1$ ) in the near-infrared using ground-based facilities. Most of the studies up to date are based on visible HST data. Recent surveys start however to probe objects beyond  $z \sim 1$  and require therefore morphologies estimated in redder wavelengths in order to avoid morphological k-correction effects. The space telescope has a NIR camera (NICMOS) but the morphological studies published up to date involve few objects or galaxies in clusters.

The use of ground-based telescopes for such a work has not been too much studied, essentially because of the poor angular resolution obtained when the light crosses the atmosphere. Properly quantifying morphologies of high-redshift galaxies requires indeed high-angular resolution as well as high sensibility. In this thesis, we have nevertheless explored two different ways to tackle the problem and we have shown the scientific results obtained with both methods.

Firstly, we have explored a *real time* solution. The 8-10m class of telescopes are indeed equipped with adaptive optics systems enabling a correction of the effects caused by the atmospheric turbulence. We have therefore obtained data of 79 objects with the NAOS/CONICA system installed at the VLT. First results are quite encouraging. We have shown through extensive simulations that the accuracy of the morphological estimate is significantly higher than the one obtained on seeing limited images and comparable to what is obtained with space data. The estimated morphologies have also been employed to look at the morphological evolution from  $z \sim 1.2$  for the three main morphological types as well as to quantify the morphological k-correction effects. We observe a number of differences between the morphologies estimated using HST imaging in the I-band and the ones obtained in the Ks-band. However, the small number of objects does not enable to conclude safely about the nature of these differences. Unfortunately, increasing the sample with the existing techniques is a hard task : the PSF spatial variation implies that only small areas can be observed, so important exposure times are required. We have tried nevertheless to increase our sample using a laser guide star by proposing a large programme to ESO. Despite, the encouraging comments of the OPC, the proposal has been rejected. This leads to another interesting conclusion : in order adaptive optics to be competitive in the field of observational cosmology, it must be able to observe large areas and obtain samples comparable to the ones obtained in recent surveys, i.e. hundred thousand galaxies. This is not possible with the techniques available today in a reasonable amount of observing time. In this sense, the new generations of adaptive optics (MCAO, MOAO) should help to solve this main issue.

The difficulties found when trying to continue with the previous work made me think about alternative ways to obtain robust morphologies in the NIR. We therefore tried to modify the existing non-parametric methods in order to extract more information from seeing-limited images. The proposed method, can be seen as a generalization of the classical C/A classification but with an unlimited number of dimensions and non-linear boundaries between decision regions. It is fully automated and thus particularly well adapted to large cosmological surveys. The calibration of the parameter space is done with a local visually classified sample, chosen in the high-redshift sample's rest-frame and artificially redshifted to match the observing conditions. We have shown that a qualitative separation in two main morphological types (late type and early type) can be obtained with an error lower than 20% whereas a classical C/A classification on the same sample leads to errors around 40%. The method is optimized to solve a specific problem, offering an objective and automated estimate of errors that enables a straightforward comparison with other surveys. Selecting the training sample in the high-redshift sample rest-frame makes the results free from wavelength dependent effects and hence its interpretation in terms of evolution easier. The source code is available for download at <http://www.lesia.obspm.fr/~huertas/galsvm.html>.

The proposed method has been used to classify  $\sim 50\,000$   $K_s$  selected galaxies in the COSMOS area observed with WIRCam at CFHT. We used a 10-dimensional volume, including 5 morphological parameters, and other characteristics of galaxies such as luminosity and redshift. The obtained classification has been used to investigate the redshift distributions and number counts per morphological type up to  $z \sim 2$  and to compare to the results obtained with HST/ACS in the I-band on the same objects from other works. We have associated to every galaxy with  $K_s < 21.5$  and  $z < 2$  a probability between 0 and 1 of being late-type or early-type. This value has been used to assess the accuracy of our classification as a function of physical parameters of the galaxy and to correct from classification errors. The classification is found to be reliable up to  $z \sim 2$ . The mean probability is  $p \sim 0.8$ . It decreases with redshift and with size, especially for the early-type population but remains above  $p \sim 0.7$ . We observe an increase of the early-type population from  $z \sim 2$  ( $\sim 20\%$ ) to the present ( $\sim 30\%$ ) probably reflecting a progressive building up of the red sequence from late-type objects. The classification is globally in good agreement with the one obtained using HST/ACS for  $z < 1$ . Above  $z \sim 1$ , the I-band classification tends to find less early-type galaxies than the  $K_s$ -band one by a factor  $\sim 1.5$  which might be a consequence of morphological k-correction effects. We argue therefore that studies based on I-band HST/ACS classifications at  $z > 1$  could be underestimating the elliptical population.

The results presented above offer a large number of possibilities to continue working. First, a confirmation by a precise study of the population of elliptical galaxies at  $z \sim 1.5$  by studying their masses and their star forming activity is required.

---

This work is in progress using the multiwavelength data available from the COSMOS survey. Moreover, the Ks-band classification is intended as a framework for future studies of the evolution of counts, luminosities, luminosity densities, correlation function for each morphological type over several redshift bins.

Moreover, galSVM is being used for other applications. In the framework of the ALHAMBRA collaboration, I am trying to use the amazing amount of available filters together with their homogeneity to obtain rest-frame morphologies up to  $z \sim 1$ . I am also working on clusters of galaxies at intermediate redshift observed with the CFHT. Thanks to the work on MS 1621.5+2640 I have shown that the first results are in good agreement with the ones obtained with the space telescope. Therefore, I am thinking of significantly increasing the sample.

Consequently, despite the fact that I have to put an end to this PhD thesis, a number of projects stay opened offering several possibilities to continue the work.



# *Redshifts* photométriques

---

Les *redshifts* photométriques sont une estimation du *redshift* d'une galaxie en employant uniquement des données photométriques (filtres larges) au lieu de faire de la spectroscopie. Les premières applications datent des années 60, mais ce n'est que récemment avec la prolifération des relevés de galaxies offrant un nombre trop grand d'objets pour réaliser des observations spectroscopiques dans un temps raisonnable que la technique a pris de l'ampleur.

La méthode est basée essentiellement sur l'identification de sauts majeurs (*breaks*) dans la distribution d'énergie spectrale de la galaxie (SED) qui restent toujours visibles même après convolution par la fonction de transmission des filtres. La précision de l'estimation est bien entendu moins bonne que si des données spectroscopiques étaient employées ( $\Delta z/(1+z) \sim 0.1$ ) mais pour la plupart des applications cosmologiques elle s'avère suffisante. C'est pour cela que son développement a été très rapide durant les dernières années et est devenu maintenant un outil essentiel dans le domaine de la cosmologie observationnelle. Il existe donc dans l'actualité plusieurs codes en libre accès pour la communauté permettant de réaliser le travail (e.g. HyperZ, LePhare).

## A.1 Généralités sur la méthode

Il existe plusieurs méthodes permettant de réaliser l'estimation. Une des plus répandues et qui est utilisée dans le cadre du consortium COSMOS c'est la méthode d'ajustement de distributions spectrales d'énergie (*SED fitting*). Il s'agit essentiellement d'ajuster la SED observée avec une librairie de spectres de référence (construits en employant les mêmes filtres que pour les observations). Les spectres de référence peuvent être théoriques ou observés. Dans le cas du code *Le Phare* employé pour l'estimation des *redshifts* dans ce manuscrit, les spectres employés sont issues de la librairie GISSEL.

Le *redshift* photométrique d'une galaxie est alors le résultat du meilleur ajustement entre le spectre observé et les spectres de référence :

$$\chi^2(z) = \sum_{i=1}^{N_{filters}} \left[ \frac{F_{obs,i} - b \times F_{temp,i}(z)}{\sigma_i} \right]^2 \quad (\text{A.1})$$

avec  $F_{obs,i}$ ,  $F_{temp,i}$  et  $\sigma_i$  étant respectivement les flux observés, les flux des spectres de référence et les incertitudes des mesures. La qualité de l'estimation obtenue dépend du nombre de filtres employés.

## A.2 COSMOS

Les *redshifts* photométriques employés tout le long de ce manuscrit ont été estimés dans le cadre du consortium COSMOS. Les mesures photométriques permettant l'ajustement des distributions d'énergie ont été réalisées dans des ouvertures de 3" sur 15 filtres différents :

- Les filtres  $B_j, V_j, g+, r+, i+, z+$  du télescope Subaru ;
- Les filtres  $u*, i*$  du CFHT ;
- Les filtres  $u', g', r', i', z'$  su relevé SDSS ;
- Le filtre  $K_s$  du KPNO/CTIO ;
- Le filtre  $F816W$  du télescope spatial Hubble.

Afin d'améliorer les performances, un étalonnage en employant 1095 *redshifts* spectroscopiques a été réalisé.

La qualité de l'estimation se mesure généralement en calculant l'écart moyen entre le *redshift* calculé et la mesure spectroscopique lorsque celle-ci est disponible. Dans le cas de COSMOS, la précision obtenue est :  $\sigma_{\Delta z}/(1 + z_s) = 0.031$ .

Une autre grandeur importante, permettant de quantifier la qualité, est la fraction d'*identifications catastrophiques*, c'est-à-dire des estimations qui diffèrent significativement de la mesure spectroscopique. Dans notre cas, les cas catastrophiques sont définis comme ceux présentant un écart de :  $\Delta z/(1 + z_s) > 0.15$ . La fraction mesurée est de  $\eta = 1.0\%$ .

# Morphologies des 79 galaxies observées avec NACO

---

Je présente ici un tableau résumant les résultats obtenus sur les galaxies observées avec NACO (chapitre 4). Pour chaque objet il est indiqué les magnitudes I et K, le rayon contenant la moitié du flux ainsi que les morphologies obtenues avec les méthodes paramétriques et non-paramétriques. Pour la deuxième on montre les résultats dans le deux bandes (I et K).

Object ID	RA	DEC	K <sub>s</sub>	I	rh(arcsec)	ZPHOT	G2D(K)	C-A(K)	C-A(I)
NHzG J100016.4+021555	150.069	2.26541	20.1822	99.5000	0.296730	99.9000	DD	DD	N/A
NHzG J100016.3+021643	150.068	2.27872	20.7045	22.4291	0.333504	0.711700	DD	DD	I
NHzG J100017.2+021637	150.072	2.27703	22.8723	24.1406	0.205416	0.247200	DD	DD	DD
NHzG J100016.3+021631	150.068	2.27552	19.7637	24.2207	0.566514	1.78470	DD	DD	I
NHzG J100015.0+021629	150.063	2.27500	18.7053	99.5000	0.489564	99.9000	DD	DD	N/A
NHzG J100014.7+021630	150.061	2.27509	22.1654	23.3978	0.209682	0.832200	BD	DD	DD
NHzG J100014.8+021629	150.062	2.27487	22.0163	23.7523	0.169614	0.793700	I	I	DD
NHzG J100014.8+021629	150.062	2.27478	21.3202	99.5000	0.292356	99.9000	I	I	N/A
NHzG J100015.9+021629	150.067	2.27476	21.7583	22.6455	0.246294	0.0400000	BD	DD	DD
NHzG J100017.2+021624	150.072	2.27355	20.0909	23.4449	0.243324	0.956100	BD	DD	DD
NHzG J100017.3+021620	150.072	2.27223	20.2320	99.5000	0.403920	99.9000	I	I	N/A
NHzG J100014.8+021624	150.062	2.27360	19.8198	99.5000	0.231174	99.9000	BD	BD	N/A
NHzG J100016.7+021618	150.070	2.27174	19.0273	20.9238	0.640170	0.219100	DD	DD	DD
NHzG J100016.7+021610	150.070	2.26963	20.3879	22.7999	0.271026	0.740000	BD	DD	BD
NHzG J100014.9+021607	150.062	2.26870	18.5753	21.8364	0.392418	0.875600	BD	BD	DD
NHzG J100014.9+021603	150.062	2.26755	22.2439	22.7490	0.235656	0.475700	BD	DD	BD
NHzG J100017.3+021601	150.072	2.26712	20.0121	23.8199	0.266760	1.33790	DD	DD	I
NHzG J100053.2+021934	150.222	2.32621	18.5348	21.3553	0.531684	0.886100	DD	DD	DD
NHzG J100051.9+022011	150.216	2.33658	22.6880	22.6407	0.183600	0.810100	BD	DD	DD
NHzG J100052.0+022008	150.217	2.33579	19.2141	21.9404	0.250614	0.832600	DD	BD	BD
NHzG J100051.8+022006	150.216	2.33511	21.9112	23.6353	0.254880	0.818800	DD	DD	I
NHzG J100050.8+022002	150.212	2.33398	20.2005	21.8439	0.471420	0.514000	DD	DD	DD
NHzG J100050.8+022000	150.212	2.33337	17.6825	19.9774	0.690012	0.400600	DD	DD	DD
NHzG J100051.9+021944	150.217	2.32890	18.7411	20.8672	0.396252	0.670000	DD	DD	DD
NHzG J100051.0+021942	150.213	2.32853	18.3509	20.8194	0.614520	0.739400	DD	DD	DD
NHzG J100052.4+021941	150.219	2.32814	20.8822	22.7814	0.404622	0.799500	I	I	DD

TAB. B.1: continued.

Object ID	RA	DEC	Ks	I	rh(arcsec)	ZPHOT	G2D(K)	C-A(K)	C-A(I)
NHzG J100011.1+020629	150.047	2.10808	17.7249	19.9124	0.348300	0.506000	BD	BD	DD
NHzG J100010.0+020624	150.042	2.10673	19.5877	22.7680	0.200826	0.987200	DD	BD	DD
NHzG J100011.6+020617	150.048	2.10486	19.5207	21.0617	0.346680	0.270200	DD	DD	DD
NHzG J100010.2+020612	150.043	2.10360	20.3156	99.5000	0.155034	99.9000	DD	DD	N/A
NHzG J100010.6+020612	150.045	2.10348	20.5515	99.5000	0.167130	99.9000	BD	DD	N/A
NHzG J100009.7+020610	150.041	2.10279	21.2964	99.5000	0.178092	99.9000	DD	DD	N/A
NHzG J095951.4+020514	149.964	2.08732	20.8670	21.1830	0.379188	0.671400	I	I	DD
NHzG J095953.9+020507	149.975	2.08549	21.8739	23.2931	0.170046	0.703700	DD	DD	DD
NHzG J095952.3+020459	149.968	2.08319	19.9646	99.5000	0.358668	99.9000	DD	DD	N/A
NHzG J095952.3+020458	149.968	2.08286	21.7629	99.5000	0.202284	99.9000	BD	DD	N/A
NHzG J095952.3+020448	149.968	2.08006	21.3778	99.5000	0.178254	99.9000	BD	DD	N/A
NHzG J095953.0+020446	149.971	2.07970	20.2794	23.9269	0.192024	0.754600	DD	DD	I
NHzG J095951.5+020444	149.965	2.07890	21.1099	23.6728	0.160164	1.37440	DD	BD	DD
NHzG J100013.8+020838	150.058	2.14397	22.4294	22.7827	0.185220	0.240300	DD	BD	DD
NHzG J100015.3+020923	150.064	2.15644	19.5222	21.1499	0.395604	0.979400	BD	BD	DD
NHzG J100013.9+020920	150.058	2.15569	21.0909	99.5000	0.189000	99.9000	BD	DD	N/A
NHzG J100014.0+020919	150.058	2.15551	21.2097	23.5849	0.247158	0.994000	BD	DD	DD
NHzG J100014.1+020918	150.059	2.15517	19.5826	23.5148	0.193482	1.26360	BD	BD	DD
NHzG J100015.1+020915	150.063	2.15427	23.0568	99.5000	0.139968	99.9000	BD	DD	N/A
NHzG J100016.3+020912	150.068	2.15348	21.6753	23.0442	0.156600	0.935900	DD	DD	DD
NHzG J100013.9+020909	150.058	2.15258	21.4149	22.7981	0.306612	1.08450	DD	DD	DD
NHzG J100013.9+020903	150.058	2.15105	21.4370	23.5968	0.185112	1.11800	BD	DD	DD
NHzG J100015.4+020854	150.064	2.14850	20.9069	23.9274	0.283014	1.16300	DD	DD	I
NHzG J100016.1+020854	150.067	2.14840	22.1741	99.5000	0.168156	99.9000	BD	DD	N/A
NHzG J100003.4+020711	150.015	2.11984	20.5216	22.3691	0.432864	0.714900	DD	DD	DD

TAB. B.1: continued.

Object ID	RA	DEC	Ks	I	rh(arcsec)	ZPHOT	G2D(K)	C-A(K)	C-A(I)
NHzG J100003.6+020706	150.015	2.11842	19.0519	22.4869	0.408024	1.17460	DD	BD	DD
NHzG J100001.7+020704	150.007	2.11780	17.0060	18.8229	0.862596	0.338100	DD	DD	DD
NHzG J100002.5+020701	150.011	2.11709	20.4333	99.5000	0.220644	99.9000	DD	DD	N/A
NHzG J100000.9+020701	150.004	2.11697	21.7613	22.4334	0.235710	0.787500	I	I	DD
NHzG J100003.2+020700	150.013	2.11682	18.8483	22.2691	0.440532	1.06140	DD	DD	I
NHzG J100002.6+020659	150.011	2.11655	17.6231	99.5000	0.397062	99.9000	DD	BD	N/A
NHzG J100000.9+020652	150.004	2.11461	19.4302	20.6107	0.372168	0.698200	BD	DD	DD
NHzG J100002.1+020650	150.009	2.11392	18.8301	21.3752	0.445932	0.689900	DD	DD	DD
NHzG J100003.1+020648	150.013	2.11341	20.3555	21.9584	0.402084	0.422300	DD	DD	I
NHzG J100003.1+020642	150.013	2.11180	21.3786	23.4545	0.283662	0.770500	I	I	DD
NHzG J100002.2+020641	150.009	2.11151	20.8944	21.5886	0.404406	0.334700	I	I	DD
NHzG J100002.1+020635	150.009	2.10981	21.4976	99.5000	0.294678	99.9000	DD	DD	N/A
NHzG J095956.8+020423	149.987	2.07308	19.8462	23.6537	0.332424	1.00700	DD	DD	DD
NHzG J095956.7+020421	149.987	2.07260	22.6499	22.7361	0.321570	1.10000	DD	DD	I
NHzG J095955.6+020420	149.982	2.07236	20.8560	22.4354	0.202986	0.689900	DD	BD	BD
NHzG J095955.6+020419	149.982	2.07204	21.3234	22.7710	0.204984	0.624900	DD	DD	BD
NHzG J095956.4+020416	149.985	2.07137	20.8485	22.3323	0.404082	0.0400000	DD	DD	BD
NHzG J095956.4+020412	149.985	2.07027	19.3325	99.5000	0.389070	99.9000	DD	BD	N/A
NHzG J095955.8+020412	149.983	2.07019	22.9091	23.4010	0.194778	0.525700	I	I	BD
NHzG J095956.8+020406	149.987	2.06838	20.3433	23.9523	0.259470	1.54810	BD	BD	DD
NHzG J095955.8+020407	149.983	2.06874	18.9586	22.0782	0.264924	1.20850	DD	BD	DD
NHzG J095955.7+020404	149.982	2.06801	20.4066	23.8483	0.391338	1.48830	I	I	DD
NHzG J095956.2+020401	149.984	2.06707	21.2254	99.5000	0.249372	99.9000	DD	DD	N/A
NHzG J095955.8+020358	149.983	2.06630	21.2308	23.8229	0.434916	0.881200	DD	DD	I
NHzG J095956.5+020355	149.986	2.06550	18.0002	21.4034	0.622458	0.794100	DD	DD	DD

TAB. B.1: continued.

Object ID	RA	DEC	Ks	I	rh(arcsec)	ZPHOT	G2D(K)	C-A(K)	C-A(I)
NHzG J095956.2+020353	149.984	2.06479	19.3081	21.7327	0.394146	0.668900	DD	DD	DD
NHzG J095956.4+020353	149.985	2.06474	22.1424	23.8900	0.188568	0.743900	DD	DD	DD
NHzG J095957.2+020347	149.989	2.06312	21.8247	23.8017	0.184248	0.214600	DD	DD	DD



ANNEXE C

# Proposition de Large Programme ESO

---



# European Organisation for Astronomical Research in the Southern Hemisphere

Organisation Européenne pour des Recherches Astronomiques dans l'Hémisphère Austral  
Europäische Organisation für astronomische Forschung in der südlichen Hemisphäre

VISITING ASTRONOMERS DEPARTMENT • Karl-Schwarzschild-Straße 2 • D-85748 Garching bei München • e-mail: visas@eso.org • Tel. : +49-89-32 00 64 73

APPLICATION FOR OBSERVING TIME

LARGE PROGRAMME

PERIOD: **81A**

## Important Notice:

By submitting this proposal, the PI takes full responsibility for the content of the proposal, in particular with regard to the names of CoIs and the agreement to act according to the ESO policy and regulations, should observing time be granted

1. Title Category: **A-1**  
The building up of the Hubble sequence of galaxies since  $z \sim 1$  from deep K-band laser guide stars AO observations

## 2. Abstract / Total Time Requested

Total Amount of Time:

Total Number of Semesters:

We propose to study the building up of the Hubble sequence along cosmic time using the exquisite morphological information offered by the unsurpassed sharpness of AO imaging in the K band. The proposal is resubmitted with argued answers to the OPC concerns. We will observe a sample of  $\sim 800$  objects with redshifts up to  $z = 1.5$  in a  $\sim 70(')^2$  area at the unprecedented resolution of  $\sim 0.1''$ , complete to  $K_s = 22(AB = 24)$  within the COSMOS treasury survey area. Thanks to a Pilot Programme on  $\sim 7(')^2$  we were able to prove that classical adaptive optics can provide morphology, in an automated way, at a spatial resolution significantly better than HST-NICMOS, and indeed comparable with HST visible cameras. The K-band has the key advantage to probe old stellar populations at these redshifts, enabling to determine galaxy morphological types unaffected by recent star formation, better linked to the underlying mass than classical optical morphology studies (HST). Combining this data set with publicly available spectroscopic redshifts (zCOSMOS), and multi wavelength information (COSMOS) will enable to study, for the first time, the evolution of counts, luminosities, luminosity densities, correlation function and other evolution estimators for three main morphological types (E/S0,S,Irr), determined in the near-IR, over two redshift bins ( $0.4 < z < 0.8$ ,  $0.8 < z < 1.5$ ). Data will be made public.

3. Run	Period	Instrument	Time	Month	Moon	Seeing	Sky Trans.	Obs.Mode
A	81	NACO	66h	apr	n	$\leq 0.6''$	THN	s
B	82	NACO	66h	mar	n	$\leq 0.6''$	THN	s
C	83	NACO	66h	apr	n	$\leq 0.6''$	THN	s

## 4. Principal Investigator: **M. Huertas-Company** (Paris Observatory, F, marc.huertas@obspm.fr)

Col(s): D. Rouan (Paris Observatory, F), G. Soucail (Toulouse Observatory, F), O. Le Fèvre (Marseille Observatory, F), L. Tasca (Marseille Observatory, F), S. Paltani (Geneve Observatory, CH), B. Rocca-Volmerange (IAP, F), Y. Clénet (Paris Observatory, F), M. Moles (IAA, E), I. Márquez (IAA, E), E. Zucca (Bologna Observatory, I), E. Gendron (Paris Observatory, F), G. Rousset (Paris Observatory, F), E. Thiebaud (Lyon Observatory, F), T. Contini (Toulouse Observatory, F), J.P. Kneib (Marseille Observatory, F), L. Tresse (Marseille Observatory, F), L. Mugnier (ONERA, F), G. Cresci (MPE, D), R. Davies (MPE, D)

## 5. Description of the proposed programme

**A) Scientific Rationale:** The process of galaxy formation and the way galaxies evolve is still one of the most important unresolved problems in modern astrophysics. In the currently popular hierarchical picture of structure formation, galaxies are thought to be embedded in massive dark halos that grow from density fluctuations in the early universe (Fall & Efstathiou 1980, MNRAS, 193, 189) and initially contain baryons in a hot gaseous phase. This gas subsequently cools, and some fraction eventually condenses into stars (Madau et al. 1998, ApJ, 498, 106). However, many of the physical details remain uncertain, in particular the process and history of mass assembly. A central issue is the building up of the Hubble sequence. One classical observational way to answer this question is to classify galaxies according to morphological criteria defined in the nearby Universe and to follow the evolution in numbers and luminosity of different galaxy types across time (Abraham et al. 1996, ApJS, 107, 1 ; Simard et al. 2002, ApJS, 142, 1). Comparison of distant populations with the ones found in the nearby Universe might help to clarify the role of merging that seems to represent the main driver in galaxy evolution (Cole et al. 2000, MNRAS, 319, 168). To this effect, large samples of galaxies are required with a robust estimate of distances, luminosities and morphological types. As a consequence, over the last ten years, progress in understanding faint distant galaxies has come from observing deeper and larger samples, but also from obtaining higher spatial resolution at a given flux and at given redshift. In the visible, progress has been simultaneous on both fronts, thanks in particular to the Hubble Deep Fields (HDF). Deep number counts are obtained up to AB magnitudes of 28, and morphological information up to approximately  $AB = 26$ . In this context, near infrared observations are particularly important because the K band flux at  $z \sim 1$  is less dependent on the recent history of star formation, which peaks in the UV in rest frame, and gives thus a galaxy type from the distribution of old stars, more closely related to the underlying total mass than optical observations. That's why K-band surveys have been carried out using ground-based telescopes in order to perform cosmological tests by means of galaxy counts essentially (e.g. Gardner et al. 1993, ApJ, 415, L9; Maihara et al. 2001, PASJ, 53, 25). In all cases the morphological information is poor because of the seeing-limited spatial resolution. This is the case even with the excellent image quality of WIRCam at CFHT : since effective radii of local galaxies (except for compact dwarf galaxies) range from  $\sim 1.0$  to  $\sim 10 kpc$ , depending on their luminosity (Bender et al. 1992, ApJ, 399, 462; Impey et al. 1996, ApJS, 105, 209), a typical seeing-limited spatial resolution of  $0.8''$  ( $\approx 8 kpc$  at  $z \sim 1$ ), makes impossible the separation between de Vaucouleurs and exponential profiles (Fig. 1). There are today two possibilities to obtain high angular resolution data in the near-infrared: NICMOS on HST or Adaptive Optics on ground based telescopes. Indeed, 7% of the COSMOS field is being observed with NIC3 in the H band ( $1.6\mu m$ ). With an initial pixel size of  $0.2''$ , NIC3 data have been dithered and combined to a pixel size of  $0.1''$  leading to an image resolution of  $0.28''$  (FWHM). On the other side, adaptive optics data from our Pilot Programme (see below for details) have a mean FWHM narrower by a factor 2 ( $0.1''$ ) and a pixel scale of  $0.054''$ . A reduction procedure (Huertas-Company et al., A&A, 468, 937, hereafter PaperI) allows to recover a PSF with a pixel scale of  $0.027''$ , which is 4 times smaller than NIC3 scale. This represents a *critical improvement* since, as shown in the histogram of Fig. 2, the majority ( $\sim 60\%$ ) of the objects observed in our Pilot Programme have half light radii lower than  $0.3''$  and would be consequently hardly analyzed by NICMOS in terms of morphology. As shown below and in Fig. 3, using LGS with bright tip-tilt stars ( $V \sim 16$ ) will also lead to a significant gain over NICMOS data. Adaptive optics installed on ground based telescopes is consequently an excellent complement to NICMOS data and probably the only way today to measure quantitative optical rest-frame morphology for a *complete* sample at  $z \sim 1$ .

**2. Current AO surveys :** Several surveys using AO have been recently conducted:

- Minowa et al. (2005, ApJ, 629, 29) who published the first results, achieved an impressive limiting magnitude of  $K \sim 24.7$  with the Subaru Adaptive Optics system with a total integration time of  $26.8 hr$  over one single field of  $1' \times 1'$ . They proved that the use of adaptive optics significantly improves detection of faint sources but they did not perform morphological analysis.
- In a recently published paper, Cresci et al. (2006, A&A, 458, 385) performed this task with AO data for the first time. They observed a  $15 arcmin^2$  in the  $K_s$  band with NACO (SWAN survey) and classified distant galaxies into two morphological bins (late type and early type) by performing a model fitting with a Sersic law. They conclude that hierarchical models are not consistent with the observed number counts of elliptical galaxies and that PLE models are preferred, showing the unique power of AO observations to study the morphology of faint compact galaxies in the near-IR. However, redshifts of most of the sources are lacking, preventing a more detailed analysis of the evolution of the source parameters as a function of  $z$ .
- **The Pilot Programme:** We made a first attempt to study morphology vs. redshift with NACO (P73.A-0814A and P75.A-0569). Results are described in a recently published paper (PaperI). We observed 7 fields of  $1' \times 1'$  around bright stars in the COSMOS area and obtained a sample of 79 objects with a mean resolution of  $0.1''$ . Specially, we showed, using two independent automated morphological classifications, (GIM2D and Concentration/Asymmetry) that adaptive optics can provide reliable high resolution morphological information in an automated way. Galactic parameters (disk scale length, bulge effective radius and bulge fraction) can be estimated with a random error lower than 20% for the bulge fraction up to  $K = 19$  ( $AB = 21$ ) and classification into three main morphological bins (Bulge Dominated, Disk Dominated, Irregular) can be done up to  $K = 20$  ( $AB = 22$ ) with at least 70% of correct identifications. We then obtained for the first time the galaxy distribution

## 5. Description of the proposed programme (continued)

as measured from the near-infrared at  $z \sim 1$  in three main morphological types (E/S0, S, Irr). The distribution was also compared to the one obtained with HST/ACS in the I band (from COSMOS (Scoville et al. 2005, BAAS, 1309) imaging). We found some differences in the morphologies due to band shifting. In particular there seems to be a larger fraction of early types at high redshift, when derived from the near-infrared. The small number of objects in each redshift bin implies too high uncertainties to conclude in a safe way about the cause of this difference. A reliable study of evolution vs redshift definitely requires more objects.

**2. The proposed survey :** Here we propose to make a significant step forward with respect to the aforementioned surveys. The main goal is to bring new light on the assembly of the Hubble sequence of galaxies from  $z = 1.5$  using laser based AO observations of about 800 galaxies in the K band. To achieve a good estimate of redshift, the fields are selected within the on-going COSMOS survey where a wealth of multi-wavelength data is already available or being obtained. This ensures a superb redshift estimate for all our objects (see *Immediate Objective* for more details). In addition, the zCOSMOS survey (Lilly et al., 2006, Msngr, 121, 42) will provide the spectra of 2/3 of all the galaxies with  $I_{AB} < 22.5$ . Since the redshift distribution of  $17 < K < 22$  galaxies is centered around  $z \sim 0.8$  (Mignoli et al., 2005, A&A, 437, 883) the sample will be separated into 2 redshift bins ( $0.4 < z < 0.8$ ,  $0.8 < z < 1.5$ ). **The scientific goals and the chosen strategy were clearly supported by the OPC in the previous submission:** *"There is a clear need for morphological information for galaxies well above  $z=1$ , and the panel recognises this and feel that the goals of the project are valuable and the COSMOS field is clearly the right field to study in view of the supporting information. This proposal also aims to make good use of the adaptive optics imaging facilities at ESO."* However two main concerns were exposed: a) The first one was about the low redshift bin (*"it is still not clear why the low redshift bin is required given the existence of ACS imaging with considerably higher resolution"*). Indeed, this bin is not needed since the morphological k-correction effect is not relevant at those redshifts and ACS data can therefore be used. That is the reason why we estimated the exposure time considering exclusively the high-redshift bin (see BOX 9). Those low-redshift objects will be in our sample anyway, that's why we plan to analyze them. This will increase the homogeneity and turn the interpretation in terms of evolution easier. Low- $z$  ACS objects can however be also used, as suggested by the OPC, and compare the results to give more robustness to the interpretation. b) The second one was about the reduced Strehl when using a LGS instead of a NGS (*"there is some concern that the significantly reduced Strehl will lead to less good classification and hence considerably reduced gains over the NICMOS data"*). If the Strehl ratio is too low the gain over the NICMOS data can be indeed considerably reduced. In our Pilot Programme we obtained however a mean Strehl ratio of about 30% using  $R \sim 14$  NGS and as shown in Fig 2 this is largely enough to resolve all the objects with sizes lower than 0.3" that couldn't be analyzed with NICMOS. This Strehl ratio can be achieved with LGS and bright tip-tilt stars ( $V \sim 16$ ) as stated in the NACO-LGS commissioning report: *"NACO LGS corrected image quality is of the order 20%-30% K-band Strehl ratio in median to good seeing conditions. LGS AO outperforms NGS AO for stars fainter than about  $R=13$ ."* This is indeed our case. Therefore, combining the accurate redshift information from the COSMOS and z-COSMOS surveys with the objective morphological classification from evolved stellar populations, will allow to tackle these main questions:

- *Evolution of the star formation rate for each morphological type:* Are galaxies still forming between  $z \sim 1$  and today from merging and accretion of smaller galaxies in a hierarchical scenario or is the population already in place at  $z \sim 1$  and evolving passively? The study of the Luminosity Function (LF) for each morphological type (E/S0,S,Irr) vs. redshift is a powerful approach to answer this question since there are now some observational evidences that the history of the star formation strongly depends on both the galaxy mass (Cowie et al. 1996, AJ, 112, 839, L9) and the morphological type (Ilbert et al. 2006, A&A, 453, 809). Indeed, more massive systems appear to have star formation histories that peak at higher redshifts whereas less massive systems peak at progressively lower redshifts and are extended over a longer time interval (*downsizing*). This is consistent with the fact that the population of massive E/S0 galaxies is still in place at  $z \sim 1$  and evolves passively towards lower redshifts. (Rocca-Volmerange et al. 2004, A&A, 415, 931; Conselice et al. 2005, ApJ, 628, 160). Moreover, this result has been recently predicted by means of a large high-resolution simulations. (De Lucia et al. 2006, MNRAS, 366, 499). However, most of these works are based on spectral classifications of galaxies or on UV-rest frame morphologies: The direct interpretation of these results in the framework of galaxy formation scenario is not straightforward since galaxies move from a spectral class to another by a passive evolution of the stellar population: a morphological classification using structural parameters is indeed less sensitive to the band shifting effects than a classification based on the spectral energy distribution. A reliable and complete morphological classification obtained with near-infrared AO observations will thus allow to determine whether there is a true evolution of the star forming rate from  $z \sim 1$  without being biased by band shifting effects.

- *Evolution of the merger rate:* Were galaxies interacting more frequently in the past? Although the evidence for a cold dark-matter dominated universe in which structure grows hierarchically is overwhelming (e.g. Spergel, 2003, ApJS, 148, 175), the role of mergers in galaxy assembly and star-formation remains unclear. Previous

## 5. Description of the proposed programme (continued)

studies of the evolution of the merger rate, carried out based on rest frame UV or B band HST morphologies or with seeing limited observations, have obtained different and still uncertain results (e.g. Conselice et al. 2003, ApJ, 126, 1183; Le Fèvre et al. 2000, MNRAS, 311, 565). The use of AO near-IR imaging in the COSMOS fields will provide an unique sample to measure the evolution of the merger rate, avoiding the strong biases introduced by wavelengths dependent effects as spurious starbursting irregular candidates. Moreover, the unprecedented angular resolution in the near-IR will allow to select merger candidates with much higher completeness, being able to detect later merger stages with respect to previous observations.

- *Evolution of the distribution of morphological types vs. environment:* Do galaxies evolve faster in dense environments? There is a well known connection between galaxy morphology and the local environment wherein galaxies reside. (e.g. Weinmann et al. 2006, MNRAS, 366, 2, Ilbert et al., 2006, arXiv:astro-ph/0602329). Cucciati et al. (2006, A&A, 458, 39) used a sample of  $\sim 6500$  galaxies from the VVDS to investigate the colour-density relation at different epochs. They found that it shows a dramatic change as a function of cosmic time in the sense that the fraction of red galaxies increases as a function of density: at previous epochs ( $z \sim 1$ ) blue and red galaxies seem to be insensitive to the surrounding environment. Another approach is based on the measure of the correlation function for each morphological type. Meneux et al. (2006, A&A, 452, 387) used  $\sim 6000$  spectroscopic redshifts from the VVDS, combined with a classification also based on the spectral type, to conclude that the clustering of the early-spectral-type galaxies is markedly stronger than of late-type at all redshifts up to  $z = 1.2$ . All those results seem to show that the quenching of star formation was more efficient in high density regions. However, this conclusion supposes that redder objects are indeed early type objects. This issue can only be solved with an unbiased morphology from evolved stellar populations.

**B) Immediate Objective:** We propose to combine objective morphological classifications from deep K band AO imaging with accurate redshift estimates to study for the first time the evolution of the fraction of morphological types, the evolution of the star formation rate, the morphology-density evolution, the galaxy sizes and the merger rate evolution between  $z \sim 0$  and  $z \sim 1.5$  for three main morphological types (E/S0,S,Irr). Thanks to the *Pilot Programme* we were able to prove that classical adaptive optics, despite of its limits, is a powerful technique to obtain reliable morphological information in the K-band. We are now able, thanks to the developed utilities to reduce and analyze data with high efficiency. The immediate work will be as follows:

- We will perform *morphological classification* using the bulge fraction estimator. For the brighter objects a quantitative analysis with Gim2D will provide the bulge fraction with a random error lower than 20% even for the more compact ones (PaperI, Fig. 3). For the faintest ones, an improved method based on concentration and asymmetry (GALSVM, Huertas-Company, 2008a, A&A, Huertas-Company et al. 2008b, in prep) will provide a more qualitative classification in 3 main types with about 80% accuracy. Interacting and merging objects will be identified using both a visual approach and one based on asymmetry, Gini and M20 coefficients.

The final classification will be combined with the also public zCOSMOS data to obtain the spectroscopic redshifts of 2/3 of the objects with  $I_{AB} < 22.5$ . For the other objects, multi- $\lambda$  COSMOS data will be used to obtain photometric redshifts with the Le Phare code that gives a fraction of catastrophic errors ( $\Delta z/(1+z) > 0.1$ ) of 1.1% and an accuracy of measurement of  $\sigma_{\Delta z/(1+z)} = 0.031$ . We will this way build a complete catalog of morphological distribution vs. redshift that will be made public as soon as possible.

- In a second step, the sample will be separated in two redshift bins ( $0.4 < z < 0.8$ ,  $0.8 < z < 1.5$ ). We will then measure the *LF and the correlation function for each morphological type and link the morphology to the underlying large scale structures distribution*.

For the LF computation, the ALF code (algorithm for luminosity function) developed by the VVDS LF working group, will be used. The underlying galaxy density field, for the morphology-density study, will be computed from the observed 3D distribution of galaxies using the spectroscopic redshifts (Marinoni et al. 2005, A&A, 442, 801). We will build the evolution of the correlation function for each of the morphological types using the  $w_p(r_p)$  projection of  $\xi(r_p, \pi)$  from existing tools (Pollo et al., 2005, A&A, 439, 887). With the complete sample we expect to measure the correlation length  $r_0$  to about 15% accuracy in each (redshift,type) bin. In order to constrain the luminosity function we need at least 30 objects in each bin (Zucca et al. 2006, A&A, 455, 879). This will constrain the exposure time as shown in BOX 9.

**C) Telescope Justification:** NAOS/CONICA is the only system that can achieve a mean resolution of 0.1" in the  $K_s$  band, which is necessary to estimate morphology. The decision of using NGS with  $V \sim 16$  tip-tilt stars is motivated for several reasons: a) First, it reduces the necessary exposure time to observe a given number of objects (see BOX 9) since there is no area loss due to the natural guide star (NGS) and fields with too few galaxies can be avoided. b) Second, we can expect to recover more spectroscopic redshifts from the zCOSMOS survey than with NGS since regions too close from bright stars are more frequently masked for a spectroscopic survey. c) Third, at least the same correction as with NGS can be achieved as stated in the commissioning report and shown on Fig. 3. d) Fourth the PSF will be more homogeneous.

**D) Observing Mode Justification (visitor or service):** These observations are operationally straightforward, and we only require standard modes.

## 5. Attachments (Figures)

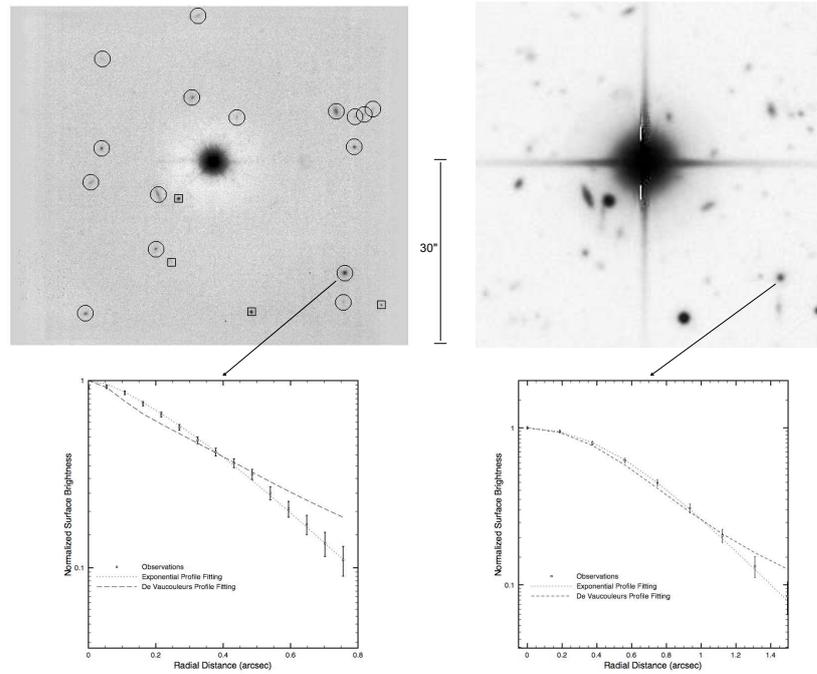


Fig. 1: Image of the field centered at  $\alpha = 10 : 00 : 16, \delta = +02 : 16 : 22$ . **Top Left:** NAOS/CONICA  $K_s$  band image. Total integration time of 10350s. The field size is  $1' \times 1'$  with a pixel scale of  $54mas$ . Circles are detected galaxies and boxes are stars. The stellar FWHM was measured to be  $0.08''$ . The bright star at the center of the image was used as the AO guide star. **Top Right:** Megacam I band image. The pixel scale is  $0.187''$ . The stellar FWHM is  $0.7''$ . **Bottom:** The surface brightness profile was fitted with both a PSF-convolved de Vaucouleurs profile and an exponential profile. With seeing limited observations, it is more difficult to establish a clear separation between both profiles. (see PaperI for details)

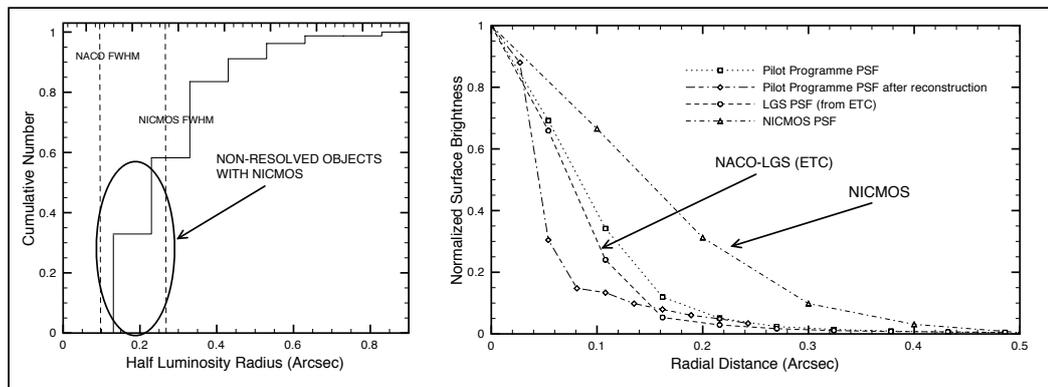


Fig. 2: **Left:** Half light radii cumulative distribution of the 79 galaxies observed in the Pilot Programme. The objects have a median redshift of 0.7. More than 50% of the galaxies have a size lower than  $0.3''$ . **Right:** Comparison between NACO ( $K_s$  band) and NICMOS(H band) PSFs radial profiles. The NICMOS PSF has been taken from a real NICMOS tile from the COSMOS survey. For NACO we plot the real PSF from the Pilot Programme, the recovered PSF after reconstruction (PaperI) and the expected obtained PSF with a Laser Guide Star (from ESO ETC).

## 5. Attachments (Figures)

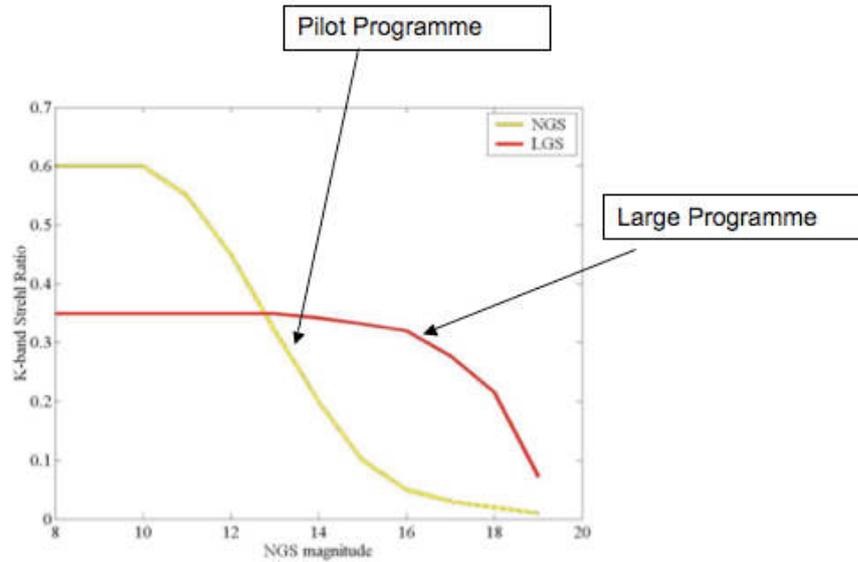


Fig. 3: NACO AO performance in NGS mode (measured) and LGS mode (predicted) for good seeing conditions ( $\sim 0.7''$ ). The measured performance on LGS mode is slightly lower than the predicted one but the commissioning report still recommends the use of LGS when the NGS are fainter than  $R \sim 13$ : "NACO LGS corrected image quality is of the order 20%-30% K-band Strehl ratio in median to good seeing conditions. This performance is in reasonable agreement with the expectations. LGS AO outperforms NGS AO for stars fainter than about  $R=13$ . LGS AO should be used for guide stars fainter than  $R=13.5$ ". The same performance as in our Pilot Programme can therefore be reached since we are planning to use  $V \sim 16$  tip-tilt stars.

## 6. Experience of the applicants with telescopes, instruments and data reduction

The LESIA team has been involved in the building and integration of several adaptive optics systems (ADONIS, PUEO at CFHT, NAOS), and developed a solid expertise in optimizing observations based on this technique and in reducing the data they produced. This is especially true in extragalactic astronomy where several GTO and normal programs were successfully conducted on AGN, Galactic Center and starburst galaxies. More recently, dedicated pieces of software have been developed to reduce the data of the *Pilot Programme*, especially to handle undersampling, recentering and automated extraction of the morphological parameters.

The LAM team and more generally the VIRMOS team, is known as one of the leaders in the field of deep imaging and spectroscopic cosmological surveys. The know-how in producing unbiased and well calibrated catalogs is of course an important part of this experience, specially relevant here.

The LAT team shares somewhat similar capabilities and was involved for several years in programs similar to the one proposed here, using adaptive optics (at CFHT and at ESO) for deep infrared surveys.

The expertise at CRAL and ONERA is mainly in the field of image restoration using a posteriori techniques (MAP), including the myopic approach ; this could be an important added value in this program where the best angular resolution is aimed at, while the actual PSF can only be approximated.

The IAA team members have experience in using AO systems at CFHT and ESO. Presently they are conducting a large survey (the ALHAMBRA project) aimed at determining precise photometric redshifts for a large number of objects.

The MPE group has a large experience in dealing with AO data and space varying PSF (SWAN survey) as well as with SINFONI infrared AO assisted spectroscopy.

Globally, the team involved in this project gather the largest experience in deep wide field extragalactic AO imaging.

## 7. Resources available to the team, such as: computing facilities, research assistants, etc.

Reduction of adaptive optics data require few specific procedures as explained in PaperI, especially because of undersampling. The same software tools, specially developed for that purpose, will be used. They have proven their efficiency to identify morphology with a fair degree of reliability. Moreover an improved method based on learning classification machines is being developed (Huertas-Company et al., 2007, A&A submitted, arXiv:astro-ph/0709.1359, [www.lesia.obspm.fr/~huertas/galsvm.html](http://www.lesia.obspm.fr/~huertas/galsvm.html)). Some improvement is however foreseen, particularly in terms of image restoration using MAP techniques where a regularization term is added in the deconvolution process. This will allow us to deal more accurately with the PSF variations over the field.

Despite the fact that the program is indeed requiring a significant amount of observing time, it doesn't need specially high computation resources. For instance, the last complete processing of the pilot programme has taken a week, using a standard work station, so that the processing of one semester should not take more than one month with the same equipment. If needed, additional resources will be available in the frame of the computing facility at Observatoire de Paris. The other partners will be involved mainly in cross-checking and processing of data sets from other surveys, that will be used for the analysis (COSMOS, z-COSMOS, etc.). With respect to this task, the COSMOS and VVDS team experience is essential. Statistical tools are already available by the VVDS and zCOSMOS investigators.

This research is a central topic in the PhD work of Huertas-Company. 100% of his time will be dedicated to the reduction and analysis of the data. Two post-docs (L.Tasca and G.Cresci) are also involved in this programme. A solid expertise in analysis of this type of surveys is shared by most of the co-Is.

## 8. Special remarks:

In the spirit of the COSMOS Legacy, we will waive all proprietary rights to these data.

## 9. Justification of requested observing time and lunar phase

**Lunar Phase Justification:** There is in principle no need for a limit on the lunar phase because K-band observations do not require such, provided that the fields are located more than 30 degrees from the Moon.

**Time Justification: (including seeing overhead)** Time estimate and instrument configuration has been performed based on our experience with the pilot programme and the obtained results. We propose to use the C54S scale (54 mas pixels). This maximizes the field of view (hence the number of galaxies) even if it implies an undersampling that can be treated with reduction utilities. In order to study the evolution of the LF we need at least 30 objects per bin. Given that we are going to divide galaxies in two redshift bins ( $0.4 < z < 0.8$ ,  $z > 0.8$ ) and in three morphological types (E/S0, S, Irr), the limiting bin will be the irregular population at high redshift, because of the low surface brightness of those objects. Using our results in morphological classification and redshift distribution (PaperI) we deduce that the high  $z$  population represents 43% of the sample among which 3% is expected to be irregular. So, if we want 30 objects in this bin this implies that we need  $N_{total} = 30/(0.43 \times 0.03) \sim 900$  objects in total. (820 given that we have already 79 objects). Based on our previous experience we know that we are able to recover about 11 objects per field (79 objects in 7 observed fields). Thanks to the use of LGS we can expect to recover 15 galaxies per field. We thus need to observe 55 fields. For each field, we will observe for  $DIT = 60sec$  and  $NDIT = 180$  (3 hours) which is necessary to be complete at  $K_s = 22.5$ . Including overheads for readout and offsetting (10 sec per telescope move or 3 minutes in total, for a move every 5 minutes), and 30 minutes for pointing and closing the loop on the star (LGS), this corresponds to a total of 3.6 hours per field. We therefore ask for a total 198 hours of service mode observations to be distributed over three observing periods. (66 hours per period).

**Calibration Request:** Standard Calibration

## 9b. Convert to a normal programme?

Yes. The strategy for this semester would basically remain the same with possibly less fields to observe. A reduced observing time this semester would translate into subsequent proposals on more semesters since the number of objects per bin cannot be reduced if one wants to reach the programme goals.

## 10. Report on the use of ESO facilities during the last 2 years

P73.A-0814A and P75.A-0569A : K imaging of  $7' \times 1'$  fields within the COSMOS area with NACO. Proved that classical adaptive optics is powerful to estimate morphology in an automated way. Results are presented in a paper recently published in A&A (PaperI) P73.B-0665 : NACO measurements of  $4.8 \mu m$  flares from Sgr A\*(Galactic Center). Constraints on the emission mechanism. Clénet et al., 2005, A&A 439, 9. 177.A-0837: VVDS Large Programme (see publications below).

## 11. Applicant's publications related to the subject of this application during the last 2 years

Huertas-Company M., Rouan D. et al., 2007, A&A, 937, Morphological Evolution of  $z \sim 1$  galaxies from deep K-band AO imaging in the COSMOS deep field

Cresci G., Davies R. et al, 2006, A&A, 458, 385: Galaxy morphology and evolution from SWAN adaptive optics imaging

Huertas-Company M., Rouan D., Tasca L., Soucail G., Le Fèvre, O., A&A submitted, arXiv:astro-ph/0709.1359 : A robust morphological classification of high-redshift galaxies on seeing limited images

Zucca E., Ilbert O. et al., 2006, A&A, 455, 879: The VIMOS VLT Deep Survey. Evolution of the luminosity functions by galaxy type up to  $z = 1.5$  from first epoch data

Ilbert O., Lauger S. et al., 2006, A&A, 453, 809: The VIMOS-VLT Deep Survey. Galaxy luminosity function per morphological type up to  $z = 1.2$

Ilbert O., Cucciati C. et al., 2006, A&A, submitted, astro-ph/0602329: Evidence for environment-dependent galaxy Luminosity Function up to  $z=1.5$  in the VIMOS-VLT Deep Survey

Tresse L., Ilbert, O. et al., 2006, A&A, submitted, astro-ph/0609005: The cosmic star formation rate evolution from  $z=5$  to  $z=0$  from the VIMOS VLT Deep Survey

Meneux B., Le Fèvre, O. et al., 2006, A&A, 452, 387: The VIMOS-VLT Deep Survey. The evolution of galaxy clustering per spectral type to  $z=1.5$

Gratadour D., Rouan, D. et al., 2006, A&A, 446, 813: Near-infrared adaptive optics dissection of the core of NGC 1068 with NAOS-CONICA

Le Fèvre O., Paltani, S. et al., 2005, Nature, 437, 519: A large population of galaxies 9 to 12 billion years back in the history of the Universe

Ilbert O., Tresse, L. et al., 2005, A&A, 439, 863: The VIMOS-VLT deep survey. Evolution of the galaxy luminosity function up to  $z = 2$  in first epoch data

## 12. List of targets proposed in this programme

Run	Target/Field	$\alpha$ (J2000)	$\delta$ (J2000)	ToT	Mag.	Diam.	Additional info	Reference star
A	STAR7	10:00:08	+01:47:21	3.6	Ks=22		Vtts=15.32	N20212301460
A	STAR8	09:58:28	+02:28:21	3.6	Ks=22		Vtts=15.33	N20212333
A	STAR9	10:01:36	+01:43:21	3.6	Ks=22		Vtts=15.35	N2021211180
A	STAR10	10:01:52	+02:51:21	3.6	Ks=22		Vtts=15.47	N2021230195
A	STAR11	10:02:08	+02:22:21	3.6	Ks=22		Vtts=15.56	N20212301909
A	STAR12	09:58:24	+02:27:21	3.6	Ks=22		Vtts=15.57	N20212334
A	STAR13	10:00:44	+01:54:21	3.6	Ks=22		Vtts=15.67	N2021230122
A	STAR14	09:58:28	+02:01:21	3.6	Ks=22		Vtts=15.69	N202123390
A	STAR15	10:01:08	+02:06:21	3.6	Ks=22		Vtts=15.71	N202123097
A	STAR16	09:58:36	+02:02:21	3.6	Ks=22		Vtts=15.75	N202123385
A	STAR17	09:59:56	+02:07:21	3.6	Ks=22		Vtts=15.81	N202123364
A	STAR18	09:59:04	+02:22:21	3.6	Ks=22		Vtts=15.90	N202123312
A	STAR19	10:01:04	+02:27:21	3.6	Ks=22		Vtts=15.92	N20212302259
A	STAR20	09:59:00	+01:38:21	3.6	Ks=22		Vtts=15.95	N2021233196
A	STAR21	09:59:16	+02:33:21	3.6	Ks=22		Vtts=15.96	N2021233278
A	STAR22	10:00:28	+02:06:21	3.6	Ks=22		Vtts=16.00	N2021233299
A	STAR23	10:01:36	+01:56:21	3.6	Ks=22		Vtts=16.00	N2021230705
A	STAR24	10:01:00	+02:39:21	3.6	Ks=22		Vtts=16.02	N2021230267
A	STAR25	09:58:36	+02:07:21	3.6	Ks=22		Vtts=16.02	N20212333326
B	STAR26	10:01:28	+01:42:21	3.6	Ks=22		Vtts=16.03	N2021211185
B	STAR27	10:02:32	+02:32:21	3.6	Ks=22		Vtts=16.04	N2021230317
B	STAR28	10:02:08	+02:14:21	3.6	Ks=22		Vtts=16.05	N202123075
B	STAR29	10:01:44	+02:24:21	3.6	Ks=22		Vtts=16.05	N202123029
B	STAR30	09:59:36	+02:51:21	3.6	Ks=22		Vtts=16.06	N2021233255
B	STAR31	09:59:48	+02:31:21	3.6	Ks=22		Vtts=16.06	N2021233282
B	STAR32	10:01:32	+02:32:21	3.6	Ks=22		Vtts=16.06	N2021230314
B	STAR33	10:02:36	+01:37:21	3.6	Ks=22		Vtts=16.07	N2021211871
B	STAR34	09:58:24	+02:01:21	3.6	Ks=22		Vtts=16.11	N202123388
B	STAR35	10:01:40	+02:35:21	3.6	Ks=22		Vtts=16.11	N20212302823
B	STAR36	10:00:20	+02:19:21	3.6	Ks=22		Vtts=16.11	N202123049
B	STAR37	10:01:44	+02:45:21	3.6	Ks=22		Vtts=16.13	N2021230236
B	STAR38	10:02:32	+01:39:21	3.6	Ks=22		Vtts=16.14	N2021211202
B	STAR39	10:00:28	+02:09:21	3.6	Ks=22		Vtts=16.16	N202123090
B	STAR40	09:59:28	+02:04:21	3.6	Ks=22		Vtts=16.19	N20212333075
B	STAR41	09:58:20	+01:41:21	3.6	Ks=22		Vtts=16.20	N2021233183

*Following targets moved at the end of the document ...*

**Target Notes:** Targets have been selected **randomly** selected using the ACS I band COSMOS catalog over the fields containing a star brighter than  $V=17$  for the tip tilt correction. The R magnitude should be slightly brighter since most of the stars are "red".

12b. ESO Archive - Are the data requested by this proposal in the ESO Archive (<http://archive.eso.org>)? If yes, explain why the need for new data.

No

### 13. Scheduling requirements

### 14. Instrument configuration

Period	Instrument	Run ID	Parameter	Value or list
81	NACO	A	IMG 54 mas/px IR-WFS	2.2
82	NACO	B	IMG 54 mas/px IR-WFS	2.2
83	NACO	C	IMG 54 mas/px IR-WFS	2.2

## 15. List of targets proposed in this programme

Run	Target/Field	$\alpha$ (J2000)	$\delta$ (J2000)	ToT	Mag.	Diam.	Additional info	Reference star
<i>...continuing from box 3</i>								
B	STAR42	09:59:08	+01:46:21	3.6	Ks=22		Vtts=16.22	N2021233139
C	STAR43	10:00:48	+02:12:21	3.6	Ks=22		Vtts=16.24	N202123080
C	STAR44	09:58:12	+01:58:21	3.6	Ks=22		Vtts=16.25	N20212332673
C	STAR45	10:01:32	+02:35:21	3.6	Ks=22		Vtts=16.32	N2021230302
C	STAR46	10:02:48	+02:23:21	3.6	Ks=22		Vtts=16.32	N202123038
C	STAR47	09:58:56	+01:42:21	3.6	Ks=22		Vtts=16.33	N202123374
C	STAR48	09:59:56	+02:05:21	3.6	Ks=22		Vtts=16.33	N20212333145
C	STAR49	10:01:40	+01:56:21	3.6	Ks=22		Vtts=16.36	N2021230120
C	STAR50	09:58:28	+02:38:21	3.6	Ks=22		Vtts=16.37	N2021231325
C	STAR51	09:58:20	+01:54:21	3.6	Ks=22		Vtts=16.39	N20212332315
C	STAR52	10:01:36	+01:55:21	3.6	Ks=22		Vtts=16.40	N2021230705
C	STAR53	10:00:32	+02:21:21	3.6	Ks=22		Vtts=16.42	N20212301787
C	STAR54	09:58:16	+02:36:21	3.6	Ks=22		Vtts=16.42	N20212312700
C	STAR55	09:59:00	+01:52:21	3.6	Ks=22		Vtts=16.44	N20212332236
C	STAR56	10:01:16	+02:10:21	3.6	Ks=22		Vtts=16.45	N202123085
C	STAR57	09:59:52	+02:01:21	3.6	Ks=22		Vtts=16.46	N20212332880
C	STAR58	09:58:16	+02:35:21	3.6	Ks=22		Vtts=16.50	N20212312645
C	STAR59	09:58:40	+02:14:21	3.6	Ks=22		Vtts=16.57	N202123337391
C	STAR60	10:00:48	+02:19:21	3.6	Ks=22		Vtts=16.57	N20212301722
C	STAR61	10:01:52	+02:16:21	3.6	Ks=22		Vtts=16.58	N202123067

ANNEXE D

# Documentation pour galSVM

---

GALSVM  
Beta Release

Marc Huertas-Company

September 19, 2007

# Chapter 1

## Introduction

### 1.1 What is GALSVM?

GALSVM is a library written in IDL that combined with the free available library LibSVM allows to perform a morphological classification of high-redshift galaxies in an automated way using support vector machines. For more details on how GALSVM works please read the reference paper submitted to A&A: Huertas-Company et al., 2007, arXiv:astro-ph:0709.1359.

### 1.2 Installation

#### 1.2.1 Requirements

You need to have python ([www.python.org](http://www.python.org)) and IDL (<http://www.itervis.com/index.asp>) installed on your computer. GALSVM also uses the IDL astronomy library (<http://idlastro.gsfc.nasa.gov/>) and SExtractor ([http://terapix.iap.fr/rubrique.php?id\\_rubrique=91](http://terapix.iap.fr/rubrique.php?id_rubrique=91)).

#### 1.2.2 GALSVM installation and compilation

1. Download the compressed tar file: `galsvm_beta.tar.gz`. Untar and uncompress it somewhere in your IDL path.
2. Open IDL and compile the following procedures:

- (a) `.r aux_funct`
- (b) `.r galSVM_adapt_catalog`
- (c) `.r IO_libSVM`
- (d) `.r morphology`
- (e) `.r read_config_file`
- (f) `.r show_progress2`
- (g) `.r easy_galSVM`

If no error occurs you have successfully installed GAL SVM in your computer.

### **1.2.3 LIBSVM installation**

Please refer to the libsvm manual: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

# Chapter 2

## Quick start

The complete GALSVM pipeline is divided in 5 steps

1. Build the training catalog. The input is a catalogue of nearby galaxies visually classified. The bigger is the catalogue, the best will be your classification. If you do not have such a catalogue please contact me at: [marc.huertas@obspm.fr](mailto:marc.huertas@obspm.fr) and I probably can provide you with a test one.
2. Compute morphological parameters
3. Run LibSVM
4. Test
5. Classify the real sample

I have written a procedure called "easy\_galSVM.pro" that allows to test GALSVM on your data in a simple way. To use it you just have to configure your input parameters using the configuration file (see below).

### 2.0.4 The configuration file

#### Parameters description

This is where you define the input parameters for GALSVM. You can find an example in `/examples/config_file.txt`.

The main parameters are:

1. PATHS

- LOCAL\_CATALOG\_PATH: path of the local images
- HIGHZ\_CATALOG\_PATH: path where the high redshift simulated catalogue will be stored (it must exist)
- REAL\_CATALOG\_PATH: path where all the real images are stored

## 2. REAL THINGS

- FIELD\_IMAGE: high redshift image where the fake galaxies will be dropped
- SEGMENTATION\_IMAGE: SExtractor segmentation map of the FIELD\_IMAGE
- REAL\_CATALOG: SExtractor catalog of the real image
- OUTPUT\_CATALOG: Final catalog with the morphological classification
- Z\_HISTO: name of a file containing the redshift histogram of the high redshift sample
- MAG\_HISTO: name of a file containing the magnitude histogram of the high redshift sample

## 3. SIMULATION CONFIGURATION

- SEED: integer for random generation
- LOCAL\_CATALOG: name of the local catalogue
- HZ\_PIX\_SCALE: pixel scale of the high redshift image (FIELD\_IMAGE)
- LZ\_PIX\_SCALE: pixel scale of the low redshift images
- HZ\_ZEROP: zeropoint of FIELD\_IMAGE
- HZ\_FWHM: FWHM of FIELD\_IMAGE
- LZ\_FWHM: FWHM of the low redshift images
- SIGMA\_REF: sigma of the FIELD\_IMAGE background
- HIGH\_Z\_CATALOG: name of the simulated catalogue
- DROPPED\_CATALOG: name of the simulated catalogue once dropped
- DROPPED\_IMAGES: root name of the simulated images
- SEX\_FILE\_ROOT: root name of the SExtractor files for each simulated image

- MORPHO\_CATALOG: name of the simulated catalogue once the morphology is computed
- MAX\_OBJECTS: maximum number of objects per simulated image
- SEX\_FILE: name of the SExtractor configuration file
- SEX\_COMMAND: command for SExtractor

#### 4. LIBSVM CONFIGURATION

- TRAIN\_FILE: name of the training file to be given to libSVM
- TEST\_FILE: name of the testing file to be given to libSVM
- LIBSVM\_PATH: path of the libSVM installation
- LIBSVM\_PARAMETERS MOCK: list of the parameters to be used for the training
- LIBSVM\_PARAMETERS\_REAL: list of the parameters to be used for the classification
- LIBSVM\_MORPHO\_POS: position of the morphological class in MORPHO\_CATALOG
- LIBSVM\_MORPHO\_BORDERS: boundaries for the morphological classes
- TRAIN\_OBJECTS: number of training objects

#### Reading the file

**Objective:** Reads the configuration file and stores all the information in a structure called "input\_parameters"

#### Used routines:

- *galSVM\_read\_config\_file* in *read\_config\_file.pro*

**Running:** Just type in the IDL command line: "galSVM\_read\_config\_file,FILENAME=name of your configuration file". A file called "input\_parameters\_batch" is created in your current directory. Type ".r input\_parameters\_batch".

To verify that everything is correct type: "print, input\_parameters". A structure containing all the parameters should have been created. All the input\_parameters are now stored in a variable, you can therefore modify the parameters at any moment.

## 2.1 Building the catalogue

**Objective:** Creates a mock catalog from the local one and the corresponding galaxy stamps.

**Used routines:**

- *galSVM\_adapt\_catalog* in *galSVM\_adapt\_catalog.pro*

**Running:** The first step is to build the training catalog. You need for that purpose:

- a set of stamps of local galaxies with known morphology in fits format ((selected in the rest-frame of the sample you want to analyze to reduce the k-correction effects))
- a segmentation mask (from SExtractor) of the stamps in fits format
- a catalog in ascii format. (see below for more details)
- a PSF for each galaxy (optional)

All those elements must be in the LOCAL\_CATALOG\_PATH.

The catalog must have five entries per objects in this order:

1. ID
2. REDSHIFT
3. MORPHOLOGY (Numerical type)
4. TOTAL FLUX
5. HALF LUMINOSITY RADIUS

The filenames of the stamps must be: "root\_ID.fits", where root must be the same for every image. The filenames of the masks must be: "root\_m\_ID.fits".

**I am working to put it more friendly but for the moment. Feel free to modify the routine to your needs.**

Once you have all this together just type: "easy\_galSVM, input\_parameters, FLAG=1"

**Files created:** Simulated galaxy stamps.

**BE CAREFUL: AN ERROR MESSAGE CAN APPEAR IF YOU ARE NOT USING THE LAST VERSIONS OF QSIMP.PRO, TRAPZD.PRO, ZANG.PRO FROM THE IDL ASTRONOMY LIBRARY. IF IT IS THE CASE JUST DOWNLOAD THE LAST VERSION.**

## 2.2 Dropping the images

**Objective:** Drops the simulated stamps in the real background

**Used routines:**

- *galSVM\_drop\_galaxies* in *galSVM\_adapt\_catalog.pro*

**Running:** Just type: "easy\_galSVM,input\_parameters, FLAG=2".

**Files created:** At the end of this step a file called "DROPPED\_CATALOG" , a set of fits images and a .sh file should have been created.

## 2.3 Running SExtractor

**Objective:** Just runs SExtractor on the simulated images

**SExtractor configuration:** In this first version you need to use SExtractor with the output parameters given in the file "se\_param\_full.param" in the same order. I am working to change it and make it more robust!!!!

**Running:** Type "easy\_galSVM,input\_parameters, FLAG=3". If this is not working type in a terminal window "source your\_file.sh". We assume that SExtractor is installed in your computer and that the command is "sex". If it is not the case, edit the file and change the command.

**Files created:** A mask image and a SExtractor catalog for every dropping image

## 2.4 Selecting the simulated objects

**Objective:** Separates the simulated galaxies from the real ones in the SExtractor catalog

**Used routines:**

- *galSVM\_select\_simu\_galaxies* in *morphology.pro*

**Running:**Type: "easy\_galSVM, input\_parameters,FLAG=4"

**Files created:** A set of .gfxt files containing the simulated galaxies

## 2.5 Computing morphological parameters

**Objective:** Measure the morphological parameters on every simulated galaxy. It basically computes the concentration, the asymmetry, the gini, the m20 and the smoothness.

**Used routines:**

- *galSVM\_mock\_morpho\_catalog* in *galSVM\_adapt\_catalog.pro*

**Running:**Type: "easy\_galSVM, input\_parameters,FLAG=5"

**Files created:** A catalog containing the measured morphological parameters in the simulated objects.

**THIS STEP IS VERY EXPENSIVE IN COMPUTING TIME. I AM WORKING TO OPTIMIZE THE ROUTINES!!**

## 2.6 Writing to LIBSVM format

**Objective:** Writes the morphological catalog created in the previous step in the LibSVM format.

**Used routines:**

- *clean\_morpho* in *galSVM\_adapt\_catalog.pro*
- *libSVM\_write* in *galSVM\_adapt\_catalog.pro*

**Running:**Type: "easy\_galSVM, input\_parameters,FLAG=6"

**Files created:** A directory called LIBSVM containing a train and test file.

## 2.7 Running LIBSVM

This step uses exclusively the LIBSVM library.

**Running:** Try "easy\_galSVM,input\_parameters,FLAG=7".

## 2.8 Testing

**Objective:** Tests the accuracy of the morphological classification

**Used routines:**

- *libSVM\_errors* in *IO\_libSVM.pro*
- *libSVM\_plot* in *IO\_libSVM.pro*
- *divide\_sample\_forLibSVM* in *IO\_libSVM.pro*
- *divide\_sample\_forLibSVM2* in *IO\_libSVM.pro*

**Running:** Several tests can be performed:

- The simplest one is to look at the global accuracy. The output is a table showing the fraction of correct classifications. Just type: "easy\_galSVM,input\_parameters,FLAG=8"
- You can also look at the accuracy as a function of a parameter (magnitude, redshift...). You should use for that purpose the procedures *divide\_sample\_forLibSVM* in *IO\_libSVM.pro* or *divide\_sample\_forLibSVM2* in *IO\_libSVM.pro*. I have written an example of use in the *easy\_galSVM.pro* file but I am still testing it.

## 2.9 Classifying the real sample

**Objective:** Uses the trained machine to classify the real sample

**Used routines:**

- *galSVM\_morpho\_real* in *morphology.pro*

**Running:** Type "easy\_galSVM,input\_parameters,FLAG=9"

**Files created:** Catalog with the morphological classifications



ANNEXE E

**Articles publiés dans des revues  
à comité de lecture**

---

# Morphological evolution of $z \sim 1$ galaxies from deep $K$ -band AO imaging in the COSMOS deep field<sup>\*,\*\*\*,\*\*\*\*</sup>

M. Huertas-Company<sup>1,4</sup>, D. Rouan<sup>1</sup>, G. Soucail<sup>2</sup>, O. Le Fèvre<sup>3</sup>, L. Tasca<sup>3</sup>, and T. Contini<sup>2</sup>

<sup>1</sup> LESIA-Paris Observatory, 5 place Jules Janssen, 92195 Meudon, France  
e-mail: marc.huertas@obspm.fr

<sup>2</sup> Laboratoire d'Astrophysique de Toulouse-Tarbes, CNRS-UMR 5572 and Université Paul Sabatier Toulouse III, 14 avenue Belin, 31400 Toulouse, France

<sup>3</sup> LAM-Marseille Observatory, Traverse du Siphon, Les trois Lucs, BP 8, 13376 Marseille Cedex 12, France

<sup>4</sup> IAA-C/ Camino Bajo de Huétor, 50-18008 Granada, Spain

Received 31 October 2006 / Accepted 13 March 2007

## ABSTRACT

**Context.** We present the results of an imaging program of distant galaxies ( $z \sim 0.8$ ) at high spatial resolution ( $\sim 0.1''$ ) aiming at studying their morphological evolution. We observed 7 fields of  $1' \times 1'$  with the NACO Adaptive Optics system (VLT) in  $K_s$  ( $2.16 \mu\text{m}$ ) band with typical  $V \sim 14$  guide stars and 3 h integration time per field. Observed fields are selected within the COSMOS survey area, in which multi-wavelength photometric and spectroscopic observations are ongoing. High angular-resolution  $K$ -band data have the advantage of probing old stellar populations in the rest-frame, enabling a determination of galaxy morphological types unaffected by recent star formation, which are more closely linked to the underlying mass than classical optical morphology studies (HST). Adaptive optics on ground based telescopes is the only method today for obtaining such a high resolution in the  $K$ -band, but it suffers from limitations since only small fields are observable and long integration times are necessary.

**Aims.** In this paper we show that reliable results can be obtained and establish a first basis for larger observing programs.

**Methods.** We analyze the morphologies by means of B/D (bulge/disk) decomposition with GIM2D and C-A (concentration-asymmetry) estimators for 79 galaxies with magnitudes between  $K_s = 17$ –23 and classify them into three main morphological types (late type, early type and irregulars). Automated and objective classification allows precise error estimation. Simulations and comparisons with seeing-limited (CFHT/Megacam) and space (HST/ACS) data are carried out to evaluate the accuracy of adaptive optics-based observations for morphological purposes.

**Results.** We obtain the first estimate of the distribution of galaxy types at redshift  $z \sim 1$  as measured from the near infrared at high spatial resolution. We show that galaxy parameters (disk scale length, bulge effective radius, and bulge fraction) can be estimated with a random error lower than 20% for the bulge fraction up to  $K_s = 19$  ( $AB = 21$ ) and that classification into the three main morphological types can be done up to  $K_s = 20$  ( $AB = 22$ ) with at least 70% of the identifications correct. We used the known photometric redshifts to obtain a redshift distribution over 2 redshift bins ( $z < 0.8$ ,  $0.8 < z < 1.5$ ) for each morphological type.

**Key words.** galaxies: fundamental parameters – galaxies: high-redshift – galaxies: evolution – instrumentation: adaptive optics

## 1. Introduction

The process of galaxy formation and the way galaxies evolve is still one of the key unresolved problems in modern astrophysics. In the currently accepted hierarchical picture of structure formation, galaxies are thought to be embedded in massive dark halos that grow from density fluctuations in the early universe (Fall & Efstathiou 1980) and initially contain baryons in a hot gaseous phase. This gas subsequently cools, and some fraction eventually

condenses into stars (Lilly et al. 1996; Madau et al. 1998). However, many of the physical details remain uncertain, in particular the process and history of mass assembly. One classical observational way to test those models is to classify galaxies according to morphological criteria defined in the nearby Universe (Hubble 1936; de Vaucouleurs 1948; Sandage 1961), which can be related to physical properties, and to follow this classification across time. (Abraham et al. 1996; Simard et al. 2002; Abraham et al. 2003). Indeed, since galaxies were recognized as distinct physical systems, one of the main goals in extragalactic astronomy has been to characterize their shapes. Comparison of distant populations with the ones found in the nearby Universe might help to clarify the role of merging as one of the main drivers in galaxy evolution. (Cole et al. 2000; Baugh et al. 1996). Progress in this field has come from observing deeper and larger samples, but also from obtaining higher spatial resolution at a given flux and at a given redshift. In the visible, progress has been simultaneous on those two fronts, thanks in particular to the ultra-deep HDF fields observed with the Hubble Space Telescope. HST imaging has brought observational evidence that galaxy evolution is differentiated with respect to morphological type and

\* Based on ESO observations at the VLT. Programmes P73.A-0814A and P75.A-0569A and 175.A-0839 (zCOSMOS).

\*\* Based on observations obtained with MegaPrime/MegaCam, a joint project of the CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of the NRC and CNRS.

\*\*\* Appendix A and Table 6 are only available in electronic form at <http://www.aanda.org>

that a large fraction of distant galaxies have peculiar morphologies that do not fit into the elliptical-spiral Hubble sequence. (Brinchmann et al. 1998; Wolf et al. 2003; Ilbert et al. 2006b). These results can however be biased by the fact that most of the sampled galaxies are at large redshift and are analyzed through their UV rest-frame emission, which is more sensitive to star formation processes and to extinction. Moreover, it now seems clear that evolution strongly depends on the galaxy mass in the sense that massive systems appear to have star formation histories that peak at higher redshifts, whereas less massive systems have star formation histories that peak at progressively lower redshifts and are extended over a longer time interval (downsizing scenario; Cowie et al. 1996; Brinchmann & Ellis 2000; Bundy et al. 2005). This could explain the fact that the population of massive E/S0 seems to be in place at  $z \sim 1$  and evolve passively towards lower redshifts (De Lucia et al. 2006; Zucca et al. 2006). However, most of these studies are based on spectral type classifications, and their interpretation in the framework of galaxy formation is not straightforward since galaxies move from one spectral class to another by passive evolution of the stellar populations.

In this context, high-resolution near-infrared observations are particularly important because the  $K$ -band flux is less dependent on the recent history of star formation, which peaks in the UV in rest frame, and thus gives a galaxy type from the distribution of old stars that is more closely related to the underlying total mass than optical observations. This is why a large number of  $K$ -band surveys have been carried out using ground-based telescopes with different spatial coverages and limiting magnitudes in order to perform cosmological tests by means of galaxy counts essentially (Gardner et al. 1993; Glazebrook et al. 1995; McLeod et al. 1995; Bershadsky et al. 1998; McCracken et al. 2000; Maihara et al. 2001). However, no morphological information can be found in particular due to the seeing limited resolution, even with superb image quality as in the ongoing CFHT/WIRCAM survey. Conselice et al. (2000) prove that only when the ratio of the angular diameter subtending  $0.5 h_{75}^{-1}$  kpc at a given distance to the angular resolution of the image is around 1 can reliable morphological estimators such as the asymmetry be obtained.

Consequently, adaptive optics (AO) installed on ground-based telescopes will be a powerful method for obtaining near-infrared high-resolution data in the future and an excellent complement to space observations, such as those that will be performed with the HST (WFC3, 2008) and with the JWST (2014).

The use of classical AO for deep surveys suffers, however, from inherent limitations such as the non-stationary PSF on long integration times and the finite isoplanetic field. This is why preliminary studies to probe the accuracy of adaptive optics are required, before launching very wide surveys. In particular, it is important to determine whether automated morphological classifications can be performed. Indeed, given the large number of galaxies, such automated methods for morphological classifications are desirable.

Minowa et al. (2005) published first results based on adaptive optics observations. They achieved the impressive limiting magnitude of  $K \sim 24.7$  with the Subaru Adaptive Optics system with a total integration time of 26.8 h over one single field of  $1' \times 1'$ . They proved that the use of adaptive optics significantly improves detection of faint sources but did not obtain morphological information. In a recently submitted paper, Cresci et al. (2006) perform a morphological analysis based on AO data for the first time. They observed a  $15 \text{ arcmin}^2$  in the  $K_s$  band with NACO (SWAN survey) and classified distant galaxies into two morphological bins (late type, early type) by performing a

model fitting with a Sersic law. They compared number counts and size-magnitude relation, for early and late-types separately, with hierarchical and pure luminosity evolution (PLE) models, respectively. They conclude that hierarchical models are not consistent with the observed number counts of elliptical galaxies and that PLE models are preferred. However, as discussed in several studies (Gardner 1998), despite galaxy counts still being one of the classical cosmological tests, their interpretation remains difficult. It is thus not realistic to expect galaxy counts alone to strongly constrain the cosmological geometry or even to constrain galaxy evolution. A more complete study needs redshift estimates, which is lacking in the SWAN survey. That is the main reason for having select the COSMOS (Scoville & COSMOS Team 2005) field to conduct our pilot program since multi-wavelength photometric and spectroscopic observations are performed. This ensures a reliable redshift estimate for all our objects.

In this paper, we continue this AO validation task by morphologically classifying a sample of 79 galaxies observed using parametric (GIM2D, Simard et al. 2002) and non-parametric (C-A, Abraham et al. 1994) methods and comparing them. Fields, observed with NAOS/CONICA adaptive optics system, are distributed over a  $7 \text{ arcmin}^2$  area. We obtain for the first time an estimate of the distribution of galaxies in three morphological types (E/S0, S, Irr) at redshift  $z \sim 1$  as measured from the near infrared at high spatial resolution. We then use the photometric redshifts to look for evolution clues as a function of morphology.

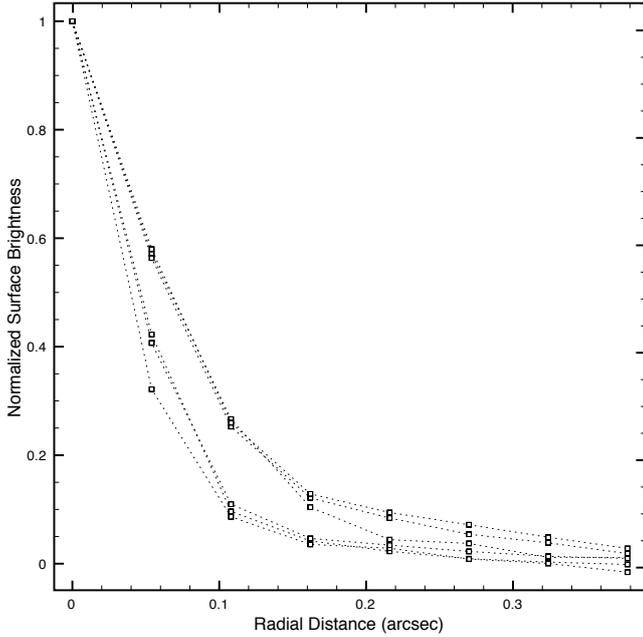
The paper proceeds as follows: the data set and the reduction procedures are presented in the next section. In Sect. 3, we focus on the detection procedures and the sample completeness. In Sect. 4 we discuss the estimate of redshifts using the multi-wavelength photometric data from COSMOS. The morphological analysis is discussed in Sect. 5 using bulge/disk decompositions and concentration and asymmetry estimators. Simulations for error characterization are carried out for both classification methods, and comparisons between classifications are shown. In Sect. 6 we compare the data with ground and space observations and use those comparisons to discuss the morphological evolution in the last section. Throughout this paper magnitudes are given in the Vega system. We use the following cosmological parameters:  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $(\Omega_M, \Omega_\Lambda, \Omega_k) = (0.3, 0.7, 0.0)$ .

## 2. The data set

Seven fields of  $1' \times 1'$  were observed in  $K_s$  band ( $2.16 \mu\text{m}$ ) with the NAOS/CONICA-assisted infrared camera installed on the VLT<sup>1</sup>. The fields were selected within the COSMOS survey area<sup>2</sup>. In order to ensure a reliable AO correction, relatively bright stars ( $V \sim 14$ ) were selected. We added a color criterium ( $B - R < 1.0$ ) in order to benefit from a large attenuation of the flux in the near-IR and thus a smaller occultation of the central region of the images. The pixel scale ( $0.054''$ ) was chosen to be twice the Nyquist-Shannon requirements with respect to the telescope diffraction limit in order to have larger fields. With such a pixel sampling, the PSF  $FWHM$  would only be one pixel width in the limit of perfect AO correction. However, we will show in Sect. 5 that our PSF reconstruction procedure can handle such undersampled data. We also note that only partial AO correction was actually achieved during our observations, and our PSF reconstruction is thus more than adequate here. Figure 1 compares

<sup>1</sup> P73.A-0814A and P75.A-0569A

<sup>2</sup> <http://www.astro.caltech.edu/cosmos/>



**Fig. 1.** Radial profile of all detected stellar objects with  $K_s < 19$ . The mean resolution is  $\sim 0.1''$ , broader than the telescope diffraction limit, as a consequence of the long integration time ( $\sim 3$  h) and the large fields.

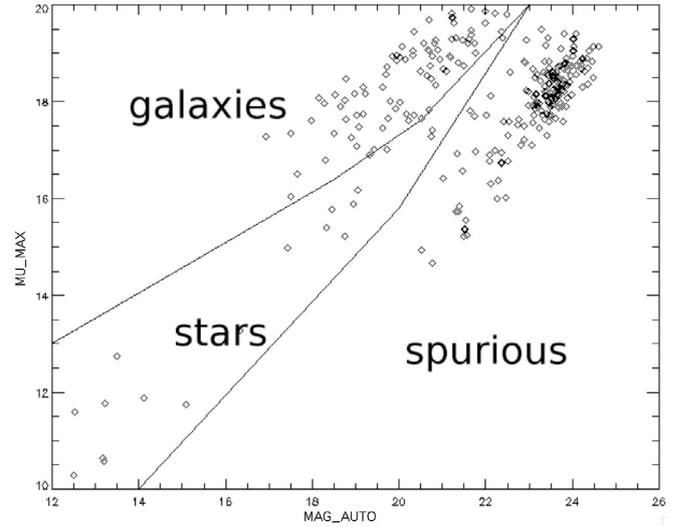
**Table 1.** Summary of observations for the seven analyzed fields. The mean seeing is estimated when faint stars were detected.

Field	$\alpha$	$\delta$	Exp. time (s)	Seeing (arcsec)
STAR1	10:00:16	+02:16:22	10 350	0.09
STAR2	10:00:52	+02:19:52	7650	–
STAR3	10:00:10	+02:06:08	7650	0.12
STAR4	09:59:52	+02:05:00	7170	0.08
STAR5	10:00:14	+02:09:09	10 200	0.09
STAR6	10:00:02	+02:06:57	7650	0.13
STAR10	09:59:56	+02:04:07	10 000	–

the radial profiles of 5 detected stellar objects. Indeed, this program is pushed to its limits in terms of field size, exposure time, and brightness of the guide star.

Data are reduced in a standard way: exposures are taken in the auto-jitter mode, which means that the pointing is randomly shifted within a  $7''$  box, in order to improve flat fielding, bad pixel correction, and sky background withdrawal. The sky value in each pixel is estimated by performing a clipped median of all the exposure frames: the 10% faintest and brightest values are removed from the computation. Cosmic ray and flat corrections were applied, and recentering was done using a standard cross-correlation method. Recentering is done at a sub-pixel level. For that purpose a cubic interpolation of resampling was performed. After stacking, a global estimate of the sky background was performed by computing the stacked image spatial median. The final image was obtained after subtraction of this value.

Photometric zeropoints were determined using 2MASS stars (Kleinmann 1992). We performed aperture photometry on the guide stars and compared it to 2MASS data to deduce the zeropoints. Note that the change of detector between periods P73 and P75 resulted in different zeropoint values for each run. The average zeropoint for the first period is:  $22.82 \pm 0.06$  and for the second period  $23.29 \pm 0.06$ . We also used the ESO calibration data set standard stars (Persson et al. 1998) to validate our measurements (ESO pipeline computations and our own measurements



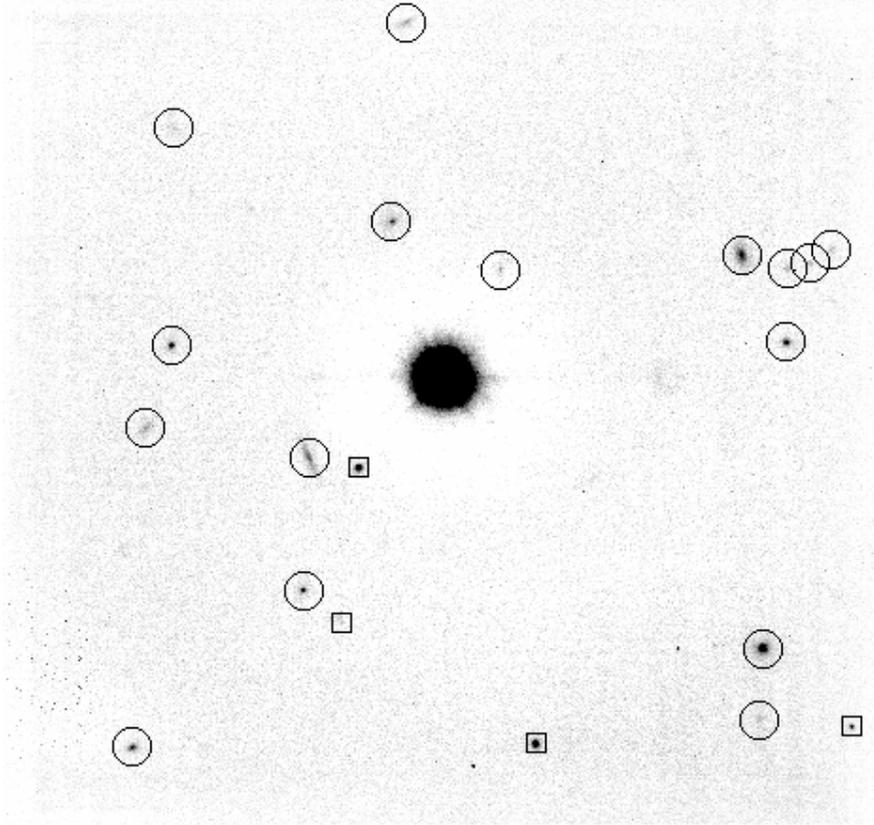
**Fig. 2.** Objects classification using the MAG\_AUTO-MU\_MAX plane.

on the standard stars images). There is good agreement between all these values.

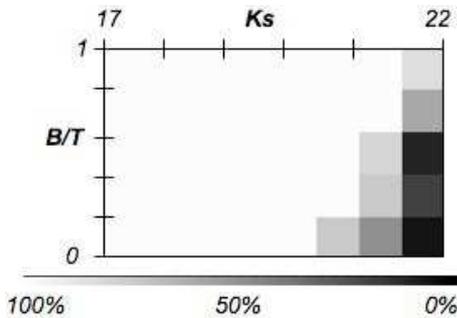
### 3. Detection and completeness

All objects having a  $3\sigma$  signal above sky, over 4 four contiguous pixels are detected using SEXTRACTOR (Bertin & Arnouts 1996). We decided to apply this low detection threshold, even if the main goal of this paper is to perform a morphological analysis, for two main reasons. First we want to test the ability of adaptive optics based observations to obtain morphology, so we are seeking the limits; and second, we wanted to be sure that no objects are lost when computing number counts to compare with other near-infrared observations (see Appendix A). We find 285 objects over the 7 fields. We then performed a cleaning task in order to separate galaxies from stars and spurious detections. This was made using the SEXTRACTOR MU\_MAX and MAG\_AUTO parameters that give the peak surface brightness above the background and the Kron-like elliptical aperture magnitude, respectively. The distribution of objects in this parameter space clearly defines three regions that separate extended sources from point-like or non-resolved sources and from spurious detections (Fig. 2). In this separation scheme, objects with very faint magnitude and high peak surface brightness are considered as false detections. Boundaries were drawn manually and a visual inspection confirms that known stars in the field are indeed identified as point sources. We consequently identified 79 galaxies, 19 stars (or unresolved objects), and 187 spurious objects in the whole field.

The sample completeness for point sources was estimated by creating artificial point sources from fields stars (see Sect. 5 for detailed explanations) with apparent magnitudes ranging from  $K_s = 18$  to  $K_s = 24$  and placing them at random positions. We ran SEXTRACTOR with the same configuration as for real sources and looked for the fraction of detected objects. We find that the sample is 50% complete at  $K_s = 22.5$  (or  $AB = 24.5$ ) for point sources. Completeness for extended sources is estimated in a similar way: we generate 1000 galaxies with exponential and de Vaucouleurs profiles of different morphological types (bulge fraction ranging from 0 to 1) and with galaxy parameters uniformly distributed. In particular, the sizes of disks and bulges are distributed uniformly between  $0'' < r_d < 0.7''$



**Fig. 3.** NAOS/CONICA  $K_s$ -band image of the field centered at  $\alpha = 10:00:16$ ,  $\delta = +02:16:22$ . The total integration time is 10 350 s. The field size is  $1' \times 1'$  with a pixel scale of 54 mas. Circles are detected galaxies and boxes are stars. The stellar  $FWHM$  is measured to be  $0.1''$ . The bright star at the center of the image is used as the AO guide star.



**Fig. 4.** Completeness for extended sources. Galaxies with parameters  $(r_d, r_e, B/T)$  uniformly distributed are simulated and placed at random positions in the fields.

and  $0'' < r_e < 0.7''$  as detailed in Sect. 5. This leads to half-luminosity radii ranging from  $0'' < r_h < 1''$ . We find that the sample is 50% complete at  $K_s = 21.5$  (or  $AB = 23.5$ ) for this population of extended sources (Fig. 4). We used this completeness to compute number counts and to compare it to other near-infrared surveys in Appendix A.

#### 4. Photometric redshifts

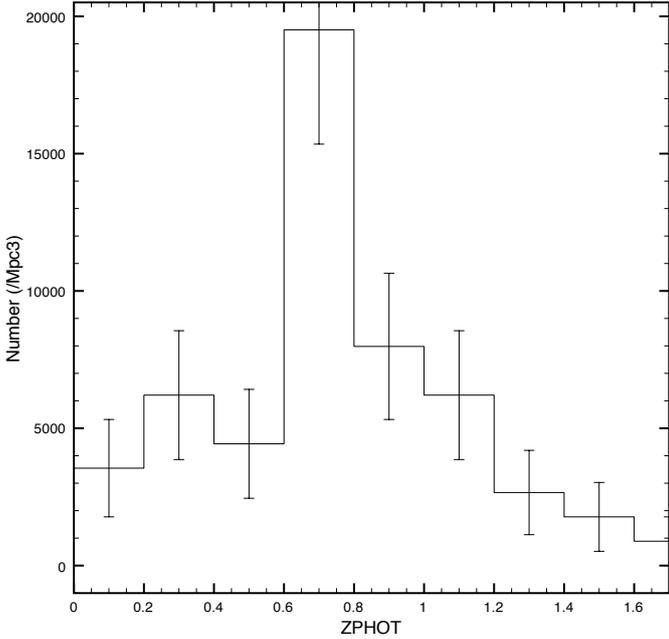
Galaxy number counts provide a useful description of galaxy populations but suffer from numerous degeneracies when trying to trace the evolution of galaxy populations. Model predictions are subject to uncertainties in the spectral energy distributions and evolution of galaxies and in the free parameters specifying the luminosity function, the cosmological geometry, the number and distribution of galaxy types, and the effect of dust and merging. The need to have redshift information is therefore the

reason for selecting the NACO fields within the ongoing Cosmic Evolution Survey (COSMOS) in which multi- $\lambda$  and spectroscopic observations are performed. COSMOS is designed to probe the correlated evolution of galaxies, star formation, active galactic nuclei (AGN) and dark matter (DM) with large-scale structure (LSS) over the redshift range  $z = 0.5$  to 3. The survey includes multi-wavelength imaging and spectroscopy from X-ray to radio wavelengths covering a 2 square deg area, including HST imaging of the entire field.

All these data are used for a direct estimate of the photometric redshifts of the galaxies detected in the NACO fields, computed with the code *Le Phare*<sup>3</sup>. A standard  $\chi^2$  method is implemented, including an iterative zero-point refinement combined with a template optimization procedure and the application of a Bayesian approach (Ilbert et al. 2006a). We used 1095 spectroscopic redshifts taken from the zCOSMOS Survey (Lilly & The Zcosmos Team. 2006) to measure the photometric redshifts. This method allows to reach an accuracy of  $\sigma_{\Delta z}/(1+z_s) = 0.031$  with 1.0% catastrophic errors, defined as  $\Delta z/(1+z_s) > 0.15$ .

The multi-color catalog of the COSMOS survey (Capack et al. 2006) consists of photometry measurements over 3 arcsec diameter apertures for deep  $B_j, V_j, g+, r+, i+, z+$  Subaru data taken with SuprimeCam,  $u^*, i^*$  bands with MegaCam (CFHT),  $u', g', r', i', z'$  information from the Sloan Digital Sky Survey (SDSS),  $K_s$  magnitude from KPNO/CTIO, and  $F816W$  HST/ACS magnitude. Objects were matched between the COSMOS and the NACO catalogs within a radius of  $2''$  which takes possible astrometry differences between the catalogues into account. We matched 60 objects out of the 79 detected in the NACO fields. Figure 5 shows the photometric redshift distribution for these 60 matched objects. As expected for a galaxy

<sup>3</sup> [http://www.lam.oamp.fr/arnouts/LE\\_PHARE.html](http://www.lam.oamp.fr/arnouts/LE_PHARE.html)



**Fig. 5.** Le Phare photometric redshift distribution for the 60 matched objects. The distribution is peaked around  $z \sim 0.8$ , in good agreement with the predictions of simple PLE models. Error bars show poissonian errors.

sample limited to  $K_s = 22$ , the redshift distribution peaks around  $z \sim 0.8$  (Mignoli et al. 2005).

## 5. Automated morphology classifications

The 79 objects identified as extended sources are morphologically classified using two automated methods based on direct model fitting and on learning classification. Automated classifications have the fundamental properties of being objective, thus reproducible, and they allow a precise error characterization. We proceeded in two steps: first we detected irregular objects using asymmetry estimators, and then we separated the regular objects between early and late type objects.

Throughout this section, we use extensive simulations for error estimates and calibration of the automated classifications as explained below. For all the simulations, we assumed that bulges are pure de Vaucouleurs profiles ( $n = 4$ ) and that disks are exponential profiles. We then generate galaxies with parameters uniformly distributed in the following parameter space:  $0 < B/T < 1$ ,  $0'' < r_d < 0.7''$ ,  $0'' < r_e < 0.7''$ ,  $0^\circ < i < 70^\circ$ , and  $0 < e < 0.7$ . Both bulge and disk position angles were fixed to  $90^\circ$ . The goal of these simulations is to characterize biases and errors; the uniformity of the parameter distributions adopted here is therefore perfectly suitable, even though real galaxy parameters may not be so distributed. For the same reason, we do not take any redshift effect into account. Each simulation was convolved with the reconstructed PSF as explained in 5.2.2. The same PSF was used in both creating and analyzing the simulations, so the results will not include any error due to PSF mismatch. In order to simulate background noise, objects are embedded at random positions in the fields and detected with the same SExtractor parameters as for the real sources.

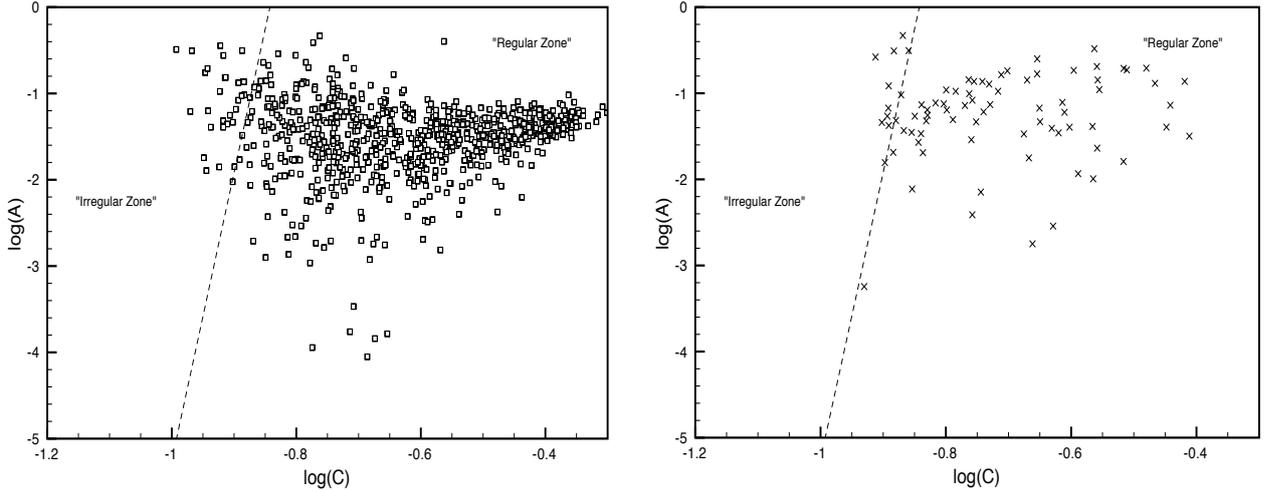
### 5.1. Irregular objects

The detection of objects presenting irregularities is made using the concentration and asymmetry estimators (Abraham et al. 1994; Abraham et al. 1996). Concentration is computed as the ratio of the flux between the inner and outer isophotes of normalized radii 0.3 and 1 within the isophotal area enclosed by pixels  $3\sigma$  above the sky level. The corresponding limiting surface brightness varies between  $\mu = 18.76\text{--}20.39$  mag arcsec $^{-2}$  because of the variations in the exposure time between the different fields and the intrinsic variation of the sky level, which is important for infrared ground-based observations. Indeed, the value of  $C$  is quite sensitive to the estimate of the background level since an error in this value will result in different limiting isophotes, and a fraction of the galaxy flux can be lost. To estimate the error in  $C$  introduced by the different isophote levels, we computed the variations in the  $C$  value for variations in the limiting surface brightness of  $\Delta\mu = 1$ . We found a fairly small error ( $\Delta C \sim 0.06$ ), so we decided not to apply any corrections. Asymmetry ( $A$ ) was obtained by rotating the galaxy image about its center by  $180^\circ$  and self-subtracting it to the unrotated image after sky subtraction. Local sky level is estimated using SExtractor output parameter *BACKGROUND* that gives the background level at the galaxy centroid position. The center of rotation was determined by first smoothing the galaxy image with a Gaussian kernel of  $\sigma = 1$  and then choosing the location of the maximum pixel as the center, as explained in Abraham et al. (1996). Since the absolute value for the residual light is used, noise in the images shows up as a small positive  $A$  signal even, in perfectly symmetrical objects (Conselice et al. 2000). This is why we applied a noise correction to the computation: the value of  $A$  in a portion of sky with area equal to that enclosed by the galaxy isophote. The definition of  $A$  that is used in the rest of the paper is:

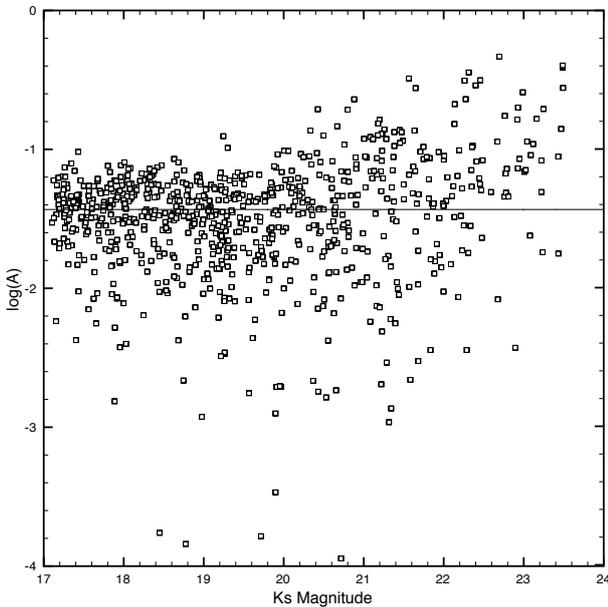
$$A = \frac{1}{2} \left( \frac{\sum |I(i, j) - I_{180}(i, j)|}{\sum I(i, j)} - \frac{\sum |B(i, j) - B_{180}(i, j)|}{\sum I(i, j)} \right). \quad (1)$$

To establish the boundaries between regular and irregular objects, a calibration of the C-A plane is needed. We thus simulated a set of galaxies with different galaxy parameters and magnitudes ranging between  $17 < K_s < 23$  that we embedded in the real images. We computed the  $C$  and  $A$  parameters of these objects and plotted the C-A plane. Since irregular objects cannot be simulated in a meaningful way, a kind of extrapolation was employed, based on two facts: a) irregular galaxies have flatter photometric profiles (less concentrated) and to be more asymmetric than regular objects; b) those objects are not present in the simulated sample. We thus defined the irregular zone as the upper left corner of the C-A plane where no simulated objects are found (Fig. 6).

As said previously, rotational asymmetry is affected by noise even after correction. This means that fainter objects might appear more asymmetric and can thus induce a bias in the number of irregular objects at the faint end of the sample. To estimate this error, we plotted the asymmetry parameter for a sample of 1000 simulated galaxies with magnitudes ranging between  $17 < K_s < 23$  (Fig. 7). The plot shows that asymmetry begins to be affected by noise only at magnitudes greater than 22.2, which is the magnitude limit of our working sample. In summary, we found the location of the irregular/peculiar objects by simulating a set of regular galaxies and defining the peculiar area as the area outside. Then, we plotted the observed data on this plane and count the galaxies in this peculiar area. We counted 10 observed objects in this zone, i.e. 12% of the sample. We can



**Fig. 6.** Separation between regular and irregular objects. *Left:* simulated objects (empty squares), *right:* real objects (crosses).



**Fig. 7.** Asymmetry versus magnitude: asymmetry begins to grow only at magnitudes greater than 22.2 which is greater than the limiting magnitude.

attempt to quantify the error in this classification by considering the regular simulated objects that fall in the irregular zone. This gives the fraction of regular objects that are misclassified. We counted 27 objects out of 1000. we conclude that  $12\% \pm 2.7$  of our sample corresponds to peculiar objects, in the magnitude range  $17 < K_s < 22$ .

## 5.2. Regular objects: disk dominated – bulge dominated separation

### 5.2.1. C-A morphology

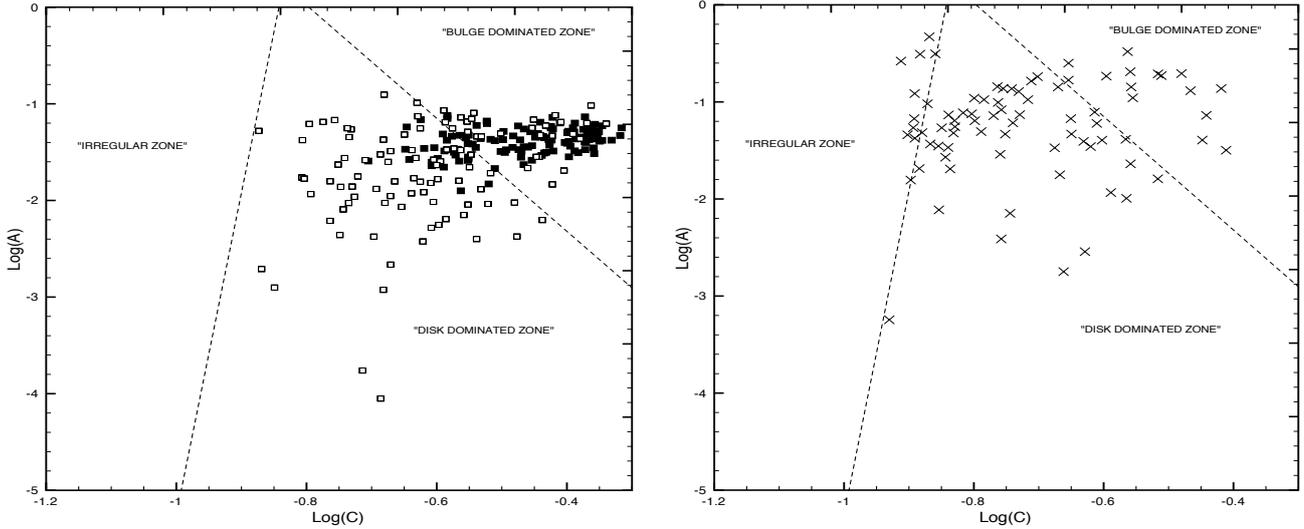
The positions of galaxies in the C-A plane are used to separate the bulge and disk-dominated galaxies as follows:

**Calibration:** A calibration of the C-A plane is needed before classification to investigate where the objects exactly fall. The strategy followed in this paper is two-fold: first we draw the

irregular border as defined in Sect. 5.1. The border between bulge-dominated and disk-dominated objects can be deduced in a more automated way thanks to the analysis of simulated galaxies. We took the same 1000 galaxies as above, for which the morphological type is known, and draw their positions in the C-A plane (Fig. 8). The border is then defined with a classification method based on support vector machines (Vapnik 1995)<sup>4</sup>. We decided not to use classical boundaries employed in previous works because: a) those boundaries were not obtained in the K-band and b) we were looking for an objective method that did not require visual inspections. Support vector machines (SVM) non-linearly map their n-dimensional input space into a “high dimensional feature space”. In this high-dimensional feature space, a linear classifier is constructed. SVM have two main parameters that can be changed: the kernel function and the tolerance  $C$ . The kernel function corresponds to the expected shape of the border, for instance, if the objects are distributed with a Gaussian distribution, then a Gaussian kernel will be used. The adjustment of the border will also take a tolerance factor  $C$  into account. If  $C$  is high, the machine will not allow any object to be on the wrong side of the border. As a consequence, if the objects are strongly mixed in a given plane, the border can have a very complex shape. In contrast, if  $C$  is too low the machine will not reach an optimal separation. In a first approach, a linear kernel was used, assuming that the two families of objects can be separated with a linear function, in order to be coherent with previous works. Thus the  $C$  parameter is set to be infinite. The results of this separation are displayed in Fig. 8.

**Accuracy:** Automated classifications are useful because they allow a characterization of errors. Once the boundaries are drawn, we generated another set of 500 fake objects with known morphology that we place again in the C-A plane and that we classified in the three morphological types we have defined. We then compared the results of our classification scheme with the initial morphological type. Errors were estimated in magnitude bins, from  $K_s = 18$  to  $K_s = 22$  (Table 2). We achieved 70% of good identifications up to  $K_s = 21$  ( $AB = 23$ ) and 66% up to our magnitude limit  $K_s = 22$ . That means that we are able to classify galaxies in the three main morphological types with a reliable accuracy at least up to  $K_{AB} = 23$ .

<sup>4</sup> <http://www.isis.ecs.soton.ac.uk/resources/svminfo/>



**Fig. 8.** C-A calibration and classification. Boundaries are drawn using an automated classification method (SVM) that avoids the use of a nearby sample and subjective visual classifications. *Left:* simulated objects, open squares: objects with  $B/T < 0.2$ , filled squares: objects with  $B/T > 0.8$ . *Right:* real objects.

**Table 2.** Error estimates of the C-A classification. Fraction of misclassified objects for several magnitude ranges.

Magnitude	Correct Identifications
$K_s < 19$	80%
$K_s < 20$	73%
$K_s < 21$	70%
$K_s < 22$	66%

Some words about the C-A plane: At first sight, the distribution of galaxies in our C-A plane looks significantly different that what has been reported in previous works using this techniques (Abraham et al. 1996; Brinchmann et al. 1998). Indeed, the slope of the separating border between bulge and disk-dominated galaxies has been found to be positive, whereas the one found here is negative, although previous classifications are somewhat arbitrary. As a consequence, bulge-dominated galaxies lie in the top right corner of the plane in our classification rather than in the bottom right corner in most other studies. There might be several reasons to explain this effect:

- As NACO images are undersampled, highly concentrated objects only cover a few pixels, consequently, even a small mismatch in the determination of the rotation center can lead to large asymmetries. To check this effect we computed concentration and asymmetry parameters for the detected point-like sources. We find that, indeed, they appear to have higher asymmetric values than the galaxies.
- It might also be a consequence of the method used for the boundary estimate. Abraham et al. (1996) and Brinchmann et al. (1998) used a visual inspection based on a local survey and, in order to account for the effects of redshift, they applied corrections to the concentration parameter. In this work, we used a fully automated method based on simulations that reproduce the observational conditions very closely and on learning classification methods, so no correction is needed in principle, but this gives a different border. The question that consequently arises is whether the differences in the boundaries produce important differences in the classification procedure. All this points are fully discussed in

Sect. 6 thanks to detailed comparisons with space observations of the sample.

### 5.2.2. Model-fitting morphology

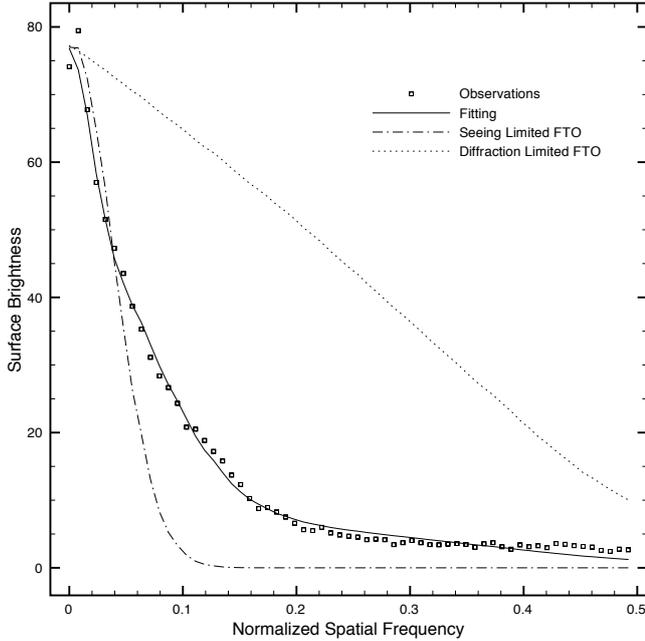
The second method is based on a direct two components fitting with exponential and de Vaucouleurs profiles, using GIM2D (Simard et al. 2002). The 2D galaxy model used by GIM2D has 11 parameters that are fitted to the real data. The most important ones are the total flux and the bulge fraction  $B/T$  ( $=0$  for pure disk systems). Other parameters are the (i) disk scale radius  $r_d$ ; (ii) the disk inclination  $i$ ; (iii) the effective radius  $r_e$ ; (iv) the ellipticity  $e$  of the bulge component, and other geometric parameters for the center and orientation of both components. As GIM2D also estimates the local sky level using image statistics, this is the value that is used.

**PSF reconstruction:** To obtain reliable results, GIM2D needs a noise-free, well-sampled PSF. This is why special attention has been paid to PSF reconstruction. Here, classical methods, such as DAOPHOT or Tiny Tim, could not be used for two reasons. First, the Adaptive Optics PSF has a specific shape that is neither “seeing limited like” nor “spatial like”. An AO system operated with a guide star of moderate brightness ( $V \sim 14$ ) can only partially correct for turbulence-induced distortions. This partially-corrected PSF consists of two components: a diffraction-limited core, superimposed on a seeing halo. Second, to have a larger field and a better sensitivity, data were under-sampled by a factor 2 ( $0.054''$  pixel scale, whereas  $0.02''$  is needed to be Nyquist sampled).

We developed a simple algorithm that uses field stars to generate Nyquist-Shannon-sampled PSFs by means of a fitting procedure in Fourier space. The process is as follows: we generate an artificial PSF with a diffraction-limited core and a Gaussian halo, with the distribution

$$\text{PSF}_{\text{art}}(x, y) = k \left[ \text{SR} \times A(x, y) + (1 - \text{SR}) \times \exp\left(-\frac{x^2 + y^2}{\sigma^2}\right) \right]$$

where SR is the Strehl ratio,  $A(x, y)$  the bi-dimensional Airy function,  $\sigma$  the Gaussian dispersion that can be related to the



**Fig. 9.** Example of PSF fitting in Fourier space. Squares: observations, dashed line: seeing-limited MTF, dotted line: diffraction-limited MTF. The AO MTF contains higher frequencies than the seeing-limited one. The telescope diffraction limit is not reached however in this example due to the undersampling of the instrumental setup.

seeing and  $k$  is a normalization factor. This artificial PSF is built with a Nyquist-Shannon sampling, binned by a factor 2 to reach the real-image pixel scale and finally Fourier-transformed to create a simulated MTF (power spectrum). On the other hand, for each observed star, its Fourier transform is fitted with the simulated MTF. The parameters estimated that way ( $SR$ ,  $\sigma$ , and  $k$ ) are then used to build an estimate of the PSF with the correct Nyquist-Shannon sampling.

Working in Fourier space avoids including the background estimate and PSF position as a fit parameter, which is particularly delicate in our case, since the  $FWHM$  is less than 2 pixels large. In the few cases where the fitting procedure does not converge a second Gaussian halo is added. Figure 9 shows the result of the fitting for one star in the spatial frequency domain. In this paper, we do not consider variations in the PSF caused by adaptive optics, such as anisoplanatism, but we are working on building a complete model for PSF estimate for future observations.

**Error analysis:** We ran GIM2D on the 79 objects with magnitudes ranging between  $K_s = 17$ – $22$ , using a two components model and the artificial PSF built as described above. We used the GIM2D mode that allows use of oversampled PSFs to deal with undersampled data, since the PSF recovered with the method explained above is Nyquist sampled. The fitting converged for the whole sample, and the results are quite convincing in terms of residual images (Table 3).

Visual inspection of the models compared to the real images (Table 3) also reveals good agreement, in particular for bright sources. For the faintest objects, however, it is more difficult to estimate the fitting accuracy. Indeed, inspection of image residuals is not a robust accuracy test, since there may still be strong degeneracies even when the image residuals do not show any features. This is why objective and systematic error characterization is needed. To do so we generated a sample of 1000 synthetic

**Table 3.** GIM2D output for some objects. *Left:* original images, *middle:* GIM2D models, *right:* residuals. The small thumbnails show the real and the model galaxy. The image size is  $1.7' \times 1.7'$

Original	Model	Residual

galaxies with known galaxy parameters uniformly distributed:  $0 < B/T < 1$ ,  $0 < r_d < 0.5''$ ,  $0 < r_e < 0.5''$ ,  $0 < e < 0.7$ ,  $0^\circ < i < 85^\circ$ . The Sersic bulge index was fixed at  $n = 4$ , and both bulge and disk position angles were fixed to  $90^\circ$ . As explained in Simard et al. (2002), the goal of these simulations is to characterize biases and error. The uniformity of the parameter distributions adopted here is therefore perfectly suitable even though real galaxy parameters may not be so distributed. Each simulation is convolved with the reconstructed PSF. The same PSF is used in both creating and analyzing the simulations, so the results will not include any error due to PSF mismatch. In order to simulate background noise, objects are embedded at random positions in the fields and detected with the same SExtractor parameters as for the real sources. Finally, the GIM2D output files are processed through the same scripts to produce a catalog of final recovered structural parameters.

Following the Simard et al. (2002) procedure, we decided to represent errors in a set of two-dimensional maps giving

**Table 4.** Error analysis of the bulge fraction  $B/T$  for different recovered magnitude ranges and different bins in recovered galaxy size. The galaxy size is represented by the half-light radius and is distributed into 4 bins in  $\log(r_{\text{hl}})$ . In the top left corner bright and small objects are found whereas faint and large objects are placed in the bottom right corner.  $\overline{\Delta B/T}$  is the average difference between introduced and recovered values of  $B/T$ , while  $\sigma\Delta B/T$  is the dispersion (see text for details).  $N$  is the number of simulations used for each bin.

Magnitude	$-1 < \log(r_{\text{hl}}) < -0.75$			$-0.75 < \log(r_{\text{hl}}) < -0.5$			$-0.5 < \log(r_{\text{hl}}) < -0.25$			$-0.25 < \log(r_{\text{hl}}) < 0$		
	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	$N$	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	$N$	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	$N$	$\overline{\Delta B/T}$	$\sigma\Delta B/T$	$N$
[17–17.5]	-0.110	0.247	(19)	-0.034	0.140	(22)	-0.002	0.056	(4)	-0.030	0.060	(8)
[17.5–18]	0.165	0.088	(4)	0.037	0.277	(18)	0.027	0.157	(23)	0.010	0.110	(11)
[18–18.5]	-0.002	0.141	(19)	0.020	0.262	(58)	0.162	0.174	(72)	0.226	0.155	(28)
[18.5–19]	-0.114	0.324	(14)	0.021	0.309	(71)	0.213	0.196	(83)	0.170	0.159	(30)
[19–19.5]	-0.064	0.221	(20)	0.212	0.266	(90)	0.192	0.252	(51)	0.145	0.027	(3)
[19.5–20]	0.100	0.259	(24)	0.270	0.298	(110)	0.181	0.258	(34)	0.140	0.100	(6)
[20–20.5]	0.105	0.430	(36)	0.163	0.332	(107)	0.113	0.281	(13)	N/A	N/A	(0)
[20.5–21]	0.050	0.476	(56)	0.148	0.351	(81)	0.194	0.054	(3)	N/A	N/A	(0)

systematic and random errors at each position. The GIM2D parameter space is a complex space with 11 dimensions, so these maps can only offer a limited representation of the complex multidimensional error functions but makes interpretation much simpler. The error analysis presented in this paper focuses on the error made on the main morphological estimator, the bulge fraction, as a function of two main parameters: apparent magnitude and half light radius. Systematic errors are computed as the mean difference between the introduced and the recovered value:  $\overline{\Delta B/T} = \frac{\sum(B/T_r - B/T_l)}{N}$  and random errors as the square root

of the variance of the difference:  $\sigma\Delta B/T = \sqrt{\frac{\sum(\Delta B/T - \overline{\Delta B/T})^2}{N-1}}$ . Table 4 precisely shows in details the sources of errors on  $B/T$  as a function of galaxy magnitude and half-light radius.

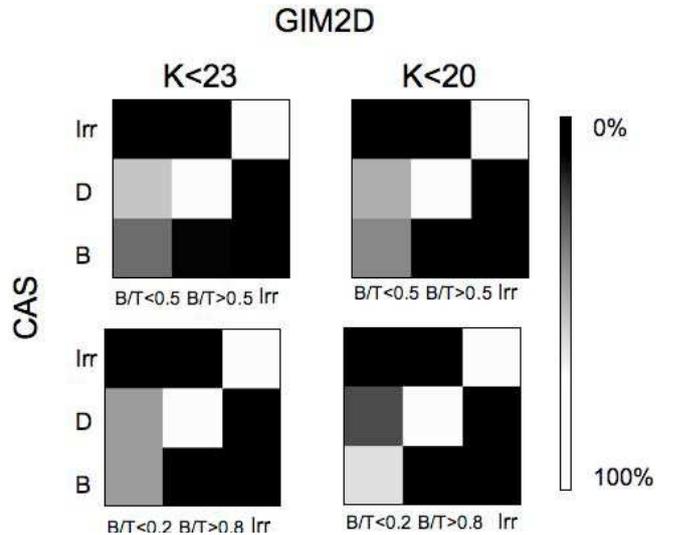
The main result after looking at the results of simulations is that, for objects brighter than  $K_s \sim 19$  ( $AB \sim 21$ ), the bulge fraction is recovered with a bias close to zero and a random error around 20%. This is true even for small objects ( $-1 < \log(r_{\text{hl}}) < -0.75$ ), and it is comparable to what is obtained for the brightest objects in the  $I$ -band with HST (Simard et al. 2002). For fainter magnitudes, we can see two main effects:

- first, an increasing bias in the bulge fraction ( $\sim 20\%$ ), in particular for large objects ( $\log(r_{\text{hl}}) > -0.75$ ). This is a well-known GIM2D effect for low S/N objects (Simard et al. 2002): the outer wings of steep surface brightness profiles, such as the  $r^{1/4}$  profile, are hidden in low S/N objects which artificially decreases the recovered bulge fraction. This can also be a consequence of errors in the sky-level estimate that causes that a fraction of the flux is hidden by noise.
- second, a slight increase in the random error, which becomes closer to 30%. This is a consequence of an increasing degeneracy of the parameters space with decreasing S/N.

But for most of the objects brighter than  $K_s = 19$  galaxy parameters can be estimated correctly ( $\sigma \sim 0.2$  and  $b < 0.1$ ), even for small objects ( $r_{\text{hl}} < 0.3''$ ) of size comparable the limits of space observations (see Sect. 6).

### 5.3. Results of the analysis and comparison of classifications

We classified the galaxies into three main morphological types according to the fitting results. One of the main results is that about  $12\% \pm 2.7\%$  of our sample corresponds to peculiar or irregular objects (10 objects out of 79). For the rest of the sample, the GIM2D analysis finds 21 ( $\sim 26\%$ ) bulge-dominated

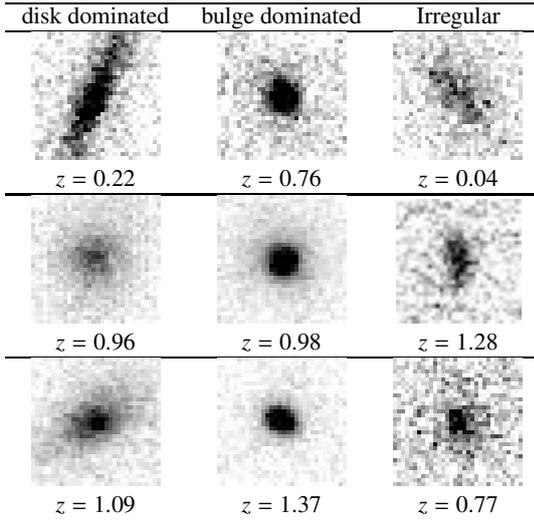


**Fig. 10.** Comparison of classification methods, show the probability that a galaxy classified with GIM2D is classified in the same morphological type by C-A. (see text for details).

galaxies ( $B/T > 0.5$ ) and 48 ( $\sim 60\%$ ) disk-dominated ( $B/T < 0.5$ ) while for the C-A classification, we find 54 ( $\sim 67\%$ ) disk-dominated galaxies and 15 ( $\sim 19\%$ ) bulge-dominated ones.

Looking in more detail into the reliability of the two classification schemes, we did a one-to-one comparison of the morphological types assigned by the GIM2D analysis or the C-A one (Fig. 10): we computed the probability that a galaxy classified using the GIM2D classification is classified with the same morphological type by C-A. The probability was computed by dividing the number of galaxies in each morphological C-A bin by the total number of galaxies of the same type selected with GIM2D. Overall, there is good agreement between both classifications in the whole sample. The probability that a disk dominated galaxy identified by GIM2D has the same morphological type in C-A classification is  $p = 0.81$ , but only  $p = 0.30$  for bulge-dominated galaxies including the faintest objects ( $K_s < 23$ ). For irregulars, it is obviously  $p = 1$  since the detection procedure is the same in both methods.

There might be two reasons why the classifications are not exactly the same. First, the S/N might cause discrepancies. Indeed, as we show in Sect. 5.2.2, at low S/N, GIM2D tends to under estimate the bulge fraction. This implies that some



**Fig. 11.** Example of classification in the three main morphological types at different redshifts. The image size is  $1.7' \times 1.7'$ .

galaxies detected by GIM2D as disk-dominated are in fact detected as bulge-dominated by C-A. Figure 10 shows the effect of reducing the limiting magnitude to  $K_s = 20$ : the fraction of objects classified as bulge dominated by GIM2D and C-A rises up to 0.67. Second, it might be a problem of definition. Indeed, the morphological bins are not exactly the same in both classifications. In particular, objects with intermediate morphological type (i.e.  $B/T \sim 0.5$ ) might cause discrepancies. If we remove those objects from the sample, 80% of the bulge-dominated objects and 95% of the disk-dominated objects detected by GIM2D are also detected by C-A with the same classification.

Either way, the comparison of both classifications allows a quantification of the error in classification of regular galaxies in the sense that it seems reasonable to think that the true value should be somewhere between the two results. The GIM2D estimate thus gives a lower limit for the early-type fraction and C-A the upper-limit and vice versa for the late-type fraction. This way, we conclude that the mixing of population in our sample is:  $24\% \pm 4\%$  of early-type galaxies,  $64\% \pm 4\%$  of late-type galaxies, and  $12\% \pm 2.7\%$  of irregular/peculiar galaxies.

Our results offer the first direct measurement of the distribution of galaxy in three morphological types at  $z \simeq 1$  from high spatial-resolution imaging in the  $K$ -band. We observe that the fraction of 12% of irregular objects at  $z = 1$  is significantly higher than the fraction of these objects in the local Universe, confirming from rest-frame data at  $\sim 1$  microns the well documented trend of this population increasing with redshift (e.g. Brinchmann et al. 1998). However, this result must be taken with caution. Indeed GIM2D accuracy decreases for objects fainter than  $K_s = 19$ , which represents 80% of the sample. Moreover at the faint end, the fraction of irregular objects can be overestimated because of the low S/N. But there are good reasons to consider this result significant. Even though there is an over estimation of disks in the faint end, the morphological classification bins are large enough to reduce the number of false classifications. Indeed, even in the zones where the random error in the bulge fraction estimate is  $\sim 0.3$  or larger, we do not classify a pure bulge ( $B/T \sim 1$ ) as a disk.

## 6. Comparison with ground-based and HST observations

In this section we compare our AO observations with ground-based and space observations.

### 6.1. Ground-based observations

Effective radii of local galaxies, except for compact dwarf galaxies, range from  $\sim 1.0$  to  $\sim 10$  kpc depending on their luminosity (Bender et al. 1992; Impey et al. 1996). Our spatial resolution of  $\sim 0.1''$  corresponds to about 1 kpc at  $z \sim 1$  and we should be able to determine morphological types even at  $z > 1$ . Thus, in order to estimate the performance of AO deep imaging and to justify the automated morphology classification, we compared our images with deep  $I$ -band seeing-limited images taken from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS)<sup>5</sup>. One of the so-called deep fields is centered on the COSMOS field, although it is smaller than the total COSMOS area (1 square degree out of 2). Here we used the release T0003 images (March 2006)<sup>6</sup>, especially the deep  $i'$  one, corresponding to a total integration time of 20 h, with an average  $FWHM$  of  $\sim 0.7''$ .

We compared real data by selecting a galaxy classified as a disk by GIM2D and C-A in the NACO data and by comparing it to the results obtained with CFHTLS data. We computed the surface brightness profile within the isophotal area enclosed by pixels  $3\sigma$  above the sky level. The corresponding limiting surface brightness is  $\mu = 20$  mag arcsec<sup>-2</sup> for the NACO image and  $\mu = 25$  mag arcsec<sup>-2</sup> for the MegaCam image. Sky levels and the corresponding isophotal areas were both determined using SEXTRACTOR.

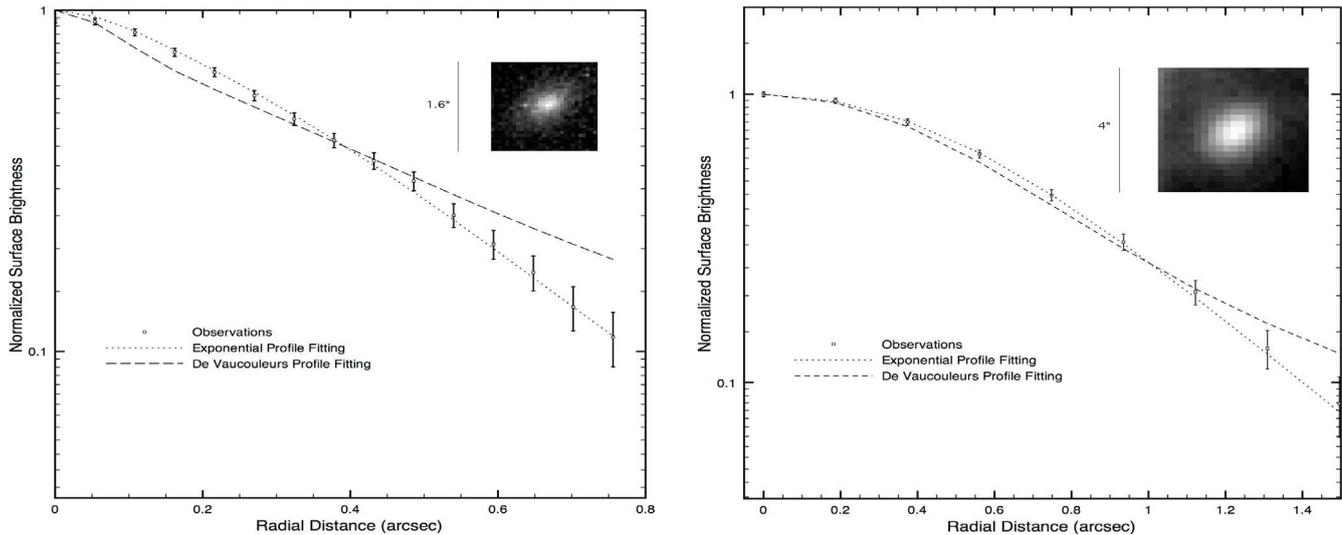
The surface brightness profile was fitted with both a PSF-convolved de Vaucouleurs profile and a PSF-convolved exponential profile. Figure 12 shows that, with seeing-limited observations, it is more difficult to establish a clear separation between both profiles at small distances from the galaxy center (i.e.  $\sim 0.2''$ ), even if the determination of the brightness profile is possible at much larger distance (i.e.  $\sim 1''$ ) thanks to the depth of the images and the low noise level of the sky background. This supports the results obtained with GIM2D, which show that a correct estimate of the bulge fraction is possible for small objects. Although ultra-deep sub-arcsecond imaging is powerful in terms of high number statistics, thanks to the wide field coverage, we consider that it is more rewarding to look at a smaller sample of galaxies, but with more reliable morphology determinations thanks to the spatial gain of the AO.

### 6.2. Space observations

We compare our images with space data taken from the COSMOS survey. Since our observed fields were selected within the COSMOS area, the same objects were observed with the HST-ACS in the  $I$ -band at high spatial resolution. We thus morphologically classified the 60 objects for which the photometric redshift are known (Sect. 4). We used those results to both estimate the effect of the observation band on morphology and to validate our method to divide the C/A plane. The C-A estimators were calibrated with simulations using the same method as for the  $K$  band data. Standard boundaries, from other existing works, were used to divide the C-A plane. Figure 16 shows the

<sup>5</sup> <http://www.cfht.hawaii.edu/Science/CFHTLS/>

<sup>6</sup> <http://terapix.iap.fr>



**Fig. 12.** Comparison with ground-based observations. We performed a profile fitting on the same real galaxy observed with NACO (*left*) and MegaCam (CFHTLS- $i'$  band, *right*). The galaxy magnitude is  $K_{AB} = 20.5$  and  $i' = 21.3$ . Surface brightness profiles were computed within the isophotal areas enclosed by pixels  $3\sigma$  above the sky level. The corresponding limiting surface brightness is  $\mu = 20$  mag arcsec $^{-2}$  for the NACO image and  $\mu = 25$  mag arcsec $^{-2}$  for the MegaCam image. The fit was done with a pure de Vaucouleurs and exponential profile.

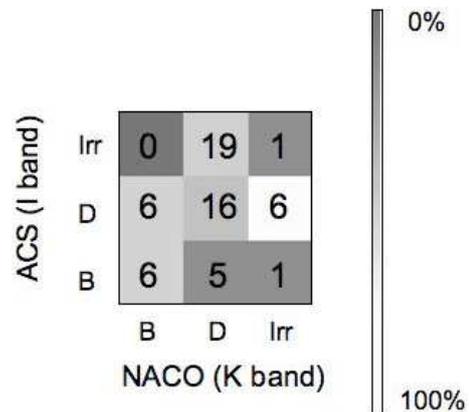
C-A plane cut. The figure also shows the border between bulge-dominated and disk-dominated galaxies obtained with the automatic method described in Sect. 5.2.1 for this population. We again find a negative slope for the border between disk and bulge dominated objects. We find for the whole sample  $32\% \pm 1.6\%$  irregulars,  $47\% \pm 1.5\%$  disk-dominated, and  $21\% \pm 2.5\%$  bulge dominated.

### 6.2.1. About boundaries

As said, the computed boundaries of the C-A plane are different from what can be found in the literature. Previous works have been done in the  $I$ -band using HST imaging (Abraham et al. 1996; Brinchmann et al. 1998). As we have a sample observed in the  $I$ -band, we are able to establish whether the change in the boundaries has significant consequences in the morphological classifications. To do so, we classified the  $I$ -band sample using the Brinchmann et al. (1998) boundaries and compared the results to the ones obtained with our method (Fig. 14). We find that there are no significant discrepancies between both classifications. We conclude that our method is valid and moreover has the key advantage being free of subjective judgments.

### 6.2.2. Rest-frame morphologies

We observe some discrepancies in the global morphological distributions between the  $I$  and  $K$  bands, in particular more perturbed morphologies are seen in the  $I$  band. When we look at each object individually (Fig. 13) we confirm this trend: there are uncertainties between  $K$  irregulars and  $I$  disks and between  $I$  disks and  $K$  bulges. Indeed an important fraction of bulge-dominated objects and disk-dominated objects detected by NACO are seen as disk-dominated and irregulars respectively, by ACS, as if there was a systematic trend that moves objects to later types when we move to shorter wavelengths. Certainly, the number of objects is small and a few mismatches cause high discrepancies in Fig. 13. However, this is an expected effect since ACS probes younger stellar populations. A visual inspection of some of the objects that present different morphologies reveals

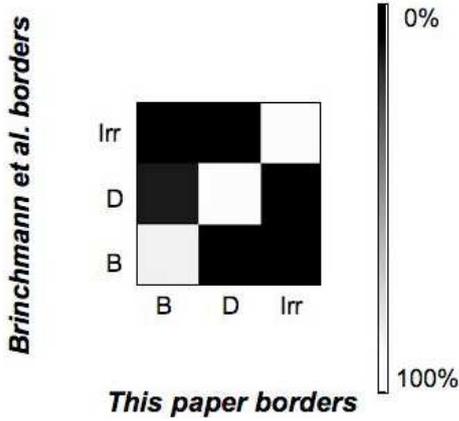


**Fig. 13.** Galaxy distribution: comparison between  $K$ -band and  $I$  band C-A classifications. The figure shows the probability that a galaxy in the  $K$ -band is classified in the same morphological type in the  $I$ -band.

that some of the ACS irregulars are in fact well-resolved spiral galaxies with inhomogeneities that probably increase the asymmetry indices.

In order to correctly compare both classifications we need to correct the measurements to estimate how galaxies would look if they were observed locally in the same photometric band. As a matter of fact, Brinchmann et al. (1998) showed that high- $z$  galaxies imaged by HST differ in appearance from their local counterparts because of their reduced apparent size and sampling characteristics, a lower S/N and reduced surface brightness with respect to the sky background and a shift in the rest wavelength of the observations. These effects combine to give some uncertainty in the morphological classifications of galaxies.

The first effect is a change in the concentration value measured at low redshift. Indeed, Brinchmann et al. (1998) draw the boundaries in the C-A plane using a local sample (Frei et al. 1996) visually classified. However, the concentration value depends on redshift, since the threshold is defined relative to the sky. Thus, less of the galaxy is sampled because of cosmological dimming. The solution they adopt is to correct  $C$  for this effect. We do not need a correction of the concentration in this paper



**Fig. 14.** Comparison of classifications with different boundaries. We repeat the morphological classification with the boundaries used by Brinchmann et al. (1998). We conclude that the results do not change significantly which supports the validity of the employed method.

because we use simulations that reproduce exactly the observing conditions to calibrate the C-A plane. The result is that boundaries are moved with respect to a local classification instead of changing the C value.

The shift in the rest-frame wavelength of observations is however more difficult to estimate. Indeed the question that arises here is whether the morphological type estimated at high redshift would be the same if observed at low redshift. When observing a galaxy in the  $K$ -band at redshift  $z \sim 1$ , the equivalent rest-frame wavelength will be around the  $z$  band, however, when observed in the  $I$ -band, the rest-frame will be around the  $B$  band. That implies that a mismatch can exist in the morphological classification since we are not probing the same morphological blocks. To correct for this effects we need to “move the objects” into a common rest-frame wavelength. This is the called morphological  $k$  correction. The method employed by Brinchmann et al. (1998) consists in determining the morphology from a local sample, redshifting the objects using SED models, and looking at the fraction of galaxies that move in to an other morphological class. A drift coefficient that characterizes the drift from category  $X$  to category  $Y$  is thus defined as

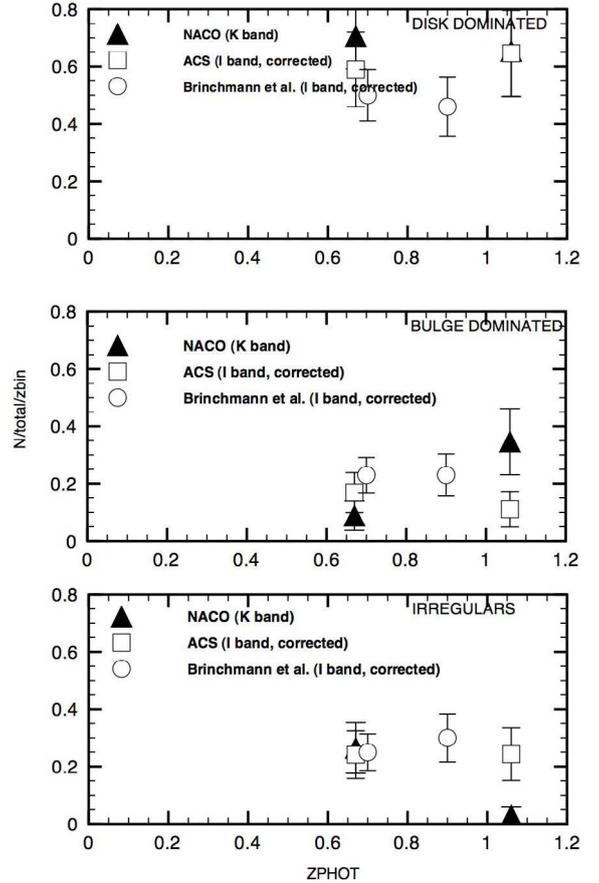
$$D_{XY} = \frac{N_{X \rightarrow Y}}{N_X}. \quad (2)$$

Once the fraction of missclassified objects is determined, the observed number of objects in class  $X$  can be related to the true number:

$$N_X^{\text{obs}} = N_X + \sum N_Y D_{YX} - N_X \sum D_{XY}. \quad (3)$$

Here we proceed as follows: Brinchmann et al. (1998) computed the coefficients to shift from the  $I$  observed morphology to the  $R$  rest frame morphology, we use those coefficients to correct the observed HST morphology of our sample to the one observed in the  $R$  rest frame band, since the filter used for observations is the same ( $F814W$ ). Once we have this corrected morphology, we can compare it to the NACO uncorrected morphology. This can be done because the observed sample is exactly the same in the  $K$  and in the  $I$ -band. If we were in the same rest-frame band, we should find the same morphology.

We use the coefficients computed by Brinchmann et al. (1998) to correct the ACS morphology and divide the sample into two redshift bins ( $z < 0.8$  and  $z > 0.8$ ). Results are shown in Fig. 15.

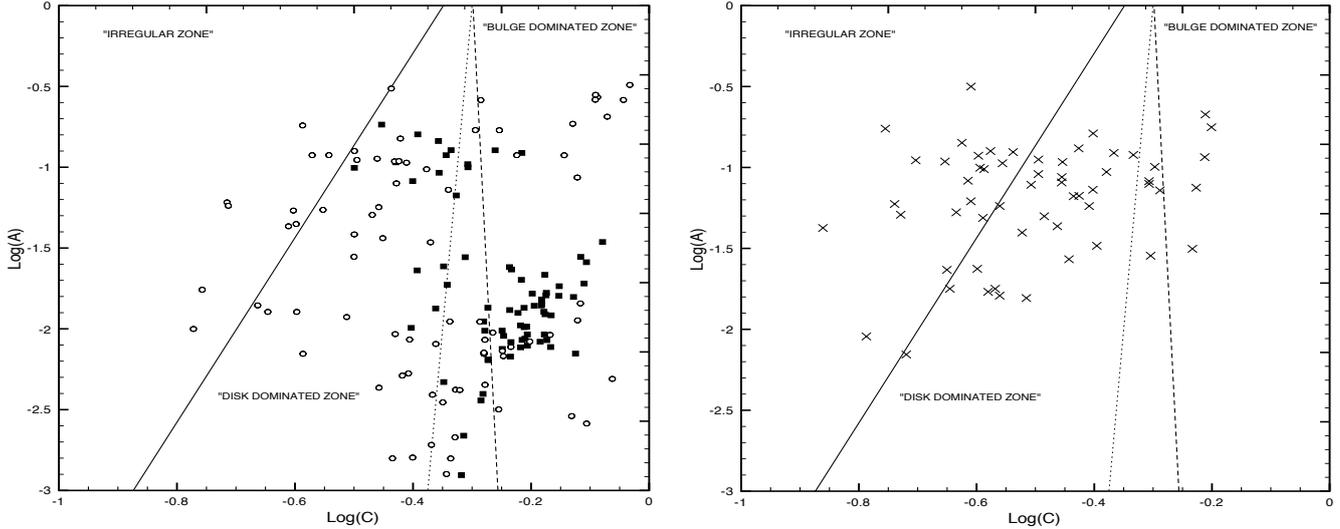


**Fig. 15.** Redshift distribution for the three morphological types. We plotted the Brinchmann et al. (1998) sample (circles) and our sample observed with ACS (squares) and with NACO (triangles). Brinchmann et al. (1998) and ACS data are corrected to the  $R$  rest-frame band. The NACO sample is observed from the  $K$ -band and no correction has been applied. The ACS and NACO samples have been separated into two redshift bins ( $z < 0.8$  and  $z > 0.8$ ). The represented redshifts are the median redshifts of each bin.

- At low redshift ( $z < 0.8$ ) the corrected  $I$ -band distributions and the uncorrected  $K$ -band are similar. The fact that there is no significant difference between the rest-frame  $R$  band and the uncorrected  $K$ -band population indicates that, when observing in the  $K$ -band, there is no need for *morphological  $k$  correction*.
- At higher redshift, the  $K$ -band distribution tends to give a larger fraction of bulge-dominated objects and a lower fraction of irregular objects than the  $I$ -band data. This can be caused by the fact that the drift coefficients computed by Brinchmann et al. (1998) might contain errors. Since they are based on SED fitting, it is logical to think that errors are more important for high-redshift objects. This would explain why the results are consistent at low redshift. It can also be a consequence of the size of the sample: as the number of objects per bin is small, errors in the classification of a small fraction can lead to discrepancies between the two bands. This issue will certainly be solved with a larger sample.

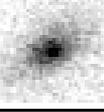
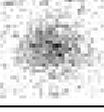
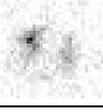
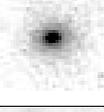
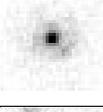
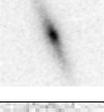
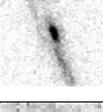
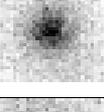
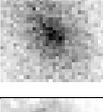
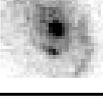
## 7. Summary and conclusions

We analyzed the morphologies of a sample of 79 galaxies in the near-infrared thanks to adaptive optics imaging at a resolution



**Fig. 16.** C-A cut for the ACS images. The same classification procedure has been repeated for the same sample observed with ACS in the  $I$ -band. *Left*: simulated objects. *Right*: real objects. Circles:  $B/T < 0.2$ , Filled squares:  $B/T > 0.8$ , crosses: real objects. Dotted line is the border used in Brinchmann et al. (1998) to separate bulge dominated from disk dominated, dashed line is the one computed in the paper.

**Table 5.** *Morphological  $k$  correction*: morphological differences when observing in the  $K$  and  $I$ -bands. The same objects observed in the  $K$  and  $I$  bands present different morphologies. The images size is  $1.7' \times 1.7'$ .

NACO		Z <sub>phot</sub>	ACS	
morph. type	image		image	morph. type
Disk dom.		1.090		Irr.
Irr.		0.4223		Irr.
Disk dom.		0.6689		Irr.
Bulge dom.		1.17460		Disk dom.
Disk dom.		0.6689		Irr.
Disk dom.		0.8861		Irr.
Disk dom.		0.7394		Irr.

of  $0.1''$ . Thanks to extensive simulations, we showed that adaptive optics can be used to obtain reliable high-resolution 4mm

morphological information in an automated way and is thus adapted to large observation programmes:

- Galaxy parameters (bulge fraction, disk scale length, bulge effective radius) can be estimated by means of model fitting with errors lower than 20% up to  $K_s = 19$ . This is comparable with space data.
- Galaxies can be separated into three main morphological bins up to  $K_s = 21$ , with at least 70% of good identifications and 66% up to  $K_s = 22$ .
- For fainter galaxies ( $22 < K_s < 23$ ) the accuracy of the morphological classification decreases. Simulations show that bulge-dominated and disk-dominated galaxies can be separated with only 55% accuracy at  $K_s = 23$ .

We obtain, for the first time, an estimate of the mix of morphological types of the galaxy population up to  $z \approx 1$  from ground-based  $K$ -band observations with high spatial resolution comparable or better than visible imaging from space. We demonstrate that estimating morphology from  $K$ -band data at  $z \approx 1$  is not affected by *morphological  $k$  correction*, as there is no significant difference between our population and the corrected  $I$ -band population. We find that the fraction of irregulars at  $z \approx 1$  is about  $12\% \pm 2.7\%$  using automated classification methods. This is higher than what is found in local surveys, confirming the well-established trend toward an increasing fraction of irregular galaxies with redshift as observed from surveys in the visible. Our small sample does not allow us to reach firm conclusions on the evolution of the fraction of late-type or early-type galaxies, but classifying galaxies from  $K$ -band AO imaging data is demonstrated to be reliable.

From this work it seems clear that adaptive optics can be used for observational cosmology with reliable accuracy, and consequently data of this type should contribute to a better understanding of galaxy evolution in the future. However, it is still a new technique and technical difficulties exist, such as variable PSF, small fields, subsampling and the need of guide stars that make the use of classical reduction methods more difficult. This is now easier with laser guide stars becoming available and new sets of utilities like the ones we are developing to enable easy data processing and analysis of adaptive optics data for the community. This opens the way to observing the large samples

required to reach a robust statistical accuracy. We are planning to enlarge our sample by observing a large number of areas around bright stars in the COSMOS field, which will strongly reduce uncertainties in the study of morphological evolution.

*Acknowledgements.* The authors want to thank the anonymous referee for many useful comments that greatly improved this paper.

## References

- Abraham, R., van den Bergh, S., Glazebrook, K., et al. 1996, *ApJS*, 107, 1  
 Abraham, R. G., Valdes, F., Yee, H. K. C., & van den Bergh, S. 1994, *ApJ*, 432, 75  
 Abraham, R. G., van den Bergh, S., & Nair, P. 2003, *ApJ*, 588, 218  
 Baker, A. J., Davies, R. I., Lehnert, M. D., et al. 2003, *A&A*, 406, 593  
 Baugh, C. M., Cole, S., & Frenk, C. S. 1996, *MNRAS*, 283, 1361  
 Bender, R., Burstein, D., & Faber, S. M. 1992, *ApJ*, 399, 462  
 Bershad, M. A., Lowenthal, J. D., & Koo, D. C. 1998, *ApJ*, 505, 50  
 Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393  
 Brinchmann, J., & Ellis, R. S. 2000, *ApJ*, 536, L77  
 Brinchmann, J., Abraham, R., Schade, D., et al. 1998, *ApJ*, 499, 112  
 Bundy, K., Ellis, R. S., & Conselice, C. J. 2005, *ApJ*, 625, 621  
 Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, *MNRAS*, 319, 168  
 Conselice, C. J., Bershad, M. A., & Jangren, A. 2000, *ApJ*, 529, 886  
 Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, *AJ*, 112, 839  
 Cresci, G., Davies, R., Baker, F., et al. 2006 [[arXiv:astro-ph/0607221](https://arxiv.org/abs/astro-ph/0607221)]  
 De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, *MNRAS*, 366, 499  
 de Vaucouleurs, G. 1948, *Annales d'Astrophysique*, 11, 247  
 Fall, S. M., & Efstathiou, G. 1980, *MNRAS*, 193, 189  
 Frei, Z., Guhathakurta, P., Gunn, J. E., & Tyson, J. A. 1996, *AJ*, 111, 174  
 Gardner, J. P. 1998, *PASP*, 110, 291  
 Gardner, J. P., Cowie, L. L., & Wainscoat, R. J. 1993, *ApJ*, 415, L9  
 Glazebrook, K., Peacock, J. A., Miller, L., & Collins, C. A. 1995, *MNRAS*, 275, 169  
 Hubble, E. 1936, *ApJ*, 415  
 Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006a, *ArXiv Astrophysics e-prints*  
 Ilbert, O., Lauger, S., Tresse, L., et al. 2006b, *A&A*, 453, 809  
 Impey, C. D., Sprayberry, D., Irwin, M. J., & Bothun, G. D. 1996, *ApJS*, 105, 209  
 Kleinmann, S. G. 1992, in *Robotic Telescopes in the 1990s*, ed. A. V. Filippenko, ASP Conf. Ser., 34, 203  
 Lilly, S., & The Zcosmos Team. 2006, *ApJ*, in press  
 Lilly, S. J., Le Fevre, O., Hammer, F., & Crampton, D. 1996, *ApJ*, 460, L1  
 Madau, P., Pozzetti, L., & Dickinson, M. 1998, *ApJ*, 498, 106  
 Maihara, T., Iwamuro, F., Tanabe, H., et al. 2001, *PASJ*, 53, 25  
 McCracken, H. J., Metcalfe, N., Shanks, T., et al. 2000, *MNRAS*, 311, 707  
 McLeod, B. A., Bernstein, G. M., Rieke, M. J., Tollestrup, E. V., & Fazio, G. G. 1995, *ApJS*, 96, 117  
 Mignoli, M., Cimatti, A., Zamorani, G., et al. 2005, *A&A*, 437, 883  
 Minowa, Y., Kobayashi, N., Yoshii, Y., et al. 2005, *ApJ*, 629, 29  
 Persson, S. E., Murphy, D. C., Krzemiński, W., Roth, M., & Rieke, M. J. 1998, *AJ*, 116, 2475  
 Sandage, A. 1961, *The Hubble atlas of galaxies* (Washington: Carnegie Institution)  
 Scoville, N. Z., & COSMOS Team. 2005, in *BAAS*, 1309  
 Simard, L., Willmer, C. N. A., Vogt, N. P., et al. 2002, *ApJS*, 142, 1  
 Vapnik, V. 1995, Springer-Verlag, 536  
 Wolf, C., Meisenheimer, K., Rix, H.-W., et al. 2003, *A&A*, 401, 73  
 Zucca, E., Ilbert, O., Bardelli, S., et al. 2006, *A&A*, 455, 879

# A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images

## I. Method description<sup>★</sup>

M. Huertas-Company<sup>1,4</sup>, D. Rouan<sup>1</sup>, L. Tasca<sup>3</sup>, G. Soucail<sup>2</sup>, and O. Le Fèvre<sup>3</sup>

<sup>1</sup> LESIA-Paris Observatory, 5 place Jules Janssen, 92195 Meudon, France  
e-mail: marc.huertas@obspm.fr

<sup>2</sup> Laboratoire d'Astrophysique de Toulouse-Tarbes, CNRS-UMR 5572 and Université Paul Sabatier Toulouse III, 14 avenue Belin, 31400 Toulouse, France

<sup>3</sup> LAM, Marseille Observatory, Traverse du Siphon, Les trois Lucs, BP 8, 13376 Marseille Cedex 12, France

<sup>4</sup> IAA-C/ Camino Bajo de Huétor, 50, 18008 Granada, Spain

Received 6 September 2007 / Accepted 7 November 2007

### ABSTRACT

**Context.** Morphology is the most accessible tracer of the physical structure of galaxies, but its interpretation in the framework of galaxy evolution still remains a problem. Its dependence on wavelength renders the comparison between local and high redshift populations difficult. Furthermore, the quality of the measured morphology being strongly dependent on the image resolution, the comparison between different surveys is also a problem.

**Aims.** We present a new non-parametric method to quantify morphologies of galaxies based on a particular family of learning machines called support vector machines. The method, which can be seen as a generalization of the classical C/A classification but with an unlimited number of dimensions and non-linear boundaries between decision regions, is fully automated and thus particularly well adapted to large cosmological surveys. The source code is available for download at <http://www.lesia.obspm.fr/~huertas/galsvm.html>

**Methods.** To test the method, we use a seeing limited near-infrared ( $K_s$  band, 2, 16  $\mu\text{m}$ ) sample observed with WIRCam at CFHT at a median redshift of  $z \sim 0.8$ . The machine is trained with a simulated sample built from a local visually classified sample from the SDSS, chosen in the high-redshift sample's rest-frame ( $i$  band, 0.77  $\mu\text{m}$ ) and artificially redshifted to match the observing conditions. We use a 12-dimensional volume, including 5 morphological parameters, and other characteristics of galaxies such as luminosity and redshift. A fraction of the simulated sample is used to test the machine and assess its accuracy.

**Results.** We show that a qualitative separation in two main morphological types (late type and early type) can be obtained with an error lower than 20% up to the completeness limit of the sample ( $KAB \sim 22$ ), which is more than 2 times better than what would be obtained with a classical C/A classification on the same sample and indeed comparable to space data. The method is optimized to solve a specific problem, offering an objective and automated estimate of errors that enables a straightforward comparison with other surveys. Selecting the training sample in the high-redshift sample rest-frame makes the results free from wavelength dependent effects and hence its interpretation in terms of evolution easier.

**Key words.** galaxies: fundamental parameters – galaxies: high-redshift – methods: data analysis

## 1. Introduction

The process of galaxy formation and the way galaxies evolve is still one of the key unresolved problems in modern astrophysics. In the currently accepted hierarchical picture of structure formation, galaxies are thought to be embedded in massive dark halos that grow from density fluctuations in the early universe (Fall & Efstathiou 1980) and initially contain baryons in a hot gaseous phase. This gas subsequently cools, and some fraction eventually condenses into stars (Lilly et al. 1996; Madau et al. 1998). However, many of the physical details remain uncertain, in particular the process and history of mass assembly. One

classical observational way to test those models is to classify galaxies according to morphological criteria, i.e., the organization of their brightness as projected on the sky's plane and observed at a particular wavelength, defined in the nearby Universe (Hubble 1936; de Vaucouleurs 1948; Sandage 1961), and to follow this classification across time (Abraham et al. 1996; Simard et al. 2002; Abraham et al. 2003). Comparison of distant populations with those found in the nearby Universe might help to clarify the formation history of the galaxy (Cole et al. 2000; Baugh et al. 1996). Progress in this field has come from observing deeper and larger samples, but also from obtaining higher spatial resolution at a given flux and at a given redshift. In the visible, progress has been simultaneous on those two fronts, thanks in particular to the ultra-deep HDF fields observed with the Hubble Space Telescope (HST). HST imaging has brought observational evidence that galaxy evolution is differentiated with respect to morphological type and that a large fraction of distant galaxies

<sup>★</sup> Based on observations obtained at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council of Canada, the Institut National des Sciences de l'Université of the Centre National de la Recherche Scientifique of France, and the University of Hawaii.

have peculiar morphologies that do not fit into the elliptical-spiral Hubble sequence (Brinchmann et al. 1998; Wolf et al. 2003; Ilbert et al. 2006b).

However, the difficulty in quantifying morphology of high redshift objects with a few simple, reliable measurements is still a major obstacle. Indeed, with the increasing number of cosmological surveys available today, classical visual classifications become useless and automated methods must be employed. Globally there exist two main approaches. The first one, known as parametric, consists in modeling the distribution of light with an analytic model and fitting it to the real galaxy. A commonly used parameter in this approach is the bulge-to-disk (B/D) light ratio that correlates with qualitative Hubble type classifications, and can be obtained by fitting a two-component profile (Simard et al. 2002; Peng et al. 2002). The main advantage of such a method is that the fitting output provides a quantitative morphology, i.e. a complete set of parameters that describe the galaxy's shape (e.g. disk scale length, bulge effective radius). Results are, unfortunately, often degenerated because of the high number of parameters to be adjusted (Huertas-Company et al. 2007), even when the residuals are almost null, and the obtained accuracy strongly depends on the observing conditions (e.g. angular resolution, S/N). Moreover, this approach assumes that the galaxy is well described by a simple, symmetric profile, which is not true for irregular or well resolved objects.

The second approach is called non-parametric and basically consists in measuring a set of well-chosen parameters that correlate with the Hubble type. The main advantage of this method is that it does not assume a particular analytic model and can therefore be used to classify regular as well as irregular galaxies. The resulting morphology however will be more qualitative. Abraham et al. (1996) first proposed this method by defining the concentration and asymmetry (C and A) parameters. They showed that plotting those values in a 2D plane results in a quite good separation between the three main morphological types (early type, late type and irregulars). Subsequent authors then modified the original definitions to make C and A more robust to surface-brightness selection, centering errors or redshift dependence (Brinchmann et al. 1998; Wu 1999; Bershadsky et al. 2000; Conselice et al. 2000), and introduced new parameters. In particular a third parameter, the smoothness (S), was proposed by Conselice et al. (2003) and gave its name to the CAS morphological classification system. More recently Abraham et al. (2003) and Lotz et al. (2004) proposed two new parameters: the Gini coefficient that correlates with concentration, and the M20 moment. Each of these parameters brings a different amount of information concerning the galaxy shape. However, because of the so-called "curse of dimensionality", adding additional parameters does not always result in a significant gain. The information added should indeed outweigh the penalty of complicating the classification procedure, and this depends on the analyzed sample. Consequently, one interesting question is under which conditions it becomes useful to use all these parameters simultaneously. There is no way, however, with classical approaches to use more than 3 parameters simultaneously. Bershadsky et al. (2000) made a first attempt to do a multi-parameter analysis on a nearby sample using a 4 dimensional space, including concentration and asymmetry as well as luminosity and color information. They found correlations between those parameters and defined six 2D planes resulting from the combinations of those parameters. The classification however, was done independently in each plane without considering all the information simultaneously. In the framework of the COSMOS consortia (Scoville & COSMOS Team 2005), Scarlata et al. (2006) have recently

made a step forward by proposing a multi-parameter classification scheme (ZEST) based on the positions of galaxies in a three dimensional space, resulting from a principal component analysis on a 5 dimensional space. The method uses almost all the information contained in the 5 parameters, but the final calibration is done in 3 dimensions.

Another key point in this kind of analysis is to correctly calibrate the volume, i.e. to correctly estimate the decision regions. One approach is to use boundaries defined in the nearby universe from a visually classified sample and assume that they will remain unchanged for a sample at high redshift, observed at a different wavelength and with another instrument (Abraham et al. 1996). However, it is well known that the galaxy morphology depends on wavelength (K-correction) and on the observing conditions, which is why some corrections should be applied to take these effects into account (Brinchmann et al. 1998). Another approach consists in visually classifying a fraction of the sample and plotting the boundaries according to the positions of galaxies in the space (Menanteau et al. 2006; Scarlata et al. 2006). This of course takes into account the observing conditions of the sample but requires enough resolution and S/N to be able to decide the galaxy morphology visually. This is possible for space observations but becomes more difficult for ground-based observations, where the low resolution does not allow a reliable visual classification. In all these approaches, boundaries are forced to be linear (2D lines or hyper-planes) and are generally plotted manually, which introduces a subjective element that makes a correct estimate of errors more difficult.

In this paper, therefore, we propose a generalization of the non-parametric classification that uses an unlimited number of dimensions and non-linear separators, enabling us to use all the information brought by the different morphological parameters simultaneously. The approach uses a particular class of learning machines (called support vector machines) that finds the optimal decision regions in a volume using a training set. Here, we build this training set from a local sample that is transformed to reproduce the physical and instrumental properties of the science sample, allowing one to use it even on seeing-limited observations. The algorithm defines, in an automated way, the optimal decision regions using multi-dimensional hyper-surfaces as boundaries. It therefore allows a straightforward comparison between different science samples. The classification scheme that we propose is intended as a framework for future studies on large cosmological fields.

The paper proceeds as follows: generalities on pattern recognition, and in particular on support vector machines (SVM), are described in the next section. In Sect. 3 we make sure that SVM work properly when applied to a nearby sample, and investigate the effect of adding dimensions to a well-resolved sample. In Sect. 4, we describe the general steps of the proposed method to classify high-redshift objects. We show, in particular, how the training set is built to reproduce the real sample properties (4.1), we define the parameters measured for the morphological classification (4.3) and we finally describe several tests performed to probe the accuracy of the method (4.4).

We use the following cosmological parameters throughout the paper:  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $(\Omega_M, \Omega_\Lambda, \Omega_k) = (0.3, 0.7, 0.0)$ .

## 2. Generalities on pattern recognition

Suppose we take a set of observations of a given phenomenon, in which each observation consists of a vector  $\mathbf{x}_i \in \mathbb{R}^n$ ,  $i = 1, \dots, l$

and an associated “truth”  $y_i$ . For instance, in a classical concentration and asymmetry classification plane,  $x_i$  would be a 2D vector whose components are the concentration and the asymmetry, and  $y_i$  would be 0 if the galaxy is irregular, 1 if it is disk dominated, and 2 if it is bulge dominated. We then call learning machine, a machine whose task is to learn the mapping  $x_i \mapsto y_i$  defined by a set of possible mappings  $x \mapsto f(x, \alpha)$ . A particular choice of  $\alpha$  generates what is called a “trained machine”.

### 2.1. Support vector machines

Support vector machines are a particular family of learning machines, first introduced by Vapnik (1995) as an alternative to neural networks, and that have been successfully employed to solve clustering problems, specially in biological applications.

In order to simplify the description of the most important points concerning SVM we will focus on a 2 class classification problem:  $\{x_i, y_i\}, i = 1, \dots, l, y_i \in \{-1, 1\}, x_i \in \mathbb{R}^d$  without loss of generalization. The basic idea is to find a hyperplane that separates the positive from the negative examples. If this plane exists, the points  $x$  that lie on the hyperplane satisfy  $w \cdot x + b = 0$ , with  $w$  normal to the hyperplane;  $|b|/\|w\|$ , the perpendicular distance from the hyperplane to the origin.  $d_+(d_-)$ , will then be the shortest distance from the separating hyperplane to the closest positive (negative) example. The “margin” is defined:  $d_+ + d_-$ . The algorithm will then simply look for the separating hyperplane with the largest margin. This can be formulated as follows:

1.  $x_i \cdot w + b \geq +1$  for  $y_i = +1$
2.  $x_i \cdot w + b \leq -1$  for  $y_i = -1$ .

The training points for which the equalities hold and whose removal would change the solution are called support vectors (Fig. 1).

It is possible (and it is the most common case) that there is no linear hyperplane that perfectly separates the two data sets. In this case we can relax constraints by introducing a positive slack variable  $\xi_i, i = 1, \dots, l$  and the equations become then:

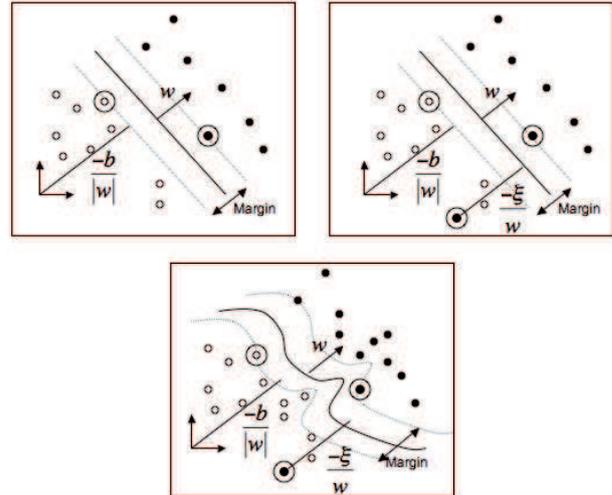
1.  $x_i \cdot w + b \geq +1 - \xi_i$  for  $y_i = +1$
2.  $x_i \cdot w + b \leq -1 + \xi_i$  for  $y_i = -1$ .

The global effect is to change the objective function to be minimized to  $\|w\|^2/2 + T(\sum \xi_i)$ , where  $T$  is a parameter to be chosen by the user, a larger  $T$  corresponding to assigning a higher penalty to errors.

Another feature that can be added to solve more complex problems are non linear decision functions. To do this, we map the data to some other (possibly infinite dimensional) Euclidian space  $H: \Phi: \mathbb{R}^d \mapsto H$  where the data can be linearly separable by some hyperplane. Since the only way in which the data appear in the training problem is in the form of dot products  $x_i \cdot x_j$  then the training algorithm would only depend on the data through dot products in  $H$ , i.e. on functions of the form  $\Phi(x_i) \cdot \Phi(x_j)$ . If there is a “kernel function”  $K$  such that  $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ , we would never need to explicitly know what  $\Phi$  is. Examples of kernels are:  $K(x, y) = (x \cdot y + 1)^p$  (Polynomial),  $K(x, y) = e^{-g\|x-y\|^2}$  (Gaussian RBF).

In summary, SVM are a particular family of learning machines that:

- for linearly separable data, simply look for the optimal separating hyperplane between distributions by maximizing the margin;
- for non separable data, a “tolerance” parameter  $C$  must be added that controls the tolerance to errors;



**Fig. 1.** 2D illustration of the three cases of SVM classification. *Top left:* linearly separable data with linear boundaries. *Top right:* non-linearly non-separable data with linear boundaries. *Bottom:* non-linearly non-separable data with non-linear borders.

- for non linear non separable data, a kernel function is built that maps the space into a higher dimensional space where the data are linearly separable. The kernel parameters must be then adjusted too.

### 2.2. Application to galaxies

Abraham et al. (1996) proposed the idea of measuring some well-chosen parameters on a galaxy image that can be easily correlated with its morphology. In their paper they introduced the concentration, which basically measures the fraction of light contained in an inner isophote, and the asymmetry, which measures the degree of symmetry of the galaxy. They showed that plotting the logarithms of those values in a 2D plane results in quite a good separation between the three main morphological populations: early-type, late-type and irregulars. They consequently plotted linear separators to define the regions and classified a set of galaxies with unknown morphology according to their positions in the so-called C/A plane. In other words, they tried to maximize the margins between 3 populations in a 2 dimensional space using linear separators and a logarithmic kernel. The same task can be done in a 3 dimensional space (CAS, Conselice et al. 2003) but it becomes simply impossible with more than 3 dimensions. In this sense, SVM offer a straightforward generalization of this method since they can separate samples with an unlimited number of dimensions and use non-linear boundaries.

Previous works (Abraham et al. 1996; Brinchmann et al. 1998) have shown that morphological classification is far from being a linearly separable problem, since the contamination in the C/A plane is quite high. We have chosen therefore to use the most general SVM, i.e. a non linear machine with an RBF kernel. A machine is thus parameterized with two parameters: the tolerance ( $T$ ) and the kernel exponential factor ( $g$ ). Each possible combination of those two parameters generates a family of functions  $f_{T,g}(\alpha, x_i)$ .  $T$  and  $g$  must be fixed before performing the training and  $\alpha$  is the result of the training procedure. There exist several techniques for finding the best  $T$  and  $g$  values for a given problem; here we will use a cross-validation method described in Chang & Lin (2001). This simply consists in performing a

**Table 1.** Comparison of the accuracy of three classifications of the SDSS sample: classical C/A, SVM C/A and 4D SVM. The table shows for each method the relations between the visual and the predicted morphological classes. The number of objects are enclosed in parentheses. (See text for details.)

	Classical C/A		SVM C/A		SVM 4D	
	Early-type	Late-type	Early-type	Late-type	Early-type	Late-type
Visual early-type	0.80 (254)	0.09 (17)	0.79 (256)	0.08 (15)	0.79 (251)	0.10 (20)
Visual late-type	0.20 (65)	0.91 (172)	0.21 (72)	0.92 (166)	0.21 (67)	0.90 (171)

systematic search over a grid of possible values and selecting the pair that gives the best results.

Our goal is therefore to train a support vector based machine to estimate the morphology of a high redshift sample. Throughout the paper we use the free available package *libSVM* (Chang & Lin 2001). The procedure is basically the same as in a classical C/A classification but using a trained SVM to plot the optimal boundaries. As we show below, this allows us to use more than two morphological parameters simultaneously, and also to measure errors in an automated and objective way, which is essential for comparing different classifications.

### 3. Test on a well-resolved nearby sample

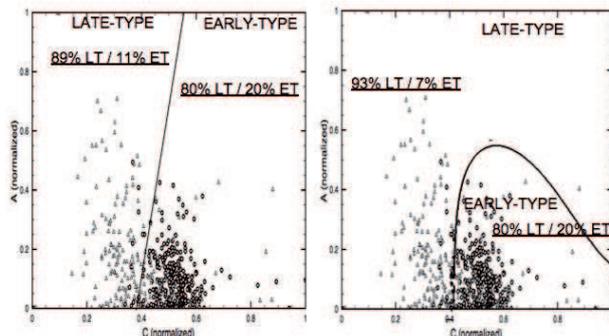
#### 3.1. Classical C/A classification versus 2D SVM

In order to verify that SVM work properly when applied to morphological classification of galaxies, we start with a simple test, i.e. classifying a local sample from the Sloan Digital Sky Survey (SDSS) in the *i* band, that has been visually classified (Tasca & White 2005). Galaxies are nearby and consequently well-resolved and with a high S/N. Classical C/A classifications have been proved to give good results in such cases (e.g. Abraham et al. 1996; Menanteau et al. 2006), therefore the idea is to verify that we get at least the same results using SVM and that no extra-biases are introduced.

We therefore measure concentration and asymmetry parameters (see Sect. 4.3 for details). On the one hand we try to plot the best linear boundary by eye as usually done to separate galaxies into two classes (late type and early type), while on the other hand we train a SVM, finally, we compare the outputs. Figure 2 shows the two resulting boundaries. The shape of the boundaries are quite different since SVM does not produce a linear boundary, but when looking at the global accuracy we see that both methods are fully consistent. Indeed, the completeness (the fraction of visual classified galaxies that are correctly recovered) and the contaminations (the fraction of visual classified galaxies that are misclassified) are practically the same for the two methods (see Table 1).

To confirm this consistency and to verify that no extra biases are introduced, we also made a one-to-one comparison of all the galaxies classified with the two methods. We find that 98% (94%) of the early-type (late-type) galaxies classified with the classical C/A method are also classified as early-type (late-type) using the trained SVM.

We conclude that for a high S/N well-resolved sample, such as the SDSS sample, the use of SVM to plot the boundaries is equivalent to using classical procedures. The major advantage, however, is that the boundary is plotted automatically to minimize the errors.



**Fig. 2.** C/A classical classification (left) versus C/A SVM classification (right) of 500 nearby objects. Triangles are galaxies visually classified as late-type and circles are galaxies visually classified as early-type. Numbers show the probability that the predicted morphological type is the same as the visual one.

#### 3.2. Classical C/A classification versus 4D SVM

Since one of the main advantages of using SVM is that it can work with an unlimited number of parameters, we investigate the effect of adding dimensions to the SVM classification. We thus classify the same sample as above but with four morphological parameters instead of two: concentration, asymmetry, smoothness and gini (see Sect. 4.3 for details on how they are calculated) and compare the outputs. Results are shown in Table 1. We see that there is no significant gain for this particular case. This suggests that, as proven in previous works (e.g. Abraham et al. 1996; Menanteau et al. 2006), when dealing with a well-resolved and high S/N sample, concentration and asymmetry are enough to obtain an accurate morphological classification.

### 4. Going to higher redshift

When observing objects at higher redshift with a ground-based telescope the S/N decreases, galaxies become poorly resolved and consequently more symmetric and less concentrated (e.g. Conselice et al. 2000). The separation in the C/A plane turns out to be less clear. For this reason, space data such as HST imaging are widely used for those purposes, and classifications based on colors are usually adopted for ground-based data (e.g. Zucca et al. 2006). It is known however (e.g. Arnouts et al. 2007) that a classification based only on colors is highly contaminated by the presence, for instance, of an important population of “blue” early-type galaxies, especially at high redshift where the red sequence is building up. This is one of the reasons why classifications based on morphological criteria are preferred. Indeed, with the increasing amount of data coming from ground-based surveys becoming available today, it would be interesting to know if it is possible to obtain at least a rough morphological classification from these observations. In the following sections we therefore investigate whether the possibilities of using a large number

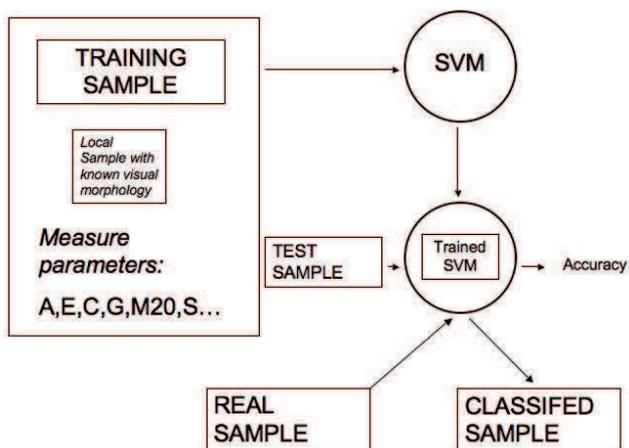


Fig. 3. Steps for morphological classification (see text for details).

of parameters and non-linear boundaries offered by support vector machines can help to increase the accuracy of “pure” morphological classifications on high-redshift ground-based data.

#### 4.1. Description of the employed method

The proposed procedure can be summarized in 4 main steps (Fig. 3):

1. Build a training set: for that purpose, we select a nearby visually classified sample at a wavelength corresponding to the rest-frame of the high redshift sample to be analyzed. We then move the sample to the proper redshift and image quality and drop it in the high  $z$  background. This is fully described in Sect. 4.2.
2. Measure a set of morphological parameters on the sample.
3. Train a support vector based learning machine with a fraction of the simulated sample and use the other fraction to test and estimate errors.
4. Classify real data with the trained machine and correct for possible systematic errors detected in the testing step.

In the following sections, we describe each of the steps enumerated above.

#### 4.2. The training set

The most important step in obtaining the morphology with a non-parametric method is to correctly calibrate the volume filled by the data in the multi-dimensional space. This is a critical step since it will determine the decision regions that will be used to perform the classification. Indeed, galaxy morphology depends on the physical properties of the galaxy (luminosity, redshift, wavelength) and on the observing conditions (background level, resolution). A suitable calibration set should consequently reproduce closely all the properties of the sample to be analyzed. One classical approach consists in visually classifying a fraction of the sample and using it as a training set to optimize boundaries (Menanteau et al. 1999, 2006). However, this is not possible for seeing-limited data where the resolution is too low to enable a reliable visual classification. Here, we decide to simulate the high redshift sample from a visually classified local catalog, selected in the rest-frame color of the high redshift sample. This has three main advantages: first, it is free from K-correction effects, second it does not introduce any modeling effect, since the used galaxies are real and finally, the training set is built to reproduce

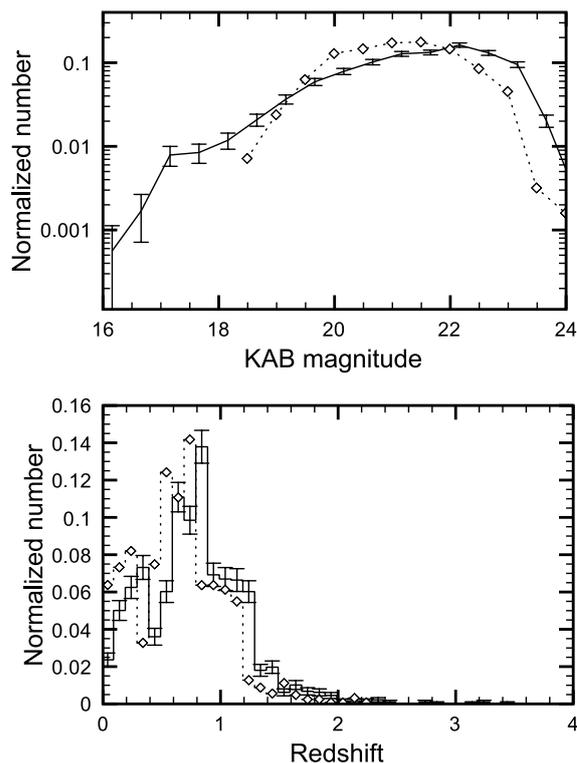


Fig. 4. Magnitude and redshift distributions of the real and simulated sample. Solid line: real sample. Dotted line: simulated sample. Error bars show poissonian errors for the real sample. See text for explanations concerning the differences between the simulated and real distributions.

the observing and physical properties of the sample to be analyzed, but it is classified locally, so it does not require a specially high resolution.

##### 4.2.1. Real sample

In order to test the method, we work on a sample of galaxies observed with WIRCam at CFHT in the near infrared  $K_s$  band. The field is part of the Canada-France Hawaii Telescope Legacy Survey (CFHTLS) Deep survey and its near infrared follow-up, and it is centered on the COSMOS area (Scoville & COSMOS Team 2005). We use a cutout of  $10' \times 10'$  to perform all the tests. The sample is complete up to  $K(AB) = 22$  and the median photometric redshift is  $\sim 0.8$  (Fig. 4). Images are reduced with the Terapix pipeline<sup>1</sup> and have a pixel scale of  $0.15''$  with a mean  $FWHM$  of  $0.7''$ . These data are particularly interesting because  $K$ -band data have the advantage of probing old stellar populations in the rest-frame, enabling a determination of galaxy morphological types unaffected by recent star formation. Moreover, no space data in this wavelength range are available today. Photometric redshifts come from the publicly available catalogue from Ilbert et al. (2006a), computed with the *LePhare*<sup>2</sup> code on the CFHTLS Deep survey Terapix release T0003 and its multi-color photometric catalogs.

<sup>1</sup> <http://terapix.iap.fr>

<sup>2</sup> <http://cencos.oamp.fr/cencos/CFHTLS/>

#### 4.2.2. Building the sample

We therefore used a local catalog of 1472 objects from the Sloan Digital Sky Survey in the  $i$  band, which roughly corresponds to the rest-frame of the  $K$ -band at  $z \sim 1$  and has been visually classified (Tasca & White 2005).

We first generate a random pair of (magnitude, redshift) values with a probability distribution that matches the real magnitude and redshift distribution of the sample to be simulated (see Fig. 4).

Then, for every galaxy stamp, we proceed in four steps:

1. First, we remove all the foreground stars and all other sources that do not belong to the galaxy itself. For that purpose we use the SExtractor segmentation map (Bertin & Arnouts 1996) and replace all the surrounding sources with a random noise with the same statistics (mean value and variance) as the real background noise. The foreground stars that fall within the galaxy are replaced with the mean value in the galaxy area.
2. Second, we degrade the resolution to reach that at high redshift: we measure the  $FWHM$  at high redshift ( $f_{hz}$ ), convert it to Kpc using a standard  $\Lambda$ CDM cosmology and deduce the resolution that the local galaxy must have ( $f_l$ ). Then the image is convolved with a 2D Gaussian function of  $FWHM = \sqrt{(f_{hz}^2 - f_l^2)}$ , where  $f_l$  is the local galaxy's initial resolution.
3. Third, the image is binned to reach the expected angular size at high redshift with the  $0.15''$  pixel scale. In this step, the image is also scaled to its new magnitude. In the scaling procedure we force the final mean background level of the simulated stamp to be at least 3 times lower than the real background. This is to avoid the local noise dominates over the high-redshift noise when dropping the galaxy in a real background. This implies that objects that are too bright (typically  $K_s < 17$ ) cannot be simulated since the necessary scaling factor is too small, and explains the difference between the real and the simulated magnitude distribution in Fig. 4. The difference in the faint end is due to the fact that some simulated objects are not detected by SExtractor.
4. Finally, we drop the galaxy in a real background image.

Figure 5 illustrates the entire procedure for a spiral galaxy.

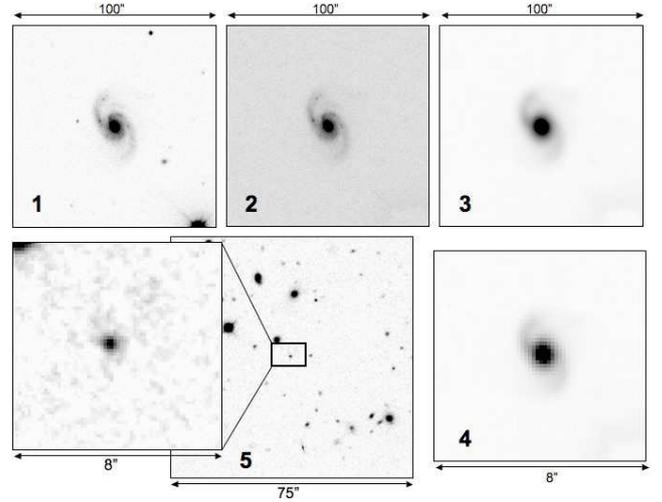
In summary, we simulate a high redshift sample from a local sample, selected in the high redshift sample's rest-frame to avoid  $K$ -correction effects. The sample reproduces the observing conditions (background level, noise, resolution) and physical properties (redshift and magnitude distribution) of the sample to be analyzed.

#### 4.3. Measuring morphological parameters

Once the simulated galaxies are dropped in a real background, we measure the following 5 morphological parameters:

- Concentration: basically, this measures the ratio of light within a circular or elliptical inner aperture to the light within a circular or elliptical outer aperture. Generally, it is defined in slightly different ways by different authors. Here we adopt the Bershady et al. (2000) definition as for the ratio of the circular radii containing 20% and 80% of the “total flux”:

$$C = 5 \log(r_{80}/r_{20}). \quad (1)$$



**Fig. 5.** Example of simulation for a galaxy. 1: SDSS  $i$  band image; 2: image after subtraction of foreground stars; 3: image after convolution; 4: image after binning; 5: final simulated field with real and simulated galaxies.

We use Conselice's (2003) definition of the total flux as the flux contained within  $1.5r_p$  (Petrosian radius). For the concentration measurement, the galaxy's center is that determined by the asymmetry minimization (see below).

- Asymmetry: this quantifies the degree to which the light of a galaxy is rotationally symmetric. It is measured by subtracting the galaxy image rotated by  $180^\circ$  from the original image:

$$A = \frac{1}{2} \left( \frac{\sum |I(i, j) - I_{180}(i, j)|}{\sum I(i, j)} - \frac{\sum |B(i, j) - B_{180}(i, j)|}{\sum I(i, j)} \right), \quad (2)$$

where  $I$  is the galaxy image and  $I_{180}^\circ$  is the galaxy image rotated by  $180^\circ$  about the galaxy's central pixel, and  $B$  is the average asymmetry of the background. The central pixel is determined by minimizing  $A$ .

- Smoothness: developed by Conselice et al. (2000), this quantifies the degree of small-scale structure. The galaxy image is smoothed by a boxcar of given width and then subtracted from the original image:

$$S = \frac{1}{2} \left( \frac{\sum |I(i, j) - I_S(i, j)|}{\sum I(i, j)} - \frac{\sum |B(i, j) - B_S(i, j)|}{\sum I(i, j)} \right), \quad (3)$$

where  $I_S$  is the galaxy's image smoothed by a boxcar of width  $0.25r_p$ .

- Moment of Light: introduced by Lotz et al. (2004), the total second-order moment  $M_{\text{tot}}$  is the flux in each pixel  $f_i$  multiplied by the squared distance to the center of the galaxy, summed over all the galaxy pixels assigned by the SExtractor segmentation map:  $M_{\text{tot}} = \sum f_i [(x_i - x_c)^2 + (y_i - y_c)^2]$ , where  $x_c$  and  $y_c$  is the galaxy's center. The second-order moment of the brightest regions of the galaxy traces the spatial distribution of any bright nuclei, bars, spiral arms and off-center star clusters. We define  $M_{20}$  as the normalized second-order moment of the 20% brightest pixels of the galaxy.
- Gini Coefficient: this is a statistic based on the Lorentz curve, i.e. the rank-ordered cumulative distribution function of a population's wealth, or in this case a galaxy's pixel values (Abraham et al. 2003). For the majority of local galaxies, the Gini coefficient is correlated with the concentration index and increases with the fraction of light in a central

**Table 2.** Comparison of the accuracy of three classifications of the WIRCam sample: classical C/A, SVM C/A and 12D SVM. The table shows for each method the relations between the visual and the predicted morphological classes. The number of objects are enclosed in parentheses. (See text for details.)

	Classical C/A		SVM C/A		SVM 12D	
	Early-type	Late-type	Early-type	Late-type	Early-type	Late-type
Visual early-type	0.59 (96)	0.51 (321)	0.57 (304)	0.45 (113)	0.75 (365)	0.18 (52)
Visual late-type	0.41 (65)	0.49 (309)	0.43 (236)	0.55 (138)	0.25 (149)	0.82 (225)

component. However, unlike C, G is independent of the large-scale spatial distribution of galaxy’s light. Therefore, G differs from C in that it can distinguish between galaxies with shallow light profiles (which have both low C and G) and galaxies where much of the flux is located in a few pixels but not at the center (having low C but high G).

Each of the above parameters measures different properties of a galaxy and therefore give a different amount of information concerning the galaxy’s morphological type. For instance, Lotz et al. (2004) used the  $M_{20}/\text{Gini}$  plane to identify merger candidates, whereas the C/A plane is typically used to separate late from early type galaxies. A multi-dimensional analysis allows one to use all the information brought by each of the morphological parameters simultaneously to increase the accuracy of the classification. Moreover, previous works have shown that the measured parameters might also depend on the size, the luminosity or the redshift of the galaxy (Brinchmann et al. 1998; Bershadsky et al. 2000). Therefore, including non-morphological parameters should help the machine to take into account systematic trends in the morphological parameters due to luminosity or size variations. We thus measure 7 more parameters that we distribute in 4 classes: shape, size, luminosity and distance, according to the kind of information they measure:

- Shape: we include 2 shape parameters, the galaxy ellipticity as measured by SExtractor, i.e. the ratio of the minor and major axes of the isophotal ellipses describing the galaxy, and the CLASS\_STAR parameter, also from SExtractor. This parameter is intended to separate galaxies from stars and results from a neural network classification. Since it spans a continuum range between 0 and 1, it can be interpreted as a measure of the galaxy’s compactness.
- Size: the size parameters include the isophotal galaxy area and the petrosian radius.
- Luminosity: we use the apparent magnitude of the galaxy and the mean surface brightness.
- Distance: we adopt the photometric redshift as a measure of the distance.

#### 4.4. Training and testing

We perform several tests to probe the accuracy of the proposed method. For all the tests we adopt the same procedure: we use a fraction of the simulated catalogue (typically 500 galaxies) to train the machine, and the remaining 1000 objects to test it by looking at the fraction of galaxies that are correctly classified. We limit the analysis to only 2 broad morphological classes (late type and early type). The main reason for this choice is that there are too few irregular galaxies in the employed local sample to define a class. There is, however, no loss of generalization and the same analysis can be performed with an unlimited number

of classes, provided of course that they are correlated with measured parameters.

##### 4.4.1. Classical C/A classification versus 2D SVM

The first point we try to answer is how good would the classification of this sample be using a classical linear C/A classification. We therefore take the brightest objects of the sample (with known visual morphology) ( $K_s < 20$ ) and try to plot a linear boundary between the two distributions. As expected, the distributions are now poorly separated and plotting a linear boundary becomes extremely difficult. This is confirmed when trying to classify the whole sample (Table 2): the completeness and contaminations are basically the same as we would have obtained with a random choice. We conclude that concentration and asymmetry alone cannot be used on this sample to obtain a reliable morphological classification.

In a second step, we classify this sample with a SVM machine with the same two parameters. Results are shown in Table 2. We observe a slight gain due to the fact that SVM can adapt boundaries in a non-linear way, but the accuracy is still comparable with a random choice.

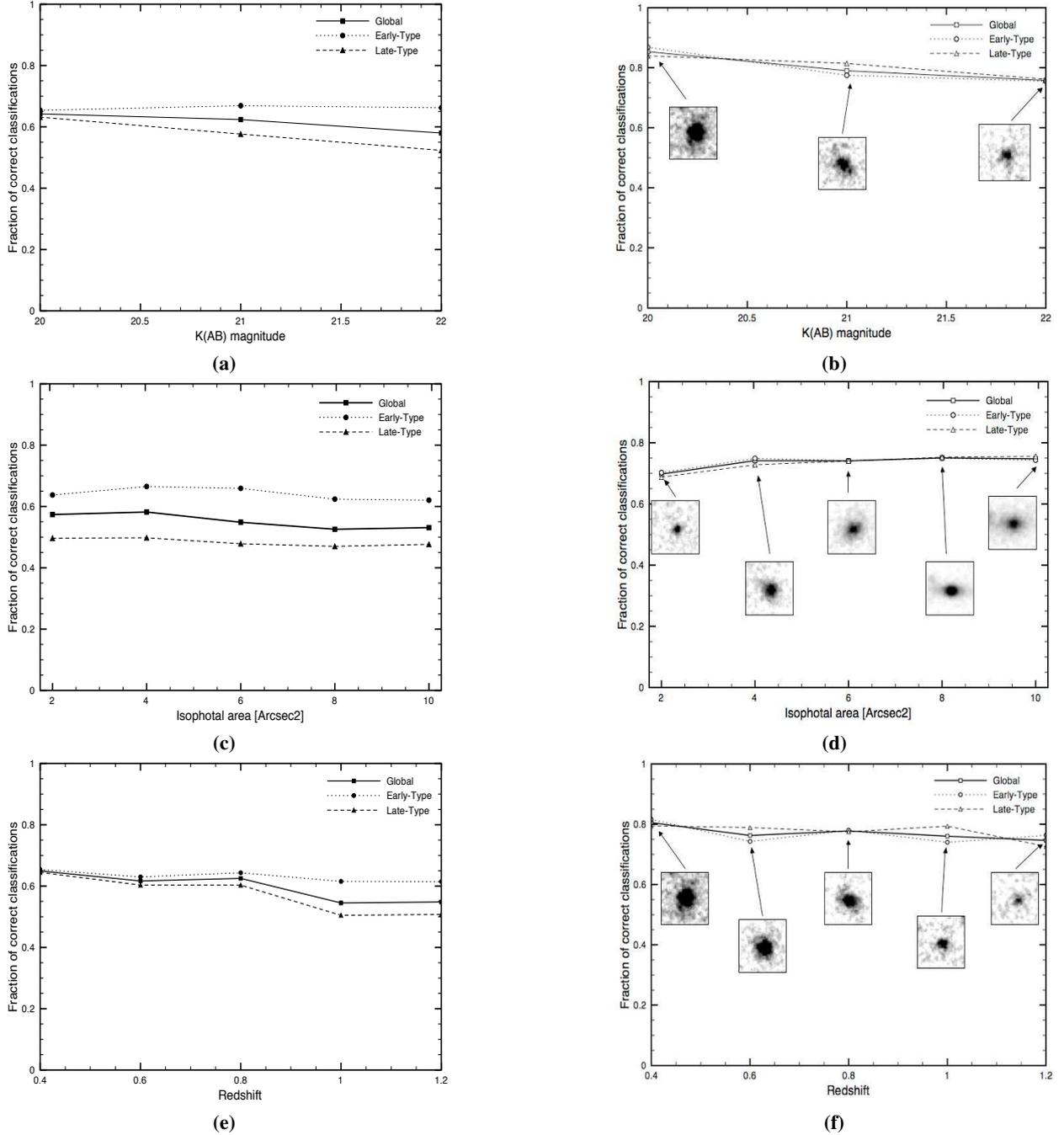
##### 4.4.2. nD versus 2D SVM

**Global effect.** We then trained 2 machines: the first one, with only 2 parameters (C and A), which should globally give the same results as a classical C/A classification as shown in Sect. 3, and the second one with 12 parameters described above. We then tested both machines by looking at the fraction of galaxies that are correctly classified. Results for the whole sample are summarized in Table 2. We observe that including more than two parameters in the classification results in a significant gain for this sample where C/A alone has accuracy comparable to a random choice. Indeed we almost recover the same accuracy that was obtained on the nearby sample (Table 1).

**Robustness.** We now try to establish the robustness of this effect. For this purpose, we look at the accuracy of the classification as a function of 3 main properties of the galaxies: luminosity, distance and area (Fig. 6). We progressively add objects and measure each time: a) the global accuracy, i.e. the fraction of galaxies that are classified correctly by the machine, and b) the accuracy per morphological type, i.e. the fraction of predicted early (late) type galaxies that are visually classified as early (late) type respectively:  $N_{E \rightarrow E}$  ( $N_{S \rightarrow S}$ ).

Several conclusions can be made from this comparison:

- First, using more than two parameters simultaneously clearly increases the global accuracy of the classification in all the redshift, area or luminosity ranges. Indeed, the mean fraction of correct classifications in the 2 dimensional machine



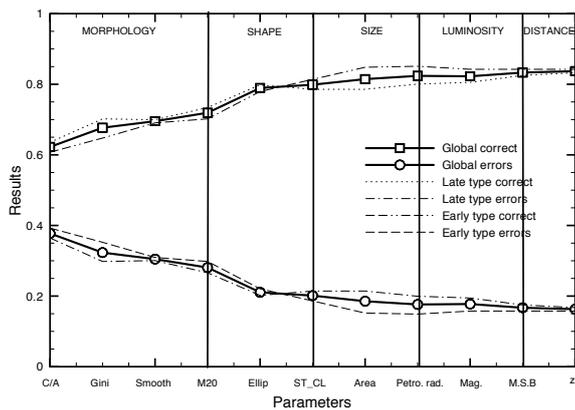
**Fig. 6.** Cumulative accuracy of classifications for a 2D machine (*left column*) and a 12D one (*right column*) as a function of magnitude (**a** and **b**), area (**c** and **d**) and redshift (**e** and **f**). A sample of  $\sim 1000$  galaxies simulated from real SDSS galaxies is used (see text for details). Solid lines show the global accuracy, i.e. the number of galaxies correctly identified, dotted and dashed lines show the fraction of early type and late type galaxies classified correctly, respectively. Stamps in the right column show a typical galaxy for every magnitude, area and redshift range.

is around  $\sim 60\%$  and decreases to  $\sim 50\%$ , which means that there is a high contamination in the C/A plane, whereas it rises to more than  $\sim 80\%$  when using a 12 dimensional machine, which is comparable to what is obtained in space observations (Brinchmann et al. 1998; Menanteau et al. 2006).

- Second, the gain is even higher when looking at the  $N_{E \rightarrow E}$  and  $N_{S \rightarrow S}$  coefficients. For the C/A classification, there is indeed an asymmetric response of the machine: early type galaxies are better identified ( $\sim 65\%$ ) whereas the fraction of late type is significantly lower ( $\sim 50\%$ ), which means that an

important fraction of late type galaxies are classified as early type. This will result, when doing the classification, in an important bias towards too a high fraction of elliptical galaxies. However, in the 12 parameter classification, the accuracies are almost perfectly symmetric for the two morphological types.

- Third, when looking at the evolution as a function of distance, size and luminosity, the 12 dimensions machine results in a more stable response, in particular as a function of magnitude and redshift.



**Fig. 7.** Accuracy of the classification as a function of the number of parameters. The first point corresponds to a classical C/A classification and each new point adds a dimension. Parameters are classified in 5 classes: morphology, shape, size, luminosity, and distance.

#### 4.4.3. How to fix the number of parameters?

In Sect. 4.4.2 it is shown that the use of more than 2 dimensions to obtain morphology clearly increases the accuracy of the global classification. However, this raises two questions. Are all these parameters necessary? Might some parameters introduce a degeneracy and consequently reduce the machine's accuracy?

To try to answer these questions we make a single test that consists in training several machines with an increasing number of parameters. We thus start with a classical 2 parameter machine (C and A) and we progressively add dimensions until we reach the 12 dimensions described in the previous section. Results are plotted in Fig. 7. As above, we plotted the global accuracy and the accuracy per morphological type. The dimensions are separated into 5 categories (morphology, shape, size, luminosity and distance).

Two important points arise at first sight. First, not all the parameters bring the same amount of useful information. The morphology and the shape carry practically the amount of information necessary to reach 80% accuracy. Second, the accuracy is a monotonic function of the number of parameters; adding a parameter can result in an almost unchanged accuracy (for instance, the magnitude) but never reduces it. This is particularly important, since it means that including more parameters than necessary does not result in a degeneracy. In addition, adding dimensions does not result in a significant increase of the computing time.

#### 4.4.4. Influence of the training set

The method we adopted here for building the training set aims to reproduce the observing conditions and physical properties of the sample in order to reduce errors due to the difference between the training and the science samples. The machine is therefore trained to solve a specific problem and should be trained differently for every science sample. We now measure the importance of this effect by simulating the same sample as if it was observed by the adaptive optics system NACO installed on the VLT. We use NACO data that have been observed in the  $K_s$  band with 2 to 3 h exposure time for each pointing (Huertas-Company et al. 2007). The total area covered by these data reaches 7 arcmin<sup>2</sup> and the mean resolution is 0.1". We therefore repeated the same procedure but dropped the simulated catalogue in a real NACO background. We then trained the machine with this sample and

**Table 3.** Accuracy of the classification when using a machine trained with a sample with different properties than the science sample – see text for details.

	WIRCam model	NACO model
Global	0.83	0.62
$N_{E \rightarrow E}$	0.81	0.96
$N_{S \rightarrow S}$	0.84	0.24

tried to classify the WIRCam simulated sample with the trained machine.

Results are shown in Table 3. The global accuracy of the classification falls to 62%, i.e. 40% of contaminations when using the NACO model to classify WIRCam galaxies. In particular, there is a systematic drift from late to early type galaxies. The training set must therefore be carefully built to take into account all the observing conditions.

## 5. Summary and conclusions

We have presented a new method to perform morphological classification of cosmological samples based on support vector machines. This method can be seen as a generalization of the classical non-parametrical C/A classification method but with an unlimited number of dimensions and non linear boundaries between the decision regions. The method is specially adapted to be used on large cosmological surveys since it is fully automated and errors are estimated objectively, allowing an easy comparison between surveys with different properties. Furthermore, since the calibration sample is built from a nearby sample, visually classified and adapted to reproduce the physical and instrumental properties, the method can even be employed on seeing-limited data. Selecting the calibration sample in the high redshift sample's rest-frame makes the results robust towards wavelength dependent effects and makes it easier to interpret them in terms of evolution.

As a test, we use our method to classify a near-infrared seeing-limited sample observed with WIRCam at CFHT with a training set of  $\sim 1500$  objects from the SDSS. We show that increasing the number of parameters in the analysis reduces errors by more than a factor 2, leading to a mean accuracy of  $\sim 80\%$  of correct classification up to the sample completeness limit ( $K_{AB} \sim 22$ ). Furthermore, the accuracy is a monotonic function of the number of parameters.

The presented method is intended as a framework for future studies. In particular, it can be used to look for luminosity and color evolution as a function of the morphology. However this method is far more general and can be applied to many other samples of galaxies observed with ground-based data, with or without AO correction. Several applications are intended in order to study the effects of local environment and galaxy density on the morphological evolution of galaxies both in the field and in rich clusters of galaxies. The library is available for download at <http://www.lesia.obspm.fr/~huertas/galsvm.html>

*Acknowledgements.* The authors want to thank the referee Dr. R. Abraham for his useful suggestions that helped to improve this paper.

## References

- Abraham, R., van den Bergh, S., Glazebrook, K., et al. 1996, ApJS, 107, 1
- Abraham, R. G., van den Bergh, S., & Nair, P. 2003, ApJ, 588, 218
- Arnouts, S., Walcher, C. J., Le Fevre, O., et al. 2007, ArXiv e-prints, 705

- Baugh, C. M., Cole, S., & Frenk, C. S. 1996, *MNRAS*, 283, 1361
- Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, *AJ*, 119, 2645
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Brinchmann, J., Abraham, R., Schade, D., et al. 1998, *ApJ*, 499, 112
- Chang, C.-C., & Lin, C.-J. 2001, LIBSVM: a library for support vector machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, *MNRAS*, 319, 168
- Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, *AJ*, 126, 1183
- Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, *ApJ*, 529, 886
- de Vaucouleurs, G. 1948, *Annales d'Astrophysique*, 11, 247
- Fall, S. M., & Efstathiou, G. 1980, *MNRAS*, 193, 189
- Hubble, E. 1936, *ApJ*, 415
- Huertas-Company, M., Rouan, D., Soucail, G., et al. 2007, *A&A*, 468, 937
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006a, *A&A*, 457, 841
- Ilbert, O., Lauger, S., Tresse, L., et al. 2006b, *A&A*, 453, 809
- Lilly, S. J., Le Fevre, O., Hammer, F., & Crampton, D. 1996, *ApJ*, 460, L1
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163
- Madau, P., Pozzetti, L., & Dickinson, M. 1998, *ApJ*, 498, 106
- Menanteau, F., Ellis, R. S., Abraham, R. G., Barger, A. J., & Cowie, L. L. 1999, *MNRAS*, 309, 208
- Menanteau, F., Ford, H. C., Motta, V., et al. 2006, *AJ*, 131, 208
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *AJ*, 124, 266
- Sandage, A. 1961, *The Hubble atlas of galaxies* (Washington: Carnegie Institution)
- Scarlata, C., Carollo, C. M., Lilly, S. J., et al. 2006, *ArXiv Astrophysics e-prints*
- Scoville, N. Z., & COSMOS Team. 2005, *BAAS*, 1309
- Simard, L., Willmer, C. N. A., Vogt, N. P., et al. 2002, *ApJS*, 142, 1
- Tasca, L., & White, S. 2005 [[arXiv:astro-ph/0507249](https://arxiv.org/abs/astro-ph/0507249)]
- Vapnik, V. 1995, *The nature of statistical learning theory*, 536 (Springer-Verlag)
- Wolf, C., Meisenheimer, K., Rix, H.-W., et al. 2003, *A&A*, 401, 73
- Wu, K. L.-K. 1999, Ph.D. Thesis, AA (University of California, Santa Cruz)
- Zucca, E., Ilbert, O., Bardelli, S., et al. 2006, *A&A*, 455, 879

# A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images

## II. Morphological evolution from $z \sim 2$ in the COSMOS field from Ks-band imaging <sup>★</sup>

M. Huertas-Company<sup>1,3</sup>, L. Tasca<sup>2</sup>, D. Rouan<sup>1</sup>, J.P. Kneib<sup>2</sup>, O. Le Fèvre<sup>2</sup>, D. Pelat<sup>4</sup>, H. J. McCracken<sup>5</sup>, M. Salvato<sup>6</sup>, P. Capak<sup>7</sup>, D. Sanders<sup>8</sup>, C. Willott<sup>9</sup>, A. Koekemoer<sup>10</sup>

<sup>1</sup> LESIA-Paris Observatory, 5 Place Jules Janssen, 92195 Meudon, France  
e-mail: marc.huertas@obspm.fr

<sup>2</sup> LAM, CNRS-Université de Provence, 38, rue Frédéric Joliot-Curie, 13388 Marseille cedex 13, France

<sup>3</sup> IAA-C/ Camino Bajo de Huétor, 50 - 18008 Granada, Spain

<sup>4</sup> LUTH-Paris Observatory, 5 Place Jules Janssen, 92195 Meudon, France

<sup>5</sup> IAP, CNRS-Université Pierre et Marie Curie, 98, Boulevard Arago, F-75014 Paris, France

<sup>6</sup> Caltech, Pasadena, CA 91125, 105-24 Caltech, Pasadena, CA 91125

<sup>7</sup> Spitzer Science Center, 314-6 Caltech, Pasadena, CA 91125, 105-24 Caltech, Pasadena, CA 91125

<sup>8</sup> Institute for Astronomy, 2680 Woodlawn Drive, University of Hawaii, Honolulu, HI 96822

<sup>9</sup> Physics Department, University of Ottawa, 150 Louis Pasteur, MacDonald Hall, Ottawa, ON K1N 6N5, Canada

<sup>10</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore MD 21218

Received [date] / Accepted [date]

### ABSTRACT

**Context.** Morphology is the most accessible tracer of galaxies physical structure, but its interpretation in the framework of galaxy evolution still remains a problem. Its quantification at high redshift requires deep high angular resolution imaging, that is why space data (HST) are normally employed. At  $z > 1$ , the HST visible cameras probe however the UV flux, dominated by the emission of young stars, which could bias the estimated morphologies towards late-type systems.

**Aims.** In this paper we quantify the effects of this *morphological k-correction* at  $1 < z < 2$  by comparing morphologies measured in the K and I-bands in the COSMOS area. Ks-band data have indeed the advantage of probing old stellar populations in the rest-frame for  $z < 2$ , enabling a determination of galaxy morphological types unaffected by recent star formation.

**Methods.** In paper I we presented a new non-parametric method to quantify morphologies of galaxies on seeing limited images based on support vector machines. Here we use this method to classify  $\sim 50\,000$  Ks selected galaxies in the COSMOS area observed with WIRCam at CFHT. We use a 10-dimensional volume, including 5 morphological parameters, and other characteristics of galaxies such as luminosity and redshift. The obtained classification is used to investigate the redshift distributions and number counts per morphological type up to  $z \sim 2$  and to compare to the results obtained with HST/ACS in the I-band on the same objects from other works.

**Results.** We associate to every galaxy with  $Ks < 21.5$  and  $z < 2$  a probability between 0 and 1 of being late-type or early-type. We use this value to assess the accuracy of our classification as a function of physical parameters of the galaxy and to correct from classification errors. The classification is found to be reliable up to  $z \sim 2$ . The mean probability is  $p \sim 0.8$ . It decreases with redshift and with size, especially for the early-type population but remains above  $p \sim 0.7$ .

**Conclusions.** The classification is globally in good agreement with the one obtained using HST/ACS for  $z < 1$ . Above  $z \sim 1$ , the I-band classification tends to find less early-type galaxies than the Ks-band one by a factor  $\sim 1.5$  which might be a consequence of morphological k-correction effects. We argue therefore that studies based on I-band HST/ACS classifications at  $z > 1$  could be underestimating the elliptical population. Using our method, we observe an increase of the early-type population from  $z \sim 2$  ( $21.9\% \pm 8\%$ ) to the present ( $32.0\% \pm 5\%$ ) probably reflecting a progressive building up of the red sequence from late-type objects.

**Key words.** galaxies: fundamental parameters – galaxies: high redshift

## 1. Introduction

In the local Universe, the distribution of galaxies is bimodal, primarily reflecting a relationship between color and morphol-

ogy. On the one hand the spiral-like galaxies are gas-rich, form stars and are supported by the rotation of their stars and on the other hand the elliptical-like galaxies are gas-poor, do not form stars anymore and are supported by the velocity dispersion of the stars. This is the so-called elliptical-spiral Hubble sequence. A fundamental question in observational cosmology is how this bimodality appears across the history of the Universe. Classical approaches to tackle this question consist in studying the evolution of the luminosity (e.g. Ilbert et al. 2006b), the star forming

<sup>★</sup> Based on observations obtained at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii.

rate or the mass assembly (e.g. Arnouts et al. 2007) for different morphological types. For that purpose, large samples of galaxies are required with a robust estimate of distances, luminosities and morphological types.

However, the difficulty in quantifying morphology of high redshift objects with a few simple, reliable measurements is still a major obstacle. The dependence on angular resolution and wavelength turns indeed the interpretation in terms of evolution difficult. In order to overcome these difficulties, people have found alternative solutions such as classifying galaxies by spectral type (Madgwick et al. 2002) or by spectro-photometric type (Zucca et al. 2006). However, the direct interpretation of these results in the framework of galaxy evolution is not straightforward since galaxies move from one spectral class to another by a passive evolution of their stellar populations. A classification based on structural parameters is nevertheless less sensitive to the star formation history and hence more robust to follow similar galaxies at different redshifts.

In the visible, progress over the last ten years have come in particular from the Hubble Deep Fields (HDF) observed with the Hubble Space Telescope. They brought observational evidence that galaxy evolution is differentiated with respect to morphological type and that a large fraction of distant galaxies have peculiar morphologies that do not fit into the elliptical-spiral Hubble sequence (Brinchmann et al. 1998; Wolf et al. 2003; Ilbert et al. 2006b). There is some evidence indeed that most of the stellar mass assembly is performed around  $z \sim 1.5 - 2$  (e.g. Arnouts et al. 2007). A better understanding of the physical processes that lead to the present Hubble sequence should come therefore from observations in this redshift range. In this context, near infrared observations are particularly important because the Ks-band flux at  $z \sim 1$  is less dependent on the recent history of star formation, which peaks in the UV in rest frame, and gives thus a galaxy type from the distribution of old stars, more closely related to the underlying total mass than optical observations. A number of Ks-band surveys have been carried out using ground-based telescopes with different spatial coverages and limiting magnitudes (e.g. Gardner et al. 1993; McCracken et al. 2000b; Maihara et al. 2001). In all cases the morphological information is poor because of the seeing-limited spatial resolution. Some morphological analysis have been performed using NICMOS on HST which is the only space instrument available in this wavelength range. All the results are on small areas involving a few tens of galaxies (e.g. Conselice & GNS Team 2007; Saracco et al. 2008) or in clusters (e.g. Zirm et al. 2008) but there are no extensive morphological classifications of field galaxies.

In paper I we proposed therefore a generalization of the non-parametric morphological classification methods that uses an unlimited number of dimensions and non-linear separators, enabling us to use all the information brought by the different morphological parameters simultaneously. We showed that when applied to seeing limited data it reduces errors by more than a factor 2 compared to classical non-parametric methods, leading to a mean accuracy of  $\sim 80\%$  of correct classifications.

In this paper, we use this method to quantify the morphologies of  $\sim 50\,000$  galaxies based on structural parameters measured in the near infrared. Galaxies are observed in the COSMOS field with WIRCam at CFHT in Ks-band. We perform basic statistics such as redshift and magnitude counts per morphological type and compare them to the ones obtained on the same objects using HST/ACS imaging (I-band) in order to quantify morphological k-correction effects. This classification is intended as a framework for future studies of the evolution of

counts, luminosities, luminosity densities, correlation function for each morphological type over several redshift bins.

The paper proceeds as follows: the data set and the sample selection are presented in the next section. In Sect. 3 we describe the technique used to quantify morphologies. In Sect. 4, we analyze the results of the classification and show the first statistics. We finally compare the results with the ones obtained with HST/ACS in Sect. 5 and discuss the effects of morphological k-correction.

We use the following cosmological parameters throughout the paper:  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $(\Omega_M, \Omega_\Lambda) = (0.3, 0.7)$  and the AB system for magnitudes.

## 2. The data

### 2.1. Description

The Ks-band data were taken with WIRCam (Thibault et al. 2003) installed at CFHT in the near infrared  $K_s$  band ( $2.2\mu\text{m}$ ). The field is part of the Canada-France Hawaii Telescope Legacy Survey (CFHTLS) Deep survey and its near infrared follow-up, and covers the entire COSMOS area ( $1.4^{(o)^2}$ ) (Scoville & COSMOS Team 2005). Images are reduced with the Terapix pipeline<sup>1</sup> and have a pixel scale of  $0.15''$  with a mean FWHM of  $0.7''$ . Figure 1 shows a cutout of the final reduced image. A more detailed description of the data set can be found in McCracken et al. 2008 (in preparation).

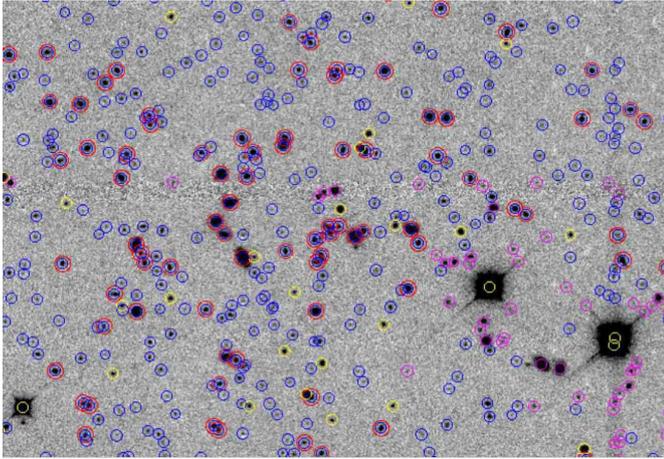
The I-band data used in § 5 are part of the COSMOS HST/ACS field (Koekemoer et al. 2007). The data set consists of a contiguous  $1.64 \text{ deg}^2$  field covering the entire COSMOS field. The Advanced Camera for Surveys (ACS) together with the F814W filter (“Broad I”) were employed. More precisely, in this paper we will use the morphological catalogue described in Tasca et al. 2008 (in preparation).

### 2.2. Building the Ks-band catalogue

#### 2.2.1. Detection and cleaning

All objects having a  $1.5\sigma$  signal above sky, over four contiguous pixels are detected using SExtractor (Bertin & Arnouts 1996). We then performed a cleaning task in order to separate galaxies from stars and spurious detections. This was made using the SExtractor MU\_MAX and MAG\_AUTO parameters that give the peak surface brightness above the background and the Kron-like elliptical aperture magnitude, respectively. The distribution of objects in this parameter space clearly defines three regions that separate extended sources from point-like or non-resolved sources and from spurious detections. In this separation scheme, objects with very faint magnitude and high peak surface brightness are considered as false detections. The final reduced tiles have several strips where the noise is considerably higher than in the other regions of the image. Therefore, objects that fall in these regions were masked. Furthermore a mask was also applied to the objects that are too close to bright stars ( $K_s < 15$ ). We finally obtain, after cleaning and masking, 282 122 non-spurious sources over the whole field

<sup>1</sup> <http://terapix.iap.fr>



**Fig. 1.**  $3' \times 2'$  cutout of the observed area. Yellow circles are stars, magenta circles are masked objects, blue circles are galaxies and red circles are galaxies with computed morphology. (see text for details)

### 2.2.2. Star/galaxy separation

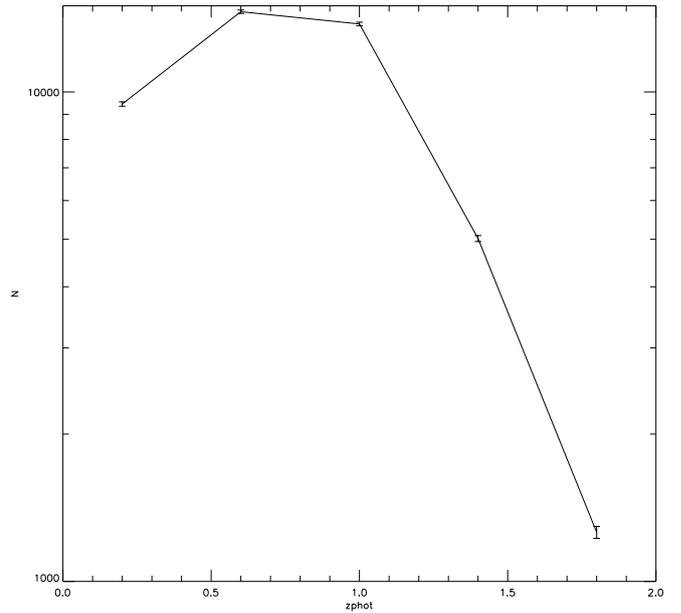
In order to select galaxies from the total K-selected photometric sample, we have used a number of photometric parameters to remove candidate stars, as described below. Some of the parameters that can be used are: (i) the CLASS\_STAR parameter given by SExtractor providing the “stellarity-index” for each object, (ii) the MU\_CLASS, i.e. the position in the MU\_MAX-MAG-AUTO plane described above, reliable up to  $K_s \sim 20$  (iii) the  $\chi^2$  of the SED fitting carried out during the photometric redshift estimate (Ilbert et al. 2006a), with templates SEDs of both stars and galaxies.

We decided to use a combination of these criteria. For objects brighter than  $K_s = 20$ , we selected as stars objects having a  $CLASS\_STAR > 0.95$  and a photometric redshift lower than 0.05 or a stellar spectral class from the SED fitting in addition to the MU\_CLASS parameter. For objects with  $K_s > 20$  the MU\_CLASS parameter was not used and the spectral class was only used when available. The final sample consists in 27 343 point-like sources and 254 779 galaxies.

### 2.2.3. The photometric redshift catalogue

We used the code *Le Phare* (Ilbert et al. 2006a) to measure the photometric redshift of 278 526 objects in the COSMOS field with  $I_{AB} < 25$  and multi-wavelength information. To compute the photometric redshifts a standard  $\chi^2$  method (Arnouts et al. 1999) is implemented according to Ilbert et al. (2006a). The multi-color catalog consists of photometry measurements over a 3 arcsecond diameter apertures for deep  $B_j, V_j, g+, r+, i+, z+$  Subaru data,  $u*, i*$  Canada-France-Hawaii Telescope bands (CFHT),  $u', g', r', i', z'$  information from the Sloan Digital Sky Survey (SDSS),  $K_s$  magnitude from KPNO/CTIO, and  $F816W$  HST/ACS magnitude. We use 1095 spectroscopic redshifts taken from the zCOSMOS Survey (Lilly & The Zcosmos Team. 2005) to calibrate the magnitude zero-points. The method applied allows to reach an accuracy  $\sigma_{\Delta z}/(1+z_s) = 0.031$  with  $\eta = 1.0\%$  of catastrophic errors, defined as  $\Delta z/(1+z_s) > 0.15$ .

Our final catalogue has 198684 objects with reliable photometric redshift measurements.



**Fig. 2.** Redshift distribution for the 44 089 analyzed galaxies. Error bars are calculated using Poisson  $\sqrt{n}$  statistics.

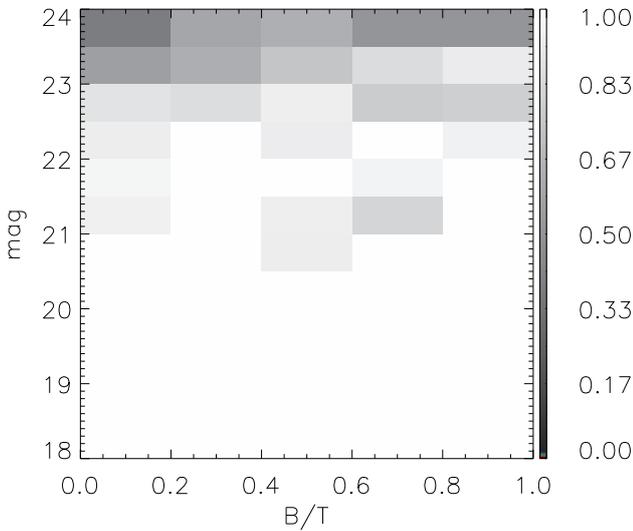
### 2.2.4. The morphological catalogue

The morphological analysis is made in a subsample of the initial catalogue: first we select only the galaxies which have a measured photometric redshift. Then we cut the catalogue to  $K_s < 21.5$  and  $z_{phot} < 2$ . This decision is based in a visual inspection; objects fainter than 21.5 have a S/N per pixel lower than 5, so we decided not to include them in the morphological study. Simulations of those objects show indeed that the morphological classifications obtained are highly contaminated. Photometric redshift above  $z \sim 2$  are not reliable enough. This selection results in a final morphological catalogue of 44 089 galaxies. Fig 2 shows the redshift distribution of the final catalogue.

In order to verify precisely whether the sample is complete for all the morphological types and to quantify the selection effects, we generated 5000 fake galaxies with exponential (Freeman 1970) and de Vaucouleurs profiles (de Vaucouleurs 1948) of different morphological types (bulge fraction uniformly distributed between 0 and 1) and with galaxy sizes uniformly distributed between  $0''$  and  $1.5''$  and dropped them in the real background images. We then tried to detect them with SExtractor. Results are shown in Figure 3. As we can see, the sample is complete for all the bulge fraction values up to our magnitude limit ( $K_s = 21.5$ ). We also looked at the completeness as a function of size. We generated for that purpose pure bulge and pure disk profiles with different sizes and detect them with SExtractor (Fig. 4). As expected bulges are detected up to fainter magnitudes, however the sample is complete up to  $K_s = 21.5$  for both bulges and disks for sizes ranging from  $0''$  to  $1.5''$ .

## 3. Morphology

Galaxies in the catalogue have been separated into two main morphological types (late-type and early-type) using the



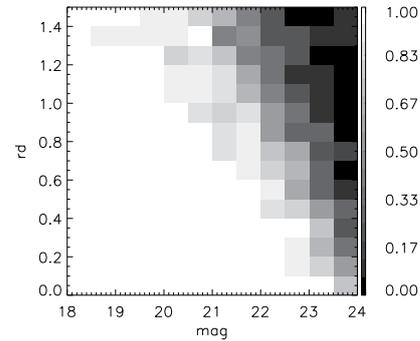
**Fig. 3.** Completeness for extended sources as a function of the bulge fraction as assessed from a mock sample of 5000 galaxies dropped in the field images.

free available code `galSVM`<sup>2</sup> (Huertas-Company et al. 2007). `galSVM` is a non-parametric N-dimensional code based on support vector machines (SVM) that uses a training set built from a local visually classified sample. The employed procedure can be summarized in 4 main steps (see Paper I for more details):

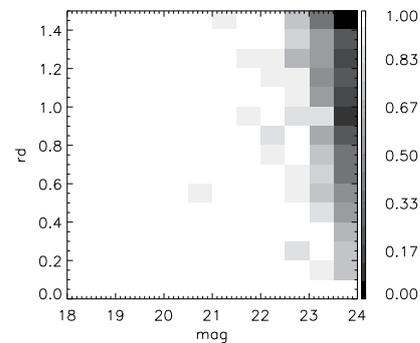
1. Build a training set: we select a nearby visually classified sample at wavelengths corresponding to the rest-frame of the high redshift sample to be analyzed. In our case, we want to simulate Ks-band observations. We used therefore an SDSS local sample observed in the i and z bands which roughly corresponds to the rest-frame wavelengths between  $z \sim 1$  and  $z \sim 2$  for Ks-band observations. We then move the sample to the proper redshift and image quality and drop it in the real background.
2. Measure a set of morphological parameters on the sample.
3. Train a support vector based learning machine with a fraction of the simulated sample and use the other fraction to test and estimate errors.
4. Classify real data with the trained machine and correct for possible systematic errors detected in the testing step.

### 3.1. The training sample

The most important step in obtaining the morphology with a non-parametric method is to correctly calibrate the volume filled by the data in the multi-dimensional space. This is a critical step since it will determine the decision regions that will be used to perform the classification. Indeed, galaxy morphology derivation depends on the physical properties of the galaxy (luminosity, redshift, wavelength) and on the observing conditions (background level, resolution). A suitable calibration set should consequently reproduce closely all the properties of the sample to be analyzed. One classical approach consists in visually classifying



(a)



(b)

**Fig. 4.** Completeness as a function of size for simulated disks (a) and bulges (b) as assessed from a mock sample of 5000 galaxies dropped in the field images. The size is represented by the disk scale length for disks ( $r_d$ ) and by the bulge effective radius ( $r_e$ ) for bulges.

a fraction of the sample and use it as a training set to optimize boundaries (Menanteau et al. 1999, 2006). However this is not possible for seeing limited data where the resolution is too low to enable a reliable visual classification. Here, we then decide to simulate the high redshift sample from a visually classified local catalog, selected in the rest-frame color of the high redshift sample. This has three main advantages: first, it is less affected by K-correction effects, second it does not introduce any modeling effect, since the used galaxies are real and finally, the training set is built to reproduce the observing conditions and physical properties of the sample to be analyzed, but it is classified safely on well-resolved images, so it does not need to have a specially high resolution.

We use a catalog of 1319 objects from the Sloan Digital Sky Survey observed in two photometric bands (z and i) and visually classified (Tasca & White 2005). As explained in paper I, for every galaxy stamp we first generate a random pair of (magnitude, redshift) values with a probability distribution that matches the real magnitude and redshift distribution of the sample to be simulated and then we proceed in four steps: a) removal of foreground stars, b) degradation of the resolution according to a  $\Lambda$ CDM cosmology, c) binning to reach the desired pixel scale and d) dropping in a real background image.

The photometric band (z or i) used to create the galaxy depends on the associated redshift. We choose indeed the one that is closer to the rest-frame band at this given z.

<sup>2</sup> <http://www.lesia.obspm.fr/~huertas/galsvm.html>

We did not take into account any variation of the PSF within the field for performing the simulations. We expect indeed this variation to be small and consequently not to induce strong changes in the morphology since the analyzed galaxies are significantly larger than the PSF.

Among all the simulated objects, 450 were used as training and the remaining 869 as test.

### 3.2. Classification procedure

We use 450 simulated galaxies as training sample. The morphological mixing is fixed to 50/50, i.e. 50% of early-type galaxies and 50% of late-type. Even if this is not a realistic distribution it is required to minimize the errors in the SVM classification. The classification is made in a 10-D volume with a Radial Basis Function Kernel (see Paper I for more details). The measured parameters include 6 morphological parameters: Asymmetry, Concentration (Conselice et al. 2000 and Abraham et al. 1996 definitions), Gini (Abraham et al. 2003), M20 (Lotz et al. 2004), Smoothness (Conselice et al. 2003), a distance parameter (photometric redshift), a shape parameter (elongation), 2 luminosity parameters (surface brightness and magnitude). See Paper I for details on how these parameters are calculated.

In Paper I we show, thanks to the test sample, that increasing the number of parameters of the classification never results in a degeneracy that decreases the accuracy. We realized however that when dealing with the real sample, some parameters might produce a bias if they are not realistic. This is the case of the magnitude. Indeed when doing the simulations we use the same magnitude distribution for early and late-type galaxies. This creates an artificial excess of early-type galaxies at faint magnitudes in the simulated sample that is not seen in the real world. Therefore if the magnitude is used as input parameter for the classification the number of early-type galaxies are overestimated. To avoid this bias we proceed in 2 steps: first we make a 9-D classification (without the magnitude) and then we use the measured magnitude distribution per morphological type to generate a new simulated sample with this magnitude distribution. We use this second sample to make the final 10-D classification.

### 3.3. The output catalogue

Classical support vector classifiers only predict class label (i.e. early-type or late-type) but not probability information. Recently some authors (Platt 2000; Wu et al. 2004) have proposed different methods to estimate a posteriori probability, i.e. given  $k$  classes of data, for any  $\mathbf{x}$  the goal is to estimate  $p_i = p(y = i|\mathbf{x})$ ,  $i = 1, \dots, k$ . The free available package libSVM (Chang & Lin 2001) implements the method described in Wu et al. (2004). Therefore we added this new feature to our classification output: We associate to every galaxy in the morphological catalogue, a class label and a probability of belonging to the given class. Since we are dealing with a 2-class problem, the probability  $p(\text{galaxy}=\text{early-type})=1-p(\text{galaxy}=\text{late-type})$ . In the following, we use this parameter to assess the accuracy of our classification.

### 3.4. Accuracy

As shown in paper I, one of the main advantages of the employed method is that the reliability of the classifications can be quantified using a test sample simulated in the same way as the training one.

Here we use the probability as the main estimator. We first looked at the evolution of the correct classifications as a function of different probability thresholds. Results are shown in table 1. As we can see, there is a clear correlation between the probability threshold and the number of correct identifications: the accuracy clearly increases when the considered probability is higher. If we select only objects with a probability between 0.5 and 0.6 the mean accuracy is only around 58%. However, objects with probabilities greater than 0.8 are classified with nearly 90% accuracy. The contamination is around  $\sim 20\%$  for the whole sample ( $p > 0.5$ ).

This relations between the probability parameter and the success rate enable the use of the probability as an estimate of the classification accuracy as a function of physical parameters of the galaxies directly on the real sample. We decided to represent 2D maps of the mean probability values (Fig. 5) for different magnitude, redshift and size bins. We observe several interesting trends:

- Globally, late-type galaxies have higher probabilities for all redshift, magnitude and size values. This indicates that late-type objects are easily isolated. It is probably a consequence of the ellipticity parameter, used in the classification procedure. Indeed, objects with high ellipticity are identified as late-type galaxies with high probability so that the mean probability increases.
- As expected there is a clear trend with size for both morphological types, small objects ( $r_{half} < 0.6''$ ) have lower probabilities ( $p \sim 0.7$ ) while large ones ( $r_{half} > 0.6''$ ) have higher values ( $p > 0.8$ ) (Fig. 5 c,d).
- There is also a trend with redshift, especially for early-type objects: below  $z \sim 1$  the mean probability is around  $p \sim 0.75$  and it decreases to  $p \sim 0.7$  for  $z > 1$ . The number of galaxies with low probability ( $p \sim 0.6$ ) also increases (Fig. 5 a,b).
- Finally, the magnitude is also important for determining the quality of the classification. Above  $K_s \sim 20$  the mean probability for early-type objects is around  $p \sim 0.7$  (Fig. 5 a,e). Interestingly, this trend does not appear for the late-type population (Fig. 5 b,f). This is probably a consequence of the way the training sample is built (§ 3.2): at faint magnitudes, the number of early-type galaxies is low so the machine tends to find late-type galaxies with higher probability.

How can we take into account these trends for correcting our classification and performing statistical analysis? The simplest way seems to establish *probability thresholds* but, what galaxies do we keep when selecting objects with a given probability or, in other words, do we introduce any bias when selecting a sample with a given probability cut?

To answer this question we selected objects with 4 probability thresholds ( $p > 0.6$ ,  $p > 0.7$ ,  $p > 0.8$  and  $p > 0.9$ ) and examined the completeness of the selected sample as a function of the magnitude, the redshift the size and the morphological type. Figure 6 shows the results. If no biases were introduced all the lines should be *flat* (i.e. “we keep all the galaxies in the same way”). We observe in Fig. 6 that this is not the case. This is seen in particular when looking at the completeness as a function of size and morphological type: large objects are preferred as expected and early-type objects are more penalized.

Therefore a sample selected by performing a probability cut should not be used for global statistical analysis. It can be used however to select particular class of objects for which a very good accuracy is needed. In the next section we show however a way of using the information brought by probabilities for statistical purposes.

**Table 1.** Accuracy of the classifications based on the test sample for a 10 parameters SVM training (C, A, G, Ell, S, M20, C2, SB, z, mag) as a function of probability. The table shows the fraction of correct classifications when a probability bin  $[P_{min}, P_{max}]$  is considered. Numbers between brackets give the number of objects per bin. Top: early-type galaxies, bottom: late-type galaxies.

$P_{max} \rightarrow$ $P_{min} \downarrow$	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1.0
0.5	–	–	0.56 (45)	0.55 (63)	0.53 (82)	0.6 (111)	0.66 (157)	0.7 (195)	0.73 (236)	0.76 (315)	0.78 (352)
0.55	–	–	–	0.58 (46)	0.56 (65)	0.62 (94)	0.68 (140)	0.72 (178)	0.76 (219)	0.80(298)	0.82 (335)
0.6	–	–	–	–	0.52 (37)	0.62 (66)	0.70 (112)	0.74 (150)	0.78 (191)	0.83(270)	0.84 (307)
0.65	–	–	–	–	–	0.63 (48)	0.75 (94)	0.78 (132)	0.82 (173)	0.87(252)	0.87 (289)
0.7	–	–	–	–	–	–	0.85 (75)	0.86 (113)	0.9 (154)	0.92(233)	0.92 (270)
0.75	–	–	–	–	–	–	–	0.89 (84)	0.92 (125)	0.94(204)	0.94 (241)
0.8	–	–	–	–	–	–	–	–	0.93 (79)	0.96(158)	0.94 (195)
0.85	–	–	–	–	–	–	–	–	–	0.98(120)	0.96 (157)
0.9	–	–	–	–	–	–	–	–	–	–	0.95 (116)

$P_{max} \rightarrow$ $P_{min} \downarrow$	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1.0
0.5	–	–	0.58 (39)	0.62 (68)	0.62 (82)	0.64 (114)	0.67 (155)	0.7 (189)	0.72 (224)	0.74 (271)	0.75 (309)
0.55	–	–	–	0.68 (53)	0.66 (67)	0.68 (99)	0.70 (140)	0.73 (174)	0.74 (209)	0.76(256)	0.77 (294)
0.6	–	–	–	–	0.68 (46)	0.70 (87)	0.70 (116)	0.74 (150)	0.75 (185)	0.78 (232)	0.78 (270)
0.65	–	–	–	–	–	0.74 (73)	0.75 (94)	0.75 (121)	0.76 (156)	0.78 (203)	0.79 (241)
0.7	–	–	–	–	–	–	0.85 (75)	0.78 (107)	0.78 (142)	0.81(189)	0.82 (227)
0.75	–	–	–	–	–	–	–	0.80 (75)	0.80 (110)	0.83 (157)	0.83 (195)
0.8	–	–	–	–	–	–	–	–	0.84 (69)	0.86(116)	0.86 (154)
0.85	–	–	–	–	–	–	–	–	–	0.85 (82)	0.85 (120)
0.9	–	–	–	–	–	–	–	–	–	–	0.87 (85)

### 3.4.1. Best estimator

Considering that we have  $N$  galaxies with a probability  $p_k$  of being in class A (e.g. early-type or late-type), we can define two functions to estimate the number of objects in a given class.

The first one (hereafter *counts estimator*) uses a function:

$$Z_k = \begin{cases} 0 & \text{if } p_k < 0.5 \\ 1 & \text{if } p_k > 0.5 \end{cases}$$

Then the number of objects in class A is simply:  $N_A = \sum Z_k$ . Objects with a probability greater than 0.5 of belonging to a given class are simply added. With this estimator we consider therefore in the same way a galaxy with  $p = 0.51$  and one with  $p = 0.99$ ; it is just added to the corresponding class and the probability is ignored.

The second one (hereafter *probability estimator*) tries to make use of the information contained in the probability parameter. It has been shown indeed in the previous section that galaxies with higher probabilities have lower classification errors. However making arbitrary probability cuts introduces biases in the global statistics in the sense that not all the objects are removed uniformly. It would be interesting therefore to find a way of using the information contained in this parameter without introducing biases. For that purpose, we define a random variable as follows:

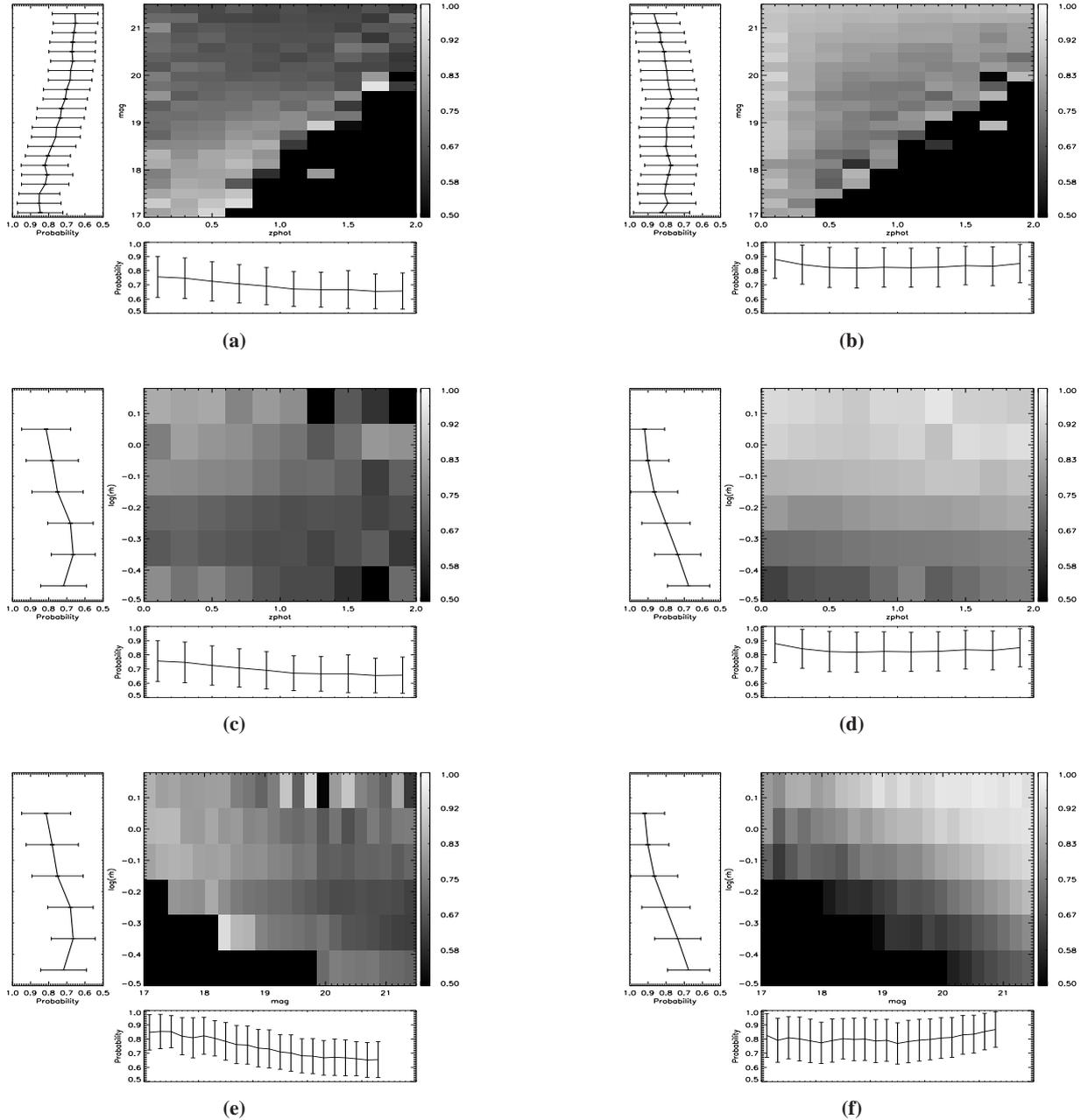
$$Y_k = \begin{cases} 0 & \text{with a probability } 1 - p_k \\ 1 & \text{with a probability } p_k \end{cases}$$

Then we can estimate the number of objects in class A as the mathematical expectation of this variable:  $E(Y) = \sum E(Y_k) = \sum p_k$  and the  $1-\sigma$  error on the number as the square root of the variance of the variable:  $Var(Y) = E(X^2) - E(X)^2 = \sum p_k(1 - p_k)$ . This way, a galaxy with a probability  $p = 0.51$  of being in class A counts as 0.51 in this class but also as 0.49 in class B.

Therefore, the larger the number of galaxies with probability values close to 0.5, the more the differences between the two estimators will be important. Indeed, if all the probabilities are equal to one, both estimators give the same results whereas if all the probabilities are equal to 0.5, the probability estimator gives half the value of the count estimator. In this sense the comparison of the results furnished by the two estimators is a kind of measure of the classification accuracy.

### 3.4.2. Errors due to the training set

The estimator presented above takes into account the information brought by the probability parameter to estimate at best the number of galaxies of each morphological type and correct from misclassifications. However, there is another source of error which has to be considered. It is related to the training set itself. Indeed, even if the training sample is built to reproduce at best the parameters of the real one it can contain errors or there can exist galaxies which are not well represented in the parameter space. In order to estimate these errors, we performed Monte Carlo simulations: we randomly removed elements from the training sample and generated therefore multiple training samples with fewer galaxies. We then used these samples to train different machines and classify the real sample. The generated samples have similar properties than the original one but do not fill the parameter space in the same way. They can be used consequently to estimate the effect in the final classification of *missing objects* in the parameter space. The differences found in the classifications are then employed to estimate a kind of *confidence region* of our classification scheme in the following sections.



**Fig. 5.** Bi-dimensional maps of the mean probabilities for different redshift, magnitude and size bins of the real sample. The size is represented by the half light radius as measured by SExtractor. The plots show the mean probability distribution as a function of a single parameter. Error bars show the dispersion of the probability distribution. On the left column we show the probability maps for early-type objects and on the right the maps for late-type objects. The black cells indicate that there are no objects in the considered bin.

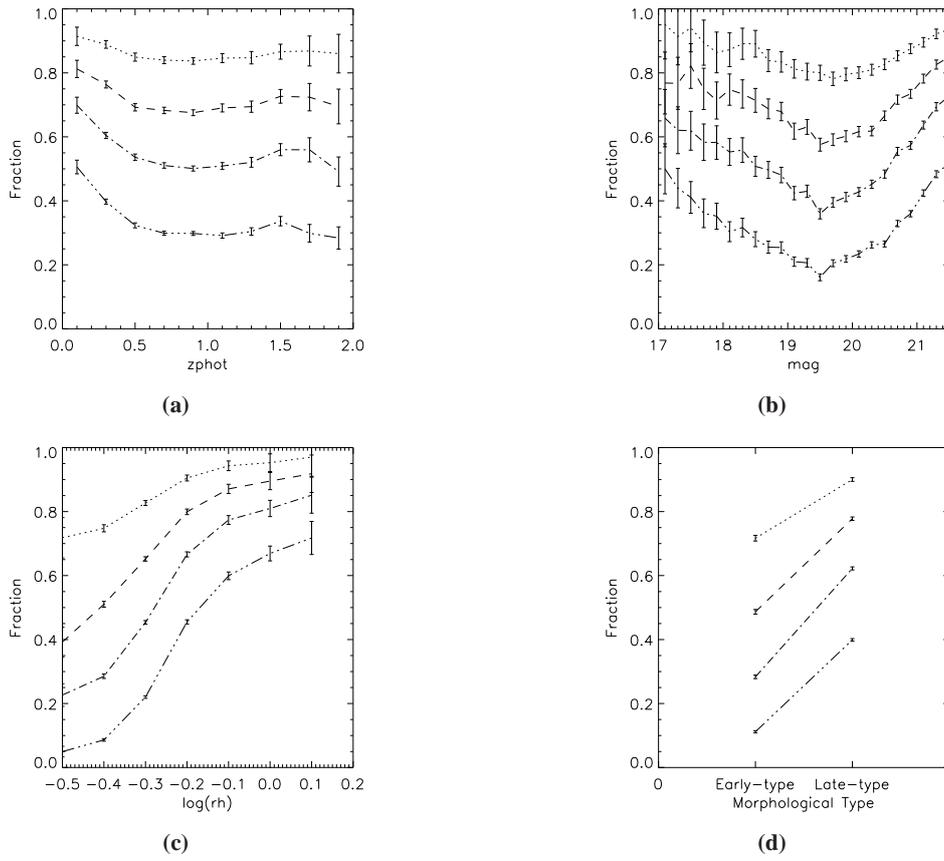
## 4. Results and discussion

### 4.1. Global statistics

We find 33291 (~75%) late-type galaxies and 10798 (~25%) early-type galaxies with the counts estimator and  $30711.7 \pm 1952$  ( $70\% \pm 5\%$ ) and  $13376.3 \pm 2014$  ( $30\% \pm 5\%$ ) early-type galaxies with the probability estimator. In the next sections we try to locate more precisely the differences.

### 4.2. Number counts

Figure 7 shows the number counts per morphological type using the two estimators described above. As expected, early-type objects dominate the bright end of the magnitude distribution while late-type objects are more frequent at the faint-end. The effect of using the probability estimator is seen clearly in the faint elliptical population ( $K_s > 20$ ). In particular the number of faint early-type galaxies increases when using the probability estimator. This reflects that at faint magnitudes early-type galaxies are classified with lower probabilities as seen from the 2D maps



**Fig. 6.** Sample completeness as a function of redshift (a), magnitude (b), size (c) and morphological type (d) for different probability cuts. Dotted line: galaxies with  $p > 0.6$ , dashed line: galaxies with  $p > 0.7$ , dashed-dotted line: galaxies with  $p > 0.8$  and dashed line with three dots: galaxies with  $p > 0.9$ . Error bars are calculated using Poisson  $\sqrt{n}$  statistics.

(Fig: 5). The use of the probability parameter in performing the counts enables to correct for the classification errors by adding galaxies that fell in the late-type class but had a significant probability of being elliptical.

#### 4.3. Morphological evolution

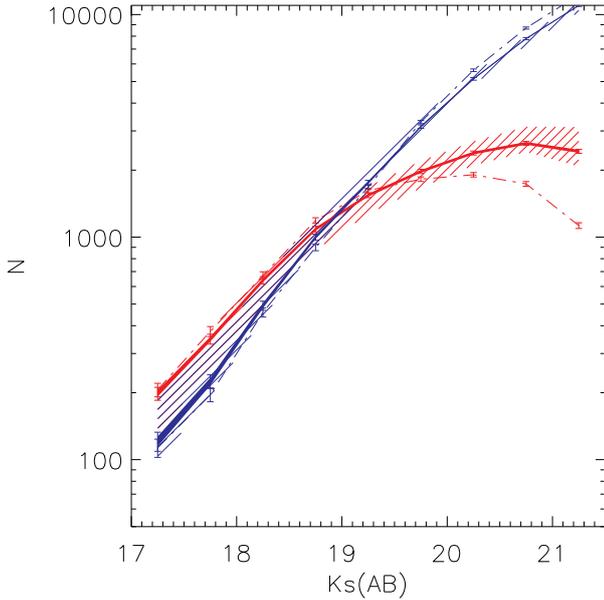
Figure 8 shows the morphological mixing evolution up to  $z \sim 2$ . Both estimators reveal an increase of the early type fraction from  $z \sim 2$ , however the effect is less important when taking into account the probability. Indeed, with the probability estimator we find  $21.9\% \pm 8\%$  early-type objects at  $z \sim 2$  while the local fraction is  $32.0\% \pm 5\%$ . The *counts estimator* predicts a fraction of  $\sim 15\%$  at  $z \sim 1.5$ . Considering the probability helps therefore to correct from the incompleteness of the early-type population at high  $z$  due to the lower probability values.

This variation in the early-type population is a well-known effect which has been detected using rest-frame morphologies from HST (e.g. Brinchmann et al. 1998; Cassata et al. 2005) and probably reflects the building-up of the red sequence from late-type systems. It could be argued however that this effect was a consequence of morphological k-correction since at higher redshifts we probe the UV galaxy emission, i.e. young stellar populations. The fact that we still observe this trend with NIR data (which probe older stellar populations) seems to point that this is a real effect and not a morphological k-correction effect. However, the fraction found here at  $z \sim 1.5$  is significantly

higher than the one obtained in Cassata et al. (2005). Can these differences be a consequence of band shifting? This is a crucial point of this study and the main reason for performing morphological analysis in the NIR. The following sections are therefore focused on this important point.

#### 5. Investigating the morphological k-correction effect

In the previous section, we have obtained a morphological classification from NIR imaging. A crucial point is to understand the differences (if there are some) between morphologies quantified in the visible and the ones computed here. We perform for that purpose a match between the Ks selected morphological catalogue and the morphologies measured from HST/ACS data (I-band) in an independent way. The algorithm used to morphologically classify galaxies in the I-band is described in Tasca et al. 2008 (in preparation) and we refer the reader to that paper for details on the method. Here it is enough to stress that the high resolution of ACS images combined with a depth of  $I \sim 27$  allow for accurate morphological measurements well further the limits imposed by this study. We match 34 165 objects of our K-selected sample. The remaining objects were too faint in the ACS data to perform reliable morphological classifications. The global morphological mixing from HST/ACS data is in good agreement with the one obtained with WIRCam for the same objects: 9163 (I-band), 10458.8 (Ks-band) early-type galaxies and



**Fig. 7.** Number counts per morphological type for the 44 089 analyzed galaxies. The solid line shows the results obtained with the probability estimator; the line width indicates the  $1 - \sigma$  confidence band estimated from the probability distribution and the dashed region is the confidence region deduced from Monte Carlo simulations of the training sample (see text for details). The dashed line shows the results obtained with the counts estimator. Red lines stand for early-type galaxies and blue lines for late-type. Error bars are calculated using Poisson  $\sqrt{n}$  statistics.

25002 (I-band), 23706.2 (Ks-band) late-type galaxies respectively. However, where are the differences localized precisely?

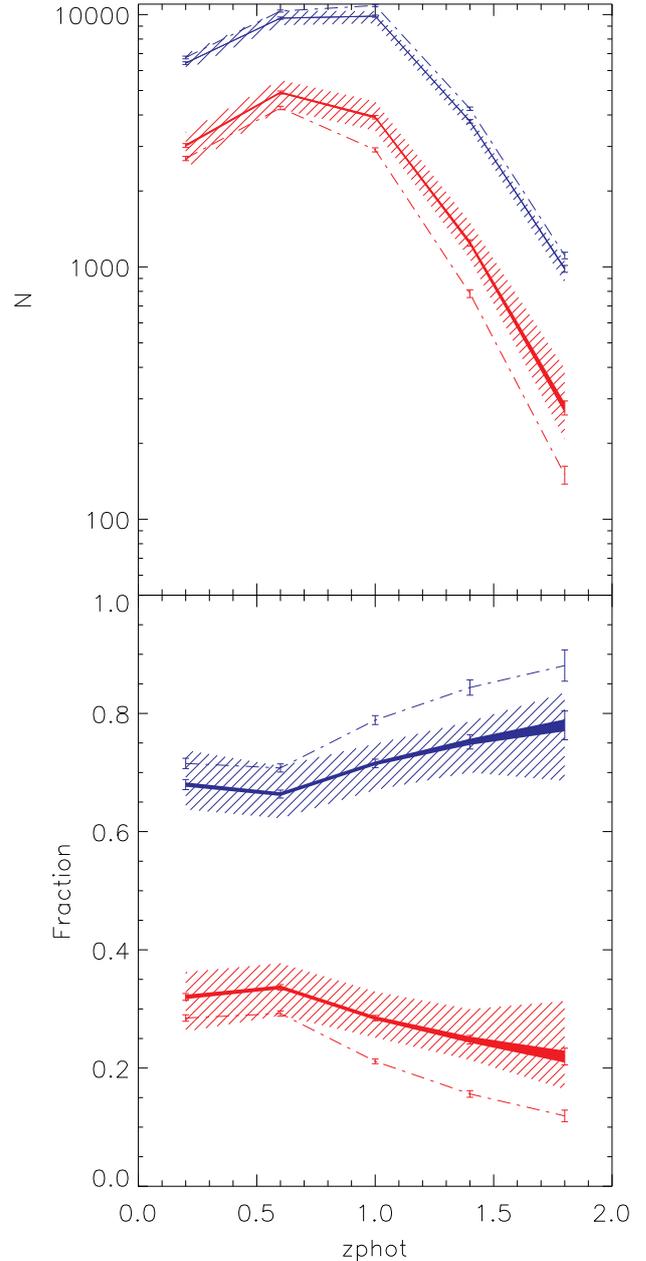
### 5.1. One-to-one comparison

Figure 9 shows the mean probability for a galaxy classified as early (late) in the I-band to be classified as early (late) in the Ks-band as a function of redshift.

We observe, that globally, there is no ambiguity in the identification of late-type objects. A galaxy which is found to be late-type in the I-band is also a late-type system in the Ks-band with a probability around  $p \sim 0.9$  up to  $z \sim 2$ .

For early-type objects, there is clear trend with redshift: below  $z \sim 0.8$  a galaxy classified as early-type in the I-band has a probability of  $p \sim 0.7$  of being early-type in the Ks-band. Above  $z \sim 1$  the discrepancies between the two classifications become higher and an elliptical galaxy is classified as early in both classifications only with a probability of  $p \sim 0.4$ . Above  $z \sim 1$  the I-band filter starts to probe the UV flux and therefore the obtained morphology is determined from young stars whereas the Ks-band filter stills probe the visible spectra of the galaxy. We are thus probably seeing a k-correction effect: late-type objects are identified in the same way since their stellar populations are younger whereas for early-type objects the ambiguity is higher and a fraction of objects tend to move to later morphological types.

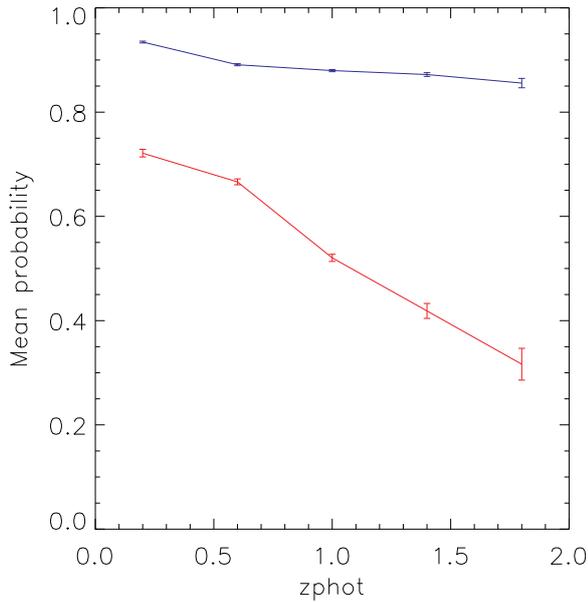
What are then the differences of morphologically selecting a sample in the Ks-band or in the I-band above  $z \sim 1$ ?



**Fig. 8.** Redshift distribution per morphological type for the 44 089 analyzed galaxies. The solid line shows the results obtained with the probability estimator; the line width indicates  $1 - \sigma$  confidence band estimated from the probability distribution and the dashed region is the confidence region deduced from Monte Carlo simulations of the training sample (see text for details). The dashed line shows the results obtained with the counts estimator. Red lines stand for early-type galaxies and blue lines for late-type. Error bars are calculated using Poisson  $\sqrt{n}$  statistics.

### 5.2. Comparing redshift distributions

Figure 10 shows the redshift distributions obtained from both classifications. The distributions for late-type objects present small differences as expected from the comparison performed in the previous section. A Kolmogorov-Smirnov test (KS-test) reveals that the two functions arise from the same statistic with 96% confidence. The early-type redshift distributions present



**Fig. 9.** One-to-one comparison of the ACS and WIRCam classifications as a function of redshift. Red line: mean probability for a galaxy classified as early-type in the I-band (HST) to be classified as early-type in the Ks-band. Blue line: mean probability for a galaxy classified as late-type in the I-band (HST) to be classified as late-type in the Ks-band. Error bars show the dispersion of the probability distribution.

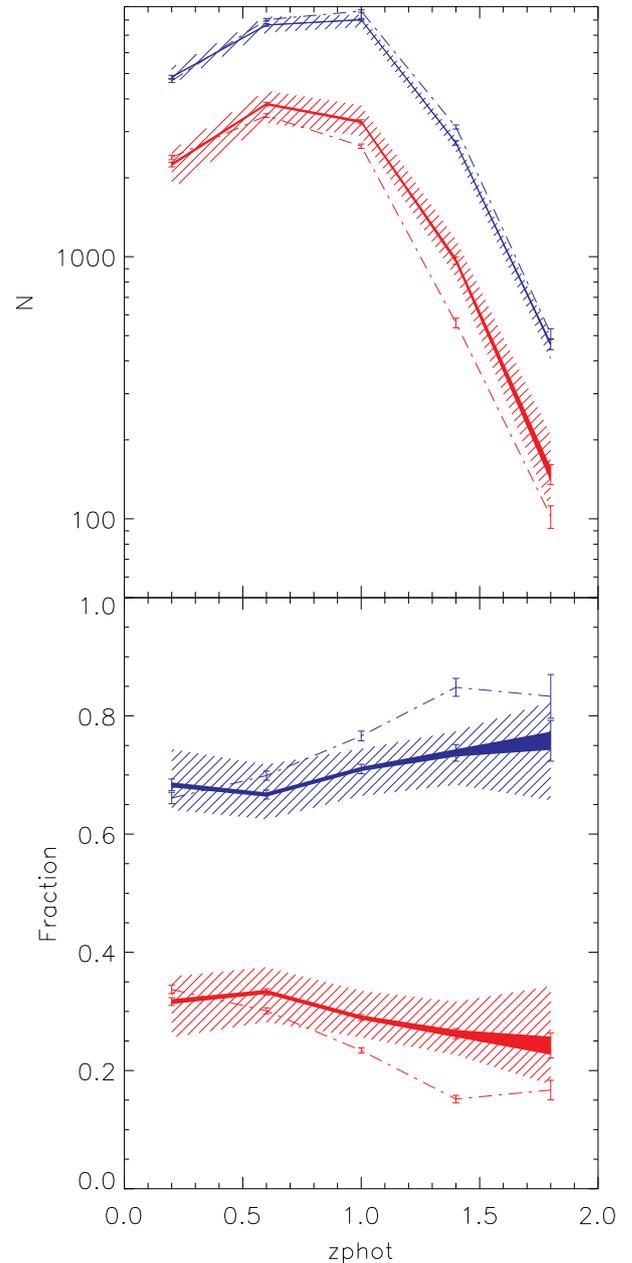
more differences, as expected from the comparisons of the previous section, in particular above  $z \sim 1$ . The match between the two distributions as computed from the KS-test is high at  $z < 1$  (97%) but clearly decreases above  $z \sim 1$  (67%). Interestingly, the I-band estimate goes out of the confidence region at  $z > 1$ . We find in particular an excess of early-type galaxies by a factor  $\sim 1.5$  in the Ks-band when compared to the I-band. Even if at these redshifts, the Ks-band classification is less accurate (§ 3.4), the difference is significant after correcting for the classification errors. The differences are probably due to morphological k-correction effects. It seems therefore that a morphological classification based on HST/ACS I-band data at those redshifts tends to under-estimate the elliptical population. Nevertheless, this result have to be confirmed with a more precise study of the early-type population at  $z \sim 1.5$  by studying their star forming histories and mass distributions.

## 6. Summary and conclusions

We have presented a morphological classification in two main morphological types of 44 089 galaxies within the COSMOS field from seeing limited near-infrared imaging. Morphologies are estimated with the non-parametric N-dimensional code galSVM using a 10 dimensional volume and non-linear boundaries.

The final output catalogue includes for every galaxy, a class label (early or late) and a probability of belonging to the class. The probability is proved to be highly correlated with the success rate and can therefore be used to assess the accuracy of the classifications.

This classification method has been used to obtain the number counts and the redshift distribution per morphological



**Fig. 10.** Redshift distribution per morphological type for the same 34 165 Ks-selected galaxies observed in the I-band (dotted lines) and in the Ks-band (solid lines). Red lines are early-type objects and blue lines are late-type. The line width is the  $1 - \sigma$  confidence interval computed from the probability distribution, the dashed region is the confidence region deduced from Monte Carlo simulations of the training sample and the error bars are calculated using Poisson  $\sqrt{n}$  statistics (see text for details).

type up to  $z \sim 2$  and to compare the results with the ones obtained from HST/ACS imaging on 34 165 galaxies in order to quantify morphological k-correction effects.

Our main conclusions are hereafter summarized:

Concerning the reliability of our morphological classification:

- (i) According to the simulated test sample, the average success rate is  $\sim 80\%$  for the whole sample for both morphological classes, leading to  $\sim 20\%$  contaminations.
- (ii) The probability parameter which results from the classification procedure is a good estimator of the reliability of the classifications since it is highly correlated with the success rate. Objects with probabilities greater than 0.8 are indeed identified with nearly 90% of confidence.
- (iii) The study of the probability distributions as a function of the magnitude, the size and the redshift of the galaxies reveals that the most difficult class to isolate are faint ( $K_s > 20$ ), small ( $\log(r_{half}) < -0.2$ ) early-type objects above  $z \sim 1$ . However, even for this class of objects, the average probability is around  $p \sim 0.7$ .
- (iv) We also showed that selection based on a probability threshold does lead to a biased sample towards late-type systems. We proposed however a way of integrating the information brought by the probability parameter to perform statistical analysis.
- (v) Errors due the training sample dominate the error source.

Concerning the morphological distribution of our sample:

- (i) The global morphological mixing is  $30711.7 \pm 1952$  ( $\sim 70\% \pm 5\%$ ) late-type galaxies and  $13376.3 \pm 2014$  ( $\sim 30\% \pm 5\%$ ) early-type galaxies. These results take into account the classification errors thanks to the use of the probability information. The galaxy population is therefore dominated by late-type systems up to  $z \sim 2$ .
- (ii) We detect a decrease of the fraction of early-type objects with redshift. There are  $\sim 32\% \pm 5\%$  early-type galaxies at  $z < 0.4$  in our sample whereas the fraction at  $z \sim 1.5$  is  $21.9\% \pm 8\%$ . This effect persists even after correcting for classification errors using the probability parameter and therefore confirms a result obtained in previous works using HST imaging. This result probably reflects a progressive building up of the red sequence from late-type objects.

The comparison of the morphologies with the ones obtained with HST/ACS in the I-band for 34 165 objects reveals several interesting trends:

- (i) The global morphological mixings are globally consistent. We find 9 163 early-type galaxies and 25 002 late-type galaxies in the I-band and 10 458 early-type for 23 706 late-type galaxies in the Ks-band.
- (ii) A galaxy classified as late-type in the I-band has a mean probability of  $p \sim 0.9$  of being classified as late-type in the Ks-band.
- (iii) The match between the two photometric bands for early-type galaxies depends on redshift: Below  $z \sim 1$ , an early-type galaxy in the I-band has a probability of  $p \sim 0.7$  of being early-type in the Ks-band. Above this redshift, where the HST cameras are probing the UV flux, the probability decreases and reaches  $p \sim 0.4$  at  $z \sim 1.5$ .
- (iv) The comparison of the redshift distributions reveals also a redshift dependence: below  $z \sim 1$  the two redshift distributions match quite well. The Kolmogorov-Smirnov test gives a probability of match of  $p = 0.97$ . Above  $z \sim 1$ , the I-band classification tends to find less early-type galaxies than the Ks-band one by a factor  $\sim 1.5$ . This probably reflects a morphological k-correction effect. Therefore, studies based on HST classifications at those redshifts could underestimate the elliptical population.

The results presented here quantify the bias in morphological classification due to morphological band shifting between I-band and K-band imaging data. We estimate that the fraction of missing early type galaxies at  $z \sim 1$  from I-band ACS data is about 19%, increasing to 31% at  $z \sim 1.5-2$ . From our K-band data we estimate that the fraction of early-type galaxies has increased by 10% between  $z=1.5$  and  $z=0$ , a further confirmation of the gradual build up of the elliptical galaxy population at the expense of late-type galaxies. The Ks-band classification performed here is intended as a framework for future studies of the evolution of counts, luminosities, luminosity densities and correlation function for each morphological type over several redshift bins.

## References

- Abraham, R., van den Bergh, S., Glazebrook, K., et al. 1996, *ApJ Supplement*, 107, 1
- Abraham, R. G., van den Bergh, S., & Nair, P. 2003, *ApJ*, 588, 218
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
- Arnouts, S., Walcher, C. J., Le Fevre, O., et al. 2007, *ArXiv e-prints*, 705
- Baker, A. J., Davies, R. I., Lehnert, M. D., et al. 2003, *A&A*, 406, 593
- Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
- Brinchmann, J., Abraham, R., Schade, D., et al. 1998, *ApJ*, 499, 112
- Cassata, P., Cimatti, A., Franceschini, A., et al. 2005, *MNRAS*, 357, 903
- Chang, C.-C. & Lin, C.-J. 2001, LIBSVM: a library for support vector machines, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Conselice, C. & GNS Team. 2007, in *American Astronomical Society Meeting Abstracts*, Vol. 211, *American Astronomical Society Meeting Abstracts*
- Conselice, C. J., Bershad, M. A., Dickinson, M., & Papovich, C. 2003, *AJ*, 126, 1183
- Conselice, C. J., Bershad, M. A., & Jangren, A. 2000, *ApJ*, 529, 886
- Cresci, G., Davies, R. I., Baker, A. J., et al. 2006, *Astronomy and Astrophysics*, 458, 385
- de Vaucouleurs, G. 1948, *Annales d'Astrophysique*, 11, 247
- Freeman, K. C. 1970, *ApJ*, 160, 811
- Gardner, J. P., Cowie, L. L., & Wainscoat, R. J. 1993, *ApJ*, 415, L9
- Huertas-Company, M., Rouan, D., Soucail, G., et al. 2007, *A&A*, 468, 937
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006a, *ArXiv Astrophysics e-prints*
- Ilbert, O., Lauger, S., Tresse, L., et al. 2006b, *A&A*, 453, 809
- Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196
- Lilly, S. & The Zcosmos Team. 2005, *The Messenger*, 121, 42
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163
- Madgwick, D. S., Lahav, O., Baldry, I. K., et al. 2002, *MNRAS*, 333, 133
- Maihara, T., Iwamuro, F., Tanabe, H., et al. 2001, *PASJ*, 53, 25
- McCracken, H. J., Metcalfe, N., Shanks, T., et al. 2000a, *MNRAS*, 311, 707
- McCracken, H. J., Metcalfe, N., Shanks, T., et al. 2000b, *MNRAS*, 311, 707
- Menanteau, F., Ellis, R. S., Abraham, R. G., Barger, A. J., & Cowie, L. L. 1999, *MNRAS*, 309, 208
- Menanteau, F., Ford, H. C., Motta, V., et al. 2006, *AJ*, 131, 208
- Platt, J. 2000, *Advances in Large Margin Classifiers*
- Saracco, P., Longhetti, M., Andreon, S., & Mignano, A. 2008, *ArXiv e-prints*, 801
- Scoville, N. Z. & COSMOS Team. 2005, in *Bulletin of the American Astronomical Society*, 1309–+
- Tasca, L. & White, S. 2005, *astro-ph/0507249*
- Thibault, S., Cui, Q., Poirier, M., et al. 2003, in *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, Vol. 4841, *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. Edited by Iye, Masanori; Moorwood, Alan F. M. *Proceedings of the SPIE*, Volume 4841, pp. 932-943 (2003), ed. M. Iye & A. F. M. Moorwood, 932–943
- Wolf, C., Meisenheimer, K., Rix, H.-W., et al. 2003, *A&A*, 401, 73
- Wu, T. F., J., L. C., & C., W. R. 2004, *Journal of machine learning research*, 5, 975
- Zirm, A. W., Stanford, S. A., Postman, M., et al. 2008, *ArXiv e-prints*, 802
- Zucca, E., Ilbert, O., Bardelli, S., et al. 2006, *A&A*, 455, 879