



**HAL**  
open science

# Modélisation Bayésienne des mesures de vitesses particulières dans le projet CosmicFlows

Romain Graziani

► **To cite this version:**

Romain Graziani. Modélisation Bayésienne des mesures de vitesses particulières dans le projet CosmicFlows. Cosmologie et astrophysique extra-galactique [astro-ph.CO]. Université de Lyon, 2018. Français. NNT : 2018LYSE1143 . tel-02068966v2

**HAL Id: tel-02068966**

**<https://theses.hal.science/tel-02068966v2>**

Submitted on 18 Mar 2019

**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.



N° d'ordre NNT : xxx

**THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON**  
opérée au sein de  
**l'Université Claude Bernard Lyon 1**

**École Doctorale ED52**  
**École Doctorale de Physique et Astrophysique**

**Spécialité de doctorat : Physique**  
**Discipline : Astrophysique**

Soutenue publiquement le 14/09/2018, par :  
**Romain Graziani**

---

**Modélisation Bayésienne des mesures de vitesses particulières dans  
le projet *CosmicFlows***

---

Devant le jury composé de :

Mme DAVIS Tamara  
M. KIM Alex  
Mme EALET Anne  
M. TULLY R. Brent  
M. REGNAULT Nicolas  
Mme COURTOIS Hélène

Rapporteure  
Rapporteur  
Examinatrice  
Examineur  
Examineur  
Directrice de thèse



## Résumé

Le modèle de concordance de la cosmologie moderne repose entre autre sur l'existence de matière dite « noire », matière qui n'interagirait que gravitationnellement et qui ne pourrait donc pas être observée directement. Les vitesses particulières des galaxies, puisqu'elles tracent le champ de gravité, sont des sondes non-biaisées de la matière dans l'Univers. Ainsi, l'étude de ces vitesses particulières permet non seulement de cartographier l'Univers proche (matière noire comprise), mais aussi de tester le modèle  $\Lambda$ CDM via la vitesse d'expansion de l'Univers et le taux de formation des structures.

Observationnellement, la mesure de la vitesse particulière d'une galaxie s'obtient à partir de la mesure de sa distance, mesure très imprécise pour les données extragalactiques. Mal modélisées, les incertitudes conduisent à des analyses biaisées des vitesses particulières, et ainsi détériorent la qualité de cette sonde cosmologique.

Dans ce contexte, cette thèse s'intéresse aux erreurs systématiques statistiques des analyses de vitesses particulières. D'abord en étudiant puis en modélisant ces erreurs systématiques. Ensuite en proposant de nouveaux modèles pour les prendre en compte. En particulier, y est développé un modèle permettant, à partir des mesures de la vitesse de rotation des galaxies, de reconstruire le champ de densité de l'Univers Local. Ce modèle s'appuie sur l'analyse des corrélations de vitesses données par le modèle de concordance, et la modélisation de la relation de Tully-Fisher, qui lie la vitesse de rotation des galaxies à leur luminosité. Le modèle développé est appliqué au catalogue de distances extragalactiques *CosmicFlows-3*, permettant ainsi une nouvelle cartographie de l'Univers proche et de sa cinématique.

## Abstract

The cosmological concordance model relies on the existence of a « dark » matter which hypothetically only interacts through gravity. Hence, the dark matter could not be observed directly with standard techniques. Since they directly probe gravity, peculiar velocities of galaxies are an unbiased tool to probe the matter content of the Universe. They can trace the total matter field and constrain the Local Universe's expansion rate and growth of structures.

The peculiar velocity of a galaxy can only be measured from its distance, which determination is very inaccurate for distant objects. If not correctly modeled, these uncertainties can lead to biased analyses and poor constraints on the  $\Lambda$ CDM model.

Within this context, this PhD studies the systematic and statistical errors of peculiar velocity analyses. First by investigating and modeling these errors. Then by building Bayesian models to include them. In particular, a model of the Local Universe's velocity field from the observations of the rotational velocity of galaxies is presented. This model relies on the  $\Lambda$ CDM's peculiar velocity correlations and on a Tully-Fisher relation model. The model has then been applied to the *CosmicFlows-3* catalog of distances and provides a new kinematic map of the Local Universe.



# Introduction générale

L'Univers est en expansion accélérée. Il y a environ 13 milliards d'années, il était homogène et isotrope, et pourtant on observe aujourd'hui des structures à grande échelle telles que les amas de galaxies ou les vides cosmiques. Ces structures nous renseignent sur les processus cosmologiques en jeu : expansion, constante cosmologique et interaction gravitationnelle. Ces processus sont dynamiques et sont donc analysables à travers la cinématique des objets extragalactiques. Les vitesses particulières des galaxies offrent un outil idéal pour tester à la fois les lois de la relativité générale et le modèle de concordance de la cosmologie moderne. Cependant, les mesures de vitesses particulières sont difficiles et peu précises, provoquant biais observationnels et statistiques dans leur analyse. Pour éviter ces biais, l'association des nouveaux outils statistiques avec les puissances de calcul actuelles offrent une solution : la modélisation bayésienne. Cette thèse s'applique à offrir une vision de cette modélisation des mesures de vitesses particulières dans le projet *CosmicFlows*.

Le modèle de concordance de la cosmologie moderne suppose que la matière dans l'Univers est composée d'environ 85% de matière « noire », composante invisible dont la nature n'est pas connue dans le modèle standard de la physique des particules. La matière noire a été mise en évidence par l'étude des courbes de rotation des galaxies spirales (Rubin and Ford, 1970; Rubin et al., 1980; Roberts and Rots, 1973), mais aussi par l'étude de la cinématique des amas de galaxies (Einasto et al., 1974; Hubble and Humason, 1931; Zwicky, 1933) et l'observation de la collision de l'Amas de la Balle (*Bullet Cluster*, Brownstein and Moffat (2007)). Dans le modèle de concordance, la matière noire est dite « froide » et ne semble interagir que gravitationnellement. Cette propriété fondamentale rend difficile une analyse directe de ses propriétés, et il faut s'appuyer sur des sondes indirectes comme l'étude de la distribution des galaxies (Percival et al., 2007), du cisaillement gravitationnel (Mellier, 2010; Nusser et al., 2013) ou des vitesses particulières de galaxies (Lahav et al., 1990; Lynden-Bell et al., 1988; Vittorio and Juszkievicz, 1989).

## Du CMB aux grandes structures de l'Univers

Les observations du fond diffus cosmologique (Hinshaw et al., 2013; Ade et al., 2014) montrent que l'Univers était, il y a 13 milliards d'années, pratiquement homogène et isotrope. Depuis, les petites instabilités gravitationnelles présentes au moment de la recombinaison ont évolué vers la formation de grandes structures que l'on peut observer aujourd'hui. Ces structures sont la conséquence d'un compromis entre l'effondrement des surdensités de ma-

tière (Blumenthal et al., 1984) et l’expansion de l’Univers (Hubble and Humason, 1931). La présence de matière baryonique au sein de ces surdensités, bien que mal connue, fait qu’aujourd’hui les galaxies se trouvent approximativement au centre de ces surdensités. L’observation de la distribution des galaxies (Gott et al., 2005) dans l’Univers Local montre que les grandes structures de l’Univers, au-delà de l’échelle galactique, sont réparties en amas, vides cosmiques et filaments.

Comment, à partir des fluctuations primordiales (Langlois, 2005), ces grandes structures se sont-elles formées ? Cette question de la formation des structures est centrale dans la cosmologie moderne. D’un point de vue théorique, une réponse paraît difficile du fait de la nature hautement non linéaire des équations décrivant l’évolution des champs cosmiques qui décrivent la matière dans l’Univers (Doroshkevich et al., 1973). L’évolution non linéaire de la matière dans l’Univers peut toutefois être modélisable en utilisant des simulations numériques qui permettent d’accéder aux mécanismes de formation des grandes structures. Ces simulations de matière noire retrouvent avec succès la plupart des propriétés des grandes structures. Elles échouent toutefois à décrire entièrement les observations et soulèvent des problèmes comme la sur-estimation du nombre de galaxies satellites d’une galaxie comme la Voie Lactée (Klypin et al., 1999; Moore et al., 1999), ou la cinématique inexplicée de ses satellites (Boylan-Kolchin et al., 2011).

L’autre grande difficulté dans l’étude du mécanisme de formation des grandes structures est observationnelle : comment nos observations, lumineuses, nous renseignent-elles sur la distribution de matière, et donc de matière noire ? L’étude directe de la distribution des galaxies comme traceurs de la matière dans l’Univers est en effet problématique car elle requiert l’introduction du biais galaxie-matière (Weinberg et al., 2004) qui s’appuie sur une modélisation empirique (Peacock and Smith, 2000) qui elle-même suppose le modèle  $\Lambda$ CDM.

## Les vitesses particulières

La matière, qu’elle soit visible ou non, du fait de la gravité, provoque le mouvement. Ainsi, c’est en étudiant le mouvement de la matière ordinaire que nous pouvons retracer la distribution de matière totale dans l’Univers. Les vitesses particulières des galaxies sont de ce point de vue un traceur non biaisé de la matière et peuvent donc être considérées comme sondes du modèle  $\Lambda$ CDM.

Les vitesses particulières interviennent dans différents régimes. Au sein d’amas de galaxies, elles créent des distorsions de l’espace des décalages vers le rouge (Kaiser, 1987) qui sont utilisées comme sonde du modèle  $\Lambda$ CDM (Percival and White, 2009). Dans le régime linéaire ou quasi non-linéaire, elles ont déjà permis la cartographie de l’Univers Local (Courtois et al., 2012; Tully et al., 2014), mais aussi l’étude statistique des mouvements des galaxies proches (Tully et al., 2008; Hong et al., 2014; Hoffman et al., 2015; Scrimgeour et al., 2016). En particulier, elles auraient montré que le dipôle observé dans le fond diffus cosmologique serait trop important comparé aux prédictions du modèle  $\Lambda$ CDM (Kashlinsky et al., 2008).

Malgré tout, l’étude des vitesses particulières, bien que datant d’il y a quelques dizaines d’années, n’est toutefois pas encore compétitive en tant que sonde cosmologique. La raison est strictement observationnelle. La mesure d’une vitesse particulière requiert deux observables :

la mesure du décalage vers le rouge, qui est un estimateur de la vitesse totale de la galaxie, et la distance, qui permet d'évaluer la contribution de l'expansion à la vitesse totale. La mesure de distance extragalactique est cependant imprécise, avec une incertitude relative de près de 20%, ce qui rend les analyses de vitesses particulières difficiles.

La situation n'est cependant pas désespérée! L'émergence de nouveaux outils statistiques permet, par une modélisation des erreurs, de maximiser les informations extraites des mesures. Cette thèse s'intègre dans cette optique. La modélisation précise des observations dans un cadre bayésien est indispensable pour l'analyse des vitesses particulières.

## Plan de la thèse

Le but de cette thèse est de contribuer à l'accession des analyses de vitesses particulières au rang de sonde cosmologique. Comme nous l'avons vu, cet objectif ambitieux repose sur la modélisation statistique de toutes les observations liées aux mesures de vitesses particulières dans un cadre Bayésien. Bien sûr, il est impensable qu'une thèse puisse suffire à achever seule un tel projet, et celle-ci échoue à remplir totalement son objectif. Cependant, elle propose une base solide pour les analyses à venir et explore les possibilités qu'offrent les vitesses particulières. D'abord en étudiant les erreurs systématiques des analyses usuelles, ensuite en développant des algorithmes qui incluent les observations dans les analyses statistiques.

Cette thèse s'appuie sur un outil nommé le filtre de Wiener (Zaroubi et al., 1995), qui permet de reconstruire le champ de vitesse *linéaire* sous-jacent à l'observation de vitesses particulières (Hoffman and Ribak, 1991). Le filtre de Wiener est le filtre idéal pour des observations gaussiennes dont on connaît le spectre de puissance. Dans le contexte des vitesses particulières, le spectre de puissance est donné par le modèle  $\Lambda$ CDM.

Le premier chapitre introduit les notions de base nécessaires à la lecture de cette thèse. En particulier sont détaillées les équations du modèle  $\Lambda$ CDM et de la théorie linéaire de formation de structures. Les observations utiles à l'étude des vitesses particulières y sont décrites et les outils statistiques qui seront utilisés dans la suite sont introduits.

Le deuxième chapitre analyse la technique du filtre de Wiener. Il montre comment une mauvaise modélisation des observations résulte en l'apparition d'erreurs systématiques sur le champ de vitesse reconstruit. Une description Bayésienne des observations de vitesses particulières qui met naturellement en évidence les biais des analyses existantes y est décrite. Sa conclusion est sans appel : le modèle doit être revu.

C'est donc dans le troisième chapitre qu'est décrit l'évolution naturelle du modèle. Ce modèle, qui s'appuie sur les travaux de Lavaux (2016), permet la modélisation de la mesure des distances ainsi que, bien que partiellement, des effets de sélection. Son implémentation y est décrite ainsi que les techniques d'échantillonnage du modèle. Le modèle est ensuite testé sur des catalogues simulés.

Le quatrième chapitre est consacré à l'application de la méthode au catalogue *CosmicFlows-3* contenant la mesure d'environ 17,000 distances et décalages vers le rouge. Une analyse préliminaire du catalogue y est donnée et les champs de vitesse et surdensités linéaires sous-jacents sont reconstruits. Il en résulte une cartographie de la matière dans l'Univers Local et une analyse cosmographique préliminaire est proposée. Le chapitre 4 souligne aussi les limites

de l'analyse. Ces limites concernent principalement les effets de sélection et la propagation des erreurs de mesure depuis les observations primaires.

Le chapitre 5 intègre la relation de Tully-Fisher (Tully and Fisher, 1977) dans l'analyse des vitesses particulières. La relation de Tully-Fisher, qui relie la vitesse de rotation d'une galaxie spirale à sa luminosité, est l'un des deux indicateurs de distance principaux des données *CosmicFlows*. Ce chapitre montre comment les vitesses particulières peuvent être modélisées directement depuis les observations primaires afin d'inclure tous les effets observationnels inhérents à l'indicateur de distance utilisé : effets de sélection, biais statistiques, effets de couleur etc. La méthode est ensuite appliquée avec succès au catalogue SFI++ (Springob et al., 2009).

Enfin, le chapitre 6 résume toutes ces études et proposent celles à venir.

# Table des matières

<b>1</b>	<b>Mesure de vitesses particulières dans le modèle <math>\Lambda</math>CDM</b>	<b>13</b>
1.1	Perturbations d'un Univers homogène en expansion . . . . .	14
1.1.1	Expansion . . . . .	14
1.1.2	Lien avec le champ de densité . . . . .	14
1.1.3	Univers homogène et modèle $\Lambda$ CDM . . . . .	15
1.1.4	Univers inhomogène . . . . .	16
1.1.5	Approximation linéaire . . . . .	17
1.2	Mesure de vitesses particulières . . . . .	18
1.2.1	Le décalage vers le rouge . . . . .	18
1.2.2	Distances . . . . .	19
1.2.3	La magnitude absolue . . . . .	19
1.2.4	Les vitesses particulières radiales . . . . .	21
1.2.5	Cadre bayésien pour l'étude des vitesses particulières . . . . .	21
1.3	La cosmologie et les champs gaussiens . . . . .	23
1.3.1	Le champ de surdensité . . . . .	23
1.3.2	Corrélations des vitesses particulières . . . . .	24
<b>2</b>	<b>Filtre de Wiener et réalisations contraintes</b>	<b>27</b>
2.1	Technique des réalisations contraintes . . . . .	28
2.1.1	Approche fréquentiste . . . . .	28
2.1.2	Approche probabiliste . . . . .	30
2.2	Données simulées . . . . .	31
2.2.1	Génération d'un catalogue simulé . . . . .	31
2.2.2	Application de la méthode WF/CR . . . . .	32
2.2.3	Cas des vitesses particulières mesurées . . . . .	39
2.3	Biais statistiques . . . . .	42
2.3.1	Biais de densité . . . . .	44
2.3.2	Biais de sélection . . . . .	44
2.3.3	Biais de positionnement . . . . .	46
2.3.4	Distribution log-normale des vitesses particulières . . . . .	47
2.4	Reconstructions . . . . .	49
2.4.1	Solutions partielles . . . . .	50
2.5	Conclusion sur la méthode WF/CR . . . . .	52



<b>3</b>	<b>Modèle statistique</b>	<b>57</b>
3.1	Le modèle . . . . .	58
3.1.1	Introduction . . . . .	58
3.1.2	Vraisemblance . . . . .	58
3.1.3	<i>A priori</i> bayésiens . . . . .	64
3.2	Algorithme . . . . .	67
3.2.1	Probabilités conditionnelles . . . . .	67
3.2.2	Marginalisation et algorithme de Gibbs effondré . . . . .	69
3.2.3	Implémentation en PYTHON . . . . .	69
3.3	Tests . . . . .	71
3.3.1	Paramètres globaux et convergences . . . . .	71
3.3.2	Champs de vitesse et de surdensité . . . . .	72
3.4	Conclusion . . . . .	75
<b>4</b>	<b>Champ de vitesse de <i>CosmicFlows-3</i></b>	<b>79</b>
4.1	Le catalogue <i>CosmicFlows-3</i> . . . . .	80
4.1.1	Analyse du catalogue . . . . .	80
4.1.2	Intercalibration . . . . .	83
4.1.3	Mesure des décalages vers le rouge . . . . .	84
4.1.4	Groupes . . . . .	84
4.2	Application de la méthode au catalogue . . . . .	85
4.2.1	MCMC et convergence . . . . .	85
4.2.2	Irrotationnalité du champ reconstruit . . . . .	86
4.2.3	Paramètres ajustés . . . . .	88
4.3	Cosmographie de l'Univers Local . . . . .	89
4.3.1	Structures de l'Univers Local . . . . .	93
4.3.2	Comparaison aux catalogues de décalages vers le rouge . . . . .	94
4.3.3	La toile cosmique des vitesses . . . . .	95
4.3.4	Comparaison filtre de Wiener, <i>CosmicFlows-2</i> , <i>CosmicFlows-3</i> . . . . .	101
4.3.5	Études possibles . . . . .	103
4.4	Erreurs systématiques . . . . .	103
4.4.1	Idée générale . . . . .	103
4.4.2	Influence des <i>a priori</i> . . . . .	105
4.4.3	Influence de la calibration des indicateurs de distance . . . . .	107
4.4.4	Dipôle externe . . . . .	110
4.5	Conclusion . . . . .	110
<b>5</b>	<b>Modélisation bayésienne de la relation de Tully-Fisher</b>	<b>113</b>
5.1	Introduction . . . . .	114
5.1.1	La relation de Tully-Fisher . . . . .	114
5.1.2	De la nécessité d'une modélisation précise . . . . .	115
5.2	Le modèle . . . . .	116
5.2.1	Vraisemblance . . . . .	116
5.2.2	<i>A priori</i> . . . . .	119

5.2.3	L'algorithme . . . . .	120
5.3	Tests . . . . .	120
5.3.1	Catalogue simulé . . . . .	120
5.3.2	Relation de Tully-Fisher . . . . .	121
5.3.3	Couplage avec le champ de vitesse . . . . .	122
5.3.4	Avec la calibration . . . . .	125
5.4	SFI++ . . . . .	128
5.5	Discussion . . . . .	129
5.6	Conclusion . . . . .	131
<b>6</b>	<b>Perspectives</b>	<b>137</b>
6.1	Vers un modèle non linéaire . . . . .	137
6.2	Les vitesses particulières comme sondes cosmologiques . . . . .	138
6.3	Les Supernovae de type Ia et les vitesses particulières . . . . .	140
	<b>Conclusion</b>	<b>143</b>



# Chapitre 1

## Mesure de vitesses particulières dans le modèle $\Lambda$ CDM

Nous abordons dans ce chapitre les notions indispensables à la lecture de ce manuscript. Cette thèse repose sur le modèle  $\Lambda$ CDM de concordance de la cosmologie ainsi que sur la théorie linéaire de formation des structures. Dans la section 1.1 nous décrivons comment les vitesses particulières peuvent être définies comme déviation au flot de Hubble, conséquence de l'expansion. Nous soulignons notamment le lien entre les vitesses particulières et le champ de densité de matière sous-jacent. Dans la section 1.2 nous abordons le principe de mesure d'une vitesse particulière. Cette mesure étant peu précise, on motivera l'usage de statistiques bayésiennes dans toute cette thèse. Enfin, nous détaillons dans la section 1.3 comment le modèle  $\Lambda$ CDM décrit les propriétés statistiques des champs de vitesse et de densité de l'Univers.

### Sommaire

---

<b>1.1</b>	<b>Perturbations d'un Univers homogène en expansion . . . . .</b>	<b>14</b>
1.1.1	Expansion . . . . .	14
1.1.2	Lien avec le champ de densité . . . . .	14
1.1.3	Univers homogène et modèle $\Lambda$ CDM . . . . .	15
1.1.4	Univers inhomogène . . . . .	16
1.1.5	Approximation linéaire . . . . .	17
<b>1.2</b>	<b>Mesure de vitesses particulières . . . . .</b>	<b>18</b>
1.2.1	Le décalage vers le rouge . . . . .	18
1.2.2	Distances . . . . .	19
1.2.3	La magnitude absolue . . . . .	19
1.2.4	Les vitesses particulières radiales . . . . .	21
1.2.5	Cadre bayésien pour l'étude des vitesses particulières . . . . .	21
<b>1.3</b>	<b>La cosmologie et les champs gaussiens . . . . .</b>	<b>23</b>
1.3.1	Le champ de surdensité . . . . .	23
1.3.2	Corrélations des vitesses particulières . . . . .	24

---

## 1.1 Perturbations d'un Univers homogène en expansion

### 1.1.1 Expansion

La première révolution observationnelle cosmologique est l'observation de l'expansion de l'Univers par Hubble en 1929 (Hubble, 1929). Du fait de cette expansion, toutes les distances *physiques* entre les galaxies augmentent avec le temps, et on modélise cette dépendance en introduisant le facteur d'échelle  $a(t)$  tel que la distance *physique*  $r = \|\mathbf{r}\|$  d'une galaxie par rapport à la Voie Lactée évolue proportionnellement à une constante  $\chi = \|\boldsymbol{\chi}\|$ , sa distance comobile :

$$r(t) = a(t)\chi \quad (1.1)$$

Du fait de la gravité, les coordonnées comobiles des galaxies ne sont en fait pas constantes. La vitesse particulière est définie comme la variation de sa coordonnée comobile  $\boldsymbol{\chi}$  avec le temps. La vitesse *apparente*  $\mathbf{V}$  d'une galaxie est donc :

$$\begin{aligned} \mathbf{V} &\triangleq \dot{\mathbf{r}} \\ &= \dot{a}\boldsymbol{\chi} + a\dot{\boldsymbol{\chi}} \\ &= a(H(t)\boldsymbol{\chi} + \dot{\boldsymbol{\chi}}) \end{aligned} \quad (1.2)$$

où on a défini le paramètre de Hubble  $H(t) \triangleq \frac{\dot{a}}{a}$  comme le taux d'expansion. L'équation (1.2) comprend deux termes :  $H(t)\boldsymbol{\chi}$  correspond à la vitesse de récession due à l'expansion seule,  $\dot{\boldsymbol{\chi}}$  correspond à la vitesse particulière de la galaxie. En pratique, le deuxième terme est de l'ordre de  $\sim 500 \text{ km s}^{-1}$  alors que  $H(t) \sim H_0 \sim 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (Riess et al., 2016; Planck Collaboration et al., 2016) où  $H_0 = H(t = t_0)$  est la constante de Hubble. La vitesse de récession est donc dominante pour les galaxies distantes de plus de quelques  $\chi \gtrsim 10 \text{ Mpc}$ .

### 1.1.2 Lien avec le champ de densité

Supposons que nous ayons accès aux vitesses particulières  $\dot{\boldsymbol{\chi}}$  des galaxies de l'Univers. Comment à partir de ces vitesses particulières est-il possible de remonter au champ de densité  $\rho$  de la matière dans l'Univers ?

Pour cela, on suppose d'abord que cette matière se comporte comme un fluide parfait de pression nulle. Pour alléger les notations, on note  $\mathbf{v} = \dot{\boldsymbol{\chi}}$  la vitesse particulière de la matière et on rappelle que  $\mathbf{V}$  dénote sa vitesse totale. Les équations qui lient la vitesse et la densité sont les équations couplées de conservation de la masse, d'Euler et de Poisson :

$$\frac{d\rho}{dt} = 0, \quad (1.3)$$

$$\frac{d\mathbf{V}}{dt} = -\nabla_{\mathbf{r}}\phi, \quad (1.4)$$

$$\nabla_{\mathbf{r}}^2\phi = 4\pi G\rho. \quad (1.5)$$

Le potentiel  $\phi$  représente ici le potentiel gravitationnel total de l'Univers. Dans cette thèse nous nous intéressons aux déviations du champ de densité et de vitesse par rapport à un

Univers homogène où la vitesse des galaxies est expliquée par l'expansion seule. Pour résoudre les équations (1.3), (1.4) et (1.5), il convient donc de définir quelle est la dynamique d'un Univers homogène uniquement gouverné par l'expansion.

### 1.1.3 Univers homogène et modèle $\Lambda$ CDM

La première supposition de la cosmologie moderne est le *principe cosmologique* qui stipule que l'Univers doit être homogène et isotrope à grande échelle. Dans toute cette thèse, on néglige les effets de courbure en la considérant nulle.

Dans l'Univers homogène et isotrope on peut montrer que la métrique de l'espace-temps est définie par la métrique Friedmann-Lemaître-Robertson-Walker (Friedman, 1922; de Sitter, 1917; Lemaître, 1927) :

$$ds^2 = c^2 dt^2 - a^2(t) (d\chi^2 + \chi^2 (d\theta^2 + \sin^2 \theta d\phi^2)). \quad (1.6)$$

À partir de cette métrique, les équations d'Einstein appliquées à un Univers homogène de masse volumique  $\rho_{\text{tot}}$  et de pression  $p$ , donnent les équations de Friedmann :

$$H^2 \triangleq \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \rho_{\text{tot}} + \frac{\Lambda c^2}{3}, \quad (1.7)$$

$$\dot{H} + H^2 = \frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2} (\rho_{\text{tot}} c^2 + 3p) + \frac{\Lambda c^2}{3} \quad (1.8)$$

où  $\Lambda$  est la constante cosmologique. Ces équations de Friedmann donnent l'évolution temporelle du facteur d'échelle en fonction du contenu de l'Univers, ce qui nous sera utile pour la résolution du système d'équations (1.3), (1.4) et (1.5).

L'énergie dans l'Univers peut être décrite en terme de matière non-relativiste,  $\bar{\rho}(t) \propto a^{-3}$ , d'énergie de radiation,  $\rho_r(t) \propto a^{-4}$  et constante cosmologique,  $\rho_\Lambda = Cte$ . En définissant les densités réduites :

$$\Omega_X = \frac{8\pi G \rho_X}{3H_0^2}, \quad (1.9)$$

on peut réécrire l'équation (1.7) :

$$H^2(t) = H_0^2 (\Omega_R a^{-4} + \Omega_m a^{-3} + \Omega_\Lambda). \quad (1.10)$$

Les paramètres cosmologiques  $\Omega_\Lambda$  et  $\Omega_m$  ont été récemment mesurés par Betoule et al. (2014); Planck Collaboration et al. (2016) aux valeurs  $\Omega_m = 0.295 \pm 0.034$  et  $\Omega_\Lambda = 0.6911 \pm 0.0062$ .

On a donc noté  $\bar{\rho}(t)$  la densité de matière de cet Univers homogène à un temps  $t$ . Elle est liée à la densité d'énergie réduite  $\Omega_m$  (définie au temp  $t = t_0$ ) par :

$$\bar{\rho}(t) = \Omega_m \frac{3H^2}{8\pi G} \quad (1.11)$$

où le facteur  $a^3$  correspond à la dilution géométrique de la matière avec l'expansion. On lui associe le potentiel gravitationnel  $\bar{\phi}$  tel qu'il respecte l'équation de Poisson :

$$\nabla_r^2 \bar{\phi} = 4\pi G \bar{\rho}. \quad (1.12)$$

### 1.1.4 Univers inhomogène

Aujourd'hui l'Univers est inhomogène jusqu'à des échelles de l'ordre d'une centaine de Mpc, et cette thèse s'emploie à caractériser ces inhomogénéités à partir de la mesure des vitesses particulières. Nous avons déjà introduit les déviations des vitesses par rapport à l'expansion par l'équation (1.2). Les déviations au champ de densité sont introduites par le champ de surdensité :

$$\delta(\boldsymbol{\chi}, t) = \frac{\rho_{\chi}(\boldsymbol{\chi}, t) - \bar{\rho}_{\chi}}{\bar{\rho}_{\chi}} \quad (1.13)$$

où la densité comobile est définie par  $\rho_{\chi}(\boldsymbol{\chi}, t) = a^3(t)\rho(\mathbf{r}, t)$ . On introduit de même le potentiel  $\Phi$  tel que :

$$\phi = \Phi + \bar{\phi} \quad (1.14)$$

où  $\Phi$  est le potentiel gravitationnel cosmologique.

On peut finalement écrire les équations liant les vitesses particulières  $\mathbf{v}$ , la surdensité  $\delta$  et le potentiel gravitationnel  $\Phi$  à partir des équations (1.13), (1.3), (1.4) et (1.5). Plutôt que d'écrire ces équations dans l'espace des coordonnées physiques  $\mathbf{r}$ , on souhaite travailler en coordonnées comobiles. Pour cela on utilise les transformations suivantes :

$$\mathbf{V} = a(H(t)\boldsymbol{\chi} + \mathbf{v}), \quad (1.15)$$

$$\frac{df}{dt} = \left. \frac{\partial f}{\partial t} \right|_{\chi} + \nabla_{\chi} \cdot (f\mathbf{v}), \quad (1.16)$$

$$\nabla_{\chi} f = \frac{1}{a} \nabla_r f. \quad (1.17)$$

Il vient :

$$\frac{\partial \rho_{\chi}}{\partial t} + \nabla_{\chi} \cdot (\rho_{\chi} \mathbf{v}) + 3H(t)\rho_{\chi} = 0, \quad (1.18)$$

$$\frac{d\mathbf{v}}{dt} + aH(t)\mathbf{v} + \ddot{a}\boldsymbol{\chi} = -\frac{1}{a} \nabla_{\chi} \phi, \quad (1.19)$$

$$\nabla_{\chi}^2 \phi = 4\pi G a^2 \rho. \quad (1.20)$$

En injectant l'équation de Friedmann (1.8) et le potentiel moyen défini par l'équation (1.11), on peut réécrire ces équations en fonction de la surdensité  $\delta$  et du potentiel gravitationnel cosmologique  $\Phi$  :

$$\frac{\partial \delta}{\partial t} + \nabla_{\chi} \cdot \mathbf{v} = 0, \quad (1.21)$$

$$\frac{d\mathbf{v}}{dt} + aH(t)\mathbf{v} = -\frac{1}{a} \nabla_{\chi} \Phi, \quad (1.22)$$

$$\nabla_{\chi}^2 \Phi = 4\pi G a^2 \bar{\rho} \delta. \quad (1.23)$$

Ces équations fournissent donc le lien entre les vitesses particulières  $\mathbf{v}$  et le champ de densité  $\rho_{\chi}$  de l'Univers. Cependant, ces équations différentielles couplées sont hautement non linéaires et n'admettent pas de solutions analytiques dans le cas général. On peut néanmoins développer  $\delta$  de façon perturbative afin de le résoudre. La prochaine section s'intéresse à l'approximation linéaire dans laquelle on considère que  $\delta \ll 1$  à l'ordre 1.

### 1.1.5 Approximation linéaire

Dans l'approximation linéaire, on suppose que  $\delta \ll 1$ , et on peut ainsi linéariser les équations (1.21), (1.22), (1.23) :

$$\frac{\partial \delta}{\partial t} + \nabla_{\chi} \cdot \mathbf{v} = 0 \quad (1.24)$$

$$\frac{\partial \mathbf{v}}{\partial t} + aH(t)\mathbf{v} = -\frac{1}{a}\nabla_{\chi}\Phi \quad (1.25)$$

$$\nabla_{\chi}^2 \Phi = \frac{3}{2}\Omega_m a H^2 \delta \quad (1.26)$$

où on a remplacé  $\bar{\rho}$  par son expression (1.11). En prenant la divergence de l'équation (1.25) et en injectant les équations (1.24) et (1.26), on obtient l'équation différentielle :

$$\ddot{\delta} + 2aH\dot{\delta} = \frac{3}{2}\Omega_m a^2 H^2 \delta \quad (1.27)$$

Cette équation différentielle peut être résolue en séparant les dépendances temporelle et spatiale :

$$\delta(\chi, t) = D(t)\delta_0(\chi) \quad (1.28)$$

Cette équation signifie que dans l'approximation linéaire, la forme du champ de densité reste constante à un facteur multiplicatif  $D(t)$  près, que l'on nomme le facteur de croissance linéaire.  $D(t)$  est solution de l'équation (1.27) et la solution générale est donnée par une combinaison linéaire d'un mode croissant et d'un mode décroissant :

$$D_+(a) = \frac{1}{D_0} \frac{\dot{a}}{a} \int_0^a \frac{da}{\dot{a}^3} \quad ; \quad D_-(a) \propto \frac{\dot{a}}{a} \quad (1.29)$$

On ne retient ici que le mode croissant  $D_+$  en considérant que le mode décroissant a eu le temps de devenir négligeable depuis  $a = 0$ . Le préfacteur  $D_0$  est conventionnellement pris tel que :

$$D_0 = H_0 \int_0^1 \frac{da}{\dot{a}^3} \quad (1.30)$$

Enfin, l'équation de continuité (1.24) nous permet de lier la vitesse particulière  $\mathbf{v}$  et le champ de surdensité  $\delta$  dans l'approximation linéaire :

$$\nabla \cdot \mathbf{v} = -a\dot{D}\delta_0 = \frac{d \ln D}{d \ln a} \dot{a} \delta \quad (1.31)$$

On introduit le taux de croissance des structures  $f$  défini par :

$$f = \frac{d \ln D}{d \ln a} = \frac{1}{\dot{a}^2} \left( \Omega_{\Lambda} a^2 - \frac{\Omega_m}{2a} \right) + a \left( \dot{a}^3 \int_0^a \frac{da}{\dot{a}^2} \right)^{-1} - 1 \quad (1.32)$$

tel que finalement :

$$\nabla \cdot \mathbf{v} = -aHf\delta \quad (1.33)$$

On voit donc ici comment relier le champ des vitesses particulières  $\mathbf{v}$  au champ de surdensité  $\delta$  en théorie linéaire. Il s'agit de savoir maintenant comment accéder à la mesure des vitesses particulières des galaxies, ce qui fait l'objet de la prochaine section.



## 1.2 Mesure de vitesses particulières

Dans cette section, nous nous intéressons à la mesure des vitesses particulières définies par l'équation (1.2). À ce jour, il n'existe aucune méthode pour mesurer directement la vitesse particulière d'une galaxie. Cependant, il est possible d'utiliser l'équation (1.2) pour en avoir une estimation. En effet, en mesurant la vitesse apparente  $\mathbf{V}$  et la position  $\chi$ , on peut calculer la vitesse particulière à  $t = t_0$  :

$$\mathbf{v} = \mathbf{V} - H_0\chi. \quad (1.34)$$

La vitesse apparente  $\mathbf{V}$  est mesurée à partir du décalage vers le rouge d'un rayonnement, et la distance  $\chi$  à partir du flux lumineux et de la luminosité.

### 1.2.1 Le décalage vers le rouge

Le décalage vers le rouge d'une longueur d'onde est défini par le rapport entre la longueur d'onde d'émission  $\lambda$  et celle mesurée  $\lambda_0$ . Le décalage vers le rouge  $z$  est défini par :

$$1 + z = \frac{\lambda_0}{\lambda} \quad (1.35)$$

En cosmologie observationnelle, le décalage vers le rouge cosmologique  $\bar{z}$  est celui dû à l'expansion de l'Univers seulement. Il est déduit de la variation du facteur d'échelle durant le temps de propagation des photons :

$$1 + \bar{z} = \frac{1}{a} \quad (1.36)$$

En plus de l'effet de l'expansion de l'Univers, il faut rajouter l'effet des vitesses particulières des galaxies qui créent un effet Doppler supplémentaire. En pratique, le décalage vers le rouge  $z$  est mesuré dans le référentiel héliocentrique, on le note  $z_{hel}$  (Davis and Scrimgeour, 2014; Hogg, 1999) :

$$(1 + z_{hel}) = (1 + \bar{z})(1 + z_p)(1 + z_\odot) \quad (1.37)$$

où  $\bar{z}$  est le décalage vers le rouge cosmologique défini par l'équation (1.36);  $z_\odot = v_\odot^r/c$  est le décalage vers le rouge dû à la vitesse du soleil par rapport au CMB le long de la ligne de visée; et  $z_p$  est le décalage vers le rouge dû à la vitesse de la galaxie observée par rapport à l'observateur. En général on note le décalage vers le rouge *dans le référentiel du CMB*  $z$  tel que :

$$1 + z = \frac{1 + z_{hel}}{1 + z_\odot} \quad (1.38)$$

et ainsi

$$1 + z = (1 + \bar{z})(1 + z_p) \quad (1.39)$$

Cette équation correspond à la décomposition de l'origine du décalage vers le rouge en un terme d'expansion et un terme de vitesse particulière. Notons que si l'on interprète le décalage vers le rouge comme un effet Doppler, on peut identifier cette équation à l'équation (1.2) et noter  $\mathbf{V} \cdot \hat{\chi} = cz$ . Mais nous retenons pour la suite de cette thèse l'équation générale (1.39).

### 1.2.2 Distances

Le terme « distance » en cosmologie regroupe plusieurs notions différentes que nous détaillons ici. Nous avons déjà introduit la distance comobile  $\chi$ . Pour exprimer la distance  $\chi$  en fonction du décalage vers le rouge cosmologique on utilise :

$$\chi = \int_{t(\bar{z})}^{t_0} d\chi \quad (1.40)$$

$$= \frac{c}{H_0} \int_0^{\bar{z}} \frac{dz}{E(z)} \quad (1.41)$$

où  $E(z)$  peut être déduit de l'équation (1.10) :

$$E(z) = \frac{H(t)}{H_0} = \left( \Omega_R (1+z)^4 + \Omega_m (1+z)^3 + \Omega_\Lambda \right)^{\frac{1}{2}} \quad (1.42)$$

La distance lumineuse  $d_L$  est définie par le lien entre la luminosité d'un objet  $L$  et le flux reçu  $f$  :

$$f = \frac{L}{4\pi d_L^2} \quad (1.43)$$

En plus de la dilution en  $\chi^{-2}$ , l'énergie des photons diminue avec l'expansion de l'Univers, de même que la fréquence de réception de ces photons. Il vient :

$$d_L = (1+z)\chi = \frac{c(1+z)}{H_0} \int_0^{\bar{z}} \frac{dz}{E(z)} \quad (1.44)$$

Dans la suite on notera  $d \triangleq d_L$  la distance lumineuse. Pour déterminer une distance lumineuse, il faut donc mesurer le flux d'une source ainsi que sa luminosité.

#### 1.2.2.1 Le flux lumineux

Le flux lumineux est exprimé en magnitude apparente  $m = -2.5 \log_{10} \frac{f}{f_0}$ . L'observation du flux est le résultat de l'émission de lumière par une galaxie dont on note  $M = -2.5 \log_{10} \frac{L}{L_0}$  sa magnitude absolue, et de la dilution du flux en  $r^{-2}$  tel que par définition :

$$m = \mu + M \quad (1.45)$$

où  $\mu$  le module de distance est lié à la distance lumineuse de la galaxie :

$$\mu = 5 \log \left( \frac{d}{10 \text{ pc}} \right) \quad (1.46)$$

#### 1.2.3 La magnitude absolue

La magnitude absolue  $M$  d'une galaxie n'est pas une observable mais un paramètre dérivé. Plusieurs méthodes, que l'on appelle des *indicateurs de distance*, existent pour déterminer la magnitude absolue, et quelques-uns utiles à cette thèse sont décrit dans le tableau 1.1. Il est à noter que la relation de Tully-Fisher sera décrite plus précisément dans le chapitre 5.

TABLE 1.1 – Principaux indicateurs de distances extragalactiques utilisés par la collaboration *CosmicFlows*.

Méthode	Abrev.	Description	Portée	$\sigma_\mu$	Exemple d'utilisation
Relation de Tully-Fisher	TFFR	Relation d'échelle entre la vitesse de rotation et la luminosité d'une galaxie spirale	[0, 1] Gpc	$\sim 0.4$	Lagattuta et al. (2013)
Plan fondamental	FP	Relation d'échelle entre le rayon, la dispersion des vitesses et la luminosité des galaxies elliptiques	[0, 1] Gpc	$\sim 0.5$	Magoulas et al. (2012)
Supernovae de type Ia	SNIa	Explosion d'une supernovae de type Ia.	$z \sim 1$	$\sim 0.14$	Betoule et al. (2014)
Céphéïdes	Ceph.	Relation période-luminosité des Céphéïdes	[0, 10] Mpc	$\sim 0.1$	Riess et al. (2018)
Sommet de la branche des géantes rouges	TRGB	Luminosité des étoiles de la branche des géantes rouges les plus lumineuses	[0, 10] Mpc	$\sim 0.1$	
Fluctuations de brillance de surface	SBF	Brillance de surface des galaxies elliptiques	[0, 100] Mpc	$\sim 0.25$	Tonry et al. (2000)

### 1.2.4 Les vitesses particulières radiales

La vitesse particulière radiale est également un paramètre dérivé. Elle peut se calculer en combinant les équations (1.44) et (1.39) :

$$v^r = c \frac{z - \bar{z}(d)}{1 + \bar{z}(d)}. \quad (1.47)$$

Une approximation souvent utilisée à faible décalage vers le rouge est la suivante :

$$v^r \sim cz - H_0 d_L. \quad (1.48)$$

L'erreur sur  $v^r$  défini par l'équation (1.48) est :

$$\sigma_{v^r} = H_0 \sigma_\mu \frac{\log 10}{5} d_L \quad (1.49)$$

qui est proportionnelle à la distance. En pratique l'erreur sur les modules de distance sont de l'ordre de 0.4 en magnitudes, ce qui donne par exemple une erreur sur la vitesse particulière d'environ  $1500 \text{ km s}^{-1}$  à une distance de  $d_L = 100 \text{ Mpc}$ , bien supérieure aux vitesses particulières attendues.

Ces grandes erreurs de mesure font partie des spécificités des mesures de vitesses particulières. Que faire d'un ensemble de mesures chacune compatible avec zéro ? En général, devant un tel problème expérimental, on répète la mesure un nombre suffisant de fois afin d'avoir accès à la distribution postérieure de cette mesure et en déduire une estimation précise de sa moyenne à travers la moyenne d'ensemble sur l'échantillon de mesures. Toutefois, la cosmologie se heurte au problème fondamental suivant : l'Univers n'offre qu'une seule réalisation de cette mesure. C'est pour cette raison fondamentale que la cosmologie moderne s'appuie toujours davantage sur les statistiques bayésiennes afin de modéliser les observations extragalactiques.

### 1.2.5 Cadre bayésien pour l'étude des vitesses particulières

Les mesures de vitesses particulières sont donc peu précises, et leur signal sur bruit est dans le cas général inférieur à 1. La statistique bayésienne offre une possibilité pour tirer la plus grande information possible de ces mesures. Elle s'appuie sur le théorème de Bayes, qui peut s'écrire de façon figurée

$$\mathcal{P}(\text{modèle}|\text{données}) \propto \mathcal{P}(\text{données}|\text{modèle}) \times \mathcal{P}(\text{modèle}) \quad (1.50)$$

Cette équation signifie qu'en modélisant les observations dans le cadre d'un modèle à travers la fonction de vraisemblance  $\mathcal{P}(\text{données}|\text{modèle})$ , et en y ajoutant des informations *a priori* connue  $\mathcal{P}(\text{modèle})$ , on peut estimer la probabilité postérieure du modèle. Cette thèse s'interroge sur la possibilité d'écrire un modèle qui décrirait le champ de vitesse linéaire  $\mathbf{v}$  de l'Univers à partir d'un certain nombre d'observations extragalactiques (décalages vers le rouge, magnitudes, distances, largeur HI, ...). Dans cette section, nous introduisons quelques notions statistiques indispensables à la lecture de cette thèse.

En théorie des probabilités,  $X$  est une variable aléatoire réelle si elle admet une densité de distribution  $f_X : x \rightarrow \mathcal{R}$  telle que son intégrale sur  $\mathcal{R}$  est bornée :

$$\int f_X(X = x)dx = 1 \quad (1.51)$$

Dans toute cette thèse, on note :

$$\mathcal{P}(x) = f_X(X = x), \quad (1.52)$$

en rendant implicite que  $\mathcal{P}$  est une notation commune à des fonctions différentes. Par exemple,  $(x \rightarrow \mathcal{P}(x)) \neq (y \rightarrow \mathcal{P}(y))$ .

On utilisera aussi les probabilités conditionnelles de deux variables aléatoires continues  $X$  et  $Y$  qui sont définies par :

$$\mathcal{P}(x, y) = \mathcal{P}(x|y)\mathcal{P}(y) \quad (1.53)$$

$$= \mathcal{P}(y|x)\mathcal{P}(x). \quad (1.54)$$

Et sa conséquence directe, le théorème de Bayes :

$$\mathcal{P}(x|y) = \frac{\mathcal{P}(y|x)\mathcal{P}(x)}{\mathcal{P}(y)}. \quad (1.55)$$

L'expression de ces probabilités conditionnelles est généralisable aux probabilités conditionnelles de  $N$  variables aléatoires  $\{X_i\}_{i \in [1, N]}$  :

$$\forall j \in [1, N] \quad \mathcal{P}(\{x_i\}_{i \in [1, N]}) = \mathcal{P}(x_j | \{x_i\}_{i \neq j})\mathcal{P}(\{x_i\}_{i \neq j}) \quad (1.56)$$

Enfin, on se servira de l'axiome des probabilités totales :

$$\mathcal{P}(x) = \int \mathcal{P}(x|y)\mathcal{P}(y)dy \quad (1.57)$$

que l'on nomme aussi « marginalisation » de la variable  $X$  par rapport à  $Y$ .

La distribution normale joue un rôle central dans cette thèse et on spécifie ici nos notations. La loi normale de moyenne  $\mu$  et variance  $\sigma^2$  sera notée  $\mathcal{N}(\mu, \sigma^2)$ . Nous adoptons la même notation pour une loi normale multi-dimensionnelle, c'est-à-dire que  $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$  désigne la loi normale multivariée de moyenne  $\boldsymbol{\mu}$  et de matrice de covariance  $\mathbf{C}$ . De la même façon, la loi log-normale sera notée  $\log\text{-}\mathcal{N}(\mu, \sigma^2)$  et la loi normale asymétrique skew- $\mathcal{N}(\mu, \sigma^2)$ . La troncature d'une loi sera dénotée par le suffixe  $\text{T}[x_{\min}, x_{\max}]$ . Par exemple :

$$\mathcal{N}(\mu, \sigma^2) \text{T}[a, b](x) = \begin{cases} \left(\frac{\pi\sigma^2}{2}\right)^{-\frac{1}{2}} \frac{\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)}{\text{erf}\left(\frac{b-\mu}{\sqrt{2}\sigma}\right) - \text{erf}\left(\frac{a-\mu}{\sqrt{2}\sigma}\right)} & \text{si } x \in [a, b], \\ 0 & \text{sinon.} \end{cases} \quad (1.58)$$

Ces définitions étant posées, on revient sur la question de l'estimation du champ de vitesse  $\mathbf{v}$  à partir des mesures de vitesses particulières. Nous l'avons vu, les mesures sont bruitées. Il nous faut donc un *a priori* informatif pour modéliser le champ de vitesse. Cet *a priori* vient des corrélations du champ de vitesse données par le modèle  $\Lambda$ CDM, comme nous allons le voir dans la section suivante.

## 1.3 La cosmologie et les champs gaussiens

### 1.3.1 Le champ de surdensité

L'étude des fluctuations quantiques (Langlois, 2005) durant la période d'inflation montre que les fluctuations primordiales de densité sont pratiquement gaussiennes (on utilisera le terme *normales* de façon équivalente). Ainsi, les observations cosmologiques (CMB, grandes structures, etc.) sont presque idéalement décrites par des champs dont les statistiques sont données par une loi gaussienne multivariée (aussi appelée loi normale multivariée) :

$$\mathcal{P}(\mathbf{f}|\boldsymbol{\mu}, \mathbf{C}) = |2\pi\mathbf{C}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{f} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{f} - \boldsymbol{\mu})\right) \quad (1.59)$$

$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}) \quad (1.60)$$

où  $\boldsymbol{\mu}$  et  $\mathbf{C}$  sont la moyenne et la covariance du champ :

$$\langle \mathbf{f} \rangle = \int_{\mathbf{f}} \mathbf{f} \mathcal{P}(\mathbf{f}|\boldsymbol{\mu}, \mathbf{C}) d\mathbf{f} = \boldsymbol{\mu} \quad (1.61)$$

$$\langle (\mathbf{f} - \boldsymbol{\mu})(\mathbf{f} - \boldsymbol{\mu})^T \rangle = \int_{\mathbf{f}} (\mathbf{f} - \boldsymbol{\mu})(\mathbf{f} - \boldsymbol{\mu})^T \mathcal{P}(\mathbf{f}|\boldsymbol{\mu}, \mathbf{C}) d\mathbf{f} = \mathbf{C} \quad (1.62)$$

En général, le vecteur  $\mathbf{f}$  correspond à un champ échantillonné en différents points de l'espace physique  $\mathbf{r}$ , et on note :

$$f(\mathbf{r}) \sim \mathcal{N}(\mu(\mathbf{r}), \mathbf{C}(\mathbf{r})) \quad (1.63)$$

et

$$\xi(\mathbf{r}_i, \mathbf{r}_j) = \mathbf{C}_{i,j} \quad (1.64)$$

Le principe cosmologique rajoute en plus l'hypothèse d'homogénéité du champ, et donc  $\mu(\mathbf{r}) = \mu$ ,  $\xi(\mathbf{r}_i, \mathbf{r}_j) = \xi(\mathbf{r}_i - \mathbf{r}_j)$ . Dans ces conditions, la transformée de Fourier  $\hat{f}(\mathbf{k})$  (qui sera parfois notée simplement  $f(\mathbf{k})$ ) est aussi un champ gaussien dont la matrice de covariance  $\hat{\mathbf{C}}$  est donnée par :

$$\hat{\mathbf{C}}_{i,j} = (2\pi)^3 \delta_D(\mathbf{k}_i - \mathbf{k}_j) P(k) \quad (1.65)$$

et

$$P(k) = |\hat{\xi}(\mathbf{k})| \quad (1.66)$$

Après cette brève introduction, nous revenons sur la cosmologie en supposant que  $f$  est le champ de surdensité  $\delta$  de matière dans l'Univers. Le modèle  $\Lambda$ CDM décrit dans la section 1.1.3 associé à la théorie linéaire des structures décrite dans la section 1.1.4 prédit que le champ de surdensité est gaussien et que sa statistique est donc donnée par une fonction unique, son spectre de puissance<sup>1</sup>. Celui-ci est déduit à la fois du CMB et de l'observation de la distribution des galaxies dans l'Univers Local (Percival et al., 2007; Planck Collaboration et al., 2016) et on le trace sur le panneau en haut à droite de la figure 1.1.

Le spectre de puissance  $P(k)$  est défini par sa forme  $T(k)$  et son amplitude  $A_s$  :

$$P(k) = A_s k^{n_s} T^2(k) \quad (1.67)$$

---

1. Par définition, la moyenne du champ de surdensité est nulle, et donc  $\boldsymbol{\mu} = \mathbf{0}$

Les paramètres  $A_s$  et  $n_s$  sont déduit du CMB, et la fonction de transfert de la résolution d'équations de Boltzmann pour les différents composants énergétiques de l'Univers (Eisenstein and Hu, 1999). Historiquement, un autre paramètre équivalent à la normalisation  $A_s$  est la variance du champ de surdensité  $\sigma_8^2$  dans une sphère de 8 Mpc de rayon :

$$\sigma_8^2 = \frac{1}{2\pi^2} \int W_{sth}(kR)P(k)dk \quad (1.68)$$

où  $R = 8$  Mpc et  $W_{sth}$  est le filtre défini par :

$$W_{sth}(x) = \frac{3j_1(x)}{x} \quad (1.69)$$

$$= \frac{3 \left( \frac{\sin x}{x^2} - \frac{\cos x}{x} \right)}{x} \quad (1.70)$$

La mesure de  $\sigma_8$  par la mission Planck (Planck Collaboration et al., 2016) donne  $\sigma_8 = 0.8158 \pm 0.0086$ .

### 1.3.2 Corrélations des vitesses particulières

L'équation 1.33 montre la relation linéaire entre le champ de vitesse linéaire, que l'on notera  $\mathbf{v}^{\text{lin}}$ , et le champ de surdensité. Corrélairement,  $\mathbf{v}^{\text{lin}}$  est également un champ gaussien dont les caractéristiques peuvent être écrites dans l'espace de Fourier. La transformée de Fourier de l'équation 1.33 est

$$\mathbf{v}^{\text{lin}}(\mathbf{k}) = i \frac{\mathbf{k}}{k^2} H f \delta(\mathbf{k}) \quad (1.71)$$

et donc le tenseur (3,3) donnant les corrélations du champ de vitesse est donné par

$$\langle \mathbf{v}^{\text{lin}}(\mathbf{k}' + \mathbf{k}) \mathbf{v}^{\text{lin}}(\mathbf{k}') \rangle = (Hf)^2 \frac{\mathbf{k}\mathbf{k}^T}{k^4} P(k) \quad (1.72)$$

Dans l'espace réel, on note ce tenseur  $\Psi$  :

$$\Psi(\mathbf{r}) = \frac{(Hf)^2}{(2\pi)^3} \int \frac{\mathbf{k}\mathbf{k}^T}{k^4} P(k) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k} \quad (1.73)$$

On peut aussi calculer la corrélation  $\zeta$  entre le champ  $\delta$  et le champ de vitesse linéaire  $\mathbf{v}^{\text{lin}}$  :

$$\zeta(\mathbf{r}) = \frac{Hf}{(2\pi)^3} \int \frac{i\mathbf{k}}{k^2} P(k) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k} \quad (1.74)$$

Notons que la variance du champ de vitesse  $\Lambda$ CDM est facilement calculable à partir de l'équation (1.73) :

$$\sigma_v^2 = \frac{(Hf)^2}{2\pi^2} \int P(k) dk \quad (1.75)$$

$$= \Psi(0) \quad (1.76)$$

$$\sim (300 \text{ km/s})^2 \quad (1.77)$$

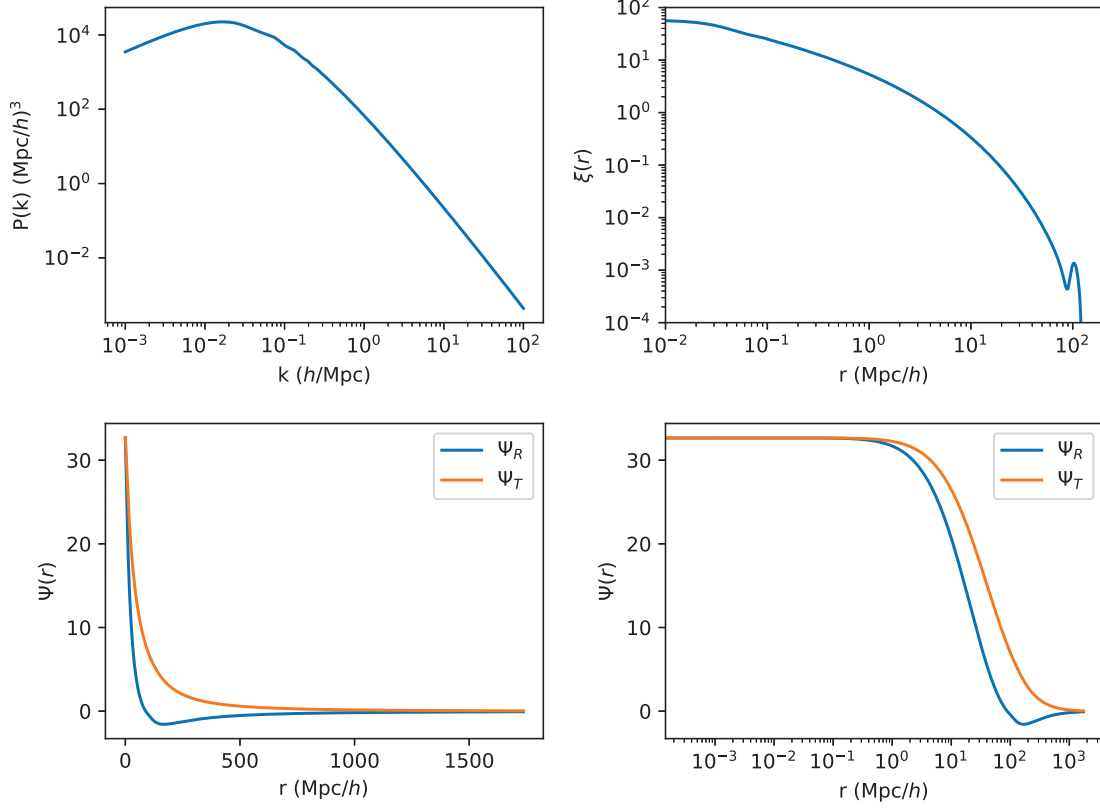


FIGURE 1.1 – En haut à droite : Spectre de puissance pour un Univers plat et  $\Omega_\Lambda = 0.7$ . En haut à gauche : fonction de corrélation associée. On peut noter le pic BAO à  $r \sim 100 \text{Mpc}/h$ . Bas : fonctions de corrélations  $\Psi_{R,T}$  radiale et transverse calculées à partir de l'équation (1.73).

On trace sur la figure 1.1 les fonctions  $P(k)$ ,  $\xi$  et  $\Psi_{R,T}$ , ces dernières étant définies par :

$$\Psi_R = \frac{1}{2\pi^2} \int \left( j_0(kx) - \frac{2j_1(kx)}{kx} \right) P(k) dk \quad (1.78)$$

$$\Psi_T = \frac{1}{2\pi^2} \int \frac{j_1(kx)}{kx} P(k) dk \quad (1.79)$$

$$\Psi_{\alpha,\beta} = (Hf)^2 (\Psi_T \delta_{\alpha,\beta}^K + (\Psi_R - \Psi_T) \hat{x}_\alpha \hat{x}_\beta) \quad (1.80)$$

Dans cette thèse, les corrélations  $\Psi$  et  $\zeta$  sont calculées par intégration numérique directe des équations (1.73) et (1.74). Les corrélations sont ensuite évaluées en n'importe quelle position  $\mathbf{r}$  en interpolant linéairement la fonction ainsi intégrée.

L'utilisation de ces corrélations de vitesse pour estimer le champ de vitesse sous-jacent n'est pas une idée nouvelle. Parmi les techniques existantes on peut noter les techniques VELMOD (Willick and Strauss, 1998) et POTENT (Dekel et al., 1999), les analyses de



2MTF (Springob et al., 2016) ou encore la technique du filtre de Wiener (Zaroubi et al., 1995). Ces méthodes partagent de nombreux points communs et nous choisissons d'axer notre analyse des biais statistiques et observationnels sur le filtre de Wiener, outil principal de la collaboration *CosmicFlows*. Le prochain chapitre tente de démontrer la nécessité d'évoluer vers un modèle complètement bayésien pour prendre en compte les erreurs sur les mesures de vitesses particulières sans introduire d'erreurs systématiques dans l'analyse.

## Chapitre 2

# Filtre de Wiener et réalisations contraintes

Cette thèse s’inscrit dans le cadre des collaborations *CosmicFlows* (Tully et al., 2008; Courtois et al., 2011) et CLUES (Yepes et al., 2009; Gottloeber et al., 2010; Courtois and Tully, 2012; Sorce et al., 2016) (Simulations contraintes de l’Univers Local) dont les objectifs sont d’étudier la structure de l’Univers Local à travers l’étude des vitesses particulières des galaxies. *CosmicFlows* d’abord, en observant, compilant et analysant les données de vitesses apparentes et de modules de distance; CLUES ensuite qui, à partir de cette analyse, simule l’Univers Local et étudie son histoire de formation. Notre analyse des vitesses particulières se base sur le catalogue *CosmicFlows-3* (CF3 ci-après) qui est un regroupement cohérent d’environ 18,000 distances et décalages vers le rouge. Nous avons vu dans le chapitre précédent comment, à partir de ces mesures, on peut en déduire les vitesses particulières radiales des galaxies. Nous détaillons dans ce chapitre la méthode d’analyse utilisée dans la collaboration au commencement de cette thèse, à savoir la technique du filtre de Wiener et réalisations contraintes (WF/CR), dans la section 2.1. Si cette technique a permis des avancées en termes de cosmographie, comme par exemple la découverte du superamas Laniakea (Tully et al., 2014), la taille et la profondeur du catalogue CF3 ont fait échouer toutes les tentatives d’analyse utilisant cette méthode. Dans la section 2.2 nous construisons un catalogue simulé afin d’analyser les systématiques de l’analyse via le filtre de Wiener et souligner les limites de la méthode dans la section 2.3. On y montre en particulier le rôle central des distances mesurées dans l’analyse. De ces distances dépend en effet l’estimation des vitesses particulières, des fonctions de corrélation et de la position des galaxies dans le volume de reconstruction. L’incertitude sur ces distances, qui n’est pas prise en compte dans la méthode du filtre de Wiener, biaise ainsi le champ de vitesse reconstruit. Cette étude préliminaire motive le développement d’un modèle plus complet qui sera décrit dans le chapitre 3 et appliqué au catalogue *CosmicFlows-3* dans le chapitre 4.

### Sommaire

---

<b>2.1</b>	<b>Technique des réalisations contraintes . . . . .</b>	<b>28</b>
2.1.1	Approche fréquentiste . . . . .	28

2.1.2	Approche probabiliste . . . . .	30
<b>2.2</b>	<b>Données simulées . . . . .</b>	<b>31</b>
2.2.1	Génération d'un catalogue simulé . . . . .	31
2.2.2	Application de la méthode WF/CR . . . . .	32
2.2.3	Cas des vitesses particulières mesurées . . . . .	39
<b>2.3</b>	<b>Biais statistiques . . . . .</b>	<b>42</b>
2.3.1	Biais de densité . . . . .	44
2.3.2	Biais de sélection . . . . .	44
2.3.3	Biais de positionnement . . . . .	46
2.3.4	Distribution log-normale des vitesses particulières . . . . .	47
<b>2.4</b>	<b>Reconstructions . . . . .</b>	<b>49</b>
2.4.1	Solutions partielles . . . . .	50
<b>2.5</b>	<b>Conclusion sur la méthode WF/CR . . . . .</b>	<b>52</b>

---

## 2.1 Technique des réalisations contraintes

Les données *CosmicFlows* ont systématiquement été analysées via la technique dite des réalisations contraintes développée par Hoffman and Ribak (1991, 1992). Dans cette section, nous présentons les aspects théoriques de la méthode dans le cas d'une reconstruction du champ de vitesse sur un nombre fini de points. Le formalisme généralisé aux espaces de dimensions infinies peut être trouvé dans van de Weygaert and Bertschinger (1996).

### 2.1.1 Approche fréquentiste

Afin d'analyser les observations de vitesses particulières radiales dans le cadre du modèle  $\Lambda$ CDM, la technique des réalisations contraintes a été développée et utilisée par Hoffman and Ribak (1991, 1992). Nous dérivons ici les principales équations de la technique dans un cadre de traitement du signal. Soit  $\mathbf{c}$  le résultat de l'addition d'un signal  $\mathbf{s}$  et d'un bruit gaussien  $\mathbf{n}$  :

$$\mathbf{c} = \mathbf{s} + \mathbf{n}. \quad (2.1)$$

On suppose en plus que le signal  $\mathbf{s}$  est une fonctionnelle du champ recherché, champ que l'on note  $\delta$  en anticipant le fait qu'il s'agira du champ de surdensité :

$$\mathbf{s} = \mathbf{R}\delta \quad (2.2)$$

où  $\mathbf{R}$  est la matrice de réponse. Par exemple, si  $s$  est la vitesse particulière radiale, on peut utiliser l'équation (1.71) pour exprimer  $R$  comme :

$$\mathbf{R} = \mathbf{DFT} \times \hat{\mathbf{R}}, \quad [\hat{\mathbf{R}}]_{i,j} = iH_0 f \frac{\mathbf{k}_i \cdot \hat{\mathbf{k}}_j^r}{k_i^2} \quad (2.3)$$

où  $\mathbf{DFT}$  est la transformée de Fourier discrète. On cherche à filtrer linéairement les données avec un filtre  $\mathbf{F}$  tel que :

$$\delta^F = \mathbf{F}\mathbf{c}. \quad (2.4)$$

Le résidu entre le champ estimé et le champ réel est noté  $\mathbf{D}$  :

$$\mathbf{D} = \boldsymbol{\delta} - \boldsymbol{\delta}^F \quad (2.5)$$

et sa variance :

$$\langle \mathbf{D}\mathbf{D}^T \rangle_{ij} = \langle (\boldsymbol{\delta} - \boldsymbol{\delta}^F)(\boldsymbol{\delta} - \boldsymbol{\delta}^F)^T \rangle_{ij}. \quad (2.6)$$

On peut montrer (Hoffman and Ribak (1991)) que le filtre minimisant la variance des résidus ci-dessus est le filtre de Wiener  $\mathbf{WF}$  :

$$\mathbf{WF} = \mathbf{C}_{\delta c} \mathbf{C}_{cc}^{-1} \quad (2.7)$$

où  $\mathbf{C}_{\delta c}$  est la matrice de corrélation entre le champ et les contraintes, et  $\mathbf{C}_{cc}$  est la matrice d'autocorrélation des contraintes. Elles dérivent de la matrice  $\mathbf{R}$  est de la matrice de covariance du champ :

$$\mathbf{C}_{\delta c} = \langle \boldsymbol{\delta} \boldsymbol{\delta}^T \rangle \mathbf{R} \quad (2.8)$$

$$\mathbf{C}_{cc} = \mathbf{R} \langle \boldsymbol{\delta} \boldsymbol{\delta}^T \rangle \mathbf{R}^T + \mathbf{N} \quad (2.9)$$

où  $\mathbf{N} = \langle \mathbf{n}\mathbf{n}^T \rangle$ , et où on a supposé que le champ et les erreurs ne sont pas corrélés.

S'ensuit l'estimation :

$$\boldsymbol{\delta}^{WF} = \mathbf{C}_{\delta c} \mathbf{C}_{cc}^{-1} \mathbf{c}. \quad (2.10)$$

En réinjectant dans l'équation 2.6, on obtient les résidus :

$$\langle \mathbf{D}\mathbf{D}^T \rangle_{ij} = \mathbf{C}_{\delta\delta} - \mathbf{C}_{\delta c} \mathbf{C}_{cc}^{-1} \mathbf{C}_{\delta c}. \quad (2.11)$$

Hoffman and Ribak (1992) remarquent alors que ce résidu ne dépend pas explicitement de la valeur des contraintes  $\mathbf{c}$ <sup>1</sup>. En particulier, il est tout à fait possible de générer un champ  $\boldsymbol{\delta}^{CR}$  (on notera  $CR$  pour *constrained realization*) dont le résidu  $\mathbf{D}$  respecte l'équation 2.11, de la façon suivante : à partir d'une réalisation aléatoire  $\boldsymbol{\delta}^{RR}$  (on notera  $RR$  pour *Random Realization*) du champ générée à partir de sa matrice de covariance  $\mathbf{C}_{\delta\delta}$ , on génère des contraintes  $\mathbf{c}^{RR}$  (par exemple, les vitesses radiales du champ) en tirant la contrainte du champ aléatoire aux mêmes positions que les contraintes observées  $\mathbf{c}$ . La réalisation contrainte définie par

$$\boldsymbol{\delta}^{CR} = \boldsymbol{\delta}^{RR} + \mathbf{C}_{\delta c} \mathbf{C}_{cc}^{-1} (\mathbf{c} - \mathbf{c}^{RR}) \quad (2.12)$$

produit donc une réalisation du champ  $\boldsymbol{\delta}$  sous les contraintes  $\mathbf{c}$ . En particulier, puisque  $\langle \boldsymbol{\delta}^{RR} \rangle = \mathbf{0}$ ,  $\langle \boldsymbol{\delta}^{CR} \rangle = \boldsymbol{\delta}^{WF}$ ; et la variance des  $\mathbf{D} = \boldsymbol{\delta}^{CR} - \boldsymbol{\delta}^{WF}$  est bien donnée par l'équation 2.11. L'équation (2.12) permet ainsi d'échantillonner le champ  $\boldsymbol{\delta}$ . En effet, en construisant un ensemble de réalisations contraintes  $\boldsymbol{\delta}^{CR}$ , on aura accès à la distribution postérieure du champ  $\boldsymbol{\delta}$ , de moyenne  $\boldsymbol{\delta}^{WF}$ . Cet algorithme est connu sous le nom d'algorithme d'Hoffman-Ribak.

Incarnons notre propos en considérant que l'on veuille reconstruire le champ de surdensité  $\boldsymbol{\delta}$  à partir de l'observation des vitesses particulières radiales des galaxies. Dans ce cas là,

1. Nous verrons cependant plus tard que dans le cas où les contraintes sont les vitesses particulières radiales, la matrice  $\mathbf{C}_{\delta c}$  dépend de la distance entre les traceurs, qui est corrélée avec ces mêmes vitesses.

$\mathbf{c} = \mathbf{v}_\delta(\mathbf{r}) \cdot \hat{\mathbf{r}}$  où  $\mathbf{v}_\delta$  est le champ de vitesse associé au champ de surdensité et défini par l'équation (1.33), et  $\mathbf{r}$  est le vecteur position des galaxies. La variance du champ  $\mathbf{C}_{\delta\delta}$  est donnée par la transformée de Fourier du spectre de puissance  $P(k)$  (équation (1.65)); la covariance entre les contraintes et le champ  $\mathbf{C}_{\delta c}$  est donnée par l'équation (1.74); et la matrice de covariance des données  $\mathbf{C}_{cc}$  est donnée par la fonction de corrélation  $\Psi$  définie par l'équation (1.73) à laquelle on ajoute la matrice d'erreur, ici diagonale :

$$[\mathbf{C}_{cc}]_{ij} = \langle v_i^r v_j^r \rangle + \langle \epsilon_i \epsilon_j \rangle \delta_{ij} \quad (2.13)$$

où  $\epsilon$  est l'erreur sur la vitesse particulière. En général, les vitesses particulières sont mesurées à  $z = 0$  où le champ de vitesse n'est plus linéaire. Pour modéliser l'écart à la non linéarité (nous tentons ici de reconstruire le champ de vitesse linéaire uniquement), on introduit le paramètre  $\sigma_{\text{NL}}$  qu'on ajoute à la diagonale de la matrice de covariance :

$$[\mathbf{C}_{cc}]_{ij} = \langle v_i^r v_j^r \rangle + \langle \epsilon_i \epsilon_j \rangle \delta_{ij} + \sigma_{\text{NL}}^2 \delta_{ij} \quad (2.14)$$

Notons dès à présent que toutes ces corrélations dépendent de la position des traceurs. Ainsi, la position des traceurs joue à la fois sur l'estimation de la vitesse radiale  $v^r$  mais aussi sur l'estimation des corrélations  $\mathbf{C}_{cc}$  et  $\mathbf{C}_{\delta c}$ .

On peut également directement échantillonner le champ de vitesse  $\mathbf{v}^{\text{lin}}$  plutôt que le champ de surdensité. L'équation (2.12) devient alors :

$$\mathbf{v}_\alpha^{\text{lin},CR} = \mathbf{v}_\alpha^{\text{lin},RR} + \mathbf{C}_{v_\alpha c} \mathbf{C}_{cc}^{-1} (\mathbf{c} - \mathbf{c}_{RR}) \quad (2.15)$$

pour chaque composante cartésienne  $\alpha$  du champ de vitesse. La corrélation  $\mathbf{C}_{v_\alpha c}$  est alors donnée par la fonction  $\zeta$  définie par l'équation (1.74) appliquée à toutes les paires de contraintes. L'avantage d'échantillonner directement le champ de vitesse plutôt que le champ de surdensité permet d'avoir une estimation de celui-ci sans l'utilisation de la transformée de Fourier pour intégrer l'équation (1.71).

### 2.1.2 Approche probabiliste

L'autre approche pour dériver la technique WF/CR est probabiliste. Elle est en tout point équivalente à la dérivation faite ci-dessus, cependant elle est plus aisément généralisable à un champ de dimension infinie (voir van de Weygaert and Bertschinger (1996)). Formellement, les contraintes  $\mathbf{c}$  définies ci-dessus (les vitesses particulières radiales dans notre cas, mais elles peuvent être définies comme n'importe quelle fonctionnelle du champ  $\delta$ ) sont tirées d'une distribution de moyenne nulle et dont la covariance est notée  $\mathbf{C}_{cc}$  et qui sont corrélées au champ de densité via la corrélation  $\mathbf{C}_{\delta c}$ . Le champ  $\delta$  est aussi un champ gaussien de moyenne nulle et de covariance  $\mathbf{C}_{\delta\delta}$  (correspondant à la transformée de Fourier du spectre de puissance). Si ce champ est de dimension finie, on peut formellement décrire le modèle de la façon suivante :

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}) \quad (2.16)$$

où  $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{c} \end{pmatrix}$  et  $\mathbf{C} = \begin{pmatrix} \mathbf{C}_{\delta\delta} & \mathbf{C}_{\delta c} \\ \mathbf{C}_{\delta c} & \mathbf{C}_{cc} \end{pmatrix}$ . Il est possible de montrer qu'il s'ensuit que la probabilité conditionnelle de  $\boldsymbol{\delta}|\boldsymbol{c}$  est une loi normale multivariée de moyenne et d'écart-type :

$$\boldsymbol{\mu}_{\delta|\boldsymbol{c}} = \mathbf{C}_{\delta c} \mathbf{C}_{cc}^{-1} \boldsymbol{c}, \quad (2.17)$$

$$\mathbf{C}_{\delta|\boldsymbol{c}} = \mathbf{C}_{\delta\delta} - \mathbf{C}_{\delta c} \mathbf{C}_{cc}^{-1} \mathbf{C}_{\delta c}. \quad (2.18)$$

On reconnaît la moyenne donnée par le filtre de Wiener de l'équation (2.10) et la variance donnée par l'équation (2.11). Notons au passage que la probabilité marginale de  $\boldsymbol{c}$  est :

$$\boldsymbol{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{cc}). \quad (2.19)$$

Cette équation nous servira par la suite.

Afin d'analyser l'efficacité de la méthode WF/CR présentée ci-dessus pour reconstruire les champs de vitesse, nous construisons dans la prochaine section un catalogue de données simulées et appliquons l'équation 2.15 au catalogue dans diverses configurations.

## 2.2 Données simulées

### 2.2.1 Génération d'un catalogue simulé

Afin de tester la méthode WF/CR, nous construisons un catalogue de données simulées comprenant des décalages vers le rouge et modules de distance comme il est nécessaire d'avoir pour étudier les vitesses particulières. Le catalogue sera une observation de  $N = 3000$  points dans une réalisation aléatoire générée à partir de son spectre de puissance  $P(k)$  défini dans la section 1.3.1. La RR (*Random Realization*) est générée sur une grille de taille  $(512)^3$  représentant une boîte de taille  $(1000 \text{ Mpc})^3$ . Cette taille est suffisante pour limiter les effets des conditions aux limites périodiques dans la zone centrale (sphère de 200 Mpc de rayon) de la boîte. La taille de la grille de  $512^3$  est quant à elle suffisante pour décrire le champ de vitesse à une échelle linéaire d'environ 4 Mpc. Elle est également générée avec un spectre de puissance tronqué à  $k_{\max} = 0.07 \text{ Mpc}^{-1}$  afin de s'affranchir de tout effet non linéaire dû aux modes d'ordre supérieur. La Figure 2.1 montre les champs de vitesse et de surdensité utilisés pour générer les données simulées. À partir de cette RR, on génère le catalogue de la façon suivante :

- des positions angulaires sont générées uniformément à partir de la distribution de coordonnées supergalactiques suivante :

$$\cos\left(\frac{\pi}{2} - \text{SGB}\right) \sim \text{Uniform}(0, 1), \quad (2.20)$$

$$\text{SGL} \sim \text{Uniform}(0, 2\pi); \quad (2.21)$$

- des distances lumineuses sont générées à partir de la distribution :

$$r \sim \left(\frac{r}{1 \text{ Mpc}}\right)^3 \exp - \left(\frac{r}{80 \text{ Mpc}}\right)^2 \quad (2.22)$$

qui permet de simuler l'effet de volume localement, et les effets de sélection à plus grande distance ;

- les vitesses particulières de chaque point du catalogue sont calculées en interpolant trilinéairement le champ de vitesse généré aux points de données. On ajoute également une dispersion gaussienne de déviation standard  $\sigma_{\text{NL}}$  aux vitesses particulières pour simuler l’effet des non-linéarités ;
- les décalages vers le rouge sont calculés à partir des vitesses particulières et des distances lumineuses en utilisant l’équation (1.47).

Il résulte un catalogue de 3000 points dans lequel on a les positions et décalages vers le rouge de chacun des points. À partir de ce catalogue, on teste la méthode WF/CR afin d’évaluer si le champ initial est correctement retrouvé.

### 2.2.2 Application de la méthode WF/CR

Dans cette section, nous appliquons la méthode WF/CR au catalogue simulé. Ces données simulées sont idéales dans le sens où elles reflètent exactement le champ de vitesse sous-jacent puisqu’aucune erreur observationnelle n’a été introduite. Dans la suite, nous introduisons des erreurs observationnelles et étudions comment elles affectent la reconstruction. Le volume de reconstruction est pris de taille  $L = 500$  Mpc afin que toutes les données soient incluses dans ce volume, et la taille de la grille  $M = 64^3$ , ce qui correspond à une résolution de  $8 \text{ Mpc} < l = 14 \text{ Mpc}$  où  $l$  est la distance moyenne entre deux traceurs. La reconstruction est faite en calculant 100 réalisations contraintes à partir de l’équation 2.12, et le champ moyen est calculé à partir de la moyenne d’ensemble de ces 100 réalisations.

La Figure 2.2 montre une reconstruction idéale (*i.e.* sans aucune erreur observationnelle) via l’algorithme des réalisations contraintes. Le panneau à gauche correspond à la tranche  $z = 0$  Mpc de la valeur moyenne des réalisations contraintes. La comparaison des Figures 2.2 et 2.1 permet d’évaluer qualitativement la qualité de la reconstruction. On peut remarquer que la plupart des structures sont retrouvées, même au delà de la limite des données figurée par le cercle pointillé noir<sup>2</sup>. Si l’on retrouve du signal au delà de la zone de données, c’est dû au fait que les vitesses corrélient à grande distance, et que la vitesse radiale des traceurs au centre est encore influencée par le champ de surdensité lointain. Puisque la technique WF/CR permet d’échantillonner la probabilité postérieure du champ de vitesse, la déviation standard de l’ensemble des réalisations contraintes permet d’avoir une estimation de la précision de la reconstruction. On peut ainsi estimer l’erreur sur les trois composantes cartésiennes du champ de vitesse (tridimensionnel, rappelons-le) reconstruit. Cependant, nous représenterons souvent l’erreur  $\sigma_{v^r}$  sur la composante radiale  $v^r = \mathbf{v} \cdot \mathbf{r}$  de ce champ puisque d’une part, la représentation d’une quantité scalaire est plus aisée, et d’autre part parce que les contraintes de la reconstruction sont les vitesses radiales et c’est donc cette composante qui est la plus contrainte et donc la moins bruitée. Il est ainsi plus aisé de mettre en évidence un biais statistique en étudiant la distribution des vitesses radiales reconstruites.

**Précision** Le panneau de droite de la figure 2.2 montre donc la déviation standard des vitesses radiales reconstruites dans la tranche  $z = 0$ . On peut remarquer que la dispersion est d’environ  $75 \text{ km s}^{-1}$  sur le volume contenant des données, et que cette dispersion tend vers

---

2. Attention, l’échelle des couleurs pour montrer la surdensité n’est pas la même entre la figure 2.2 et 2.1



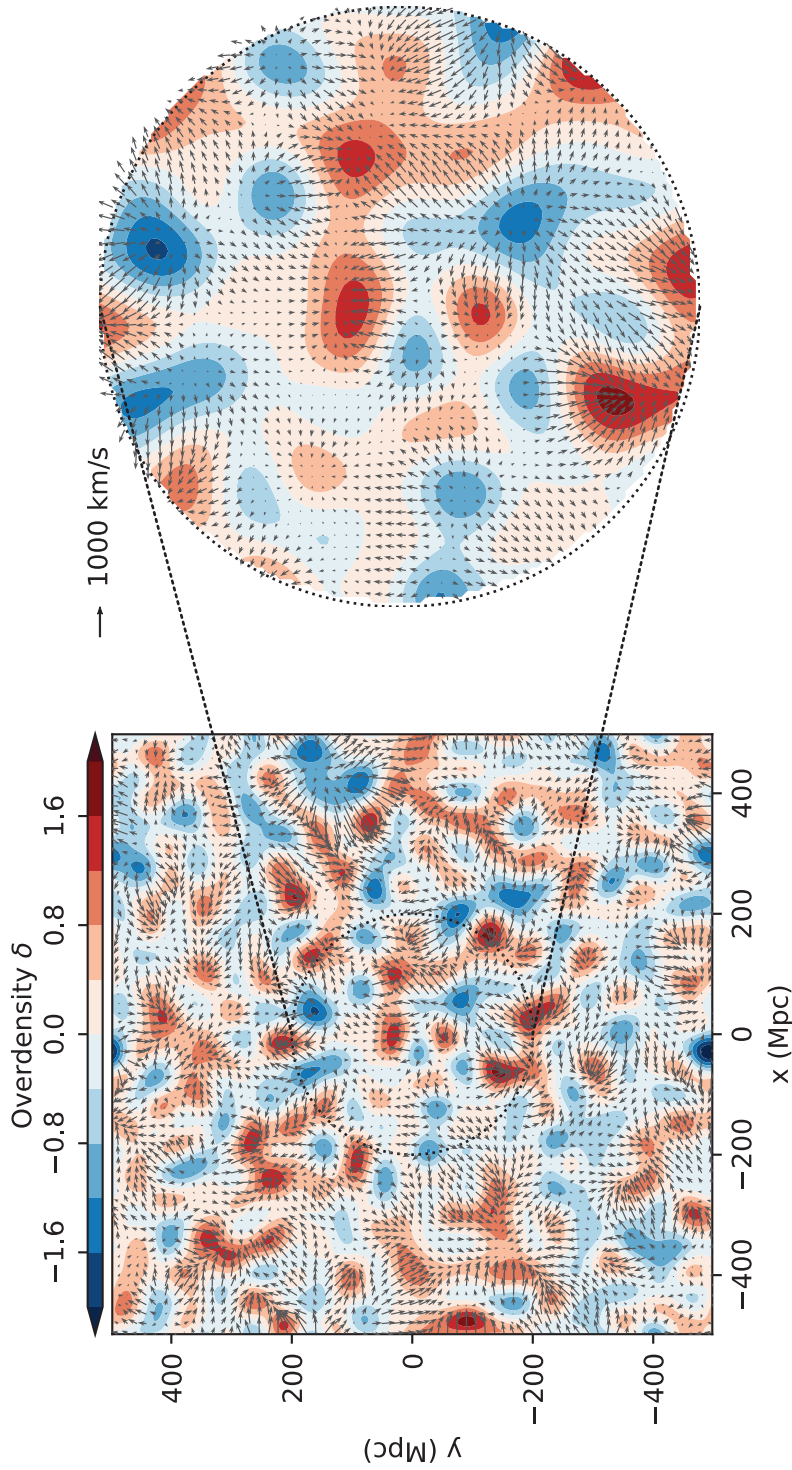


FIGURE 2.1 – Tranche en  $z = 0$  Mpc de la réalisation contrainte utilisée pour générer le premier catalogue de données simulées. La couleur correspond au champ de surdensité (en bleu, les régions sous-denses, ou vides; en rouge, les régions surdenses, ou pleines). Le cercle en pointillé symbolise la limite en distance des données. À droite figure un zoom sur la partie centrale. Les flèches noires représentent la projection des vitesses particulières simulées dans le plan  $z = 0$ . On voit notamment que les régions surdenses (resp. sous-denses) correspondent bien aux points d'attraction (resp. repulsion) du champ de vitesse. À partir de cette réalisation aléatoire,  $N = 3000$  points sont tirés uniformément en distance dans le volume afin de simuler un catalogue d'observations de vitesses particulières.



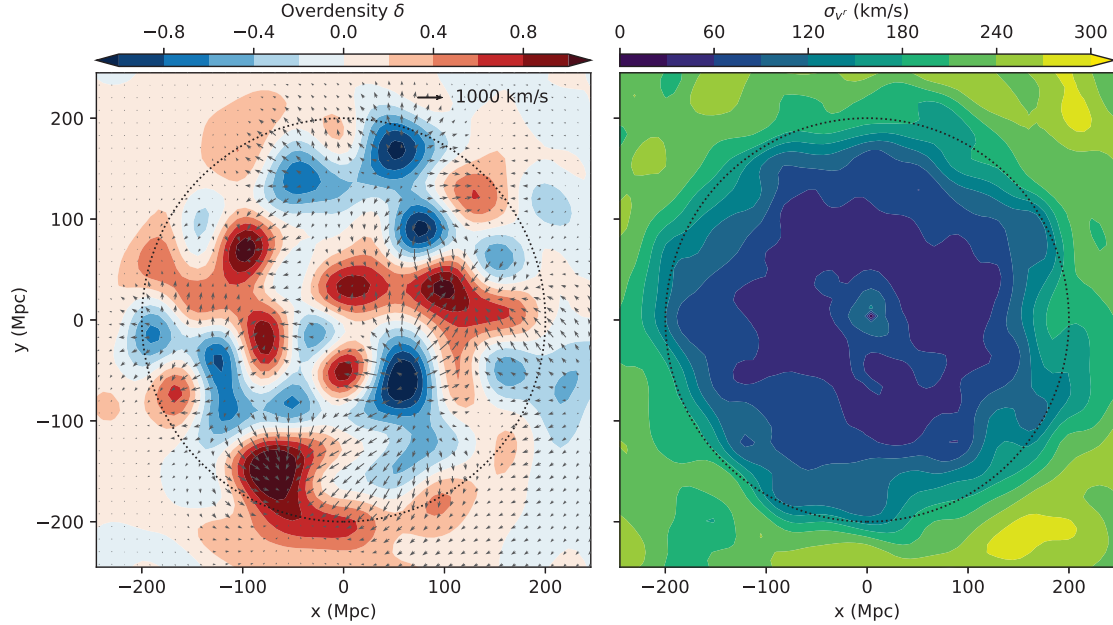


FIGURE 2.2 – À gauche : tranche en  $z = 0$  du champ de surdensité reconstruit (dans le cadre du catalogue simulé) à partir de la moyenne de 100 réalisations contraintes dans le cas idéal, c'est-à-dire sans erreurs observationnelles. À droite : déviation standard des vitesses radiales reconstruites. La déviation standard montre la précision des vitesses radiales reconstruites. La valeur  $\sigma_{vr} = 300 \text{ km s}^{-1}$  correspond à la déviation standard prédite par le modèle  $\Lambda\text{CDM}$ . Les valeurs moindres au centre montrent que le champ a bien été contraint par les vitesses radiales observationnelles. Le cercle pointillé noir montre la limite des données simulées  $r = 200 \text{ Mpc}$ .

$300 \text{ km s}^{-1}$  à grande distance, déviation standard prédite par le modèle  $\Lambda\text{CDM}$  comme décrit par les équations (1.68) et (1.75). On peut également comparer la précision de la reconstruction du champ des vitesses radiales à celle du champ de surdensité. La Figure 2.3 montre la tranche  $z = 0$  de la déviation standard pour les deux champs. Les limites de couleur (surdensité 0.6 et vitesses  $300 \text{ km s}^{-1}$ ) sont données par la déviation standard du modèle  $\Lambda\text{CDM}$ . On peut remarquer que le champ de vitesse radiale (puisque c'est lui qui est le plus contraint) tend moins vite vers cette variance que le champ de surdensité, confirmant notre remarque précédente sur les composantes radiales du champ de vitesse. En général, le champ de surdensité est reconstruit avec plus de bruit que le champ de vitesse radiale, et nous montrerons pour les prochaines reconstructions les déviations standards sur le champ de vitesses radiales uniquement, comme dans la Figure 2.2.

**Justesse** Les deux Figures précédentes donnent une représentation du champ moyen reconstruit ainsi que la précision sur celui-ci. Cependant, ce qui nous intéresse dans ce chapitre est la *justesse*, ou *fidélité*, de la reconstruction par rapport à la simulation initiale, indépendamment de sa précision. Nous introduisons ici des outils et des représentations pour estimer cette jus-

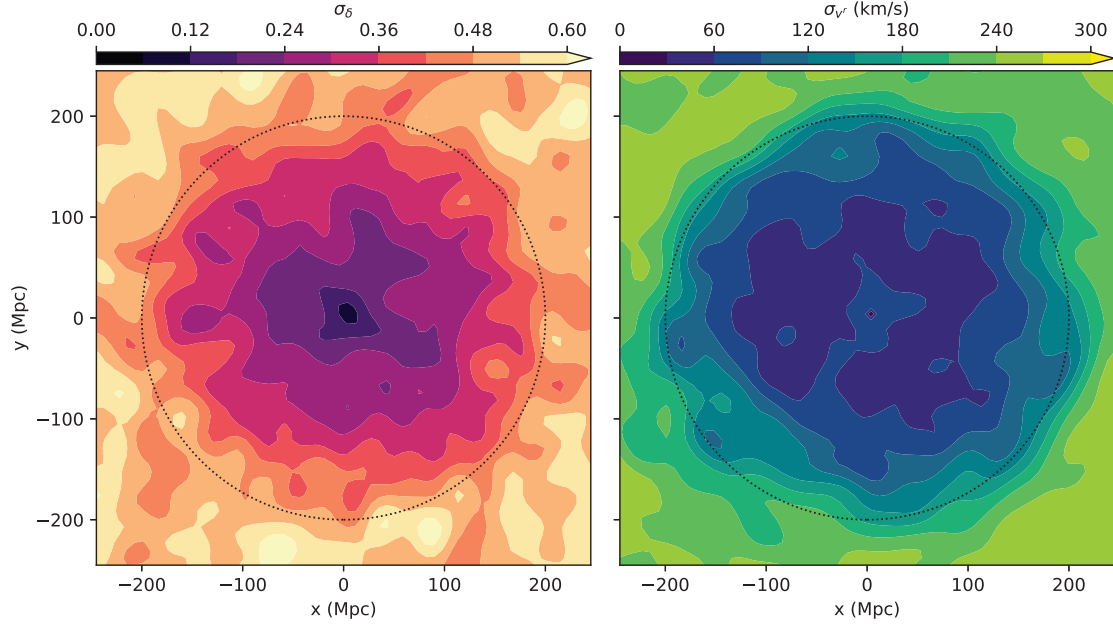


FIGURE 2.3 – À gauche : déviation standard du champ de surdensité  $\sigma_\delta$  reconstruit dans le cadre de la reconstruction de la réalisation aléatoire et le catalogue simulé associé idéal (sans l’introduction d’erreurs observationnelles). À droite : identique au panneau de gauche mais pour le champ de vitesse radiale  $\sigma_{v_r}$ . Le cercle noir montre la limite des données simulées à  $R = 200$  Mpc. Ces deux panneaux montrent la précision de la reconstruction. Les valeurs extrêmes  $\sigma_\delta = 0.6$  et  $\sigma_{v_r} = 300 \text{ km s}^{-1}$  correspondent aux déviations standards prédites par le modèle  $\Lambda$ CDM.

tesse. On introduit d’abord les résidus normalisés (aussi appelés « pull »)  $p$  du champ  $f$  défini comme

$$p = \frac{\bar{f} - f^{\text{simulation}}}{\sigma_f}. \quad (2.23)$$

Cette quantité donne l’écart entre la reconstruction moyenne et la simulation initiale, en normalisant par les erreurs de la reconstruction. Ainsi, si en un point de la grille le résidu vaut par exemple  $p = 3$ , alors la valeur reconstruite sera compatible avec le champ initial à  $3\sigma$ . La Figure 2.4 montre la valeur de ces résidus pour le champ de surdensité  $\delta$  dans la tranche  $z = 0$  Mpc (à gauche) et l’histogramme de ces résidus sur l’ensemble du volume de reconstruction (à droite). On peut observer que la reconstruction n’exhibe aucune structure à  $\pm 3\sigma$  et que la distribution des résidus est très proche d’une loi gaussienne centrée et unitaire (notée  $\mathcal{N}(0, 1)$  sur la figure). En effet, la moyenne de la distribution est de  $m = -0.01$  et la déviation standard  $s = 0.98$ . Le fait que la distribution du *pull* coïncide avec la loi normale est une estimation de la justesse de la reconstruction. En effet, cela signifie qu’en moyenne les surdensités reconstruites sont les mêmes que dans la simulation de départ (c’est le caractère centré de la distribution des résidus) et que les écarts à cette moyenne sont bien modélisés par

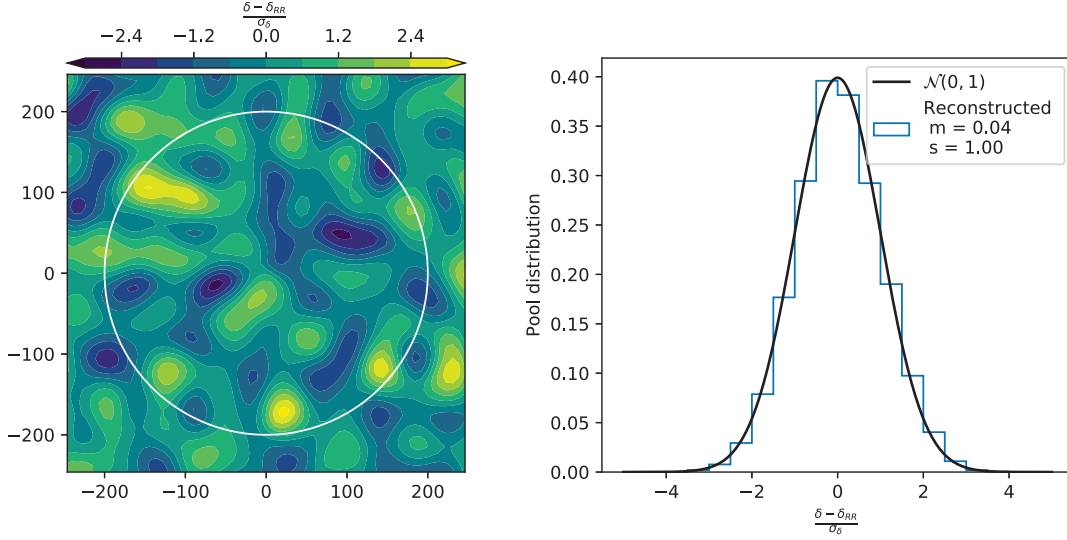


FIGURE 2.4 – Gauche : tranche en  $z = 0$  Mpc des résidus  $\frac{\bar{\delta} - \delta_{RR}}{\sigma_{\delta}}$  du champ de densité reconstruit par rapport au champ initial simulé  $\delta_{RR}$ , dans le cas du catalogue simulé optimal, c'est-à-dire sans avoir introduit d'erreurs observationnelles. Le cercle blanc symbolise la limite des données à  $R = 200$  Mpc. Droite : Histogramme des résidus normalisés pour tous les points de la grille à une distance inférieure à  $R = 200$  Mpc. La courbe noire montre la distribution normale unitaire  $\mathcal{N}(0, 1)$ .

les erreurs sur la reconstruction (c'est la largeur de la distribution). La Figure 2.5 est identique à la Figure 2.4 pour les vitesses radiales reconstruites, *i.e.*  $f = v^r$ . Là encore la distribution du *pull* est très proche de la loi normale, ce qui suggère une reconstruction juste et des erreurs correctement estimées.

Si les résidus donnent une bonne idée de la justesse de la reconstruction, l'ensemble des réalisations contraintes calculées permet en fait de remonter à un bon échantillonnage de la distribution de probabilité postérieure du champ. Par exemple, au point  $(x, y, z) = (250, 250, 250)$  Mpc, la distribution des vitesses radiales reconstruites est montrée sur la Figure 2.6 sur laquelle la valeur dans la simulation initiale  $v^r = -98 \text{ km s}^{-1}$  est figurée par le trait noir. Graphiquement, il est difficile de tracer l'ensemble des distributions postérieures pour chaque point de la grille de reconstruction. Cependant, on peut sélectionner certaines régions et montrer la somme des distributions de probabilité correspondantes comme proposé par Lavaux (2016). Par exemple, la figure 2.7 montre les distributions postérieures des vitesses radiales reconstruites pour tous les points de la grille où la vitesse radiale de la simulation initiale est comprise entre 200 et 250  $\text{km s}^{-1}$ . En procédant ainsi pour toutes les tranches de vitesses radiales originales entre  $-1000$  et  $1000 \text{ km s}^{-1}$  par pas de  $40 \text{ km s}^{-1}$ , on obtient la figure 2.8, sur laquelle la couleur code l'intensité de la somme des distributions de probabilité et les tranches de vitesses radiales de la simulation initiale sont données par l'axe des ordonnées. Si la reconstruction est non biaisée et si les vitesses reconstruites ne sont pas corrélées, les densités de probabilité doivent

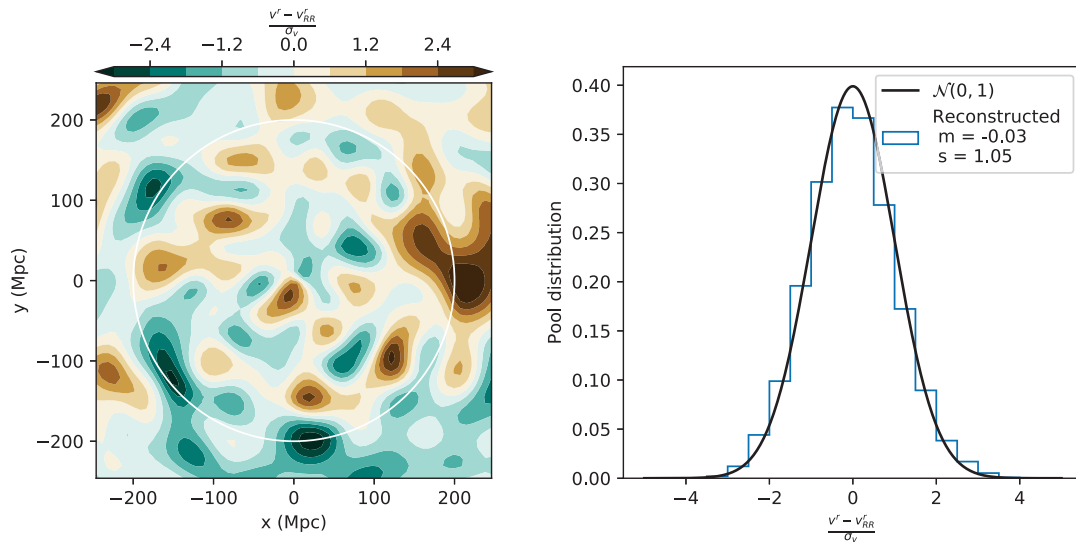


FIGURE 2.5 – Identique à la figure 2.4 pour les résidus normalisés des vitesses radiales reconstruites  $v^r$  dans le cas de la reconstruction optimale du catalogue simulé.

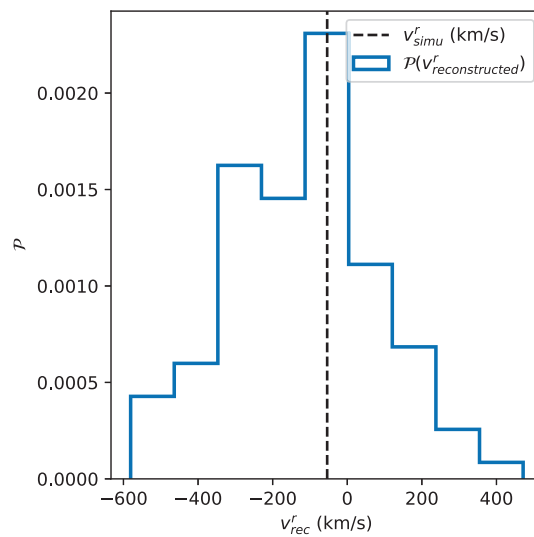


FIGURE 2.6 – Distribution postérieure des vitesses radiales reconstruites en  $(x, y, z) = (250, 250, 250)$  Mpc dans le cas de la reconstruction optimale du catalogue simulé pour laquelle aucune erreur n'a été introduite. Le trait en pointillés correspond à la vitesse radiale de la simulation de départ.

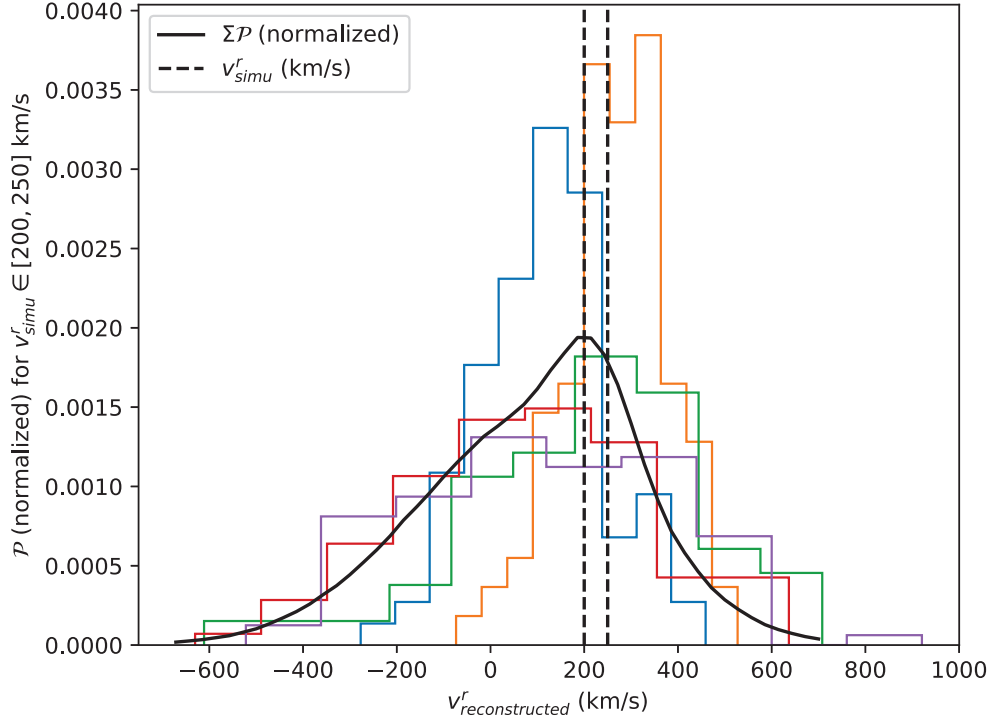


FIGURE 2.7 – Distributions postérieures des vitesses radiales reconstruites pour les points dont les vitesses radiales de la simulation originale sont comprises dans l’intervalle  $[200, 250]\text{km s}^{-1}$ , intervalle figuré par les traits en pointillés. Pour plus de clarté, seules cinq distributions tirées au hasard sont représentées. Le courbe noire figure la somme normalisée de toutes les distributions de probabilité. C’est cette somme qui est représentée horizontalement en couleur sur la Figure 2.8.

théoriquement être centrées sur la droite  $x = y$  et leur largeur donne une estimation de l’erreur sur la reconstruction. Puisqu’en pratique les vitesses reconstruites sont (très) corrélées, un décalage systématique est possible<sup>3</sup>. Le panneau de gauche correspond à la comparaison sur l’intégralité du volume de reconstruction tandis que le panneau de droite correspond à la comparaison sur un cube de côté  $L/2$  (et donc un volume de  $L^3/8$ ) au centre du volume. On constate que toutes les distributions sont compatibles avec les valeurs initiales, ce qui signifie que la reconstruction est juste. On peut voir sur le panneau de droite que la quantité d’information retrouvée est moindre quand on considère tout le volume, puisque les données sont confinées au centre. À elles trois, les Figures 2.8, 2.4 et 2.5 permettent de jauger la justesse d’une reconstruction.

3. Est-il possible d’imaginer une comparaison complète entre la reconstruction et la simulation initiale qui prend en compte cette covariance ?

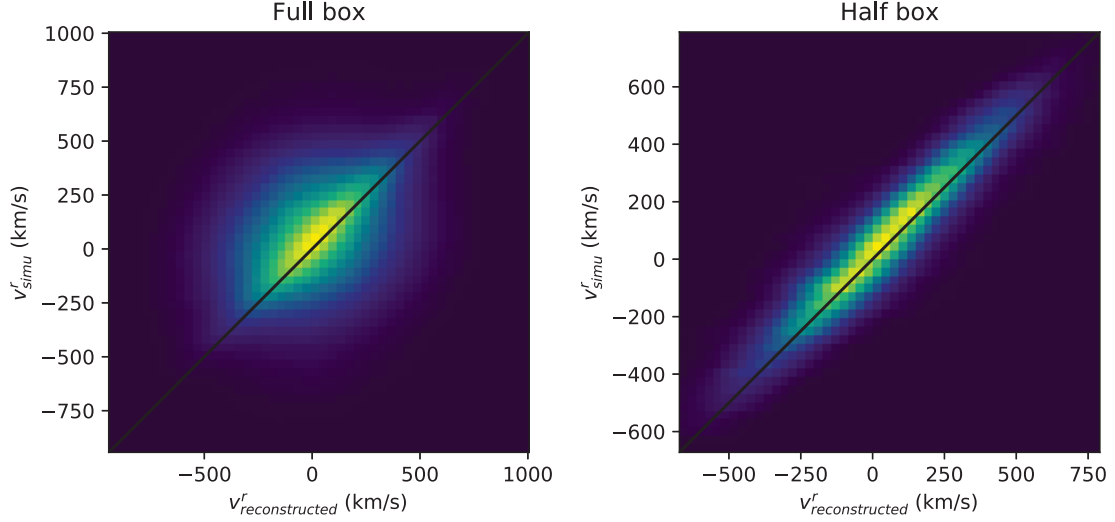


FIGURE 2.8 – Histogramme de la somme des densités de probabilité des vitesses radiales reconstruites par intervalle de vitesses radiales de la simulation. L’ordonnée représente des tranches de vitesses radiales dans la simulation originale. L’abscisse montre les vitesses reconstruites aux point correspondants. La couleur représente pour la somme des densités de probabilité, calculée comme expliquée par les figures 2.6 et 2.8. Cette représentation permet de jauger la justesse de la reconstruction. Toutes les distributions devraient être compatibles avec la droite  $x = y$ . La largeur des distributions permet d’estimer la précision de la reconstruction.

### 2.2.3 Cas des vitesses particulières mesurées

La reconstruction précédente est idéale donc irréaliste. Pour simuler l’effet des erreurs sur les vitesses particulières, nous ajoutons des erreurs dans le catalogue simulé. On suppose que l’on mesure en pratique un ensemble de modules de distance  $\mu$  mesurés avec une erreur  $\sigma_\mu = 0.3$ , comme c’est typiquement le cas pour une mesure de distance extragalactique. Les modules de distance  $\mu$  sont donc générés par :

$$\mu \sim \mathcal{N}\left(5 \log \frac{r}{10 \text{ pc}}, \sigma_\mu^2\right) \quad (2.24)$$

De ces modules de distances peuvent être déduites les distances  $d$  observées des galaxies :

$$d(\text{Mpc}) = 10^{\frac{\mu}{5} - 5} \quad (2.25)$$

Pour le moment, nous n’utilisons pas ces valeurs en tant qu’indicateurs de la position des galaxies (*i.e.* on suppose toujours que l’on connaît les véritables distances  $r$  des traceurs pour les positionner dans le volume de reconstruction), mais dispersons néanmoins les vitesses

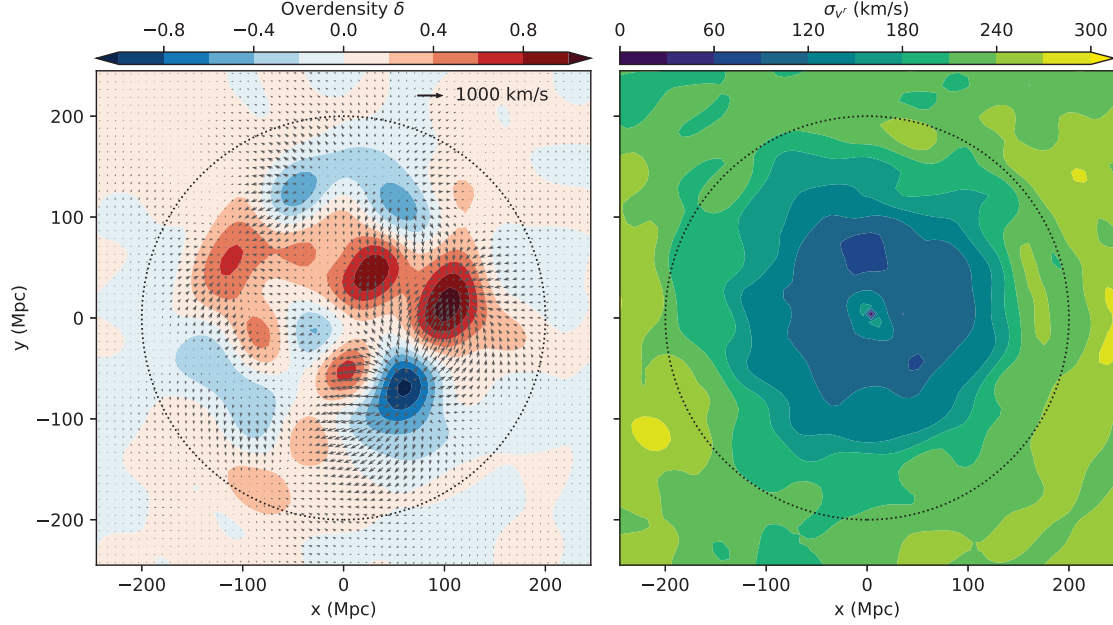


FIGURE 2.9 – Identique à la Figure 2.2 dans le cas où les vitesses particulières simulées ont été dispersées normalement via l'équation (2.26). Les structures reconstruites (panneau de gauche) sont moins contrastées que sur la figure 2.2 qui montre le cas de la reconstruction sans erreurs. On remarque aussi que la dispersion des vitesses radiales augmente avec la distance.

radiales simulées autour de leur vraie valeur selon la densité de probabilité :

$$\mathcal{P}(v^r) = \mathcal{N}\left(v^r, \left(\frac{\log 10}{5} \sigma_\mu H_0 r\right)^2\right) \quad (2.26)$$

comme suggéré par l'équation (1.49). Le terme d'erreur  $\frac{\log 10}{5} \sigma_\mu H_0 r$  provient du fait que les vitesses radiales ont été mesurées à partir des modules de distance  $\mu$  plutôt qu'à partir des vraies distances  $r$ .

Les Figures 2.9 et 2.10 sont identiques aux Figures 2.2 et 2.5 et montrent le résultat sur les données normalement dispersées. L'effet des erreurs proportionnelles à la distance est spectaculaire. Le champ reconstruit révèle beaucoup moins de structures. La distance joue comme un paramètre doublement dégradant : d'abord à cause de la diminution de la densité de données ; ensuite à cause de la proportionnalité de l'erreur sur les vitesses particulières avec la distance. La dispersion des vitesses radiales reconstruites est passée d'environ  $50 \text{ km s}^{-1}$  à une distance de 100 Mpc, à  $200 \text{ km s}^{-1}$ . Les Figures 2.10 et 2.11 montrent toujours l'aspect non biaisé de la reconstruction. En particulier, la plus grande dispersion autour de la droite  $x = y$  sur la Figure 2.11 montre que l'information retrouvée est plus faible à cause des erreurs observationnelles, mais que toutes les vitesses reconstruites sont compatibles avec leurs valeurs originales.

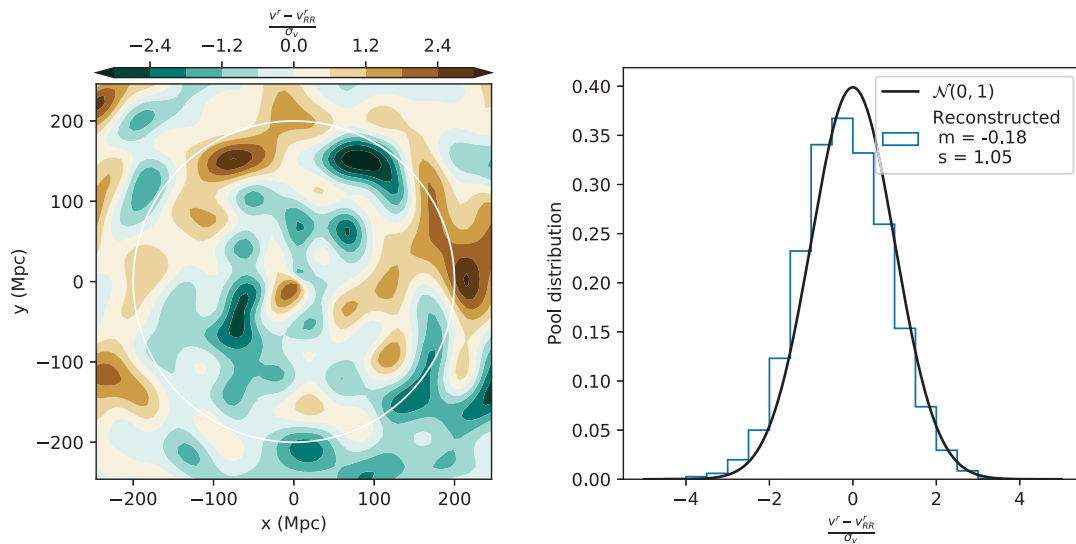


FIGURE 2.10 – Identique à la figure 2.5 dans le cas où les vitesses particulières simulées ont été dispersées normalement via l'équation (2.26).

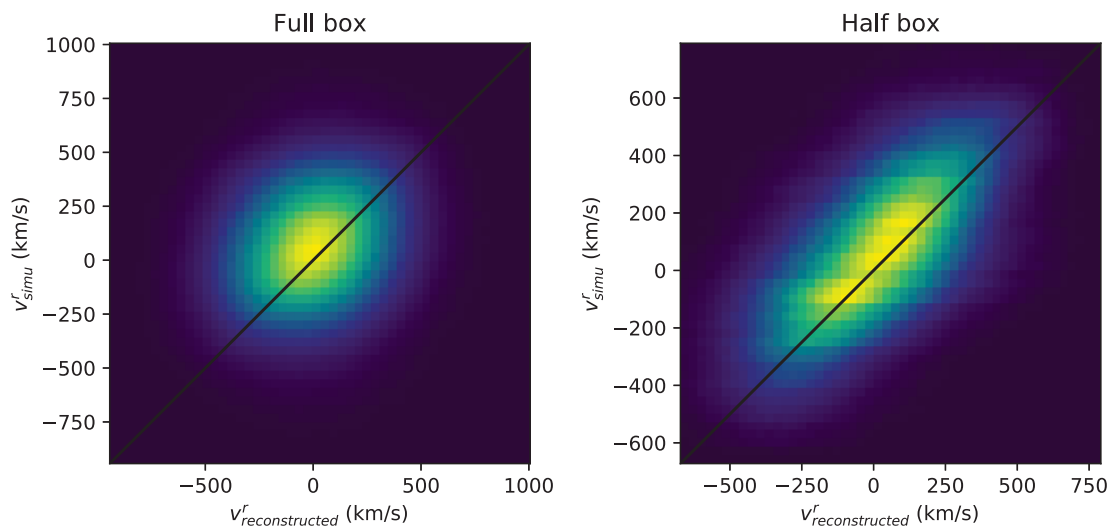


FIGURE 2.11 – Histogramme de la somme des densités de probabilité des vitesses radiales reconstruites par intervalle de vitesse radiale de la simulation dans le cas où les vitesses particulières simulées ont été dispersées normalement via l'équation (2.26).



Nous avons vu ici comment, à partir de la mesure de vitesses radiales, le champ des vitesses particulières (et de façon équivalente le champ de surdensité) pouvait être reconstruit avec l'algorithme de Hoffman-Ribak. Cependant, le test précédent (et la méthode WF/CR en général) repose sur deux suppositions fondamentales :

- la distance  $r$  des traceurs est connue,
- la distribution de chaque vitesse particulière observée est distribuée selon une loi normale.

La première supposition n'est en pratique jamais valable, puisque la distance des galaxies est mesurée et non connue. Cette distance intervient à trois niveaux : l'évaluation de la vitesse particulière radiale des galaxies, le positionnement de la galaxie dans le volume de reconstruction et l'évaluation des fonctions de corrélation entre les galaxies. Nous allons voir dans la suite quel sera l'impact de l'erreur sur la position des galaxies dans le cadre de la technique WF/CR.

### 2.3 Biais statistiques

Nous avons vu dans la section précédente comment la méthode WF/CR pouvait reconstruire le champ tridimensionnel de vitesse linéaire sous-jacent aux contraintes de vitesses radiales mesurées. Nous nous sommes cependant appuyés sur une analyse naïve dans laquelle les erreurs sur les vitesses particulières étaient gaussiennes et où nous connaissions la distance  $r$  des traceurs. Nous nous tournons maintenant vers une étude des biais statistiques du filtre de Wiener. *Dans cette section nous notons  $r$  la position réelle d'une galaxie,  $\mu$  la mesure de son module de distance, et  $d$  la mesure de distance qui découle de  $\mu$ .*

Pour comprendre comment l'erreur sur la position des traceurs rentre en compte dans l'analyse, considérons le fait que les modules de distance sont mesurés avec des erreurs gaussiennes autour du logarithme de la distance réelle  $r$  :

$$\mathcal{P}(\mu|r) = \mathcal{N}\left(5 \log \frac{r}{10 \text{ pc}}, \sigma_\mu^2\right). \quad (2.27)$$

Pour le moment nous limitons la question à l'inférence de la distance réelle des galaxies  $r$  à partir de la mesure du module de distance  $\mu$ . En toute généralité, le théorème de Bayes nous dit que la probabilité postérieure de  $r$  est :

$$\mathcal{P}(r|\mu) = \frac{\mathcal{P}(\mu|r) \mathcal{P}(r)}{\mathcal{P}(\mu)}. \quad (2.28)$$

En général la distribution  $\mathcal{P}(r)$  n'est pas constante et dépend de la densité numérique moyenne des galaxies en fonction de la distance  $r$ ,  $n(r)$ . Dans un univers isotrope :

$$\mathcal{P}(r) \propto r^2 n(r). \quad (2.29)$$

En plus de la distribution intrinsèque des galaxies en fonction de la distance, il faut prendre en compte les effets de sélection sur la mesure de  $\mu$ . En particulier, il faut prendre en compte la probabilité d'avoir effectivement observé la galaxie, événement que l'on note  $\mathcal{O}$ . Ainsi, on veut

déterminer la probabilité  $\mathcal{P}(r|\mu, \mathcal{O})$  que la galaxie soit effectivement à une distance  $r$  sachant qu'on l'a observée à un module de distance  $\mu$ . Cette probabilité peut être déduite de la façon suivante :

$$\begin{aligned}
 \mathcal{P}(r|\mu, \mathcal{O}) &= \frac{\mathcal{P}(\mathcal{O}, \mu, r)}{\mathcal{P}(\mu, \mathcal{O})} \\
 &= \frac{\mathcal{P}(\mathcal{O}|\mu, r)\mathcal{P}(\mu, r)}{\mathcal{P}(\mu, \mathcal{O})} \\
 &= \frac{\mathcal{P}(\mathcal{O}|\mu, r)\mathcal{P}(\mu|r)\mathcal{P}(r)}{\mathcal{P}(\mathcal{O}|\mu)\mathcal{P}(\mu)} \\
 &\propto \mathcal{P}(\mathcal{O}|\mu, r)\mathcal{P}(\mu|r)\mathcal{P}(r)
 \end{aligned} \tag{2.30}$$

où on appellera  $\mathcal{P}(\mathcal{O}|\mu, r)$  la fonction de sélection et où les termes indépendants de  $r$  ont été omis. La signification de l'équation (2.30) est qu'on ne peut pas déduire la distance  $r$  de la mesure  $\mu$  seule. Considérer que  $\mathcal{P}(r|\mu) = \mathcal{P}(\mu|r)$  implique l'apparition de deux biais différents : le biais de sélection, correspondant à l'omission du terme  $\mathcal{P}(\mathcal{O}|\mu, r)$  ; et le biais de Malmquist, correspondant à l'omission du terme  $\mathcal{P}(r)$ .

Notons que l'équation (2.30) n'est pas tout à fait juste, puisqu'on a considéré ici l'observation d'une galaxie  $\mathcal{O}$  comme une variable aléatoire. Une dérivation rigoureuse de cette distribution postérieure est donnée dans Hinton et al. (2017). Bien qu'inexacte, l'équation (2.30) est néanmoins très utile pour illustrer<sup>4</sup> les effets de sélection : elle montre que le lien entre la probabilité postérieure de la distance, les effets de sélection et l'*a priori* sur  $d$ .

Historiquement l'appellation *biais de Malmquist* (Strauss and Willick, 1995) recouvrait originellement les effets de sélection dans des observations à flux limité, puis a progressivement pris le sens des effets de la distribution des distances réelles  $r$  dans l'estimation de celles-ci. Pour éviter toute confusion concernant le terme *biais de Malmquist* nous nommons le premier effet *biais de sélection*, et le deuxième *biais de densité*. On suppose aussi que la vitesse radiale s'estime à partir de la distance inférée  $r$  via l'équation

$$v^r(r) = cz - H_0 r \tag{2.31}$$

qui est approximative mais suffisante pour l'étude du mécanisme des biais que nous présentons dans cette section.

Il y a enfin un troisième biais, aussi appelé *biais de Malmquist*, mais qui est propre cette fois à l'analyse des vitesses particulières. Supposons qu'on a la mesure non biaisée d'une vitesse particulière  $v^r$  à la distance mesurée  $d = 10^{\mu/5-5}$  Mpc. Le but ultime est de trouver le champ tridimensionnel  $\mathbf{v}^{\text{lin}}(\mathbf{r})$  tel qu'il est contraint par l'observation d'un ensemble de vitesses radiales  $v^r(d)$ . Ne connaissant pas les positions réelles des galaxies, la question est de savoir comment projeter les vitesses observées  $v^r(d)$  dans l'espace des distances réelles (mais inconnues)  $v^r(r) = \mathbf{v}^{\text{lin}}(r) \cdot \hat{\mathbf{r}}$ . Nous verrons que supposer  $\mathbf{v}^{\text{lin}}(d) \cdot \hat{\mathbf{r}} = v^r(d)$  est sujet à un biais que nous appellerons *biais de positionnement*.

Nous verrons également l'effet de la distribution log-normale des vitesses particulières radiales dans la section 2.3.4.

---

4. En particulier, cette probabilité postérieure n'est pas utilisée directement dans le modèle développé dans le chapitre 3.

### 2.3.1 Biais de densité

Oublions pour le moment les effets de sélection en considérant que  $\mathcal{P}(\mathcal{O}|\mu, r) = 1$ , c'est-à-dire que toutes les galaxies peuvent être observées indépendamment de leur distance et de leur module de distance mesuré. On veut quantifier l'impact du terme  $\mathcal{P}(r)$  sur l'estimation des distances  $r$ , puis sur les vitesses et enfin sur la technique WF/CR. Anticipons un peu et supposons que l'on connaisse le profil de surdensité  $\delta(r)$  dans la direction du superamas Coma<sup>5</sup>. On suppose aussi que  $n(r) = n_0(\delta(r) + 1)$  (la constante de proportionnalité est sans importance puisqu'on étudie les distributions de probabilité normalisées) ce qui est évidemment une approximation mais reste suffisant pour illustrer le biais de densité. On considère qu'on observe une galaxie à module de distance  $\mu = 5 \log \frac{d}{10 \text{ pc}}$ ,  $d = 80 \text{ Mpc}$ , avec erreur  $\sigma_\mu = 0.4$  et on trace alors sur la Figure 2.12 les probabilités  $\mathcal{P}(\mu|r)$  et  $\mathcal{P}(r|\mu)$  en prenant en compte les inhomogénéités dans l'équation (2.30). Le terme  $\mathcal{P}(\mu|r)$  représente la probabilité d'observer la valeur  $\mu = 5 \log \frac{d}{10 \text{ pc}}$  en fonction de la distance réelle de la galaxie et est donnée par l'équation (2.27). Elle est naturellement maximale (courbe bleue) quand  $r = d = 80 \text{ Mpc}$ . Le terme  $\mathcal{P}(r|\mu)$  représente quant à lui la probabilité que la galaxie soit à la distance  $r$  sachant qu'elle est mesurée avec un module de distance  $\mu$  et est donnée par l'équation générale (2.28). On voit que dans un univers homogène  $n(r) = n_0$  la mesure seule a tendance à sous-estimer la valeur de la distance  $r$  à cause de l'effet de volume  $\mathcal{P} \propto r^2$ , c'est le biais de Malmquist homogène. On peut montrer (Strauss and Willick, 1995) que l'effet du biais de Malmquist homogène est proportionnel à la distance mesurée  $d$  et  $\bar{r} = d \times 10^{\frac{\tau \sigma_\mu}{10}}$ . Dans un univers inhomogène, l'effet est encore plus fort puisque la distribution de la vraie distance se recentre dans les régions surdenses, c'est le biais de Malmquist inhomogène. Ce biais de densité est directement transposable dans l'espace des vitesses particulières inférées puisqu'il y a une bijection entre l'espace des distances  $r$  et des vitesses  $v^r(r)$  via l'équation (2.31). On aura ainsi tendance à, en général, surestimer les vitesses particulières radiales à cause de l'effet de volume. Les vitesses particulières seront surestimées devant un amas, et sous-estimées derrière, ce qui provoque un effondrement apparent trop grand autour de cet amas.

### 2.3.2 Biais de sélection

On a supposé dans la sous-section précédente que  $\mathcal{P}(\mathcal{O}|\mu, r) = 1$  et qu'il n'y avait aucun effet de sélection en jeu. On veut voir ici quel est l'impact de ce terme. Pour modéliser le terme  $\mathcal{P}(\mathcal{O}|\mu, r)$ , on peut d'abord supposer que les effets de sélection ne dépendent que de la distance réelle  $r$  de la galaxie. Ceci est tout à fait plausible puisque l'effet principal est la sélection en magnitude apparente, et cette magnitude apparente croît avec la distance  $r$ . La distance peut ainsi être décomposée en une magnitude apparente  $m$  et une magnitude absolue

---

5. Ce profil vient de la reconstruction du champ de surdensité avec *CosmicFlows-3* qui sera faite dans le chapitre 4

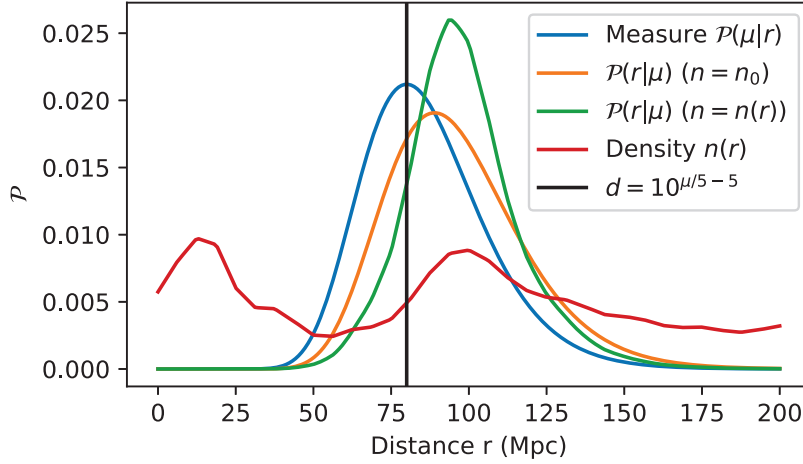


FIGURE 2.12 – Distributions de probabilité de la distance réelle  $r$  d'une galaxie si l'on a mesuré son module de distance  $\mu = 5 \log_{10} 80/10$  pc.  $\mathcal{P}(r|\mu)$  correspond à la probabilité réelle tandis que  $\mathcal{P}(\mu|r)$  correspond à la probabilité de  $r$  si on ne prend en compte que la mesure. On voit que même dans un univers homogène, la distance des galaxies est en général sous-estimée : la courbe orange est décalée par rapport à la courbe bleue, c'est le biais de Malmquist homogène. L'inhomogénéité du super-amas Coma (en rouge à  $r \sim 100$  Mpc) affecte encore plus la distribution de  $r$  (courbe verte) en la décalant vers la surdensité et la resserrant autour d'elle, c'est le biais de Malmquist inhomogène.

$M$  de la galaxie, telles que  $5 \log \frac{r}{10 \text{ pc}} = m - M$ . On aura alors :

$$\begin{aligned}
 \mathcal{P}(\mathcal{O}|\mu, r) &= \int \mathcal{P}(\mathcal{O}|m, r) P(m|r) dm \\
 &= \int \mathcal{P}(\mathcal{O}|m, r) \left( \int \mathcal{P}(m|r, M) \mathcal{P}(M) dM \right) dm \\
 &= \int \mathcal{P}(\mathcal{O}|m, r) \left( \int \delta_D \left( 5 \log \frac{r}{10 \text{ pc}} - m + M \right) \mathcal{P}(M) dM \right) dm \\
 &= \int \mathcal{P}(\mathcal{O}|m, r) \mathcal{P} \left( M = m - 5 \log \frac{r}{10 \text{ pc}} \right) dm. \tag{2.32}
 \end{aligned}$$

La deuxième égalité vient de la somme sur toutes les magnitudes potentiellement observables, et la troisième vient du fait que le module de distance  $5 \log \frac{r}{10 \text{ pc}}$  et la magnitude  $m$  de la galaxie sont liées par  $5 \log \frac{r}{10 \text{ pc}} = m - M$ . Les observations sont en général à flux limité et on peut par exemple écrire :

$$\mathcal{P}(\mathcal{O}|m, r) = \mathcal{P}(\mathcal{O}|m) = \frac{1}{m_0} \begin{cases} 1 & \text{si } 0 \leq m \leq m_0, \\ 0 & \text{sinon,} \end{cases} \tag{2.33}$$

où  $m_0$  est la magnitude limite et on a considéré que les galaxies n'étaient sélectionnées que sur leur magnitude apparente  $m$ . La magnitude absolue mesurée dépend quant à elle des effets

de sélection des indicateurs de distances définis en section 1.2.3 qui eux-mêmes dépendent considérablement de l'indicateur de distance utilisé. Il y a en plus l'effet naturel de la distribution intrinsèque des luminosités des galaxies, qui peut être décrite par une fonction de Schechter (Schechter, 1976) :

$$\mathcal{P}(M) \propto 10^{0.4(\alpha+1)(M^0-M)} \exp\left(-10^{0.4(M-M^0)}\right) \quad (2.34)$$

où  $\alpha$  et  $M^0$  sont évalués expérimentalement. On peut ainsi déterminer (numériquement, car le résultat n'est pas analytique) la fonction de sélection  $\mathcal{P}(\mathcal{O}|r, \mu)$ . On peut alors calculer la probabilité  $\mathcal{P}(r|\mu, \mathcal{O})$  grâce à l'équation (2.30). On trace alors sur la Figure 2.3.2 les probabilités  $\mathcal{P}(\mathcal{O}|r, \mu)$ ,  $\mathcal{P}(r|\mu, \mathcal{O})$  et  $\mathcal{P}(r|\mu)$  pour un module de distance fixé à  $\mu = 5 \log d/10$  pc,  $d = 150$  Mpc, qui représentent respectivement la fonction de sélection en fonction de la distance  $r$ , la distribution de probabilité de la distance  $r$  sachant qu'on a observé cette galaxie et que son module de distance est mesuré à  $\mu$ , et cette même probabilité si l'on pense naïvement qu'il n'y a pas d'effets de sélection. Notons que pour calculer  $\mathcal{P}(r|\mu)$  on a négligé les effets précédents de volume et d'inhomogénéités. On voit que les effets de sélection, s'ils ne sont pas pris en compte, tendent à faire surestimer les distances. Il s'agit donc d'un effet inverse au biais de Malmquist homogène et qui apparaît cette fois lorsqu'on écrit la probabilité  $\mathcal{P}(r|\mu)$ , c'est à dire la vraisemblance (le biais de Malmquist étant un problème d'*a priori* sur les distances). Il est difficile d'estimer numériquement l'impact de ce biais, puisqu'il dépend par exemple des valeurs choisies pour la magnitude limite  $m_0$  et pour les paramètres  $\alpha$ ,  $M^0$  de la fonction de Schechter. Dans notre cas raisonnable, on trouve un effet d'environ 10%. Ce biais de sélection est directement transposable dans l'espace des vitesses particulières inférées puisqu'il y a une bijection entre l'espace des distances  $r$  et des vitesses  $v^r(r)$  via l'équation (2.31). On aura en général tendance à sous-estimer les vitesses particulières radiales à grande distance.

### 2.3.3 Biais de positionnement

Le but final est de contraindre le champ de vitesse  $\mathbf{v}(\mathbf{r}) \cdot \hat{\mathbf{r}}$  à partir de mesures de vitesses radiales  $v^r$  à des distances mesurées  $d$ . La question est de savoir quelle région du champ  $\mathbf{v}(\mathbf{r}) \cdot \hat{\mathbf{r}}$  est contrainte par de telles observations. On pourrait *a priori* penser que les contraintes se font aux positions données par  $r = d\hat{\mathbf{r}}$ , cependant nous avons vu que les mesures  $d$  étaient biaisées, et qu'en moyenne ces positions sont décalées par rapport aux positions réelles  $r$  auxquelles sont mesurées les vitesses. L'effet est double puisqu'en plus de contraindre le champ de vitesse aux mauvaises positions, il est contraint par des vitesses radiales biaisées elles aussi. Le biais de positionnement que nous présentons ici est difficile à formaliser puisqu'il est la conséquence des biais de densité et de sélection associés à la projection des vitesses radiales mesurées dans l'espace des distances réelles. On peut néanmoins avoir une idée de son ordre de grandeur de la façon suivante. Supposons qu'on ait une continuité d'observations  $v^r(d)$  aux distances  $d$ . L'estimation du champ  $\mathbf{v}(\mathbf{r}) \cdot \hat{\mathbf{r}}$  à la position  $r$  peut être déterminée par la moyenne sur les observations :

$$\mathbf{v}(\mathbf{r}) \cdot \hat{\mathbf{r}} = \int v^r(d) \mathcal{P}(r|d) dd. \quad (2.35)$$

Ici,  $\mathcal{P}(r|d)$  peut être interprété comme le poids de la contribution  $v^r(d)$  au champ de vitesse à la distance  $r$ . Dans le cas du filtre de Wiener, la contrainte sur le champ  $\mathbf{v}$  est discrète et

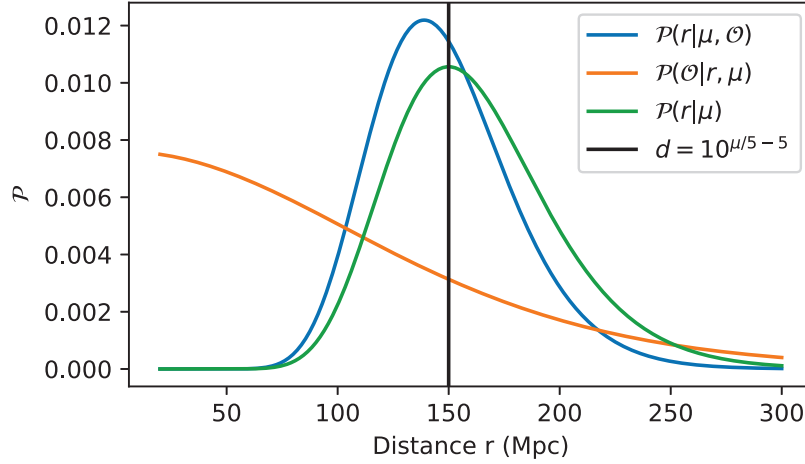


FIGURE 2.13 – Quelques distributions de probabilité montrant l’effet de sélection. En orange, la probabilité  $\mathcal{P}(\mathcal{O}|\mu, r)$  qui représente la probabilité d’observer une galaxie à distance  $r$ . Cette probabilité a été calculée numériquement à partir de l’équation (2.32). La probabilité  $\mathcal{P}(r|\mu)$  en vert montre la distribution des distances inférées sans prendre en compte les effets de sélection. La probabilité  $\mathcal{P}(r|\mu, \mathcal{O})$  en bleu montre la distribution réelle des distances. En pratique on observe une réalisation de la courbe bleue, et on la modélise comme la courbe verte, c’est-à-dire que les distances sont généralement surestimées.

est donnée par une série de vitesses  $\{v^r(d_i)\}_{i=1\dots n}$  à des distances  $\{d_i\}_{i=1\dots n}$ . La supposition sous-jacente est que  $\mathcal{P}(r|d_i) = \delta_D(r - d_i)$ . Cependant, nous avons vu que la distance observée  $d$  était en général un estimateur biaisé de la position, et que la vitesse  $v^r(d)$  était aussi un estimateur biaisé de la vitesse radiale. Là où l’effet de volume domine, les distances mesurées sous-estiment les distances réelles et les vitesses sont surestimées, il y aura ainsi apparition d’un flot sortant radialement par rapport à l’observateur. Au contraire, dans la région où les effets de sélection dominent, il y aura apparition d’un flot entrant radial. Ces effets sont schématisés dans la figure 2.3.3.

### 2.3.4 Distribution log-normale des vitesses particulières

Indépendamment de la connaissance parfaite de la position des galaxies, le modèle statistique sur lequel repose la méthode des réalisations contraintes suppose que la distribution statistique des vitesses particulières radiales déduites des observations est gaussienne autour de celle donnée par le champ sous-jacent. C’est évidemment une approximation : puisque ce sont les modules de distance qui ont une statistique gaussienne, les vitesses particulières déduites de l’Eq. 1.47 ont, au premier ordre en  $z$ , une distribution Log-normale. En effet, soit le module de distance  $\mu'$  distribué normalement autour d’une vraie valeur  $\mu$  avec dispersion  $\sigma_\mu$  :

$$\mathcal{P}(\mu'|\mu, \sigma_\mu) = \mathcal{N}(\mu, \sigma_\mu^2) \quad (2.36)$$

En effectuant le changement de variable  $x = \frac{cz-v^r}{H_0} = 10^{\mu'/5-5}$  (on a approximé ici la valeur de  $v^r$  par  $cz - H_0d$ ) :

$$\mathcal{P}(x|\mu, \sigma_\mu^2) = \frac{5}{x \log 10} \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2} \frac{\left(\log x - \frac{\log 10}{5}(\mu - 25)\right)^2}{\left(\frac{\log 10}{5}\sigma_\mu\right)^2}\right) \quad (2.37)$$

qui est bien une loi log-normale. L'espérance de cette loi est donnée par :

$$E(x) = \exp\left(\frac{\log 10}{5}(\mu - 25) + \frac{\left(\frac{\log 10}{5}\sigma_\mu\right)^2}{2}\right) \quad (2.38)$$

$$= 10^{\frac{\mu}{5}-5} \exp\left(\frac{\left(\frac{\log 10}{5}\sigma_\mu\right)^2}{2}\right) \quad (2.39)$$

$$\neq 10^{\frac{\mu}{5}-5}. \quad (2.40)$$

Ainsi :

$$\langle v^{r'} \rangle = \langle cz - H_0 10^{\frac{\mu'}{5}-5} \rangle \quad (2.41)$$

$$\neq v^r \quad (2.42)$$

ce qui montre bien que l'estimateur est biaisé<sup>6</sup>.

Un autre effet dû à la distribution log-normale des vitesses particulières est l'estimation des erreurs sur les vitesses particulières. À partir de l'équation (2.31), on peut en effet penser que l'erreur sur  $v^r(d)$  est naïvement :

$$\sigma_{v^r}(d) = \frac{\log 10}{5} \sigma_\mu H_0 d \quad (2.43)$$

Cependant, cette équation repose sur la propagation d'erreur qui provient d'un développement en série de Taylor. En effet, pour toute fonction  $f$  continue et dérivable :

$$f(x+a) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \dots \quad (2.44)$$

La subtilité réside dans le fait que la fonction  $f'$  doit être évaluée en  $a$ , et non en  $x$ , correction négligeable quand l'erreur sur  $x$  est faible. Mais dans le cas des vitesses particulières, la valeur de  $\sigma_{v^r}(d)$  peut être significativement différente de sa valeur réelle  $\sigma_{v^r}(r)$ . Pour se figurer cet effet, imaginons un univers sans mouvement où  $\mathbf{v} = \mathbf{0}$ . On mesure la distance  $d$  d'une galaxie

6. Un estimateur non biaisé pourrait être, par exemple :

$$v^{r'} \triangleq cz - H_0 10^{\frac{\mu'}{5}-5} \exp\left(-\frac{\left(\frac{\log 10}{5}\sigma_\mu\right)^2}{2}\right)$$

située à une distance  $r$ . La vitesse particulière mesurée sera  $v^r = H_0(r - d)$  et son erreur  $\frac{\log 10}{5} \sigma_\mu H_0 d$ . On aura alors :

$$d > r \Rightarrow v^r < 0 \quad (2.45)$$

$$\Rightarrow \sigma_{v^r}(d) > \sigma_{v^r}(r) \quad (2.46)$$

$$d < r \Rightarrow v^r > 0 \quad (2.47)$$

$$\Rightarrow \sigma_{v^r}(d) < \sigma_{v^r}(r) \quad (2.48)$$

et donc :

$$\sigma_{v^r}(v^r < 0) > \sigma_{v^r}(v^r > 0). \quad (2.49)$$

Et par conséquent, les vitesses radiales sont biaisées vers les valeurs positives.

## 2.4 Reconstructions

Nous avons souligné dans la section précédente les biais statistiques susceptibles d'apparaître lors d'une reconstruction via la technique WF/CR. Nous testons ici diverses reconstructions à partir du catalogue défini en section 2.2 pour tester la technique WF/CR face aux biais mentionnés ci-dessus. À chaque fois, nous traçons la tranche  $z = 0$  de la reconstruction, la distribution du *pull* des vitesses radiales reconstruites et leur comparaison bayésienne. Ces représentations sont décrites et expliquées dans la section 2.2.2. Cinq reconstructions sont faites, et leurs caractéristiques sont les suivantes :

1. La position réelle  $r$  des galaxies est connue, les vitesses radiales sont déduites des positions mesurées et les erreurs sont estimées à partir de la distance réelle  $r$ . Cette reconstruction est sujette au biais de densité et de sélection.
2. La position des galaxies est inconnue, les vitesses radiales sont déduites des positions mesurées  $d$  et les erreurs sont estimées à partir de la distance réelle  $r$ . Cette reconstruction est sujette au biais de densité, de sélection et de positionnement.
3. La position des galaxies est connue, les vitesses radiales sont déduites des positions réelles  $r$  avec un bruit gaussien et les erreurs sont estimées à partir de la distance mesurée  $d$ . Cette reconstruction est sujette au biais des erreurs lognormales.
4. La position des galaxies est inconnue, les vitesses radiales sont déduites des distances mesurées  $d$  et les erreurs sont estimées à partir de la distance mesurée  $d$ . Cette reconstruction est sujette au biais de densité, de sélection, de positionnement et des erreurs lognormales. C'est le cas usuel en pratique.

Les différents cas sont rappelés dans la Table 2.4 et les reconstructions sont montrées sur la Figure 2.15. Le cas référence est montré tout à gauche. Dans le cas des vitesses particulières radiales estimées à partir de la distance mesurée (cas # 1), la reconstruction est biaisée vers les vitesses radiales négatives, provoquant un flot entrant sur le volume de reconstruction. C'est la conséquence du biais de sélection expliqué section 2.3.2. Le biais de densité, qui est principalement dû à l'effet de volume, n'est pas visible sur cette reconstruction. En effet, les erreurs sur les vitesses particulières radiales étant proportionnelles à la distance, elles sont trop



TABLE 2.1 – Récapitulatif des reconstructions étudiées dans la section 2.4.  $r$  représente la distance réelle de la galaxie et  $d$  sa distance observée.

#	Position	Vitesses radiales	Erreurs	Biais
Référence	$r$	$cz - H_0 r$	$H_0 \frac{\log 10}{5} \sigma_\mu r$	Aucun
1	$r$	$cz - H_0 d$	$H_0 \frac{\log 10}{5} \sigma_\mu r$	Densité, sélection
2	$d$	$cz - H_0 d$	$H_0 \frac{\log 10}{5} \sigma_\mu r$	Densité, sélection, positionnement
3	$r$	$\mathcal{N}(cz - H_0 r, \sigma_v)$	$H_0 \frac{\log 10}{5} \sigma_\mu d$	Erreurs log-normales
4	$d$	$cz - H_0 d$	$H_0 \frac{\log 10}{5} \sigma_\mu d$	Tous

faibles là où l'effet de volume domine pour que le biais dû à l'effet de volume soit dominant. Le biais de sélection est amplifié par le biais de positionnement lorsque l'on place incorrectement les galaxies dans le volume de reconstruction, comme le montre le panneau du milieu (cas # 2). La situation est dégradée et on décèle même la présence d'un flot sortant au centre de la boîte, manifestation de l'effet de volume et du biais de densité. Le cas # 4 montre l'effet de la mésestimation des erreurs sur les vitesses particulières radiales. Comme attendu après la description faite en section 2.3.4, les vitesses reconstruites sont biaisées positivement et on observe un flot sortant radial. Finalement, le panneau de droite montre l'analyse du catalogue simulé par la technique WF/CR si l'on procède naïvement en ne prenant aucun de ces biais en compte. Il est important de noter que le biais des erreurs lognormales semble compenser partiellement les biais de sélection et de positionnement, en tout cas si l'on ne regarde que la moyenne des vitesses radiales reconstruites.

### 2.4.1 Solutions partielles

Plusieurs solutions apparaissent dans la littérature pour pallier ces biais dans le cadre des reconstructions de champ de vitesse. Nous passons ici en revue qualitativement ces « corrections du biais de Malmquist ».

**Le biais de positionnement** Le biais de positionnement est d'apparence le plus simple à corriger. En effet, ce biais provient du mauvais positionnement de la galaxie dans le volume de reconstruction. Cependant, nous avons aussi accès à un autre indicateur de distance que la mesure de distance, celle du décalage vers le rouge. En effet, si la vitesse particulière est faible devant le décalage vers le rouge,  $v^r \ll cz$ , alors  $cz \sim H_0 d$  et ainsi :

$$d \sim \frac{cz}{H_0} \triangleq d_z. \quad (2.50)$$

Le modèle  $\Lambda$ CDM suppose en plus que la distribution des vitesses est gaussienne avec une déviation standard  $\sigma_v = 300 \text{ km s}^{-1}$ . Ainsi l'erreur (gaussienne) sur l'estimateur  $d_z$  est :

$$\sigma_{d_z}^2 = \frac{(c\sigma_z)^2 + \sigma_v^2}{H_0^2} \quad (2.51)$$

$$\sim (4 \text{ Mpc})^2 \quad (2.52)$$

ce qui est plus précis que tous les indicateurs de distance à partir de quelques dizaines de Mpc. En plaçant les galaxies à leur position dérivée de  $d_z$ , on réduit le biais de positionnement cité ci-dessus<sup>7</sup>. Cependant, trois problèmes subsistent. Le premier est la non-correction des biais de densité et de sélection. Le deuxième provient des non-linéarités. En effet, si une galaxie se trouve proche d'un groupe ou un amas, sa vitesse particulière peut être bien au-dessus de  $300 \text{ km s}^{-1}$ . Ainsi, l'erreur sur sa position sera potentiellement d'une dizaine de Mpc. Si l'on envisage le cas de deux galaxies, l'une devant un amas ayant une vitesse radiale positive, et l'autre derrière l'amas ayant une vitesse négative, alors leur positionnement avec  $d_z$  provoquera la situation suivante : la galaxie apparemment devant l'amas aura une vitesse radiale négative, tandis que celle derrière l'amas aura une vitesse positive. S'ensuit ainsi une reconstruction erronée des structures denses. Enfin, quand les distances sont faibles, en dessous de la dizaine de Mpc, l'estimateur  $d_z$  est très imprécis et peut amener à une mauvaise reconstruction du champ de vitesse local, particulièrement utile pour les simulations contraintes du groupe local.

**Distribution log-normale des vitesses particulières** Watkins and Feldman (2015) proposent de traiter ce biais en utilisant l'estimateur (à statistique gaussienne) des vitesses suivant :

$$v_{wf}^r = \frac{cz}{1 + \bar{z}} \ln \frac{cz}{H_0 d} \quad (2.53)$$

avec erreur :

$$\sigma_{v^r} = \frac{cz}{1 + \bar{z}} \frac{\log 10}{5} \sigma_\mu. \quad (2.54)$$

On peut facilement se convaincre que  $v_{wf}^r \rightarrow v^r$  quand  $z \rightarrow \infty$ , et que sa distribution est bien gaussienne. Cependant, cet estimateur revêt les mêmes problèmes que précédemment : il est incorrect localement et ne permet pas de modéliser les biais de sélection, densité et positionnement.

**Corrections du biais de Malmquist** Deux autres méthodes ont été proposées pour minimiser l'effet du « biais de Malmquist » (qui regroupe en fait les trois biais dont nous avons parlé précédemment) dans l'étude des vitesses particulières. Citons Tully et al. (2014) qui propose de replacer les galaxies en utilisant le champ de vitesse reconstruit et le décalage vers le rouge des traceurs :

$$\hat{d} = \frac{cz - \alpha v_{WF}^r}{H_0} \quad (2.55)$$

où  $v_{WF}^r$  est la composante radiale du champ de vitesse reconstruit par le filtre de Wiener, et  $\alpha$  est un paramètre *ad-hoc* qui est fixé en minimisant le monopole du champ de vitesse. Si cette méthode règle partiellement le biais de positionnement, elle néglige les autres effets décrits ci-dessus.

Un autre méthode est proposée par Sorce (2015). Elle suppose que la distribution des vitesses particulières radiales doit être gaussienne avec une déviation standard de  $300 \text{ km s}^{-1}$  comme prédit par le modèle  $\Lambda$ CDM. Il est alors d'estimer les vraies distances sous-jacentes en « gaussianisant » la distribution des vitesses observées. Cette méthode propose donc une correction globale sans prendre en compte les différents phénomènes physiques sous-jacents

7. Cette méthode est historiquement appelée *Methode II* (Strauss and Willick (1995)).

à la présence d'erreurs systématiques. Il est de plus à noter que supposer que la distribution des vitesses radiales est une gaussienne unidimensionnelle, c'est négliger toutes les covariances entre les points de données, qui sont précisément celles que l'on utilise pour estimer le champ de vitesse.

## 2.5 Conclusion sur la méthode WF/CR

Cette étude préliminaire sur la technique du filtre de Wiener et réalisations contraintes pose les bases de cette thèse. Nous avons vu dans les sections 2.3 et 2.3.4 que l'incertitude sur la position des galaxies joue un rôle déterminant dans la reconstruction du champ de vitesse. En particulier, le mauvais placement des galaxies provoque un flot entrant conséquent dans la région (éloignée) où les effets de sélection jouent un rôle non négligeable. L'équation principale pour décrire ces effets est l'équation (2.30) :

$$\mathcal{P}(r|\mu, \mathcal{O}) \propto \mathcal{P}(\mathcal{O}|\mu, r)\mathcal{P}(\mu|r)\mathcal{P}(r).$$

Pour prendre en compte les biais dûs aux termes  $\mathcal{P}(\mathcal{O}|\mu, r)$  et  $\mathcal{P}(r)$ , il faut échantillonner cette distribution de probabilité, ou une distribution approchée. Il est donc naturel de construire un modèle qui introduit la distance  $r$  comme paramètre libre plutôt que la vitesse radiale observée directement. Échantillonner sur la distance  $r$  permettra de prendre en compte *a priori* sur  $r$ , et d'échantillonner en même temps la vitesse radiale puisqu'elle est déduite directement de  $r$ , et donc d'éviter le biais de distribution log-normale. Le modèle qui découle naturellement de la technique WF/CR et de ces études sur les biais sera décrit dans le Chapitre 3.

Une autre limite de la méthode WF/CR que nous n'avons pas soulignée précédemment est l'inclusion des autres paramètres libres du modèle. Le premier de ces paramètres,  $\sigma_{\text{NL}}$ , a été introduit dans la section 2.1.1. Il modélise l'écart à la non-linéarité par une distribution d'erreur gaussienne supplémentaire sur les vitesses. Il a été fixé à sa valeur nominale dans tous les cas d'étude précédents. En général, il est ajusté itérativement pour que :

$$\chi^2 = \mathbf{v}^{r,T} \mathbf{C}_{cc} \mathbf{v}^r \sim 1/\text{d.o.f.} \quad (2.56)$$

où d.o.f. correspond au nombre de degrés de liberté du modèle. L'autre paramètre est la constante de Hubble  $H_0$ , qui permet de lier les distances aux vitesses. Ce paramètre est très corrélé avec le champ de vitesse puisqu'il est dégénéré (à  $z \ll 0$ ) avec un monopole  $\nabla \cdot \mathbf{v}$  du champ de vitesse. Il dépend en plus de la calibration absolue des indicateurs de distance utilisés. En général, ce paramètre est fixé de telle sorte que le monopole global à une distance fixée soit nulle. Cette méthode souffre d'un défaut majeur : les biais de Malmquist provoquent aussi un monopole sur le champ de vitesse, et la valeur de  $H_0$  est donc corrélée à ces biais, sans pour autant pouvoir les corriger de façon *ad hoc*. Il est donc nécessaire d'inclure l'estimation de ce paramètre, avec  $\sigma_{\text{NL}}$ , en tant que partie intégrante de l'analyse du champ de vitesse. Le modèle présenté dans le Chapitre 3 offre également une solution à ce problème. Le modèle sera ensuite appliqué aux données de *CosmicFlows-3* dans le Chapitre 4.

Enfin, l'analyse des biais faite ici a supposé que les modules de distances étaient mesurés directement avec une erreur gaussienne. C'est cependant une approximation. En règle générale

le module de distance vient d'un indicateur de distance qui donne la magnitude absolue d'un objet  $M$  en fonction d'un paramètre  $\eta$  indépendant, comme la vitesse de rotation pour la relation de Tully-Fisher. La relation de calibration s'écrit en général :

$$M = ZP + s\eta \quad (2.57)$$

où  $ZP$  et  $s$  sont le point zéro et la pente de la relation. Les mesures sont donc en pratique l'ensemble  $(M, \eta)$  en plus des calibrateurs qui permettent de définir la valeur de  $ZP$  et  $s$ . La distribution des modules de distances déduits de :

$$\mu = m - M \quad (2.58)$$

n'est donc pas nécessairement gaussienne. En effet, elle dépend de l'erreur sur les paramètres de calibration, de la dispersion intrinsèque de la relation, de l'erreur sur le(s) paramètre(s) indépendant(s)  $\eta$ , des calibrateurs, etc. Il en va de même pour les effets de sélection qui dépendent des mêmes paramètres. Il reste que cette thèse s'emploie aussi à essayer d'inclure la relation de calibration des indicateurs de distance dans le modèle de vitesses particulières. Le cas de la relation de Tully-Fisher sera présenté dans le Chapitre 5.

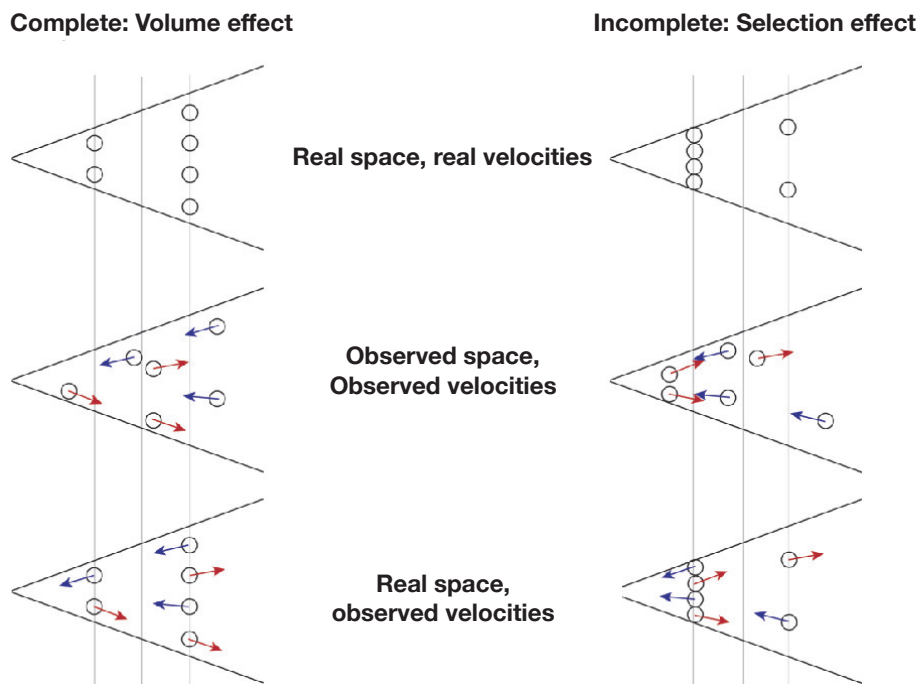


FIGURE 2.14 – Illustration du biais de positionnement dans le cas où l’effet de volume domine (à gauche), et dans le cas où les effets de sélection dominant (à droite). Dans cet exemple, les galaxies n’ont aucune vitesse particulière réelle (schéma du haut), mais les incertitudes sur les distances créent des vitesses particulières observées. En rouge (resp. bleu) les vitesses apparentes positives (resp. négatives) (distances sous-estimées (resp. surestimées)). En regardant la moyenne des vitesses radiales observées (milieu) au centre du volume d’observation, on voit que les vitesses particulières sont surestimées où l’effet de volume domine, et sous-estimées quand les effets de sélection dominant.

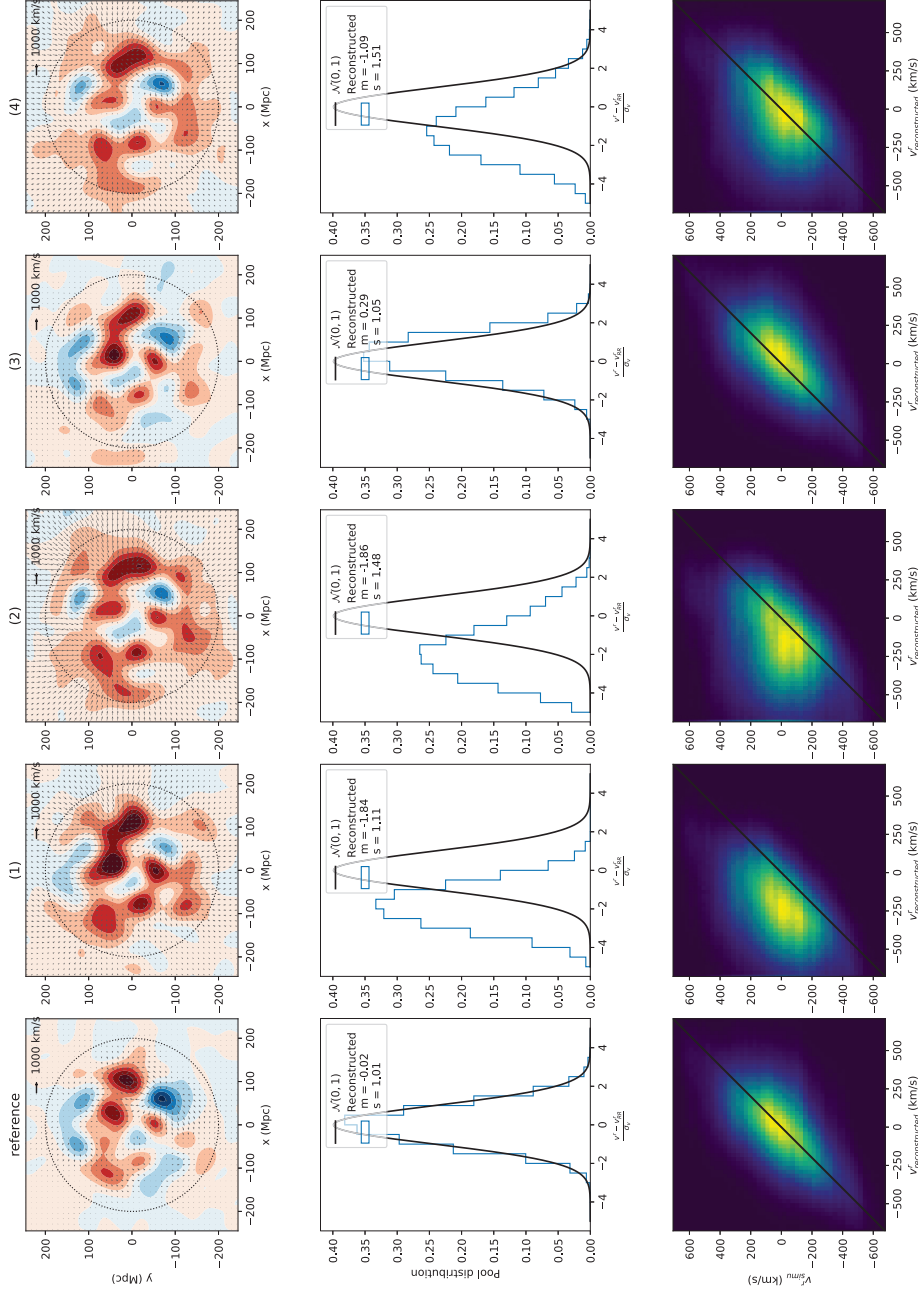


FIGURE 2.15 – Reconstructions exhibant les biais de densité, de sélection, de positionnement et de log-normalité décrits en section 2.3. Les colonnes correspondent aux cas référencés dans le Tableau 2.4. # 1 : Estimation des vitesses particulières radiales à partir des distances mesurées ; # 2 : Positionnement des galaxies à la distance mesurée ; # 3 : Estimation des erreurs individuelles sur les vitesses radiales à partir de la distance mesurée ; # 4 : Tous les effets cités précédemment. Haut : Tranche en  $z = 0$  de la reconstruction. Milieu : distribution des résidus normalisés des vitesses particulières radiales sur le volume de reconstruction. Bas : comparaison bayésienne entre la simulation originale et les vitesses radiales reconstruites.



# Chapitre 3

## Modèle statistique

À partir des données du catalogue *CosmicFlows-3* (CF3) qui regroupe les distances lumineuses et décalages vers le rouge de quelques 18,000 galaxies, notre objectif est de reconstruire les champs de vitesse et de surdensité linéaires sous-jacents. nous avons vu dans le chapitre 2 quels biais pouvaient apparaître suite à une mauvaise modélisation de la mesure des modules de distance. La méthode des réalisations contraintes appliquée à un catalogue tel CF3 résulte en une reconstruction irréaliste avec notamment un flot entrant radialement de plusieurs milliers de  $\text{km s}^{-1}$ . Afin de prévenir les différents biais de Malmquist, l'incertitude sur la mesure du module de distance doit être propagée à la fois à la mesure de la vitesse particulière radiale et à la position d'une galaxie. L'idée de base pour modéliser les vitesses particulières est de laisser la distance des galaxies comme un paramètre libre. En parallèle de l'étude préliminaire faite dans le chapitre 2, Lavaux (2016) a développé un tel modèle bayésien dont nous présentons une adaptation et implémentation dans les sections 3.1 et 3.2. Dans la section 3.3, le modèle est finalement testé sur le catalogue simulé défini dans le chapitre précédent.

### Sommaire

---

<b>3.1</b>	<b>Le modèle</b>	<b>58</b>
3.1.1	Introduction	58
3.1.2	Vraisemblance	58
3.1.3	<i>A priori</i> bayésiens	64
<b>3.2</b>	<b>Algorithme</b>	<b>67</b>
3.2.1	Probabilités conditionnelles	67
3.2.2	Marginalisation et algorithme de Gibbs effondré	69
3.2.3	Implémentation en PYTHON	69
<b>3.3</b>	<b>Tests</b>	<b>71</b>
3.3.1	Paramètres globaux et convergences	71
3.3.2	Champs de vitesse et de surdensité	72
<b>3.4</b>	<b>Conclusion</b>	<b>75</b>

---



### 3.1 Le modèle

Dans cette section est présenté le modèle utilisé pour analyser le champ de vitesse linéaire de l'Univers Local à partir de l'observation de décalages vers le rouge et de modules de distance.

#### 3.1.1 Introduction

Le modèle est motivé par le traitement des différents biais de Malmquist ainsi que la distribution log-normale des vitesses particulières. La clef du modèle réside dans le fait que la position des galaxies intervient à trois niveaux : l'estimation des vitesses particulières avec l'équation (1.47), l'estimation des fonctions de corrélations  $\Psi$  et  $\zeta$  (équations (1.73), (1.74)) et l'emplacement des galaxies dans le volume de reconstruction. Au bruit de mesure près, ces trois niveaux doivent être cohérents : il faut trouver les distances pour lesquelles l'écart entre les vitesses reconstruites et observées soit minimal, sous la contrainte que les statistiques du champ reconstruit soient données par les fonctions de corrélations  $\Psi$  et  $\zeta$ . L'idée est la suivante :

- À partir de l'estimation initiale du champ de vitesse sur une grille, on tire une estimation des vitesses aux positions estimées des galaxies ;
- on remplace ainsi les galaxies à partir de ces vitesses particulières et de leur décalage vers le rouge avec l'Équation (1.47) ;
- on reconstruit le nouveau champ de vitesse ;
- on boucle jusqu'à convergence.

Statistiquement, cette façon de faire est un échantillonnage dit *de Gibbs*. Cet échantillonnage consiste à tirer au hasard chacun des paramètres du problème à partir de leur probabilité conditionnelle, en fixant tous les autres paramètres. Notre problème contient  $2N + 2 + M^3$  paramètres : les décalages vers le rouge ( $N$ ), les mesures de modules de distance ( $N$ ), la dispersion due aux non-linéarités (1), la constante de Hubble (1) et le champ reconstruit sur une grille ( $M^3$ ). Le modèle bayésien qui s'ensuit et que nous allons décrire ici a été développé par Lavaux (2016) en parallèle de notre étude préliminaire décrite section 2.1. Ce modèle rejoint nos attentes puisqu'il permet d'estimer les paramètres sus-mentionnés dans un contexte bayésien. Nous choisissons donc d'adapter cette méthode à notre cas particulier, l'étude des données de *CosmicFlows*.

#### 3.1.2 Vraisemblance

Pour construire la fonction de vraisemblance il est nécessaire de modéliser les observations (modules de distance et décalages vers le rouge) à partir des paramètres du problème. Les paramètres sont les suivants :

- les distances lumineuses des galaxies  $d$  (*Attention, contrairement au chapitre précédent,  $d$  ne dénote plus la distance dérivée du module de distance, mais est un paramètre ajusté*) ;
- le champ de vitesse (linéaire et tridimensionnel) reconstruit sur une grille  $\mathbf{v}_g$  ;

- la constante de Hubble  $H_0$ , ou la constante de Hubble effective réduite  $h_{\text{eff}}$  qui sera introduite dans la section 3.1.2.1 ;
- et la dispersion  $\sigma_{\text{NL}}$  due aux non-linéarités.

Tandis que les observations sont :

- les modules de distances  $\mu$  avec erreur  $\sigma_\mu$  ;
- les décalages vers le rouge  $z$  avec erreur  $\sigma_z$ .

Les paramètres et données du modèle sont résumés dans la Table 3.1. Nous décrivons ci-après la construction de la fonction de vraisemblance  $\mathcal{L}$ . À chaque fois, les limites de la modélisation sont décrites dans un paragraphe séparé.

### 3.1.2.1 Modules de distance

Le module de distance d'une galaxie est directement lié à la distance lumineuse via l'Équation (1.46). En considérant que l'erreur sur la mesure du module de distance  $\mu$  est gaussienne avec une déviation standard notée  $\sigma_\mu$ , la probabilité conditionnelle de  $\mu$ ,  $\mathcal{P}(\mu|d)$ , s'écrit :

$$\mathcal{P}(\mu|d) = \mathcal{N}\left(5 \log \frac{d}{10 \text{ pc}}, \sigma_\mu^2\right). \quad (3.1)$$

Cependant, le lien entre le module de distance et la distance lumineuse n'est connu qu'à une constante près : le point zéro de la calibration de l'indicateur de distance considéré. Si celui-ci peut être connu, il est néanmoins notable qu'il est fortement corrélé à la valeur de la constante de Hubble qui est également inconnue. Ainsi doit-on supposer la valeur de l'un ou de l'autre mais pas des deux. Nous choisissons ici de fixer la valeur de  $H_0$  et de laisser le point zéro varier. Nous introduisons une constante de Hubble effective réduite  $h_{\text{eff}}$  telle que :

$$\bar{\mu} = 5 \log \frac{d}{10 \text{ pc}} - 5 \log h_{\text{eff}}. \quad (3.2)$$

Ce paramètre permet de modéliser la non-compatibilité potentielle du point zéro de l'indicateur de distance avec la valeur choisie de la constante de Hubble. En effet, si on définit  $H_{\text{eff}}$  par  $h_{\text{eff}} = \frac{H_{\text{eff}}}{H_0}$ , on a, au premier ordre :

$$H_0 d = H_0 10^{\frac{\bar{\mu} + 5 \log h_{\text{eff}}}{5} - 5} \quad (3.3)$$

$$= H_0 h_{\text{eff}} 10^{\frac{\bar{\mu}}{5} - 5} \quad (3.4)$$

$$= H_{\text{eff}} 10^{\frac{\bar{\mu}}{5} - 5}. \quad (3.5)$$

La valeur de  $h_{\text{eff}}$  permet donc de fixer la valeur de la constante de Hubble compatible avec les données.

En pratique, nous n'introduisons pas le paramètre  $h_{\text{eff}}$  directement dans la vraisemblance définie par l'équation 3.2. Nous ajoutons plutôt le facteur multiplicatif  $h_{\text{eff}}$  à chaque fois qu'il est nécessaire d'évaluer une quantité dépendante de la distance. Par exemple, à partir de la valeur  $d$  de la distance d'une galaxie, la vitesse particulière reconstruite pour cette galaxie  $\mathbf{v}_g(dh_{\text{eff}})$  sera évaluée sur la grille à la distance  $dh_{\text{eff}}$ .

À partir de la vraie distance lumineuse  $d$  et des positions angulaires des galaxies, que nous supposons mesurées parfaitement, on peut déduire la position spatiale des galaxies, notée  $\mathbf{r}$ .

TABLE 3.1 – Données et paramètres utilisés dans le modèle développé dans le chapitre 3.

Données	Description	Nombre	Valeurs pour CF3 et <i>a priori</i>
$N$	Nombre de galaxies et groupes	1	$N = 11483$
$L$	Côté du volume de reconstruction	1	$L = 800 \text{ Mpc}/h_{75}$
$M$	Taille de la grille de reconstruction	1	$M = 128$
$\mu$	Modules de distance	$N$	Distribués normalement avec erreur $\sigma_\mu$ .
$\sigma_\mu$	Erreurs sur le module de distance	$N$	
$z$	Décalages vers le rouge	$N$	Distribués normalement avec erreur $\sigma_z$
$\sigma_z$	Erreurs sur les décalages vers le rouge	1	$\sigma_z = \frac{\sigma_{az}}{c} = 1.7 \times 10^{-4}$
Paramètres ajustés			
$\mathbf{v}_g$	Champ de vitesse linéaire échantillonné sur une grille	$M^3$	<i>a priori</i> $\Lambda$ CDM défini par l'équation (3.14)
$d$	Distances lumineuses	$N$	<i>a priori</i> empiriques définis par les équations 3.17 et 3.16 et/ou uniformes
$h_{\text{eff}}$	Constante de Hubble effective réduite	1	<i>A priori</i> uniforme sur $[0.5, 1.5]$
$\sigma_{\text{NL}}$	Dispersion gaussienne modélisant l'écart des observations à la linéarité	1	<i>A priori</i> uniforme sur $[50, 1500] \text{ km/s}$
$(a, b, c)$	Hyperparamètres définissant la forme des fonctions d' <i>a priori</i> sur les distances	3	<i>A priori</i> uniformes

Ceci permet de modéliser l'observation du module de distance en connaissant la vraie distance sous-jacente. Nous discutons ci-dessous les limites de cette modélisation.

**Limite : indicateur de distance et calibration** Le modèle de mesure de module de distance défini par l'Équation 3.1 est en réalité très simplifié. En effet, il repose sur le fait que les erreurs sur les modules de distance sont gaussiennes, ce qui est probablement faux pour au moins deux raisons. La principale provient des erreurs sur la calibration de l'indicateur de distance utilisé. Par exemple, pour la relation de Tully-Fisher, les modules de distances sont déduits des vitesses de rotation des galaxies selon la loi linéaire calibrée  $M = a + b(\log \frac{W}{\text{km.s}^{-1}} - 2.5)$  où  $W$  est la vitesse de rotation de la galaxie déduite du spectre HI. Si dans notre modèle l'erreur sur le point zéro  $a$  est modélisée avec l'introduction du paramètre  $h_{\text{eff}}$ , l'erreur sur la pente  $b$  ne l'est pas. Cette erreur étant systématique dans le catalogue, elle n'est pas modélisable par une erreur statistique. Par exemple, dans Sorce et al. (2013), la pente mesurée pour la relation de Tully-Fisher (TFR) des données Spitzer est  $b = 9.13 \pm 0.22$ . L'erreur (supposée gaussienne) qui s'ensuit sur le module de distance est :

$$\sigma_{\mu}^b = 0.22 \left| \log \frac{W}{\text{km.s}^{-1}} - 2.5 \right| \quad (3.6)$$

En supposant une valeur raisonnable comme :  $\log \frac{W}{\text{km.s}^{-1}} - 2.5 = 0.5$ , on obtient  $\sigma_{\mu}^b = 0.11$ . Cette erreur dépend également de la vitesse de rotation de la galaxie.

La deuxième raison de la non-gaussianité du module de distance est que l'erreur sur la mesure de la vitesse de rotation  $V$  n'est pas propagée individuellement. Cette erreur (supposée gaussienne) est :

$$\sigma_{\mu}^W = \frac{9.13}{\ln 10} \frac{\sigma_W}{W} \quad (3.7)$$

Typiquement,  $\sigma_{\mu}^W \sim 0.37$ . Cette erreur dépend aussi de la vitesse de rotation de la galaxie, et décroît avec celle-ci. Enfin, l'erreur sur la mesure de l'inclinaison de la galaxie n'est pas prise en compte, et celle-ci dépend aussi de la valeur de la vitesse de rotation. Au final, l'erreur sur la mesure d'un module de distance dépend de plusieurs paramètres : vitesse de rotation, couleur, etc. Ces paramètres étant corrélés au champ de surdensité (les grosses galaxies rouges se trouvant préférentiellement dans les surdensités), il est probable qu'attribuer une erreur gaussienne identique à tous les modules de distance mésestime l'erreur sur les galaxies en fonction de leur environnement. Nous tenterons dans le Chapitre 5 d'inclure la calibration de la TFR dans le modèle présenté ici. En outre, chaque indicateur de distance dépend de propriétés spécifiques, et l'analyse devrait être faite en tenant en compte ces spécificités.

### 3.1.2.2 Décalages vers le rouge

Le décalage vers le rouge  $z$  d'une galaxie résulte de la combinaison du décalage vers le rouge cosmologique et de l'effet Doppler dû à la vitesse particulière, combinaison décrite par l'équation 1.47. Si l'on suppose que l'erreur sur le décalage vers le rouge est gaussienne de

déviations standard notées  $\sigma_z$ , la probabilité  $\mathcal{P}(z|\mathbf{v}_g, d, \sigma_{\text{NL}}, H_0, \mathbf{r})$  est définie par<sup>1</sup>

$$\mathcal{P}(z|\mathbf{v}_g, d, \sigma_{\text{NL}}, H_0, \mathbf{r}) = \frac{(1 + \bar{z})^{-1}}{\sqrt{2\pi (\sigma_{\text{NL}}^2 + \sigma_{cz}^2 (1 + \bar{z})^{-2})}} \exp\left(-\frac{1}{2} \frac{(v^r(z, d) - \mathbf{v}_g(\mathbf{r}) \cdot \hat{\mathbf{r}})^2}{\sigma_{\text{NL}}^2 + \sigma_{cz}^2 (1 + \bar{z})^{-2}}\right) \quad (3.9)$$

Dans l'Équation 3.9 certaines dépendances sont implicites : la fonction  $v^r$  dépend de la cosmologie (et donc de  $H_0$ ) et le décalage vers le rouge cosmologique  $\bar{z}$  dépend de  $H_0$  et de la distance lumineuse  $d$ . La probabilité conditionnelle écrite ci-dessus repose sur une modélisation des écarts à la linéarité des vitesses observées comme distribution gaussienne de valeur moyenne nulle et de déviation standard  $\sigma_{\text{NL}}$ .

**Limite : le modèle de non-linéarité** Les vitesses particulières observées à faible décalage vers le rouge ne tracent pas uniquement le champ de vitesse linéaire. Comme la formation des structures devient une théorie non linéaire à partir d'une valeur de la surdensité  $|\delta| \ll 1$ , le modèle linéaire est une approximation. Pour modéliser cet écart, il est d'usage d'ajouter une dispersion gaussienne autour du champ linéaire, telle que la probabilité d'observer une vitesse particulière  $v^r$  (de fait *non linéaire*) à la position  $\mathbf{r}$  de ce champ  $\mathbf{v}$  (linéaire) soit :

$$\mathcal{P}(v^r|\mathbf{v}) = \mathcal{N}(\mathbf{v}(\mathbf{r}) \cdot \hat{\mathbf{r}}, \sigma_{\text{NL}}^2) \quad (3.10)$$

L'avantage de cette modélisation est la possibilité d'estimer le champ de vitesse via l'algorithme de Hoffman-Ribak (voir 2.1). Cependant, il faut garder à l'esprit que cette modélisation est une bonne approximation pour les galaxies de champ, pour lesquelles la non-linéarité est homogène et isotrope, mais devient mauvaise au fur et à mesure que l'on sonde les galaxies localisées près des fortes non-linéarités (les amas notamment). En effet, dans ces régions, les non-linéarités deviennent dépendantes d'au moins l'échelle et la taille de l'amas, ce qui rend donc obsolète l'utilisation d'une unique dispersion. Le spectre de puissance non linéaire n'est plus isotrope, et donc non diagonal dans l'espace de Fourier, ce qui complique drastiquement l'analyse.

Dans le cadre de *CosmicFlows*, la stratégie est de "grouper" les données, comme il sera expliqué dans la section 4.1.4. Ceci a l'avantage de s'affranchir des non-linéarités dans les amas ou groupes de galaxies. Cependant, utiliser des données groupées risquent de biaiser le champ reconstruit. En effet, prendre la moyenne sur un groupe des décalages vers le rouge pour estimer la vitesse particulière dudit groupe, comme décrit par l'Équation (4.9), revient à négliger les covariances d'ordre supérieur à zéro entre ces vitesses. Il est de ce fait probable que des corrélations linéaires soient négligées en procédant ainsi. De plus, le degré d'approximation fait en groupant les galaxies dépend de la technique de construction des groupes. En prenant par exemple une limite de groupe défini par  $nR_{\text{vir}}$ ,  $n$  étant un entier et  $R_{\text{vir}}$  le rayon du Viriel de l'amas, les corrélations linéaires négligées dépendront de  $n^2$ . Peu de modèles sont proposés

1. Le facteur  $(1 + \bar{z})^{-2}$  provient de :

$$\partial_z \left( \frac{z - \bar{z}}{1 + \bar{z}} \right) = (1 + \bar{z})^{-1}. \quad (3.8)$$

2. Pour tester cette hypothèse, il conviendrait de fabriquer des catalogues simulés réalistes et appliquer une technique pour grouper les données en faisant varier le  $n$  ci-dessus. Malheureusement, dans le cadre de la

alternativement. Lavaux (2016) propose par exemple de modéliser le champ par un typage des galaxies et un modèle de  $k$  mélanges gaussiens :

$$\mathcal{P}(v^r|\mathbf{v}, t) = \mathcal{N}(\mathbf{v}(\mathbf{r}) \cdot \hat{r}, \sigma_{\text{NL},t}^2) \quad (3.11)$$

où  $t \in [0, k]$  est le type de la galaxie. Nous n'utiliserons toutefois pas ce modèle ici.

### 3.1.2.3 Vraisemblance

En combinant les équations (3.9) et (3.1), et en factorisant sur toutes les galaxies, indexées par la lettre  $i$ , nous obtenons la fonction de vraisemblance suivante :

$$\begin{aligned} \mathcal{L}(\mathcal{M}, \mathcal{Z}|\mathcal{D}, h_{\text{eff}}, \sigma_{\text{NL}}, \mathbf{v}) = & \prod_i \frac{(1 + \bar{z}_i(d_i h_{\text{eff}}))^{-1}}{2\pi\sigma_{\mu,i} \left( \sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i(d_i h_{\text{eff}}))^{-2} \right)} \\ & \exp \left( -\frac{1}{2} \left( \frac{\left( \mu_i - 5 \log_{10} \frac{d_i}{10 \text{ pc}} \right)^2}{\sigma_{\mu,i}^2} \right) - \frac{1}{2} \left( \frac{(v^r(z, \bar{z}_i(d_i h_{\text{eff}})) - \mathbf{v}_g(d_i h_{\text{eff}} \hat{\mathbf{r}}_i) \cdot \hat{\mathbf{r}}_i)^2}{\sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i(d_i h_{\text{eff}}))^{-2}} \right) \right) \end{aligned} \quad (3.12)$$

où  $\mathcal{M} = \{\mu_i\}_{i \in [1, N]}$ ,  $\mathcal{Z} = \{z_i\}_{i \in [1, N]}$  et  $\mathcal{D} = \{d_i\}_{i \in [1, N]}$ . Le premier terme donne le lien entre les distances ajustées et les modules de distance mesurés tandis que le deuxième terme donne le lien entre les vitesses particulières calculées à partir des distances et le champ de vitesse reconstruit. Afin d'estimer la probabilité postérieure des paramètres du modèle  $\mathcal{P}(\mathcal{D}, h_{\text{eff}}, \sigma_{\text{NL}}, \mathbf{v}|\mathcal{M}, \mathcal{Z})$  en utilisant le théorème de Bayes, nous devons attribuer des *a priori* à ces paramètres.

### 3.1.2.4 Graphe orienté acyclique (*directed acyclic graph* ou DAG)

Une méthode pour calculer la vraisemblance (3.12) est l'utilisation de graphes orientés acycliques, ou DAG. Un DAG peut représenter un modèle hiérarchique Bayésien en traçant les liens statistiques entre les différents paramètres du modèle. Le DAG pour le modèle de ce chapitre est présenté dans la figure 3.1. Dans cette figure, chaque noeud représente une distribution de probabilité.

1.  $\mathcal{P}(d'|d, h_{\text{eff}}) = \delta_D(d' h_{\text{eff}} d)$ ,
2. see (1.44),
3.  $\mathcal{P}(v^l|\mathbf{v}, d') = \delta_D(v^l - \mathbf{v}(d' \hat{\mathbf{r}}))$ ,
4.  $v^{\text{NL}} \sim \mathcal{N}(v^l, \sigma_{\text{NL}}^2)$ ,
5.  $\mathcal{P}(z'|\bar{z}, v^{\text{NL}}) = \delta_D(v^{\text{NL}} - c \frac{z' - \bar{z}}{1 + \bar{z}})$ ,
6.  $z \sim \mathcal{N}(z', \sigma_z^2)$ .

---

collaboration *CosmicFlows*, les techniques pour grouper les données ne sont pas systématisées ni répétables. On note cependant qu'il existe d'autres algorithmes qui ont comme paramètre libre la longueur de lien ("linking length", en anglais).

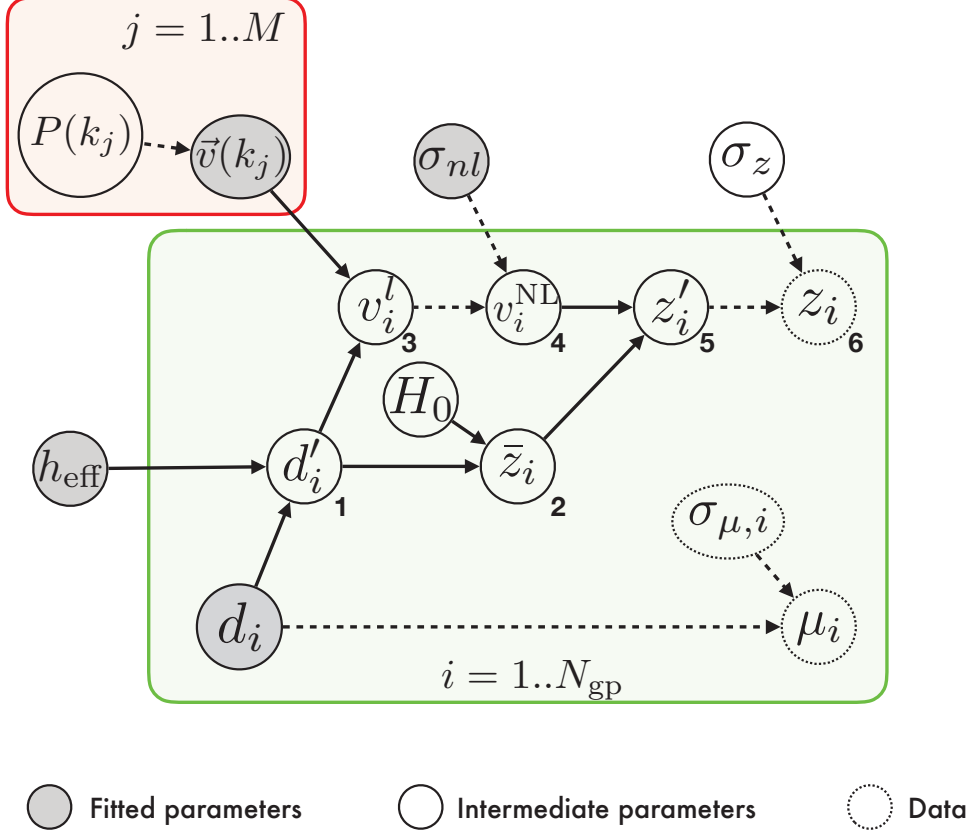


FIGURE 3.1 – Graphe orienté acyclique du modèle. Les numéros correspondent aux distributions de probabilités définies dans le texte.

La fonction de vraisemblance est définie par :

$$\mathcal{L} = \prod_i \int dd'_i dz'_i dv_i^{\text{NL}} dv_i^l d\bar{z}_i \mathcal{P}(z_i|z'_i) \mathcal{P}(z'|z_i, v_i^{\text{NL}}) \mathcal{P}(v_i^{\text{NL}}|v_i^l, \sigma_{\text{NL}}) \mathcal{P}(v_i^l|\mathbf{v}, d'_i) \mathcal{P}(d'_i|d_i, h_{\text{eff}}) \quad (3.13)$$

qui donne par intégration directe l'équation (3.12).

### 3.1.3 A priori bayésiens

#### 3.1.3.1 Sur le champ de vitesse

Du fait de la reconstruction du champ de vitesse sur une grille, le nombre de paramètres ajustés (environ  $128^3$ ) est bien supérieur à la quantité de données disponibles (environ  $2 \times 11000$ ). Il y a donc *a priori* un manque d'information au problème. L'autre source d'information qui est utilisée dans ce modèle provient du modèle  $\Lambda$ CDM décrit Section 1.3.1. En effet, le modèle prédit précisément la corrélation entre les vitesses particulières de deux galaxies

en fonction de leur position. Mathématiquement, cette supposition s'exprime sous la forme d'un *a priori* (voir Section 1.3.1) sur la transformée de Fourier  $\hat{\delta}(\mathbf{k})$  du champ de surdensité reconstruit :

$$\mathcal{P}(\hat{\delta}(\mathbf{k})|\Lambda\text{CDM}) = \pi(\delta) = \prod_{\mathbf{k}} \mathcal{N}(0, P(k)); \quad (3.14)$$

où  $P(k)$  est le spectre de puissance des fluctuations de matière. On rappelle que cet *a priori* est valide dans le domaine linéaire, ce qui n'est pas le cas à  $z = 0$ . Le prolongement du modèle pour prendre en compte le spectre de puissance non linéaire est un projet complexe et est discuté dans la section 6.1 du dernier chapitre.

Cet *a priori* sur le champ de surdensité est directement transposable en *a priori* sur le champ de vitesse en utilisant les corrélations  $\Psi$  définies par l'équation (1.73).

### 3.1.3.2 Sur les distances lumineuses

Comme décrit dans la Section 2.3, les effets de sélection en distance provoquent l'apparition de biais tels les biais de Malmquist homogène et inhomogène. Il y a en général deux régimes de sélection différents dans un catalogue de distances extragalactiques. Le premier, proche, est l'effet de volume. Puisque la répartition des galaxies est en première approximation homogène, la probabilité d'observer une galaxie à une distance  $d$  croît comme le carré de la distance  $d$  :

$$\mathcal{P}(d) \propto d^2 \quad (3.15)$$

Si aucun *a priori* n'est appliqué aux distances ajustées (ce qui revient à dire qu'un *a priori* uniforme est appliqué), l'effet de volume sera négligé et la distance moyenne (resp. la vitesse particulière) des galaxies sera sous-estimée (resp. surestimée), biaisant le champ reconstruit. Il en va de même à grande distance : à cause des effets de sélection, la probabilité d'observer une galaxie croît avec la luminosité apparente de celle-ci, et donc décroît avec la distance. Idéalement, cet effet est modélisable si l'on connaît la magnitude apparente de chacune des galaxies ainsi que la façon dont elles ont été sélectionnées en fonction de celle-ci. En pratique il est cependant très difficile d'avoir une expression simple (et analytique) des effets de sélection. Dans le cas de la relation de Tully-Fisher par exemple, la sélection est généralement faite dans différentes bandes et s'appuie sur des données extérieures au sondage (dans la majorité des cas, les sondages pour la relation TFR se font par pointage et ne sont pas systématiques). Il y a aussi une sélection sur l'inclinaison des galaxies et la masse d'hydrogène neutre HI. S'ajoute à cela la répartition intrinsèque des luminosités absolues des galaxies qui n'est pas uniforme et est généralement modélisée par une fonction de Schechter (Schechter, 1976).

La modélisation précise de la fonction de sélection d'un échantillon de distances extragalactiques est donc une tâche ardue, et nous nous appuyons donc ici sur une modélisation empirique de ces effets. On considère que la fonction de sélection est une fonction analytique qui remplit les quatre critères suivants :

- la fonction doit être normalisable ;
- croissante à petite distance (effet de volume) ;
- décroissante à grande distance (effet de sélection) ;
- asymétrique.



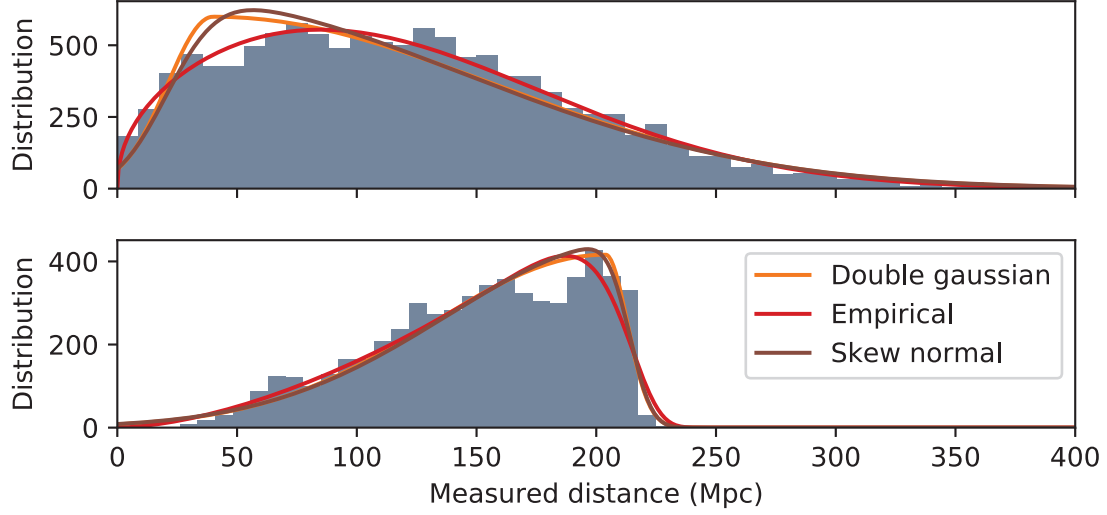


FIGURE 3.2 – Ajustements des fonctions d'*a priori* sur les distances mesurées de CF3 (haut), et sur les distances mesurées à partir du décalage vers le rouge des données de 6dF seulement (bas).

Ainsi considère-t-on trois fonctions ayant ces propriétés et dépendantes de trois paramètres libres  $a$ ,  $b$  et  $c$ .

- Une fonction empirique proposé par Lavaux (2016) :

$$\mathcal{P}^{(1)}(d|a, b, c) = \frac{1}{N(a, b, c)} (d)^a \exp(- (d/b)^c) \quad (3.16)$$

où  $N(a, b, c)$  est un facteur de normalisation ;

- une double gaussienne définie par morceaux :

$$\mathcal{P}^{(2)}(d|a, b, c) = \frac{1}{\sqrt{2\pi}(b+c)} \begin{cases} \exp\left(-\frac{1}{2}\frac{(d-a)^2}{b^2}\right) & \text{si } d_i \leq a \\ \exp\left(-\frac{1}{2}\frac{(d-a)^2}{c^2}\right) & \text{sinon.} \end{cases} \quad (3.17)$$

- et une loi gaussienne asymétrique :

$$\mathcal{P}^{(3)}(d|a, b, c) = \frac{1}{\sqrt{2\pi}b^2} \exp\left(-\frac{1}{2}\frac{(d-a)^2}{b^2}\right) \left(1 + \operatorname{erf}\left(c\left(\frac{d-a}{b\sqrt{2}}\right)\right)\right) \quad (3.18)$$

où  $\operatorname{erf}$  est la fonction d'erreur.

La figure 3.2 montre l'ajustement de ces trois fonctions empiriques sur les distances mesurées de CF3 (panneau du haut) et sur les distances inférées par les décalages vers le rouge de l'échantillon 6dF. Cette figure montre la flexibilité des fonctions d'*a priori* proposées.

### 3.1.3.3 Sur les autres paramètres

Aux autres paramètres, nommément la dispersion des vitesses due aux non-linéarités  $\sigma_{\text{NL}}$ , la constante de Hubble effective réduite  $h_{\text{eff}}$  et les paramètres des *a priori* sur les distances  $(a, b, c)$ , sont attribués des *a priori* uniformes. Ce choix, le plus simple et direct, n'est pas forcément optimal. On pourrait par exemple analyser indépendamment la fonction de sélection et imposer des *a priori* plus informatifs sur les coefficients  $(a, b, c)$ . On pourrait également imposer que la dispersion due aux effets non linéaires soit proche d'une certaine valeur (typiquement 150 km/s) afin de maximiser le poids des données, etc. L'échantillonnage de Gibbs étant très flexible, il n'est pas exclu de modifier ces *a priori* dans le futur.

## 3.2 Algorithme

Nous décrivons ici comment le modèle précédent est échantillonné. Le but est de maximiser la probabilité :

$$\mathcal{P}(\mathbf{v}_g, \mathcal{D}, \sigma_{\text{NL}}, h_{\text{eff}} | \mathcal{M}, \mathcal{Z}) = \mathcal{L} \times \mathcal{P}(\mathcal{D}) \mathcal{P}(\mathbf{v}_g) \quad (3.19)$$

L'algorithme qui est proposé par Lavaux (2016) est l'échantillonnage de Gibbs par bloc. L'échantillonnage de Gibbs est une méthode Monte Carlo Markov Chain (MCMC) qui permet d'échantillonner une distribution de probabilité. Soient  $n$  paramètres  $\{x_i^k\}$  que l'on veut échantillonner  $m$  fois ( $k \in [0, m]$ ), l'échantillonnage se fait de la façon suivante :

- on initialise les paramètres  $\{x_i^0\}$  ;
- chaque paramètre  $x_i^k$  pour le pas  $k$  est généré à partir de la probabilité conditionnelle  $\mathcal{P}(x_i^k | x_0^k, \dots, x_{i-1}^k, x_{i+1}^{k-1}, \dots, x_n^{k-1})$  ;
- et ce jusqu'à  $k = m$ .

Au final nous avons accès à  $m$  réalisations des  $n$  paramètres  $\{x_i^k\}_{i \in [0, n], k \in [0, m]}$ .

Dans le modèle présenté ici, l'espace des paramètres est :

$$\theta = \{h_{\text{eff}}, \{\mathbf{v}_{g,j}\}, \sigma_{\text{NL}}, \mathcal{D}, \{a, b, c\}\}, \quad (3.20)$$

les indices  $j$  et  $i$  faisant référence aux points de la grille ( $j \in [0, M^3]$ ) et aux traceurs du catalogue ( $i \in [0, N]$ ). Puisqu'on veut tirer successivement, à chaque pas de Markov, chacun des paramètres de sa probabilité conditionnelle, nous détaillons ci-après ces probabilités conditionnelles.

### 3.2.1 Probabilités conditionnelles

Nous donnons dans cette section l'expression des probabilités conditionnelles utilisées pour échantillonner les paramètres de l'ensemble  $\theta$ .

### 3.2.1.1 Constante de Hubble effective $h_{\text{eff}}$

$$\begin{aligned} \mathcal{P}(h_{\text{eff}}|\mathcal{D}, \sigma_{\text{NL}}, \mathbf{v}_g) &= \frac{\mathcal{L} \times \mathcal{P}(\mathcal{D})\mathcal{P}(\mathbf{v}_g)}{\int \mathcal{L} \times \mathcal{P}(\mathcal{D})\mathcal{P}(\mathbf{v})d\mathcal{D}d\mathbf{v}_gdh_{\text{eff}}d\sigma_{\text{NL}}} & (3.21) \\ &\propto \prod_i \frac{1}{\left(\sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i)^{-2}\right)} \exp\left(-\frac{1}{2} \left(\frac{(v^r(z, \bar{z}_i) - \mathbf{v}_g(d_i h_{\text{eff}} \hat{\mathbf{r}}_i) \cdot \hat{\mathbf{r}}_i)^2}{\sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i)^{-2}}\right)\right) \end{aligned}$$

### 3.2.1.2 Vitesses $\mathbf{v}_g$ échantillonnées sur la grille de reconstruction

$$\begin{aligned} \mathcal{P}(\mathbf{v}_g|\mathcal{D}, \sigma_{\text{NL}}, h_{\text{eff}}) &= \frac{\mathcal{L} \times \mathcal{P}(\mathcal{D})\mathcal{P}(\mathbf{v}_g)}{\int \mathcal{L} \times \mathcal{P}(\mathcal{D})\mathcal{P}(\mathbf{v})d\mathcal{D}d\mathbf{v}_gdh_{\text{eff}}d\sigma_{\text{NL}}} & (3.22) \\ &\propto \mathcal{P}(\mathbf{v}_g) \prod_i \frac{1}{\left(\sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i)^{-2}\right)} \exp\left(-\frac{1}{2} \left(\frac{(v^r - \mathbf{v}_g(d_i h_{\text{eff}} \hat{\mathbf{r}}_i) \cdot \hat{\mathbf{r}}_i)^2}{\sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i)^{-2}}\right)\right) \end{aligned}$$

Échantillonner cette distribution revient à calculer une réalisation contrainte du champ  $\mathbf{v}_g$  sous la contrainte donnée par les vitesses particulières radiales. On utilise l'algorithme de Hoffman-Ribak décrit dans la Section 2.1. Notons qu'il est possible d'échantillonner à la fois le champ de vitesse  $\mathbf{v}_g$  et le champ de surdensité  $\delta_g$ . En pratique, nous échantillonons les deux à partir d'une réalisation aléatoire commune  $\delta^{RR}$ .

### 3.2.1.3 Dispersion gaussienne $\sigma_{\text{NL}}$ due aux non-linéarités

La probabilité conditionnelle pour  $\sigma_{\text{NL}}$  est donnée par l'Équation (3.21) où cette fois  $h_{\text{eff}}$  est fixé et  $\sigma_{\text{NL}}$  est laissé libre.

### 3.2.1.4 Distances ajustées $\mathcal{D}$

Dans ce modèle, les probabilités conditionnelles pour les distances  $\mathcal{D} = \{d_i\}$  sont toutes indépendantes. Les corrélations entre ces distances se font via les paramètres de vitesse de la grille  $\mathbf{v}_g$ . La probabilité conditionnelle pour une distance s'écrit :

$$\mathcal{P}(d_i|\mathcal{D}, \sigma_{\text{NL}}, \mathbf{v}_g) = \frac{\mathcal{L} \times \mathcal{P}(\mathcal{D})\mathcal{P}(\mathbf{v}_g)}{\int \mathcal{L} \times \mathcal{P}(\mathcal{D})\mathcal{P}(\mathbf{v})d\mathcal{D}d\mathbf{v}_gdh_{\text{eff}}d\sigma_{\text{NL}}} \quad (3.23)$$

$$\propto \mathcal{P}^{(p)}(d_i) \times \mathcal{L} \quad (3.24)$$

où  $p = (1, 2, 3)$  correspond à la fonction d'*a priori* sur les distances définie par les Équations (3.16), (3.17) et (3.18).

### 3.2.2 Marginalisation et algorithme de Gibbs effondré

On note que les paramètres  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$  sont très corrélés avec le champ reconstruit, ce qui peut mener à des temps de convergence très longs. Ainsi, plutôt que de tirer ces paramètres de leurs probabilités conditionnelles complètes, il est plus judicieux de marginaliser sur les valeurs du champ de vitesse sur la grille. Les probabilités conditionnelles s'écrivent alors :

$$\begin{aligned} \mathcal{P}(h_{\text{eff}}|\mathcal{D}, \sigma_{\text{NL}}) &= \int d\mathbf{v}_g \mathcal{P}(h_{\text{eff}}|\mathcal{D}, \sigma_{\text{NL}}, \mathbf{v}_g) \\ &\propto \int d\mathbf{v}_g \mathcal{P}(\mathcal{D}|h_{\text{eff}}, \sigma_{\text{NL}}, \mathbf{v}_g) \frac{\mathcal{P}(\mathcal{D})}{\mathcal{P}(h_{\text{eff}})} \\ &\propto \mathcal{N}(\mathbf{v}^r(h_{\text{eff}}); \mathbf{0}, \mathbf{C}_{v^r v^r}(h_{\text{eff}})) \end{aligned} \quad (3.25)$$

où  $\mathbf{C}_{v^r v^r}$  est la matrice d'autocorrélation des vitesses particulières radiales et est définie par :

$$\mathbf{C}_{v^r v^r} = \hat{\mathbf{r}} \cdot \Psi(h_{\text{eff}} d \hat{\mathbf{r}}) \cdot \hat{\mathbf{r}} + \sigma_{\text{NL}}^2 \mathbf{1} + \left( \frac{c\sigma_z}{1 + \bar{z}} \right)^2 \mathbf{1}. \quad (3.26)$$

La dernière égalité dans l'équation (3.25) provient de la probabilité marginale du système décrit par l'Équation (2.16). De la même façon, on a :

$$\begin{aligned} \mathcal{P}(\sigma_{\text{NL}}|\mathcal{D}, h_{\text{eff}}) &= \int d\mathbf{v}_g \mathcal{P}(\sigma_{\text{NL}}|\mathcal{D}, h_{\text{eff}}, \mathbf{v}_g) \\ &\propto \mathcal{N}(\mathbf{v}^r; \mathbf{0}, \mathbf{C}_{v^r v^r}(\sigma_{\text{NL}})). \end{aligned} \quad (3.27)$$

Cette méthode d'échantillonnage est l'échantillonnage de Gibbs effondré (*collapsed* en anglais) et se fait par bloc :

- On tire d'abord tous les paramètres marginalisés de leurs probabilités conditionnelles (ici,  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$ ),
- on tire le paramètre sur lequel on a marginalisé (ici,  $\mathbf{v}_g$ ),
- on tire le reste des paramètres de leurs probabilités conditionnelles.

Toutefois, il faut prendre garde à la cohérence du modèle. Dans le cas d'un champ de vitesse reconstruit sur une grille comme c'est le cas ici, l'écart entre les vitesses radiales déduites des distances  $d$  et le champ de vitesse reconstruit  $\mathbf{v}_g$  dépend d'une part de l'écart des données à la non-linéarité mais aussi d'effets de discrétisation. Si la grille n'est pas assez fine certaines corrélations à petites échelles seront négligées dans l'estimation de la vitesse particulière en certains points (par interpolation linéaire par exemple). Ces approximations ne seront pas propagées dans le terme  $\sigma_{\text{NL}}$  si celui-ci est tiré de sa distribution marginalisée. Il y aura ainsi une incohérence entre le modèle d'estimation des distances via l'équation (3.23) et l'estimation de  $\sigma_{\text{NL}}$  via l'équation (3.27). En pratique, les tailles de grille raisonnables considérées dans la suite sont de l'ordre de  $128^3$  pour une largeur de  $L = 1$  Gpc.

### 3.2.3 Implémentation en PYTHON

L'algorithme décrit ci-dessus nécessite l'inversion de grandes matrices pour l'évaluation des matrices  $\mathbf{C}_{v^r v^r}$  dans les équations (3.25) et (3.27) (typiquement de dimension  $N^2$ ) et

l'évaluation des fonctions de corrélation pour chaque paire entre les traceurs et les points de la grille. Il a été implémenté en PYTHON et sera rendu public. Les principales bibliothèques utilisées sont NUMPY, SCIPY, MATPLOTLIB et PANDAS. Nous détaillons ici quelques spécificités d'implémentation.

### 3.2.3.1 Inversion de matrices

Pour échantillonner les densités de probabilité (3.27) et (3.25) ainsi que pour l'algorithme de Hoffman-Ribak, il est nécessaire d'inverser la matrice  $\mathbf{C}_{v^r v^r}$ . La matrice  $\mathbf{C}_{v^r v^r}$  est par construction symétrique, définie positive et à valeurs réelles. De plus, du fait de la forme des fonctions de corrélation de vitesse, la diagonale est numériquement dominante devant les autres termes. Par conséquent, la matrice est toujours inversible. L'algorithme optimal pour inverser une telle matrice est la décomposition de Cholesky, qui décompose une matrice en un produit matriciel de matrices triangulaires :

$$\mathbf{C} = \mathbf{L}\mathbf{L}^T \quad (3.28)$$

où  $\mathbf{L}$  est la matrice triangulaire inférieure de la décomposition. La matrice triangulaire supérieure est notée  $\mathbf{U} = \mathbf{L}^T$ . De l'équation 3.28 on peut directement déduire l'inverse :

$$\mathbf{C}^{-1} = (\mathbf{L}^T)^{-1} \mathbf{L}^{-1} \quad (3.29)$$

L'avantage de la méthode réside dans le fait que l'inversion d'une matrice triangulaire est rapide et peut se faire en place. L'inversion est faite par la bibliothèque LAPACK intégrée à la bibliothèque SCIPY.LINALG.

### 3.2.3.2 Spectre de puissance et fonctions de corrélation

L'algorithme nécessite le calcul du spectre de puissance  $P(k)$  ainsi que les fonctions de corrélation  $\Psi_R$ ,  $\Psi_T$  et  $\zeta$ . Le spectre de puissance est généré à partir du module python CAMB (Lewis and Bridle, 2002)<sup>3</sup> pour une cosmologie donnée. Le code est basé sur les relations d'échelle données par Eisenstein and Hu (1998) et Eisenstein and Hu (1999). Les fonctions de corrélations sont, quant à elles, estimées par intégration numérique directe des intégrales (1.73) et (1.74). Le spectre est renormalisé en fonction de la valeur de  $\sigma_8$  supposée au départ. Cette intégration est faite via le programme IceCore développé par Doumler (2012).

### 3.2.3.3 Grille de vitesse

Le champ est reconstruit sur une grille régulière de dimension  $M^3$ . Le problème se pose d'évaluer le terme  $\mathbf{v}_g(\mathbf{r})$  de l'Équation 3.12 en n'importe quel point du volume reconstruit. L'implémentation faite ici utilise une interpolation trilineaire entre les points de la grille. Ceci biaise le champ puisque les corrélations à une échelle inférieure au pas de la grille ne sont plus données par les fonctions de corrélations  $\Lambda$ CDM. Les conséquences précises de cette interpolation restent à étudier. À noter aussi que l'estimation du champ de vitesse via l'algorithme de

---

3. <http://camb.readthedocs.io/>

Hoffman-Ribak nécessite la génération d'une réalisation aléatoire. Cette réalisation aléatoire se fait en générant un bruit blanc dans l'espace des configurations :

$$w(\mathbf{r}) \sim \mathcal{N}(0, 1) \quad (3.30)$$

puis les modes de Fourier du champ de surdensité aléatoire sont calculés par :

$$\hat{\delta}^{RR}(\mathbf{k}) = \hat{w}(\mathbf{k}) \times \sqrt{P(\mathbf{k})} \quad (3.31)$$

Les modes du champ de vitesse sont calculés à partir de l'Équation (1.33). Enfin, les champs  $\delta$  et  $\mathbf{v}$  sont calculés par transformée de Fourier inverse. Cette procédure implique que les champs de vitesse et surdensité sont à conditions aux limites périodiques. Cependant, les réalisations contraintes, qui sont générées via l'Équation (2.15), n'ont pas cette propriété puisque le champ moyen ne suppose pas de condition aux limites périodiques.

### 3.3 Tests

#### 3.3.1 Paramètres globaux et convergences

Nous testons l'implémentation de l'algorithme sur le catalogue simulé défini dans la Section 2.2. Les champs de vitesse et de surdensité ont été reconstruits sur 2000 pas de Markov dans une boîte de côté  $L = 500$  Mpc et une grille de pas  $M = 64$ . Le mode maximal de cette grille est  $\frac{2\pi M}{L} = 0.8 \text{ Mpc}^{-1} > k_{\text{max}}$  qui est supérieur au mode le plus élevé utilisé pour générer la réalisation aléatoire pour créer le catalogue simulé.

Bien que la reconstruction ait été calculée sur 2000 pas, ces pas peuvent être corrélés entre eux et le nombre d'échantillons indépendants de la distribution postérieure n'est pas égal au nombre de pas de la chaîne. Pour étudier cet effet, nous traçons l'autocorrélation des chaînes pour les paramètres globaux  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$  en fonction de la longueur d'autocorrélation sur la figure 3.3. L'autocorrélation d'une chaîne de longueur  $N$  pour une longueur de corrélation  $\tau$  est définie par :

$$c(\tau) = \frac{1}{N - \tau} \sum_{i=1}^{N-\tau} (f_i - \bar{f})(f_{i+\tau} - \bar{f}) \quad (3.32)$$

La première intersection de l'autocorrélation avec zéro correspond à la longueur de première décorrélation, et donne approximativement la longueur des chaînes indépendantes.

Dans le cas présent, on peut considérer que la chaîne MCMC contient une centaine d'échantillons indépendants. Les valeurs des paramètres  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$  le long de la chaîne sont tracées sur la figure 3.4 et les histogrammes correspondants sur la figure 3.5. Nous rappelons que les valeurs utilisées pour générer les données simulées sont  $h_{\text{eff}} = 1$  et  $\sigma_{\text{NL}} = 150 \text{ km s}^{-1}$ . On retrouve, comme attendu, les valeurs de  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$  initiales à moins de une déviation standard. On note qu'on a utilisé ici l'échantillonnage de Gibbs effondré défini en section 3.2.2, c'est-à-dire qu'on a utilisé les probabilités conditionnelles marginalisées sur le champ  $\mathbf{v}_{\mathbf{g}}$  pour échantillonner les paramètres  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$ . Pour voir l'effet de cette marginalisation, on trace sur la figure 3.6 l'évolution de la chaîne avec (avant le pas  $p = 1000$ ) et après marginalisation. On voit que la convergence du paramètre  $h_{\text{eff}}$  est beaucoup plus rapide, ce qui suggère une grande

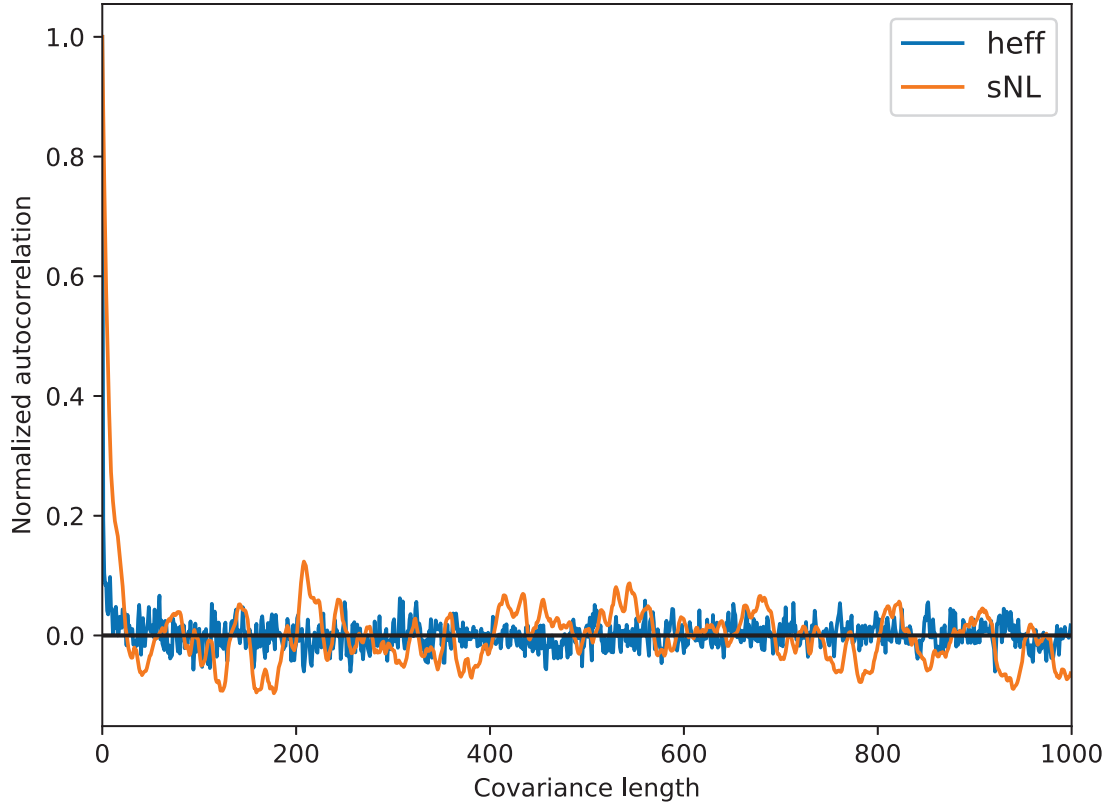


FIGURE 3.3 – Autocorrélation normalisée  $c(\tau)/c(0)$  des chaînes pour les paramètres  $\sigma_{\text{NL}}$  et  $h_{\text{eff}}$  en fonction de la longueur d'autocorrélation.

corrélation avec le champ de vitesse reconstruit et donc une longueur de décorrélation beaucoup plus longue sans la marginalisation. Concernant le paramètre  $\sigma_{\text{NL}}$ , il n'y a visuellement pas d'amélioration notable.

Finalement, on montre aussi sur la figure 3.7 l'ajustement de la fonction d'*a priori* et les distances ajustées.

### 3.3.2 Champs de vitesse et de surdensité

On vérifie d'abord que le champ de vitesse reconstruit est bien irrotationnel comme le suppose la théorie linéaire. Ceci est garanti par l'équation (1.33) qui s'écrit dans l'espace de Fourier :

$$\mathbf{v}(\mathbf{k}) = iHf \frac{\mathbf{k}}{k^2} \delta(\mathbf{k}) \quad (3.33)$$

et qui donne :

$$\mathbf{k} \times \mathbf{v}(\mathbf{k}) \propto \mathbf{k} \times \mathbf{k} = \mathbf{0}. \quad (3.34)$$

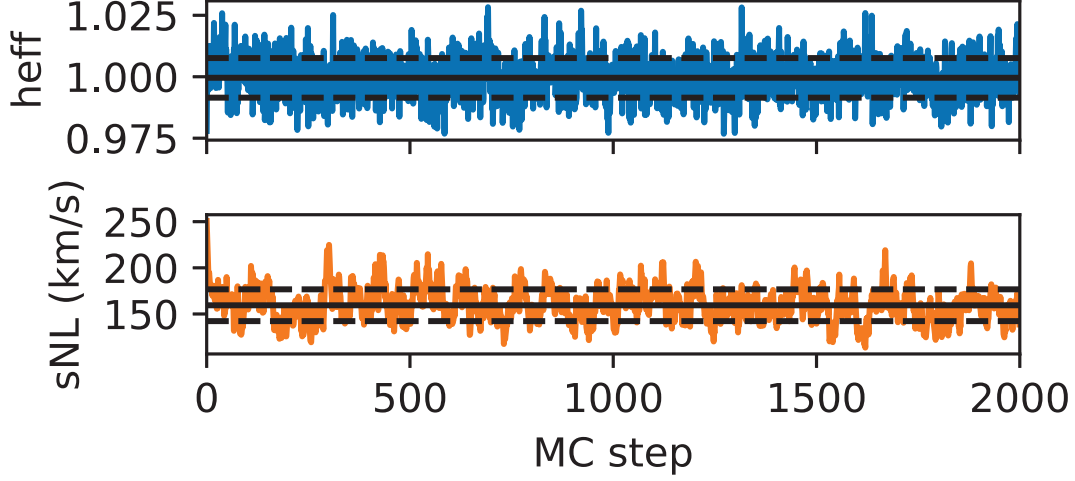


FIGURE 3.4 – Évolution des chaînes pour les paramètres  $\sigma_{\text{NL}}$  et  $h_{\text{eff}}$  dans le cadre de la reconstruction du catalogue simulé. Les paramètres utilisés pour générer le catalogue simulé sont  $h_{\text{eff}} = 1$  et  $\sigma_{\text{NL}} = 150$  km/s. Les traits pointillés et plein donnent les 26<sup>e</sup>, 50<sup>e</sup> et 84<sup>e</sup> centiles des distributions.

Cependant, la reconstruction étant faite dans l'espace des configurations, l'utilisation d'interpolations pour évaluer les fonctions de corrélation n'assure pas ce caractère irrotationnel. On vérifie donc *a posteriori* que le champ de vitesse est bien irrotationnel en pratique. Pour ce faire, il est possible d'utiliser la décomposition de Gibbs-Helmholtz. Le théorème de Helmholtz montre que tout champ vectoriel  $\mathbf{F}$  peut être décomposé en deux parties, rotationnelle et irrotationnelle, telles que :

$$\begin{aligned} \mathbf{F} &= \mathbf{F}_{\Phi} + \mathbf{F}_A \\ &= -\nabla\Phi + \nabla \times \mathbf{A} \end{aligned} \quad (3.35)$$

Dans l'espace de Fourier, cette équation s'écrit :

$$\hat{\mathbf{F}}(\mathbf{k}) = -i\mathbf{k} \cdot \hat{\mathbf{F}}_{\phi}(\mathbf{k}) + i\mathbf{k} \times \hat{\mathbf{F}}_A(\mathbf{k}) \quad (3.36)$$

où

$$\hat{\mathbf{F}}_{\phi}(\mathbf{k}) = i \frac{\mathbf{k} \cdot \hat{\mathbf{F}}(\mathbf{k})}{k^2} \quad (3.37)$$

$$\hat{\mathbf{F}}_A(\mathbf{k}) = i \frac{\mathbf{k} \times \hat{\mathbf{F}}(\mathbf{k})}{k^2} \quad (3.38)$$

Les normes de  $\mathbf{F}_{\{\Phi, A\}}$  dans la tranche  $z=0$  de la reconstruction, sont montrées sur la figure 3.8 dans le cas  $\mathbf{F} = \mathbf{v}$ . On voit que, hormis sur les bords, la contribution du champ



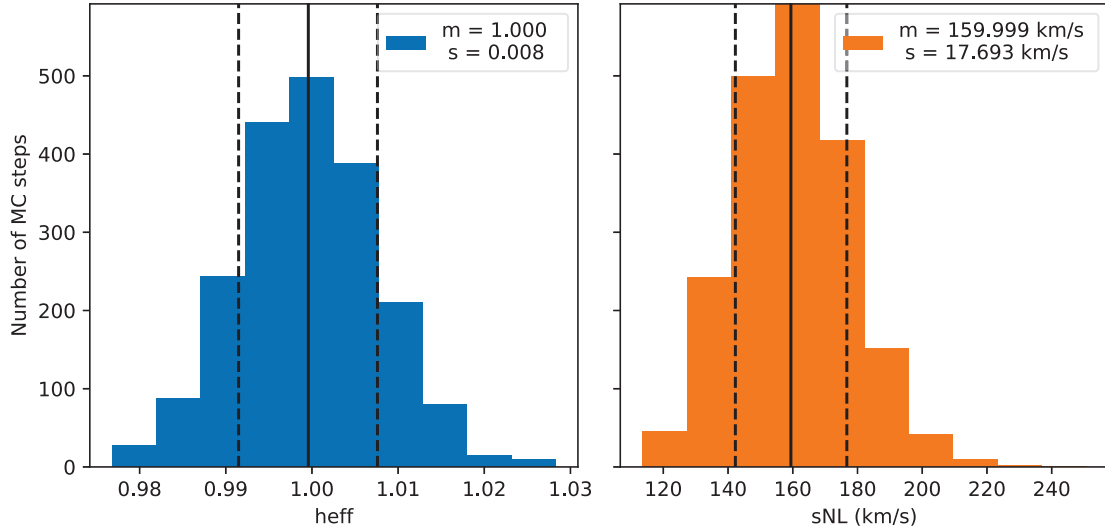


FIGURE 3.5 – Histogrammes des distributions postérieures des paramètres  $\sigma_{\text{NL}}$  et  $h_{\text{eff}}$  dans le cadre de la reconstruction du catalogue simulé. Les paramètres initiaux utilisés pour générer la simulation sont  $h_{\text{eff}} = 1$  et  $\sigma_{\text{NL}} = 150$  km/s. La valeur numérique  $m$  (resp.  $s$ ) donne la moyenne (resp. la déviation standard) des distributions. Les traits pointillés et pleins donnent les 26<sup>e</sup>, 50<sup>e</sup> et 84<sup>e</sup> centiles des distributions.

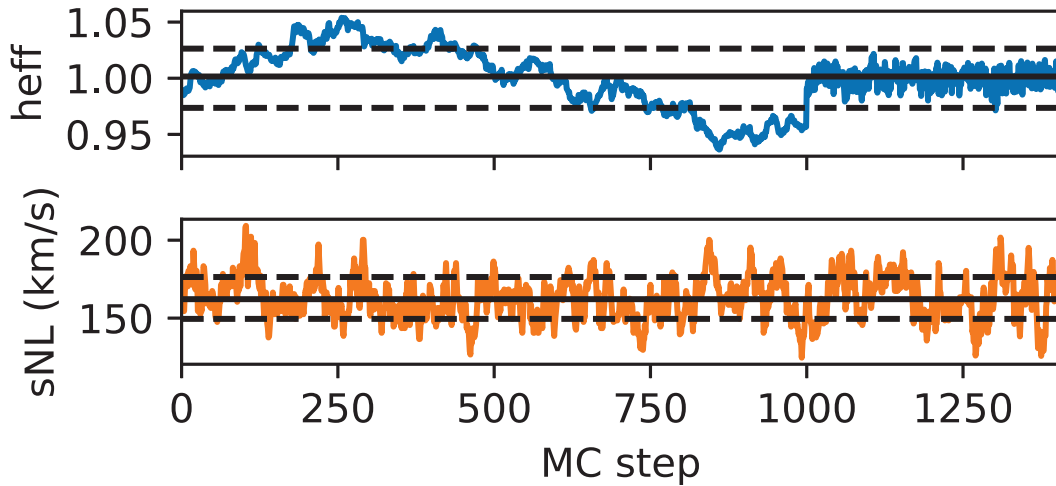


FIGURE 3.6 – Évolution des chaînes pour les paramètres  $\sigma_{\text{NL}}$  et  $h_{\text{eff}}$  dans le cadre de la reconstruction du catalogue simulé. À partir du pas  $p = 1000$ , les deux paramètres ont été tirés de leurs distributions marginalisées sur les vitesses reconstruites  $v_g$ .

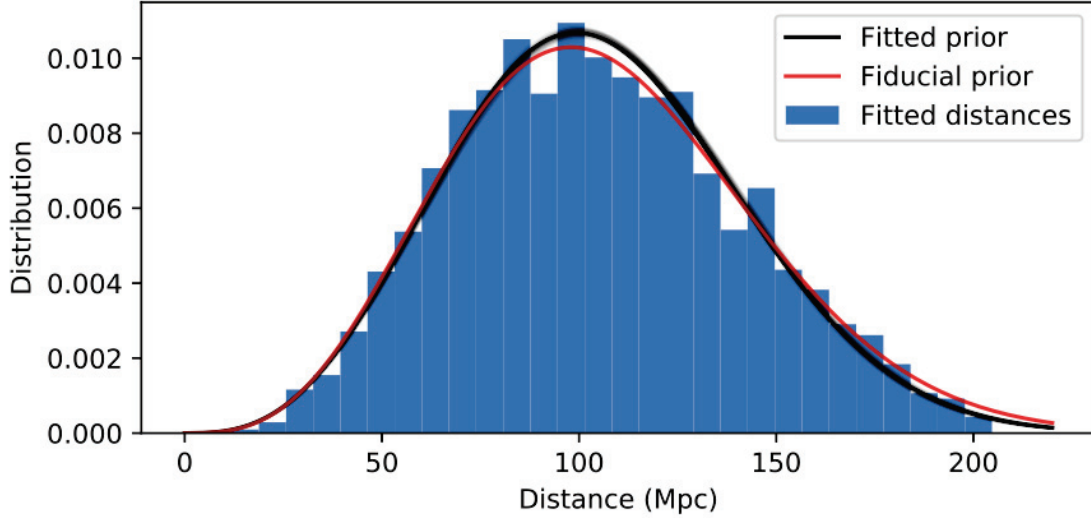


FIGURE 3.7 – Distribution des distances ajustées par le modèle. La courbe rouge représente l'*a priori* initial. La courbe noire montre l'*a priori* ajusté et la dispersion associée.

irrotationnel est négligeable. Sur les bords, la transformée de Fourier discrète impose des conditions aux limites périodiques, ce qui affecte les champs de vitesse rotationnel et irrotationnel calculés à partir de l'équation (3.37).

Finalement, la reconstruction est montrée en figure 3.9. Les figures 3.10 et 3.11 montrent la comparaison entre les vitesses radiales reconstruites et celles de la simulation. On voit que la reconstruction est non biaisée, d'abord par le fait que le *pull* est proche d'une normale unitaire, ensuite parce que les densités de probabilité sont toutes compatibles avec la droite  $x = y$ , même sur l'ensemble du volume de reconstruction. La comparaison entre les figures 3.10 (cas présent) et 2.11 (cas des erreurs gaussiennes où les distances sont parfaitement connues), montre que les reconstructions sont quasiment identiques et le manque d'information lié à l'incertitude sur la position des traceurs est fortement réduit par l'utilisation de nos *a priori*. À noter cependant que la distribution du *pull* a une déviation standard de 0.85 plutôt que 1 dans le cas idéal, ce qui suggère une surestimation des erreurs sur la reconstruction.

### 3.4 Conclusion

Dans le chapitre précédent nous avons vu pourquoi la méthode des réalisations contraintes n'est plus adaptée aux nouveaux catalogues de distances extragalactiques, qui par leur extension en distance et en nombre de galaxies sont sujets à de nombreux biais statistiques (biais de Malmquist, distribution log-normales des vitesses etc.). Le principal défaut de la méthode WF/CR réside dans l'absence de modélisation des distances mesurées des galaxies. Ces distances interviennent dans l'estimation des vitesses particulières radiales, l'estimation des fonctions de corrélation entre les galaxies, et l'estimation de la position de la galaxie dans

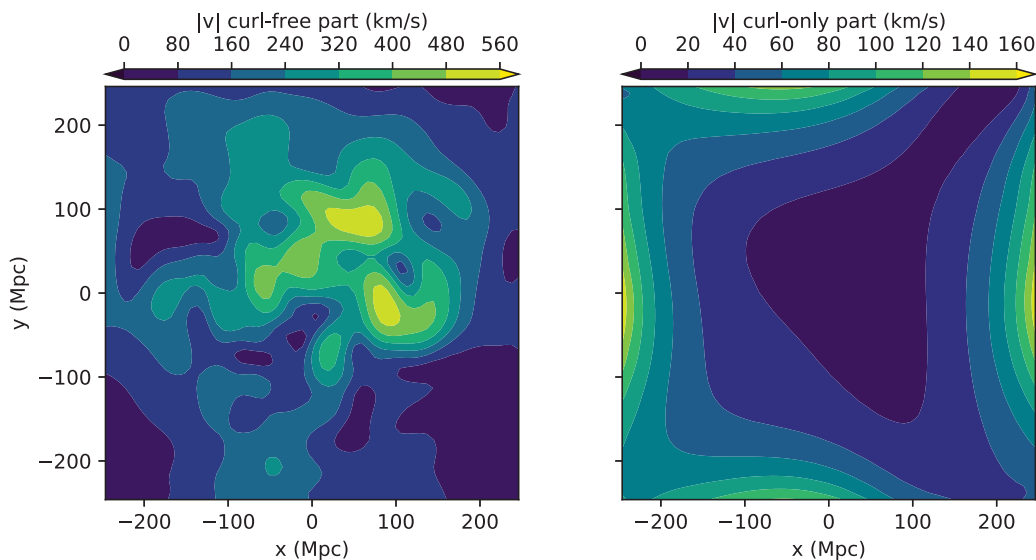


FIGURE 3.8 – Tranche  $z=0$  des normes du champ de vitesse reconstruit. À gauche, la partie irrotationnelle du champ. À droite, la partie purement rotationnelle. Attention, l'échelle des couleurs n'est pas la même sur les deux panneaux.

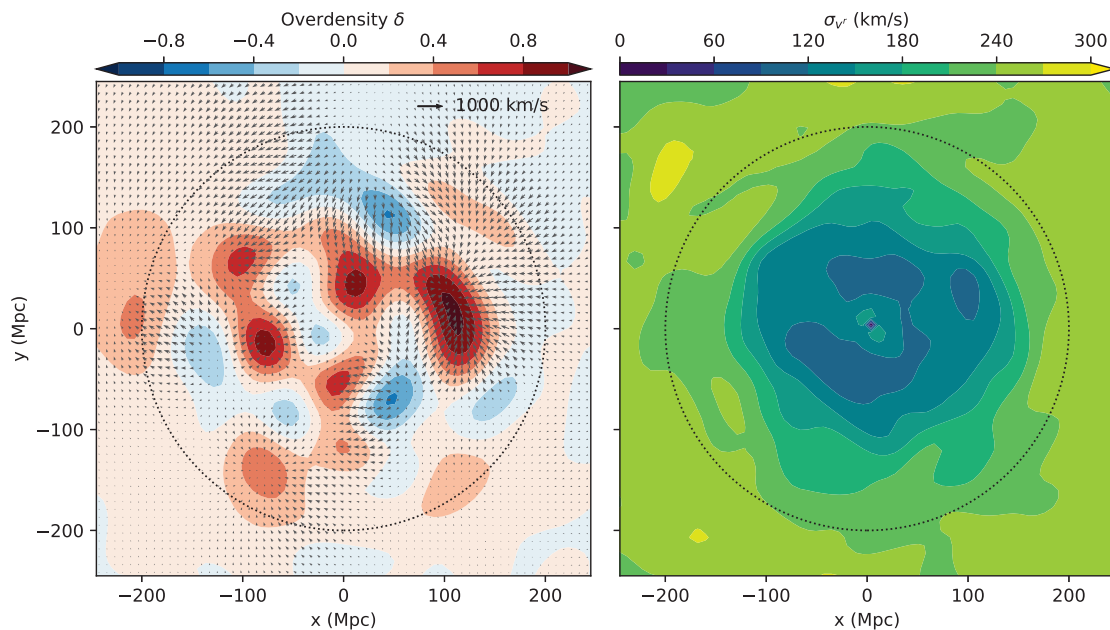


FIGURE 3.9 – Gauche : tranche en  $z=0$  Mpc du champ de surdensité reconstruit (dans le cadre du catalogue simulé) par l'échantillonnage de Gibbs sur 2000 pas de Markov. Droite : déviation standard des vitesses radiales reconstruites. Le cercle pointillé noir montre la limite des données simulées à  $r = 200$  Mpc. Cette figure est à comparer avec la figure 2.2.

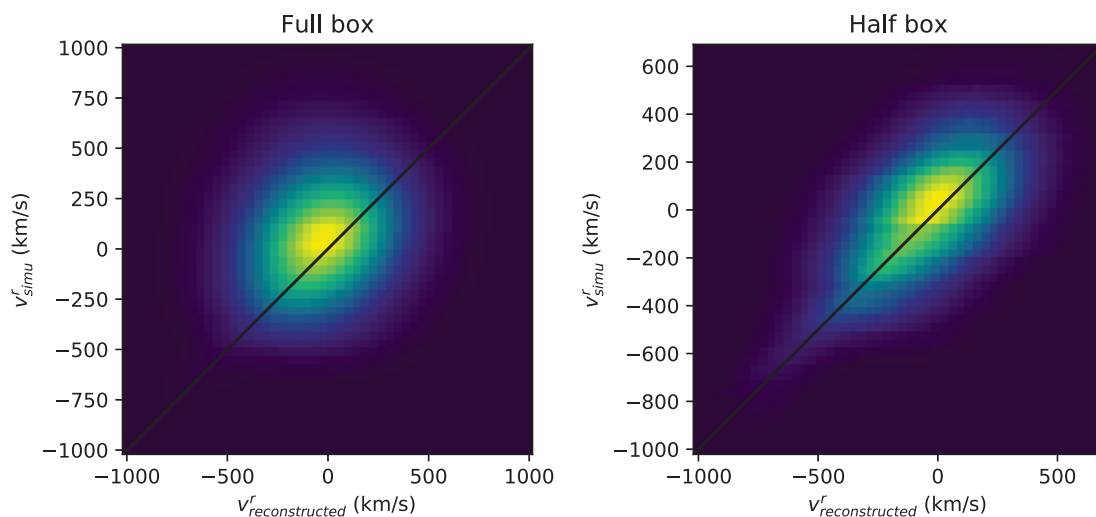


FIGURE 3.10 – Histogramme de la somme des densités de probabilité des vitesses radiales reconstruites par intervalle de vitesse radiale de la simulation. Cas de la reconstruction via l’algorithme de Gibbs. Cette représentation est expliquée dans la section 2.2.

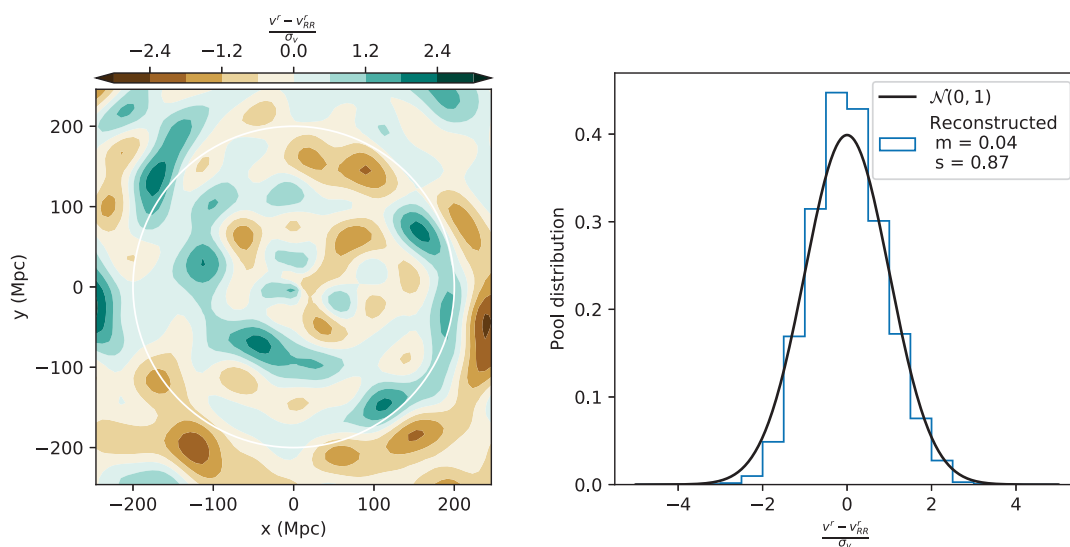


FIGURE 3.11 – Résidus du champ des vitesses radiales entre la reconstruction via l’algorithme de Gibbs et la simulation originale. Les résidus sont définis par l’Équation (2.23). À gauche est représenté la tranche  $z=0$  de ces résidus, et à droite l’histogramme des résidus (le *pull*) sur l’ensemble du volume couvert par les données simulées et limité par le cercle pointillé à  $r = 200$  Mpc.

le volume de reconstruction. L'étude présentée dans le chapitre 2 a motivé l'introduction du modèle bayésien présenté dans ce chapitre.

Nous avons vu dans la section 3.1.2 comment la fonction de vraisemblance était construite et qu'elle prenait en compte l'incertitude sur la mesure des modules de distance en introduisant les paramètres  $d$  représentant la véritable distance sous-jacente aux données. Les *a priori* sur le champ de vitesse sous forme de fonctions de corrélation ont été introduits dans la section 3.1.3 et nous avons vu comment modéliser les effets de sélection sous la forme d'*a priori* sur les paramètres  $d$ . Enfin, nous avons décrit l'algorithme d'échantillonnage de Gibbs dans la section 3.2 et testé la méthode sur un catalogue simulé dans la Section 3.3.

Ce chapitre a finalement montré comment une modélisation bayésienne correcte des observations permet de remonter aux champs initiaux de vitesse et surdensité. En introduisant les distances comme paramètres libres du modèle, l'algorithme est capable de retrouver le champ initial avec la même précision que si ces distances étaient parfaitement connues (voir la section 2.2.2). Naturellement, nous tentons d'appliquer la méthode au catalogue de distances *CosmicFlows-3* dans le chapitre suivant.

# Chapitre 4

## Champ de vitesse de *CosmicFlows-3*

Dans le chapitre 3 nous avons présenté un modèle statistique qui relie les observations de modules de distance et décalages vers le rouge au champ de vitesse linéaire de l'Univers Local. Dans ce chapitre, nous appliquons la méthode au catalogue *CosmicFlows-3*, que nous présentons dans la première section 4.1. Une fois la reconstruction calculée dans la section 4.2.1, nous analysons la cosmographie de l'Univers Local dans la section 4.3. Nous étudions dans la section 4.4 les erreurs systématiques potentielles de la reconstruction, systématiques provenant du modèle de calibration et du choix des *a priori* sur les distances.

### Sommaire

---

<b>4.1</b>	<b>Le catalogue <i>CosmicFlows-3</i></b>	<b>80</b>
4.1.1	Analyse du catalogue	80
4.1.2	Intercalibration	83
4.1.3	Mesure des décalages vers le rouge	84
4.1.4	Groupes	84
<b>4.2</b>	<b>Application de la méthode au catalogue</b>	<b>85</b>
4.2.1	MCMC et convergence	85
4.2.2	Irrotationnalité du champ reconstruit	86
4.2.3	Paramètres ajustés	88
<b>4.3</b>	<b>Cosmographie de l'Univers Local</b>	<b>89</b>
4.3.1	Structures de l'Univers Local	93
4.3.2	Comparaison aux catalogues de décalages vers le rouge	94
4.3.3	La toile cosmique des vitesses	95
4.3.4	Comparaison filtre de Wiener, <i>CosmicFlows-2</i> , <i>CosmicFlows-3</i>	101
4.3.5	Études possibles	103
<b>4.4</b>	<b>Erreurs systématiques</b>	<b>103</b>
4.4.1	Idée générale	103
4.4.2	Influence des <i>a priori</i>	105
4.4.3	Influence de la calibration des indicateurs de distance	107
4.4.4	Dipôle externe	110
<b>4.5</b>	<b>Conclusion</b>	<b>110</b>

---

## 4.1 Le catalogue *CosmicFlows-3*

### 4.1.1 Analyse du catalogue

Comme nous l’avons vu dans le chapitre d’introduction, la mesure des vitesses particulières se fait à partir de deux observations que sont le décalage vers le rouge et la distance. Nous avons vu également que les méthodes de détermination des distances extragalactiques ne donnaient accès qu’au module de distance. C’est donc à partir du décalage vers le rouge et du module de distance que l’analyse débute. À ce dessein, *CosmicFlows-3*<sup>1</sup>(Tully et al., 2016) est le troisième catalogue de données de la collaboration *CosmicFlows*. Il contient 17648 mesures de décalage vers le rouge et autant de mesures de module de distance de galaxies distribuées, en première approximation, uniformément dans le ciel. La Figure 4.1 montre la distribution en module de distance, en position angulaire et en décalage vers le rouge du catalogue. Les histogrammes en traits pointillés correspondent à la version groupée du catalogue dont nous parlerons dans la Section 4.1.4. Sur le panneau de droite est montré l’histogramme des distances estimées à partir du décalage vers le rouge :

$$d_z = \frac{cz}{h100 \text{ km s}^{-1} \text{ Mpc}^{-1}} \quad (4.1)$$

On observe une coupure quasiment nette à une distance de  $\sim 160 \text{ Mpc}/h$ . Ceci est dû à une coupure dans les données dont il sera question ci-après. En revanche, cette coupure n’existe pas pour les distances mesurées comme le montre la courbe violette. Cet écart entre décalages vers le rouge et distances ne peut être expliqué uniquement par les vitesses particulières, et provient presque intégralement des erreurs de mesure sur les distances.

La collaboration *CosmicFlows* joue différents rôles dans la conception d’un tel catalogue :

- la compilation de données déjà existantes ;
- la prise de nouvelles données (exclusivement via la relation de Tully-Fisher) ;
- l’intercalibration de l’ensemble de ces données.

Ainsi le catalogue est-il composé de modules de distance provenant de six méthodologies différentes présentées Section 1.2.3 : relation de Tully-Fisher des galaxies spirales ; plan fondamental des galaxies elliptiques ; relation période-luminosité des Céphéides ; sommet de la branche des géantes rouges ; supernovae de type Ia et fluctuations de brillance de surface. Marginalement, le catalogue comprend d’autres sources de mesure de distances telles que RR Lyrae, branche horizontale des Géantes Rouges et les binaires à éclipse, qui ne seront pas mentionnées ici. Pour un catalogue tel que *CosmicFlows-3*, la classification en origine méthodologique n’est pas suffisante. En effet, les données proviennent en plus de différentes campagnes d’observation par différents télescopes, ce qui complexifie l’analyse. Nous retenons ici une classification en cinq échantillons :

**6dF** Échantillon observé par la collaboration 6dF par la méthode du plan fondamental (Campbell et al., 2014),

**TF-Arecibo** Échantillon dont les mesures de distance par la relation de Tully-Fisher ont été faites via le télescope Arecibo,

---

1. Disponible sur <http://edd.ifa.hawaii.edu>.

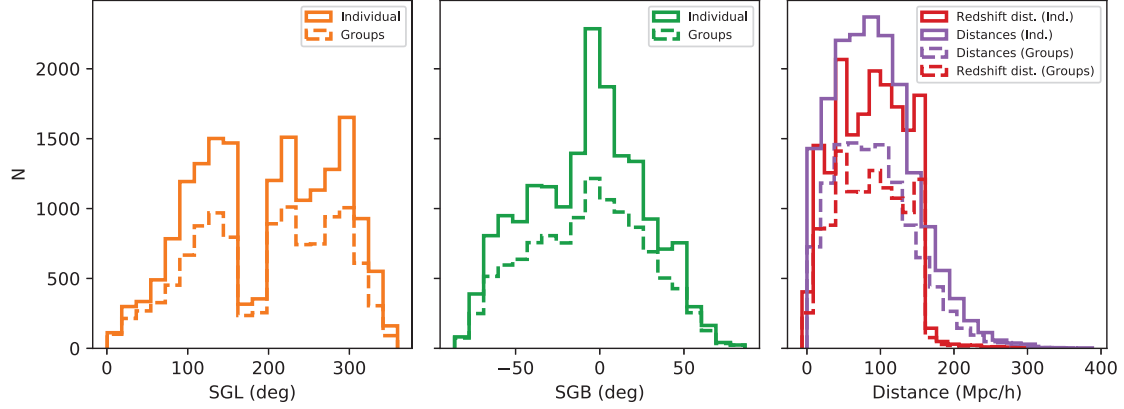


FIGURE 4.1 – Distribution des longitudes (gauche) et latitudes (milieu) extragalactiques du catalogue CF3, ainsi que des distances mesurées (droite) via le module de distance (violet) et le redshift (rouge). Les traits pointillés correspondent à la version groupée du catalogue (voir la section 4.1.4).

**TF-Spitzer** Échantillon dont les mesures de distance par la relation de Tully-Fisher ont été faites via le télescope spatial Spitzer (Sorice et al., 2014),

**TF-Autres** Autres données dont les mesures de distance ont été faites via la relation de Tully-Fisher.

**Autres** Le reste qui contient principalement des mesures proches (Céphéides, sommet de la branche des géantes rouges etc.) et des mesures provenant de la technique du plan fondamental.

La figure 4.2 montre la distribution des décalages vers le rouge pour les différents sous-échantillons décrits ci-dessus. On peut remarquer tout d’abord la coupure présente dans les données de 6dF. Cette coupure, qui était déjà notable dans la Figure 4.1, est faite à  $cz \sim 16\,000 \text{ km s}^{-1}$ . L’échelle des ordonnées montre également que l’échantillon 6dF est le plus conséquent du catalogue avec 8800 galaxies. La principale différence entre les échantillons provenant de la relation de Tully-Fisher est leur portée : elle augmente dans le sens Spitzer, TF-Autres, TF-Arecibo, avec une longue queue de distribution pour ce dernier sous-échantillon. Enfin, le sous-échantillon regroupant le reste des données a une distribution quasiment plate comparée aux autres sous-échantillons. À noter aussi la présence de nombreux groupes locaux (*i.e.* à petite distance) dans cet échantillon. La Figure 4.3 montre la répartition angulaire des galaxies dans une projection de Mollweide en coordonnées équatoriales selon leur appartenance aux sous-échantillons susmentionnés. Elle montre par exemple que les données 6dF sont principalement localisées dans l’hémisphère Sud. Comme nous le verrons plus tard, l’existence d’une corrélation entre l’origine observationnelle d’un point de donnée et sa localisation spatiale pose un problème d’intercalibration.



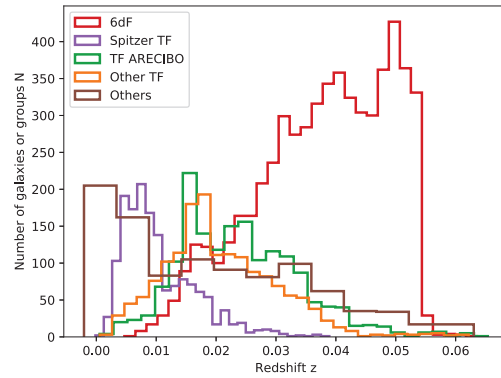


FIGURE 4.2 – Distributions des décalages par le rouge pour les différents sous-échantillons de CF3. De haut en bas : 6dF, TF Spitzer, TF Arcibo, TF Autres, Autres.

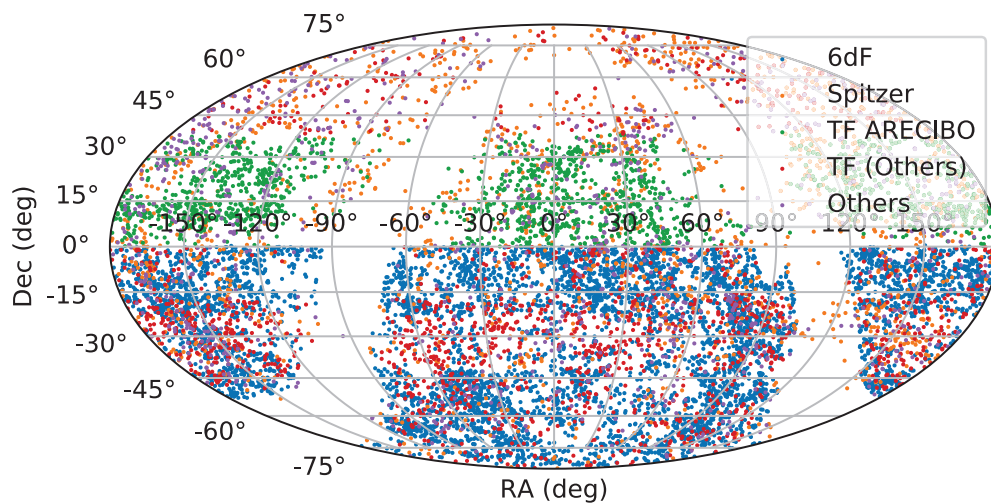


FIGURE 4.3 – Projection du ciel en coordonnées équatoriales. Les points correspondent aux galaxies observées dans CF3, et les couleurs aux échantillons définis dans le texte.

### 4.1.2 Intercalibration

Puisque les modules de distance n'ont pas une origine commune, des désaccords peuvent exister entre les échelles de distance de chacun des échantillons. Ceci s'explique par la différence de calibration de chacun des télescopes, ainsi que par l'erreur de calibration de l'indicateur de distance (voir Section 1.2.3). Tully et al. (2016) proposent d'utiliser les galaxies communes aux différents échantillons afin de les réajuster sur une même échelle de la façon suivante :

- En partant des méthodologies couvrant les galaxies les plus proches, comme la relation période-luminosité des Céphéides ou le sommet de la branche des géantes rouges, on regarde le recouvrement entre les différents échantillons en mesurant la valeur moyenne des différences entre les modules de distance :  $\Delta\mu = \langle \mu^{(1)} - \mu^{(2)} \rangle_{\text{overlap}}$ ,
- on ajuste les modules de distances pour que  $\Delta\mu \sim 0$ ,
- et on réitère pour chaque méthodologie que l'on souhaite incorporer dans le catalogue.

S'ajoute à cela que l'ajustement des modules de distance dépend souvent d'un paramètre tiers. Par exemple, l'intercalibration de l'échantillon 6dF et des données du catalogue *CosmicFlows-2* présente un décalage de  $\Delta\mu = -0.55$  pour toutes les distances également issues de la relation de Tully-Fisher. Pour y remédier, Tully et al. (2016) proposent une correction empirique en fonction du type morphologique  $M_t$  des galaxies :

$$\mu_{6dF} += \begin{cases} -0.042 + 0.085(M_t - 1.6) & \text{si } M_t > 1.6 \\ -0.042 & \text{sinon.} \end{cases} \quad (4.2)$$

Un autre exemple est l'intégration de l'échantillon SFI++ (Springob et al., 2009) dans le catalogue proposé dans Tully et al. (2013). La correction proposée entre les modules de distance de  $\mu_{SFI}^{100}$  où on a supposé  $H_0 = 100$  et ceux de *CosmicFlows-2* est :

$$\mu_{cf2} - \mu_{SFI}^{100} = 0.492 + 0.000031V_{LS} \quad (4.3)$$

où  $V_{LS}$  représente la vitesse de la galaxie dans le référentiel du plan local (LS pour *Local Sheet*).

Ces procédures d'intercalibration empiriques sont des sources d'erreurs systématiques de l'ordre de quelques 0.1 mag, et s'ajoutent au fur et à mesure de l'intégration de nouvelles données dans les catalogues *CosmicFlows*. Il est évident qu'avec l'arrivée de nouvelles données, cette procédure *ad hoc* souffre de non-répétabilité et doit être revue. Idéalement, la calibration de chacun des indicateurs de distance doit être incluse dans le modèle de vitesses particulières utilisé, l'intercalibration se faisant naturellement en supposant dans le modèle qu'une galaxie possède une unique distance réelle. On souligne ici que l'avantage d'un modèle comme ceux proposés dans cette thèse (dans les chapitres 3 et 5) est d'être modulaire (dans sa construction, et dans son échantillonnage). Ainsi, il est tout à fait possible d'implémenter un autre modèle, dépendant de l'indicateur de distance par exemple, pour tirer les distances  $d$  de l'équation (3.1). Nous verrons dans le chapitre 5 comment un tel modèle peut être implémenté dans le cas de la relation de Tully-Fisher.

Pour la suite, nous supposons cependant que l'intercalibration des différentes méthodologies du catalogue CF3 est non biaisée et sans erreurs.

### 4.1.3 Mesure des décalages vers le rouge

Le catalogue *CosmicFlows-3* fournit également des mesures de décalages vers le rouge. Ces mesures proviennent de catalogues déjà construits, notamment du « 2MASS Redshift Survey » (2MRS, Huchra et al. (2012)), du « Lyon Extragalactic DataBase » (LEDA, Makarov et al. (2014)) et « 2M++ » Lavaux and Hudson (2011). En outre, lorsque des mesures de spectre HI utiles à la construction de la relation de Tully-Fisher sont disponibles, les décalages vers le rouge sont mesurés à partir de ceux-ci (Tully, comm. privée).

Les décalages vers le rouge sont en général donnés dans le référentiel héliocentrique, et peuvent être calculés dans un référentiel  $\mathcal{R}$  via l'Équation 1.37 :

$$1 + z_{\mathcal{R}} = \frac{1 + z_{hel}}{1 + z_{\odot/\mathcal{R}}} \quad (4.4)$$

Plutôt que du décalage vers le rouge, il est parfois d'usage de parler de vitesse apparente, en interprétant le décalage vers le rouge comme un effet Doppler :

$$V_{\mathcal{R}} = cz_{\mathcal{R}} \quad (4.5)$$

Ainsi CF3 fournit la vitesse apparente dans les référentiels du CMB ( $V_{cmb}$ ), de la galaxie ( $V_{gsr}$ ), et de la feuille locale (« Local Sheet »,  $V_{LS}$ ). Les transformations de référentiel sont données dans Fixsen et al. (1996) et Tully et al. (2008). Les vitesses apparentes sont données sans erreurs individuelles, sans doute du fait qu'elles sont négligeables devant celles sur les distances. Dans la suite de cette thèse, nous considérons que l'erreur individuelle sur les vitesses apparentes est donnée par :

$$\sigma_{cz} = 50 \text{ km s}^{-1} \quad (4.6)$$

### 4.1.4 Groupes

Le cadre théorique de cette thèse est le modèle  $\Lambda$ CDM associé à la théorie *linéaire* de croissance des structures. Cependant, à  $z \ll 1$ , les champs de surdensité et de vitesse ne sont plus linéaires, et les observations de ce champ ne peuvent pas être directement interprétées dans ce contexte. Ceci est particulièrement vrai dans les groupes de galaxies.

On définit un groupe comme un ensemble de galaxies qui sont en équilibre gravitationnel et dont les mouvements sont de grande amplitude (déviations standard de l'ordre de grandeur de 500 km/s). Ces objets étant très compacts, on peut supposer que toutes les galaxies d'un groupe se situent à un même décalage vers le rouge cosmologique  $\bar{z}$ . L'hypothèse d'équilibre gravitationnel donne :

$$\left\langle \frac{v_i^r}{c} \right\rangle_i = \frac{v_{gp}^{r,lin}}{c} \quad (4.7)$$

$$\left\langle \frac{1 + z_i}{1 + \bar{z}} \right\rangle_i = \frac{1 + \langle z \rangle}{1 + \bar{z}} \quad (4.8)$$

$$\langle z \rangle = \frac{1}{n} \sum_i z_i \quad (4.9)$$

ce qui permet de définir le décalage vers le rouge moyen  $\langle z \rangle$  d'un groupe tel que sa vitesse particulière soit affranchie des non-linéarités internes. Si l'on néglige les corrélations internes, l'erreur sur ce décalage vers le rouge est :

$$\sigma_{\langle z \rangle} = \frac{\sigma_{vr}}{\sqrt{n}} \quad (4.10)$$

De la même façon, le décalage vers le rouge cosmologique  $\bar{z}$  d'un groupe peut être dérivé des mesures individuelles de modules de distance  $\mu_i$  :

$$\bar{z}_{\text{gp}} = \bar{z} \left( \mu_{\text{gp}} = \frac{\sum_i w_i \mu_i}{\sum_i w_i} \right) \quad (4.11)$$

où  $w_i = \frac{1}{\sigma_{\mu,i}^2}$  et la fonction  $\bar{z}(\mu)$  est définie par les équations (1.44) et (1.46).

Il est ainsi possible de dériver un catalogue groupé CF3 : à partir de relevés de décalages vers le rouge, on identifie des groupes de galaxies, puis on attribue aux galaxies CF3 leur appartenance à ces groupes. On définit ensuite le décalage vers le rouge moyen de chaque groupe à partir de l'équation 4.9 appliquée à tous les membres connus (qu'ils soient inclus ou non dans CF3), et le redshift cosmologique moyen à partir de l'équation 4.11.

Plusieurs méthodes d'identification de groupes existent dans la littérature (Kourkchi and Tully, 2017; Tully, 2015; Tempel et al., 2012; Crook et al., 2007) et le catalogue groupé est à l'origine construit à partir de Tully (2015). Dans la version de CF3 utilisée ici, nous avons rajouté en plus les groupes locaux ( $V < 3500$  km/s) décrits dans Kourkchi and Tully (2017). Les distributions résultantes en positions angulaires et distances sont montrées en traits pointillés sur la figure 4.1.

## 4.2 Application de la méthode au catalogue

Dans cette section l'application de la méthode détaillée dans le chapitre 3 au catalogue *CosmicFlows-3* est présentée.

### 4.2.1 MCMC et convergence

Le modèle est appliqué aux données CF3 dans le cadre simple et conservateur suivant :

- la cosmologie pré-supposée est déterminée par  $(H_0, \sigma_8, \Omega_m) = (75 \text{ km s}^{-1} \text{ Mpc}^{-1}, 0.8, 0.3)$  ;
- les calibrations des indicateurs de distance sont cohérents et on ajuste une seule constante de Hubble effective réduite  $h_{\text{eff}}$  ;
- les transformations de référentiel données dans la Section 4.1.3 sont supposées exactes ;
- les non-linéarités sont modélisées par le groupage des galaxies comme présenté dans la Section 4.1.4, et l'écart restant à la non-linéarité est modélisé par une dispersion gaussienne de variance  $\sigma_{\text{NL}}^2$  ;
- les fonctions d'*a priori* sur les distances sont attribuées comme suit aux différents sous-échantillons définis dans la Section 4.1.1. Les distances du sous-échantillon (i) (6dF) sont ajustées avec une loi double gaussienne par morceaux (cf. Éq. 3.17). Les distances des sous-échantillons (ii,iii,iv) sont ajustées avec une loi empirique (cf. Éq. 3.16). Enfin, les distances du sous-échantillon (v) sont supposées distribuées uniformément.

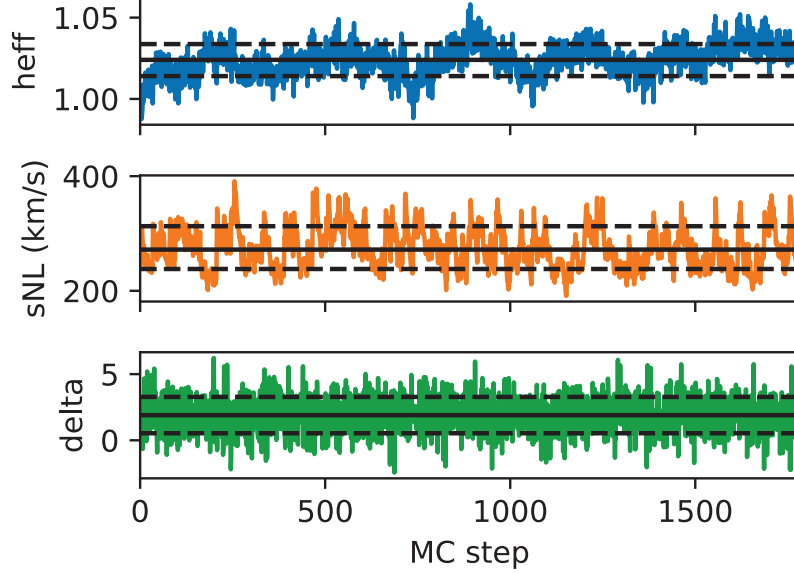


FIGURE 4.4 – Évolution des paramètres  $h_{\text{eff}}$ ,  $\sigma_{\text{NL}}$  et  $\delta_{\text{Virgo}}$  le long de la chaîne de Markov. Les traits pointillés et plein donnent les 26, 50 et 84ème centiles des distributions respectivement.

La reconstruction est faite dans une boîte de largeur  $L = 800 \text{ Mpc}/h_{75}$  et sur une grille de taille  $M^3 = 128^3$ . La chaîne de Markov finale contient 1780 pas. Les chaînes pour les paramètres  $h_{\text{eff}}$ ,  $\sigma_{\text{NL}}$  et  $\delta_{\text{Virgo}}$  sont montrées sur la Figure 4.4.  $\delta_{\text{Virgo}}$  est la surdensité ajustée au point de coordonnées  $(0, 10, 0) \text{ Mpc}/h_{75}$ . Les histogrammes correspondant à ces chaînes sont montrés sur la Figure 4.5. Enfin, la convergence globale est monitorée par la convergence des deux paramètres  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$  dont les autocorrélations des chaînes sont montrées sur la figure 4.6. La Figure 4.6 montre que la longueur de décorrélation pour le paramètre  $h_{\text{eff}}$  est d'environ 100 pas et 30 pas pour  $\sigma_{\text{NL}}$ . Cela signifie que la chaîne de Markov contenant au total 1780 pas comprend (seulement) une vingtaine de pas indépendants pour sonder la distribution du paramètre  $h_{\text{eff}}$ .

Les histogrammes de la Figure 4.5 montrent que les distributions postérieures sont quasiment gaussiennes. Cela est confirmé si l'on trace l'histogramme des quantités normalisées  $\frac{p - \langle p \rangle}{\sqrt{\langle p^2 \rangle - \langle p \rangle^2}}$ , comme montré sur la Figure 4.7.

## 4.2.2 Irrotationnalité du champ reconstruit

Nous testons également l'irrotationnalité du champ comme expliqué dans la section 3.3.2. La Figure 4.8 montre l'intensité relative des parties rotationnelles et irrotationnelles du champ de vitesse reconstruit. On observe que la partie irrotationnelle est négligeable dans la région couverte par les données. Sur les bords, les champs de vitesse calculés ici sont affectés par

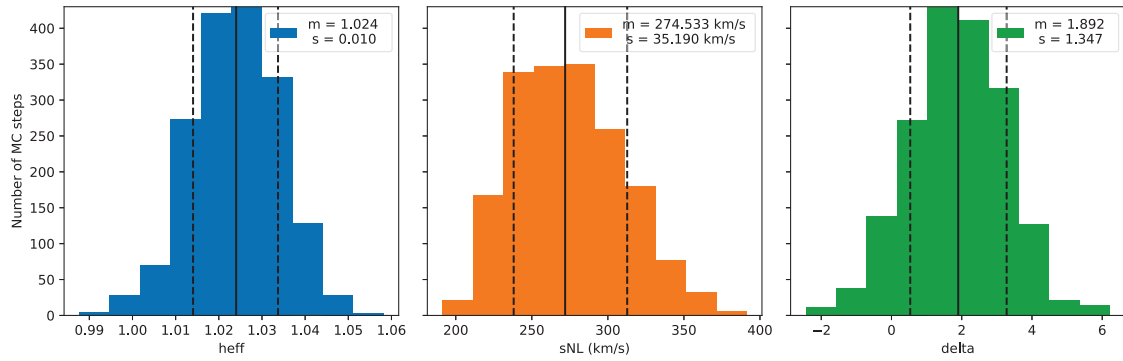


FIGURE 4.5 – Histogramme des paramètres  $h_{\text{eff}}$ ,  $\sigma_{\text{NL}}$  et  $\delta_{\text{Virgo}}$ . Les traits pointillés et plein donnent les 26, 50 et 84ème centiles des distributions respectivement.

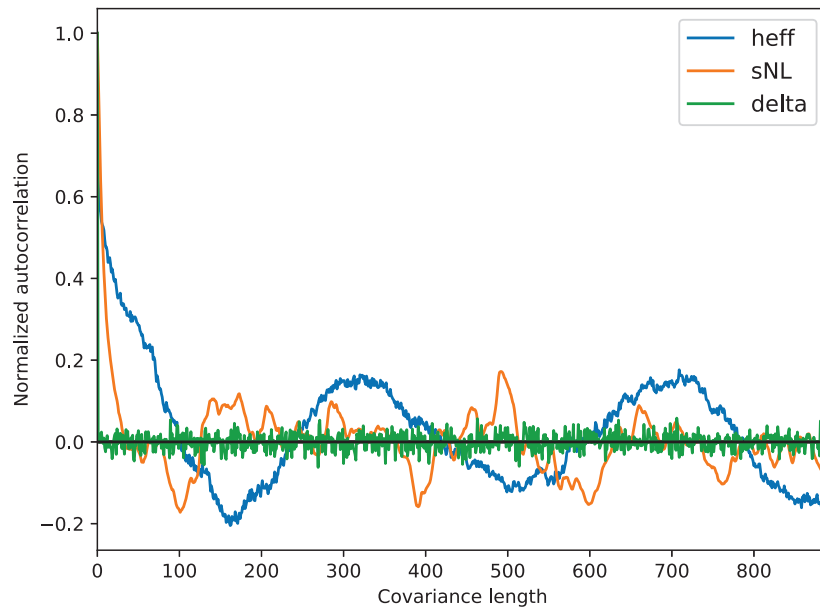


FIGURE 4.6 – Autocorrélation des chaînes des paramètres  $h_{\text{eff}}$ ,  $\sigma_{\text{NL}}$  et  $\delta_{\text{Virgo}}$  en fonction de la longueur de corrélation.

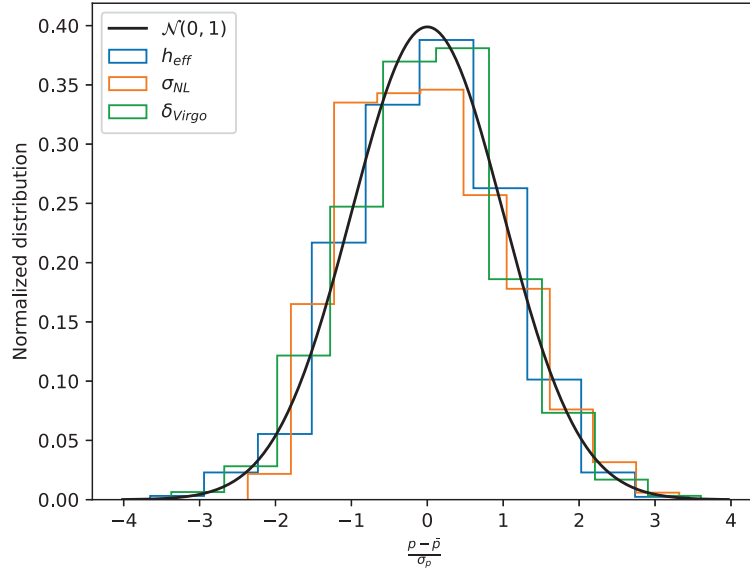


FIGURE 4.7 – Histogrammes normalisés de la distribution postérieure des paramètres  $h_{\text{eff}}$ ,  $\sigma_{\text{NL}}$  et  $\delta_{\text{Virgo}}$ .  $m$  et  $s$  désignent la moyenne des distributions et leur écart type.

les conditions aux limites périodiques via le calcul par transformée de Fourier discrète de l'équation (3.36).

### 4.2.3 Paramètres ajustés

**Constante de Hubble réduite** La constante de Hubble réduite est ajustée à  $h_{\text{eff}} = 1.02 \pm 0.01$ . Cela signifie que les données CF3 calibrées selon Tully et al. (2016) sont compatibles avec la constante de Hubble admise dans le modèle cosmologique initial, ce qui valide l'analyse donnée dans Tully et al. (2016). Cependant, ce résultat est à tempérer par la simplicité du modèle présenté ici. Une discussion sera donnée à ce sujet dans la Section 4.4. Nous verrons que cette valeur dépend notamment des fonctions choisies comme *a priori* sur les distances.

**Dispersion non linéaire** La dispersion due aux non-linéarités est trouvée à  $\sigma_{\text{NL}} = 275 \pm 35$  (stats)  $\text{kms}^{-1}$ . Cette valeur semble élevée en comparaison des valeurs usuelles tirées des simulations (Bistolos and Hoffman, 1998). Comme nous le verrons dans la section 4.4, le modèle de point-zéro sous la forme de  $h_{\text{eff}}$  peut être mis en cause, mais aussi la résolution faible qui revient à négliger les corrélations à petites distances. Aussi, l'interpolation entre les points de la grille pour estimer les vitesses particulières peut être à l'origine d'un bruit supplémentaire comme le montre Howlett et al. (2017). Également, le modèle présenté ici ne prend pas en compte les possibles données aberrantes. Ces données peuvent jouer un rôle

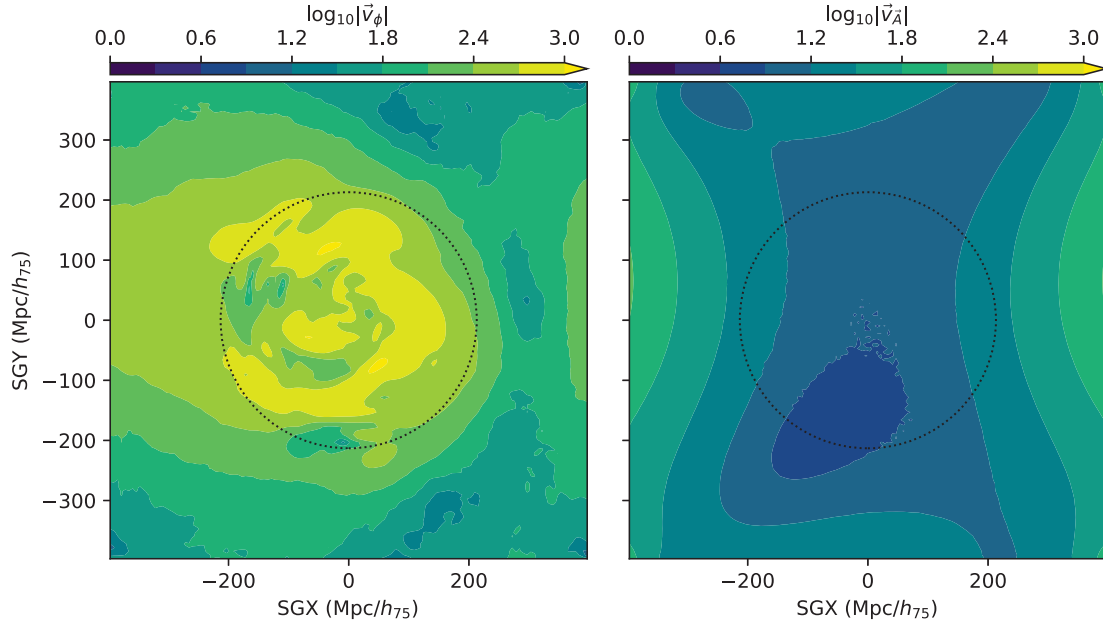


FIGURE 4.8 – À gauche : logarithme de la norme de la partie irrotationnelle du champ de vitesse reconstruit. À droite : logarithme de la norme de la partie rotationnelle du champ de vitesse reconstruit. Les deux graphiques représentent la tranche  $SGZ = 0 \text{ Mpc}/h_{75}$  du volume de reconstruction. Le cercle pointillé noir représente la limite des données.

majeur dans l'estimation du paramètre  $\sigma_{NL}$ . Nous verrons plus tard une solution possible à ce problème.

Dans cette section nous avons étudié les paramètres de convergence et le comportement de la chaîne de Markov de la reconstruction. Nous nous tournons maintenant vers une analyse cosmographique de la reconstruction.

### 4.3 Cosmographie de l'Univers Local

Les tranches  $SGZ = 0 \text{ Mpc}/h_{75}$ ,  $SGX = 0 \text{ Mpc}/h_{75}$  et  $SGY = 0 \text{ Mpc}/h_{75}$  de la reconstruction sont présentées sur les Figures 4.9, 4.11 et 4.10 respectivement. Le champ de surdensité apparaît en couleur, et les flèches noires représentent la projection dans ces plans des vitesses linéaires tridimensionnelles reconstruites. Nous présentons dans cette section une analyse préliminaire de la cosmographie de l'Univers Local à partir de cette reconstruction. Une analyse plus poussée et nécessitant de nouveaux outils de visualisation est en cours de préparation.

Avant de passer à la suite, nous rappelons que les champs de vitesse et surdensité ont également des erreurs statistiques. On trace sur la figure 4.12 les déviations standard des champs dans la tranche  $SGZ = 0$ . Comme nous l'avons vu dans la section 2.2.2, le champ de surdensité est beaucoup plus bruité que le champ de vitesse. Les déviations standard tendent



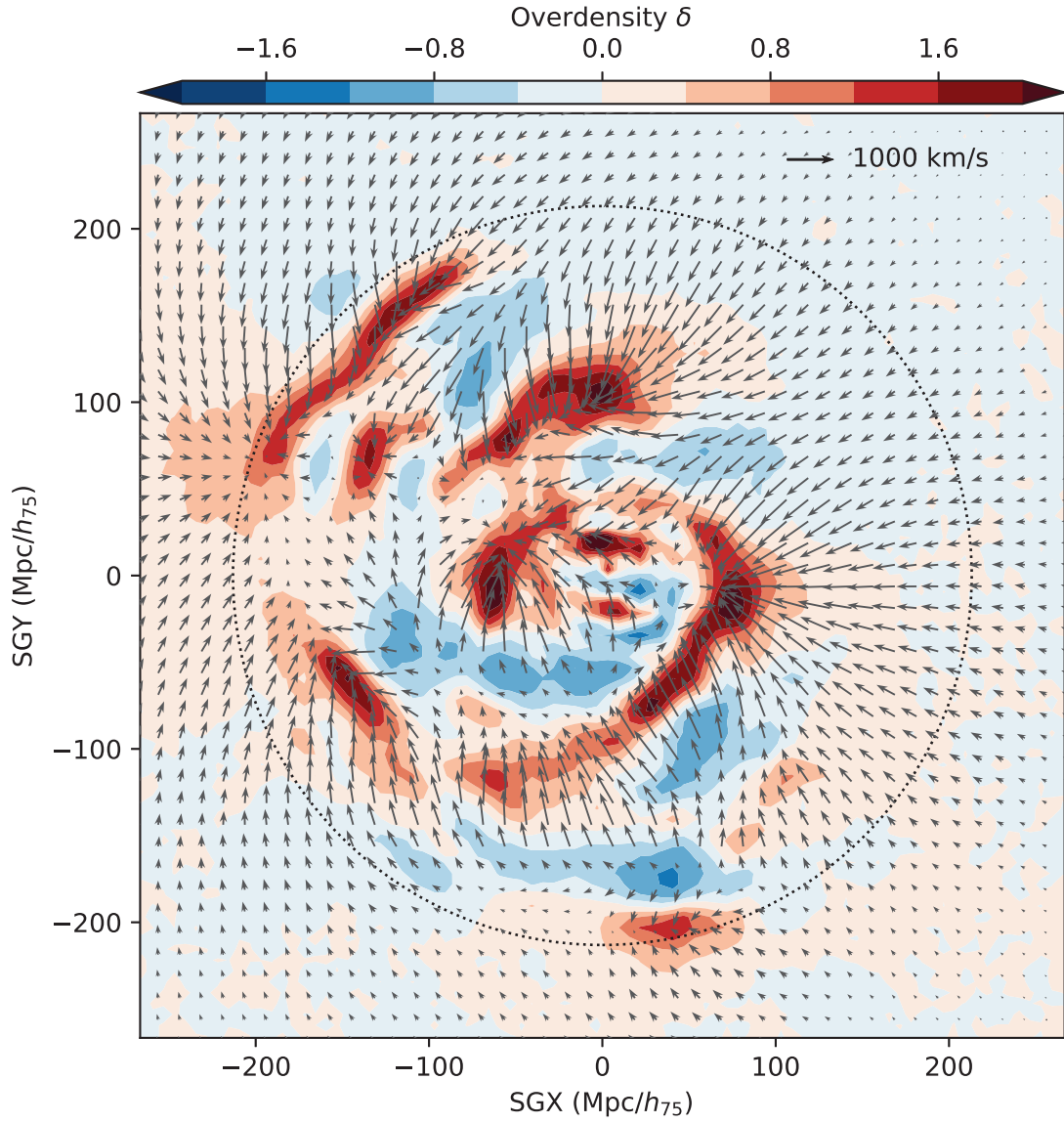
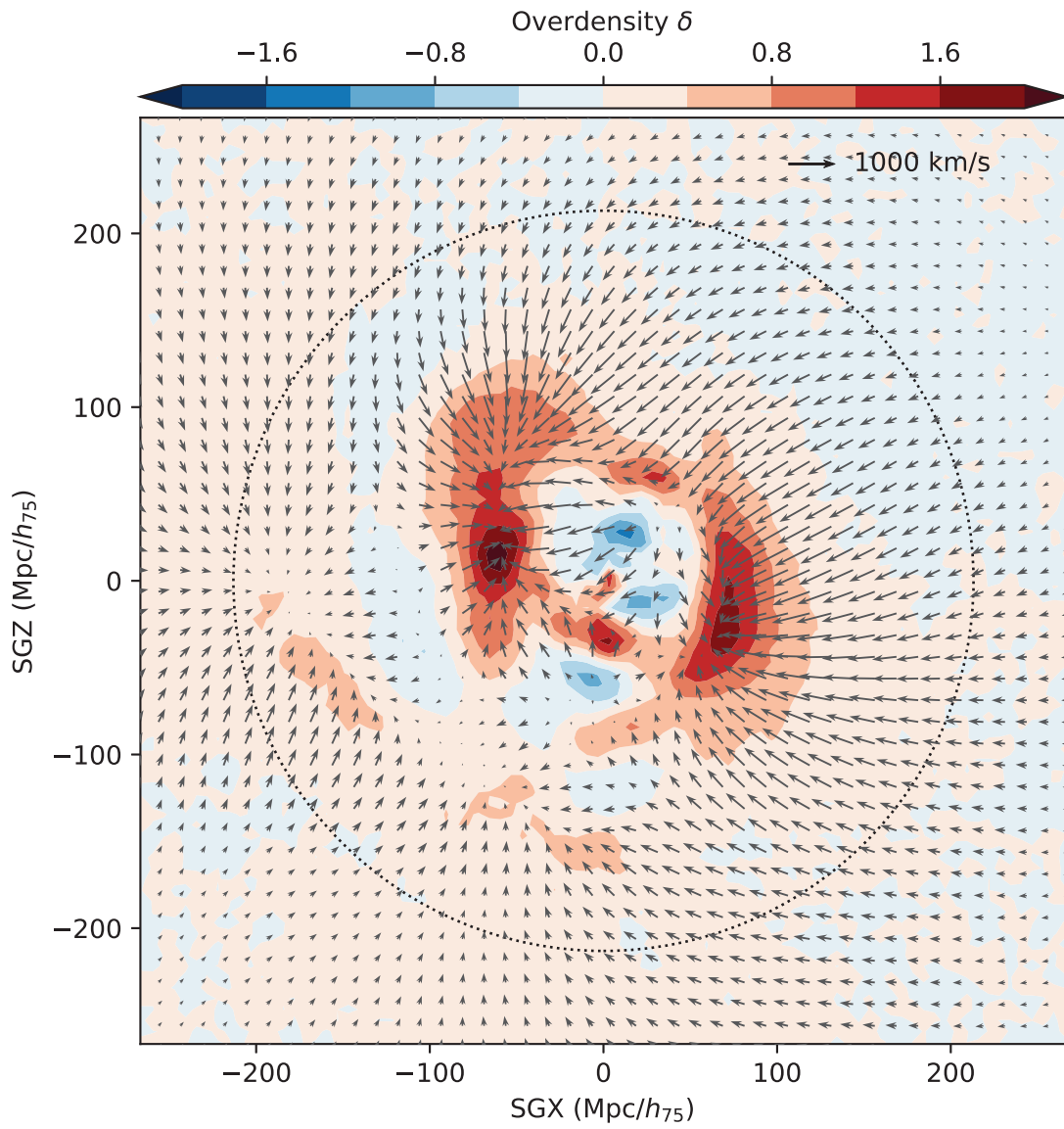
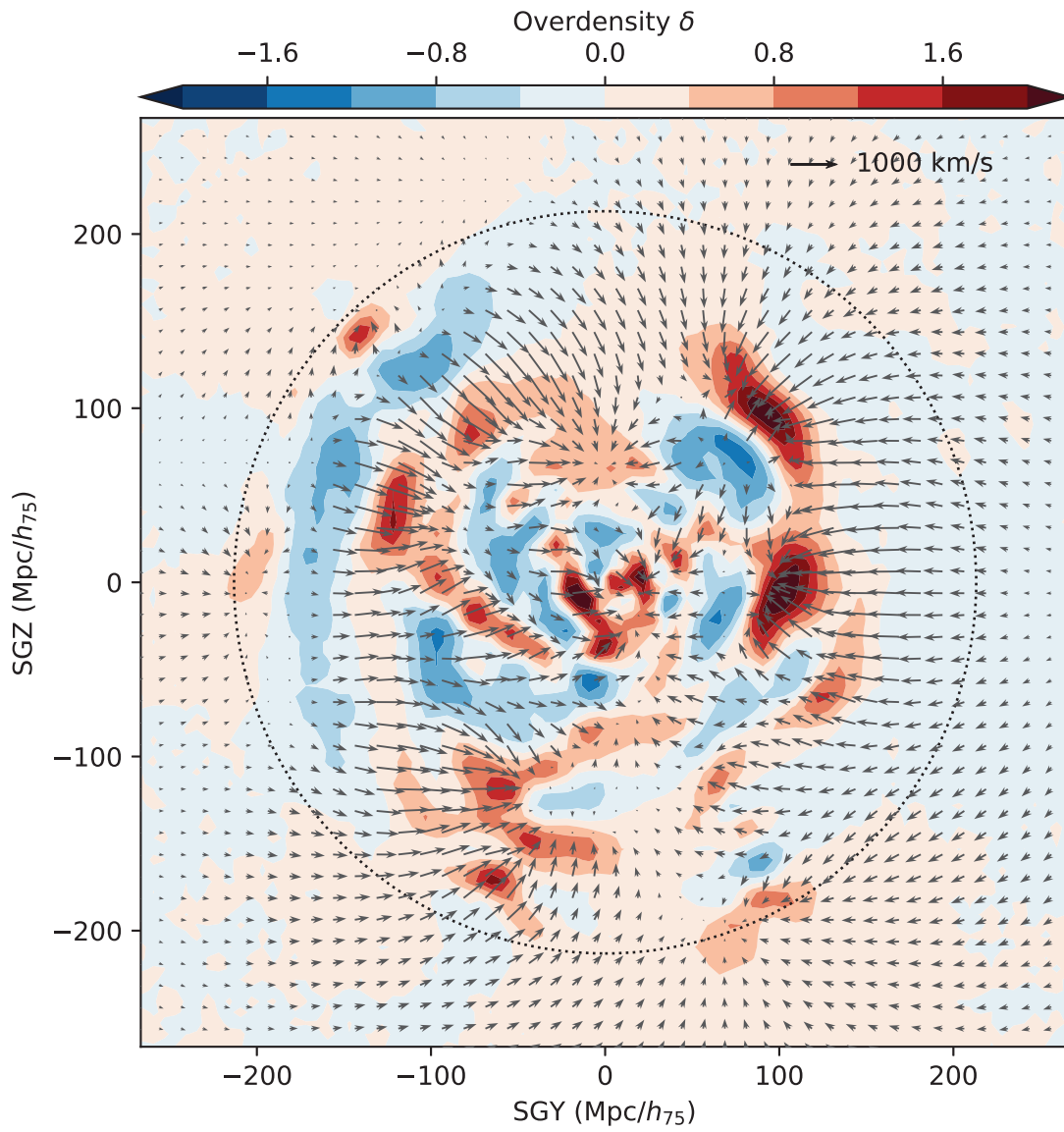


FIGURE 4.9 – Tranche en  $SGZ = 0$  du champ de surdensité et de vitesse reconstruit à partir des données *CosmicFlows-3*. La couleur correspond à la surdensité et les flèches noires aux vitesses projetées dans le plan  $SGZ = 0$ . Le cercle pointillé est centré en zéro et de rayon  $r = \frac{16000 \text{ km s}^{-1}}{H_0}$  correspondant à la limite des données 6dF.

FIGURE 4.10 – Identique à la figure 4.9 pour la tranche en  $SGY = 0$ .

FIGURE 4.11 – Identique à la figure 4.9 pour la tranche en  $SGX = 0$ .

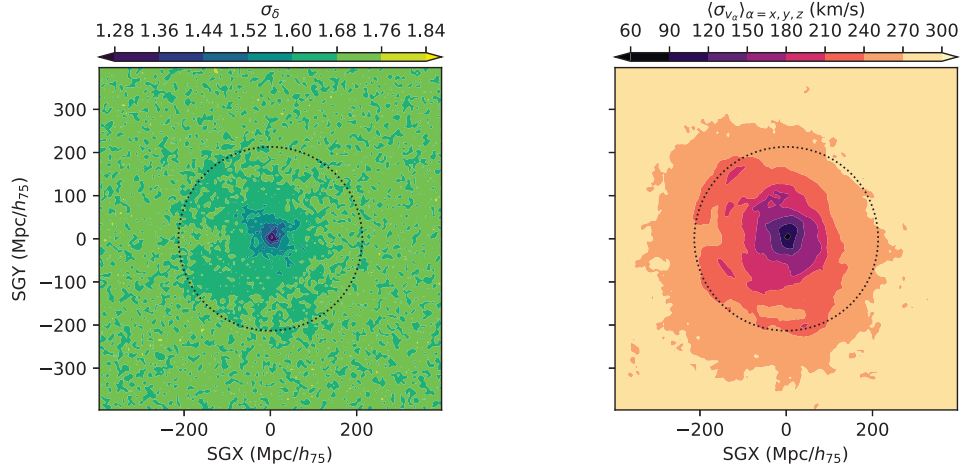


FIGURE 4.12 – Gauche : déviation standard du champ de surdensité reconstruit à partir de CF3. Droite : Moyenne des déviations standard sur les composantes cartésiennes du champ de vitesse de CF3. Les deux panneaux sont des tranches à  $SGZ = 0$ .

vers les valeurs prédites par le modèle  $\Lambda$ CDM, et la reconstruction est précise jusqu'à environ 200  $\text{Mpc}/h_{75}$  de distance du centre.

#### 4.3.1 Structures de l'Univers Local

Un moyen d'étudier la répartition de la matière est de calculer les surfaces d'iso-surdensité. Puisqu'il est difficile de représenter le champ tridimensionnel des vitesses, nous utilisons la figure interactive (D. Pomarède, comm. privée) SketchFab disponible via ce lien, et qui montre les iso-surdensités de valeurs 0.4, 0.75, 1.0, 1.25 et 1.75. Cette figure ainsi que la tranche  $SGZ = 0$  montrée sur la figure 4.9 révèlent des superamas déjà connus : Coma, Persée-Poissons, Shapley ; et la tranche  $SGX = 0$  le superamas Hercules. On confirme aussi la présence du « Grand Mur » entre l'amas Coma et le superamas Hercules.

Une nouveauté apportée par CF3 est la structure du superamas Shapley. Précédemment, le superamas Shapley était plutôt vu comme une seule grande entité, or nous séparons maintenant 3 composantes, une composante proche et une plus lointaine qui s'étend comme un filament du type de Persée-Poissons, avec deux surdensités majeures aux coordonnées  $(160, 47, 0) \text{ Mpc}/h_{75}$  et  $(-107, 53, 5) \text{ Mpc}/h_{75}$ . Ces deux régions sont au repos dans le référentiel du CMB. On peut voir que la partie Sud du filament Shapley lointain est en mouvement et converge vers la partie Nord lointaine de Shapley. Shapley serait donc un superamas encore en formation, contrairement à ce qui est habituellement relaté dans la littérature (Reisenegger et al., 2000).

Une autre nouveauté est l'apparition d'une nouvelle grande structure, que l'on nomme ici le superamas « Pôle Sud ». Cette grande structure, clairement visible, se situe aux coordonnées  $(-133, -53, -60) \text{ Mpc}/h_{75}$  et a été mentionné dans Courtois et al. (2017). De plus grandes analyses sont en cours.

En plus des surdensités positives, on peut également s'intéresser aux surdensités négatives, c'est-à-dire aux vides cosmiques. Nous reviendrons sur l'identification des vides dans la section 4.3.3, mais pouvons d'ores et déjà remarquer la délimitation du vide Local au centre du volume de reconstruction dans les trois tranches présentées par les figures 4.9, 4.10 et 4.11. Ce vide, dont la longueur est autour de  $67 \text{ Mpc}/h_{75}$ , contient en fait quatre minima locaux, dont le plus profond se situe dans la direction de Lacerta et du Cygne. Notons que c'est le vide d'Hercules qui représente le vide le plus profond de toute la reconstruction de CF3. Ceci est cohérent avec ce qui est connu dans la littérature d'après la distribution des galaxies dans les catalogues de décalage vers le rouge.

Cette cartographie des surdensités et des sous-densités sera approfondie dans les mois qui viennent. Elle conduit à un article en préparation. Nous concluons cette section en remarquant la grande variété et l'inhomogénéité des formes et des tailles des grandes structures et des vides de l'Univers Local.

### 4.3.2 Comparaison aux catalogues de décalages vers le rouge

Nous nous intéressons dans cette section à la comparaison du champ de surdensité reconstruit avec la répartition des galaxies observées dans l'Univers Local. La distribution des galaxies est, à grande échelle, semblable à celle de la matière totale. Ceci n'a rien d'évident et est difficile à vérifier expérimentalement *autrement qu'avec les vitesses particulières*. La littérature modélise généralement la différence entre la densité de matière stellaire et totale par le biais  $b$  galaxie-matière (Weinberg et al., 2004; Peacock and Smith, 2000) :

$$\delta_g = b\delta \quad (4.12)$$

La comparaison entre le champ de surdensité reconstruit et la distribution des galaxies (tirées du catalogue LEDA, Makarov et al. (2014)) est montrée pour trois tranches différentes sur la figure 4.13. Le panneau en haut à gauche montre la distribution des points de données dans la tranche  $SGZ = 0$  et d'épaisseur  $10 \text{ Mpc}/h_{75}$  et le panneau en haut à droite montre la distribution de toutes les galaxies dont on connaît le décalage vers le rouge. Qualitativement, on retrouve que les galaxies observées sont localisées dans les puits de potentiel reconstruits. À titre d'exemple on peut citer par exemple le superamas Persée-Poisson, constitué de deux voire trois attracteurs liés, et dont la surdensité maximale se trouverait dans la zone d'obscurisation. Également, la superposition des galaxies au niveau du superamas Shapley, composé de deux parties indépendantes, est très bonne. Ceci souligne la puissance des analyses de vitesses particulières : à partir d'un nombre de traceurs réduit, l'information retrouvée est semblable à celle contenue dans un catalogue de décalages vers le rouge.

On note tout de même quelques écarts. Au niveau de Poissons-Baleine (Porter and Raychaudhury, 2005) (Pisces-Cetus) et de l'Oiseau de Paradis (Apus), on observe un décalage systématique entre les positions des surdensités reconstruites et les galaxies observées. On pourrait penser qu'il s'agit là d'un mauvais ajustement de la constante de Hubble réduite  $h_{\text{eff}}$ , puisqu'elle entre en compte comme facteur multiplicatif des distances. Cependant, puisque le décalage n'est pas isotrope, et en particulier il y a une superposition cohérente du côté des  $SGY$  positifs, la raison de ce décalage pourrait plutôt se situer dans la mauvaise intercalibration des indicateurs de distance. Ce point est discuté dans la section 4.4.

Les panneaux du bas montrent les tranches  $SGX = 0$  et  $SGY = -93 \text{ Mpc}/h_{75}$  sur lesquelles on retrouve des structures connues comme le Mur du Sud, l'amas d'Hercules ou du Lièvre (Lepus). Là encore la reconstruction est remarquablement en accord avec les sondages de galaxies par décalage vers le rouge. On peut par exemple remarquer les sous-structures du superamas Lepus ou la morphologie du Mur du Sculpteur qui est retrouvée.

On observe donc que la superposition est qualitativement très correcte. La comparaison quantitative entre la distribution des galaxies et le champ de vitesse de matière totale peut en théorie mener à une estimation du biais galaxie-matière (Weinberg et al., 2004). L'étude a par exemple été tentée par Hoffman et al., 2018 (accepté à *Nature Astronomy*) mais le résultat passe par deux reconstructions différentes (Hoffman et al., 2015; Carrick et al., 2015) sujettes toutes deux à des erreurs systématiques différentes. Pour une analyse complète du biais galaxie-matière, certains effets systématiques doivent être inclus dans le modèle. Dans l'ordre d'importance :

- complétude des catalogues de décalages vers le rouge ;
- prises en compte des effets de distorsion dans l'espace des décalages vers le rouge (RSD) ;
- compatibilité du point-zéro de la reconstruction.

Le deuxième point est le plus délicat. Notre modélisation actuelle du champ de vitesse de l'Univers Local s'appuie sur la théorie linéaire de formation des structures. Pour inclure les RSD dans l'analyse, il faudrait la compléter par un modèle qui décrit les non-linéarités au sein des amas de galaxies.

### 4.3.3 La toile cosmique des vitesses

Dans cette section, on analyse les grandes structures de l'Univers Local à partir des propriétés du champ de vitesse. On s'intéresse d'abord aux profils de vitesse radiale en direction de superamas connus dans la section 4.3.3.1, puis systématisons l'étude en s'intéressant à la toile cosmique des vitesses dans la section 4.3.3.2.

#### 4.3.3.1 Profils de vitesse

Historiquement, les amas ont été étudiés à partir des courbes dites « en S » ou « en Z » (Dressler and Faber, 1990; Tonry and Davis, 1981; Mathewson et al., 1992; Sandage et al., 2010; Karachentsev et al., 2014). Physiquement, les vitesses radiales sont plus grandes avant l'amas qu'après lui, et inversement pour les vides. Ces profils permettent ainsi de localiser un superamas, ainsi que d'avoir accès à sa vitesse particulière. On trace sur la Figure 4.14 deux profils de vitesse radiale en direction de l'amas Coma et l'amas Shapley, déjà connus. Le panneau de gauche permet quant à lui de visualiser la direction du profil dans l'espace tridimensionnel.

À partir de ces profils on identifie les superamas et les vides comme étant les points où s'annulent localement la dérivée seconde de la courbe. Si la dérivée première est positive, il s'agit d'un vide, sinon d'un superamas. On constate sur le panneau de droite que cette construction coïncide bien avec la présence de surdensité dans le volume. L'avantage de visualiser les superamas à partir de la vitesse particulière est de pouvoir identifier la « portée » du superamas (ou du vide), qui peut être une interprétation de l'amplitude de la « courbe en



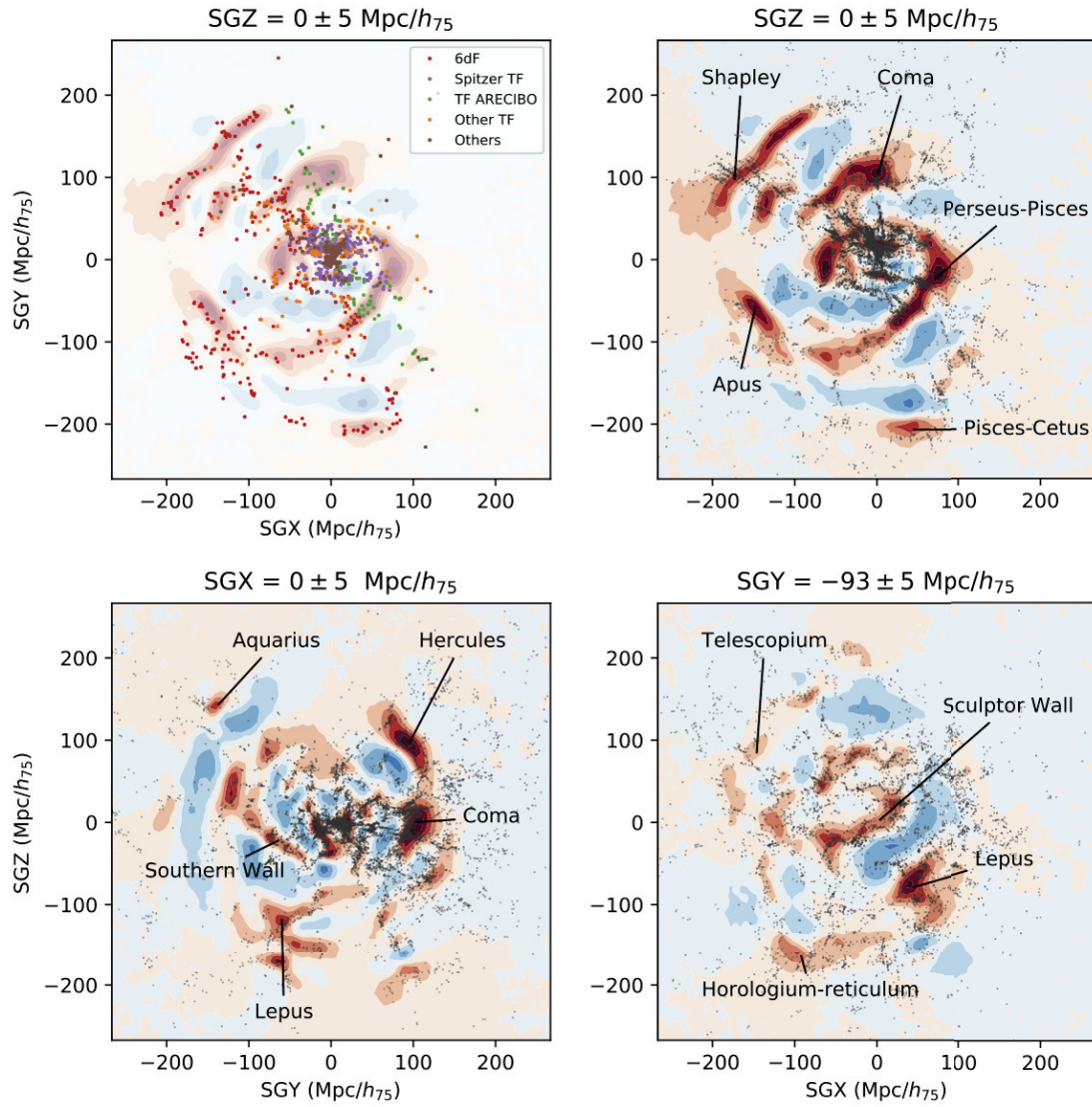


FIGURE 4.13 – Comparaison du catalogue de décalages vers le rouge LEDA avec la reconstruction du champ de surdensité à partir de *CosmicFlows*. La comparaison est faite dans trois tranches  $SGZ = 0$ ,  $SGX = 0$  et  $SGY = -93$   $Mpc/h_{75}$ . Le panneau en haut à gauche montre la distribution des galaxies de *CosmicFlows-3* utilisées pour la reconstruction.

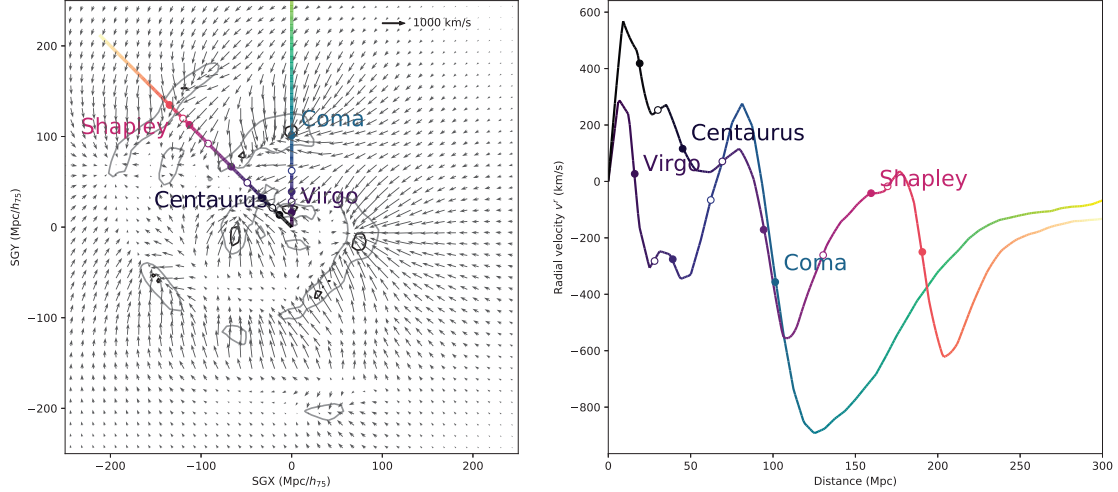


FIGURE 4.14 – Profils de vitesses radiales dans les directions de l’amas Coma et Shapley. À gauche figure le champ de vitesse reconstruit à partir des données CF3 (les contours représentent les surdensités positives). À droite est montrée la vitesse radiale en fonction de la distance à notre galaxie. Les points pleins correspondent à l’identification d’amas de matière tandis que les points creux correspondent à l’identification de vides.

S » autour du point d’inflexion. Cette façon de faire est issue des analyses à partir de données observationnelles qui ne fournissent que la vitesse radiale. Nous avons cependant reconstruit l’intégralité du champ tridimensionnel et pouvons ainsi accéder à une analyse plus systématique et mathématiquement précise de ces superamas, en particulier avec la toile cosmique des vitesses.

#### 4.3.3.2 Toile cosmique des vitesses

En anglais, *Cosmic V-web*. La dérivée d’un champ de vitesse irrotationnel comme le champ linéaire reconstruit ici est complètement définie par le tenseur de cisaillement  $\Sigma$  (Landau and Lifshitz, 2013) :

$$\Sigma_{\alpha,\beta} = -\frac{1}{2} \left( \frac{\partial v_\beta}{\partial r_\alpha} - \frac{\partial v_\alpha}{\partial r_\beta} \right) \quad (4.13)$$

où  $\alpha$  et  $\beta$  dénotent les composantes cartésiennes du champ de vitesse. Le tenseur  $\Sigma$  est réel et symétrique et donc diagonalisable. Les vecteurs propres de  $\Sigma$  correspondent aux directions préférentielles du champ de vitesse, et les valeurs propres à la dérivée des vitesses le long de ces axes. En particulier, le signe des valeurs propres du tenseur de cisaillement nous renseigne sur le caractère extensif ou compressif des structures sous-jacentes. Par exemple, trois valeurs propres positives signifient qu’il y a accrétion de matière selon les trois directions principales, et donc que la structure sous-jacente est un attracteur.

Exhaustivement, les cas sont les suivants :



- si les trois valeurs propres sont positives, le point est un *nœud* du champ de vitesse, correspondant à un effondrement selon toutes les directions ;
- si deux valeurs propres sont positives, le point fait partie d’un *filament*, puisque l’effondrement se fait selon deux directions, perpendiculaires à la droite définie par le vecteur propre associé à la valeur propre positive ;
- si une seule valeur propre est positive, il s’agit d’un *plan* : effondrement selon une seule direction ;
- enfin, les *vides* cosmiques peuvent se définir comme les positions où toutes les valeurs propres du tenseur de cisaillement sont négatives (expansion dans les trois directions).

Pour résumer ces propriétés on peut tracer la somme des signes des valeurs propres du tenseur de cisaillement. La valeur 3 correspond ainsi à un noeud, 1 à un filament, -1 à un plan et -3 à un vide (la somme de trois nombres de valeur absolue 1 ne peut avoir une norme de deux). Toutefois, on introduit en général (Pomarède et al., 2017) un minimum positif pour le signe des valeurs propres. En effet, pour théoriquement identifier un effondrement il faudrait rajouter le flot de Hubble autour du superamas, ce qui réduit l’effondrement sur la structure. En pratique on fixe ces minima à  $\lambda_{1,2,3} \sim 0.1$  où les indices 1, 2, 3 correspondent au minimum pour la  $i$ -ème valeur propre dans l’ordre décroissant.

La figure 4.15 montre les noeuds et filaments de la toile cosmique des vitesses, où plusieurs superamas sont annotés, et la figure 4.16 montre les vides et les plans. Les noeuds permettent notamment d’identifier des superamas connus comme Coma Persée-Poissons, Hercules et Shapley. La figure 4.16 exhibe une structure singulière en bas à gauche du côté des  $SGZ < 0$  et  $SGX > 0$ . Ce vide très étendu est approximativement aligné avec la direction du point froid du CMB (Vielva et al., 2004; Cruz et al., 2005) et pourrait expliquer une partie de cette anomalie (Courtois et al., 2017). Cependant, plusieurs analyses semblent réfuter la possibilité que le point froid soit uniquement la conséquence d’une région sous-dense (Mackenzie et al., 2017; Naidoo et al., 2016). Dans tous les cas cette région est singulière par son extension mais se trouve assez distante de nous (environ 190 Mpc/ $h_{75}$ ) et donc sujette à des biais de reconstruction. Pour approfondir l’étude, il est nécessaire d’obtenir plus de données dans cette région.

Le V-Web étant calculé, on peut ensuite se demander quelle est la proportion de chacune des structures dans l’Univers Local ? Les résultats sont donnés dans la Table 4.1. On voit que les filaments et les feuilles dominant en volume. Les noeuds semblent occuper plus de volume que les vides, ce qui est contradictoire avec les études faites à partir de la distribution des galaxies dans les catalogues de décalages vers le rouge (Leclercq et al., 2015). Ceci est dû à l’approximation linéaire sous-jacente à la méthode. En effet, dans le domaine non linéaire, les superamas sont plus effondrés et la matière plus structurée. Du fait de ces effets non linéaires, les reconstructions à partir des vitesses particulières ne sont pas encore compétitives en terme de classification des grandes structures de l’Univers Local. La section 6.1 propose une ouverture sur cette problématique.

L’autre information que l’on peut tirer du champ de vitesse après l’identification des structures est leurs mouvements respectifs. On voit ici que, dans le référentiel du CMB, les superamas se déplacent avec des vitesses de l’ordre de  $\sim 300 \text{ km s}^{-1}$ , ce qui est prédit par le modèle  $\Lambda$ CDM. En particulier, une question souvent soulevée est le mouvement de notre Groupe

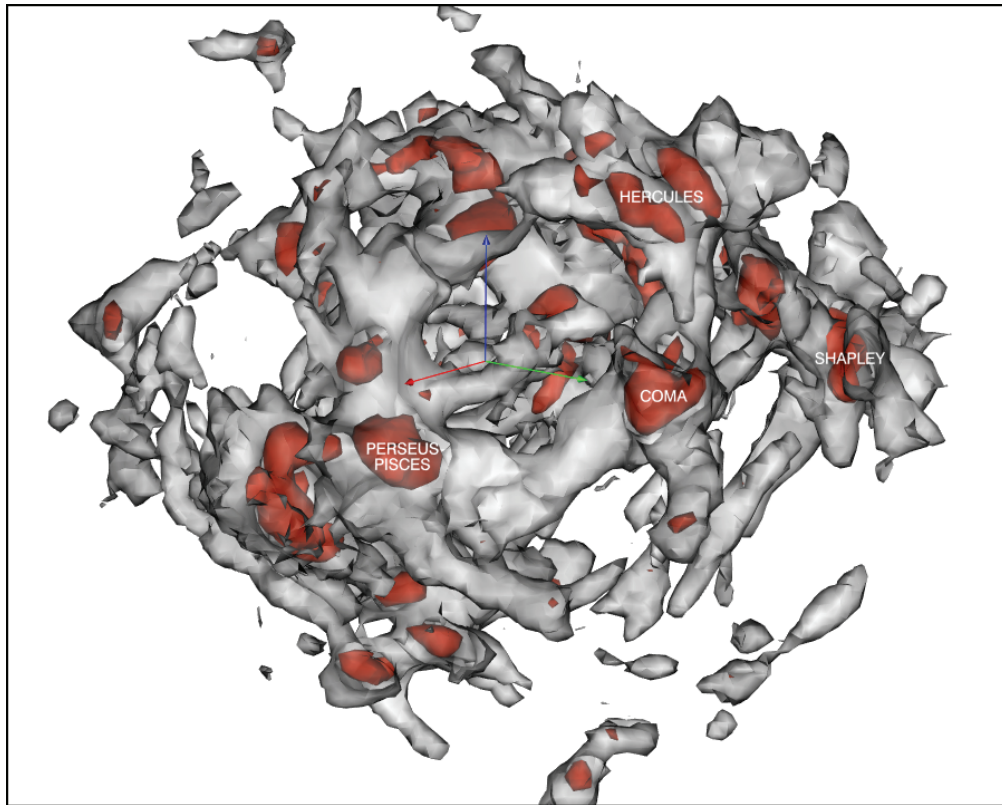


FIGURE 4.15 – Filaments (en gris) et noeuds (en rouge) de la toile cosmique des vitesses calculés à partir des valeurs propres du tenseur de cisaillement. Les flèches rouges, vertes et bleues correspondent aux directions  $SGX$ ,  $SGY$  et  $SGZ$ . La taille de ces flèches est de  $67 \text{ Mpc}/h_{75}$ .

TABLE 4.1 – Répartition volumique des différentes structures du V-Web dans la reconstruction du champ de vitesse de CF3

Structure	Densité volumique totale (%)	Densité volumique pour $R < 213 \text{ Mpc}/h_{75}$ (%)
Nœuds	2.3	9.5
Filaments	50.0	50.0
Feuilles	41.2	34.3
Vides	6.5	6.2

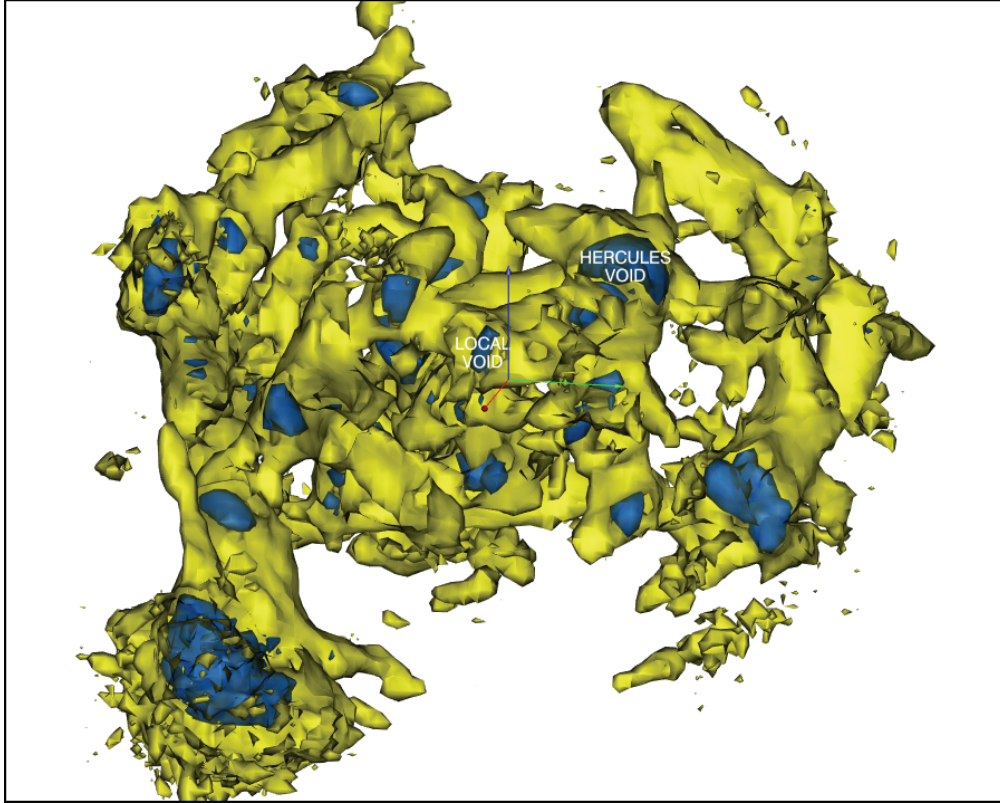


FIGURE 4.16 – Plans (en jaune) et vides (en bleu) de la toile cosmique des vitesses calculés à partir des valeurs propres du tenseur de cisaillement. Les flèches rouges, vertes et bleues correspondent aux directions  $SGX$ ,  $SGY$  et  $SGZ$ . La taille de ces flèches est de  $67 \text{ Mpc}/h_{75}$ .

Local par rapport aux grandes structures et au CMB. Puisque nous avons accès au champ tridimensionnel des vitesses particulières, on peut calculer les mouvements relatifs.

#### 4.3.3.3 Mouvement relatif du groupe local

Plus précisément, on se demande ici si le dipôle du CMB est expliquable par un effet Doppler provenant des vitesses particulières. Ce dipôle est en général donné dans le référentiel du groupe local  $LG$  (Fixsen et al., 1996; Tully et al., 2008) :

$$V_{cmb/LG} = 627 \text{ km s}^{-1} \quad \text{en direction de } (l, b) = (276, 30)^\circ \quad (4.14)$$

Est-ce que ce dipôle peut être expliqué uniquement par la présence des structures de l'Univers Local ou faut-il revoir le modèle  $\Lambda$ CDM comme certains le proposent (Scrimgeour et al., 2016; Springob et al., 2016; Carrick et al., 2015; Hong et al., 2014)? À partir de la reconstruction proposée ici, on peut calculer le mouvement relatif de notre groupe local (qui est le point au centre du volume de reconstruction) par rapport aux grandes structures connues, Virgo, Coma,

le Grand Attracteur (GA), Shapley et par rapport au CMB. Les résultats sont présentés dans la Table 4.2.

On constate que la reconstruction est compatible avec le dipôle du CMB à environ  $1\sigma$  et que les structures retrouvées expliquent bien le mouvement de notre galaxie. On remarque que notre mouvement est principalement selon les directions  $SGX$  négatives et  $SGY$  positives du fait de la présence de Virgo, du Grand Attracteur, de Coma et de Shapley. Attention toutefois, cette analyse souffre d'une limite majeure : la reconstruction est faite à partir d'un *a priori*  $\Lambda$ CDM. Même si la reconstruction est dominée par les données jusqu'à environ  $250 \text{ Mpc}/h_{75}$ , l'*a priori*  $\Lambda$ CDM interdit de fait la présence d'un dipôle trop grand dans le champ de vitesse. Il est néanmoins possible d'intégrer une déviation au dipôle prédit au modèle  $\Lambda$ CDM dans le modèle comme nous le verrons dans la section 4.4.4. Cette analyse reste à conduire. L'autre limite est le manque de données dans l'hémisphère Nord, qui rend l'analyse asymétrique et donc sujette à caution. Toutefois, cette limite n'est pas fondamentale puisque, les vitesses corrélant sur des grandes distances, la présence d'un superamas capable de compenser l'attraction gravitationnelle vers Shapley aurait été décelée, sans toutefois savoir dans quelle mesure.

#### 4.3.4 Comparaison filtre de Wiener, *CosmicFlows-2*, *CosmicFlows-3*

Dans cette section, nous comparons les champs reconstruits à partir des catalogues *CosmicFlows-2* (Tully et al., 2013) ( $\sim 8000$  galaxies) et *CosmicFlows-3* (Tully et al., 2016) ( $\sim 17000$  galaxies) avec la méthode développée dans ce chapitre. On présente aussi la reconstruction du champ CF3 obtenue par méthode WF/CR présentée dans le chapitre 2. On regarde d'abord la comparaison entre la reconstruction à partir de CF2 (panneau de gauche) et celle à partir de CF3 (panneau de droite). Cette comparaison montre l'information qu'apportent environ 8000 mesures de distances supplémentaires dans la reconstruction. La plupart de ces nouvelles données viennent du catalogue 6dF (Campbell et al., 2014), dont les données proviennent de l'hémisphère Sud. L'amélioration est en effet surtout notable dans ces régions : les superamas Shapley, Oiseau de Paradis (Apus), et Lièvre (Lepus) sont soit mieux définis, soit apparaissent tout simplement quand ils n'étaient pas présents dans CF2. On remarque aussi une amélioration dans le flot arrière des superamas. en particulier du côté des superamas Coma et du Grand Attracteur, qui sont désormais délimités complètement. Nous verrons dans le chapitre 6 de perspectives quels sont les futurs programmes d'observation de distances extragalactiques, mais nous pouvons d'ores et déjà dire qu'ils contiendront plusieurs dizaines de milliers de nouvelles données. Au vu de l'amélioration entre les reconstructions à partir des catalogues CF2 et CF3, on peut prévoir l'impact de tels programmes sur la reconstruction du champ de vitesse dans un futur proche.

Le panneau du milieu montre la technique WF/CR appliquée au catalogue CF3. Les biais détaillés dans le chapitre 2 apparaissent clairement :

- au centre, les sous-densités semblent trop faibles, c'est l'effet de volume associé au biais de positionnement ;
- à grande distance, la présence d'un flot global entrant est dû aux effets de sélection et au biais de positionnement. En particulier on peut noter que le flot entrant est plus important du côté des données 6dF, puisque les effets de sélection y sont plus forts.

TABLE 4.2 – Mouvements relatif du groupe local par rapport à quelques structures connues, en coordonnées supergalactiques.

Référentiel $\mathcal{R}$	$\mathbf{v}_g/\mathcal{R} \cdot \hat{\mathbf{x}}$ (km/s)	$\mathbf{v}_g/\mathcal{R} \cdot \hat{\mathbf{y}}$ (km/s)	$\mathbf{v}_g/\mathcal{R} \cdot \hat{\mathbf{z}}$ (km/s)	$ \mathbf{v}_g/\mathcal{R} $ (km/s)	Direction ( $l, b$ ) ( $^\circ$ )
CMB Fixsen et al. (1996)	-407	351	-322	627	(276, 30)
CMB	$-394 \pm 17$	$380 \pm 15$	$-316 \pm 21$	$579 \pm 24$	$(211, 66) \pm (7, 3)$
Virgo	$67 \pm 28$	$551 \pm 23$	$-166 \pm 34$	$833 \pm 37$	$(234, 59) \pm (6, 3)$
Coma	$-49 \pm 47$	$756 \pm 34$	$-348 \pm 50$	$560 \pm 42$	$(244, 40) \pm (5, 4)$
GA	$-121 \pm 34$	$403 \pm 41$	$-369 \pm 46$	$533 \pm 55$	$(252, 42) \pm (6, 6)$
Shapley	$-165 \pm 38$	$396 \pm 55$	$-316 \pm 58$	$639 \pm 50$	$(274, 33) \pm (5, 5)$

La comparaison entre le panneau du milieu et celui de droite souligne bien la nécessité d'inclure des modèles comme celui présenté dans ce chapitre au coeur des analyses de vitesses particulières.

#### 4.3.5 Études possibles

Les cartes présentées ici sont des cartes de matière (totale) de l'Univers Local. En dehors de l'aspect cosmographie que nous avons brièvement développé dans la section précédente, elles peuvent potentiellement servir d'outils pour tester le modèle  $\Lambda$ CDM. Une liste non exhaustive des études possibles à partir de telles reconstructions est la suivante :

- correction des vitesses particulières dans des analyses tierces (ondes gravitationnelles (Abbott et al., 2017) ou SNIa (Betoule et al., 2014)) ;
- estimation du biais galaxie-matière par comparaison de la distribution de matière reconstruite avec celle des galaxies (Nusser, 2017; Carrick et al., 2015) ;
- profils de vitesse dans le vides cosmiques (Novosyadlyj and Tsizh, 2017; Falck et al., 2018) ;
- étude de la réaction cinématique (Buchert, 2018) ;
- biais d'alignement dans les analyses de cisaillement (Ganeshiah Veena et al., 2018).

L'intégration de ces études au modèle d'analyse n'a pas fait partie de cette thèse et ces analyses seront développées dans un temps ultérieur. Nous nous tournons pour le moment vers l'analyse des erreurs systématiques du modèle.

## 4.4 Erreurs systématiques

### 4.4.1 Idée générale

Le but de cette section est d'analyser les potentielles erreurs systématiques de la reconstruction. Comme présenté dans le Chapitre 3, le modèle que nous utilisons est flexible. Il est notamment possible de changer plusieurs paramètres :

- modèle de non-linéarité des vitesses observées, ici modélisée par la dispersion gaussienne  $\sigma_{\text{NL}}$  ;
- modèle de calibration du point zéro, modélisé ici par le paramètre  $h_{\text{eff}}$  ;
- ajout d'un dipôle externe ;
- *a priori* sur les distances ;
- données utilisées (galaxies, groupes forts, faibles,...) ;
- coupure à petite échelle dans le spectre de puissance ;
- etc.

L'étude exhaustive de tous les cas de figure n'est pas nécessaire. Nous proposons d'étudier trois paramètres principaux et leur impact sur la reconstruction. Le premier paramètre est le choix du modèle d'*a priori* parmi les densités de probabilité définies dans la Section 3.1.3. Le second est le nombre de constantes de Hubble réduites définies dans la section 3.1.2.1. Et enfin, la présence ou l'absence d'un dipôle externe. Dans tous les cas étudiés nous utilisons une sous-partie du catalogue groupé de CF3 afin de limiter les temps de calcul. La sous-partie, que l'on nommera *CF3gp-50* est construite en extrayant de façon aléatoire 50% des galaxies (et groupes



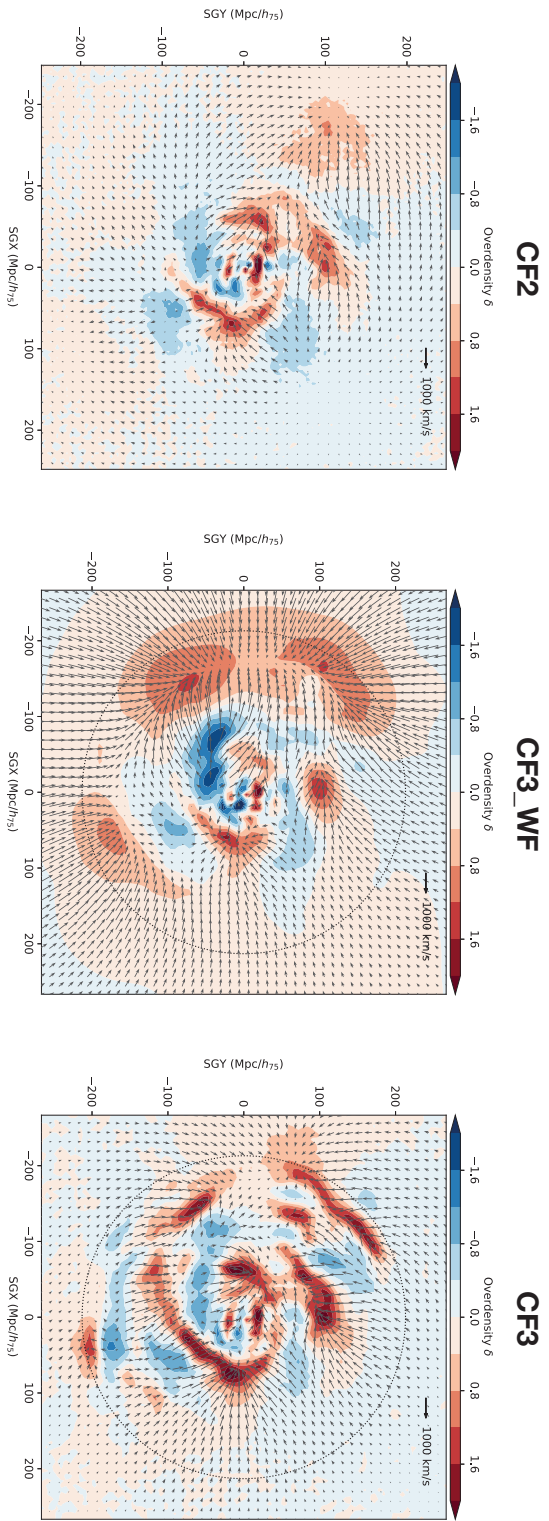


FIGURE 4.17 – Tranche  $SGZ = 0$  des champs reconstruits pour CF2 et CF3. Gauche : CF2 avec la méthode présentée dans ce chapitre. Milieu : CF3 avec la méthode WF/CR. Droite : CF3 avec la méthode présentée dans ce chapitre.

TABLE 4.3 – Résumé des différentes reconstructions faites pour estimer les erreurs systématiques sur la reconstruction du champ de vitesse de CF3gp-50.

#	Modèle $h_{\text{eff}}$	$\mathbf{v}_{\text{ext}}$	<i>A priori</i>	Section
1	Unique	Fixé	Uniforme	4.4.2
2	Unique	Fixé	Empirique	
3	Unique	Fixé	Empirique pour 6dF seulement	
4	Unique	Fixé	Asymétriques gaussiens	
5	Deux, dont un pour 6dF	Fixé	Empirique	4.4.3
6	Deux, dont un pour 6dF	Fixé	Empirique pour 6dF seulement	
7	Unique	Ajusté	Empirique	4.4.4

de galaxies) au sein de chacune des catégories définies dans la Section 4.1.1. En procédant ainsi, nous conservons les proportions relatives de chacune des méthodologies ainsi que la même répartition angulaire et en décalage vers le rouge. Les propriétés des reconstructions sont résumées dans la Table 4.3

#### 4.4.2 Influence des *a priori*

Dans cette section est testée l'influence des *a priori* sur la reconstruction du champ de vitesse. Pour reconstruire le champ de vitesse de CF3, nous avons fait le choix d'*a priori* sur les distances comme expliqué en Section 4.2.1. Nous remettons en question ce choix en effectuant la reconstruction du champ de vitesse sous-jacent à CF3gp-50 en envisageant les cas suivants :

1. *a priori* uniforme sur toutes les distances,
2. *a priori* empirique défini par l'équation (3.16) sur toutes les distances,
3. *a priori* empirique sur les distances issues du sous-échantillon 6dF, uniforme sur le reste,
4. *a priori* gaussien asymétrique défini par l'équation (3.18) sur toutes les distances.

Les résultats sont montrés sur la figure 4.18 sur laquelle on montre la tranche  $SGZ = 0$  de chacune des reconstructions, ainsi que l'histogramme des paramètres  $h_{\text{eff}}$  et  $\sigma_{\text{NL}}$ .

**Cas #1** C'est le cas de la reconstruction avec des *a priori* uniformes sur toutes les distances. À première vue, la reconstruction semble irréaliste. En particulier, la présence de deux flots, l'un partant du centre vers les SGX négatifs, l'autre de l'extérieur vers les SGX positifs est due à une mauvaise modélisation des distances de l'échantillon 6dF. La coupure en décalage vers le rouge à  $z \sim 0.054$  devrait être synonyme de coupure en distance à environ  $d \sim 230$  Mpc, mais les distances reconstruites ne respectent pas cette coupure du fait des erreurs observationnelles. S'ensuit une surestimation systématique de toutes les vitesses particulières radiales à grande distance et une sous-estimation à petite distance, qui donnent lieu au champ de la figure 4.18. Les histogrammes montrent que la valeur pour  $h_{\text{eff}}$  est  $h_{\text{eff}} = 0.8$  soit un écart de 20% avec le modèle cosmologique supposé. Notons que la technique WF/CR discutée dans le chapitre 2 suppose toujours



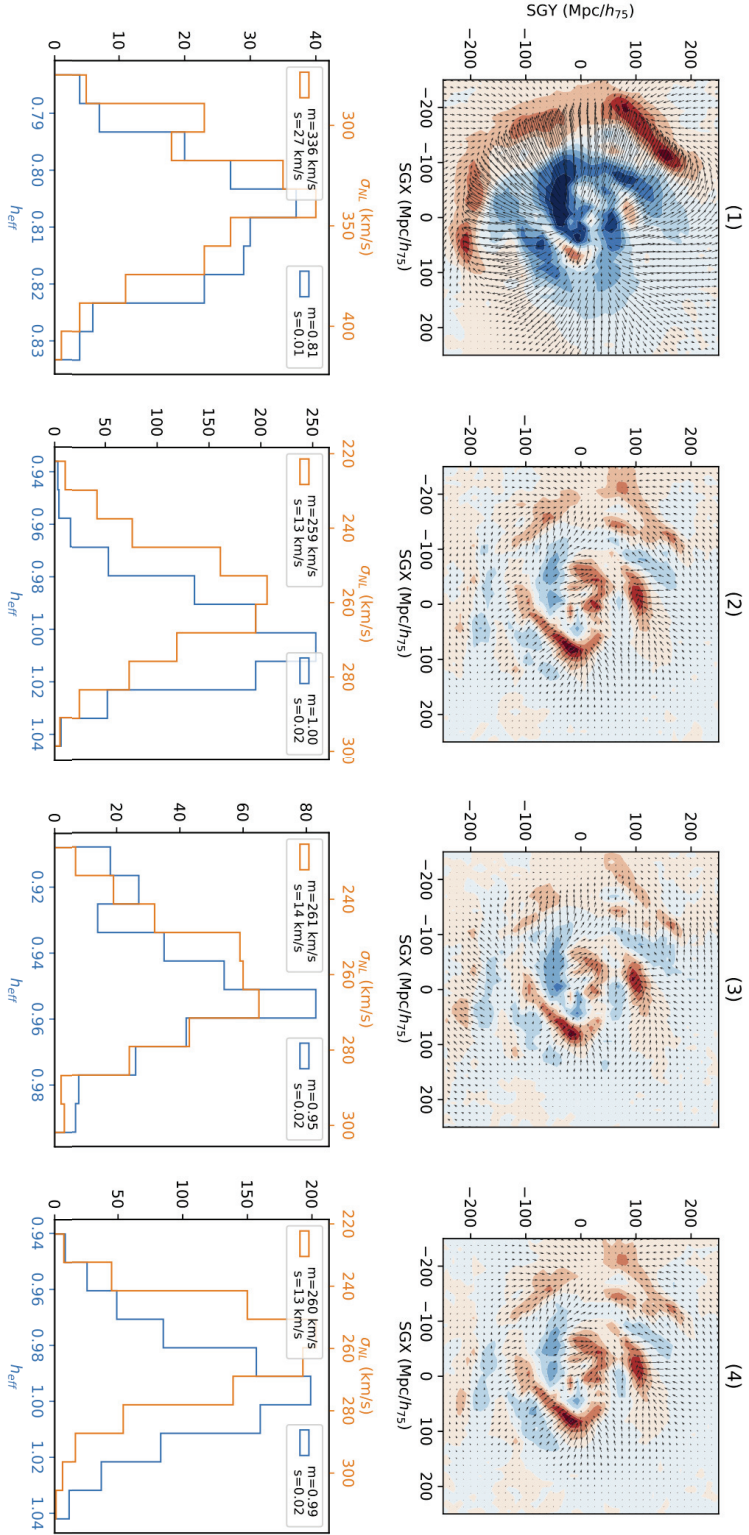


FIGURE 4.18 – La tranche  $SGZ = 0$  du champ reconstruit à partir du catalogue CF3gp-50 dans différents cas : (1) aucun  $a priori$  n'est attribué aux distances reconstruites ; (2) il s'agit du cas de référence,  $a priori$  empiriques sur les distances de chacun des sous-échantillons du catalogue ; (3)  $a priori$  uniforme sur les distances autres que 6dF ; (4)  $a priori$  asymétriques gaussiens sur les distances de chacun des sous-échantillons du catalogue.

des *a priori* uniformes, ce qui montre bien la nécessité du modèle présenté dans cette thèse.

**Cas #2** Reconstruction avec des *a priori* empiriques sur chacun des sous-échantillons définis en section 4.1.1. On remarque que la reconstruction est semblable à celle présentée dans la section 4.2. Il s’agit de la même technique de reconstruction appliquée sur le catalogue CF3gp-50.

**Cas #3** Reconstruction avec des *a priori* uniformes excepté sur les données 6dF. La reconstruction est similaire au cas précédent, les valeurs de  $\sigma_{\text{NL}}$  sont également similaires. Un examen plus précis montre un effondrement plus grand sur les structures de Coma et Persée-Poisson. Ce qui est intéressant est la différence de valeur moyenne de  $h_{\text{eff}}$ . On trouve une augmentation de 3%, soit l’équivalent de  $2.25 \text{ km s}^{-1}$  sur la constante de Hubble, simplement en ayant changé partiellement le modèle d’*a priori* sur les distances. Cela souligne l’impact des effets de sélection dans la détermination de paramètres cosmologiques à partir de l’analyse de vitesses particulières.

**Cas #4** Cas de la reconstruction avec des *a priori* asymétriques gaussiens sur les distances. On pourrait s’attendre à ce que la reconstruction soit similaire à celle du cas #2 puisqu’on a vu sur la figure 3.2 de la section 3.1.3 que les fonctions d’*a priori* étaient toutes semblables les unes aux autres. Toutefois on voit ici un léger changement dans le champ de vitesse, et en particulier un effondrement moins grand sur les amas. La valeur de  $h_{\text{eff}}$  reste autour de 0.99.

Physiquement, on comprend que plus la coupure à grande distance de l’*a priori* est nette, plus les distances ajustées seront grandes et plus la valeur du paramètre  $h_{\text{eff}}$  sera faible. Il faut rajouter à cela l’erreur totale sur le champ de vitesse qui n’est pas triviale à modéliser. L’erreur systématique sur le paramètre  $h_{\text{eff}}$  est d’environ 5%, soit l’équivalent de quelques  $\text{km s}^{-1} \text{ Mpc}^{-1}$ . Nous avons aussi confirmé ici l’indispensabilité d’attribuer des *a priori* sur les paramètres de distance pour permettre d’utiliser les vitesses particulières comme traceurs. On voit aussi, ce qui est rassurant, que les structures du champ de surdensité restent stables entre les différents choix d’*a priori*.

#### 4.4.3 Influence de la calibration des indicateurs de distance

Nous avons vu dans la section 4.1.2 que les différents indicateurs de distance ont été placés sur une échelle commune de distance, puis avons introduit une constante de Hubble effective  $h_{\text{eff}}$  pour ajuster globalement cette échelle. Il est à noter cependant que la procédure d’intercalibration des indicateurs de distance est assortie d’erreurs sur chacun des déplacements des points zéros vers le point zéro commun. Ces incertitudes jouent sur tous les modules de distance d’un indicateur de distance donné et ainsi ne peuvent plus être traitées comme des incertitudes statistiques, mais systématiques.

Le modèle décrit dans la section 3.1.2.1 permet de propager ces incertitudes en introduisant non seulement une, mais plusieurs constantes de Hubble effectives. Ainsi, chaque indicateur de distance peut être incorporé dans la reconstruction du champ de vitesse en ayant son propre point zéro. En particulier, le recouvrement entre les données 6dF et les autres est faible (environ 120 galaxies), ce qui implique une importante incertitude sur l’intercalibration de ce

sous-échantillon avec le reste de l'échantillon de CF3. Les autres sous-échantillons sont supposés bien intercalibrés. Dans cette sous-section, le champ de vitesse sous-jacent au catalogue CF3gp-50 est donc reconstruit pour trois cas différents :

2. une seule constante de Hubble effective, déjà étudié ci-dessus, c'est le modèle de référence ;
5. deux constantes de Hubble effectives : Une pour les données provenant de 6dF et une autre pour le reste des données ;
6. deux constantes de Hubble effectives, avec des *a priori* uniformes sur les distances autres que 6dF ;

Les résultats pour les différents cas sont présentés sur la figure 4.19.

**Cas # 5** Deux constantes de Hubble effectives. On constate un décalage d'environ 2% entre les deux constantes de Hubble effectives, celle attribuée au données 6dF étant la plus grande. On note aussi que l'erreur statistique sur chacune d'elle est plus grande que dans le cas optimale. Ceci peut être expliqué par le fait que les deux constantes sont corrélées et créent ainsi une confusion dans la distribution postérieure. D'un point de vue cosmographie, on ne note aucun changement majeur comparé au cas idéal.

**Cas # 6** Deux constantes de Hubble effectives, avec des fonctions d'*a priori* uniformes pour les distances autres que 6dF. On observe une incompatibilité entre les deux ajustements de point-zéro. En particulier, le point-zéro des données autres que 6dF tombe à  $h_{\text{eff}} = 0.92$ , soit  $5 \text{ km s}^{-1} \text{ Mpc}^{-1}$  de moins que pour 6dF. On voit que la valeur de la constante de Hubble dépend fortement du choix de l'*a priori* mais aussi que les données ne sont pas nécessairement correctement intercalibrées. Ce cas est particulièrement intéressant puisqu'il montre la dépendance de la reconstruction sur l'*a priori* et sur la modélisation des points-zéro. La reconstruction montre un plus faible flot entrant du côté de l'hémisphère Nord à cause de l'absence d'*a priori* dans cette région. On peut aussi remarquer l'apparition plus prononcée d'un superamas aux coordonnées  $(SGX, SGY) = (100, -150) \text{ Mpc}/h_{75}$ .

Finalement, on peut conclure que l'intercalibration des distances de 6dF avec le reste de l'échantillon peut être sujette à révision. En particulier, le changement de modèle de constante de Hubble effective fait varier le paramètre  $h_{\text{eff}}$  jusqu'à  $\Delta h_{\text{eff}} = 0.7$  soit  $5 \text{ km s}^{-1} \text{ Mpc}^{-1}$  d'erreur systématique.

Il est toutefois à noter que le modèle défini par plusieurs constantes de Hubble effectives n'est pas idéal pour au moins trois raisons :

- le catalogue CF3 contient des données venant de multiples sources et indicateurs de distance. Les données publiques ne permettent pas de remonter à ces sources, et la classification retenue en 5 sous-échantillons est approximative et ne prend pas en compte toutes les spécificités observationnelles sous-jacentes ;
- le modèle définit la constante de Hubble en terme de vitesses particulières. Puisque les sous-échantillons définissent aussi des régions différentes du volume de reconstruction, un flot systématique entre deux régions peut mener à un ajustement de points-zéro différents en effaçant artificiellement ce flot potentiellement réel ;

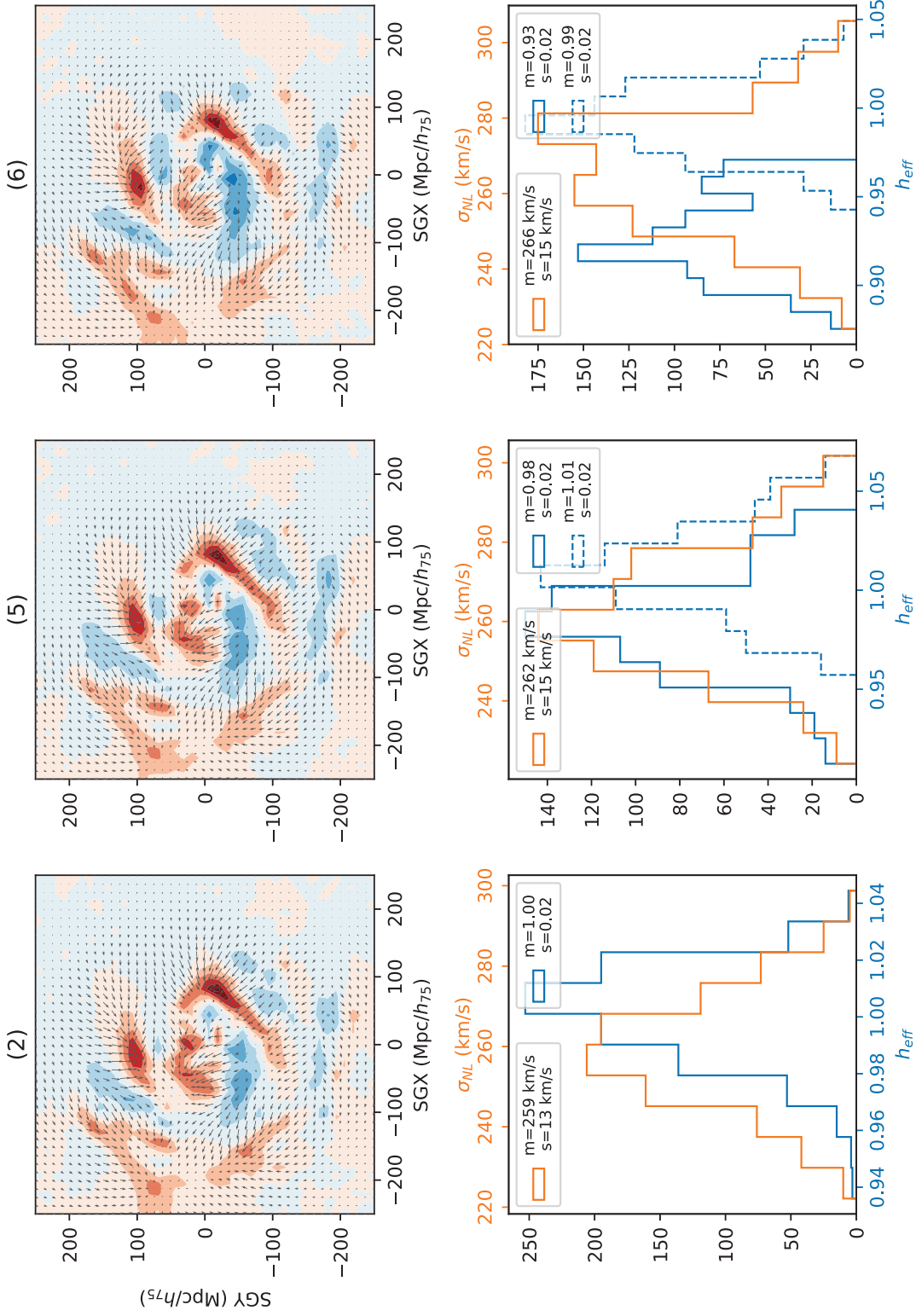


FIGURE 4.19 – La tranche  $SGZ = 0$  du champ reconstruit à partir du catalogue CF3gp-50 dans différents cas : (2) il s'agit du cas de référence, *a priori* empiriques sur les distances de chacun des sous-échantillons du catalogue; (5) *a priori* empiriques sur les distances de chacun des sous-échantillons du catalogue et une constante de Hubble effective  $h_{eff}$  spécifique pour les données 6dF (en traits pointillés); (6) *a priori* empiriques sur les distances de 6dF uniquement et une constante de Hubble effective  $h_{eff}$  spécifique pour ces mêmes données (en traits pointillés).

- enfin, le modèle ne prend pas en compte les observations multiples d'une même galaxie par différentes méthodes de mesure. C'est-à-dire qu'à chaque galaxie est attribuée une seule méthodologie et une seule constante de Hubble effective.

#### 4.4.4 Dipôle externe

On étudie enfin l'introduction d'un dipôle externe  $\mathbf{v}_{\text{ext}}$ . Le dipôle est ajouté en transformant la fonction de vraisemblance (3.12) en :

$$\mathcal{L}(\dots, \mathbf{v}_{\text{ext}}) = \prod_i \frac{1}{2\pi\sigma_{\mu,i} \left( \sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i(h_{\text{eff}}d_i))^{-2} \right)} \exp \left( -\frac{1}{2} \left( \frac{\left( \mu_i - 5 \log_{10} \frac{d_i}{10 \text{ pc}} \right)^2}{\sigma_{\mu,i}^2} + \frac{\left( v^r(z, \bar{z}_i(h_{\text{eff}}d_i)) - \mathbf{v}_{\mathbf{g}}(d_i h_{\text{eff}} \hat{\mathbf{r}}_i) \cdot \hat{\mathbf{r}}_i - \mathbf{v}_{\text{ext}} \cdot \hat{\mathbf{r}}_i \right)^2}{\sigma_{\text{NL}}^2 + c\sigma_{z,i}^2(1 + \bar{z}_i(h_{\text{eff}}d_i))^{-2}} \right) \right) \quad (4.15)$$

Ce dipôle correspond à un mouvement constant de l'ensemble du volume de reconstruction. Il peut provenir d'une mauvaise transformation de référentiel (voir section 1.2.1) ou d'un dipôle externe dû à des structures éloignées. Également, une mauvaise intercalibration des distances, par exemple entre les deux hémisphères terrestres, peut aussi être la source d'un dipôle résiduel. On compare sur la figure 4.20 les cas # 2 et # 7 correspondants au cas de référence et avec dipôle externe.

Dans le cas # 7, le dipôle est trouvé à :

$$\mathbf{v}_{\text{ext}} = (-123, -724, 372) \pm (300, 200, 280) \text{ km s}^{-1}. \quad (4.16)$$

On peut observer ce dipôle sur le panneau de droite de la figure 4.20. Ce dipôle n'a pas d'influence sur la constante de Hubble effective, mais on peut noter qu'il réduit la valeur ajustée de  $\sigma_{\text{NL}}$  d'environ  $25 \text{ km s}^{-1}$ . La présence de ce dipôle nécessite une investigation plus poussée pour en définir son origine. En particulier, on peut voir que si la composante en direction de *SGY* est relativement bien contrainte, les autres sont compatibles avec zéro. Il est possible qu'une dégénérescence dans le modèle empêche la minimisation correcte de la probabilité postérieure.

Il est finalement à noter que les structures du champ de surdensité sont extrêmement stables entre les différentes reconstructions. Seuls leur intensité et leur mouvement varient entre ces reconstructions.

## 4.5 Conclusion

Dans ce chapitre, nous avons appliqué le modèle de vitesse développé dans le chapitre 3 au catalogue *CosmicFlows-3*. Les données ont été divisées en plusieurs sous-échantillons afin de modéliser correctement les différents *a priori* sur les distances. Nous avons utilisé la version groupée du catalogue afin d'éviter les mouvements non linéaires et on a modélisé le champ de vitesse à partir des modules de distance mesurés  $\mu$  et leurs erreurs  $\sigma_{\mu}$ . Dans la section 4.2 on



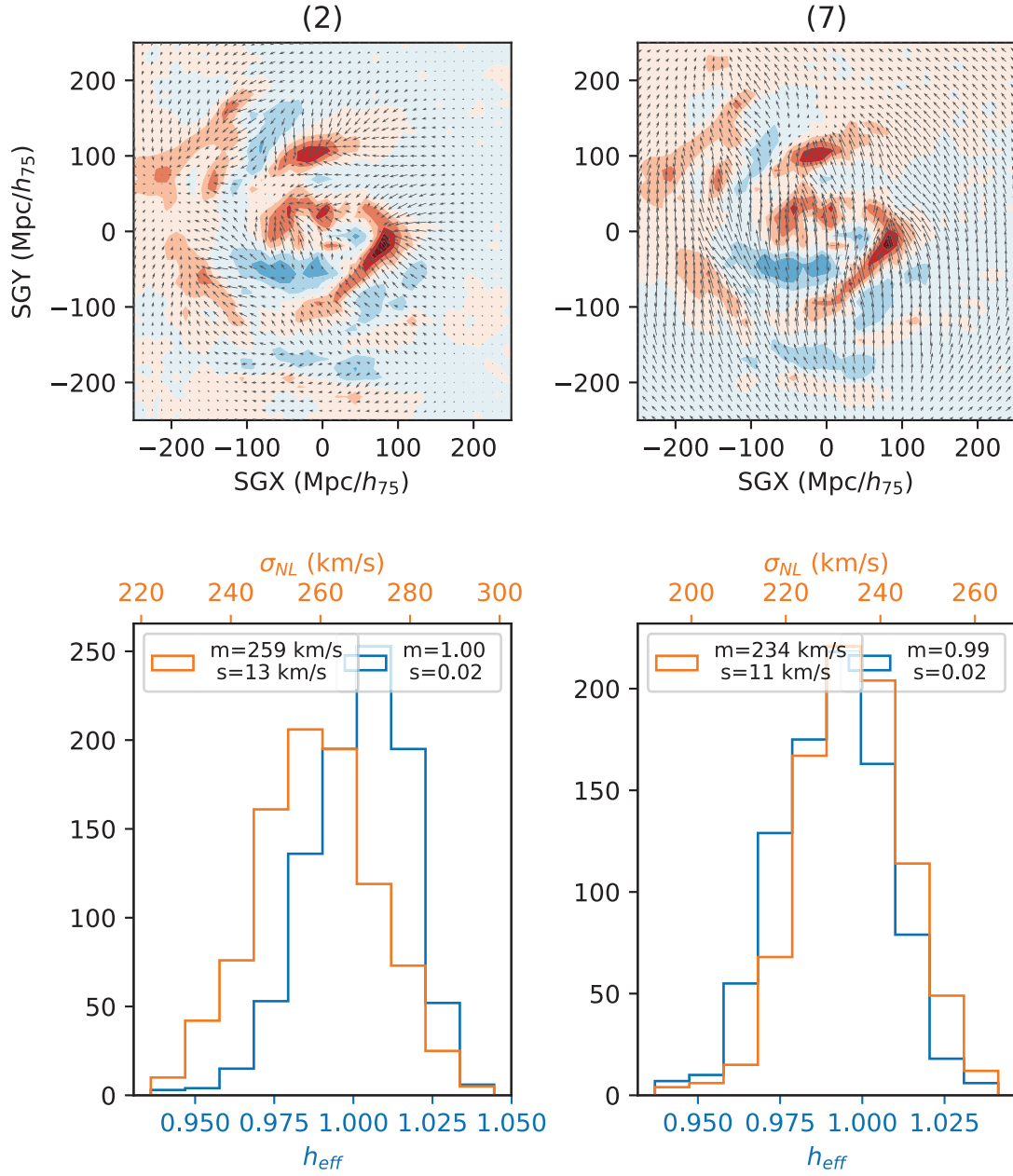


FIGURE 4.20 – La tranche  $SGZ = 0$  du champ reconstruit à partir du catalogue CF3gp-50 dans différents cas : (2) il s’agit du cas de référence, *a priori* empiriques sur les distances de chacun des sous-échantillons du catalogue ; (7) Mêmes conditions que le cas de référence (2) en ajoutant l’ajustement d’un dipôle externe  $\mathbf{v}_{ext}$ .

a brièvement analysé le champ reconstruit et cité quelques applications et analyses possibles. La suite de cette thèse aurait pu se diriger vers ces analyses.

Cependant, l'impact du modèle développé ici dépasse ce cadre et peut mener à des analyses cosmologiques plus fondamentales, en particulier l'estimation de la constante de Hubble  $H_0$  et du taux de structures des croissance  $f\sigma_8$ . En effet, ces deux paramètres ont été fixés dans notre analyse, mais peuvent être laissés libres, à condition de maîtriser la calibration absolue des données (pour  $H_0$ ) et d'avoir un modèle physique des non-linéarités (pour  $f\sigma_8$ ). Ces considérations seront développées dans le dernier chapitre de perspectives.

En attendant, pour utiliser les vitesses particulières comme sonde cosmologique, d'autres systématiques doivent être prises en compte, systématiques qui proviennent principalement de la méthode d'estimation des modules de distances à partir des paramètres intrinsèques  $\eta$  des indicateurs de distance (par exemple la largeur HI dans le cas de la TFR). En effet, rappelons-nous de l'équation (2.30) qui a motivé le modèle développé dans le chapitre 3 :

$$\mathcal{P}(r|\mu, \mathcal{O}) \propto \mathcal{P}(\mathcal{O}|\mu, r)\mathcal{P}(\mu|r)\mathcal{P}(r) \quad (4.17)$$

Cette équation donne la probabilité postérieure de la distance  $r$  d'une galaxie à partir de son observation  $\mathcal{O}$  et la mesure de son module de distance  $\mu$ . Un meilleur modèle est l'estimation de la probabilité postérieure  $\mathcal{P}(r|\eta, \mathcal{O}, z)$  où  $\eta$  sont les paramètres indépendants des indicateurs de distance. Strauss and Willick (1995) ont montré que différents biais (de l'ordre de 10% sur la distance) existaient en fonction de si l'on considère la magnitude absolue  $M$  ou le paramètre  $\eta$  comme paramètre indépendant de la relation de l'indicateur de distance. Kelly (2007) montre également que ce choix a un impact sur la distribution postérieure du paramètre inféré. De plus, les effets de sélection ne peuvent être correctement modélisés que dans l'espace des observations, c'est-à-dire  $\eta$ , et non des paramètres dérivés  $\mu$ .

Naturellement, donc, le prochain chapitre va tenter d'étendre le modèle précédent afin de modéliser la calibration des indicateurs de distance en même temps que la reconstruction du champ de vitesse. Tous les indicateurs de distance étant très spécifiques, on ne s'intéressera uniquement qu'à la relation de Tully-Fisher, qui est l'une des deux techniques majoritaires dans les catalogues *CosmicFlows*. On rappelle que l'objectif est d'estimer les paramètres  $H_0$  et  $f\sigma_8$ , ce qui requiert le plus grand soin dans la calibration des indicateurs de distance.

## Chapitre 5

# Modélisation bayésienne de la relation de Tully-Fisher

La conclusion partielle du chapitre précédent est sans appel : afin de maîtriser les effets de sélection et de contrôler les erreurs systématiques provenant de la calibration des indicateurs de distance, il faut modéliser les données observationnelles primaires plutôt que les modules de distance dérivés. Puisque cette problématique découle naturellement du modèle présenté dans le chapitre 3, ce problème a été partiellement traité au cours de cette thèse. Le premier indicateur de distance étudié est la relation de Tully-Fisher (TFR). Le choix de cet indicateur est naturel puisque l'apport de la collaboration *CosmicFlows* dans les données de vitesses particulières se fait surtout via la TFR, et les données sont contrôlées de l'observation à l'analyse. L'idéal d'une modélisation *end-to-end* pour la collaboration *CosmicFlows* est de développer un modèle statistique décrivant la spectroscopie, la photométrie, l'inclinaison et le décalage vers le rouge des galaxies en fonction du champ de vitesse sous-jacent. Ce projet n'est pas utopique mais requiert plus de temps qu'il est possible de demander pour une thèse de doctorat. Ainsi ce chapitre se consacre-t-il uniquement à la modélisation d'observables dérivées (largeur HI, magnitudes, voir ci-après), en fonction du champ de vitesse. Ce choix repose sur des hypothèses qui seront discutées en fin de chapitre. Ce chapitre est construit de la façon suivante : dans la section 5.1 nous présentons d'abord la relation de Tully-Fisher, puis motivons la nécessité d'une nouvelle modélisation statistique. La méthode est décrite dans la section 5.2 et testée sur des catalogues simulés dans la section 5.3. Enfin, elle est appliquée aux données SFI++ dans la section 5.4

### Sommaire

---

<b>5.1</b>	<b>Introduction</b>	<b>114</b>
5.1.1	La relation de Tully-Fisher	114
5.1.2	De la nécessité d'une modélisation précise	115
<b>5.2</b>	<b>Le modèle</b>	<b>116</b>
5.2.1	Vraisemblance	116
5.2.2	<i>A priori</i>	119
5.2.3	L'algorithme	120
<b>5.3</b>	<b>Tests</b>	<b>120</b>



5.3.1	Catalogue simulé . . . . .	120
5.3.2	Relation de Tully-Fisher . . . . .	121
5.3.3	Couplage avec le champ de vitesse . . . . .	122
5.3.4	Avec la calibration . . . . .	125
<b>5.4</b>	<b>SFI++ . . . . .</b>	<b>128</b>
<b>5.5</b>	<b>Discussion . . . . .</b>	<b>129</b>
<b>5.6</b>	<b>Conclusion . . . . .</b>	<b>131</b>

---

## 5.1 Introduction

### 5.1.1 La relation de Tully-Fisher

La relation de Tully-Fisher est une relation empirique mise au jour par Tully and Fisher (1977). Elle donne le lien entre la luminosité d’une galaxie spirale et sa vitesse de rotation (Kauffmann et al., 1993; Steinmetz and Navarro, 1999; Aquino-Ortíz et al., 2018). Une interprétation physique peut être faite en considérant le théorème du Viriel. Si l’on considère que la brillance de surface des galaxies spirales est constante et que leur rapport masse-luminosité aussi, on peut facilement montrer que :

$$L \propto V_{\text{rot}}^4 \quad (5.1)$$

où  $L$  est la luminosité de la galaxie et  $V_{\text{rot}}$  sa vitesse de rotation. En pratique, les observations donnent accès à la magnitude absolue  $M$  plutôt qu’à la luminosité  $L$ . De plus, la mesure de la vitesse de rotation se fait en général par radioastronomie, via l’étude du dédoublement de la raie d’hydrogène neutre (HI) dans les galaxies spirales à cause de l’effet Doppler. Puisque l’effet Doppler observé est radial, la vitesse de rotation ainsi mesurée doit être déprojetée. Si  $i$  est l’inclinaison de la galaxie, la relation de Tully-Fisher s’écrit finalement :

$$M = ZP + s \log \frac{W_{HI}}{\sin i} \quad (5.2)$$

où  $ZP$  et  $s$  sont le point zéro et la pente de la relation. Puisque la magnitude est définie par  $M = -2.5 \log_{10} L/L^0$ , la pente  $s$  doit être théoriquement telle que  $s = -2.5 \times 4 = -10$ , et on l’observe généralement autour de cette valeur bien que de nombreux effets rentrent en compte dans son évaluation. Il est à noter que la TFR est employée à d’autres fins que la mesure de distances extragalactiques, comme l’étude de la cinématique des galaxies ou comme sonde observationnelle dans la théorie MOND (*MOdified Newtonian Dynamics* Famaey and McGaugh (2012)). Nous nous restreignons ici à l’étude de la TFR en tant qu’indicateur de distance.

Dans ce cadre, la TFR est généralement utilisée de la façon suivante (Springob et al., 2009; Hong et al., 2013; Tully et al., 2013; Lagattuta et al., 2013; Sorce et al., 2014) : des magnitudes apparentes  $m$ , largeurs HI  $W_{HI}$  et inclinaisons  $i$  sont mesurées pour des galaxies dont on connaît les distances par ailleurs (galaxies que l’on appellent les *calibrateurs*), puis les paramètres  $ZP$  et  $s$  sont ajustés à partir de ces calibrateurs dont la magnitude absolue est donnée par  $M = m - 5 \log \frac{d}{10 \text{ pc}}$ . On peut alors, à partir de la mesure de n’importe quelle

largeur HI  $W_{HI}$  déterminer la magnitude absolue d'une galaxie via la relation (5.2) calibrée. On attribue en général une erreur gaussienne aux magnitudes absolues ainsi déterminées, erreur que l'on tire de la dispersion observée de la TFR.

### 5.1.2 De la nécessité d'une modélisation précise

Deux problèmes majeurs se posent en procédant ainsi. Le premier est la mésestimation de l'erreur sur la distance de certaines galaxies, puisqu'on ne prend en compte ni l'erreur sur la mesure de son inclinaison, sa magnitude ou largeur HI ; ni le fait que la dispersion intrinsèque de la TFR n'est pas égale à sa dispersion observée. Le deuxième est la perte d'information induite du découplage entre le champ de vitesse et la relation de Tully-Fisher elle-même. En effet, à partir d'un champ de vitesse estimé, il est nécessaire de replacer les galaxies sur le diagramme de Tully-Fisher et de vérifier *a posteriori* que les distances ainsi inférées sont compatibles avec sa calibration. Cela permettrait de tester à la fois le modèle de champ de vitesse et de la TFR. Après tout, l'information de base est toujours la même : les décalages vers le rouge, les magnitudes apparentes et les distances des calibrateurs. Nous argumentons dans cette sous-section la nécessité d'une modélisation précise.

#### 5.1.2.1 Propagation des erreurs individuelles

Au sein de la collaboration *CosmicFlows*, l'erreur attribuée à la magnitude absolue des galaxies est usuellement donnée par la dispersion observée de la TFR. Si ce choix est sûr et conservateur, il n'optimise pas l'utilisation des données disponibles. En effet, une galaxie ayant une grande erreur de mesure sur sa vitesse de rotation, peut devenir une donnée aberrante si on lui attribue une erreur indépendante. Modéliser complètement les observations de la TFR permettrait de propager chacune des erreurs individuelles au champ de vitesse reconstruit, et d'éviter des biais dus à la mésestimation des erreurs sur les distances. Ceci est d'autant plus vrai que le modèle présenté dans le chapitre 3 n'inclut pas de modélisation de données aberrantes, et que la distribution supposée gaussienne des observations de module de distance peut biaiser la reconstruction. Les erreurs individuelles des données utilisées pour la TFR sont les suivantes :

- l'erreur sur les magnitudes apparentes ;
- sur les vitesses de rotation mesurées ;
- sur les inclinaisons.

Le modèle présenté en section 5.2 prend en compte chacune de ces erreurs observationnelles.

#### 5.1.2.2 Erreurs systématiques

L'estimation du point zéro et de la pente de la TFR est toujours assortie d'erreurs. Ces erreurs sont systématiques dans le sens où elles affectent toutes les distances de la même manière. Ainsi, les propager simplement comme erreurs individuelles sans prendre en compte cette corrélation biaise le champ de vitesse reconstruit. L'erreur sur le point zéro est indirectement propagée via l'ajout du paramètre  $h_{\text{eff}}$  défini dans le chapitre 3. Cependant, si l'on veut estimer la constante de Hubble localement, il faut s'affranchir de ce paramètre. L'incertitude sur la pente de la relation, elle, n'est pour le moment pas propagée. Le modèle présenté en

section 5.2 permet d'ajuster le point zéro de la relation ainsi que sa pente en même temps que le champ de vitesse.

### 5.1.2.3 Effets de sélection

Les effets de sélection ont été décrits dans les chapitres 2 et 4. La conclusion de cette discussion est la nécessité d'améliorer notre modélisation des effets de sélection en prenant en compte la sélection des données initiales (largeur HI, etc.). En ajoutant la TFR dans le modèle du champ de vitesse, il est naturel d'ajouter les paramètres de la TFR dans les *a priori* et de ce fait d'imposer des *a priori* directement sur les quantités mesurées, comme les inclinaisons, vitesses de rotation et magnitudes, plutôt que sur les distances.

## 5.2 Le modèle

Nous souhaitons donc construire un modèle prenant en compte les mesures de vitesse de rotation, d'inclinaison et de magnitudes, leurs erreurs respectives et l'incertitude sur la détermination des coefficients  $ZP$  et  $s$ . Pour cela, on essaye d'estimer la probabilité conjointe de  $ZP$ ,  $s$  et le champ de vitesse  $\mathbf{v}$ .

Le problème statistique posé par la calibration de Tully-Fisher est l'ajustement d'une droite sur des données définies par une variable dépendante de plusieurs variables indépendantes. En astrophysique, le choix de la variable indépendante est souvent sujette à discussion. Il est possible de démontrer (Kelly, 2007) que si l'échantillon étudié a été sélectionné sur les variables indépendantes, et que les erreurs de mesure sur les variables dépendantes et indépendantes ne sont pas corrélées, alors l'inférence des paramètres de régression n'est pas affectée par les effets de sélection. Ce résultat est primordial dans notre modèle. Puisque l'échantillon de données pour calibrer une relation de Tully-Fisher souffre d'effets de sélection provenant principalement de (i) les effets de sélection en magnitude apparente  $m$ , et (ii) de la fonction de luminosité des galaxies ; nous supposons ici que la variable dépendante est la vitesse de rotation, ou plus exactement son logarithme  $\log W$ . En procédant ainsi, la magnitude absolue  $M$  sera la variable indépendante, et les effets de sélection sur celle-ci n'affecteront pas l'estimation des paramètres  $ZP$  et  $s$ . Ainsi la relation de Tully-Fisher dans le cadre de notre méthode s'écrit :

$$\log W = a + bM. \quad (5.3)$$

Cette équation est appelée dans la littérature la relation de Tully-Fisher inverse (ci-après ITFR) (Willick and Strauss, 1998) . Les données et paramètres du modèle sont résumés dans le tableau 5.1.

### 5.2.1 Vraisemblance

Plusieurs ajustements sont nécessaires pour compléter l'équation 5.3. Pour chaque galaxie nous introduisons son inclinaison  $i$ , sa couleur  $c$ , et pour l'ensemble de l'échantillon une correction de couleur  $c_c$  et une dispersion intrinsèque gaussienne  $\sigma_{\text{int}}$  :

$$\log \frac{W}{\sin i} = a + \mathbf{b}^T \mathbf{x} + \epsilon \quad (5.4)$$

TABLE 5.1 – Données et paramètres utilisés dans le modèle développé dans le chapitre 5.

Données	Description	Nombre	Valeurs pour CF3 et <i>a priori</i>
$N$	Nombre de galaxies et groupes	1	$N = 11483$
$L$	Côté du volume de reconstruction	1	$L = 800 \text{ Mpc}/h_{75}$
$M_g$	Taille de la grille de reconstruction	1	$M_g = 128$
$m$	Magnitudes apparentes	$N$	Distribuées normalement avec erreur $\sigma_m$ .
$i$	Inclinaisons mesurées	$N$	Distribuées normalement avec erreur $\sigma_i$ .
$W_{HI}$	Largeurs HI	$N$	Distribuées normalement avec erreur $\sigma_W$ .
$c$	Couleurs	$N$	Supposées sans erreurs.
$z$	Décalages vers le rouge	$N$	Distribués normalement avec erreur $\sigma_z$
Paramètres ajustés			
$\mathbf{v}_g$	Champ de vitesse linéaire échantillonné sur une grille	$M^3$	<i>a priori</i> $\Lambda$ CDM défini par l'équation (3.14)
$M$	Magnitudes absolues	$N$	<i>a priori</i> empiriques définis par les équations 3.17 et 3.16 et/ou uniformes
$a$	Point zéro de l'ITFR	1	<i>A priori</i> uniforme.
$\mathbf{b}$	Pente de l'ITFR	1	<i>A priori</i> uniforme.
$\sigma_{\text{int}}$	Dispersion gaussienne modélisant la dispersion intrinsèque de l'ITFR	1	<i>A priori</i> uniforme sur $\sigma_{\text{int}}^2$
$\sigma_{\text{NL}}$	Dispersion gaussienne modélisant l'écart des observations à la linéarité	1	<i>A priori</i> uniforme sur $[50, 1500] \text{ km/s}$
$(a', b', c')$	Hyperparamètres définissant la forme des fonctions d' <i>a priori</i> sur les magnitudes absolues	3	<i>A priori</i> uniformes

où  $\mathbf{b} = \begin{pmatrix} b \\ c_c \end{pmatrix}$ ,  $\mathbf{x} = \begin{pmatrix} M \\ c \end{pmatrix}$  et  $\epsilon$  est l'erreur normale due à la dispersion intrinsèque :

$$\epsilon \sim \mathcal{N}(0, \sigma_{\text{int}}^2). \quad (5.5)$$

Pour prendre en compte les erreurs de mesure sur la vitesse de rotation  $W$  et sur l'inclinaison  $i$ , nous introduisons les paramètres intermédiaires  $l\hat{W}$ ,  $\hat{i}$  tels que le modèle peut être exprimé hiérarchiquement de la façon suivante :

$$i \sim \mathcal{N}(\hat{i}, \sigma_i^2) \quad (5.6)$$

$$W \sim \mathcal{N}(10^{l\hat{W}} \sin \hat{i}, \sigma_W^2) \quad (5.7)$$

$$l\hat{W} \sim \mathcal{N}(a + \mathbf{b}^T \mathbf{x}, \sigma_{\text{int}}^2). \quad (5.8)$$

Les équations (5.6) et (5.7) décrivent les mesures tandis que l'équation (5.11) décrit l'ITFR elle-même.

Si l'on connaissait les magnitudes absolues  $M$  des galaxies, ce système suffirait à estimer les paramètres  $a$  et  $\mathbf{b}$ . Cependant, elles sont inconnues, mais peuvent être déduites du champ de vitesse. En effet, à partir de l'estimation de la vitesse particulière  $v^r$  et du décalage vers le rouge, on peut déterminer le module de distance d'une galaxie via l'équation (1.47). Depuis ce module de distance, on en déduit :

$$M = m - \mu. \quad (5.9)$$

La connaissance du champ de vitesse permet donc d'accéder à la magnitude absolue des galaxies. On tire donc partie du modèle présenté dans le chapitre 3 et introduisons la magnitude absolue  $M$  comme paramètre du modèle, de la même façon que les distances  $d$  étaient des paramètres libres dans le chapitre 3. En ne prenant en compte que le champ de vitesse et la magnitude, la distribution de  $M$  est alors :

$$M \propto \frac{1}{\sqrt{2\pi(\sigma_{\text{NL}}^2 + \sigma_{cz}^2(1 + \bar{z})^{-2})}} \exp\left(-\frac{1}{2} \frac{\left(v^r(z, m - M) - \mathbf{v}_g\left(10^{\frac{m-M}{5}-5}\hat{\mathbf{r}}\right)\right)^2}{\sigma_{\text{NL}}^2 + \sigma_{cz}^2(1 + \bar{z})^{-2}}\right) \quad (5.10)$$

où  $\bar{z}$  dépend implicitement de  $m$  et  $M$ , et  $v^r(z, \mu)$  est la fonction qui donne la vitesse radiale en fonction du décalage vers le rouge et du module de distance. Le modèle pour  $\mathbf{v}$ ,  $\sigma_{\text{NL}}$  reste le même que celui donné dans le chapitre 3.

Au final, la vraisemblance du modèle est :

$$\begin{aligned} \mathcal{P}(i, W, m, z | \alpha, \mathbf{b}, \sigma_{\text{int}}, \sigma_{\text{NL}}, \mathbf{v}_g, \hat{i}, l\hat{W}, \mathbf{x}) = & \mathcal{N}\left(v^r(z, m - M); \mathbf{v}_g\left(10^{\frac{m-M}{5}-5}\hat{\mathbf{r}}\right), \sigma_{\text{NL}}^2 + \sigma_{cz}^2(1 + \bar{z})^{-2}\right) \\ & \times \mathcal{N}(i; \hat{i}, \sigma_i^2) \mathcal{N}(W; 10^{l\hat{W}} \sin \hat{i}, \sigma_W^2) \mathcal{N}(l\hat{W}; a + \mathbf{b}^T \mathbf{x}, \sigma_{\text{int}}^2) \end{aligned} \quad (5.11)$$

### 5.2.2 *A priori*

Les *a priori* sur les inclinaisons  $\hat{i}$ , la pente  $\mathbf{b}$ , le point zéro  $a$  et la dispersion intrinsèque sont pris uniformes. L'important est de modéliser les effets de sélection et de distribution intrinsèque de la magnitude absolue  $M$ . En général, ces effets de sélection dépendent des effets de sélection sur la magnitude apparente  $m$  et de la distribution des distances réelles  $r$ . Pour être plus précis, l'*a priori* sur  $M$  peut être modélisé par la probabilité jointe suivante, où  $\mathcal{O}$  représente le fait d'observer une galaxie :

$$\mathcal{P}(M, \mathcal{O}) = \mathcal{P}(\mathcal{O}|M)\mathcal{P}(M) \quad (5.12)$$

$$= \int \mathcal{P}(\mathcal{O}|M, r)\mathcal{P}(r)dr\mathcal{P}(M) \quad (5.13)$$

$$= \int \mathcal{P}\left(\mathcal{O}|m = M + 5 \log \frac{r}{10 \text{ pc}}\right) \mathcal{P}(r)dr\mathcal{P}(M). \quad (5.14)$$

En pratique, la probabilité  $\mathcal{P}(M)$  peut être modélisée par une fonction de Schechter (2.34). Cependant, si l'on mesure la magnitude apparente  $m$  de la galaxie, l'intégrale de l'équation (5.12) n'a pas besoin d'être calculée puisque la distance découle de  $\mu = m - M$ . On a donc :

$$\mathcal{P}(M, \mathcal{O}) = \mathcal{P}(r(M, m))\mathcal{P}(M). \quad (5.15)$$

Dans la version actuelle de l'algorithme,  $\mathcal{P}(M)$  est ajustée par une des fonctions d'*a priori* présentées dans la section 3.1.3.2 (équations (3.16), (3.17) et (3.18)). Cette façon de faire est similaire au traitement des *a priori* sur les distances ajustées dans le modèle de vitesses présenté dans le chapitre 3.

On note aussi le système d'*a priori* proposé par (Kelly, 2007). Dans ce modèle flexible, la variable indépendante  $\mathbf{x}$  est supposée provenir d'un modèle de mélange gaussien :

$$\mathbf{x} \sim \sum_{k=1}^K \pi_k \mathcal{N}_2(\mu_k, \tau_k^2) \quad (5.16)$$

où  $K$  est généralement compris entre 1 et 3 et correspond au nombre de gaussiennes du mélange. Le mélange gaussien est paramétrisé de telle façon que

$\mathcal{P}(a, \mathbf{b}, \mu_0, u, w, \tau) \propto \mathcal{P}(\pi)\mathcal{P}(\mu|\mu_0, u^2) \mathcal{P}(\tau^2|w^2) \mathcal{P}(u^2|w^2)$  :

$$a, \mathbf{b} \sim \text{Unif}(-\infty, \infty) \quad (5.17)$$

$$\sigma_{\text{int}}^2 \sim \text{Unif}(0, \infty) \quad (5.18)$$

$$\pi \sim \text{Dirichlet}(1, \dots, 1) \quad (5.19)$$

$$\mu_k | \mu_0 u^2 \sim \mathcal{N}(\mu_0, u^2) \quad (5.20)$$

$$\tau_k, u^2, w^2 \sim \text{Inv} - \chi^2(1, w^2) \quad (5.21)$$

$$\mu_0 \sim \text{Unif}(-\infty, \infty) \quad (5.22)$$

$$w^2 \sim \text{Unif}(0, \infty) \quad (5.23)$$

Bien que ce modèle d'*a priori* soit intégré à l'algorithme développé ici, nous utiliserons dans la suite les *a priori* empiriques présentés dans la section 3.1.3.2.

### 5.2.3 L'algorithme

Nous choisissons d'utiliser l'échantillonnage de Gibbs pour tirer chacun des paramètres du modèle présenté ci-dessus. Ce choix s'appuie sur deux raisons : la première est la facilité à coupler le modèle de la TFR avec le modèle du champ de vitesse présenté dans le chapitre 3, qui utilise aussi l'échantillonnage de Gibbs. La deuxième est la possibilité d'utiliser l'algorithme proposé par Kelly (2007). De façon générale, l'espace des paramètres  $(M, \hat{i}, l\hat{W})$  contient potentiellement plusieurs dizaines de milliers de paramètres, et l'échantillonnage de Gibbs est le plus adapté aux modèles à grand nombre de paramètres.

Chacun des paramètres est donc tiré des probabilités conditionnelles qui sont décrites dans les sections 5.2.2 et 5.2.1. Contrairement à l'algorithme utilisé pour échantillonner le champ de vitesse décrit dans la section 3.2, la génération des paramètres du modèle de la TFR ne présente pas de difficulté de calcul particulière.

## 5.3 Tests

Dans cette section, nous testons l'algorithme d'ajustement de la TFR et son couplage avec le champ de vitesse.

### 5.3.1 Catalogue simulé

Nous utilisons le catalogue qui a été simulé dans la section 2.2. À partir des distances générées, on construit une TFR de la façon suivante :

- on fixe un point zéro  $a$ , une pente  $\mathbf{b}$  et une dispersion intrinsèque  $\sigma_{\text{int}}$ ,
- des magnitudes absolues  $M$  sont générées selon une normale asymétrique de moyenne  $-21$ , de déviation standard 2 et d'asymétrie 5 :

$$M \sim \text{skew-}\mathcal{N}(-21, 2, 5); \quad (5.24)$$

- des couleurs  $c$  sont générées à partir des magnitudes absolues :

$$c \sim \mathcal{N}\left(\frac{M - \bar{M}}{\sigma_M}, 0.25^2\right) \quad (5.25)$$

afin de compléter le vecteur  $\mathbf{x} = \begin{pmatrix} M \\ c \end{pmatrix}$ . Notons qu'on a corrélé les couleurs et les magnitudes absolues pour se rapprocher des données observationnelles : les galaxies les plus rouges sont en général aussi les plus brillantes.

- Les magnitudes apparentes  $m$  sont déduites de  $\mu = m - M$ , et on rajoute une dispersion de 0.1 mag.
- on génère  $\log W$  à partir de la loi normale  $\mathcal{N}(a + \mathbf{b}^T \cdot \mathbf{x}, \sigma_{\text{int}}^2)$ ,
- les inclinaisons sont tirées uniformément sur l'intervalle  $[\pi/6, \pi/2]$ ,
- les largeurs HI observées sont tirées de la distribution

$$W_{HI} \sim 10^{\mathcal{N}(\log W + \sin i, \sigma_{iw}^2)}, \quad (5.26)$$

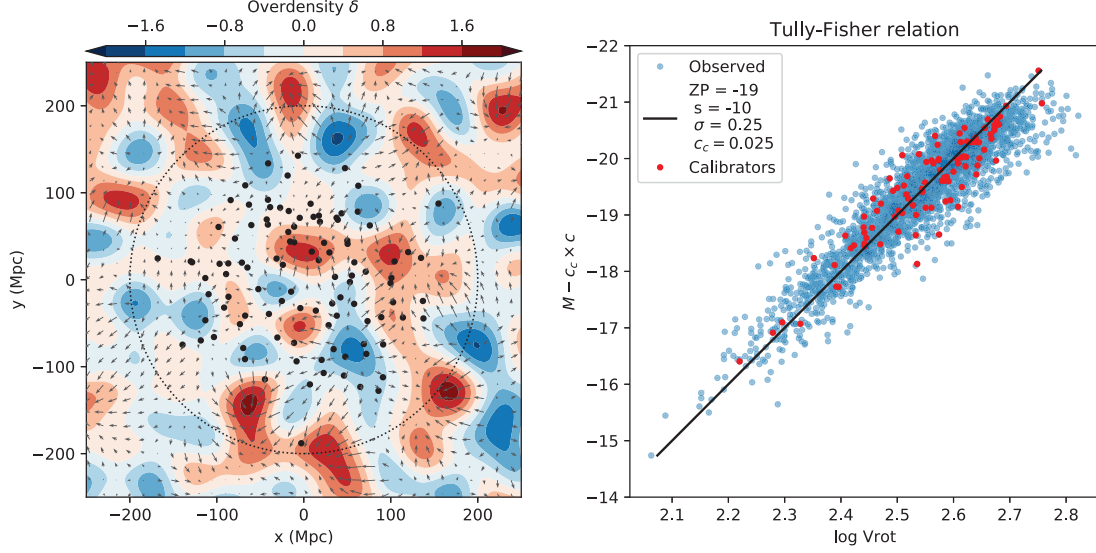


FIGURE 5.1 – À gauche : tranche  $z = 0$  du champ de surdensité simulé. À droite : relation de Tully Fisher simulée. Les points bleus correspondent aux observations, les points rouges aux calibrateurs et la courbe noire représente la TFR utilisée pour générer les données.

— enfin, 100 calibrateurs sont tirés aléatoirement. On considère qu’on connaît leurs magnitudes absolues à 0.05 mag près.

On rappelle que les notations  $ZP, s, \sigma$  correspondent aux point-zéro, pente et dispersion intrinsèque si l’on considère la droite  $M = f(\log W)$  (TFR) alors que les notations  $a, b, \sigma_{\text{int}}$  correspondent aux mêmes paramètres si l’on considère la droite  $\log W = f(M)$  (ITFR).

La figure 5.1 montre le champ simulé et la TFR associée. On choisit ici  $a = 0.6$ ,  $\mathbf{b} = (0.1, 0.025)$ ,  $\sigma_{\text{int}}^2 = 6 \times 10^{-4}$ , ce qui correspond à  $ZP = -19$ ,  $s = -10$  et  $\sigma = 0.25$ . La dispersion observée sur la TFR est  $\sigma^{\text{obs}} = 0.4$ , ce qui correspond approximativement à ce qui est généralement observé.

### 5.3.2 Relation de Tully-Fisher

Nous testons l’algorithme défini dans ce chapitre dans le cas simple où ni les distances, ni le champ de vitesse ne sont ajustés. On considère en outre que les *a priori* sont connus. L’espace des paramètres dans ce cas est  $\theta = \{\hat{l}\hat{W}, \hat{i}, a, \mathbf{b}, \sigma_{\text{int}}\}$ . Ce test permet de voir si l’on parvient à retrouver les paramètres principaux de la TFR -point zéro, pente et dispersion intrinsèque- à partir des données bruitées. La figure 5.2 montre les histogrammes des résidus normalisés (*pull*) pour les paramètres  $ZP, s$  et  $\sigma_{\text{int}}$  ainsi que la TFR ajustée. On retrouve  $ZP = -19.01 \pm 0.01$ ,  $s = -9.7 \pm 0.2$ ,  $\sigma_{\text{int}} = 0.24 \pm 0.01$  et  $c_c = 0.029 \pm 0.003$ . Les valeurs de  $ZP, s$  et  $\sigma_{\text{int}}$  sont donc au maximum à  $2\sigma$  des valeurs utilisées pour la simulation. Ces paramètres étant corrélés entre eux, on trace sur la figure 5.3 la distribution des pas de Markov dans les plans formés par les paires de paramètres dans l’espace  $(ZP, s, \sigma_{\text{int}})$ . Il est intéressant de noter la forte corrélation



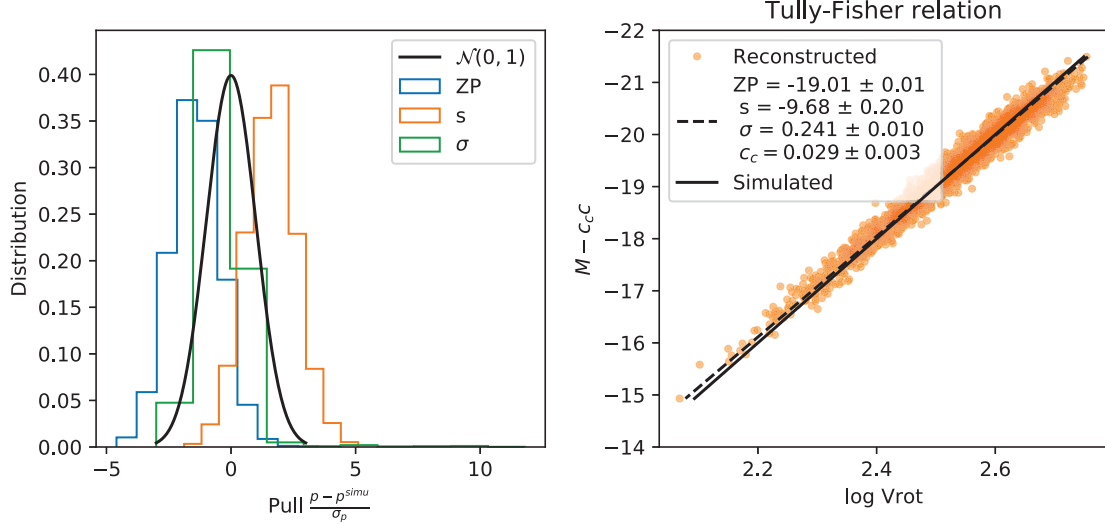


FIGURE 5.2 – Ajustement de la TFR seule pour les données simulées. À gauche : distribution des résidus normalisés pour les paramètres  $ZP$ ,  $s$  et  $\sigma_{\text{int}}$ . L’axe des abscisses représente l’écart à la valeur initiale en unité de déviation standard. À droite : TFR et données ajustées.

entre le terme de couleur  $c_c$  et les autres paramètres, qui provient du fait que les couleurs et les magnitudes absolues sont étroitement liées. Sur le panneau de droite de la figure 5.2, on voit que la droite ajustée est similaire à la droite originale. Avec 3000 points répartis dans une sphère d’environ 200 Mpc de rayon, l’erreur sur le point zéro et la pente est de quelques pourcents.

### 5.3.3 Couplage avec le champ de vitesse

Dans cette section nous étudions comment l’algorithme présenté dans ce chapitre permet de reconstruire le champ de vitesse des galaxies dont nous observons le décalage vers le rouge, la vitesse de rotation, la couleur, l’inclinaison et la magnitude apparente.

#### 5.3.3.1 Sans calibration ni ajustement des *a priori*

Supposons d’abord que la TFR soit calibrée de telle sorte que nous connaissions exactement la valeur des paramètres  $ZP$ ,  $H_0$  et  $c_c$ , qui sont fixés à leur valeur de référence. Le problème de la mesure du point zéro et de  $H_0$  sera discuté dans la section 5.3.4. L’espace des paramètres est alors  $\theta = \{l\hat{W}, \hat{i}, \mathbf{v}_g, M, \mathbf{b}, \sigma_{\text{int}}, \sigma_{\text{NL}}\}$ . On trace sur la figure 5.4 le champ reconstruit ainsi que la distribution des résidus normalisés par rapport au champ simulé. On observe à partir du panneau de droite que les résidus sont proches d’une loi normale unitaire. Le champ de surdensité reconstruit est comparable à celui de la figure 5.1 bien que le contraste ait été amoindri par les erreurs observationnelles (on rappelle à ce propos que la dispersion observée

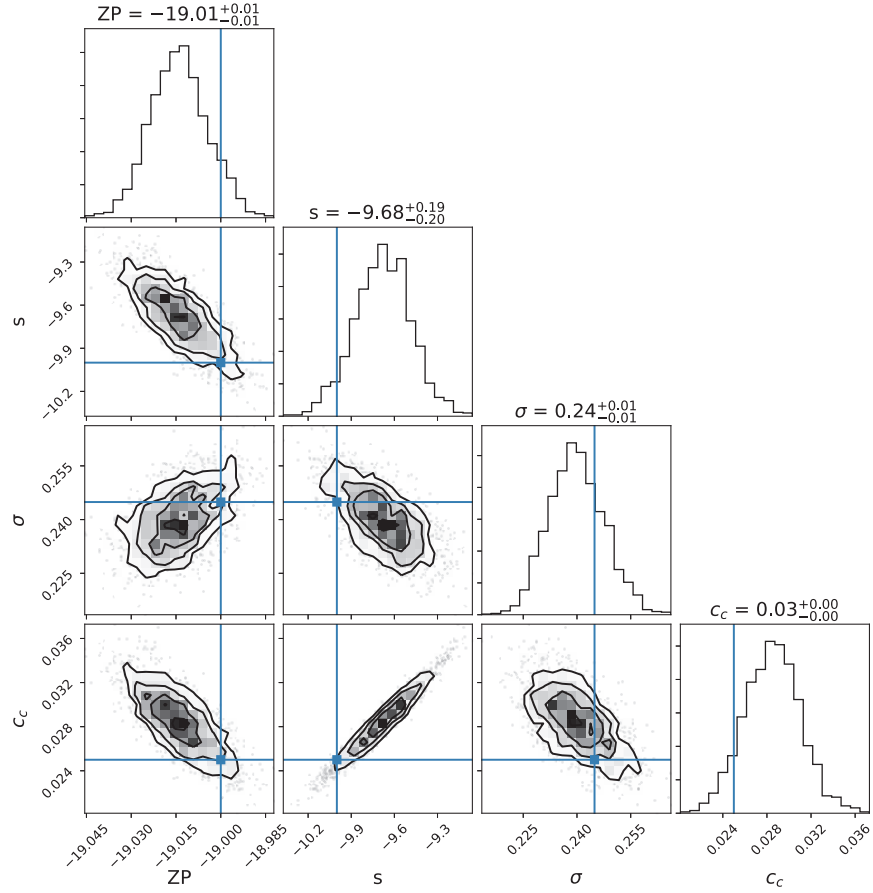


FIGURE 5.3 – Figure en coin des paramètres  $ZP$ ,  $s$ ,  $\sigma_{\text{int}}$ ,  $cc$  pour l’ajustement de la TFR seule. Chaque point correspond à un pas de Markov, et les traits bleus représentent les valeurs utilisées dans la simulation.

sur la TFR est d’environ 0.4 mag). On voit notamment que la plupart des structures ont été reconstruites.

Après avoir testé la méthode ainsi que son implémentation, on s’intéresse maintenant à l’avantage d’utiliser une telle méthode plutôt que de procéder en découplant l’estimation des paramètres de la TFR et l’estimation du champ de vitesse comme il est d’usage. En particulier, y a-t-il un gain d’information à propager toutes les erreurs depuis les vitesses de rotation jusqu’au champ de vitesse plutôt que d’utiliser les modules de distance directement ? Pour s’en convaincre, nous procédons comme suit. Étant données les mesures  $m$ ,  $W_{HI}$ ,  $i$ ,  $c$  ainsi que les paramètres (supposés connus) de la TFR  $a$  et  $\mathbf{b}$ , on évalue les modules de distances directement à partir de la TFR. On définit les modules de distances suivants :

$$\mu^{TF} \triangleq m - \frac{\log(W_{HI}/\sin i) - a - c_c c}{\mathbf{b}(0)} \quad (5.27)$$

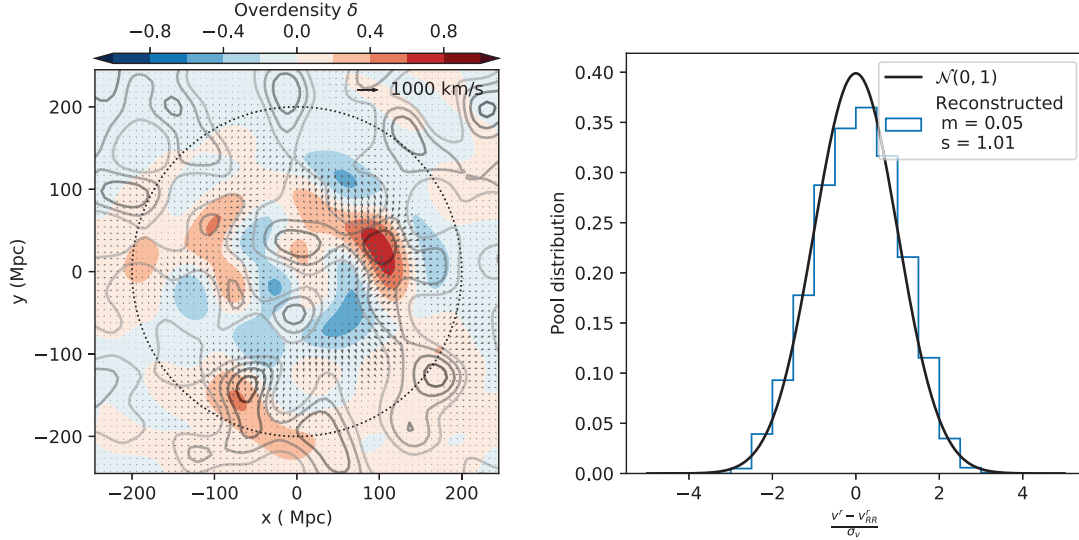


FIGURE 5.4 – Ajustement du champ de vitesse conjointement à la TFR, dans le cas où le point zéro et les *a priori* sont connus. À gauche : champs de surdensité et de vitesse reconstruits. Les contours gris correspondent au champ de surdensité du champ initial, pour comparaison. À droite : distribution des résidus normalisés des vitesses radiales pour les points de la grille à l’intérieur du cercle en traits pointillés de rayon  $R = 200$  Mpc. Si la distribution postérieure est une normale unité on peut considérer que la reconstruction n’est pas biaisée.

qui correspondent aux modules de distance déduites de la TFR. En moyenne, les distances ne sont pas biaisées puisqu’on trouve  $\langle \mu^{TF} - \mu \rangle = 0$ . Pour définir l’erreur sur ces modules de distance, on compare les modules de distance connus (ceux des calibrateurs) à ceux déduites de l’équation (5.27). En moyenne, on a  $\langle \mu^{TF} - \mu^{\text{calib}} \rangle \sim 0$  et  $\sqrt{\langle (\mu^{TF} - \mu^{\text{calib}})^2 \rangle} \sim 0.4$ . On définit donc une erreur globale sur tous les modules de distance de  $\sigma_\mu = 0.4$ . Cette façon de faire est usuelle dans la collaboration *CosmicFlows* et permet d’avoir un budget d’erreur constant pour toutes les galaxies. À partir de cette estimation, on reconstruit le champ de vitesse en utilisant la méthode développée dans le chapitre 3.

Le résultat de la reconstruction est montré sur la figure 5.5. On peut remarquer que le résultat présente un biais non négligeable, de pratiquement  $1\sigma$  sur la moyenne des vitesses radiales. On peut observer la présence d’un flot entrant radialement, provenant de la mauvaise estimation des effets de sélection. En effet, en utilisant l’équation (5.27), les effets de sélection sur les magnitudes absolues n’ont pas été modélisés. En particulier, les magnitudes absolues des galaxies à grande vitesse de rotation seront sous-estimées à cause de l’effet de sélection. Si les magnitudes absolues sont sous-estimées, les modules de distance seront par suite surestimés, et les vitesses particulières sous-estimées. Il en résulte un flot entrant radial comme observé sur la figure 5.5. En ce qui concerne le champ de surdensité reconstruit, on observe des structures factices reconstruites, c’est-à-dire des surdensités à la place de vides initiaux. De plus, les

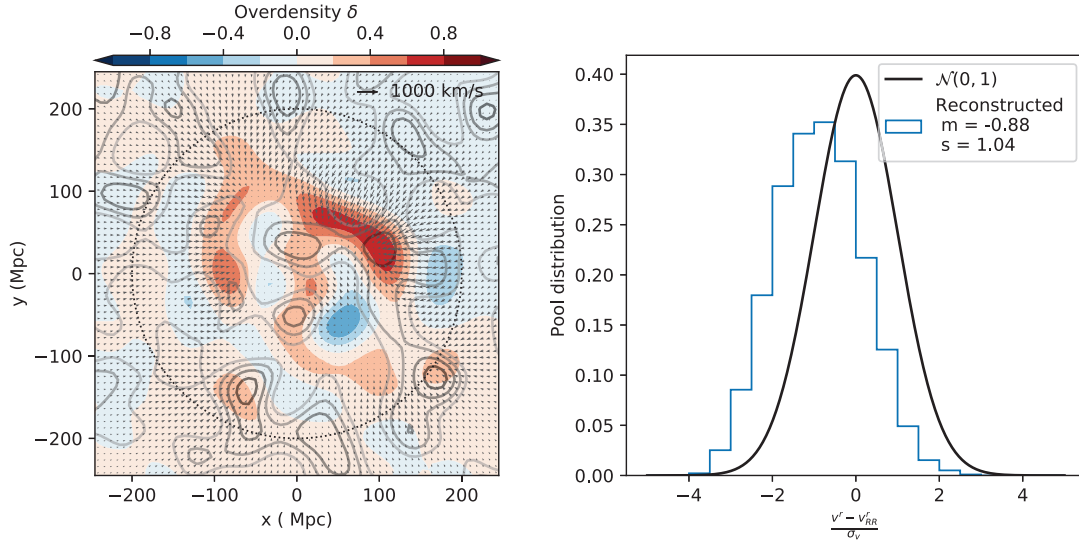


FIGURE 5.5 – Ajustement du champ de vitesse à partir des distances calculées par l'équation (5.27). À gauche : champs de surdensité et de vitesse reconstruits. Les contours gris correspondent au champ de surdensité du champ initial, pour comparaison. À droite : distribution des résidus normalisés des vitesses radiales pour les points de la grille à l'intérieur du cercle en traits pointillés de rayon  $R = 200$  Mpc.

surdensités reconstruites semblent être plus allongées orthoradialement, à cause du flot radial entrant et des biais de sélection.

En plus du biais de la reconstruction, on peut aussi comparer la précision des deux reconstructions. On trace sur la figure 5.6 la dispersion des vitesses radiales reconstruites dans le cas où la reconstruction est faite à partir des données TF et dans le cas où elle est faite à partir des modules de distance déduits de l'équation (5.27). On peut observer que la dispersion est réduite dans le cas où l'on a considéré les magnitudes et vitesses de rotation directement plutôt que les modules de distance. Comme attendu, la non-modélisation des indicateurs de distance provoque à la fois l'apparition de biais systématiques et de la perte d'information dues à la mauvaise estimation des erreurs sur les modules de distance dérivés. Nous soulignons que le biais constaté provient uniquement de la mauvaise estimation des modules de distance en utilisant la TFR telle quelle. En plus de cet effet, il est possible que des erreurs systématiques proviennent des incertitudes sur la calibration de la TFR, que nous avons supposée parfaite dans ce test.

### 5.3.4 Avec la calibration

On s'intéresse maintenant à la possibilité d'utiliser les calibrateurs pour ajuster le point zéro de la TFR et corrélairement la constante de Hubble  $H_0$ . L'intégration des calibrateurs au

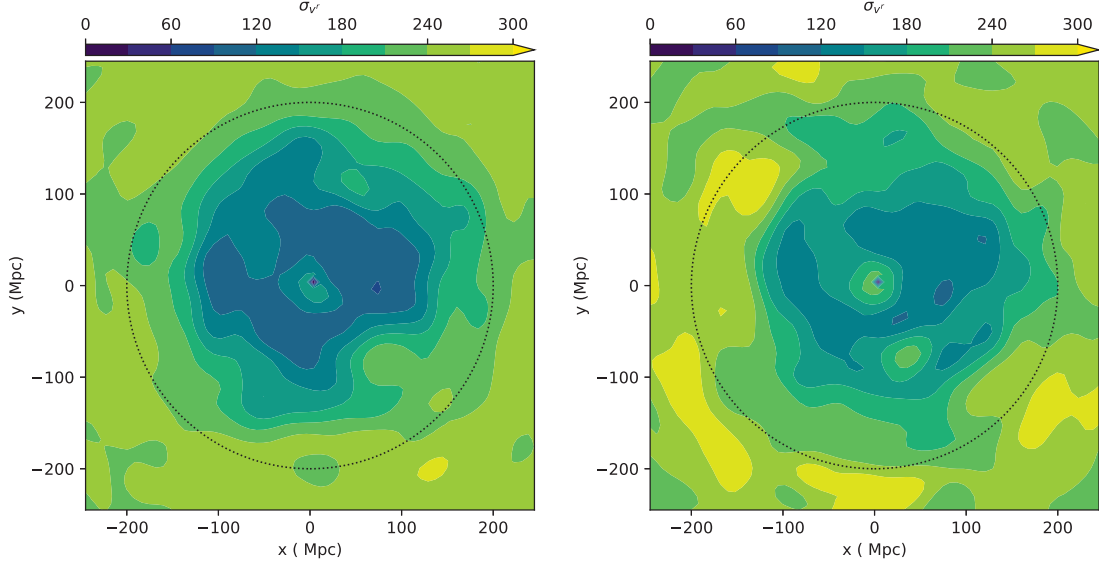


FIGURE 5.6 – Comparaison de la dispersion des vitesses radiales reconstruites entre le cas pour lequel on utilise les données TF (à gauche) et celui pour lequel on utilise les modules de distance pré-calculés (à droite). Les deux panneaux représentent la tranche  $z = 0$  de la reconstruction.

modèle se fait par l'ajout de la fonction de vraisemblance suivante :

$$\mathcal{P}^{calib}(M|\hat{M}) = \mathcal{N}(\hat{M}, \sigma_M^2) \quad (5.28)$$

où on a introduit les données  $\hat{M}$  et  $\sigma_M^2$ , mesure indépendante de la magnitude absolue du calibrateur et son erreur. Cette fonction de vraisemblance remplace l'équation (5.10) pour toutes les galaxies qui sont des calibrateurs. De cette façon la magnitude absolue des calibrateurs sera indépendante du champ de vitesse reconstruit, et les calibrateurs permettront d'ancrer la TFR et ainsi définir une échelle de distance absolue. Avec cette échelle de distance, il est alors possible d'ajuster dans le même temps la constante de Hubble qui elle est définie comme la déviation des vitesses particulières au flot de Hubble.

On prend en compte les 100 calibrateurs des données simulées et on tente de reconstruire le champ de vitesse en laissant libres tous les paramètres du modèle. L'espace des paramètres est  $\theta = \{l\hat{W}, \hat{i}, \mathbf{v}_g, M, \mathbf{b}, \sigma_{int}, \sigma_{NL}, \alpha, H_0\}$ . Le résultat de la reconstruction est montré sur la figure 5.7. On voit que la reconstruction semble sur-estimer globalement les vitesses particulières radiales. La figure 5.8 montre la répartition des pas de Markov dans l'espace des paramètres. On peut constater que le point zéro  $ZP$  est surestimé à  $3\sigma$ , ce qui provoque une surestimation de toutes les vitesses particulières et donc le comportement observé dans la distribution des résidus de la figure 5.8. Un calcul rapide de l'erreur attendue sur le point zéro à partir de 100 calibrateurs est :

$$\Delta ZP = \frac{\sigma_{obs}}{\sqrt{n_{calib}}} \sim 0.04 \text{ mag} \quad (5.29)$$

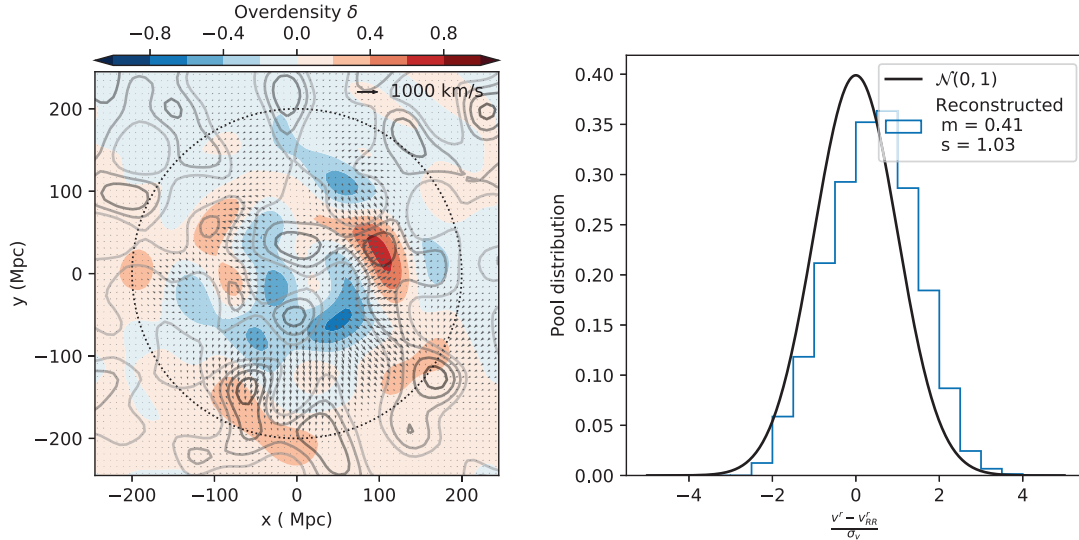


FIGURE 5.7 – Ajustement du champ de vitesse à partir de l’ajustement global du modèle aux données de la TFR. À gauche : champs de surdensité et de vitesse reconstruits. Les contours gris correspondent au champ de surdensité du champ initial, pour comparaison. À droite : distribution des résidus normalisés des vitesses radiales pour les points de la grille à l’intérieur du cercle en traits pointillés de rayon  $R = 200$  Mpc.

ce qui n’est pas suffisant pour expliquer le décalage observé. On constate aussi que la dispersion due aux non-linéarités  $\sigma_{\text{NL}}$  est aussi très surestimée,  $\sigma_{\text{NL}} = 320 \pm 50 \text{ km s}^{-1}$ . Compte tenu de la corrélation apparente entre les paramètres  $\sigma_{\text{NL}}$  et  $ZP$ , il est possible que l’explication des décalages observés soit commune. Enfin, on remarque que la pente  $s$ , la dispersion  $\sigma_{\text{int}}$  et la constante de Hubble  $H_0$  sont correctement retrouvées. On voit qu’avec 100 calibrateurs, l’erreur sur la constante de Hubble est d’environ  $\sim 1 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Toutefois cette erreur dépend aussi de la densité numérique de galaxies observées, de la distance de ces galaxies ainsi que de la dispersion de la TFR.

Surtout, le poids des calibrateurs dans l’ajustement de  $H_0$  doit être analysé. L’ancrage du point zéro de la TFR à partir des calibrateurs n’est pas direct. Certes la magnitude absolue des calibrateurs est connue, mais leur position dans le diagramme de TF est sensible aux erreurs observationnelles sur la mesure de leur vitesse de rotation, à la dispersion intrinsèque de la TFR et aussi à leur couleur. Également, les calibrateurs ancrent le champ de vitesse à leurs positions respectives, et donc influent aussi sur les vitesses particulières des autres galaxies. Une analyse complète de ces effets dépasse cependant le cadre de cette thèse, et fait partie des perspectives possibles d’amélioration du modèle.

Notons également que nous avons considéré dans ce test cent galaxies dont on connaît la magnitude absolue. En pratique, l’ancrage des indicateurs de distance se fait à partir de deux types de calibrateurs, primaires et secondaires. Les indicateurs primaires sont les indicateurs qui définissent de façon absolue la distance d’objets extragalactiques : parallaxe, binaires à

eclipse et megamaser. Les deux calibrateurs usuels sont le Petit Nuage de Magellan et la galaxie NGC4258. Les indicateurs secondaires sont des indicateurs de distance qui doivent être calibrés sur les indicateurs primaires. Les calibrateurs secondaires sont donc ceux dont on connaît la magnitude absolue à partir d'indicateurs secondaires précis, en pratique les Céphéïdes et les magnitudes de la branche des géantes rouges (TRGB). Notre test ici suppose l'existence de 100 calibrateurs, qui sont donc des calibrateurs secondaires, et ils ont été aléatoirement choisis parmi les 3000 galaxies simulées. On voit donc que nous nous sommes placés dans un cas idéaliste : en pratique tous les calibrateurs se situe à des distances faibles pour pouvoir être mesurés à partir de la méthode des Céphéïdes ou TRGB. Dans le catalogue CF3 par exemple, environ 400 distances proviennent de mesures de Céphéïdes ou TRGB, et 47 ont en commun des mesures de TFR à partir des échantillons Spitzer et SFI++. Leur distance moyenne est environ  $d \sim 9$  Mpc. On est donc bien loin d'une calibration homogène en distance et avec autant de calibrateurs. Pour toutes ces raisons, l'incorporation de la calibration absolue de la TFR au sein du modèle de vitesse nécessite des développements supplémentaires.

En définitive, cette section souligne la nécessité de construire des catalogues simulés davantage en accord avec les données observationnelles afin d'étudier l'influence des calibrateurs sur l'ajustement de la TFR. On a néanmoins montré la nécessité de propager les erreurs observationnelles et effets de sélection depuis les observations primaires plutôt que d'utiliser directement les modules de distance dérivés.

## 5.4 SFI++

Dans cette section, nous appliquons la méthode à quelques données TFR de CF3. Plus exactement, on utilise l'échantillon SFI++ (Masters et al., 2006; Springob et al., 2009) de *CosmicFlows*. SFI++ est un catalogue d'environ 5000 mesures de vitesses de rotation par spectrométrie. On utilise les magnitudes et largeurs HI supposées dans Tully et al. (2013). Les magnitudes ont été préalablement corrigées de l'extinction interne (Tully et al., 1998) et galactique (Schlegel et al., 1998).

La reconstruction du champ de vitesse sous-jacent est réalisée par les trois méthodes différentes présentées dans cette thèse :

1. filtre de Wiener appliqué aux modules de distance tels que donnés par le catalogue CF3,
2. modèle du chapitre 3 appliqué aux modules de distance tels que donnés par le catalogue CF3,
3. modèle de ce chapitre appliqué aux vitesses de rotation et magnitudes du catalogue directement.

La figure 5.9 montre les trois champs reconstruits ainsi que la déviation standard sur les vitesses radiales reconstruites. De la gauche vers la droite, on peut observer une amélioration visuelle globale des reconstructions. La modélisation des modules de distances plutôt que des vitesses particulières (grâce au modèle du chapitre 3) puis des observations de la TFR plutôt que des modules de distance (grâce au modèle de ce chapitre), permet l'optimisation de l'information contenue dans les données. On observe notamment une meilleure



résolution du superamas shapley, et l'apparition, bien que faible, de deux surdensités compatibles avec la distribution des galaxies aux coordonnées  $(SGX, SGY) = (120, -80)\text{Mpc}/h_{75}$  et  $(SGX, SGY) = (-50, -110)\text{Mpc}/h_{75}$ . Les superamas semblent généralement mieux résolus. On note aussi une légère amélioration de la dispersion sur le champ des vitesses radiales, comme attendu d'après les tests effectués dans la section 5.3.

La relation de Tully-Fisher ajustée est montrée sur la figure 5.10. On trouve les paramètres  $s = -7.74 \pm 0.15$  et  $\sigma = 0.324 \pm 0.013$  pour la pente et la dispersion intrinsèque. La valeur de la pente est comparable à celle donnée dans Masters et al. (2006)  $s = -7.85 \pm 0.1$  mais la dispersion intrinsèque est légèrement en dessous de la dispersion moyenne qu'ils trouvent à  $\sigma \sim 0.35$ . On note cependant que leur modélisation inclue une dépendance de la dispersion intrinsèque avec la vitesse de rotation telle que  $\sigma = 0.35 - 0.37(\log W - 2.5)$ . Cette modélisation de la TFR n'inclut ni le modèle de couleur développé (puisque qu'aucune couleur n'est accessible dans le catalogue SFI++) ni aucun calibrateur. La constante de Hubble a été fixée de telle sorte que le point zéro soit lui aussi bien défini<sup>1</sup>.

Finalement, la comparaison entre le panneau de gauche et le panneau de droite montre l'amélioration rendue possible par les modèles de cette thèse sur quelques 5000 points de données. Cette analyse est une première étape et d'autres restent à effectuer : définition de calibrateurs, modélisation des fonctions de sélection spécifiques aux sondages observationnels, intercalibration des différentes sources de HI, etc. L'objectif ultime est d'intégrer chacune des méthodologies de *CosmicFlows* dans une analyse commune. En commençant par les données TFR (Sorice et al., 2014; Tully and Courtois, 2012); puis en étendant l'analyse aux données du Plan Fondamental. En outre, nous soulignons ci-dessous quelques limites du modèle qui doivent être prises en compte pour une analyse plus complète de la TFR comme indicateur de distance extragalactique.

## 5.5 Discussion

Nous avons montré dans les sections précédentes l'importance de propager les erreurs depuis les mesures brutes pour estimer le champ de vitesse (et les paramètres cosmologiques) à partir de vitesses particulières tirées de la TFR. Le but de cette section est de souligner qu'il reste néanmoins des effets à modéliser pour que la TFR soit compétitive en terme de sonde cosmologique. Afin de limiter les erreurs systématiques sur la TFR ainsi que de diminuer sa dispersion intrinsèque, il est en particulier nécessaire d'identifier les paramètres dont dépendent le point zéro et la pente. Il est également nécessaire d'identifier les biais d'observation ainsi que les effets de sélection des observations. Les erreurs de la TFR proviennent donc de deux phénomènes bien distincts : les effets intrinsèques et les effets observationnels.

**Effets intrinsèques** Les effets intrinsèques correspondent aux décalages de la relation de Tully-Fisher par rapport à la relation théorique provenant du théorème du Viriel. Ces décalages ont deux effets : le premier est d'augmenter la dispersion intrinsèque de la TFR, et le deuxième de créer des erreurs systématiques sur la distance des galaxies en fonction de leurs propriétés

---

1. N'ayant imposé aucun calibrateur, les valeurs absolues supposées de la constante de Hubble et du point zéro ne sont pas importantes ici, du moment qu'elles sont compatibles.



physiques<sup>2</sup>. L'écart au théorème du Viriel correspond aux écarts à l'équilibre dynamique d'un disque en rotation : la galaxie est-elle purement un disque en rotation ? Et est-elle virialisée ?

Plusieurs traceurs de la virialisation existent. Principalement, on utilise la couleur, la morphologie ou la formation stellaire. La morphologie est typiquement un traceur du caractère virialisé ou non de la galaxie ainsi que de sa forme : la TFR n'aura pas les mêmes propriétés selon le type morphologique étudié (Giraud, 1986; Shen et al., 2009; Lapi et al., 2018). La couleur, puisque corrélée avec la morphologie, en est un autre. En général, les études de la TFR prennent toujours en compte une correction de couleur (Masters et al., 2014; Sorce et al., 2014; Desmond, 2017). Enfin, il est important de noter que la relation de Tully-Fisher dépend aussi de l'environnement des galaxies étudiées (Giuricin et al., 1986; Adami et al., 1999) puisque l'interaction avec d'autres galaxies supprime l'hypothèse d'équilibre gravitationnel isolé.

Un autre effet systématique dans l'étude de la TFR est de supposer un rapport masse-luminosité constant. En pratique, ce rapport masse-luminosité n'est pas constant (McGaugh and Schombert, 2014) et il faut de plus considérer que la masse lumineuse ne représente pas la masse cinématique de la galaxie. En particulier, la masse de gaz joue un rôle non-négligeable dans la dynamique des galaxies de masse inférieure à  $\sim 10^9 M_\odot$ . En incluant la masse de l'hydrogène neutre (calculée à partir du flux HI), il est possible de construire la relation de Tully-Fisher baryonique (BTFR) (Zaritsky et al., 2014; Desmond, 2017). Cependant, les flux HI sont en général mesurés à 15% près dans CF3, et par conséquent il n'est pas clair que les inclure dans l'analyse présentée ici l'améliorerait.

De façon plus générale, il est nécessaire d'identifier les paramètres qui sont susceptibles de réduire la dispersion intrinsèque de la TFR. En effet, la détermination expérimentale de la TFR montre que sa dispersion observée est d'environ  $\sim 0.3 - 0.4$  en magnitude alors qu'une analyse précise de cette dispersion (Lelli et al., 2015) montre que la dispersion intrinsèque, c'est-à-dire non liée aux erreurs observationnelles, se trouverait plutôt autour de 0.2 en magnitude. À noter aussi que la dispersion intrinsèque est parfois (Giovanelli, 2013) considérée comme dépendante de la vitesse de rotation bien que les erreurs observationnelles revêtent la même dépendance, ce qui rend difficile l'analyse de cette propriété.

**Effets observationnels** Notre modélisation nécessite également l'intégration d'effets observationnels qui sont résumés ci-après. La détermination expérimentale de la relation de Tully-Fisher se heurte à une difficulté majeure qu'est la détermination de la vitesse de rotation de la galaxie à partir de sa raie HI. Cette mesure se fait en deux étapes que sont la détermination de la vitesse de rotation projetée, notée  $W$ , et l'inclinaison  $i$  de la galaxie permettant de déprojeter cette vitesse de rotation.

À partir d'un spectre HI, la vitesse de rotation apparente est mesurée à partir de la distance entre les deux raies décalées par effet Doppler. Cette distance est théoriquement proportionnelle à la vitesse de rotation de la galaxie. La constante de proportionnalité est absorbée dans le paramètre de la pente de la TFR. Cela souligne néanmoins qu'il est nécessaire de mesurer toutes les largeurs de façon consistante et systématique pour éviter tout biais sur la calibration du point zéro. Également, la définition de  $W$  peut souffrir d'erreurs systématiques qui doivent

---

2. Pour fixer les idées, un exemple de ce qu'on appelle effet intrinsèque est le biais d'âge dans les SNeIa (Rigault et al., 2018)

être incluses au modèle. Par exemple, la cinématique interne d'une galaxie, qui est sans doute corrélée à la vitesse de rotation, va augmenter artificiellement la largeur HI mesurée, et ainsi biaiser positivement la vitesse de rotation inférée<sup>3</sup>. Un autre effet est celui de la grande corrélation entre l'inclinaison et la largeur HI mesurée (puisqu'elles sont reliées par  $V_{\text{rot}} = W \sin i$ ). Cette corrélation pourrait être prise en compte si la mesure de l'inclinaison et de la largeur HI d'une galaxie se faisait conjointement.

Cette corrélation, si elle influe sur la mesure des vitesses de rotation, joue surtout sur les effets de sélection d'un sondage de galaxies spirales par spectroscopie HI. En effet, si la sélection se fait à partir du flux HI uniquement, elle sera de fait corrélée avec l'inclinaison des galaxies, effet qui n'est pour l'instant pas modélisé. Si la sélection se fait par pointage, il faut prendre en compte la couleur de la galaxie, sa morphologie et son inclinaison, puisque ce sont les principaux critères de sélection de galaxie pour la TFR. On note à ce propos que l'ajustement d'une droite à des points dont les effets de sélection jouent à la fois sur les variables dépendantes (ici,  $M$ ) et indépendantes (ici,  $\log W / \sin i$ ) est un exercice difficile, comme il est expliqué dans Kelly (2007).

## 5.6 Conclusion

Ce chapitre a été motivé par la nécessité d'aller plus loin dans la modélisation des observations pour analyser le champ de vitesse de l'Univers Local. Il s'est appuyé sur un retour à la modélisation des observations primaires de la relation de Tully-Fisher, magnitudes, largeurs HI et inclinaisons. Le modèle a été décrit dans la section 5.2 et testé dans la section 5.3. Deux conclusions principales sont à tirer de ces tests. La première est la nette amélioration induite par l'utilisation d'un tel modèle. En comparant avec une reconstruction utilisant les modules de distance dérivés directement de la TFR, on a montré qu'inclure la TFR dans le modèle d'une part réduit les erreurs systématiques et d'autre part améliore la précision de la reconstruction. La deuxième conclusion est la difficulté d'inclure tous les effets de calibration au sein du modèle. La corrélation forte entre le monopole du champ de vitesse, la constante de Hubble et le point zéro de la TFR nécessite une étude complémentaire sur la modélisation du rôle des calibrateurs.

On a ensuite présenté une analyse préliminaire du catalogue SFI++, sous-échantillon des catalogues CF3. Les premiers résultats montrent qu'on retrouve la pente de l'article original Masters et al. (2006) et qu'on a pu réduire sa dispersion intrinsèque. Le champ reconstruit simultanément montre une forte amélioration depuis la méthode du filtre de Wiener. Cette étude préliminaire nous incite à améliorer le modèle et à l'appliquer aux autres données de HI disponibles. Améliorations qui sont discutées dans la section 5.5.

Le modèle présenté ici est le résultat de la confrontation d'un modèle théorique, le modèle  $\Lambda$ CDM, aux observations. L'utilisation de la technique WF/CR a montré les limites induites par l'utilisation des vitesses particulières en tant que variables indépendantes. Le chapitre 4 a quant à lui montré les limites induites par l'utilisation des modules de distance en tant que variables indépendantes. Et le modèle présenté ici conclut sur la nécessité d'une modélisation

---

3. À ce propos, il est intéressant de noter dans la littérature récente un modèle analytique de flux intégré HI par Stewart et al. (2014).

plus avancée du rôle des calibrateurs. Cette deconstruction, nécessaire à la compréhension du lien entre les données et le modèle  $\Lambda$ CDM, a inspiré des projets d'étude, qui seront décrits dans le chapitre 6. Finalement, ce chapitre nous permet de souligner l'importance de la modélisation statistique *end-to-end* dans les études cosmologiques.

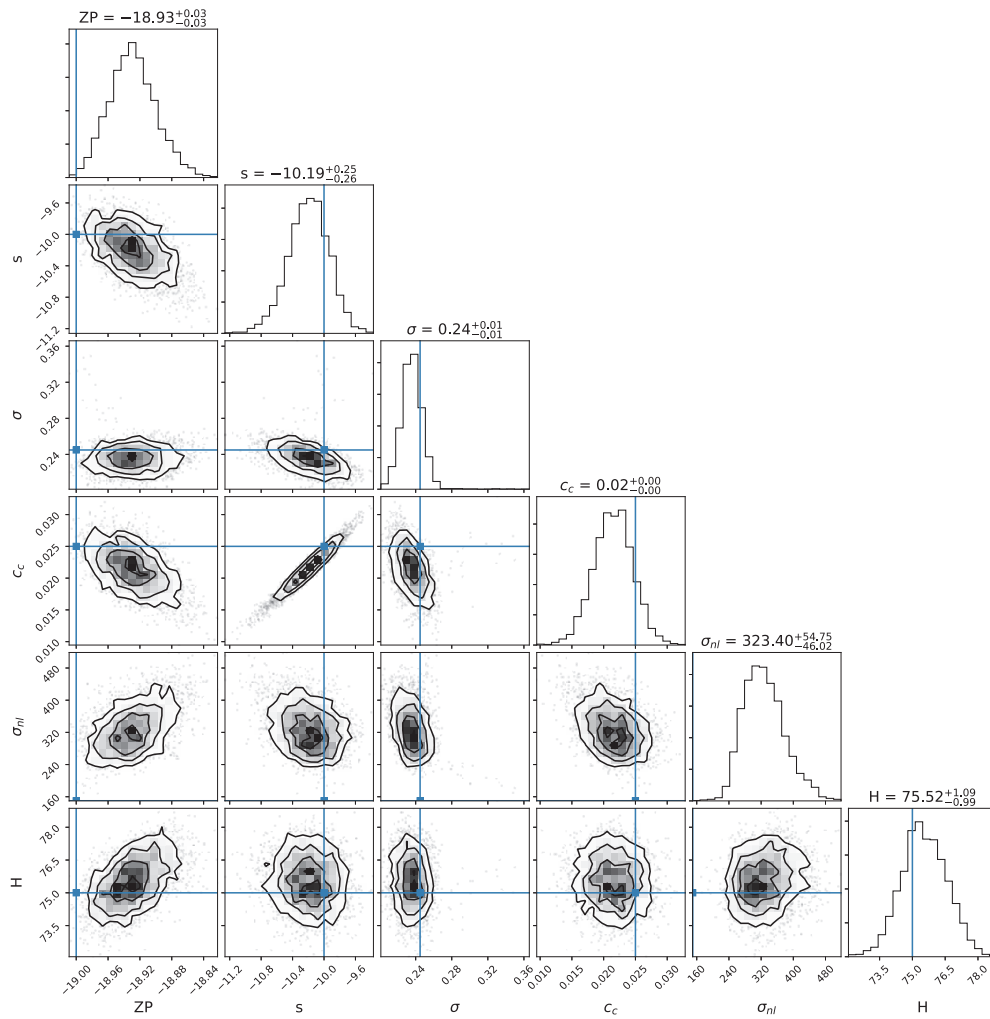


FIGURE 5.8 – Figure en coin des paramètres  $ZP$ ,  $s$ ,  $\sigma_{\text{int}}$ ,  $cc$ ,  $H_0$ ,  $\sigma_{\text{NL}}$  pour l’ajustement de la TFR avec le champ de vitesse et en utilisant 100 calibrateurs. Chaque point correspond à un pas de Markov, et les traits bleus représentent les valeurs utilisées dans la simulation.

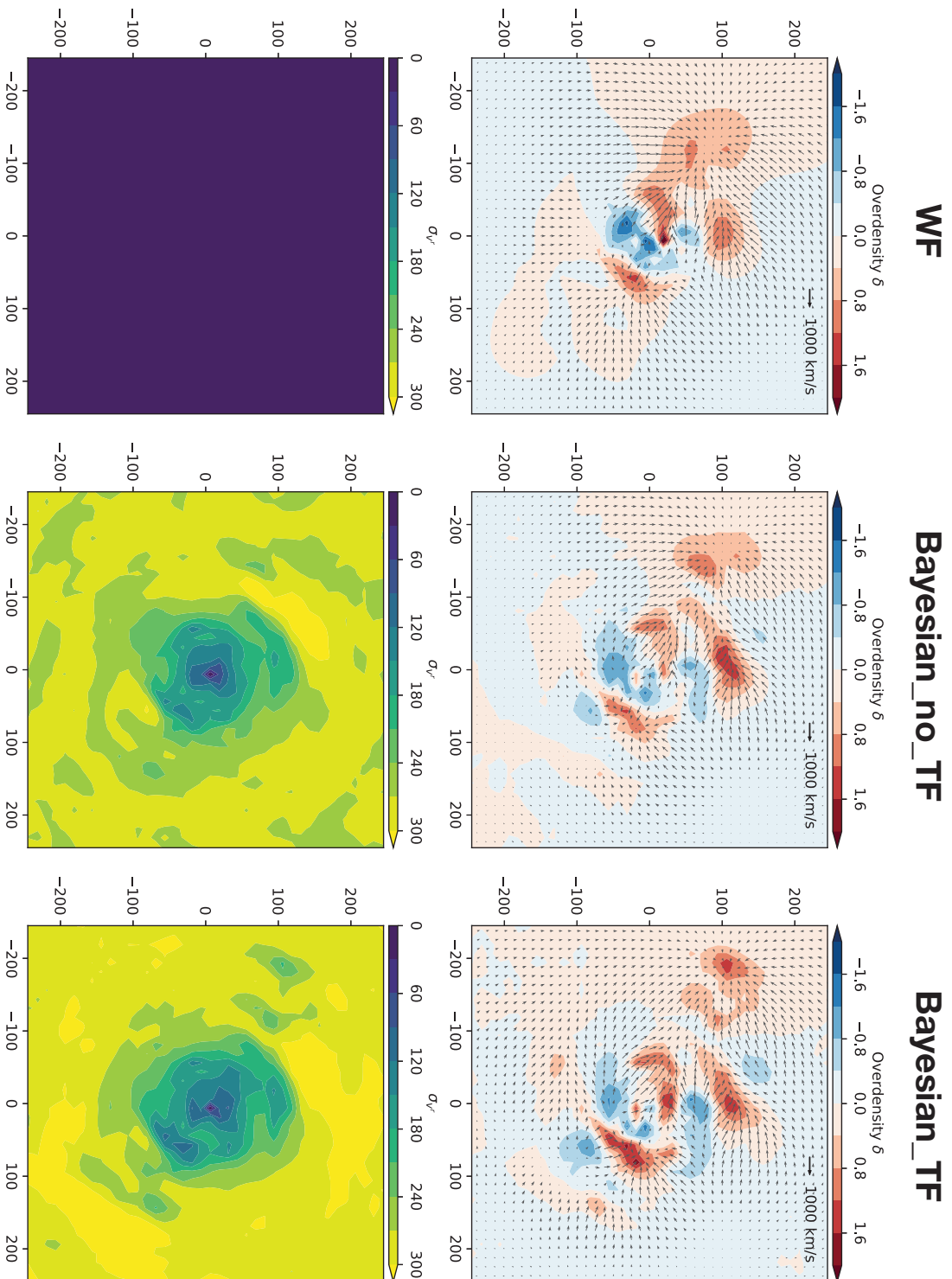


FIGURE 5.9 – Reconstruction du champ de surdensité à partir des données SFI++ en utilisant les trois méthodes mentionnées dans cette thèse. À gauche : filtre de Wiener, chapitre 2. Au milieu : échantillonnage de Gibbs, chapitre 3. À droite : échantillonnage de Gibbs et ajustement de la TFR, chapitre 5. Ligne du haut : champs de surdensité et de vitesse dans la tranche  $SGZ = 0$ . Ligne du bas : Dispersion des vitesses radiales reconstruites.

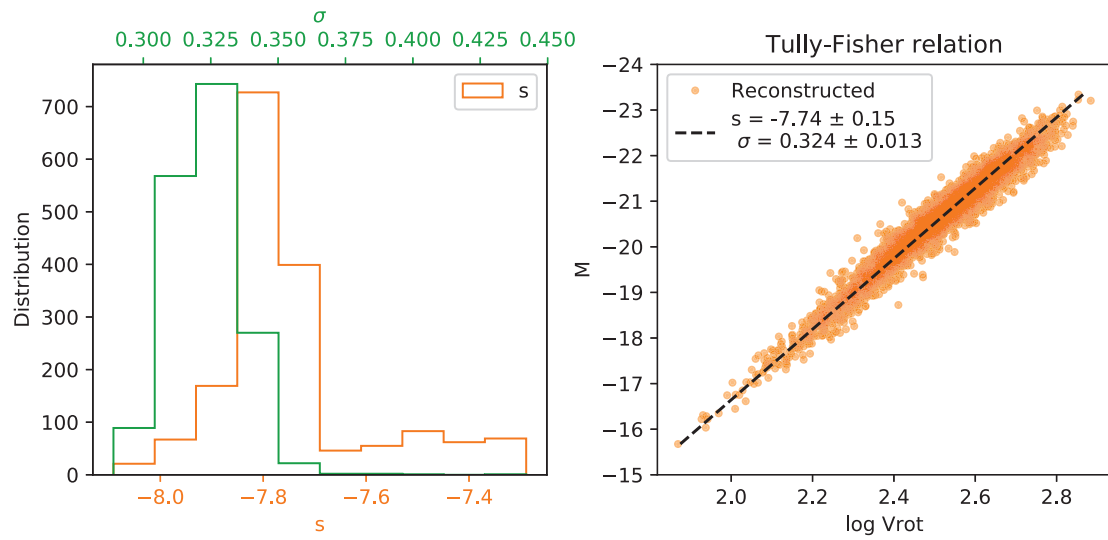


FIGURE 5.10 – Reconstruction du champ de vitesse de SFI++. À gauche : distribution des paramètres  $s$  et  $\sigma$ . À droite : TFR reconstruite.



# Chapitre 6

## Perspectives

La question sous-jacente à cette thèse est la possibilité de rendre les analyses de vitesses particulières compétitives en terme de sonde cosmologique. Dans ce chapitre, nous proposons des solutions à explorer en vue de cet objectif. La section 6.1 propose des solutions pour améliorer le modèle de non-linéarités. La section 6.2 souligne l'intérêt que revête une amélioration du modèle de la TFR. Enfin, la section 6.3 présente des perspectives pour l'analyse des vitesses particulières à partir des Supernovae de Type Ia. Ce chapitre est l'occasion de décrire les évolutions possibles du travail présenté dans cette thèse. Ces évolutions peuvent être des améliorations du modèle ou des applications de celui-ci.

### 6.1 Vers un modèle non linéaire

Cette thèse s'est concentrée sur la modélisation des données observationnelles pour l'étude des vitesses particulières. Toutefois il reste une approximation qui a été peu discutée jusqu'ici et qui est pourtant cruciale : l'approximation linéaire. Dans la section 1.1.5 nous nous sommes limités à la résolution des équations de Poisson et d'Euler à l'ordre le plus bas. Le problème d'inclure les ordres supérieurs dans notre modèle est une question de difficulté algorithmique. En effet, le spectre de puissance défini dans la section 1.3.1 en théorie linéaire a l'avantage d'être diagonal dans l'espace de Fourier, puisque les modes sont tous indépendants les uns des autres, et isotropes. Ce sont ces deux propriétés qui rendent par exemple possible la génération d'un champ aléatoire gaussien (section 3.2) et par extension l'algorithme de Hoffman-Ribak (section 2.1). De plus, le lien entre le champ de vitesse et de surdensité défini par l'équation (1.33) n'est valable que dans le régime linéaire.

**CLUES** L'évolution non-linéaire ne peut en fait qu'être résolue numériquement à travers les simulations contraintes. Le consortium CLUES (Yepes et al., 2009; Gottloeber et al., 2010; Courtois and Tully, 2012; Sorce et al., 2016) tente à cet effet de produire des simulations numériques contraintes par les observations *CosmicFlows* de la façon suivante :

- reconstruction du champ de vitesse par la technique WF/CR;
- construction de conditions initiales à partir du champ ainsi obtenu, via l'approximation de Zeldovich inverse (Doumler, 2012; Doumler et al., 2013),



— simulations contraintes à partir de ces conditions initiales.

Si ces simulations permettent de simuler bon nombre d’observations du groupe Local (Di Cintio et al., 2013; Carlesi et al., 2016), elles restent difficiles à lier aux observations puisque leurs contraintes sont statistiquement intraquables à cause de la procédure décrite ci-dessus. Toutefois, avec la puissance de calcul dont nous disposons aujourd’hui, il est possible d’optimiser l’utilisation des simulations cosmologiques pour avoir accès aux champs non linéaires de l’Univers Local.

À l’instar de l’algorithme BORG (Jasche and Wandelt, 2013), on peut imaginer le processus suivant :

- le modèle linéaire étant valable à grand décalage vers le rouge  $z_{\text{init}}$ , la technique des réalisations contraintes y est parfaitement adaptée. On peut donc construire un champ initial contraint à  $z_{\text{init}}$  *via* l’algorithme de Hoffman-Ribak,
- on fait ensuite évoluer ces conditions initiales à décalage vers le rouge nul  $z = 0$  grâce à une simulation numérique,
- puis on compare le champ de vitesse résultant aux observations en évaluant la fonction de vraisemblance définie par l’équation (3.12).
- Enfin, on génère de nouvelles conditions initiales grâce aux nouvelles distances et vitesses particulières ainsi reconstruites. CLUES utilise jusque là l’approximation de Zel-dovich mais on peut imaginer une approximation plus précise comme par exemple la théorie lagrangienne des perturbations du second ordre (2LPT) (Croce et al., 2006; Scoccimarro et al., 2012)

Ce processus répété suffisamment permet d’échantillonner le champ de vitesse linéaire à  $z_{\text{init}}$  plutôt qu’à  $z = 0$  où l’approximation linéaire n’est plus valable. Bien sûr, le modèle statistique reste à écrire complètement, et plusieurs difficultés techniques subsistent. En particulier le coût de calcul du nombre de simulations nécessaires peut se révéler très élevé pour un tel algorithme. Un tel modèle, s’il est statistiquement définissable, permettrait de reconstruire le champ de vitesse non-linéaire à  $z = 0$ .

## 6.2 Les vitesses particulières comme sondes cosmologiques

Les vitesses particulières sont un outil idéal pour les études à champ proche. Contrairement aux positions des galaxies dans l’espace des décalages vers le rouge, elles sont non biaisées vis-à-vis de la matière noire, et corréler sur de plus grandes distances. Ces deux propriétés confèrent à leur étude une supériorité de principe devant les analyses de décalages vers le rouge ou de cisaillement gravitationnel. Koda et al. (2014) souligne l’avantage des vitesses particulières comme sonde cosmologique, et prévoit que l’incertitude de la mesure du taux de croissance  $f\sigma_8$  pourra par exemple atteindre une précision de 5% dans le futur. Des travaux ont déjà exploité ce potentiel, comme Howlett et al. (2017) qui propose l’ajustement du spectre de puissance des vitesses particulières pour estimer le taux de croissance  $f\sigma_8$  à  $\sim 16\%$ .

Toutefois, ces prédictions doivent être tempérées par les erreurs systématiques des analyses de vitesses particulières. Cette thèse souligne l’importance de ces systématiques, d’abord parce que les effets de sélection sont amplifiés par le fait que les vitesses particulières sont dérivées directement des distances; et ensuite parce que la mesure de distance est une mesure très

bruitée qui nécessite une modélisation Bayésienne approfondie. On souligne à ce propos que l’ajustement de  $f\sigma_8$  est possible dans le modèle présenté dans le chapitre 3, puisque le spectre de puissance peut être recalculé à chaque étape itérative. L’implémentation détaillée de cette évolution peut être trouvée dans Lavaux (2016).

### La relation de Tully-Fisher

Dans le chapitre 5, nous avons initié le travail d’étude des erreurs systématiques des données de vitesses particulières, en nous focalisant sur la relation de Tully-Fisher, principal indicateur de distance extragalactique (en nombre). Nous avons souligné dans la section 5.5 tous les paramètres qui restent à inclure à notre analyse afin qu’elle soit précise. En particulier, la dispersion observée de la relation n’est pas entièrement expliquée à partir des erreurs observationnelles. La prochaine étape est donc d’intégrer à la relation de Tully-Fisher des paramètres observationnels qui permettent de réduire la dispersion de la relation de Tully-Fisher en standardisant la relation selon ces paramètres, de façon semblable à ce qui est fait avec les données de supernovae de type Ia. La dispersion intrinsèque est en effet prédite à environ 0.25 en magnitude par les simulations numériques de galaxies (Dutton et al., 2017), sans que l’origine de cette dispersion soit expliquée. Nous avons vu que l’un de ces paramètres est la couleur, et parmi les candidats pour les autres on peut citer le type morphologique, la métallicité ou l’âge.

L’étude systématique de la relation de Tully-Fisher est également motivée par l’arrivée massive de données dans un futur proche.

**ALFALFA et *CosmicFlows-4*.** On peut d’abord citer l’analyse en cours des données ALFALFA (Giovanelli et al., 2005; Haynes et al., 2011, 2018) par la collaboration *CosmicFlows* (thèse en cours à l’institut d’astronomie de l’Université d’Hawaï). Ce catalogue contiendra environ 5000 sources HI avec un signal sur bruit suffisant et la photométrie associée, provenant principalement des observations de SDSS (Stoughton et al., 2002; Abolfathi et al., 2018) ainsi que WISE (Bilicki et al., 2016) puisque leurs empreintes sur le ciel coïncident en grande partie.

**WALLABY** est une campagne d’observation dans l’hémisphère sud qui prévoit d’observer l’émission HI d’environ 800,000 objets (Duffy et al., 2012). Après sélection sur l’inclinaison et sur la vitesse de rotation, Koda et al. (2014) estiment qu’environ 32,000 de ces objets seront intégrables à la relation de Tully-Fisher.

Il est donc possible que d’ici quelques années environ 30,000 données TF soient disponibles à faible décalage vers le rouge. Si ce nombre est évidemment un grand gain dans l’étude des vitesses particulières, il faut noter le défi qu’il nous propose. Les algorithmes présentés dans toute cette thèse sont numériquement coûteux et leur complexité évolue approximativement en  $O(N^3)$  (due notamment à l’inversion de la matrice de corrélation) où  $N$  est le nombre de points de donnée. Pour donner un ordre de grandeur, l’étude faite au chapitre 4 ( $\sim 11,000$  points de données) a nécessité environ 30,000 heures de calcul. La question de la mise à échelle d’un tel modèle physique reste posée. Des alternatives existent, comme par exemple l’utilisation de technique de groupement des données sur une grille régulière, technique proposée par Howlett et al. (2017), mais qui génère un bruit de grenaille supplémentaire. On peut aussi penser à des

techniques d'écrantage et négliger les corrélations à partir d'une certaine distance, cependant les vitesses particulières corréleront sur des grandes distances et la perte d'information pourrait être trop importante. Dans tous les cas, l'étude de la mise à l'échelle de ce modèle est une étape importante de son développement pour les années à venir.

### Plan fondamental

Bien que le modèle développé dans cette thèse s'applique exclusivement à la relation de Tully-Fisher, il est tout à fait possible d'intégrer des modèles d'autres indicateurs de distance, à commencer par la technique du plan fondamental. La nécessité d'inclure le Plan Fondamental comme partie intégrante de l'analyse est motivée d'une part par la présence dominante des données 6dF dans le catalogue *CosmicFlows-3*, et d'autre part par l'arrivée du sondage observationnel TAIPAN (da Cunha et al., 2017). **TAIPAN** est un sondage observationnel dans l'hémisphère sud qui prévoit l'acquisition du spectre de 2 millions de galaxies en vue de l'étude des paramètres cosmologiques  $H_0$  et  $f\sigma_8$  à partir de la technique du Plan Fondamental. Parmi ces 2 millions, environ 40,000 seront des données de Plan Fondamental.

### 6.3 Les Supernovae de type Ia et les vitesses particulières

L'autre continuation naturelle de cette étude sur les vitesses particulières concerne bien sûr les Supernovae de type Ia (SNeIa). Ces données, dont les erreurs systématiques sont régulièrement évaluées (Betoule et al., 2014; Rigault et al., 2018), sont idéales pour l'étude des vitesses particulières. Elles sont peu bruitées ( $0.1^2$  en magnitude) et les effets de sélection seront mineurs dans les futures observations qui seront limitées par le décalage vers le rouge et non la brillance. Ces deux points limitent les erreurs systématiques décrites dans le chapitre 2, biais de Malmquist et effets de sélection. Le seul problème des supernovae de Type Ia est leur nombre réduit. Toutefois, pour ajuster le spectre de puissance à une échelle linéaire, il n'est pas nécessaire que la densité de traceurs soit élevée. Par exemple, 4000 SNeIa dans un volume de 1 Gpc (voir ZTF ci-dessous), permet de reconstruire les modes au dessus de  $k = 0.1 \text{ Mpc}^{-1}$ .

#### Diagramme de Hubble à faible décalage vers le rouge

Une possibilité d'utilisation des vitesses particulières dans le cadre de l'étude des SNeIa est la réduction de la dispersion du diagramme de Hubble à petit décalage vers le rouge. En effet, du fait des vitesses particulières radiales, les résidus au diagramme de Hubble augmentent quand on étudie les SNeIa proches. Il est possible d'utiliser les cartes de vitesse de CF3 pour essayer de corriger les décalages observés. Une analyse préliminaire a été faite au cours de cette thèse, avec la participation de M. Rigault (LPC) et K. Asvesta (IPNL). À partir des données de SNFactory (Wood-Vasey et al., 2004) (Rigault, comm. privée), on peut localiser une SNIa à partir de son décalage vers le rouge. À la manière de Léget et al. (2018), il est possible de recalculer le décalage vers le rouge cosmologique des SNeIa et ainsi retracer le diagramme de Hubble corrigé. La figure 6.1 montre une analyse *très préliminaire* des SNeIa de SNFactory à partir des vitesses calculées dans le chapitre 4. Notons que nous n'observons pas pour le moment de réduction systématique de la dispersion intrinsèque au diagramme de Hubble.

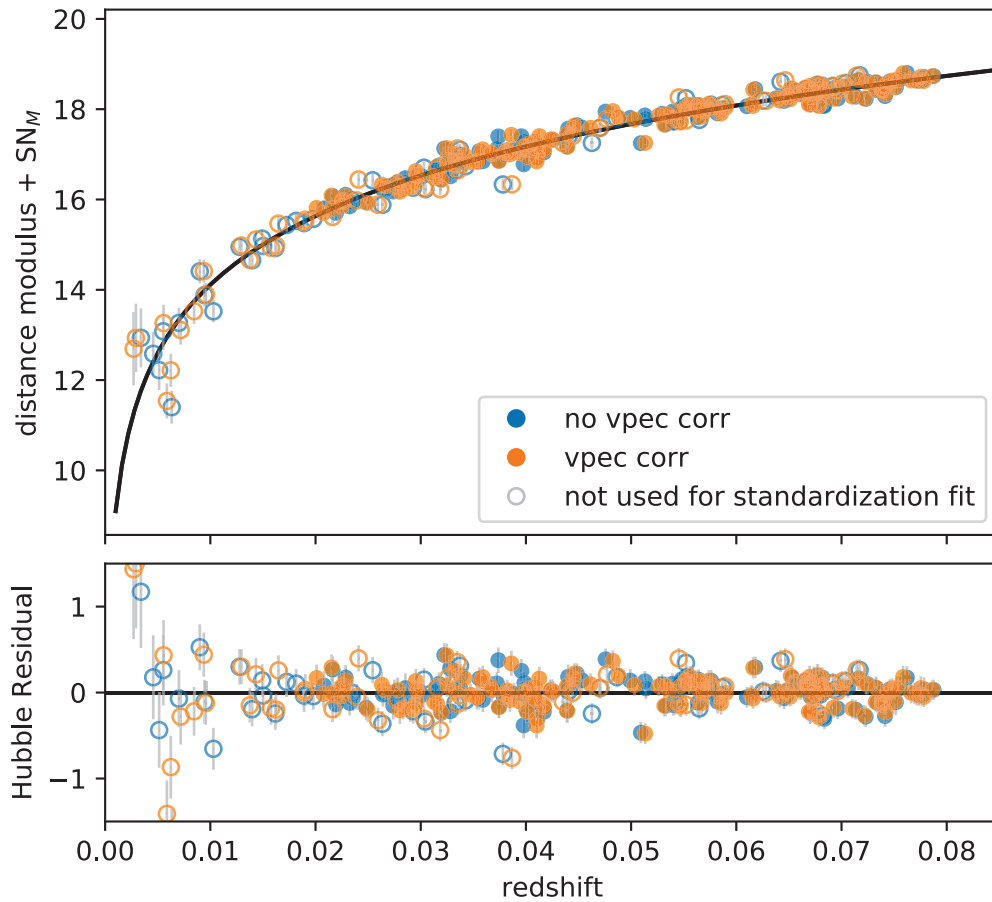


FIGURE 6.1 – Diagramme de Hubble et ses résidus à partir des SNeIa de la collaboration SNFactory.

Ces analyses pourraient par exemple permettre d'inclure plus de SNeIa dans l'analyse des biais systématiques à partir des résidus au diagramme de Hubble, ou encore d'ancrer plus facilement le diagramme avec les galaxies plus proches. À terme bien sûr, l'idéal est d'inclure complètement les SNeIa dans l'analyse du champ de vitesse afin de déduire leurs décalages vers le rouge cosmologiques de façon cohérente avec les autres galaxies.

### Sondages observationnels à venir

L'énorme potentiel des SNeIa réside surtout dans les grands sondages systématiques à venir. Parmi eux, citons ZTF et LSST.

**ZTF**<sup>1</sup> (Bellm, 2014) est un sondage de transients qui prévoit d'observer jusqu'à 1000 SNeIa (typées, donc) par an jusqu'en 2022, soit environ 4000 SNeIa. Toutes ces SNeIa seront observées à décalage vers le rouge  $z < 0.1$ , ce qui en fait des données parfaites pour étudier le champ de vitesse de l'Univers Local. En terme de vitesses particulières, les analyses présentées dans cette thèse sont parfaitement adaptables à ces nouvelles données, et il est possible d'envisager plusieurs études comme :

- la reconstruction du champ de vitesse et du spectre de puissance à partir des SNeIa ;
- l'estimation de la constante de Hubble locale et du taux de croissance ;
- l'intercalibration de la TFR et des SNeIa.

Ces analyses font partie des priorités pour le développement du modèle présenté dans cette thèse.

**LSST** observera 20,000 SNe par an entre 2022 et 2032, soit environ 200,000 SNe potentiellement typées. Ces SNeIa iront jusqu'à un décalage vers le rouge d'environ  $z = 0.3$  dépassant le cadre de l'Univers Local pour arriver à des échelles intermédiaires. Un tel nombre nécessitera bien évidemment de nouveaux outils d'analyse, mais les analyses mentionnées ci-dessus restent d'actualité. Notons qu'un tel catalogue permettra en plus la possibilité de faire des analyses directionnelles et dans différentes régions, afin de tester l'isotropie et l'homogénéité de l'Univers à grande échelle.

---

1. <https://www.ptf.caltech.edu/page/ztf>

# Conclusion

La méthode à l'origine de cette thèse, la technique WF/CR, était supposée être un outil pour reconstruire et analyser le champ de vitesse de l'Univers. Toutefois, l'analyse de la méthode et de ses résultats a transformé cet outil en objet d'étude. Les biais inhérents à la méthode ainsi que les possibilités d'évolution sont rapidement devenus une préoccupation centrale pour les trois années qui s'ensuivirent.

**Modélisation statistique** Le chapitre 2 a montré comment les modélisations probabilistes permettent de mettre en exergue les erreurs systématiques des méthodes d'analyse et les corrections possibles. En particulier, l'estimation de la vitesse particulière à partir de la seule mesure du module de distance est biaisée, et il faut prendre en compte les effets de sélection et de distribution des galaxies. Nous avons distingué trois effets différents : le biais de densité, le biais de positionnement et les biais de sélection. Le rôle du biais de positionnement est un peu différent des autres puisqu'il est spécifique aux reconstructions de champ de vitesse. L'élément principal de l'analyse est que la distance d'une galaxie rentre en compte à trois niveaux différents : son positionnement dans l'espace réel, le calcul de sa vitesse particulière radiale et le calcul de la corrélation avec les autres galaxies.

La modélisation de ces effets nous a amené à intégrer le modèle décrit dans le chapitre 3 au sein des analyses de *CosmicFlows*. Ce modèle, adapté de Lavaux (2016), permet de considérer les distances des galaxies comme paramètres libres du modèle. De cette façon, la reconstruction du champ de vitesse prend en compte les incertitudes sur les distances et permet en outre d'imposer des *a priori* sur ces distances afin de modéliser les effets de sélection.

**Champ de vitesse de *CosmicFlows-3*** On a ainsi pu reconstruire, dans le chapitre 4, le champ de vitesse et de surdensité de l'Univers Local à partir du catalogue *CosmicFlows-3*, donnant accès à la cartographie la plus étendue et la plus précise calculée à ce jour. Cette carte donne accès à la distribution de matière totale dans l'espace physique, et permet d'étudier les grandes structures de l'Univers. L'analyse cosmographique faite dans le chapitre 4 a permis l'identification de grandes structures, comme le Mur du Sud, et de vides cosmiques tels celui d'Hercule. Cette analyse cosmographique a été faite à travers l'analyse des surdensités, mais aussi à partir de la toile cosmique des vitesses, qui permet d'identifier les structures en fonction des directions principales d'accrétion. On a également mis en évidence la compatibilité du champ de surdensité ainsi retrouvé avec la distribution des galaxies de l'Univers Local, ce qui montre le fort potentiel des études de vitesses particulières en terme de cosmographie. Ce travail a notamment conduit à l'écriture d'un article en cours de soumission.

**Limites du modèle actuel** Le chapitre 4 a aussi souligné les limites de l'analyse. Cette étude importante a montré que les erreurs systématiques du modèle étaient principalement dues à deux éléments : la modélisation des effets de sélection à partir d'*a priori* empiriques, et la modélisation de l'erreur sur la calibration des indicateurs de distance à partir d'un unique paramètre  $h_{\text{eff}}$ . Ces deux éléments, calibration et effets de sélection, sont idéalement modélisés dans l'espace des observations, plutôt qu'à partir des modules de distance directement (qui sont des paramètres dérivés). Cette conclusion nous a conduit à développer le modèle de la relation de Tully-Fisher présenté dans le chapitre 5.

**Intégration de la relation de Tully-Fisher** La nécessité d'inclure les données observationnelles primaires dans l'analyse des vitesses particulières nous a amené à développer un nouvel algorithme dans la section 5. Cet algorithme inclut la propagation des données observationnelles de la relation de Tully-Fisher jusqu'à l'estimation du champ de vitesse linéaire. Il permet en théorie de maîtriser les effets de sélection ainsi que la calibration de la TFR sur les vitesses particulières. On a notamment vu que l'utilisation directe des modules de distance estimés à partir de la TFR biaisait le champ de vitesse reconstruit.

La première analyse du modèle de TFR développée montre aussi que des étapes importantes de modélisation restent à faire. Ces étapes reposent sur l'existence d'une forte corrélation entre les calibrateurs de l'indicateur de distance, la constante de Hubble, et le monopole du champ de vitesse de l'Univers Local. En particulier, le lien entre les indicateurs primaires et secondaires doit être précisément construit et le couplage au champ de vitesse soigneusement déterminé. Les deux prochaines étapes sont les suivantes : la première est l'inclusion de nouveaux paramètres physiques dans la modélisation de la TFR afin de réduire la dispersion intrinsèque de la relation. Entre autres, on peut penser au type morphologique, à la masse du gaz et à l'inclusion de toutes les couleurs disponibles. La deuxième étape est la définition des calibrateurs utilisables pour les données CF3. Enfin, il est nécessaire de développer des modèles similaires pour les autres indicateurs de distance afin d'intercalibrer précisément les différentes sources de données de vitesses particulières. La modélisation de la relation du Plan Fondamental serait en particulier une grande avancée pour le futur de *CosmicFlows*.

**Perspectives** Cette thèse a surtout mis en évidence qu'une modélisation précise des données observationnelles doit précéder les analyses des vitesses particulières. Avec l'arrivée de grands sondages cosmologiques à faible décalage vers le rouge comme TAIPAN, Wallaby ou ZTF, il est indispensable de travailler à la réduction des erreurs systématiques inhérentes aux mesures de vitesses particulières.

# Bibliographie

- B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, and et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551 :85–88, Nov. 2017. doi : 10.1038/nature24471.
- B. Abolfathi, D. S. Aguado, G. Aguilar, C. Allende Prieto, A. Almeida, T. Tasnim Ananna, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, A. Aragón-Salamanca, M. Argudo-Fernández, E. Armengaud, M. Ata, E. Aubourg, V. Avila-Reese, C. Badenes, S. Bailey, C. Balland, K. A. Barger, J. Barrera-Ballesteros, C. Bartosz, F. Bastien, D. Bates, F. Baumgarten, J. Bautista, R. Beaton, T. C. Beers, F. Belfiore, C. F. Bender, M. Bernardi, M. A. Bershad, F. Beutler, J. C. Bird, D. Bizyaev, G. A. Blanc, M. R. Blanton, M. Blomqvist, A. S. Bolton, M. Boquien, J. Borissova, J. Bovy, C. Andres Bradna Diaz, W. Nielsen Brandt, J. Brinkmann, J. R. Brownstein, K. Bundy, A. J. Burgasser, E. Burtin, N. G. Busca, C. I. Cañas, M. Cano-Díaz, M. Cappellari, R. Carrera, A. R. Casey, B. Cervantes Sodi, Y. Chen, B. Cherinka, C. Chiappini, P. Doohyun Choi, D. Chojnowski, C.-H. Chuang, H. Chung, N. Clerc, R. E. Cohen, J. M. Comerford, J. Comparat, J. Correa do Nascimento, L. da Costa, M.-C. Cousinou, K. Covey, J. D. Crane, I. Cruz-Gonzalez, K. Cunha, G. da Silva Ilha, G. J. Damke, J. Darling, J. W. Davidson, Jr., K. Dawson, M. A. C. de Icaza Lizaola, A. de la Macorra, S. de la Torre, N. De Lee, V. de Sainte Agathe, A. Deconto Machado, F. Dell’Agli, T. Delubac, A. M. Diamond-Stanic, J. Donor, J. José Downes, N. Drory, H. du Mas des Bourboux, C. J. Duckworth, T. Dwelly, J. Dyer, G. Ebelke, A. Davis Eigenbrot, D. J. Eisenstein, Y. P. Elsworth, E. Emsellem, M. Eracleous, G. Erfanianfar, S. Escoffier, X. Fan, E. Fernández Alvar, J. G. Fernandez-Trincado, R. F. Cirolini, D. Feuillet, A. Finoguenov, S. W. Fleming, A. Font-Ribera, G. Freisclad, P. Frinchaboy, H. Fu, Y. Gómez Maqueo Chew, L. Galbany, A. E. García Pérez, R. Garcia-Dias, D. A. García-Hernández, L. A. Garma Oehmichen, P. Gaulme, J. Gelfand, H. Gil-Marín, B. A. Gillespie, D. Goddard, J. I. González Hernández, V. Gonzalez-Perez, K. Grabowski, P. J. Green, C. J. Grier, A. Gueguen, H. Guo, J. Guy, A. Hagen, P. Hall, P. Harding, S. Hasselquist, S. Hawley, C. R. Hayes, F. Hearty, S. Hekker, J. Hernandez, H. Hernandez Toledo, D. W. Hogg, K. Holley-Bockelmann, J. A. Holtzman, J. Hou, B.-C. Hsieh, J. A. S. Hunt, T. A. Hutchinson, H. S. Hwang, C. E. Jimenez Angel, J. A. Johnson, A. Jones, H. Jönsson, E. Jullo, F. Sakil Khan, K. Kinemuchi, D. Kirkby, C. C. Kirkpatrick, IV, F.-S. Kitaura, G. R. Knapp, J.-P. Kneib, J. A. Kollmeier, I. Lacerna, R. R. Lane, D. Lang, D. R. Law, J.-M. Le Goff, Y.-B. Lee, H. Li, C. Li, J. Lian, Y. Liang, M. Lima, L. Lin, D. Long, S. Lucatello, B. Lundgren, J. T. Mackereth, C. L. MacLeod, S. Mahadevan, M. A. Geimba Maia, S. Ma-



- jewski, A. Manchado, C. Maraston, V. Mariappan, R. Marques-Chaves, T. Masseron, K. L. Masters, R. M. McDermid, I. D. McGreer, M. Melendez, S. Meneses-Goytia, A. Merloni, M. R. Merrifield, S. Meszaros, A. Meza, I. Minchev, D. Minniti, E.-M. Mueller, F. Muller-Sanchez, D. Muna, R. R. Muñoz, A. D. Myers, P. Nair, K. Nandra, M. Ness, J. A. Newman, R. C. Nichol, D. L. Nidever, C. Nitschelm, P. Noterdaeme, J. O'Connell, R. J. Oelkers, A. Oravetz, D. Oravetz, E. Aquino Ortíz, Y. Osorio, Z. Pace, N. Padilla, N. Palanque-Delabrouille, P. Alonso Palicio, H.-A. Pan, K. Pan, T. Parikh, I. Pâris, C. Park, S. Peirani, M. Pellejero-Ibanez, S. Penny, W. J. Percival, I. Perez-Fournon, P. Petitjean, M. M. Pieri, M. Pinsonneault, A. Pisani, F. Prada, A. Prakash, A. B. d. A. Queiroz, M. J. Raddick, A. Raichoor, S. Barboza Rembold, H. Richstein, R. A. Riffel, R. Riffel, H.-W. Rix, A. C. Robin, S. Rodríguez Torres, C. Román-Zúñiga, A. J. Ross, G. Rossi, J. Ruan, R. Ruggeri, J. Ruiz, M. Salvato, A. G. Sánchez, S. F. Sánchez, J. Sanchez Almeida, J. R. Sánchez-Gallego, F. A. Santana Rojas, B. X. Santiago, R. P. Schiavon, J. S. Schimoia, E. Schlafly, D. Schlegel, D. P. Schneider, W. J. Schuster, A. Schwobe, H.-J. Seo, A. Serenelli, S. Shen, Y. Shen, M. Shetrone, M. Shull, V. Silva Aguirre, J. D. Simon, M. Skrutskie, A. Slosar, R. Smethurst, V. Smith, J. Sobek, G. Somers, B. J. Souter, D. Souto, A. Spindler, D. V. Stark, K. Stassun, M. Steinmetz, D. Stello, T. Storchi-Bergmann, A. Streblyanska, G. S. Stringfellow, G. Suárez, J. Sun, L. Szigeti, M. Taghizadeh-Popp, M. S. Talbot, B. Tang, C. Tao, J. Tayar, M. Tembe, J. Teske, A. R. Thakar, D. Thomas, P. Tissera, R. Tojeiro, C. Tremonti, N. W. Troup, M. Urry, O. Valenzuela, R. van den Bosch, J. Vargas-González, M. Vargas-Magaña, J. A. Vazquez, S. Villanova, N. Vogt, D. Wake, Y. Wang, B. A. Weaver, A.-M. Weijmans, D. H. Weinberg, K. B. Westfall, D. G. Whelan, E. Wilcots, V. Wild, R. A. Williams, J. Wilson, W. M. Wood-Vasey, D. Wylezalek, T. Xiao, R. Yan, M. Yang, J. E. Ybarra, C. Yèche, N. Zakamska, O. Zamora, P. Zarrouk, G. Zasowski, K. Zhang, C. Zhao, G.-B. Zhao, Z. Zheng, Z. Zheng, Z.-M. Zhou, G. Zhu, J. C. Zinn, and H. Zou. The Fourteenth Data Release of the Sloan Digital Sky Survey : First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. *The Astrophysical Journal Supplement Series*, 235 :42, Apr. 2018. ISSN 0067-0049. doi : 10.3847/1538-4365/aa9e8a. URL <http://adsabs.harvard.edu/abs/2018ApJS...235...42A>.
- C. Adami, M. Marcelin, P. Amram, and D. Russeil. The intra-cluster medium influence on spiral galaxies. *Astronomy and Astrophysics*, 349 :812–818, Sept. 1999. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1999A%26A...349...812A>.
- P. a. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, C. Baccigalupi, A. J. Banday, R. B. Barreiro, R. Barrena, M. Bartelmann, J. G. Bartlett, N. Bartolo, S. Basak, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J.-P. Bernard, M. Bersanelli, B. Bertin-court, M. Bethermin, P. Bielewicz, I. Bikmaev, A. Blanchard, J. Bobin, J. J. Bock, H. Böhringer, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, H. Bourdin, J. W. Bowyer, M. Bridges, M. L. Brown, M. Bucher, R. Burenin, C. Burigana, R. C. Butler, E. Calabrese, B. Cappellini, J.-F. Cardoso, R. Carr, P. Carvalho, M. Casale, G. Castex, A. Catalano, A. Challinor, A. Chamballu, R.-R. Chary, X. Chen, H. C. Chiang, L.-Y. Chiang,

G. Chon, P. R. Christensen, E. Churazov, S. Church, M. Clemens, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, B. Comis, F. Couchot, A. Coulais, B. P. Crill, M. Cruz, A. Curto, F. Cuttaia, A. D. Silva, H. Dahle, L. Danese, R. D. Davies, R. J. Davis, P. d. Bernardis, A. d. Rosa, G. d. Zotti, T. Déchelette, J. Delabrouille, J.-M. Delouis, J. Démoclès, F.-X. Désert, J. Dick, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, O. Fabre, E. Falgarone, M. C. Falvella, Y. Fantaye, J. Fergusson, C. Filliard, F. Finelli, I. Flores-Cacho, S. Foley, O. Forni, P. Fosalba, M. Frailis, A. A. Fraisse, E. Franceschi, M. Freschi, S. Fromenteau, M. Frommert, T. C. Gaier, S. Galeotta, J. Gallegos, S. Galli, B. Gandolfo, K. Ganga, C. Gauthier, R. T. Génova-Santos, T. Ghosh, M. Giard, G. Giardino, M. Gilfanov, D. Girard, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Haissinski, J. Hamann, F. K. Hansen, M. Hansen, D. Hanson, D. L. Harrison, A. Heavens, G. Helou, A. Hempel, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, S. Ho, M. Hobson, W. A. Holmes, A. Hornstrup, Z. Hou, W. Hovest, G. Huey, K. M. Hufenberger, G. Hurier, S. Ilić, A. H. Jaffe, T. R. Jaffe, J. Jasche, J. Jewell, W. C. Jones, M. Juvela, P. Kalberla, P. Kangaslahti, E. Keihänen, J. Kerp, R. Keskitalo, I. Khamitov, K. Kiiveri, J. Kim, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, F. Lacasa, G. Lagache, A. Lähteenmäki, J.-M. Lamarre, M. Langer, A. Lasenby, M. Lattanzi, R. J. Laureijs, A. Lavabre, C. R. Lawrence, M. L. Jeune, S. Leach, J. P. Leahy, R. Leonardi, J. León-Tavares, C. Leroy, J. Lesgourgues, A. Lewis, C. Li, A. Liddle, M. Liguori, P. B. Lilje, M. Linden-Vørnle, V. Lindholm, M. López-Caniiego, S. Lowe, P. M. Lubin, J. F. Macías-Pérez, C. J. MacTavish, B. Maffei, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, D. Marinucci, M. Maris, F. Marleau, D. J. Marshall, P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, T. Matsumura, F. Matthai, L. Maurin, P. Mazzotta, A. McDonald, J. D. McEwen, P. McGehee, S. Mei, P. R. Meinhold, A. Melchiorri, J.-B. Melin, L. Mendes, E. Menegoni, A. Mennella, M. Migliaccio, K. Mikkelsen, M. Millea, R. Miniscalco, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, N. Morisset, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, M. Negrello, N. P. H. Nesvadba, C. B. Netterfield, H. U. Nørgaard-Nielsen, C. North, F. Noviello, D. Novikov, I. Novikov, I. J. O'Dwyer, F. Orioux, S. Osborne, C. O'Sullivan, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, S. Pandolfi, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, P. Paykari, D. Pearson, T. J. Pearson, M. Peel, H. V. Peiris, O. Perdureau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, P. Platania, D. Pogosyan, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, G. Prézeau, S. Prunet, J.-L. Puget, A. R. Pullen, J. P. Rachen, B. Racine, A. Rahlin, C. R  th, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, A. Riazuelo, S. Ricciardi, T. Riller, C. Ringeval, I. Ristorcelli, G. Robbers, G. Rocha, M. Roman, C. Rosset, M. Rossetti, G. Roudier, M. Rowan-Robinson, J. A. Rubi  o-Mart  n, B. Ruiz-Granados, B. Rusholme, E. Salerno, M. Sandri, L. Sanselme, D. Santos, M. Savelainen, G. Savini, B. M. Schaefer, F. Schiavon, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, K. Smith, G. F. Smoot, T. Souradeep, L. D. Spencer, J.-L. Starck, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, F. Sureau, P. Sutter, D. Sutton, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, D. Tavagnacco, D. Tay-

- lor, L. Terenzi, D. Texier, L. Toffolatti, M. Tomasi, J.-P. Torre, M. Tristram, M. Tucci, J. Tuovinen, M. Türlér, M. Tuttlebee, G. Umana, L. Valenziano, J. Valiviita, B. V. Tent, J. Varis, L. Vibert, M. Viel, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, C. Watson, R. Watson, I. K. Wehus, N. Welikala, J. Weller, M. White, S. D. M. White, A. Wilkinson, B. Winkel, J.-Q. Xia, D. Yvon, A. Zacchei, J. P. Zibin, and A. Zonca. Planck 2013 results. I. Overview of products and scientific results. *Astronomy & Astrophysics*, 571 : A1, Nov. 2014. ISSN 0004-6361, 1432-0746. doi : 10.1051/0004-6361/201321529. URL <https://www.aanda.org/articles/aa/abs/2014/11/aa21529-13/aa21529-13.html>.
- E. Aquino-Ortíz, O. Valenzuela, S. F. Sánchez, H. Hernández-Toledo, V. Ávila-Reese, G. van de Ven, A. Rodríguez-Puebla, L. Zhu, B. Mancillas, M. Cano-Díaz, and R. García-Benito. Kinematic scaling relations of CALIFA galaxies : A dynamical mass proxy for galaxies across the Hubble sequence. *Monthly Notices of the Royal Astronomical Society*, page 1451, June 2018. doi : 10.1093/mnras/sty1522. URL <https://ui.adsabs.harvard.edu/#abs/2018MNRAS.tmp.1451A/abstract>.
- E. Bellm. The Zwicky Transient Facility. pages 27–33, eprint : arXiv :1410.8185, 2014. URL <http://adsabs.harvard.edu/abs/2014htu...conf...27B>.
- M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas, P. Astier, P. El-Hage, M. Konig, S. Kuhlmann, J. Marriner, R. Pain, N. Regnault, C. Balland, B. A. Bassett, P. J. Brown, H. Campbell, R. G. Carlberg, F. Cellier-Holzem, D. Cinabro, A. Conley, C. B. D’Andrea, D. L. DePoy, M. Doi, R. S. Ellis, S. Fabbro, A. V. Filippenko, R. J. Foley, J. A. Frieman, D. Fouchez, L. Galbany, A. Goobar, R. R. Gupta, G. J. Hill, R. Hlozek, C. J. Hogan, I. M. Hook, D. A. Howell, S. W. Jha, L. Le Guillou, G. Leloudas, C. Lidman, J. L. Marshall, A. Möller, A. M. Mourão, J. Neveu, R. Nichol, M. D. Olmstead, N. Palanque-Delabrouille, S. Perlmutter, J. L. Prieto, C. J. Pritchett, M. Richmond, A. G. Riess, V. Ruhlmann-Kleider, M. Sako, K. Schahmanche, D. P. Schneider, M. Smith, J. Sollerman, M. Sullivan, N. A. Walton, and C. J. Wheeler. Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *Astronomy and Astrophysics*, 568 :A22, Aug. 2014. ISSN 0004-6361. doi : 10.1051/0004-6361/201423413. URL <http://adsabs.harvard.edu/abs/2014A%26A...568A...22B>.
- M. Bilicki, J. A. Peacock, T. H. Jarrett, M. E. Cluver, N. Maddox, M. J. I. Brown, E. N. Taylor, N. C. Hambly, A. Solarz, B. W. Holwerda, I. Baldry, J. Loveday, A. Moffett, A. M. Hopkins, S. P. Driver, M. Alpaslan, and J. Bland-Hawthorn. WISE x SuperCOSMOS Photometric Redshift Catalog : 20 Million Galaxies over  $3/\pi$  Steradians. *The Astrophysical Journal Supplement Series*, 225 :5, July 2016. ISSN 0067-0049. doi : 10.3847/0067-0049/225/1/5. URL <http://adsabs.harvard.edu/abs/2016ApJS...225...5B>.
- V. Bistolas and Y. Hoffman. Nonlinear Constrained Realizations of the Large-Scale Structure. *The Astrophysical Journal*, 492 :439–451, Jan. 1998. ISSN 0004-637X. doi : 10.1086/305080. URL <http://adsabs.harvard.edu/abs/1998ApJ...492...439B>.

- G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees. Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311 :517–525, Oct. 1984. ISSN 0028-0836. doi : 10.1038/311517a0. URL <http://adsabs.harvard.edu/abs/1984Natur.311..517B>.
- M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat. Too big to fail? The puzzling darkness of massive Milky Way subhaloes. *Monthly Notices of the Royal Astronomical Society*, 415 : L40–L44, July 2011. ISSN 0035-8711. doi : 10.1111/j.1745-3933.2011.01074.x. URL <http://adsabs.harvard.edu/abs/2011MNRAS.415L..40B>.
- J. R. Brownstein and J. W. Moffat. The Bullet Cluster 1e0657-558 evidence shows modified gravity in the absence of dark matter. *Monthly Notices of the Royal Astronomical Society*, 382 :29–47, Nov. 2007. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2007.12275.x. URL <http://adsabs.harvard.edu/abs/2007MNRAS.382...29B>.
- T. Buchert. On Backreaction in Newtonian cosmology. *Monthly Notices of the Royal Astronomical Society*, 473 :L46–L49, Jan. 2018. ISSN 0035-8711. doi : 10.1093/mnrasl/slx160. URL <http://adsabs.harvard.edu/abs/2018MNRAS.473L..46B>.
- L. A. Campbell, J. R. Lucey, M. Colless, D. H. Jones, C. M. Springob, C. Magoulas, R. N. Proctor, J. R. Mould, M. A. Read, S. Brough, T. Jarrett, A. I. Merson, P. Lah, F. Beutler, M. E. Cluver, and Q. A. Parker. The 6df Galaxy Survey : Fundamental Plane data. *Monthly Notices of the Royal Astronomical Society*, 443 :1231–1251, Sept. 2014. ISSN 0035-8711. doi : 10.1093/mnras/stu1198. URL <http://adsabs.harvard.edu/abs/2014MNRAS.443.1231C>.
- E. Carlesi, Y. Hoffman, J. G. Sorce, S. Gottlöber, G. Yepes, H. Courtois, and R. B. Tully. The tangential velocity of M31 : CLUES from constrained simulations. *Monthly Notices of the Royal Astronomical Society*, 460 :L5–L9, July 2016. ISSN 0035-8711. doi : 10.1093/mnrasl/slw059. URL <http://adsabs.harvard.edu/abs/2016MNRAS.460L...5C>.
- J. Carrick, S. J. Turnbull, G. Lavaux, and M. J. Hudson. Cosmological parameters from the comparison of peculiar velocities with predictions from the 2m++ density field. *Monthly Notices of the Royal Astronomical Society*, 450 :317–332, June 2015. ISSN 0035-8711. doi : 10.1093/mnras/stv547. URL <http://adsabs.harvard.edu/abs/2015MNRAS.450..317C>.
- H. M. Courtois and R. B. Tully. Cosmic Flows surveys and CLUES simulations. *Astronomische Nachrichten*, 333 :436, June 2012. ISSN 0004-6337. doi : 10.1002/asna.201211682. URL <http://adsabs.harvard.edu/abs/2012AN....333..436C>.
- H. M. Courtois, R. B. Tully, D. I. Makarov, S. Mitronova, B. Koribalski, I. D. Karachentsev, and J. R. Fisher. Cosmic Flows : Green Bank Telescope and Parkes H I observations. *Monthly Notices of the Royal Astronomical Society*, 414 :2005–2016, July 2011. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2011.18515.x. URL <http://adsabs.harvard.edu/abs/2011MNRAS.414.2005C>.
- H. M. Courtois, Y. Hoffman, R. B. Tully, and S. Gottlöber. Three-dimensional Velocity and Density Reconstructions of the Local Universe with Cosmicflows-1. *The Astrophysical Journal*, 744 :43, Jan. 2012. ISSN 0004-637X. doi : 10.1088/0004-637X/744/1/43. URL <http://adsabs.harvard.edu/abs/2012ApJ...744...43C>.

- H. M. Courtois, R. B. Tully, Y. Hoffman, D. Pomarède, R. Graziani, and A. Dupuy. Cosmicflows-3 : Cold Spot Repeller? *The Astrophysical Journal Letters*, 847 :L6, Sept. 2017. ISSN 0004-637X. doi : 10.3847/2041-8213/aa88b2. URL <http://adsabs.harvard.edu/abs/2017ApJ...847L...6C>.
- M. Crocce, S. Pueblas, and R. Scoccimarro. Transients from initial conditions in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 373 :369–381, Nov. 2006. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2006.11040.x. URL <http://adsabs.harvard.edu/abs/2006MNRAS.373...369C>.
- A. C. Crook, J. P. Huchra, N. Martimbeau, K. L. Masters, T. Jarrett, and L. M. Macri. Groups of Galaxies in the Two Micron All Sky Redshift Survey. *The Astrophysical Journal*, 655 : 790–813, Feb. 2007. ISSN 0004-637X. doi : 10.1086/510201. URL <http://adsabs.harvard.edu/abs/2007ApJ...655...790C>.
- M. Cruz, E. Martínez-González, P. Vielva, and L. Cayón. Detection of a non-Gaussian spot in WMAP. *Monthly Notices of the Royal Astronomical Society*, 356 :29–40, Jan. 2005. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2004.08419.x. URL <http://adsabs.harvard.edu/abs/2005MNRAS.356...29C>.
- E. da Cunha, A. M. Hopkins, M. Colless, E. N. Taylor, C. Blake, C. Howlett, C. Magoulas, J. R. Lucey, C. Lagos, K. Kuehn, Y. Gordon, D. Barat, F. Bian, C. Wolf, M. J. Cowley, M. White, I. Achitouv, M. Bilicki, J. Bland-Hawthorn, K. Bolejko, M. J. I. Brown, R. Brown, J. Bryant, S. Croom, T. M. Davis, S. P. Driver, M. D. Filipovic, S. R. Hinton, M. Johnston-Hollitt, D. H. Jones, B. Koribalski, D. Kleiner, J. Lawrence, N. Lorente, J. Mould, M. S. Owers, K. Pimbblet, C. G. Tinney, N. F. H. Tothill, and F. Watson. The Taipan Galaxy Survey : Scientific Goals and Observing Strategy. *Publications of the Astronomical Society of Australia*, 34 :e047, Oct. 2017. ISSN 1323-3580. doi : 10.1017/pasa.2017.41. URL <http://adsabs.harvard.edu/abs/2017PASA...34...47D>.
- T. M. Davis and M. I. Scrimgeour. Deriving accurate peculiar velocities (even at high redshift). *Monthly Notices of the Royal Astronomical Society*, 442 :1117–1122, Aug. 2014. ISSN 0035-8711. doi : 10.1093/mnras/stu920. URL <http://adsabs.harvard.edu/abs/2014MNRAS.442.1117D>.
- W. de Sitter. Einstein’s theory of gravitation and its astronomical consequences. Third paper. *Monthly Notices of the Royal Astronomical Society*, 78 :3–28, Nov. 1917. ISSN 0035-8711. doi : 10.1093/mnras/78.1.3. URL <http://adsabs.harvard.edu/abs/1917MNRAS...78...3D>.
- A. Dekel, A. Eldar, T. Kolatt, A. Yahil, J. A. Willick, S. M. Faber, S. Courteau, and D. Burstein. POTENT Reconstruction from Mark III Velocities. *The Astrophysical Journal*, 522(1) :1, Sept. 1999. ISSN 0004-637X. doi : 10.1086/307636. URL <http://iopscience.iop.org/article/10.1086/307636/meta>.
- H. Desmond. The scatter, residual correlations and curvature of the SPARC baryonic Tully-Fisher relation. *Monthly Notices of the Royal Astronomical Society*, 472 :L35–L39, Nov.



2017. ISSN 0035-8711. doi : 10.1093/mnrasl/slx134. URL <http://adsabs.harvard.edu/abs/2017MNRAS.472L..35D>.
- A. Di Cintio, A. Knebe, N. I. Libeskind, C. Brook, G. Yepes, S. Gottlöber, and Y. Hoffman. Size matters : the non-universal density profile of subhaloes in SPH simulations and implications for the Milky Way's dSphs. *Monthly Notices of the Royal Astronomical Society*, 431 :1220–1229, May 2013. ISSN 0035-8711. doi : 10.1093/mnras/stt240. URL <http://adsabs.harvard.edu/abs/2013MNRAS.431.1220D>.
- A. G. Doroshkevich, V. S. Ryaben'kii, and S. F. Shandarin. Nonlinear theory of the development of potential perturbations. *Astrophysics*, 9 :144–153, Apr. 1973. ISSN 0571-7256. doi : 10.1007/BF01011421. URL <http://adsabs.harvard.edu/abs/1973Ap.....9..144D>.
- T. Doumler. *Constrained local universe simulations from galaxy peculiar velocities*. Lyon 1, June 2012. URL <http://www.theses.fr/2012LY010076>.
- T. Doumler, S. Gottlöber, Y. Hoffman, and H. Courtois. Reconstructing cosmological initial conditions from galaxy peculiar velocities - III. Constrained simulations. *Monthly Notices of the Royal Astronomical Society*, 430 :912–923, Apr. 2013. ISSN 0035-8711. doi : 10.1093/mnras/sts614. URL <http://adsabs.harvard.edu/abs/2013MNRAS.430..912D>.
- A. Dressler and S. M. Faber. New measurements of distances to spirals in the great attractor - Further confirmation of the large-scale flow. *The Astrophysical Journal Letters*, 354 :L45–L48, May 1990. ISSN 0004-637X. doi : 10.1086/185719. URL <http://adsabs.harvard.edu/abs/1990ApJ...354L..45D>.
- A. R. Duffy, M. J. Meyer, L. Staveley-Smith, M. Bernyk, D. J. Croton, B. S. Koribalski, D. Gerstmann, and S. Westerlund. Predictions for ASKAP neutral hydrogen surveys. *Monthly Notices of the Royal Astronomical Society*, 426 :3385–3402, Nov. 2012. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2012.21987.x. URL <http://adsabs.harvard.edu/abs/2012MNRAS.426.3385D>.
- A. A. Dutton, A. Obreja, L. Wang, T. A. Gutcke, T. Buck, S. M. Udrescu, J. Frings, G. S. Stinson, X. Kang, and A. V. Macciò. NIHAO XII : galactic uniformity in a  $\Lambda$ CDM universe. *Monthly Notices of the Royal Astronomical Society*, 467 :4937–4950, June 2017. ISSN 0035-8711. doi : 10.1093/mnras/stx458. URL <http://adsabs.harvard.edu/abs/2017MNRAS.467.4937D>.
- J. Einasto, A. Kaasik, and E. Saar. Dynamic evidence on massive coronas of galaxies. *Nature*, 250(5464) :309–310, July 1974. ISSN 1476-4687. doi : 10.1038/250309a0. URL <https://www.nature.com/articles/250309a0>.
- D. J. Eisenstein and W. Hu. Baryonic Features in the Matter Transfer Function. *The Astrophysical Journal*, 496 :605–614, Mar. 1998. ISSN 0004-637X. doi : 10.1086/305424. URL <http://adsabs.harvard.edu/abs/1998ApJ...496..605E>.

- D. J. Eisenstein and W. Hu. Power Spectra for Cold Dark Matter and Its Variants. *The Astrophysical Journal*, 511 :5–15, Jan. 1999. ISSN 0004-637X. doi : 10.1086/306640. URL <http://adsabs.harvard.edu/abs/1999ApJ...511....5E>.
- B. Falck, K. Koyama, G.-B. Zhao, and M. Cautun. Using voids to unscreen modified gravity. *Monthly Notices of the Royal Astronomical Society*, 475 :3262–3272, Apr. 2018. ISSN 0035-8711. doi : 10.1093/mnras/stx3288. URL <http://adsabs.harvard.edu/abs/2018MNRAS.475.3262F>.
- B. Famaey and S. S. McGaugh. Modified Newtonian Dynamics (MOND) : Observational Phenomenology and Relativistic Extensions. *Living Reviews in Relativity*, 15 :10, Sept. 2012. doi : 10.12942/lrr-2012-10. URL <http://adsabs.harvard.edu/abs/2012LRR...15...10F>.
- D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright. The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set. *The Astrophysical Journal*, 473 :576, Dec. 1996. ISSN 0004-637X. doi : 10.1086/178173. URL <http://adsabs.harvard.edu/abs/1996ApJ...473..576F>.
- A. Friedman. Über die Krümmung des Raumes. *Zeitschrift für Physik*, 10(1) :377–386, Dec. 1922. ISSN 0044-3328. doi : 10.1007/BF01332580. URL <https://link.springer.com/article/10.1007/BF01332580>.
- P. Ganeshiah Veena, M. Cautun, R. van de Weygaert, E. Tempel, B. J. T. Jones, S. Rieder, and C. S. Frenk. The Cosmic Ballet : spin and shape alignments of haloes in the cosmic web. *ArXiv e-prints*, 1805 :arXiv :1805.00033, Apr. 2018. URL <http://adsabs.harvard.edu/abs/2018arXiv180500033G>.
- R. Giovanelli. On the scaling relations of disk galaxies. volume 289, pages 296–303, eprint : arXiv :1304.0346, Feb. 2013. doi : 10.1017/S1743921312021588. URL <http://adsabs.harvard.edu/abs/2013IAUS..289..296G>.
- R. Giovanelli, M. P. Haynes, B. R. Kent, P. Perillat, A. Saintonge, N. Brosch, B. Catinella, G. L. Hoffman, S. Stierwalt, K. Spekkens, M. S. Lerner, K. L. Masters, E. Momjian, J. L. Rosenberg, C. M. Springob, A. Boselli, V. Charmandaris, J. K. Darling, J. Davies, D. Garcia Lambas, G. Gavazzi, C. Giovanardi, E. Hardy, L. K. Hunt, A. Iovino, I. D. Karachentsev, V. E. Karachentseva, R. A. Koopmann, C. Marinoni, R. Minchin, E. Muller, M. Putman, C. Pantoja, J. J. Salzer, M. Scodreggio, E. Skillman, J. M. Solanes, C. Valotto, W. van Driel, and L. van Zee. The Arecibo Legacy Fast ALFA Survey. I. Science Goals, Survey Design, and Strategy. *The Astronomical Journal*, 130 :2598–2612, Dec. 2005. ISSN 0004-6256. doi : 10.1086/497431. URL <http://adsabs.harvard.edu/abs/2005AJ...130.2598G>.
- E. Giraud. The dependence of the Tully-Fisher relation on morphological type. *The Astrophysical Journal*, 309 :512–521, Oct. 1986. ISSN 0004-637X. doi : 10.1086/164621. URL <http://adsabs.harvard.edu/abs/1986ApJ...309..512G>.

- G. Giuricin, M. Mezzetti, and F. Mardirossian. The blue Tully-Fisher relation in a variety of environments. *Astronomy and Astrophysics*, 157 :400–402, Mar. 1986. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1986A%26A...157..400G>.
- J. R. Gott, III, M. Jurić, D. Schlegel, F. Hoyle, M. Vogeley, M. Tegmark, N. Bahcall, and J. Brinkmann. A Map of the Universe. *The Astrophysical Journal*, 624 :463–484, May 2005. ISSN 0004-637X. doi : 10.1086/428890. URL <http://adsabs.harvard.edu/abs/2005ApJ...624..463G>.
- S. Gottloeber, Y. Hoffman, and G. Yepes. Constrained Local Universe Simulations (CLUES). *ArXiv e-prints*, 1005 :arXiv :1005.2687, May 2010. URL <http://adsabs.harvard.edu/abs/2010arXiv1005.2687G>.
- M. P. Haynes, R. Giovanelli, A. M. Martin, K. M. Hess, A. Saintonge, E. A. K. Adams, G. Hallenbeck, G. L. Hoffman, S. Huang, B. R. Kent, R. A. Koopmann, E. Papastergis, S. Stierwalt, T. J. Balonek, D. W. Craig, S. J. U. Higdon, D. A. Kornreich, J. R. Miller, A. A. O’Donoghue, R. P. Olowin, J. L. Rosenberg, K. Spekkens, P. Troischt, and E. M. Wilcots. The Arecibo Legacy Fast ALFA Survey : The  $\alpha$ .40 H I Source Catalog, Its Characteristics and Their Impact on the Derivation of the H I Mass Function. *The Astronomical Journal*, 142 :170, Nov. 2011. ISSN 0004-6256. doi : 10.1088/0004-6256/142/5/170. URL <http://adsabs.harvard.edu/abs/2011AJ....142..170H>.
- M. P. Haynes, R. Giovanelli, B. R. Kent, E. A. K. Adams, T. J. Balonek, D. W. Craig, D. Fertig, R. Finn, C. Giovanardi, G. Hallenbeck, K. M. Hess, G. L. Hoffman, S. Huang, M. G. Jones, R. A. Koopmann, D. A. Kornreich, L. Leisman, J. R. Miller, C. Moorman, J. O’Connor, A. O’Donoghue, E. Papastergis, P. Troischt, D. Stark, and L. Xiao. The Arecibo Legacy Fast ALFA Survey : The ALFALFA Extragalactic HI Source Catalog. *ArXiv e-prints*, 1805 :arXiv :1805.11499, May 2018. URL <http://adsabs.harvard.edu/abs/2018arXiv180511499H>.
- G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Cosmological Parameter Results. *The Astrophysical Journal Supplement Series*, 208 :19, Oct. 2013. ISSN 0067-0049. doi : 10.1088/0067-0049/208/2/19. URL <http://adsabs.harvard.edu/abs/2013ApJS...208...19H>.
- S. R. Hinton, A. Kim, and T. M. Davis. Accounting for sample selection in Bayesian analyses. *ArXiv e-prints*, 1706 :arXiv :1706.03856, June 2017. URL <http://adsabs.harvard.edu/abs/2017arXiv170603856H>.
- Y. Hoffman and E. Ribak. Constrained realizations of Gaussian fields - A simple algorithm. *The Astrophysical Journal Letters*, 380 :L5–L8, Oct. 1991. ISSN 0004-637X. doi : 10.1086/186160. URL <http://adsabs.harvard.edu/abs/1991ApJ...380L...5H>.



- Y. Hoffman and E. Ribak. Primordial Gaussian perturbation fields - Constrained realizations. *The Astrophysical Journal*, 384 :448–452, Jan. 1992. ISSN 0004-637X. doi : 10.1086/170886. URL <http://adsabs.harvard.edu/abs/1992ApJ...384..448H>.
- Y. Hoffman, H. M. Courtois, and R. B. Tully. Cosmic bulk flow and the local motion from Cosmicflows-2. *Monthly Notices of the Royal Astronomical Society*, 449 :4494–4505, June 2015. ISSN 0035-8711. doi : 10.1093/mnras/stv615. URL <http://adsabs.harvard.edu/abs/2015MNRAS.449.4494H>.
- D. W. Hogg. Distance measures in cosmology. *ArXiv Astrophysics e-prints*, pages arXiv :astro-ph/9905116, May 1999. URL <http://adsabs.harvard.edu/abs/1999astro.ph..5116H>.
- T. Hong, L. Staveley-Smith, K. Masters, C. Springob, L. Macri, B. Koribalski, H. Jones, and T. Jarrett. The 2mass Tully-Fisher Survey : Mapping the mass in the Universe. volume 289, pages 312–315, eprint : arXiv :1212.2090, Feb. 2013. doi : 10.1017/S1743921312021606. URL <http://adsabs.harvard.edu/abs/2013IAUS...289..312H>.
- T. Hong, C. M. Springob, L. Staveley-Smith, M. I. Scrimgeour, K. L. Masters, L. M. Macri, B. S. Koribalski, D. H. Jones, and T. H. Jarrett. 2mtf - IV. A bulk flow measurement of the local Universe. *Monthly Notices of the Royal Astronomical Society*, 445 :402–413, Nov. 2014. ISSN 0035-8711. doi : 10.1093/mnras/stu1774. URL <http://adsabs.harvard.edu/abs/2014MNRAS.445..402H>.
- C. Howlett, L. Staveley-Smith, P. J. Elahi, T. Hong, T. H. Jarrett, D. H. Jones, B. S. Koribalski, L. M. Macri, K. L. Masters, and C. M. Springob. 2mtf - VI. Measuring the velocity power spectrum. *Monthly Notices of the Royal Astronomical Society*, 471 :3135–3151, Nov. 2017. ISSN 0035-8711. doi : 10.1093/mnras/stx1521. URL <http://adsabs.harvard.edu/abs/2017MNRAS.471.3135H>.
- E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15 :168–173, Mar. 1929. ISSN 0027-8424. doi : 10.1073/pnas.15.3.168. URL <http://adsabs.harvard.edu/abs/1929PNAS...15..168H>.
- E. Hubble and M. L. Humason. The Velocity-Distance Relation among Extra-Galactic Nebulae. *The Astrophysical Journal*, 74 :43, July 1931. ISSN 0004-637X. doi : 10.1086/143323. URL <http://adsabs.harvard.edu/abs/1931ApJ....74...43H>.
- J. P. Huchra, L. M. Macri, K. L. Masters, T. H. Jarrett, P. Berlind, M. Calkins, A. C. Crook, R. Cutri, P. Erdoğdu, E. Falco, T. George, C. M. Hutcheson, O. Lahav, J. Mader, J. D. Mink, N. Martimbeau, S. Schneider, M. Skrutskie, S. Tokarz, and M. Westover. The 2mass Redshift Survey—Description and Data Release. *The Astrophysical Journal Supplement Series*, 199 :26, Apr. 2012. ISSN 0067-0049. doi : 10.1088/0067-0049/199/2/26. URL <http://adsabs.harvard.edu/abs/2012ApJS...199...26H>.
- J. Jasche and B. D. Wandelt. Bayesian physical reconstruction of initial conditions from large-scale structure surveys. *Monthly Notices of the Royal Astronomical Society*, 432 :894–913,

- June 2013. ISSN 0035-8711. doi : 10.1093/mnras/stt449. URL <http://adsabs.harvard.edu/abs/2013MNRAS.432..894J>.
- N. Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227 :1–21, July 1987. ISSN 0035-8711. doi : 10.1093/mnras/227.1.1. URL <http://adsabs.harvard.edu/abs/1987MNRAS.227....1K>.
- I. D. Karachentsev, R. B. Tully, P.-F. Wu, E. J. Shaya, and A. E. Dolphin. Infall of Nearby Galaxies into the Virgo Cluster as Traced with Hubble Space Telescope. *The Astrophysical Journal*, 782 :4, Feb. 2014. ISSN 0004-637X. doi : 10.1088/0004-637X/782/1/4. URL <http://adsabs.harvard.edu/abs/2014ApJ...782....4K>.
- A. Kashlinsky, F. Atrio-Barandela, D. Kocevski, and H. Ebeling. A Measurement of Large-Scale Peculiar Velocities of Clusters of Galaxies : Results and Cosmological Implications. *The Astrophysical Journal Letters*, 686 :L49, Oct. 2008. ISSN 0004-637X. doi : 10.1086/592947. URL <http://adsabs.harvard.edu/abs/2008ApJ...686L..49K>.
- G. Kauffmann, S. D. M. White, and B. Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 264 :201, Sept. 1993. doi : 10.1093/mnras/264.1.201. URL <https://ui.adsabs.harvard.edu/#abs/1993MNRAS.264..201K/abstract>.
- B. C. Kelly. Some Aspects of Measurement Error in Linear Regression of Astronomical Data. *The Astrophysical Journal*, 665 :1489–1506, Aug. 2007. ISSN 0004-637X. doi : 10.1086/519947. URL <http://adsabs.harvard.edu/abs/2007ApJ...665.1489K>.
- A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada. Where Are the Missing Galactic Satellites? *The Astrophysical Journal*, 522 :82–92, Sept. 1999. ISSN 0004-637X. doi : 10.1086/307643. URL <http://adsabs.harvard.edu/abs/1999ApJ...522....82K>.
- J. Koda, C. Blake, T. Davis, C. Magoulas, C. M. Springob, M. Scrimgeour, A. Johnson, G. B. Poole, and L. Staveley-Smith. Are peculiar velocity surveys competitive as a cosmological probe? *Monthly Notices of the Royal Astronomical Society*, 445 :4267–4286, Dec. 2014. ISSN 0035-8711. doi : 10.1093/mnras/stu1610. URL <http://adsabs.harvard.edu/abs/2014MNRAS.445.4267K>.
- E. Kourkchi and R. B. Tully. Galaxy Groups Within 3500 km s<sup>-1</sup>. *The Astrophysical Journal*, 843 :16, July 2017. ISSN 0004-637X. doi : 10.3847/1538-4357/aa76db. URL <http://adsabs.harvard.edu/abs/2017ApJ...843...16K>.
- D. J. Lagattuta, J. R. Mould, L. Staveley-Smith, T. Hong, C. M. Springob, K. L. Masters, B. S. Koribalski, and D. H. Jones. WISE TF : A Mid-infrared, 3.4  $\mu$ m Extension of the Tully-Fisher Relation Using WISE Photometry. *The Astrophysical Journal*, 771 :88, July 2013. ISSN 0004-637X. doi : 10.1088/0004-637X/771/2/88. URL <http://adsabs.harvard.edu/abs/2013ApJ...771...88L>.

- O. Lahav, N. Kaiser, and Y. Hoffman. Local gravity and peculiar velocity - Probes of cosmological models. *The Astrophysical Journal*, 352 :448–456, Apr. 1990. ISSN 0004-637X. doi : 10.1086/168550. URL <http://adsabs.harvard.edu/abs/1990ApJ...352..448L>.
- L. D. Landau and E. M. Lifshitz. *Fluid Mechanics*. Elsevier, Oct. 2013. ISBN 978-1-4831-4050-6. Google-Books-ID : CeBbAwAAQBAJ.
- D. Langlois. Inflation, Quantum Fluctuations and Cosmological Perturbations. volume 188, page 235, eprint : arXiv :hep-th/0405053, Jan. 2005. URL <http://adsabs.harvard.edu/abs/2005ASIB...188..235L>.
- A. Lapi, P. Salucci, and L. Danese. Precision Scaling Relations for Disk Galaxies in the Local Universe. *The Astrophysical Journal*, 859 :2, May 2018. ISSN 0004-637X. doi : 10.3847/1538-4357/aabf35. URL <http://adsabs.harvard.edu/abs/2018ApJ...859....2L>.
- G. Lavaux. Bayesian 3d velocity field reconstruction with VIRBIUS. *Monthly Notices of the Royal Astronomical Society*, 457 :172–197, Mar. 2016. ISSN 0035-8711. doi : 10.1093/mnras/stv2915. URL <http://adsabs.harvard.edu/abs/2016MNRAS.457..172L>.
- G. Lavaux and M. J. Hudson. The 2m++ galaxy redshift catalogue. *Monthly Notices of the Royal Astronomical Society*, 416 :2840–2856, Oct. 2011. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2011.19233.x. URL <http://adsabs.harvard.edu/abs/2011MNRAS.416.2840L>.
- F. Leclercq, J. Jasche, and B. Wandelt. Bayesian analysis of the dynamic cosmic web in the SDSS galaxy survey. *Journal of Cosmology and Astro-Particle Physics*, 06 :015, June 2015. ISSN 1475-7516. doi : 10.1088/1475-7516/2015/06/015. URL <http://adsabs.harvard.edu/abs/2015JCAP...06..015L>.
- P.-F. Léget, M. V. Pruzhinskaya, A. Ciulli, E. Gangler, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, K. Barbary, S. Bongard, K. Boone, C. Buton, M. Childress, N. Chotard, Y. Copin, S. Dixon, P. Fagrelus, U. Feindt, D. Fouchez, P. Gris, B. Hayden, W. Hillebrandt, D. A. Howell, A. Kim, M. Kowalski, D. Kuesters, S. Lombardo, Q. Lin, J. Nordin, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, M. Rigault, K. Runge, D. Rubin, C. Saunders, L.-P. Says, G. Smadja, C. Sofiatti, N. Suzuki, S. Taubenberger, C. Tao, and R. C. Thomas. Correcting for peculiar velocities of Type Ia Supernovae in clusters of galaxies. *arXiv :1804.03418 [astro-ph]*, Apr. 2018. URL <http://arxiv.org/abs/1804.03418>. arXiv : 1804.03418.
- F. Lelli, S. S. McGaugh, and J. M. Schombert. The small scatter of the baryonic Tully-Fisher relation. *The Astrophysical Journal*, 816(1) :L14, Dec. 2015. ISSN 2041-8213. doi : 10.3847/2041-8205/816/1/L14. URL <http://arxiv.org/abs/1512.04543>. arXiv : 1512.04543.
- G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47 :49–59, 1927. URL <http://adsabs.harvard.edu/abs/1927ASSB...47...49L>.

- A. Lewis and S. Bridle. Cosmological parameters from CMB and other data : A Monte Carlo approach. *Physical Review D*, 66 :103511, Nov. 2002. ISSN 1550-7998. doi : 10.1103/PhysRevD.66.103511. URL <http://adsabs.harvard.edu/abs/2002PhRvD..66j3511L>.
- D. Lynden-Bell, S. M. Faber, D. Burstein, R. L. Davies, A. Dressler, R. J. Terlevich, and G. Wegner. Spectroscopy and photometry of elliptical galaxies. V - Galaxy streaming toward the new supergalactic center. *The Astrophysical Journal*, 326 :19–49, Mar. 1988. ISSN 0004-637X. doi : 10.1086/166066. URL <http://adsabs.harvard.edu/abs/1988ApJ...326...19L>.
- R. Mackenzie, T. Shanks, M. N. Bremer, Y.-C. Cai, M. L. P. Gunawardhana, A. Kovács, P. Norberg, and I. Szapudi. Evidence against a supervoid causing the CMB Cold Spot. *Monthly Notices of the Royal Astronomical Society*, 470 :2328–2338, Sept. 2017. ISSN 0035-8711. doi : 10.1093/mnras/stx931. URL <http://adsabs.harvard.edu/abs/2017MNRAS.470.2328M>.
- C. Magoulas, C. M. Springob, M. Colless, D. H. Jones, L. A. Campbell, J. R. Lucey, J. Mould, T. Jarrett, A. Merson, and S. Brough. The 6df Galaxy Survey : the near-infrared Fundamental Plane of early-type galaxies. *Monthly Notices of the Royal Astronomical Society*, 427 :245–273, Nov. 2012. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2012.21421.x. URL <http://adsabs.harvard.edu/abs/2012MNRAS.427..245M>.
- D. Makarov, P. Prugniel, N. Terekhova, H. Courtois, and I. Vauglin. HyperLEDA. III. The catalogue of extragalactic distances. *Astronomy and Astrophysics*, 570 :A13, Oct. 2014. ISSN 0004-6361. doi : 10.1051/0004-6361/201423496. URL <http://adsabs.harvard.edu/abs/2014A%26A...570A..13M>.
- K. L. Masters, C. M. Springob, M. P. Haynes, and R. Giovanelli. SFI++ I : A New I-Band Tully-Fisher Template, the Cluster Peculiar Velocity Dispersion, and H0. *The Astrophysical Journal*, 653 :861–880, Dec. 2006. ISSN 0004-637X. doi : 10.1086/508924. URL <http://adsabs.harvard.edu/abs/2006ApJ...653..861M>.
- K. L. Masters, A. Crook, T. Hong, T. H. Jarrett, B. S. Koribalski, L. Macri, C. M. Springob, and L. Staveley-Smith. 2mtf III. H I 21 cm observations of 1194 spiral galaxies with the Green Bank Telescope. *Monthly Notices of the Royal Astronomical Society*, 443 :1044–1056, Sept. 2014. ISSN 0035-8711. doi : 10.1093/mnras/stu1225. URL <http://adsabs.harvard.edu/abs/2014MNRAS.443.1044M>.
- D. S. Mathewson, V. L. Ford, and M. Buchhorn. No back-side infall into the Great Attractor. *The Astrophysical Journal Letters*, 389 :L5–L8, Apr. 1992. ISSN 0004-637X. doi : 10.1086/186335. URL <http://adsabs.harvard.edu/abs/1992ApJ...389L...5M>.
- S. S. McGaugh and J. M. Schombert. Color-Mass-to-light-ratio Relations for Disk Galaxies. *The Astronomical Journal*, 148 :77, Nov. 2014. ISSN 0004-6256. doi : 10.1088/0004-6256/148/5/77. URL <http://adsabs.harvard.edu/abs/2014AJ...148...77M>.

- Y. Mellier. Gravitational lensing and dark matter. In *Particle Dark Matter : Observations, Models and Searches*, page 56. 2010. URL <http://adsabs.harvard.edu/abs/2010pdmo.book...56M>.
- B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi. Dark Matter Substructure within Galactic Halos. *The Astrophysical Journal Letters*, 524 :L19–L22, Oct. 1999. ISSN 0004-637X. doi : 10.1086/312287. URL <http://adsabs.harvard.edu/abs/1999ApJ...524L..19M>.
- K. Naidoo, A. Benoit-Lévy, and O. Lahav. Could multiple voids explain the cosmic microwave background Cold Spot anomaly? *Monthly Notices of the Royal Astronomical Society*, 459 : L71–L75, June 2016. ISSN 0035-8711. doi : 10.1093/mnras/slw043. URL <http://adsabs.harvard.edu/abs/2016MNRAS.459L..71N>.
- B. Novosyadlyj and M. Tsizh. Voids in the Cosmic Web as a probe of dark energy. *ArXiv e-prints*, 1703 :arXiv :1703.00364, Mar. 2017. URL <http://adsabs.harvard.edu/abs/2017arXiv170300364N>.
- A. Nusser. Velocity-density correlations from the cosmicflows-3 distance catalogue and the 2mass Redshift Survey. *Monthly Notices of the Royal Astronomical Society*, 470 :445–454, Sept. 2017. ISSN 0035-8711. doi : 10.1093/mnras/stx1225. URL <http://adsabs.harvard.edu/abs/2017MNRAS.470..445N>.
- A. Nusser, E. Branchini, and M. Feix. A direct probe of cosmological power spectra of the peculiar velocity field and the gravitational lensing magnification from photometric redshift surveys. *Journal of Cosmology and Astro-Particle Physics*, 01 :018, Jan. 2013. ISSN 1475-7516. doi : 10.1088/1475-7516/2013/01/018. URL <http://adsabs.harvard.edu/abs/2013JCAP...01..018N>.
- J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *Monthly Notices of the Royal Astronomical Society*, 318 :1144–1156, Nov. 2000. ISSN 0035-8711. doi : 10.1046/j.1365-8711.2000.03779.x. URL <http://adsabs.harvard.edu/abs/2000MNRAS.318.1144P>.
- W. J. Percival and M. White. Testing cosmological structure formation using redshift-space distortions. *Monthly Notices of the Royal Astronomical Society*, 393 :297–308, Feb. 2009. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2008.14211.x. URL <http://adsabs.harvard.edu/abs/2009MNRAS.393..297P>.
- W. J. Percival, R. C. Nichol, D. J. Eisenstein, J. A. Frieman, M. Fukugita, J. Loveday, A. C. Pope, D. P. Schneider, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, I. Zehavi, N. A. Bahcall, J. Brinkmann, A. J. Connolly, and A. Meiksin. The Shape of the Sloan Digital Sky Survey Data Release 5 Galaxy Power Spectrum. *The Astrophysical Journal*, 657 :645–663, Mar. 2007. ISSN 0004-637X. doi : 10.1086/510615. URL <http://adsabs.harvard.edu/abs/2007ApJ...657..645P>.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye,

- K. Benabed, A. Benoit, A. Benoit-Lévy, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R.-R. Chary, H. C. Chiang, J. Chluba, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F.-X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, M. Farhang, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, T. Ghosh, M. Giard, Y. Giraud-Héraud, E. Giusarma, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Ho-vest, Z. Huang, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Kesitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, J. P. Leahy, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Lig- guori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marchini, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, J.-B. Melin, L. Mendes, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Pa- ladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J.-L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, B. Rouillé d'Orfeu- il, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, N. Said, V. Salvatelli, L. Sal- vati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, L. D. Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sudiwala, R. Su- nyaev, D. Sutton, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, M. Türlér, G. Umama, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wan- delt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2015 results. XIII. Cosmological parameters. *Astronomy and Astrophysics*, 594 :A13, Sept. 2016. ISSN 0004-6361. doi : 10.1051/0004-6361/201525830. URL <http://adsabs.harvard.edu/abs/2016A%26A...594A..13P>.
- D. Pomarède, Y. Hoffman, H. M. Courtois, and R. B. Tully. The Cosmic V-Web. *The Astrophysical Journal*, 845 :55, Aug. 2017. ISSN 0004-637X. doi : 10.3847/1538-4357/aa7f78. URL <http://adsabs.harvard.edu/abs/2017ApJ...845...55P>.



- S. C. Porter and S. Raychaudhury. The Pisces-Cetus supercluster : a remarkable filament of galaxies in the 2df Galaxy Redshift and Sloan Digital Sky surveys. *Monthly Notices of the Royal Astronomical Society*, 364 :1387–1396, Dec. 2005. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2005.09688.x. URL <http://adsabs.harvard.edu/abs/2005MNRAS.364.1387P>.
- A. Reisenegger, H. Quintana, E. R. Carrasco, and J. Maze. The Shapley Supercluster. III. Collapse Dynamics and Mass of the Central Concentration. *The Astronomical Journal*, 120 : 523–532, Aug. 2000. ISSN 0004-6256. doi : 10.1086/301477. URL <http://adsabs.harvard.edu/abs/2000AJ....120..523R>.
- A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko, B. E. Tucker, M. J. Reid, D. O. Jones, J. M. Silverman, R. Chornock, P. Challis, W. Yuan, P. J. Brown, and R. J. Foley. A 2.4% Determination of the Local Value of the Hubble Constant. *The Astrophysical Journal*, 826 :56, July 2016. ISSN 0004-637X. doi : 10.3847/0004-637X/826/1/56. URL <http://adsabs.harvard.edu/abs/2016ApJ...826...56R>.
- A. G. Riess, S. Casertano, W. Yuan, L. Macri, J. Anderson, J. W. MacKenty, J. B. Bowers, K. I. Clubb, A. V. Filippenko, D. O. Jones, and B. E. Tucker. New Parallaxes of Galactic Cepheids from Spatially Scanning the Hubble Space Telescope : Implications for the Hubble Constant. *The Astrophysical Journal*, 855 :136, Mar. 2018. ISSN 0004-637X. doi : 10.3847/1538-4357/aaadb7. URL <http://adsabs.harvard.edu/abs/2018ApJ...855..136R>.
- M. Rigault, V. Brinnel, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, K. Barbary, S. Bongard, K. Boone, C. Buton, M. Childress, N. Chotard, Y. Copin, S. Dixon, P. Fagrelus, U. Feindt, D. Fouchez, E. Gangler, B. Hayden, W. Hillebrandt, D. A. Howell, A. Kim, M. Kowalski, D. Kuesters, P.-F. Leget, S. Lombardo, Q. Lin, J. Nordin, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, K. Runge, D. Rubin, C. Saunders, G. Smadja, C. Sofiatti, N. Suzuki, S. Taubenberger, C. Tao, and R. C. Thomas. Strong Dependence of Type Ia Supernova Standardization on the Local Specific Star Formation Rate. *ArXiv e-prints*, 1806 :arXiv :1806.03849, June 2018. URL <http://adsabs.harvard.edu/abs/2018arXiv180603849R>.
- M. S. Roberts and A. H. Rots. Comparison of Rotation Curves of Different Galaxy Types. *Astronomy and Astrophysics*, 26 :483–485, Aug. 1973. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1973A%26A....26..483R>.
- V. C. Rubin and W. K. Ford. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *The Astrophysical Journal*, 159 :379, Feb. 1970. ISSN 0004-637X. doi : 10.1086/150317. URL <http://adsabs.harvard.edu/abs/1970ApJ...159..379R>.
- V. C. Rubin, W. K. Ford, and N. Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *The Astrophysical Journal*, 238 :471–487, June 1980. ISSN 0004-637X. doi : 10.1086/158003. URL <http://adsabs.harvard.edu/abs/1980ApJ...238..471R>.
- A. Sandage, B. Reindl, and G. A. Tammann. The Linearity of the Cosmic Expansion Field from 300 to 30,000 km s<sup>-1</sup> and the Bulk Motion of the Local Supercluster with Respect to

- the Cosmic Microwave Background. *The Astrophysical Journal*, 714 :1441–1459, May 2010. ISSN 0004-637X. doi : 10.1088/0004-637X/714/2/1441. URL <http://adsabs.harvard.edu/abs/2010ApJ...714.1441S>.
- P. Schechter. An analytic expression for the luminosity function for galaxies. *The Astrophysical Journal*, 203 :297–306, Jan. 1976. ISSN 0004-637X. doi : 10.1086/154079. URL <http://adsabs.harvard.edu/abs/1976ApJ...203..297S>.
- D. J. Schlegel, D. P. Finkbeiner, and M. Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *The Astrophysical Journal*, 500 :525–553, June 1998. ISSN 0004-637X. doi : 10.1086/305772. URL <http://adsabs.harvard.edu/abs/1998ApJ...500..525S>.
- R. Scoccimarro, L. Hui, M. Manera, and K. C. Chan. Large-scale bias and efficient generation of initial conditions for nonlocal primordial non-Gaussianity. *Physical Review D*, 85 :083002, Apr. 2012. ISSN 1550-7998. doi : 10.1103/PhysRevD.85.083002. URL <http://adsabs.harvard.edu/abs/2012PhRvD...85h3002S>.
- M. I. Scrimgeour, T. M. Davis, C. Blake, L. Staveley-Smith, C. Magoulas, C. M. Springob, F. Beutler, M. Colless, A. Johnson, D. H. Jones, J. Koda, J. R. Lucey, Y.-Z. Ma, J. Mould, and G. B. Poole. The 6df Galaxy Survey : bulk flows on 50-70 h-1 Mpc scales. *Monthly Notices of the Royal Astronomical Society*, 455 :386–401, Jan. 2016. ISSN 0035-8711. doi : 10.1093/mnras/stv2146. URL <http://adsabs.harvard.edu/abs/2016MNRAS.455..386S>.
- S. Shen, C. Wang, R. Chang, Z. Shao, J. Hou, and C. Shu. The Morphological-Dependent Tully-Fisher Relation of Spiral Galaxies. *The Astrophysical Journal*, 705 :1496–1502, Nov. 2009. ISSN 0004-637X. doi : 10.1088/0004-637X/705/2/1496. URL <http://adsabs.harvard.edu/abs/2009ApJ...705.1496S>.
- J. G. Sorce. Minimization of biases in galaxy peculiar velocity catalogues. *Monthly Notices of the Royal Astronomical Society*, 450 :2644–2657, July 2015. ISSN 0035-8711. doi : 10.1093/mnras/stv760. URL <http://adsabs.harvard.edu/abs/2015MNRAS.450.2644S>.
- J. G. Sorce, H. M. Courtois, R. B. Tully, M. Seibert, V. Scowcroft, W. L. Freedman, B. F. Madore, S. E. Persson, A. Monson, and J. Rigby. Calibration of the Mid-infrared Tully-Fisher Relation. *The Astrophysical Journal*, 765 :94, Mar. 2013. ISSN 0004-637X. doi : 10.1088/0004-637X/765/2/94. URL <http://adsabs.harvard.edu/abs/2013ApJ...765...94S>.
- J. G. Sorce, R. B. Tully, H. M. Courtois, T. H. Jarrett, J. D. Neill, and E. J. Shaya. From Spitzer Galaxy photometry to Tully-Fisher distances. *Monthly Notices of the Royal Astronomical Society*, 444 :527–541, Oct. 2014. ISSN 0035-8711. doi : 10.1093/mnras/stu1450. URL <http://adsabs.harvard.edu/abs/2014MNRAS.444..527S>.
- J. G. Sorce, S. Gottlöber, G. Yepes, Y. Hoffman, H. M. Courtois, M. Steinmetz, R. B. Tully, D. Pomarède, and E. Carlesi. Cosmicflows Constrained Local UniversE Simulations. *Monthly Notices of the Royal Astronomical Society*, 455 :2078–2090, Jan. 2016. ISSN 0035-8711. doi : 10.1093/mnras/stv2407. URL <http://adsabs.harvard.edu/abs/2016MNRAS.455.2078S>.



- C. M. Springob, K. L. Masters, M. P. Haynes, R. Giovanelli, and C. Marinoni. Erratum : "SFI++ II : A New I-Band Tully-Fisher Catalog, Derivation of Peculiar Velocities and Data Set Properties" (2007, ApJS, 172, 599). *The Astrophysical Journal Supplement Series*, 182 :474–475, May 2009. ISSN 0067-0049. doi : 10.1088/0067-0049/182/1/474. URL <http://adsabs.harvard.edu/abs/2009ApJS..182..474S>.
- C. M. Springob, T. Hong, L. Staveley-Smith, K. L. Masters, L. M. Macri, B. S. Koribalski, D. H. Jones, T. H. Jarrett, C. Magoulas, and P. Erdoğdu. 2mtf - V. Cosmography,  $\beta$ , and the residual bulk flow. *Monthly Notices of the Royal Astronomical Society*, 456 :1886–1900, Feb. 2016. ISSN 0035-8711. doi : 10.1093/mnras/stv2648. URL <http://adsabs.harvard.edu/abs/2016MNRAS.456.1886S>.
- M. Steinmetz and J. F. Navarro. The Cosmological Origin of the Tully-Fisher Relation. *The Astrophysical Journal*, 513 :555–560, Mar. 1999. ISSN 0004-637X. doi : 10.1086/306904. URL <http://adsabs.harvard.edu/abs/1999ApJ...513..555S>.
- I. M. Stewart, S.-L. Blyth, and W. J. G. de Blok. A Simple Model for Global HI Profiles of Galaxies. *Astronomy & Astrophysics*, 567 :A61, July 2014. ISSN 0004-6361, 1432-0746. doi : 10.1051/0004-6361/201423602. URL <http://arxiv.org/abs/1405.1838>. arXiv : 1405.1838.
- C. Stoughton, R. H. Lupton, M. Bernardi, M. R. Blanton, S. Burles, F. J. Castander, A. J. Connolly, D. J. Eisenstein, J. A. Frieman, G. S. Hennessy, R. B. Hindsley, Ž. Ivezić, S. Kent, P. Z. Kunszt, B. C. Lee, A. Meiksin, J. A. Munn, H. J. Newberg, R. C. Nichol, T. Nicinski, J. R. Pier, G. T. Richards, M. W. Richmond, D. J. Schlegel, J. A. Smith, M. A. Strauss, M. SubbaRao, A. S. Szalay, A. R. Thakar, D. L. Tucker, D. E. Vanden Berk, B. Yanny, J. K. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, M. Bartelmann, S. Bastian, A. Bauer, E. Berman, H. Böhringer, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, L. Carey, M. A. Carr, B. Chen, D. Christian, P. L. Colestock, J. H. Crocker, I. Csabai, P. C. Czarapata, J. Dalcanton, A. F. Davidsen, J. E. Davis, W. Dehnen, S. Dodelson, M. Doi, T. Dombeck, M. Donahue, N. Ellman, B. R. Elms, M. L. Evans, L. Eyer, X. Fan, G. R. Federwitz, S. Friedman, M. Fukugita, R. Gal, B. Gillespie, K. Glazebrook, J. Gray, E. K. Grebel, B. Greenawalt, G. Greene, J. E. Gunn, E. de Haas, Z. Haiman, M. Haldeman, P. B. Hall, M. Hamabe, B. Hansen, F. H. Harris, H. Harris, M. Harvanek, S. L. Hawley, J. J. E. Hayes, T. M. Heckman, A. Helmi, A. Henden, C. J. Hogan, D. W. Hogg, D. J. Holmgren, J. Holtzman, C.-H. Huang, C. Hull, S.-I. Ichikawa, T. Ichikawa, D. E. Johnston, G. Kauffmann, R. S. J. Kim, T. Kimball, E. Kinney, M. Klaene, S. J. Kleinman, A. Klypin, G. R. Knapp, J. Korienek, J. Krolik, R. G. Kron, J. Krzesiński, D. Q. Lamb, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, B. McLean, K. Menou, A. Merelli, H. J. Mo, D. G. Monet, O. Nakamura, V. K. Narayanan, T. Nash, E. H. Neilsen, Jr., P. R. Newman, A. Nitta, M. Odenkirchen, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. S. Peterson, D. Petravick, A. Pope, R. Pordes, M. Postman, A. Prosapio, T. R. Quinn, R. Rechenmacher, C. H. Rivetta, H.-W. Rix, C. M. Rockosi, R. Rosner, K. Ruthmansdorfer,

- D. Sandford, D. P. Schneider, R. Scranton, M. Sekiguchi, G. Sergey, R. Sheth, K. Shimasaku, S. Smee, S. A. Snedden, A. Stebbins, C. Stubbs, I. Szapudi, P. Szkody, G. P. Szokoly, S. Tabachnik, Z. Tsvetanov, A. Uomoto, M. S. Vogeley, W. Voges, P. Waddell, R. Walterbos, S.-i. Wang, M. Watanabe, D. H. Weinberg, R. L. White, S. D. M. White, B. Wilhite, D. Wolfe, N. Yasuda, D. G. York, I. Zehavi, and W. Zheng. Sloan Digital Sky Survey : Early Data Release. *The Astronomical Journal*, 123 :485–548, Jan. 2002. ISSN 0004-6256. doi : 10.1086/324741. URL <http://adsabs.harvard.edu/abs/2002AJ...123..485S>.
- M. A. Strauss and J. A. Willick. The density and peculiar velocity fields of nearby galaxies. *Physics Reports*, 261 :271–431, 1995. ISSN 0370-1573. doi : 10.1016/0370-1573(95)00013-7. URL <http://adsabs.harvard.edu/abs/1995PhR...261..271S>.
- E. Tempel, E. Tago, and L. J. Liivamägi. Groups and clusters of galaxies in the SDSS DR8. Value-added catalogues. *Astronomy and Astrophysics*, 540 :A106, Apr. 2012. ISSN 0004-6361. doi : 10.1051/0004-6361/201118687. URL <http://adsabs.harvard.edu/abs/2012A%26A...540A.106T>.
- J. L. Tonry and M. Davis. Velocity Dispersions of Elliptical and s0 Galaxies - Part Two - Infall of the Local Group to Virgo. *The Astrophysical Journal*, 246 :680, June 1981. ISSN 0004-637X. doi : 10.1086/158965. URL <http://adsabs.harvard.edu/abs/1981ApJ...246..680T>.
- J. L. Tonry, J. P. Blakeslee, E. A. Ajhar, and A. Dressler. The Surface Brightness Fluctuation Survey of Galaxy Distances. II. Local and Large-Scale Flows. *The Astrophysical Journal*, 530 :625–651, Feb. 2000. ISSN 0004-637X. doi : 10.1086/308409. URL <http://adsabs.harvard.edu/abs/2000ApJ...530..625T>.
- R. B. Tully. Galaxy Groups. *The Astronomical Journal*, 149 :54, Feb. 2015. ISSN 0004-6256. doi : 10.1088/0004-6256/149/2/54. URL <http://adsabs.harvard.edu/abs/2015AJ...149...54T>.
- R. B. Tully and H. M. Courtois. Cosmicflows-2 : I-band Luminosity-H I Linewidth Calibration. *The Astrophysical Journal*, 749 :78, Apr. 2012. ISSN 0004-637X. doi : 10.1088/0004-637X/749/1/78. URL <http://adsabs.harvard.edu/abs/2012ApJ...749...78T>.
- R. B. Tully and J. R. Fisher. A new method of determining distances to galaxies. *Astronomy and Astrophysics*, 54 :661–673, Feb. 1977. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1977A%26A...54..661T>.
- R. B. Tully, M. J. Pierce, J.-S. Huang, W. Saunders, M. A. W. Verheijen, and P. L. Witchalls. Global Extinction in Spiral Galaxies. *The Astronomical Journal*, 115 :2264–2272, June 1998. ISSN 0004-6256. doi : 10.1086/300379. URL <http://adsabs.harvard.edu/abs/1998AJ...115.2264T>.
- R. B. Tully, E. J. Shaya, I. D. Karachentsev, H. M. Courtois, D. D. Kocevski, L. Rizzi, and A. Peel. Our Peculiar Motion Away from the Local Void. *The Astrophysical Journal*, 676 : 184–205, Mar. 2008. ISSN 0004-637X. doi : 10.1086/527428. URL <http://adsabs.harvard.edu/abs/2008ApJ...676..184T>.

- R. B. Tully, H. M. Courtois, A. E. Dolphin, J. R. Fisher, P. Héraudeau, B. A. Jacobs, I. D. Karachentsev, D. Makarov, L. Makarova, S. Mitronova, L. Rizzi, E. J. Shaya, J. G. Sorce, and P.-F. Wu. Cosmicflows-2 : The Data. *The Astronomical Journal*, 146 :86, Oct. 2013. ISSN 0004-6256. doi : 10.1088/0004-6256/146/4/86. URL <http://adsabs.harvard.edu/abs/2013AJ...146...86T>.
- R. B. Tully, H. Courtois, Y. Hoffman, and D. Pomarède. The Laniakea supercluster of galaxies. *Nature*, 513 :71–73, Sept. 2014. ISSN 0028-0836. doi : 10.1038/nature13674. URL <http://adsabs.harvard.edu/abs/2014Natur.513...71T>.
- R. B. Tully, H. M. Courtois, and J. G. Sorce. Cosmicflows-3. *The Astronomical Journal*, 152(2) :50, Aug. 2016. ISSN 1538-3881. doi : 10.3847/0004-6256/152/2/50. URL <http://arxiv.org/abs/1605.01765>. arXiv : 1605.01765.
- R. van de Weygaert and E. Bertschinger. Peak and gravity constraints in Gaussian primordial density fields : An application of the Hoffman-Ribak method. *Monthly Notices of the Royal Astronomical Society*, 281 :84, July 1996. ISSN 0035-8711. doi : 10.1093/mnras/281.1.84. URL <http://adsabs.harvard.edu/abs/1996MNRAS.281...84V>.
- P. Vielva, E. Martínez-González, R. B. Barreiro, J. L. Sanz, and L. Cayón. Detection of Non-Gaussianity in the Wilkinson Microwave Anisotropy Probe First-Year Data Using Spherical Wavelets. *The Astrophysical Journal*, 609 :22–34, July 2004. ISSN 0004-637X. doi : 10.1086/421007. URL <http://adsabs.harvard.edu/abs/2004ApJ...609...22V>.
- N. Vittorio and R. Juszkiewicz. Peculiar Velocity and Gravity as Cosmological Probes. In *Large Scale Structure and Motions in the Universe*, Astrophysics and Space Science Library, pages 241–258. Springer, Dordrecht, 1989. ISBN 978-94-010-6895-6 978-94-009-0903-8. doi : 10.1007/978-94-009-0903-8\_18. URL [https://link.springer.com/chapter/10.1007/978-94-009-0903-8\\_18](https://link.springer.com/chapter/10.1007/978-94-009-0903-8_18).
- R. Watkins and H. A. Feldman. An unbiased estimator of peculiar velocity with Gaussian distributed errors for precision cosmology. *Monthly Notices of the Royal Astronomical Society*, 450 :1868–1873, June 2015. ISSN 0035-8711. doi : 10.1093/mnras/stv651. URL <http://adsabs.harvard.edu/abs/2015MNRAS.450.1868W>.
- D. H. Weinberg, R. Davé, N. Katz, and L. Hernquist. Galaxy Clustering and Galaxy Bias in a  $\lambda$ CDM Universe. *The Astrophysical Journal*, 601 :1–21, Jan. 2004. ISSN 0004-637X. doi : 10.1086/380481. URL <http://adsabs.harvard.edu/abs/2004ApJ...601...1W>.
- J. A. Willick and M. A. Strauss. Maximum Likelihood Comparison of Tully-Fisher and Redshift Data. II. Results from an Expanded Sample. *The Astrophysical Journal*, 507 :64–83, Nov. 1998. ISSN 0004-637X. doi : 10.1086/306314. URL <http://adsabs.harvard.edu/abs/1998ApJ...507...64W>.
- W. M. Wood-Vasey, G. Aldering, B. C. Lee, S. Loken, P. Nugent, S. Perlmutter, J. Siegrist, L. Wang, P. Antilogus, P. Astier, D. Hardin, R. Pain, Y. Copin, G. SMadja, E. Gangler,

- A. Castera, G. Adam, R. Bacon, J.-P. Lemonnier, A. Pecontal, E. Pecontal, and R. Kessler. The Nearby Supernova Factory. *New Astronomy Reviews*, 48(7-8) :637–640, May 2004. ISSN 13876473. doi : 10.1016/j.newar.2003.12.056. URL <http://arxiv.org/abs/astro-ph/0401513>. arXiv : astro-ph/0401513.
- G. Yepes, L. A. Martínez-Vaquero, S. Gottlöber, and Y. Hoffman. The CLUES project : Constrained Local Universe Simulations. volume 1178, pages 64–75, Oct. 2009. doi : 10.1063/1.3264558. URL <http://adsabs.harvard.edu/abs/2009AIPC.1178...64Y>.
- D. Zaritsky, H. Courtois, J.-C. Muñoz-Mateos, J. Sorce, S. Erroz-Ferrer, S. Comerón, D. A. Gadotti, A. Gil de Paz, J. L. Hinz, E. Laurikainen, T. Kim, J. Laine, K. Menéndez-Delmestre, T. Mizusawa, M. W. Regan, H. Salo, M. Seibert, K. Sheth, E. Athanassoula, A. Bosma, M. Cisternas, L. C. Ho, and B. Holwerda. The Baryonic Tully-Fisher Relationship for S4g Galaxies and the "Condensed" Baryon Fraction of Galaxies. *The Astronomical Journal*, 147 :134, June 2014. ISSN 0004-6256. doi : 10.1088/0004-6256/147/6/134. URL <http://adsabs.harvard.edu/abs/2014AJ....147..134Z>.
- S. Zaroubi, Y. Hoffman, K. B. Fisher, and O. Lahav. Wiener Reconstruction of the Large-Scale Structure. *The Astrophysical Journal*, 449 :446, Aug. 1995. ISSN 0004-637X. doi : 10.1086/176070. URL <http://adsabs.harvard.edu/abs/1995ApJ...449..446Z>.
- F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6 : 110–127, 1933. ISSN 0018-0238. URL <http://adsabs.harvard.edu/abs/1933AcHPh...6..110Z>.



# Remerciements

Je remercie d'ores et déjà les futurs thésards qui liront autre chose que les remerciements. Je remercie d'ailleurs Alex, T. Davis, Yannick et Hélène d'avoir déjà lu (et corrigé) l'intégralité de cette thèse. Pour le reste, merci à...

Hélène pour sa confiance. Pour m'avoir soutenu quand je prenais le chemin le plus tortueux. Pour m'avoir jeté dans le monde scientifique en supposant que je saurais m'y retrouver.

Yannick pour m'avoir tant appris sur la physique. Et pour nos discussions scientifiques sans fin. Est-il encore nécessaire de rappeler qu'il est un (enseignant-)chercheur exceptionnel ? Pour sa bonne humeur, aussi, parfois, mais pas tout le temps. Merci aussi de donner tant de vie au groupe. Et de parfois avoir pris le temps de préciser, après avoir pointé tous les défauts possibles d'un travail, qu'il ne faut pas « le prendre personnellement ».

Mickaël pour sa solidarité, pour ses initiatives et aussi pour sa confiance. Pour sa force de cohésion. De partager son expertise des plots.

Gérard pour m'avoir montré ce que pouvait être l'engagement d'un scientifique.

Yehuda pour sa pédagogie, sa gentillesse, pour son entêtement aussi. Pour son impact majeur dans ma thèse. Pour m'avoir expliqué la nécessité d'être pragmatique, et pour avoir compris que c'était hors de question pour moi. Enfin, pour m'avoir prévenu dès le début que « the only thing to know is that science is done by people ».

Guilhem pour m'avoir mis sur le chemin de la modélisation Bayésienne.

Mehdi pour son entrain, sa relativisation à toute épreuve et ses « yolo », « on est pas sur le même  $\lambda$  » et autres « PLS ». Pour le «  $X \Downarrow L B x L \searrow$  » et pour me faire participer à son futur papier « Slitless Spectroscopy from Bayesian Models : The SSBM method ». Pour m'avoir si souvent parlé alors que je n'écoutais pas.

Lola pour son soutien quotidien, ses commérages, son cruciverbisme de haut vol et surtout pour nos discussions à rallonge. Pour toutes nos pauses café. Pour sa perception des âmes humaines. Pour qu'il reste encore des gens *quasiment* normaux dans les laboratoires de physique. Pour ses grands sourires du fond du couloir. Pour avoir pu partager mes moments de déprime

doctorale.

Clément aussi, pour son soutien quotidien. Pour son extrême gentillesse. Pour tous ses conseils politiques, scientifiques, numériques, personnels au cours de ces trois années. Pour avoir éclairé ma vision du couple. Pour les nombreuses sessions de grimpe, pour m'avoir répété que tout n'est pas compétition. Pour ses trop nombreux « c'est ce qu'elles disent toutes ».

Mes amis Lyonnais. Géraldine pour son dynamisme, son écoute et pour m'avoir forcé à rester positif. Pour son combat contre les explications fumeuses dont je suis un grand adepte. Antoine pour sa fraîcheur dans la lutte pour l'indépendance d'esprit, son entrain à proposer des concerts ou des manifs sans que je n'y aille jamais, nos nombreuses discussions sur le problème des relations amoureuses. Thibault pour son extrême gentillesse (oui, il y en a deux comme ça). Pour m'avoir montré ce qu'est la pugnacité. Pour les pizzas-bières de la pinte douce. Vladimir et Irénée pour leur accueil à Lyon et les bons moments passés à la coloc.

Mes amis Cachanais. Brigitte pour sa bienveillance et pour sa patience. Tim, Mathias, Antoine, Adrien, Severin, Amaudric pour ne pas trop m'avoir tenu rigueur de mon manque de communication pendant ces quelques années.

Mes amis parisiens Aurélien, Eugénie, Lucas, Clara et Anne-So pour leur soutien au long cours.

Raphaël pour son aide précieuse, son soutien, ses relectures, ses discours sur la science, les arts, la politique, pour sa radicalité. Pour avoir été là tout ce temps. Pour avoir joué support pendant trois ans.

Mes parents et ma soeur pour avoir toujours été là quand il fallait, s'il fallait, où il fallait. J'ai le sentiment de pouvoir compter sur eux à chaque instant, et je les en remercie pour ça.

Corto, parce que j'espère qu'il prendra plaisir à voir son nom dans une thèse de cosmologie plus tard.