



HAL
open science

Etude de la distribution et de la cinématique des galaxies dans l'univers

Stéphane Rauzy

► **To cite this version:**

Stéphane Rauzy. Etude de la distribution et de la cinématique des galaxies dans l'univers. Cosmologie et astrophysique extra-galactique [astro-ph.CO]. Université Aix-Marseille II, 1993. Français. NNT : . tel-01705311

HAL Id: tel-01705311

<https://hal.science/tel-01705311>

Submitted on 12 Feb 2018

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.

THESE

PRESENTEE PAR Stéphane RAUZY

pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE D'AIX-MARSEILLE II
FACULTE DES SCIENCES DE LUMINY

Spécialité : Physique Théorique

"ETUDE DE LA DISTRIBUTION ET DE LA CINEMATIQUE
DES GALAXIES DANS L'UNIVERS"

Soutenue le 14 Avril 1993

Devant le jury composé de :

Albert Bijaoui	(rapporteur)
Avishai Dekel	(rapporteur)
Lucienne Gouguenheim	(président du jury)
Alex Grossmann	
Richard H. Henriksen	
Marc Lachièze-Rey	(directeur de thèse)
Pierre Mery	
Roland Triay	



REMERCIEMENTS :

Je désire remercier Marc Lachièze-Rey, mon directeur de thèse, ainsi que Dick Henriksen et Roland Triay avec qui j'ai collaboré.

Je remercie également Albert Bijaoui, Avishai Dekel, Lucienne Gougenheim, Alex Grossmann et Pierre Mery qui ont accepté d'être membre de mon jury.

Je remercie enfin l'hospitalité du Centre de Physique Théorique de Luminy, du Service d'Astrophysique du CEN Saclay et du département de Physique de Queen's University à Kingston, Ontario.



SOMMAIRE :

Partie A: Présentation du sujet	5
Partie B: Le champ des vitesses radiales propres	31
Partie C: Le tenseur de corrélation des vitesses	89
Partie D: Le "Grand Attracteur"	131
Partie E: Analyse par ondelettes du champ des vitesses	175



INTRODUCTION :

Cette thèse est composée de 5 parties. Chacune de ces parties est en principe indépendante, même si des références croisées sont parfois inévitables. Cette étude est le fruit d'un travail commun impliquant Marc Lachièze-Rey et Roland Triay pour la partie B, et Dick Henriksen et Marc Lachièze-Rey pour les parties C, D et E.

Dans la partie A, je replace le sujet de cette thèse dans son contexte. Si l'univers est réellement homogène et isotrope à très grande échelle, sa dynamique est complètement caractérisée par les trois paramètres d'univers H_0 , q_0 et Ω_0 . La hiérarchie de structures actuellement observée à "petite" échelle (typiquement inférieure à 50 Mpc) est alors étudiée dans l'approximation newtonienne comme une perturbation locale de la densité moyenne de matière présente dans l'univers. La dynamique de ces fluctuations peut être en partie étudiée par l'analyse de la distribution des galaxies dans l'espace, et complémentirement par l'étude du champ des vitesses propres associé aux perturbations locales de la densité totale de matière. L'étude cinématique des grandes structures de l'univers fait l'objet de cette thèse.

Dans la partie B, nous montrons comment déterminer le champ des vitesses propres de notre entourage à partir des données observationnelles disponibles. On n'a accès en pratique qu'à la composante radiale de la vitesse propre d'une galaxie, si on est capable d'estimer sa distance ainsi que la constante de Hubble H_0 . On ne possède en fait que des indicateurs statistiques de la distance d'une galaxie lointaine. Ces estimateurs sont souvent soumis à des biais difficiles à corriger à l'aide des techniques usuelles. Nous introduisons dans cette partie une formalisation statistique adéquate qui nous permet de traiter ces problèmes de biais qui interviennent également dans l'évaluation de H_0 . La valeur estimée de la vitesse radiale propre d'une galaxie, après un traitement correct des biais, comporte encore une incertitude intrinsèque d'environ 20%.

L'étude du tenseur de corrélation des vitesses est abordée dans la partie C. L'évaluation des composantes de ce tenseur permet en théorie d'accéder au spectre des fluctuations de la densité de matière totale présente dans l'univers. La comparaison de ce spectre avec

la fonction de corrélation spatiale de la matière lumineuse (les galaxies) dévoile alors le comportement de la masse non visible. L'estimation des composantes du tenseur à partir du champ des vitesses radiales seulement est impossible si on ne postule pas que le champ des vitesses observé est statistiquement homogène et isotrope, c'est-à-dire que l'échantillon étudié est représentatif de la cinématique de l'univers dans sa globalité. Nous montrons dans la partie D que les échantillons disponibles ne vérifient pas cette hypothèse. Pour cette raison, l'évaluation du tenseur de corrélation des vitesses est actuellement impossible. De plus, la présence d'incertitudes sur l'estimation du champ des vitesses radiales propres interdit en pratique l'obtention des composantes de ce tenseur.

La partie D est consacrée au "Grand Attracteur", un flot de vitesses convergent, cohérent à une échelle d'au moins 60 Mpc, qui domine la cinématique des catalogues de redshift-distance actuellement disponibles. Les difficultés rencontrées par les théories classiques de formation de structures afin d'expliquer l'existence d'un tel flot ont poussées certains auteurs à douter de la fiabilité de l'étude statistique de Lynden-Bell et *al.* (1988) ayant mis en évidence le "Grand Attracteur". Nous proposons dans cette partie des tests géométriques simples basés sur les propriétés d'invariance du champ des vitesses qui nous permettent de détecter la présence d'un flot convergent, ceci indépendamment du repère de référence dans lequel est effectuée l'analyse. Si le catalogue de Faber et *al.* (1989) est fiable, nos résultats confirment l'existence du "Grand Attracteur" et indiquent de plus que cet attracteur ainsi que toutes les galaxies de l'échantillon possèdent un mouvement d'ensemble par rapport au repère absolu défini par le fond diffus cosmologique.

Dans la partie E, nous proposons une méthode alternative à la méthode POTENT proposée par E. Bertschinger et A. Dekel (1989) qui permet de reconstruire totalement un champ de vitesses si il est irrotationnel (c'est à dire dérivant d'un potentiel) à partir de sa composante radiale seulement. Notre méthode est basée sur l'utilisation de la transformée en ondelettes et hérite à ce titre des propriétés puissantes dont jouie cette transformation. Notre procédure de reconstruction est ainsi linéaire (vis-à-vis du champ des vitesses), ce qui est une propriété appréciable lorsqu'il s'agit de traiter le champ de vitesses radiales propres observé, bruité par les incertitudes inhérentes à sa détermination. De plus l'utilisation de la transformée en ondelettes nous offre un contrôle naturel du lissage des données intervenant obligatoirement lors de la reconstruction du potentiel des vitesses ou d'autres quantités. L'application de notre méthode aux échantillons actuellement disponibles est en cours.

PARTIE A :

PRESENTATION DU SUJET

SOMMAIRE :

I. Les modèles d'univers :

I.1. Le principe cosmologique :

I.2. Le contenu énergétique de l'univers :

I.3. La validité du principe cosmologique :

I.3.1. Le décalage des raies spectrales :

I.3.2. Le rayonnement du fond diffus cosmologique :

I.4. Les valeurs des paramètres d'univers H_0 , q_0 et Ω_0 :

II. La dynamique locale de l'univers :

II.1. L'approximation newtonienne :

II.2. Le champ des vitesses propres :

II.2.1. Le repère absolu des vitesses :

II.2.2. La composante radiale du champ des vitesses propres :

III. L'étude cinématique des grandes structures :

III.1. Le comportement de la matière non visible :

III.2. Une information à plus grande échelle :

III.3. Une connaissance incomplète du champ des vitesses :

L'étude de l'univers dans son ensemble apparaît a priori très complexe. Toutefois si on se place à suffisamment grande échelle, on peut modéliser son contenu énergétique par un gaz de galaxies, de pression actuellement négligeable, en interaction auto-gravitationnelle. De plus, on soupçonne que ce fluide cosmique est distribué de façon homogène et isotrope à très grande échelle (c'est le principe cosmologique).

La théorie de la relativité générale nous permet alors de déduire la métrique d'espace-temps de notre univers. Il existe un système de coordonnées privilégié (le repère comobile) pour lequel le temps et l'espace sont séparables. Les galaxies au repos dans ce repère tracent les lignes d'univers et leurs distances propres évoluent au cours du temps comme le facteur d'expansion $a(t)$. La dynamique de l'univers à grande échelle est complètement caractérisée par cette fonction $a(t)$ qui dépend seulement des 3 paramètres H_0 , q_0 et Ω_0 . Un modèle d'univers est alors spécifié par les valeurs de la constante de Hubble H_0 , du paramètre de décélération q_0 et de la densité réduite de l'univers Ω_0 .

Mais localement, les galaxies ont tendance à se regrouper pour former une hiérarchie de structures (amas et super-amas de galaxies). On attribue l'origine de ces grandes structures à la présence des fluctuations initiales de la densité du fluide cosmique qui ont évoluées sous l'influence de la gravitation. Parce que ces perturbations sont de faible intensité et d'échelle caractéristique relativement modeste comparée à la taille actuelle de l'horizon, la dynamique globale de l'univers n'est pas modifiée. De plus en se plaçant dans le repère comobile, on peut définir un champ d'accélération ainsi qu'un champ des vitesses propres à ces fluctuations dont l'étude est possible dans le cadre de la mécanique purement newtonienne.

La dynamique des grandes structures actuellement observées peut être déduite de l'analyse de la distribution spatiale des galaxies et de façon complémentaire, de l'étude du champ des vitesses propres de ces galaxies qui fait l'objet de cette thèse. L'étude cinématique des fluctuations locales du fluide cosmique permet en principe d'estimer le champ de densité de masse totale à grande échelle. Néanmoins, on a accès en pratique qu'à la composante radiale du champ des vitesses propres, ce qui nous oblige à postuler certaines hypothèses que devra vérifier le champ des vitesses. De plus, la détermination de la vitesse radiale d'une galaxie est un problème complexe et est soumise à des incertitudes statistiques importantes.

I. LES MODELES D'UNIVERS :

Il s'agit de donner une description globale de l'univers, c'est-à-dire d'étudier la dynamique à grande échelle de son contenu énergétique, soumis principalement à l'action des forces de gravitation. Le principe d'équivalence stipule qu'une modification de la géométrie de l'espace-temps peut rendre compte de ces effets gravitationnels. On affecte alors à l'espace-temps une structure riemannienne et la théorie de la relativité générale nous fournit les équations gouvernant sa métrique définie en chaque point.

La modélisation de notre univers nécessite néanmoins la connaissance de son contenu énergétique à une époque donnée. Cette information n'étant pas disponible, on postule alors que le tenseur d'impulsion-énergie (ou la métrique) possède certaines symétries. Le modèle standard repose par exemple sur l'hypothèse qu'il existe une classe privilégiée d'observateurs pour lesquels l'univers apparaît spatialement homogène et isotrope, c'est le principe cosmologique (voir section I.1). La métrique de l'espace-temps adopte alors une forme bien particulière. De plus, si on postule que le contenu de l'univers est constitué d'un gaz de poussières froides, trois paramètres (H_0 , Ω_0 et q_0) suffisent alors à complètement spécifier le modèle décrivant la dynamique de notre univers.

Cette modélisation de notre univers peut apparaître quelque peu simpliste et *ad hoc* (notamment le principe cosmologique). Toutefois ces modèles prédisent des faits réellement observés comme l'existence du fond diffus cosmologique ou la dépendance linéaire (à petite échelle) du décalage des raies spectrales des galaxies en fonction de leurs distances (la loi de Hubble). L'évaluation des valeurs des paramètres d'univers reste néanmoins difficile. La constante de Hubble H_0 est en fait mal connue ($50 \text{ km.s}^{-1}/\text{Mpc} < H_0 < 100 \text{ km.s}^{-1}/\text{Mpc}$) et on ne peut seulement fixer que des limites inférieures ou supérieures pour les valeurs de la densité réduite Ω_0 et du paramètre de décélération q_0 .

I.1. Le principe cosmologique :

L'univers est peuplé d'une multitude de particules élémentaires qui sous l'action des 4 forces fondamentales interagissent pour former des structures de plus en plus complexes. Néanmoins, si on cherche à déterminer les propriétés globales de l'univers (c'est à dire à grande échelle), cette apparente complexité peut être évitée. En effet, au delà d'une certaine échelle spatiale, les galaxies peuplant notre univers apparaissent comme ponctuelles et la modélisation de l'univers est notablement simplifiée. Notre tâche se résume alors à l'étude dynamique d'un gaz cosmique de particules (les galaxies), de pression actuellement négligeable, qui se meut sous l'emprise de la gravitation (l'influence des autres forces étant négligeable à cette échelle).

L'étude de ce gaz cosmique est réalisée dans le cadre de la théorie de la relativité générale. En chaque point de l'espace-temps, le principe d'équivalence nous permet de définir le tenseur métrique $g_{\mu\nu}$ de telle façon que l'on élimine localement les effets de la gravitation. On a alors donné à la géométrie de notre espace-temps une structure riemannienne et le tenseur métrique $g_{\mu\nu}$ dépend des caractéristiques énergétiques du gaz en chaque point (par l'intermédiaire du tenseur d'énergie-impulsion $T_{\mu\nu}$). La dynamique du gaz de galaxies est complètement caractérisée par la structure de la métrique ainsi définie.

Mais les observations ne nous donnent hélas pas accès au contenu énergétique de l'univers et par conséquent à des quantités comme le tenseur d'énergie-impulsion $T_{\mu\nu}$ par exemple. Il est alors nécessaire de poser des hypothèses supplémentaires caractérisant la structure de notre univers. Il est souhaitable que ces hypothèses nous permettent de faire des prédictions sur le comportement de l'univers qui pourront par la suite être confrontées aux observations. Le modèle standard, maintenant largement accepté en cosmologie, postule qu'il existe une classe privilégiée d'observateurs pour lesquels l'univers apparaît spatialement homogène et isotrope à grande échelle. Un volume d'espace de taille suffisamment importante centré sur un tel observateur (nous par exemple) est alors représentatif de l'univers dans sa globalité. On donne en fait une symétrie maximale à l'espace et on contraint le contenu énergétique de l'univers (ou le tenseur d'énergie-impulsion $T_{\mu\nu}$) à adopter cette symétrie bien particulière. Pour chaque observateur, la

mesure invariante propre ds , associée à la métrique de l'espace-temps $g_{\mu\nu}$, prend alors la forme suivante (voir par exemple Weinberg (1971), équation 14.2.1) :

$$ds^2 = c^2 dt^2 - a(t)^2 \left(\frac{dr^2}{1 - k r^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right) \quad (A.1)$$

La fonction $a(t)$ est appelée le facteur d'expansion et la constante k prend les valeurs $-1, 0$ ou $+1$ (avec un bon choix des coordonnées $\mathbf{x} = (r, \theta, \phi)$) selon le type de géométrie que possède l'espace.

Le principe cosmologique implique que tous les observateurs pour lesquels l'équation (A.1) est vérifiée possèdent un temps propre t en commun. On introduit donc de nouveau le concept de temps absolu. Il est possible de séparer le temps et l'espace.

De plus il est facile de montrer que les trajectoires $\mathbf{x}(t)$, telles que les coordonnées (r, θ, ϕ) sont constantes définissent les géodésiques de notre espace-temps (ces points \mathbf{x} sont en chute libre). On peut alors définir le repère comobile (t, \mathbf{x}) , en prenant comme origine un des observateurs privilégiés O . Le contenu énergétique de l'univers est alors au repos par rapport à ce système de coordonnées, le principe cosmologique implique ainsi l'existence d'un repère absolu.

Pour un temps t fixé, on peut étudier les propriétés géométriques de l'espace à 3 dimensions. Cette espace est ouvert, plat ou fermé suivant que la constante k prend les valeurs $-1, 0$ ou $+1$. On peut remarquer que cet espace conserve sa courbure initiale, même si son échelle caractéristique $a(t)$ évolue au cours du temps. Le volume V d'un espace fermé ($k = +1$) est par exemple proportionnel au cube du facteur d'expansion $a(t)$ ($V \propto a(t)^3$). Enfin, la distance propre $d_p(t)$ entre deux points de coordonnées comobiles \mathbf{r}_1 et \mathbf{r}_2 varie au cours du temps. Si (r, θ, ϕ) est la position de \mathbf{r}_2 dans le repère comobile d'origine \mathbf{r}_1 , leur distance propre $d_p(t)$ s'exprime de la façon suivante :

$$d_p(t) = a(t) \int_0^r \frac{dr'}{\sqrt{1 - k r'^2}} \quad (A.2)$$

I.2. Le contenu énergétique de l'univers :

Il s'agit maintenant de spécifier le contenu énergétique de notre univers. Le principe cosmologique contraint en fait le tenseur d'énergie-impulsion à posséder la forme suivante (voir Weinberg (1971), équation 16.2.3) :

$$T_{\mu\nu} = \frac{p}{c^2} g_{\mu\nu} + \left(\frac{p}{c^2} + \rho\right) U_\mu U_\nu - \frac{\Lambda c^2}{8\pi G} g_{\mu\nu} \quad (A.3)$$

La constante c est la vitesse de la lumière, G la constante de la gravitation, U le vecteur quadri-vitesse ($U_t = 1$ et $U_r = U_\theta = U_\phi = 0$, en coordonnées comobiles), Λ la constante cosmologique et $p(t)$ et $\rho(t)$ deux fonctions du temps propre t seulement. Ce tenseur s'identifie alors avec le tenseur d'énergie-impulsion d'un fluide parfait de pression p et de densité ρ . Le principe cosmologique implique que la pression et la densité de ce fluide cosmique ne varient qu'en fonction du temps propre t .

Les équations d'Einstein qui relient l'expression du tenseur d'impulsion-énergie à la métrique d'espace-temps vont nous permettre alors de déterminer le paramètre d'expansion $a(t)$ en fonction du temps propre. Si on néglige la pression du fluide cosmique (ce qui est une approximation honnête pour le gaz de galaxies dans son état actuel), la fonction $a(t)$ vérifie en fait l'équation différentielle suivante :

$$\left(\frac{\dot{a}(t)}{a_0}\right)^2 = H_0^2 \left[1 + \Omega_0 \left(\frac{a_0}{a(t)}\right) - 1\right] \quad (A.4)$$

Le fonction $\dot{a}(t)$ est la dérivée de $a(t)$ par rapport au temps, et les constantes a_0 , H_0 et Ω_0 s'expriment de la façon suivante. Soit t_0 le temps actuel, on définit la constante de Hubble H_0 , la densité réduite Ω_0 et le facteur de décélération q_0 (ces 3 paramètres sont en principe accessible par l'observation) :

$$H_0 = H(t_0) = \frac{\dot{a}(t_0)}{a(t_0)}; \quad q_0 = q(t_0) = + \frac{\ddot{a}(t_0) a(t_0)}{\dot{a}(t_0)^2}; \quad \Omega_0 = \Omega(t_0) = \frac{8\pi G \rho(t_0)}{3 H_0^2} \quad (A.5)$$

Et le facteur d'expansion actuel $a_0 = a(t_0)$ ainsi que la constante de courbure k sont

déduits par les formules suivantes (voir Rindler (1969), équation (9.93)) :

$$a_0 = \left(\frac{k c^2}{H_0^2 (3/2 \Omega_0 + q_0 - 1)} \right)^{1/2} ; \quad k = \text{signe} \left[\left(\frac{3}{2} \Omega_0 + q_0 - 1 \right) \right] \quad (A.6)$$

L'évolution du paramètre d'expansion $a(t)$ est donc fixée par la données des trois constantes H_0 , q_0 et Ω_0 . En fait, si le fluide cosmique vérifie bien les symétries postulées par le principe cosmologique, la mesure de la constante de Hubble H_0 , du paramètre de décélération q_0 et de la densité réduite Ω_0 suffisent à complètement spécifier le modèle décrivant la dynamique à grande échelle de notre univers.

I.3. La validité du principe cosmologique :

On peut à juste titre s'interroger sur le principe cosmologique. L'hypothèse d'homogénéité et d'isotropie de l'espace à grande échelle est certainement satisfaisante (parce que elle nous permet de tirer des conclusions globales sur la dynamique de l'univers dans son ensemble à partir de l'étude locale de sa structure), mais est-elle réaliste ? En d'autres termes, les modèles d'univers que l'on a dérivés en postulant le principe cosmologique prédisent-ils des faits observationnellement vérifiables ?

I.3.1. Le décalage des raies spectrales :

Les télescopes nous permettent de collecter les ondes lumineuses que nous envoient les galaxies. La matière constituant ces galaxies possède les mêmes propriétés que les composants terrestres par exemple. Notamment, les raies d'absorption ou d'émission des différents atomes présents dans les galaxies lointaines possèdent les mêmes longueurs d'onde que sur terre. L'étude du spectre des galaxies nous permet de mesurer ces différentes longueurs d'onde, pour ensuite les comparer avec leurs équivalents terrestres.

En fait, la longueur d'onde observée λ_g pour une raie spectrale donnée d'une galaxie lointaine ne correspond généralement pas avec sa valeur déduite en laboratoire λ_l . On mesure traditionnellement cette différence de longueurs d'onde par la quantité z appelée le décalage des raies spectrales ou "redshift" définie comme suit :

$$z = \frac{\lambda_g - \lambda_l}{\lambda_l} \quad (\text{A.7})$$

Quelle est l'origine de ce décalage des raies spectrales z ? Les modèles d'univers précédemment présentés répondent à cette question.

La lumière qui nous parvient d'une galaxie lointaine doit parcourir la distance qui nous sépare de cette galaxie. Durant ce parcours, l'onde lumineuse est soumise aux variations de l'échelle caractéristique $a(t)$ de la métrique spatiale. Une onde de longueur d'onde λ_l émise au temps t_1 par une galaxie lointaine nous arrive au temps t_0 avec une longueur d'onde λ_g telle que :

$$\frac{\lambda_l}{\lambda_g} = \frac{a(t_1)}{a(t_0)} \quad (\text{A.8})$$

On peut alors exprimer le temps d'émission t_1 en fonction de la distance propre actuellement mesurée $d_p(t_0)$ de la galaxie. Si le redshift z est petit devant 1 (ou $d_p(t_0) \ll a(t_0)$), il est possible de développer le facteur d'expansion $a(t)$ en série afin d'obtenir la relation suivante (voir Weinberg (1971), équation (14.6.8)) :

$$z = \frac{H_0}{c} d_p(t_0) + (1 + 2q_0) \frac{H_0^2}{c^2} d_p(t_0)^2 + \dots \quad (\text{A.9})$$

Pour des petits redshifts, le décalage des raies spectrales z d'une galaxie est proportionnel à sa distance propre actuellement mesurée $d_p(t_0)$, c'est la loi de Hubble. Cette dépendance linéaire entre distance et redshift est effectivement observée. Une des prédictions des modèles d'univers postulant la validité du principe cosmologique est ainsi réalisée. De plus, on estime la valeur de la constante de Hubble H_0 ($H_0 \approx 100 \text{ km.s}^{-1}/\text{Mpc}$). Les raies spectrales des galaxies sont alors toutes décalées vers le rouge, notre univers est actuellement dans une phase d'expansion.

I.3.2. Le rayonnement du fond diffus cosmologique :

Les observations montrent que l'univers est actuellement dans une phase d'expansion. D'après les modèles d'univers présentés dans la section précédente, le facteur d'expansion $a(t)$ n'a alors pas cessé de croître, l'univers était plus concentré dans le passé et la métrique

d'espace-temps de l'univers admet alors une singularité, un temps $t = 0$ (c'est le "big-bang"). De nombreux indices tendent à confirmer ce scénario, soit que l'univers dans le passé était plus chaud et plus condensé qu'actuellement.

Il a alors dû exister une époque à laquelle les atomes qui peuplent l'univers étaient ionisés, la température importante ne permettant pas aux noyaux et aux électrons de s'assembler. Par exemple lorsque la température est supérieure à 3000 K environ, l'hydrogène atomique ne peut pas être formé et les protons et les électrons cohabitent en équilibre thermique avec un rayonnement électromagnétique. Lorsque la température descend en dessous de 3000 K, les atomes d'hydrogène sont fabriqués et se découplent du gaz de photons en équilibre thermique (un rayonnement électromagnétique de corps noir).

Les modèles d'univers reposant sur la validité du principe cosmologique prédisent que l'histoire de l'univers passe par une telle phase. Une fois découplé de la matière, le rayonnement de corps noir se refroidit au rythme imposé par la variation du facteur d'expansion $a(t)$. On doit donc observer actuellement ce rayonnement fossile. De plus, ce gaz de photons doit être réparti uniformément dans l'espace, comme le veut le principe cosmologique.

On observe effectivement ce rayonnement de corps noir à une température de 2.7 K. Ce rayonnement du fond diffus est sans doute d'origine cosmologique et possède une remarquable isotropie. En fait, les photons que l'on détecte maintenant ont parcouru l'espace durant 10 milliards d'années environ, et proviennent donc des régions de l'espace situées sur une sphère d'environ 3000 Mpc de rayon dont nous occupons le centre (c'est la surface de dernière diffusion). L'isotropie détectée jusqu'à cette échelle tend à confirmer observationnellement la validité du principe cosmologique.

I.4. Les valeurs des paramètres d'univers H_0 , q_0 et Ω_0 :

Les données observationnelles tendent à confirmer la validité du principe cosmologique. La dynamique de notre univers à grande échelle est alors complètement caractérisée par le facteur d'échelle $a(t)$ et donc par les valeurs des trois paramètres d'univers H_0 , q_0 et Ω_0 . L'évaluation de la valeur de ces paramètres apparaît alors fondamentale.

L'équation (A.9) nous permet en principe d'accéder aux valeurs de la constante de

Hubble H_0 et du paramètre de décélération q_0 .

On mesure les redshifts z des galaxies et en estimant par exemple leurs distances propres $d_p(t_0)$, il est possible d'évaluer statistiquement H_0 et q_0 . Malheureusement, l'estimation de la distance d'une galaxie est soumise à des incertitudes ainsi qu'à des biais difficiles à corriger (voir partie B). De plus, l'homogénéité de l'univers n'intervient qu'à de très grandes échelles (supérieures à 200 Mpc) ce qui compromet sérieusement l'estimation des paramètres d'univers. Nous verrons par la suite qu'il existe par exemple un champ des vitesses propres qui par effet Doppler-Fizeau contribue à la valeur purement cosmologique du redshift z . En fait, la valeur de la constante de Hubble estimée par ce genre de méthode fluctue d'un facteur 2 suivant les techniques et les catalogues employés ($50 \text{ km.s}^{-1}/\text{Mpc} < H_0 < 100 \text{ km.s}^{-1}/\text{Mpc}$).

La détermination du paramètre de décélération est encore plus difficile, q_0 intervenant au deuxième ordre du développement dans l'équation (A.9). Il est alors plus question d'utiliser des estimateurs de la distance des galaxies. Certaines techniques statistiques permettent néanmoins de déterminer q_0 , au prix d'hypothèses sur les caractéristiques de la population des galaxies étudiées (par exemple sur sa distribution spatiale ou son évolution au cours du temps). Actuellement, nous ne sommes pas en mesure de donner une estimation fiable de q_0 .

La détermination de la densité réduite Ω_0 peut être effectuée en estimant la masse de matière présente dans un volume donné. Mais cette approche pose de nombreux problèmes. On soupçonne en effet qu'il existe de la masse non visible, peut être même sous forme non-baryonique. On ne possède alors que des limites inférieures pour la valeur de Ω_0 . Si on ne tient compte que de la matière visible par exemple, $\Omega_0^L \approx 0.02$. L'étude de l'abondance relative des constituants de l'univers nous fournit une limite supérieure pour la densité de matière sous forme baryonique ($\Omega_0^b < 0.3$). Nous montrons dans la suite que l'étude du champ des vitesses propres donne accès en théorie à la valeur de la densité réduite totale Ω_0 . En pratique néanmoins, cette valeur est aujourd'hui inconnue.

On voit donc que la situation actuelle est loin d'être satisfaisante. Sur les 3 paramètres d'univers caractérisant le modèle, c'est la valeur de la constante de Hubble H_0 qui est la mieux connue (à un facteur 2 près).

Le modèle standard postule que la constante cosmologique Λ est nulle, le paramètre de décélération est alors égal à l'opposé de la moitié de la densité réduite Ω_0 ($\Omega_0 = -2q_0$). Cette hypothèse porte alors à deux le nombre de paramètres indépendants. De plus, on

a tendance à privilégier les modèles d'univers plat, c'est à dire possédant une densité réduite $\Omega_0 = 1$ (théorie inflationniste, etc...). Il ne reste alors plus qu'un paramètre libre, la constante de Hubble. Cette élimination progressive des paramètres libres ne possède aucune justification observationnelle. Elle semble plutôt significative de notre incapacité actuelle à déduire les 3 paramètres d'univers des observations disponibles.

II. LA DYNAMIQUE LOCALE DE L'UNIVERS :

L'hypothèse d'homogénéité et d'isotropie du contenu énergétique de l'univers semble valide à très grande échelle. L'observation montre que cette homogénéité est violée localement, typiquement pour des échelles inférieures à 100 Mpc, les galaxies ayant tendance à se regrouper pour former une hiérarchie de structures (les amas et super-amas de galaxies). On attribue l'origine de ces structures à des fluctuations initiales de la densité de matière, amplifiées au cours du temps sous l'action de la gravitation. L'étude de la dynamique à petite échelle de notre univers nous permet en principe de quantifier l'influence qu'ont ces perturbations locales sur la métrique de Robertson-Walker et ainsi d'accéder aux vraies valeurs des paramètres d'univers H_0 , q_0 et Ω_0 .

II.1. L'approximation newtonienne :

Les observations montrent que les galaxies forment des grandes structures de tailles importantes (au moins 20 Mpc pour le super-amas local) dont on attribue l'origine à des fluctuations initiales de la densité de matière ayant évoluées sous l'influence de la gravitation. La description de l'univers à partir de la métrique de Robertson-Walker (équation (A.1)) n'est alors plus valable, la densité du fluide cosmique étant dépendante de la position spatiale des galaxies ($\rho \equiv \rho(\mathbf{x}, t)$). L'étude de ces perturbations va être notablement simplifiée pour les deux raisons suivantes.

D'une part, nous étudions ces perturbations dans des volumes de taille très petite (typiquement 50 Mpc) par rapport à la taille actuelle de l'horizon (environ 3000 Mpc). Les effets de courbure peuvent alors être négligés (la structure de l'espace devient euclidienne). En effectuant le changement de coordonnées $\mathbf{y} = a(t)\mathbf{x}$ ($\mathbf{y} = (y_1, y_2, y_3)$ et $\mathbf{x} = (r, \theta, \phi)$) (voir équation (A.1)), il est possible d'écrire la métrique de l'espace-temps de la façon suivante :

$$ds^2 \approx c^2 dt^2 - (dy_1^2 + dy_2^2 + dy_3^2) \quad (\text{A.10})$$

D'autre part, les perturbations de la densité de matière ne créent pas un champ de vitesses d'amplitude importante (typiquement 500 km.s^{-1} , à comparer avec la vitesse de la lumière $c = 3 \cdot 10^5 \text{ km.s}^{-1}$). Le fluide cosmique est ainsi non relativiste ($v \ll c$), et on peut traiter les perturbations dans l'approximation linéaire des champs faibles. La composante temporelle de la métrique d'espace-temps perturbée $\hat{g}_{\mu\nu}$ devient alors (voir Peebles (1980), équation 6.12) :

$$\hat{g}_{tt} \approx c^2 + 2\Phi(\mathbf{y}, t) \quad (\text{A.11})$$

Et les équations d'Einstein imposent la relation suivante entre le potentiel $\Phi(\mathbf{y}, t)$ et la densité du fluide cosmique $\rho(\mathbf{y}, t)$ (dans le cas particulier où la pression du fluide est négligée) :

$$\nabla_{\mathbf{y}}^2 \Phi = 4\pi G \rho(\mathbf{y}, t) - \Lambda \quad (\text{A.12})$$

Les géodésiques de l'espace-temps sont alors données par l'équation suivante (voir Peebles (1980), équation 6.14)) :

$$\frac{d\mathbf{y}}{dt} = -\nabla_{\mathbf{y}} \Phi \quad (\text{A.13})$$

Si les perturbations de la densité de matière sont effectivement petites, leurs présences n'affectent alors pas le comportement dynamique global de notre univers. En particulier, l'équation (A.4) reste valable et l'expression du paramètre d'échelle $a(t)$ reste inchangée. On étudie généralement la dynamique locale de l'univers dans le repère comobile (t, \mathbf{x}) afin de séparer les contributions locales et globales à des quantités comme la vitesse \mathbf{v} ou l'accélération γ . On définit alors le contraste en densité $\delta(\mathbf{x}, t)$ à partir de la densité perturbée $\rho(\mathbf{x}, t)$ et la densité moyenne $\bar{\rho}(t)$ de l'univers au temps t :

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad (\text{A.14})$$

On introduit ensuite le potentiel gravitationnel propre $\phi_p(\mathbf{x}, t)$ vérifiant l'équation :

$$\nabla_{\mathbf{x}}^2 \phi_p = 4\pi G a(t)^2 \bar{\rho}(t) \delta(\mathbf{x}, t) \quad (\text{A.15})$$

Compte tenu des conditions imposées aux limites, ce potentiel propre admet en fait comme solution (voir Peebles (1980), équation 8.1) :

$$\phi_p(\mathbf{x}, t) = -G a(t)^2 \bar{\rho}(t) \int d^3 \mathbf{x}' \frac{\delta(\mathbf{x}', t)}{\|\mathbf{x}' - \mathbf{x}\|} \quad (\text{A.16})$$

Il reste à définir le champ des vitesses propres $\mathbf{v}_p(\mathbf{x}, t)$ ainsi que le champ d'accélération propre $\mathbf{g}_p(\mathbf{x}, t)$ engendrés par les fluctuations locales de la densité de matière :

$$\mathbf{v}_p(\mathbf{x}, t) = a(t) \frac{d\mathbf{x}}{dt} ; \quad \mathbf{g}_p(\mathbf{x}, t) = \frac{d\mathbf{v}_p}{dt} \quad (\text{A.17})$$

Et d'après l'équation (A.13), la dynamique locale dans le repère comobile est régie par l'équation suivante :

$$\mathbf{g}_p(\mathbf{x}, t) = -\frac{\nabla_{\mathbf{x}} \phi_p}{a(t)} \quad (\text{A.18})$$

Les quantités comme la distance propre $d_p(t)$ (équation (A.2)), le contraste en densité $\delta(\mathbf{x}, t)$ (équation (A.14)), le champ des vitesses propres $\mathbf{v}_p(\mathbf{x}, t)$ ou le champ d'accélération $\mathbf{g}_p(\mathbf{x}, t)$ (équation (A.17)) évoluent au cours du temps. La vitesse de la lumière étant finie, il nous est en principe impossible de connaître ces quantités au temps actuel t_0 . Pour une galaxie située à une distance de 100 Mpc par exemple, nous avons au mieux accès qu'aux valeurs que possédaient ces quantités il y a $100/c$ Mpc.m⁻¹.s, soit environ $\Delta t = 10^8$ ans. Les valeurs actuelles doivent donc être différentes. Si la galaxie possédait une vitesse propre de 500 km.s⁻¹ dirigée vers nous, sa distance actuelle est alors d'environ 100 Mpc - Δd avec $\Delta d = 500 \times 10^8 \times 3.15 \cdot 10^7$ km.

Mais ici encore, le fait que l'on étudie l'univers localement va considérablement simplifier les choses. Notre volume d'analyse est petit par rapport à la taille de l'horizon et le temps que met la lumière pour le traverser est négligeable par rapport au temps propre d'évolution des structures que nous étudions. La valeur de Δd est par exemple très faible comparée aux distances impliquées ($\Delta d = 1.8 \cdot 10^{-3}$ Mpc). Il en va de même pour le temps Δt . Il est alors possible d'identifier les valeurs des quantités au temps actuel avec

celles effectivement mesurées. De même, on considérera que la force de gravitation agit spontanément à distance.

On peut donc effectuer en toute quiétude l'étude de la dynamique locale de notre univers dans le cadre de la mécanique purement newtonienne. Il est alors nécessaire de se placer dans le repère comobile (t, \mathbf{x}) et de travailler avec les quantités propres précédemment définies.

II.2. Le champ des vitesses propres :

Le passage à la mécanique newtonienne introduit à nouveau la notion de référentiel absolu des vitesses. Un observateur quelconque est généralement en mouvement par rapport à ce repère absolu (le repère comobile). Il faut alors corriger le champ des vitesses observées de ce mouvement relatif.

Le champ des vitesses propres contribue au décalage des longueurs d'onde des raies spectrales des galaxies par effet Doppler-Fizeau. Le redshift z d'une galaxie se décompose ainsi en un terme purement cosmologique (c'est la loi de Hubble) et un terme dû aux perturbations locales du champ des vitesses. La mesure de ce redshift nous donne accès en pratique à la composante radiale (la projection le long de la ligne de visée) de la vitesse propre d'une galaxie.

II.2.1. Le repère absolu des vitesses :

Nous avons vu dans la section I.1 que le repère comobile définit en fait un repère absolu dans lequel le contenu énergétique de l'univers est globalement au repos. Pour un observateur quelconque M , le champ des vitesses propres actuel $v_p(\mathbf{x}, t_0)$ (défini précédemment par l'équation (A.17)) est alors le champ des vitesses propres observées $v_p^M(\mathbf{x}, t_0)$ corrigé du mouvement relatif de l'observateur dans le repère comobile \mathbf{v}_C^M :

$$v_p(\mathbf{x}, t_0) = v_p^M(\mathbf{x}, t_0) + \mathbf{v}_C^M \quad (\text{A.19})$$

La vitesse relative \mathbf{v}_C^M peut être déterminée par les deux techniques suivantes.

Nous avons montré section I.3.2 que les modèles d'univers du type big-bang prévoient

l'existence d'un rayonnement électromagnétique fossile effectivement détecté. Les photons de ce rayonnement d'origine cosmologique proviennent alors de la surface de dernière diffusion (voir section I.3.2) qui est en fait au repos par rapport au repère comobile (t, \mathbf{x}) . Pour un observateur M au repos par rapport au repère comobile ($\mathbf{v}_C^M = \mathbf{0}$), ce rayonnement apparaît parfaitement isotrope. Par contre si cet observateur possède une vitesse relative non nulle, cette isotropie subit une altération due à l'effet Doppler-Fizeau. La longueur d'onde des photons détectés dans la direction du mouvement doit en effet être décalée vers le bleu et celle des photons dans la direction opposée au mouvement vers le rouge. De plus, l'amplitude du décalage de ces longueurs d'onde est proportionnelle à la vitesse de l'observateur.

Une telle anisotropie dipolaire a été détectée dans le fond diffus cosmologique. Si notre modèle de big-bang est réaliste, ce dipole est alors la trace du mouvement de notre galaxie par rapport au repère comobile et nous permet de déterminer la direction ainsi que le module de cette vitesse.

Une autre possibilité existe pour estimer la vitesse relative d'un observateur. Le principe cosmologique postule en effet qu'un volume d'espace suffisamment grand est représentatif de l'univers dans son ensemble. Le champ des vitesses propres $v_p(\mathbf{x})$ (dans le repère comobile) doit donc être nul en moyenne dans ce volume. La moyenne du champ des vitesses propres observé $v_p^M(\mathbf{x})$ dans ce volume est alors égale à l'opposé de la vitesse relative de l'observateur par rapport au repère comobile \mathbf{v}_C^M . Malheureusement, nous ne possédons pas encore des échantillons de vitesses propres couvrant un volume représentatif de l'univers dans son ensemble. Cette méthode est ainsi actuellement inapplicable.

II.2.2. La composante radiale du champ des vitesses propres :

Nous avons montré dans la section I.3.1 que l'expansion de l'univers était responsable d'un décalage vers le rouge z de la longueur d'onde des raies spectrales d'absorption et d'émission des galaxies. Le champ des vitesses propres $v_p(\mathbf{x})$ va alors contribuer à ce redshift z d'origine purement cosmologique par effet Doppler-Fizeau. Dans le repère où notre galaxie est au repos, l'équation (A.9) est alors modifiée de la façon suivante (les termes d'ordre 2 sont omis) :

$$cz = H_0 D + v_p^r \quad (\text{A.20})$$

avec D la distance propre de la galaxie ($D = a(t_0) \|\mathbf{x}\|$) et v_p^r la projection de la vitesse propre de la galaxie le long de la ligne de visée ($v_p^r = \mathbf{v}_p(a(t_0)\mathbf{x}) \cdot \hat{\mathbf{x}}$). C'est en mesurant le redshift z et en estimant la distance D des galaxies que l'on obtient en pratique le champ des vitesses radiales propres $v_p^r(\mathbf{x})$ (voir partie B).

On trouve parfois dans la littérature la notion de champ de Hubble \mathbf{v}_H qui donne une interprétation newtonienne de l'expansion de l'univers. Le décalage vers le rouge z d'origine cosmologique est alors associé à un champ de vitesses \mathbf{v}_H centré sur l'observateur et d'amplitude variant avec la distance ($\mathbf{v}_H(\mathbf{x}) = H_0 \mathbf{x}$). Le champ des vitesses observées $\mathbf{v}(\mathbf{x})$ se décompose alors en 3 termes d'origines différentes :

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}_H(\mathbf{x}) + \mathbf{v}_C + \mathbf{v}_p(\mathbf{x}) \quad (\text{A.21})$$

avec $\mathbf{v}_p(\mathbf{x})$ le champ des vitesses propres exprimé dans le repère comobile (t, \mathbf{x}) (le référentiel absolu), \mathbf{v}_C la vitesse relative de l'observateur (voir section II.2.1) et $\mathbf{v}_H(\mathbf{x})$ caractérisant l'expansion de l'univers.

L'utilisation de l'équation (A.21) nécessite néanmoins une certaine prudence, parce que le champ de Hubble n'est pas un champ de vitesses à proprement parler. Notamment, le champ de Hubble ne subit pas les mêmes transformations que les champs de vitesses ordinaires (par exemple $\mathbf{v}_p(\mathbf{x})$) lorsqu'on effectue un changement de repère dans l'espace. Toutefois, l'équation (A.21) est d'ordinaire projetée le long de la ligne de visée liant notre galaxie à une galaxie observée au point \mathbf{x} , ce qui fait disparaître ce type d'ambiguïté :

$$cz = H_0 \|\mathbf{x}\| + \mathbf{v}_C \cdot \hat{\mathbf{x}} + v_p^r(\mathbf{x}) \quad (\text{A.22})$$

III. L'ETUDE CINEMATIQUE DES GRANDES STRUCTURES :

Nous avons présenté dans les sections précédentes les idées générales qui conditionnent la modélisation de notre univers. On postule en fait que le contenu massique de l'univers est distribué dans sa globalité de façon homogène et isotrope dans l'espace. Sa dynamique à grande échelle est alors complètement caractérisée par les trois paramètres d'univers H_0 , q_0 et Ω_0 . Cette uniformité est toutefois troublée localement par de faibles fluctuations de la densité de matière dont la longueur de cohérence est petite devant la taille actuelle de l'horizon. Un volume d'espace de rayon supérieur à l'échelle caractéristique des fluctuations de la densité de matière est alors représentatif de l'univers dans son ensemble.

L'étude de ces perturbations actuellement observées présente deux principaux intérêts. D'abord, elle permet d'analyser les caractéristiques propres de ces fluctuations et ainsi de découvrir leur origine. Ensuite, la détermination des paramètres d'univers est bruitée par la présence de telles perturbations. La connaissance de la dynamique de notre voisinage nous offre alors en principe la possibilité de corriger les estimations des paramètres d'univers H_0 , q_0 et Ω_0 .

Depuis une quinzaine d'années maintenant, des catalogues de galaxies dont on a mesuré le redshift et estimé indépendamment la distance sont disponibles. Ces échantillons nous donnent alors accès à la connaissance locale du champ des vitesses radiales propres. L'étude cinématique des grandes structures peuplant notre entourage est en fait complémentaire de l'analyse de la distribution des galaxies dans l'espace. Notamment, elle nous permet en principe de déduire le comportement de la matière non visible. De plus, l'obtention du champ des vitesses radiales propres n'est pas soumise au problème de complétude des échantillons. Une information est ainsi disponible à plus grande échelle.

III.1. Le comportement de la matière non visible :

L'analyse de la distribution des galaxies dans l'espace nous permet de déterminer les caractéristiques propres des fluctuations de la densité de matière lumineuse. On a calculé par exemple que la fonction de corrélation spatiale des galaxies $\xi_L(r)$ vérifiait une loi de puissance (d'exposant -1.8) pour des séparations r entre 0.1 et 10 Mpc environ. Mais tout porte à croire que la matière lumineuse (principalement les galaxies) ne représente qu'une petite partie de la matière totale dont est composé notre univers. Il est même possible que les propriétés de la distribution spatiale des galaxies ne reflètent en rien les caractéristiques de la densité de matière totale $\rho(x)$ introduite précédemment dans l'équation (A.14).

L'étude du champ des vitesses propres nous permet en fait de déduire le comportement de la matière non visible. En effet, le champ des vitesses propres $v_p(x)$ défini en (A.17) est directement relié à la densité de masse totale $\rho(x)$ par l'intermédiaire des équations (A.18) et (A.15). On peut donc en principe dériver le champ de densité de matière totale à partir de la connaissance du champ des vitesses propres (voir partie E).

Une comparaison de cette densité avec la densité de matière lumineuse (déduite de la distribution des galaxies) nous donne alors accès à la distribution spatiale de la matière non visible. Notamment, il est possible par cette méthode d'estimer le paramètre de biais b introduit dans les scénarii de formations de structures dans le but de quantifier la différence entre masse lumineuse et masse totale (voir partie C, section II.2). Nous montrons dans la suite (partie D, section I.3.1) que les densités dérivées à partir du champ des vitesses propres et de la distribution des galaxies présentent des différences remarquables.

III.2. Une information à plus grande échelle :

Les échantillons de galaxies actuellement disponibles sont soumis à des effets de sélection. Par exemple, il existe une magnitude apparente de coupure au dessus de laquelle les télescopes ne sont plus capable de détecter une galaxie (sa luminosité apparente est trop faible). A partir d'une certaine distance donc, on observe plus alors les galaxies de faible luminosité absolue. Si l'on veut déterminer la distribution des galaxies dans l'espace,

il est nécessaire de tenir compte de ces effets de sélection qui éliminent du catalogue certaines galaxies pourtant présentes en réalité. Il est possible de corriger statistiquement ces effets de sélection. Toutefois, ces extrapolations deviennent très vite incertaines lorsqu'on désire connaître la distribution des galaxies à des distances importantes ou si les effets de sélection sont mal connus. Ces problèmes de complétude des échantillons interdisent en pratique de tirer de l'information de la distribution des galaxies à grande échelle.

L'étude du champ des vitesses radiale n'est pas soumise à ces problèmes de complétude. Peu importe en effet que l'on n'ait pas accès aux vitesses radiales de toutes les galaxies. Il suffit simplement que le champ des vitesses soit mesuré assez uniformément dans le volume d'espace couvert par le catalogue. La sélection des objets de l'échantillon, si elle transforme la fonction de distribution des galaxies observées, ne va pas modifier la valeur du champ des vitesses radiales propres. Elle définit en fait le support spatial sur lequel on connaît les valeurs du champ des vitesses. Cette remarque est importante puisque en pratique, le champ des vitesses nous permet d'analyser les fluctuations de la densité de matière à plus grande échelle (voir partie C, section II.1). C'est par exemple à partir de l'étude du champ des vitesses radiales propres que le "Grand Attracteur" (une concentration de matière dominant notre cinématique locale dans un rayon de 50 Mpc environ) a été détecté (voir partie D).

III.3. Une connaissance incomplète du champ des vitesses :

L'étude de la cinématique des grandes structures de l'univers permet donc d'estimer le champ de densité de masse totale à grande échelle. Néanmoins, seule la composante radiale du champ des vitesses propres nous est observationnellement accessible. De plus la détermination de ce champ de vitesses radiales est entachée d'incertitudes intrinsèques à la procédure d'évaluation de la distance la plus probable d'une galaxie.

On estime la vitesse radiale propre d'une galaxie via l'équation (A.20). Il faut alors mesurer le redshift z , déterminer la distance D de la galaxie et connaître la constante de Hubble H_0 . La distance d'une galaxie lointaine est estimée statistiquement à partir des indicateurs secondaires de distance (voir partie B, section I). Même correctement calibrées et débiaisées, ces relations statistiques nous donnent la distance de la galaxie avec une

incertitude intrinsèque d'environ 20%. De plus, si on ajoute les erreurs commises lors de l'estimation de la constante de Hubble H_0 (voir partie B, section III), on voit que le champ des vitesses radiales propres est assez mal connu. Il sera alors nécessaire de tenir compte de ces incertitudes en vitesses lors de nos diverses analyses de la cinématique locale de l'univers.

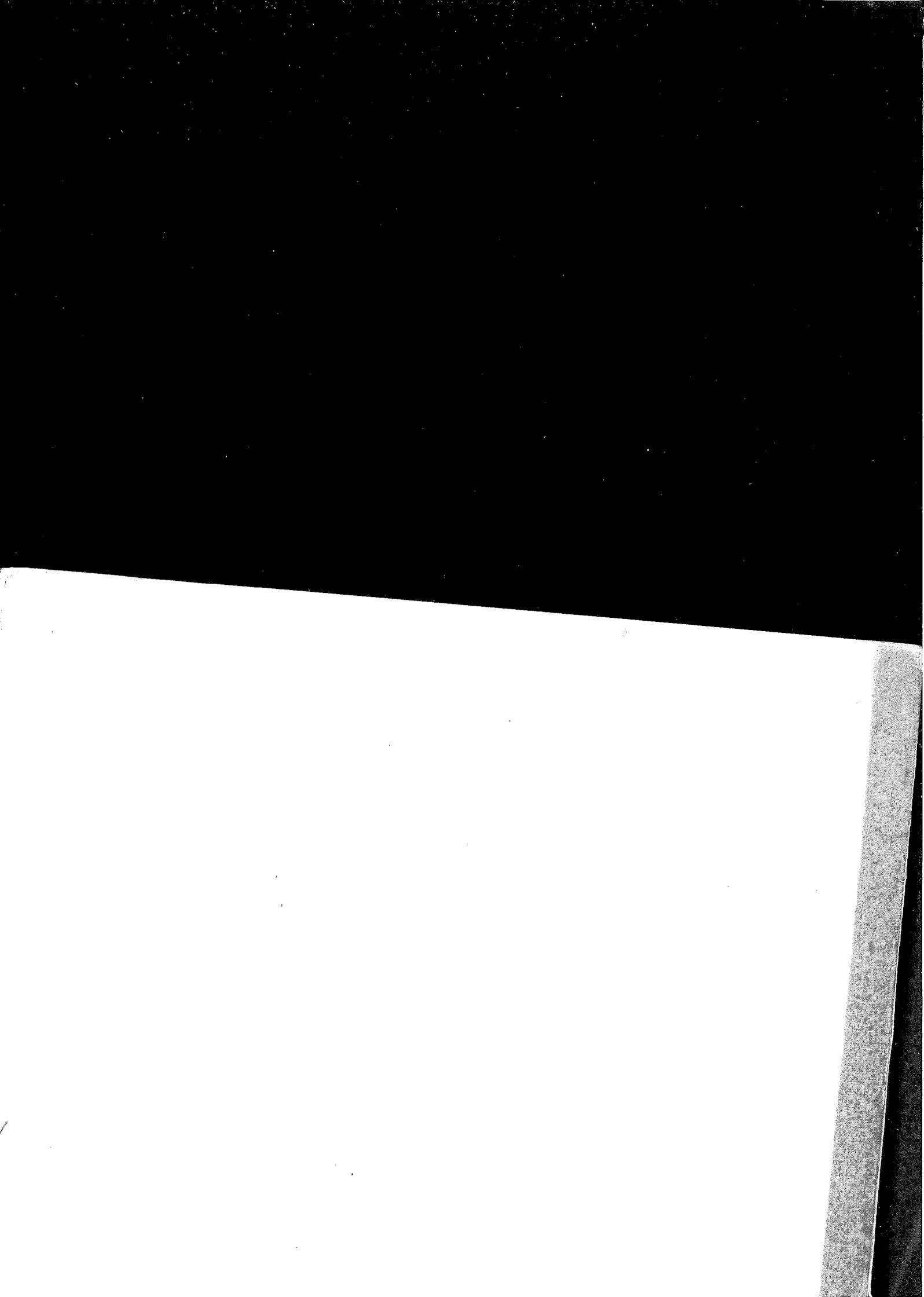
Outre ces incertitudes, seule la composante radiale du champ des vitesses propres est disponible. Si on souhaite dériver des quantités comme la densité de masse totale $\rho(\mathbf{x})$ à partir de $v_p^r(\mathbf{x})$, il est alors obligatoire de faire des hypothèses sur la géométrie du champ des vitesses propres (par exemple de postuler que ce champ est irrotationnel à grande échelle (voir partie E)). De même, il faudra supposer que le champ des vitesses est statistiquement homogène et isotrope pour évaluer le tenseur de corrélation des vitesses (voir partie C). Ces hypothèses ne pourront en aucun cas être testées dans le cadre de ces analyses.

REFERENCES :

[1] : P.J.E. Peebles : "The large scale structure of the universe", (1980) Princeton university series, N.J.

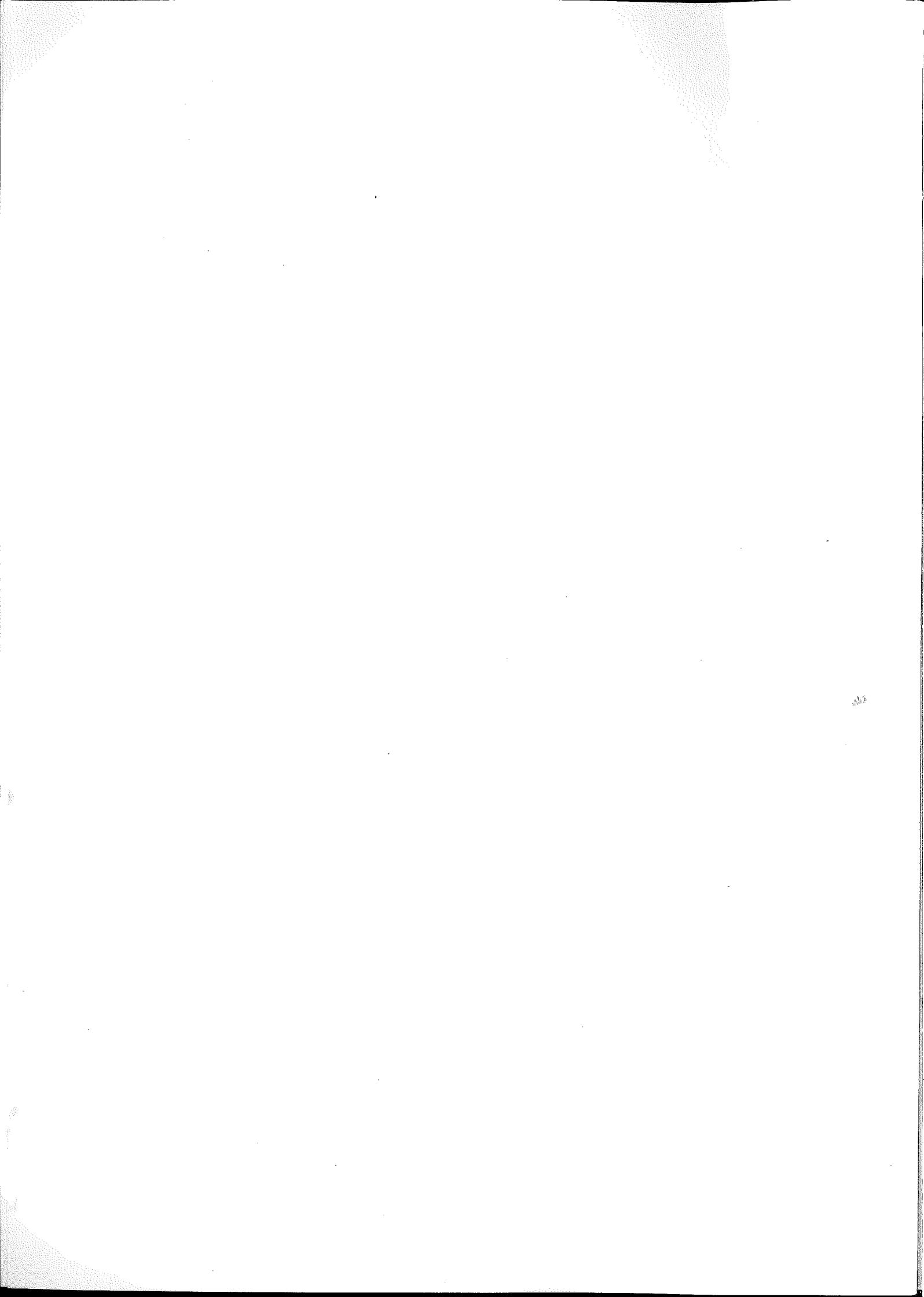
[2] : W. Rindler : "Essential relativity", Second edition, Springer-Verlag (1977)

[3] : S. Weinberg : "Gravitation and cosmology", (1972) New York: Wiley.



PARTIE B :

LE CHAMP DES VITESSES
RADIALES PROPRES



SOMMAIRE :

I. Détermination de la distance d'une galaxie :

I.1. Les indicateurs de distance :

I.1.1. Les indicateurs primaires de distance :

I.1.2. Les indicateurs secondaires de distance :

I.2. Les problèmes rencontrés lors de l'évaluation de la distance d'une galaxie :

I.2.1. Existence de plusieurs couples (a,b) :

I.2.2. Le problème des "biais" statistiques :

I.3. Formalisation statistique du problème :

I.3.1. Le formalisme statistique des densités de probabilité :

I.3.2. Correspondance entre modèles statistiques et droites de régression :

I.3.3. Le choix du modèle statistique :

I.4. Phase de calibration des indicateurs secondaires de distance :

I.4.1. Les procédures de calibration généralement adoptées :

I.4.2. La technique du maximum de vraisemblance :

I.4.3. Applications pour différents modèles statistiques :

I.4.4. Comparaison des différents modèles statistiques :

I.5. Estimation de la distance la plus probable d'une galaxie :

I.5.1. Estimation pour le modèle statistique $[u = p]$:

I.5.2. Estimation pour le modèle statistique $[u = M]$:

I.5.3. Exemples de différentes analyses :

I.5.4. Conclusion :

II. Le champ des vitesses :

II.1. Le décalage des raies spectrales d'une galaxie :

II.2. La composante radiale du champ des vitesses :

III. Estimation de la constante de Hubble H_0 :

III.1. La calibration généralement adoptée :

III.1.1. L'inhomogénéité du champ des vitesses propres :

III.1.2. Le problème des biais statistiques :

III.2. Application du formalisme statistique défini en I.3 :

IV. Le catalogue de Faber et *al.* (1989) :

IV.1. Les données observationnelles :

IV.2. Les indicateurs de distance :

IV.3. La phase de calibration :

IV.3.1. La détermination de la pente α :

IV.3.2. La détermination du point zéro :

IV.4. L'évaluation de la distance d'une galaxie :

IV.5. Le champ des vitesses radiales propres :

On attribue l'origine du décalage vers le rouge des raies spectrales d'une galaxie (le redshift z) à l'expansion de l'univers. En première approximation, ce décalage vers le rouge est proportionnel à la distance D qui nous sépare de la galaxie : c'est la loi de Hubble ($c.z = H_0.D$, avec c la vitesse de la lumière et H_0 la constante de Hubble). Mais à cette composante cosmologique du redshift vient s'ajouter un terme supplémentaire. En effet, les fluctuations primordiales de la densité de matière dans l'univers ont engendré, par l'intermédiaire de processus gravitationnels, un champ de vitesses propres v_p . Ce champ de vitesses, statistiquement homogène et isotrope à grande échelle, crée des écarts à la loi de Hubble localement. Chaque galaxie possède alors une vitesse propre qui contribue au redshift observé z par effet Doppler-Fizeau. Si v_p^r est la vitesse radiale propre d'une galaxie exprimée dans le repère des vitesses où notre galaxie est au repos, l'équation suivante est vérifiée : $c.z = H_0.D + v_p^r$.

La mesure du redshift z et la détermination de D et H_0 nous permettent ainsi d'accéder à la composante radiale du champ des vitesses propres v_p . La distance d'une galaxie est obtenue à partir de relations statistiques (les indicateurs de distance). Ces estimateurs sont souvent soumis à des biais difficiles à corriger à l'aide des techniques usuelles. Nous introduisons dans cette partie une formalisation statistique adéquate qui nous fournit une base théorique solide afin de traiter ces problèmes de biais, intervenant également dans l'évaluation statistique de la constante de Hubble H_0 . La valeur estimée de la vitesse radiale propre d'une galaxie, après un traitement correct des biais, comporte encore une incertitude intrinsèque d'environ 20 %.

A titre d'exemple, nous présentons l'échantillon de Faber et *al.* (1989), composé de 512 galaxies elliptiques pour lesquelles les vitesses radiales propres sont déduites par les techniques usuelles. Cet échantillon possède une profondeur d'environ 60 Mpc. C'est donc le catalogue en redshift-distance le plus étendu actuellement disponible.

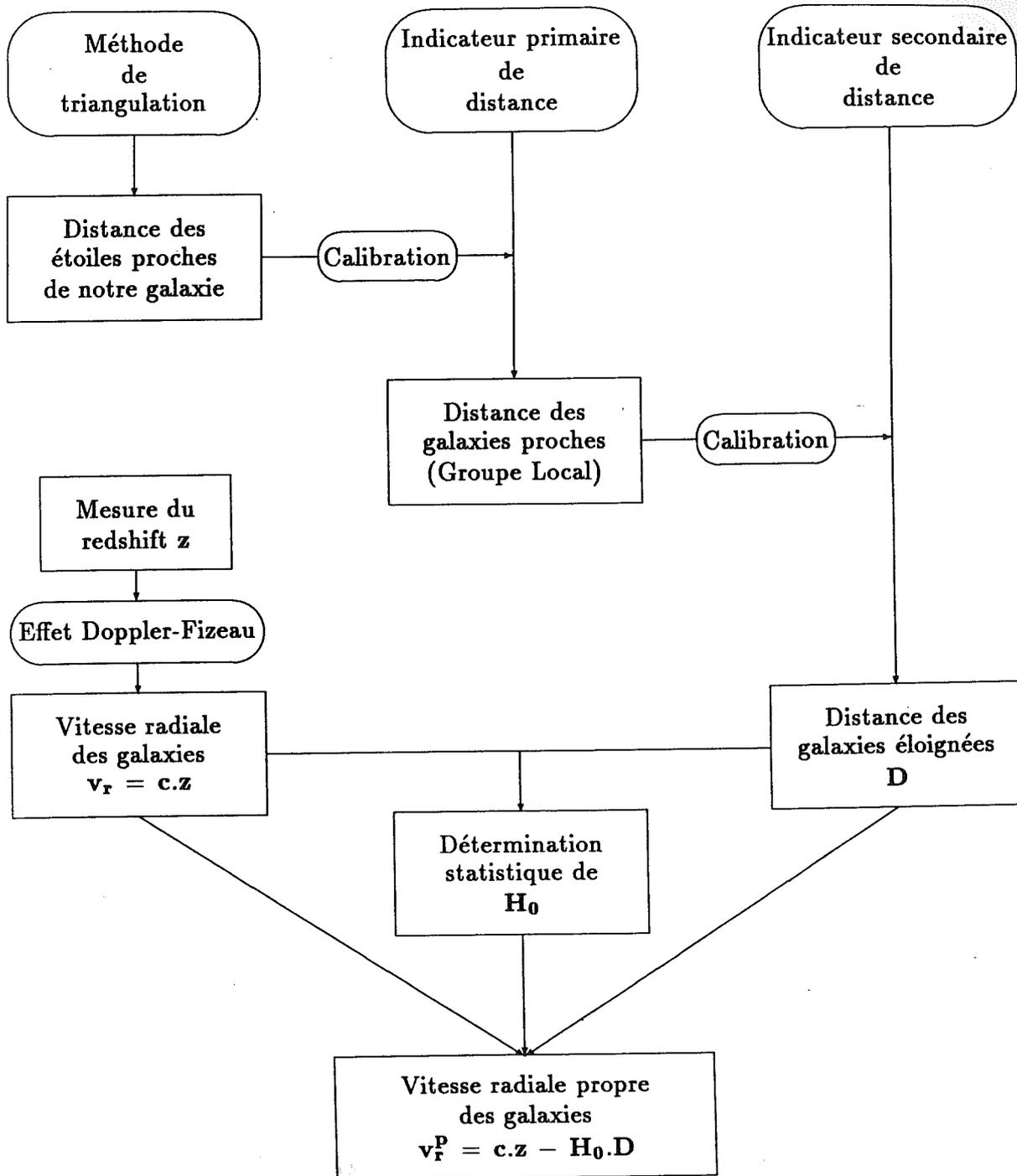


Figure B.1 : Procédure permettant l'obtention de la vitesse radiale propre d'une galaxie.

I. DETERMINATION DE LA DISTANCE D'UNE GALAXIE :

Si l'on connaît la luminosité intrinsèque d'une source lumineuse, la mesure de sa magnitude apparente permet de déterminer la distance qui nous sépare de cet objet. Mais l'estimation de la magnitude absolue M (ou de la luminosité intrinsèque) d'une galaxie n'est pas chose aisée. Il existe néanmoins des relations statistiques (les indicateurs de distance) liant la magnitude absolue M d'une source avec certains paramètres physiques mesurables (p). On distingue deux classes de relations : les indicateurs primaires de distance, qui s'appliquent à des objets lumineux (étoiles pulsantes, novas ou supernovas) présents à l'intérieur des galaxies, et les indicateurs secondaires de distance s'appliquant aux galaxies elles-mêmes.

La détermination de la distance d'une galaxie lointaine au moyen des indicateurs secondaires de distance nécessite plusieurs étapes, notamment une calibration de la relation statistique liant la magnitude absolue M et le paramètre physique observable p ($M = a.p + b$). Lors de cette phase de calibration, deux problèmes apparaissent. D'une part, sans une connaissance plus intime du processus physique sous-jacent aux indicateurs secondaires de distance, la calibration de cette relation n'est pas unique : il faut alors choisir a priori une calibration. D'autre part, la technique statistique classiquement employée dans cette phase de calibration (méthode de régression) semble mal adaptée lorsqu'il s'agit de traiter des échantillons de galaxies soumis à des effets de sélection non triviaux (cas dits "biaisés" dans la littérature).

Nous introduisons ici un formalisme statistique plus général qui nous permet de clarifier ces problèmes. Nous établissons une correspondance entre le choix d'une calibration et la description de l'échantillon par son modèle statistique associé. Nous montrons que l'utilisation d'un modèle statistique (ou le choix d'une calibration) plutôt qu'un autre introduit une erreur négligeable dans le traitement des données.

Cette propriété nous permet alors de choisir le modèle statistique le plus apte à traiter les effets de sélection auxquels est soumis l'échantillon. Nous illustrons par des exemples quels sont les avantages du formalisme introduit (notamment pour le traitement des cas "biaisés") et comment la sélection des données influence le choix du modèle statistique à

utiliser.

Nous décrivons de plus la procédure globale permettant l'obtention de la distance la plus probable d'une galaxie, connaissant sa magnitude apparente m ainsi que son paramètre physique observable p . En pratique, les données observationnelles sont soumises à des effets de sélection assez mal connus. L'estimation de la distance la plus probable d'une galaxie isolée, après une calibration et un traitement des "biais" aussi correct que possible, comporte encore une erreur intrinsèque de 20 %.

I.1. Les indicateurs de distance :

Il est possible de déterminer la distance qui nous sépare d'une source lumineuse en mesurant d'une part sa magnitude apparente m et d'autre part, en évaluant sa magnitude absolue M . La distance r de cette source est alors donnée par les relations suivantes :

$$\mu = m - M \quad \text{avec } \mu \text{ le module de distance.} \quad (B.1)$$

$$\mu = 5 \log_{10}(r) + 25 \quad (r \text{ en Mpc}) \quad (B.2)$$

On possède des relations statistiques (les indicateurs secondaires de distance) qui permettent d'estimer la magnitude absolue M d'une galaxie lointaine à partir de certains paramètres physiques observables p . Mais ces relations statistiques doivent être préalablement calibrées sur des galaxies dont on a déterminé indépendamment la distance. Pour se faire, on utilise les galaxies proches pour lesquelles la distance est connue grâce aux indicateurs primaires de distance, c'est à dire des relations statistiques liant la luminosité intrinsèque de certains types d'étoiles avec des quantités physiques mesurables. La calibration de ces indicateurs primaires de distance s'effectue au moyen des étoiles proches peuplant notre galaxie dont on évalue les distances par la méthode de triangulation. On voit donc que le processus permettant la détermination de la distance d'une galaxie lointaine n'est pas simple et présente certains inconvénients. D'une part, les erreurs commises lors des différentes phases de calibration sont cumulatives. D'autre part, du fait du caractère statistique des indicateurs de distance, on ne peut estimer que la distance la plus probable d'une galaxie. Ainsi, après un traitement idéal des données

(prise en compte de tous les problèmes statistiques inhérents à la méthode), l'incertitude sur la distance d'une galaxie reste d'environ 20 %.

I.1.1 Les indicateurs primaires de distance :

On obtient la distance des étoiles proches appartenant à notre galaxie par la méthode de triangulation. Parmi ces étoiles, certaines montrent des propriétés particulières qui permettent d'estimer leurs magnitudes absolues indépendamment de leurs magnitudes apparentes.

Les populations d'étoiles variables du type Céphéides ou RR Lyrae possèdent par exemple une période de pulsation proportionnelle à leur magnitude absolue. Ainsi la mesure de la période de pulsation de ces étoiles permet d'estimer leurs distances, via la relation période-luminosité de cette population.

La variation au cours du temps de la luminosité intrinsèque des Novas ou des Supernovas semble avoir un comportement universel. Encore un indicateur donnant accès à la distance de ces objets.

D'autres indicateurs de distance sont disponibles (nébuleuses planétaires, amas globulaires etc...). (Voir Van den Bergh (1989) ou Bottinelli et *al.* (1990) pour une revue).

Ainsi ces types d'objets possèdent des caractéristiques physiques observables qui permettent l'estimation de leurs magnitudes absolues, via les indicateurs primaires de distance. Après avoir calibré ces indicateurs primaires à l'aide des étoiles proches peuplant notre galaxie, il est possible d'appliquer ces relations aux étoiles des galaxies environnantes. Bien sûr, il est alors nécessaire de mesurer leur magnitude apparente ainsi que leur paramètre physique observable, et donc de distinguer individuellement ces objets.

Actuellement, nous pouvons déduire par ces méthodes les distances de quelques galaxies appartenant pour la plupart à notre Groupe Local. Les différents indicateurs primaires de distance, correctement calibrés et recoupés, nous donnent en fait les distances d'environ 30 galaxies, avec une erreur de moins de 10% (voir Van den Bergh (1989), Tamman (1984)). Ces galaxies dont les distances sont inférieures à 7 Mpc nous serviront ultérieurement à calibrer les indicateurs secondaires de distance.

I.1.2 Les indicateurs secondaires de distance :

Il s'agit maintenant d'estimer la distance des galaxies lointaines, c'est à dire n'appartenant pas au Groupe Local et pour lesquelles l'obtention de la distance ne peut se faire au moyen des indicateurs primaires de distance. Heureusement, certaines quantités physiques mesurables sont fortement corrélées à la luminosité intrinsèque de ces galaxies. La méthode est en fait basée sur l'existence de relations statistiques linéaires, reliant ces quantités physiques observables à la magnitude absolue, de la forme :

$$M = a.p + b + \zeta \quad (B.3)$$

avec ζ une variable aléatoire centrée

p la mesure de la quantité physique observée

a et b deux constantes

La quantité physique impliquée dans l'équation (B.3) ainsi que la physique sous-jacente à cette relation dépendent de la morphologie de la galaxie. On distingue notamment la relation de Tully-Fisher pour les galaxies spirales et la relation de Faber-Jackson pour les galaxies elliptiques.

A) RELATION DE TULLY-FISHER :

La quantité physique observée pour les galaxies spirales est W_R^i , la largeur du profil de la raie HI. Elle est en fait directement reliée à la vitesse de rotation de la galaxie spirale, en tenant compte de certaines corrections à adopter (Tully et al. (1977) ou Van Den Berg (1989)). Or, si l'on considère que les galaxies spirales sont des objets virialisés, la vitesse de rotation de la galaxie et sa masse sont reliées par une loi de puissance. Sachant par ailleurs que le rapport masse-luminosité pour une population de galaxies du même type est sensiblement constant, on déduit que la luminosité d'une galaxie spirale est approximativement proportionnelle au logarithme de sa vitesse de rotation. Le paramètre p introduit dans la relation de Tully-Fisher est le logarithme de la largeur de la raie HI.

$$M = a.\log W_R^i + b \quad (B.4)$$

Les observations de galaxies regroupées dans les amas, et donc possédant la même distance, confirment bien la linéarité de cette loi (Pierce et *al.* (1988)).

B) RELATION DE FABER-JACKSON :

La quantité mesurée est la dispersion des vitesses σ dans la galaxie elliptique. Si les galaxies elliptiques sont des corps virialisés, le logarithme du rapport masse-dispersion des vitesses est alors constant. Il s'en suit le même raisonnement que pour les galaxies spirales, p étant alors le logarithme de la dispersion des vitesses :

$$M = a \cdot \log \sigma + b \quad (B.5)$$

La relation de Faber-Jackson a été vérifiée pour plusieurs amas (Faber et *al.* (1976)).

Un équivalent de l'équation (B.5), faisant intervenir le module de distance μ , la dispersion des vitesses σ et une nouvelle quantité physique mesurable, le paramètre photométrique D_n s'exprime :

$$\frac{\mu}{5} = a \cdot \log \sigma - \log D_n + b \quad (B.6)$$

Le paramètre photométrique D_n est défini comme le diamètre d'une ouverture circulaire pour laquelle la brillance moyenne de surface de la galaxie observée est égale à une certaine valeur Σ (Dressler et *al.* (1987)). L'introduction de ce paramètre permet de réduire la dispersion intrinsèque de la relation de Faber-Jackson, la rendant ainsi quantitativement comparable à la relation de Tully-Fisher pour les galaxies spirales.

I.2. Les problèmes rencontrés lors de l'évaluation de la distance d'une galaxie :

L'équation (B.3), $M = a \cdot p + b + \zeta$, nous fournit donc la possibilité d'estimer la magnitude absolue M d'une galaxie, ou tout au moins la valeur la plus probable de M , connaissant le paramètre observé p . Toutefois, il faudra avant tout calibrer cette relation, c'est à dire estimer statistiquement la valeur des constantes a et b à partir des galaxies proches dont les distances ont été évaluées indépendamment à l'aide des indicateurs primaires de distance. Ensuite, connaissant les valeurs du paramètre p ainsi que la magnitude apparente m , on pourra estimer la valeur la plus probable de la distance d'une

galaxie.

Mais la formalisation statistique de ces problèmes réclame beaucoup d'attention. D'une part, sans une information supplémentaire sur la physique sous-jacente à la relation (B.3), le couple (a,b) n'est pas défini de façon unique. Ainsi, il faudra choisir arbitrairement une calibration plutôt qu'une autre. D'autre part, on rencontre des difficultés lorsque l'on tente d'estimer le couple (a,b) ou la distance la plus probable d'une galaxie en présence d'effets de sélection avec la technique généralement utilisée (méthode des droites de régression).

I.2.1 Existence de plusieurs couples (a,b) :

Le problème statistique soulevé lorsqu'il s'agit de déterminer les constantes a et b est connu depuis fort longtemps. En effet, nous possédons un nuage de N points $\{G_k\}$ de coordonnées (p_k, M_k) , pour lequel nous voulons trouver la droite de régression Δ , d'équation $M = a.p + b$ telle que la "distance du nuage de points à la droite Δ " soit minimale. Il reste cependant à définir la notion de distance d'un nuage de points à une droite. La littérature nous fournit les "distances" les plus généralement utilisées (droite de régression verticale, droite de régression horizontale etc...) (Tully (1988), T. Isobe et *al.* (1991)), mais d'autres définitions de la distance d'un nuage de points à une droite sont tout aussi valables. Il est alors clair qu'à chaque définition d'une distance correspondra une valeur différente du couple de constantes (a,b) . A titre d'exemple, nous montrons figure (B.2) et figure (B.3) une famille de distances possibles ainsi que la variation de la valeur de a pour un nuage de points donné.

Dans la littérature, on fait référence à ce problème en introduisant la notion de relation directe ($M = a_d.p + b_d$) et de relation inverse ($p = a_i.M + b_i$) (voir Teerikorpi et *al.* (1990), par exemple). Lorsqu'on utilise la relation directe, on fait implicitement l'hypothèse que l'incertitude intrinsèque présente dans les indicateurs secondaires de distance (la variable ζ) concerne les magnitudes absolues M_k , ce qui revient à pratiquer une régression verticale, ou choisir $\theta = -\pi/4$ dans notre exemple. La relation inverse est applicable si les erreurs sont distribuées sur le paramètre p , ce qui correspond au choix de la régression horizontale (ou $\theta = \pi/4$ pour notre exemple). Le choix d'une relation plutôt qu'une autre n'est pas évident (Lynden-Bell et *al.* (1988), Teerikorpi (1990) ou Fouqué et *al.* (1990)) et revient à introduire une hypothèse supplémentaire sur la distribution de l'échantillon de galaxies étudié.

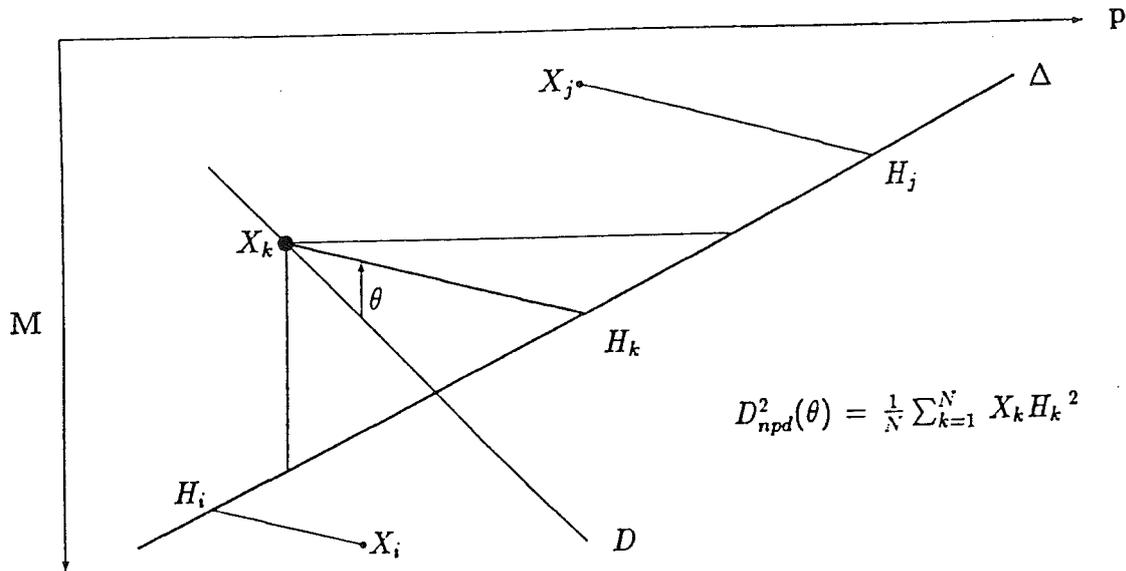


Figure B.2 : Nous définissons ici une famille de "distance d'un nuage de points à une droite" (D_{npd}) dépendant du paramètre θ , l'angle entre la bissectrice D et le segment XH , H appartenant à la droite de régression Δ . Pour un échantillon de N points $\{X_k = (p_k, M_k)\}$, $D_{npd}^2(\theta) = \frac{1}{N} \sum_{k=1}^N X_k H_k^2$. Les coefficients de la droite de régression sont alors tels que $D_{npd}(\theta)$ soit minimale

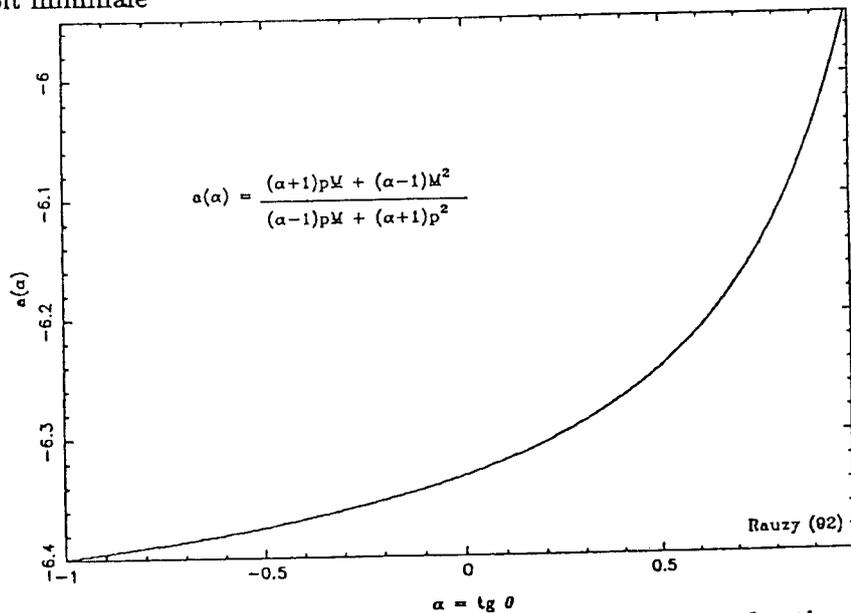


Figure B.3 : Variation de la pente a de la droite de régression en fonction de $\alpha = \text{tg } \theta$ (figure B.2) pour un échantillon donné. $a(\alpha) = \frac{(\alpha+1)\overline{p.M} + (\alpha-1)\overline{M^2}}{(\alpha-1)\overline{p.M} + (\alpha+1)\overline{p^2}}$, $\overline{M^2}$, $\overline{p.M}$ et $\overline{p^2}$ définis aux équations (B.18b, B.19b, B.20). L'estimation de a dépend de la "distance du nuage de points à la droite" utilisée.

Ainsi, la donnée du nuage de points n'est pas suffisante pour déterminer les coefficients de la droite de régression puisque l'estimation de leurs valeurs est dépendante du choix de la distance de régression (ou de l'hypothèse faite sur la distribution des erreurs). En fait, il faudrait connaître la façon dont a été engendré le nuage de points afin d'accéder à la vraie valeur des coefficients a et b . Hélas, la physique sous-jacente à la relation (B.3) est actuellement mal connue, et l'information sur le processus générateur du nuage de points observé n'est pas accessible. On est donc obligé de choisir arbitrairement une distance de régression, en espérant que l'erreur introduite par ce choix n'est pas importante.

Heureusement, nous montrerons par la suite que cette notion de vrais coefficients a et b n'est pas fondamentale. Le problème peut être mieux compris si l'on utilise un formalisme statistique plus général que celui des droites de régression. Dans la section I.3. , nous introduisons ce formalisme, en établissant une correspondance entre processus générateur d'un échantillon et modèle statistique associé.

I.2.2 Le problème des "biais" statistiques :

Les échantillons de galaxies observés sont souvent soumis à des effets de sélection. Certains effets de sélection, comme la limitation en magnitude apparente ($m < m_{lim}$) sont facilement modélisables. D'autres, par exemple le critère de sélection pour le paramètre physique observable p , sont encore assez mal connus. Le traitement d'échantillons de galaxies possédant des effets de sélection s'avère difficile, notamment avec la technique employée classiquement (méthode de régression). On sait par exemple que pour un échantillon limité en magnitude apparente, la calibration de la relation (B.3) au moyen de la régression verticale (l'erreur intrinsèque est sur les M_k) est biaisée (Teerikorpi (1987)). Des biais d'une nature différente interviennent de plus lorsqu'il s'agit d'estimer la distance la plus probable d'une galaxie connaissant la valeur de son paramètre p et de sa magnitude apparente m (Teerikorpi (1975), Teerikorpi (1984) ou Gouguenheim et *al.*). Le traitement de ces biais est complexe et est abordé dans la littérature de différentes façons, souvent empiriquement.

Nous proposons dans la section suivante de clarifier cette situation en introduisant une formalisation statistique rigoureuse du problème, qui généralise la méthode classique de régression. Nous montrons les avantages de ce formalisme, notamment en présence d'effets de sélection pour l'échantillon.

I.3 Formalisation statistique du problème :

Nous introduisons ici le formalisme des densités de probabilité afin de décrire le problème posé par la calibration des indicateurs secondaires de distance ainsi que par l'évaluation de la distance la plus probable d'une galaxie. Ce formalisme fait apparaître l'existence de plusieurs modèles statistiques possibles afin de traiter l'échantillon.

Nous établissons une correspondance entre ces différents modèles statistiques et les différentes définitions de "distance d'un nuage de points à une droite" exposées précédemment dans la section I.2.1 . Nous avons montré que sans une information supplémentaire sur la physique du phénomène, l'estimation des coefficients a et b n'était pas unique mais dépendait du modèle statistique adopté. Actuellement, aucune raison physique ne nous incite à choisir un modèle statistique plutôt qu'un autre.

En toute rigueur, il n'existe qu'un seul modèle statistique capable de décrire l'échantillon réellement observé, et le fait de traiter cet échantillon avec un "mauvais" modèle statistique introduit une erreur dans l'estimation des distances. Néanmoins, nous montrons que cette erreur s'avère négligeable comparée à l'incertitude intrinsèque présente dans la relation (B.3) par exemple. En pratique donc, on pourra choisir le modèle statistique le plus apte à traiter les données observationnelles, et nous verrons plus loin que cette propriété est très précieuse lorsqu'il s'agit de traiter des échantillons soumis à des effets de sélection.

I.3.1 Le formalisme statistique des densités de probabilité :

Nous nous proposons donc de décrire l'échantillon de galaxies observées en utilisant le formalisme mathématique des densités de probabilité (voir Bigot et *al.* (1990a)). Commençons tout d'abord par citer les variables aléatoires impliquées dans cette modélisation:

- La magnitude absolue M
- La magnitude apparente m
- Le paramètre physique observé p
- Le module de distance μ précédemment défini (équation (B.2))
- La variable aléatoire centrée ζ introduite dans l'équation (B.3), ζ étant en fait l'erreur intrinsèque présente dans la relation $M = a.p + b$

Nous posons maintenant les hypothèses suivantes :

. Hypothèse 1 : Il existe une relation liant les variables aléatoires M , p et ζ de la forme $M = a.p + b + \zeta$ telle que ζ soit une variable aléatoire centrée (indicateurs secondaires de distance). Pour des raisons de simplicité, nous ferons l'hypothèse que ζ est distribuée gaussiennement autour de 0 avec une dispersion σ_ζ . Nous introduisons ici la fonction $g(x, x_0, \sigma_x)$:

$$g(x, x_0, \sigma_x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x - x_0)^2}{2\sigma_x^2}\right) \quad (B.7)$$

Par la suite donc, nous admettrons que la densité de probabilité de la variable ζ est $g(\zeta, 0, \sigma_\zeta) d\zeta$.

. Hypothèse 2 : Les N galaxies $\{G_k\}$ de l'échantillon sont des événements indépendants, ainsi la densité de probabilité de l'échantillon s'écrit : $dP_{\text{échantillon}} = \prod_{k=1}^N dP(G_k)$

. Hypothèse 3 : Nous admettrons de plus que la population des galaxies étudiées n'est soumise à aucune évolution en luminosité, c'est à dire que les variables aléatoires M et p ne sont pas corrélées au module des distances μ . Cette hypothèse semble bien vérifiée dans la plage des distances inspectées (typiquement de 1 à 100 Mpc).

Nous sommes maintenant en mesure d'écrire la densité de probabilité pour une galaxie de l'échantillon. Nous noterons $dP_{\text{théorique}}$ la densité de probabilité en l'absence d'effets de sélection et $dP_{\text{observée}}$ la densité de probabilité réellement observée :

$$dP_{\text{théorique}} = \frac{1}{A_{\text{théorique}}} F(M, p) dM dp h(\mu) d\mu \quad (B.8)$$

- . $F(M, p)$ la fonction de distribution des variables M et p .
- . $h(\mu)$ la fonction de distribution spatiale de l'échantillon.
- . $A_{\text{théorique}}$ le facteur de normalisation.

$$A_{\text{théorique}} = \int F(M, p) h(\mu) dM dp d\mu \quad (B.9)$$

Mais la densité de probabilité observée est soumise aux effets de sélection :

$$dP_{\text{observée}} = \frac{1}{A_{\text{observée}}} \phi(m, p, \mu) F(M, p) dM dp h(\mu) d\mu \quad (\text{B.10})$$

. $\phi(m, p, \mu)$ la fonction de sélection dépendant des variables m , p et μ sous sa forme la plus générale possible. Si l'échantillon est soumis à un cut-off en magnitude apparente, la fonction de sélection s'écrit $\phi(m, p, \mu) = \Theta(m_{\text{lim}} - m)$ avec Θ la distribution de Heaveside et m_{lim} la magnitude apparente de coupure.

. $A_{\text{observée}}$ le facteur de normalisation pour la densité de probabilité réellement observée.

$$A_{\text{observée}} = \int \phi(m, p, \mu) F(M, p) h(\mu) dM dp d\mu \quad (\text{B.11})$$

Il s'agit maintenant d'introduire la variable ζ dans l'expression de la densité de probabilité $F(M, p) dM dp$. L'hypothèse 1 (la densité de probabilité de ζ s'exprime $g(\zeta, 0, \sigma_\zeta) d\zeta$) nous assure qu'il existe une variable aléatoire u telle que :

$$F(M, p) dM dp = f_u(u) du g(\zeta, 0, \sigma_\zeta) d\zeta \quad (\text{B.12})$$

. avec $u = u(M, p)$, une variable aléatoire s'exprimant en fonction de M et p telle que u et ζ soient des variables aléatoires indépendantes.

Nous voulons préciser la signification de la variable aléatoire u précédemment introduite. Si nous désirons simuler un échantillon de N galaxies vérifiant les hypothèses 1, 2 et 3 définies dans cette section, il faut engendrer simultanément deux distributions de N valeurs. D'une part, les valeurs ζ_k de la variable ζ dont la distribution doit vérifier l'hypothèse 1. D'autre part, les valeurs u_k d'une autre variable aléatoire u nécessairement indépendante de ζ . On peut ensuite calculer pour chaque événement k , les valeurs M_k et p_k vérifiant les deux équations $[M_k = a.p_k + b + \zeta_k]$ et $[u_k = u(M_k, p_k)]$. On voit alors que des fonctions $u(M, p)$ différentes vont engendrer des nuages de N points (M_k, p_k) différents. Ainsi, l'expression de la variable aléatoire u en fonction de M et p définit un processus générateur de nuages de points, particulier pour chaque u . Nous appellerons modèle statistique $[u(M, p)]$ l'expression de la densité de probabilité exprimée en (B.10) avec $u = u(M, p)$ dans l'équation (B.11). Lorsqu'on traite un échantillon avec un modèle statistique particulier, on fait donc implicitement l'hypothèse que cet échantillon a été engendré par le processus générateur associé à ce modèle statistique. Nous allons voir dans

la suite qu'il existe une correspondance entre le choix d'un modèle statistique et le choix d'une distance de régression.

I.3.2 Correspondance entre modèles statistiques et droites de régression :

Nous voyons qu'il existe plusieurs modèles statistiques, complètement définis par l'expression de la fonction $u = u(M, p)$. Le formalisme des densités de probabilité précédemment établi en I.3.1 étant une généralisation de la méthode de régression, on peut alors associer à chaque définition de la "distance d'un nuage de points à une droite", un modèle statistique.

Nous pouvons par exemple, faire correspondre à chaque "distance" dépendant du paramètre θ (définie figure (B.2) en I.2.1), le modèle statistique $[u(M, p) = \frac{1}{2}((1 + \operatorname{tg}\theta)M + (1 - \operatorname{tg}\theta)p)]$. De même, l'estimation de (a, b) pour le modèle statistique $[u(M, p) = M]$ (nous faisons donc l'hypothèse que les variables aléatoires M et ζ sont indépendantes), correspond à une régression suivant l'horizontale dans le diagramme $\{M, p\}$ (relation inverse). A la relation directe (c'est à dire la régression verticale dans le diagramme $\{M, p\}$), on peut associer le modèle statistique $[u(M, p) = p]$ (p et ζ indépendantes).

Ainsi, l'introduction de la fonction $u(M, p)$ définit complètement le modèle statistique rendant compte de la relation $M = a.p + b + \zeta$. De plus, ceci nous permettra d'estimer le couple de constantes (a, b) lorsque les méthodes classiques de régression sont inutilisables, par exemple lorsque l'échantillon étudié est soumis à des effets de sélection. Nous appliquerons pour cela la technique générale du maximum de vraisemblance.

I.3.3 Le choix du modèle statistique :

L'expression de $u(M, p)$ va donc déterminer le modèle statistique avec lequel nous allons estimer par la suite le couple de constantes (a, b) et la distance la plus probable d'une galaxie.

Pour un échantillon donné de galaxies, il existe un seul modèle statistique capable de décrire la vraie distribution de ses galaxies. Le problème auquel nous devons faire

face lorsqu'il s'agit d'analyser un échantillon observé, c'est l'ignorance de ce modèle. Il se peut que pour des raisons physiques particulières, un modèle soit mieux adapté qu'un autre. Mais notre connaissance de la physique sous-jacente aux indicateurs secondaires de distance est réduite, et ne peut nous fournir actuellement un tel modèle. En traitant les données observationnelles avec un modèle statistique quelconque, nous commettons donc théoriquement une erreur. Nous allons montrer à l'aide d'un exemple que cette erreur est en fait négligeable lorsqu'il s'agit de déterminer la distance des galaxies (Triay et *al.*, en préparation).

Faisons l'hypothèse que le processus physique générateur de la relation (B.3), $M = a.p + b + \zeta$, implique que ce sont les variables aléatoires M et ζ qui sont indépendantes. Le seul modèle statistique capable d'engendrer cette distribution des galaxies est alors le modèle [$u(M, p) = M$], et la densité de probabilité décrivant l'échantillon s'exprime alors :

$$F(M, p) dM dp = f_M(M) dM g(\zeta, 0, \sigma_\zeta) d\zeta \quad (B.13)$$

Si nous traitons maintenant cette échantillon avec le modèle [$u(M, p) = p$], nous allons alors imposer implicitement une forme à la fonction $f_M(M)$. En effet, on contraint la densité de probabilité de l'échantillon à s'écrire de la façon suivante :

$$F(M, p) dM dp = f_p(p) dp g(\zeta, 0, \sigma_\zeta) d\zeta \quad (B.14)$$

Les équations (B.13) et (B.14) impliquent alors :

$$f_p(p) = a. \int f_M(M) g(\zeta, 0, \sigma_\zeta) dM$$

$$f_M(M) = \int f_p(p) g(\zeta, 0, \sigma_\zeta) dp$$

d'où :

$$f_M(M) = a. \int f_M(X) g(M - a.p - b, 0, \sigma_\zeta) g(X - a.p - b, 0, \sigma_\zeta) dp dX$$

Et donc :

$$f_M(M) = \frac{1}{a} \int f_M(X) g(p, (M - b)/a, \sigma_\zeta/a) g(p, (X - b)/a, \sigma_\zeta/a) dp dX$$

En intégrant sur p :

$$f_M(M) = \frac{1}{a} \int f_M(X) g((X - b)/a, (M - b)/a, \sqrt{2}\sigma_\zeta/a) dX$$

Enfin :

$$f_M(M) = \int f_M(X) g(X, M, \sqrt{2} \sigma_\zeta) dX \quad (B.15)$$

Nous voyons donc que la contrainte imposée à la fonction de luminosité $f_M(M)$ est une contrainte de régularité. En effet, il suffit que la fonction de luminosité de l'échantillon soit relativement lisse (à l'échelle $\sqrt{2} \sigma_\zeta$), pour que l'équation (B.15) soit quasiment vérifiée. L'erreur commise est en tout cas négligeable pour les échantillons de galaxies réellement observées (la fonction de luminosité $f_M(M)$ peut être approchée par une gaussienne de dispersion $\sigma_M \approx 2.5 \text{ mg}$, à comparer avec la dispersion intrinsèque de la relation $M = a.p + b + \zeta$, $\sigma_\zeta \approx 0.5 \text{ mg}$ (Fouqué et al. (1990))).

Le traitement de l'échantillon avec un modèle statistique quelconque plutôt qu'avec le vrai modèle correspondant au processus physique sous-jacent introduit donc une erreur dans l'estimation des distances. Toutefois, nous montrons à l'aide de l'exemple précédent que pour les données observationnelles disponibles, l'approximation faite est raisonnable. La méconnaissance du processus physique impliqué dans les indicateurs secondaires de distance n'est donc pas importante. Ainsi, la possibilité nous est offerte de choisir le modèle statistique le plus approprié afin de traiter les données. Nous verrons notamment que lorsque l'échantillon de galaxies est soumis à des effets de sélection, le choix du modèle statistique est fondamental.

I.4. Phase de calibration des indicateurs secondaires de distance :

Il existe deux procédures possibles afin de calibrer les indicateurs secondaires de distances : soit les coefficients a et b de la relation (B.3) sont estimés directement à partir des galaxies dont on connaît la magnitude absolue M par l'intermédiaire des indicateurs primaires de distance, soit on détermine a dans les amas de galaxies et b est fixé par la distance qui nous sépare d'un amas particulier. Dans la littérature, la méthode des droites de régression est utilisée dans les deux cas.

Nous présentons ici la technique du maximum de vraisemblance qui permet d'estimer les coefficients a et b dans le formalisme statistique introduit précédemment en I.3.1 .

Nous montrons à l'aide d'exemples que contrairement à la méthode de régression

classiquement employée, notre formalisation statistique du problème rend possible la détermination des valeurs de a et b pour des échantillons de galaxies soumis à des effets de sélection non triviaux.

De plus, nous établissons que ce sont en fait les effets de sélection qui nous inciteront à utiliser un modèle statistique plutôt qu'un autre afin de traiter l'échantillon (le choix du modèle étant possible d'après la discussion en I.3.3).

I.4.1. Les procédures de calibration généralement adoptées :

Si on veut utiliser correctement les indicateurs secondaires de distance, il est nécessaire que l'échantillon à partir duquel on calibre la relation (B.3), $M = a.p + b + \zeta$, possède les mêmes propriétés statistiques que l'ensemble des galaxies dont on cherche à déterminer les distances les plus probables. Or, nous savons que les mesures des quantités impliquées dans la relation (B.3) sont entachées d'erreurs et que ces erreurs de mesure varient hélas suivant les techniques d'observations employées. Ainsi, les indicateurs secondaires de distances ne sont valides que lorsque qu'ils sont appliqués à des données observationnelles homogènes, c'est à dire souffrant du même type d'erreurs de mesure (Fouqué et *al.* (1990)). L'inhomogénéité des données observationnelles est certainement l'un des principaux problèmes rencontrés lors de l'évaluation de la distance des galaxies.

A) A PARTIR DES INDICATEURS PRIMAIRES DE DISTANCE :

L'estimation des coefficients a et b peut se faire à partir des galaxies proches dont on a déterminé les magnitudes absolues à partir des indicateurs primaires de distance. Cette calibration a été effectuée à de nombreuses reprises (par exemple Aaronson et *al.* (1986), Bottinelli et *al.* (1986), Kraan-Korteweg et *al.* (1988), Fouqué et *al.* (1990) ou Pierce et *al.* (1992)) et les valeurs de a et b varient suivant le choix des galaxies considérées et les magnitudes absolues que les auteurs leurs attribuent.

Le problème du choix de la droite de régression est abordé (Bottinelli et *al.* (1986), Lynden-Bell et *al.* (1988) ou Fouqué et *al.* (1990)) mais il est rare de trouver dans la littérature une discussion approfondie sur le sujet. Il semble pourtant que la droite de régression verticale (l'erreur intrinsèque de la relation $M = a.p + b$ est sur M , ce qui correspond aussi à utiliser le modèle statistique [$u(M, p) = p$]) soit la plus fréquemment

employée. En fait, puisque les indicateurs primaires de distance fournissent la valeur de la magnitude absolue M des galaxies proches avec une incertitude de 10 % environ, il paraît naturel d'utiliser la relation directe qui postule justement que les erreurs viennent de M .

Néanmoins, nous voulons souligner que l'incertitude intrinsèque (la variable ζ) présente dans les indicateurs secondaires de distance s'ajoute à cette erreur sur la valeur de M . De plus les erreurs de mesure sur M ne sont pas de même nature que l'incertitude intervenant dans la relation (B.3). A ce titre, ces deux types d'erreurs réclament des traitements statistiques radicalement différents (Triay et *al.*, en préparation). Cet argument est ainsi trompeur et ne peut en aucun cas justifier l'utilisation de la relation de Tully-Fisher directe afin de déterminer les valeurs des coefficients a et b . De surcroît, nous avons montré en I.3.3 que l'erreur commise en traitant le problème avec un modèle statistique (ou une distance de régression) quelconque était négligeable.

B] A PARTIR DES AMAS DE GALAXIES :

Une procédure différente permet d'estimer a d'après les données observationnelles que nous possédons pour les amas de galaxies (Dressler et *al.* (1987) par exemple). En effet, il est raisonnable de supposer que les galaxies groupées dans un amas possèdent toutes la même distance. On peut alors remplacer dans la relation (B.3) la magnitude absolue M d'une galaxie par sa magnitude apparente m . Ainsi pour toutes les galaxies d'un même amas, on a :

$$m = a.p + b_1 + \zeta \quad (B.16)$$

$$\text{avec } b_1 = b - \mu_{\text{amas}}$$

et μ_{amas} le module de distance de l'amas

On voit donc qu'il est possible d'appliquer tout le formalisme précédemment développé dans la section I.3, cette fois-ci en utilisant la relation (B.16).

Les données observationnelles actuellement disponibles concernent en fait plusieurs amas. Il faudra alors calculer un a moyen, tenant compte des estimations de a dans chaque amas individuel.

Il faut cependant rester vigilant lorsqu'on évalue a car la sélection des galaxies dans les amas se fait en magnitude apparente m . L'estimation de la constante a est alors

soumise à un biais, plus connu sous le nom de biais de Malmquist (Teerikorpi (1975), Teerikorpi (1984) ou Gouguenheim et *al.*). L'importance de cet effet varie suivant les auteurs, l'erreur commise étant négligeable pour certains (la sélection dans les amas se fait surtout en fonction du module de distance μ (Aaronson et *al.* (1986), Pierce et *al.* (1988)), appréciable pour d'autres (par exemple Teerikorpi (1987)). Dans le cas où l'échantillon est libre d'effets de sélection en p , l'utilisation de la relation inverse (c'est à dire une régression horizontale ou l'emploi du modèle statistique [$u(M, p) = M$]) est alors recommandée, puisque elle est insensible aux effets de sélection en magnitude apparente m (Schechter (1980), Teerikorpi (1990)).

le coefficient a étant connu, il suffit alors seulement de quelques galaxies dont on connaît la distance pour que la statistique liée à l'estimation de b soit solide. Cependant, il faut remarquer que c'est dans cette étape de calibration que se répercutent les erreurs intrinsèques contenues dans la phase primaire de calibration (Pierce et *al.* (1992)). C'est pourquoi certains auteurs (Lynden-Bell et *al.* (1988)) préfèrent exprimer les distances des galaxies en fonction de la distance d'un amas particulier. Ceci évite en fait l'utilisation des indicateurs primaires de distance. Toutefois, la détermination de la valeur de b au moyen des indicateurs primaires de distance s'avère nécessaire, par exemple pour le calcul de la constante de Hubble H_0 .

I.4.2. La technique du maximum de vraisemblance :

Nous présentons ici la technique générale du maximum de vraisemblance (voir Bigot et *al.* (1990b)) qui nous permettra d'estimer la valeur des coefficients a et b , après avoir fixé le modèle statistique (la fonction $u(M, p)$) ainsi que la fonction de sélection ($\phi(m, p, \mu)$). Ce calcul est réalisé dans le cadre du formalisme statistique introduit en I.3.1 .

La densité de probabilité en fonction des variables observées s'exprime à partir de l'équation (B.10) de la façon suivante :

$$dP_{\text{observée}} = \frac{1}{A_{\text{observée}}} \phi(m, p, \mu) f_u(u(m - \mu, p)) g(m - \mu - a.p - b, 0, \sigma_\zeta) h(\mu) \cdot \mathcal{J}(m, p, \mu) dm dp d\mu \quad (B.17)$$

avec $\mathcal{J} = \left| \frac{\partial u}{\partial p} + a \cdot \frac{\partial u}{\partial m} \right|$, le Jacobien du changement de variable.

Nous avons alors accès au logarithme de la fonction de vraisemblance :

$$\mathcal{L}(a, b, \sigma_\zeta) = -\ln A_{\text{observée}} + \frac{1}{N} \sum_{k=1}^N \ln f_u(u(m_k - \mu_k, p_k)) + \ln \mathcal{J} + \frac{1}{N} \sum_{k=1}^N \ln h(\mu_k) - \ln \sqrt{2\pi} \sigma_\zeta - \frac{1}{2N\sigma_\zeta^2} \sum_{k=1}^N (a \cdot p_k + b - (m_k - \mu_k))^2 \quad (B.18)$$

Et les équations de vraisemblance ($\partial \mathcal{L} / \partial a = 0$, $\partial \mathcal{L} / \partial b = 0$, $\partial \mathcal{L} / \partial \sigma_\zeta = 0$) impliquent :

$$\frac{1}{N\sigma_\zeta^2} \sum_{k=1}^N p_k \cdot (a \cdot p_k + b - (m_k - \mu_k)) = \frac{\partial}{\partial a} (\ln \mathcal{J}) - \frac{\partial}{\partial a} (\ln A_{\text{observée}}) \quad (B.19)$$

$$\frac{1}{N\sigma_\zeta^2} \sum_{k=1}^N (a \cdot p_k + b - (m_k - \mu_k)) = \frac{\partial}{\partial b} (\ln \mathcal{J}) - \frac{\partial}{\partial b} (\ln A_{\text{observée}}) \quad (B.20)$$

$$\frac{1}{N\sigma_\zeta^3} \sum_{k=1}^N (a \cdot p_k + b - (m_k - \mu_k))^2 = \frac{1}{\sigma_\zeta} + \frac{\partial}{\partial \sigma_\zeta} (\ln A_{\text{observée}}) \quad (B.21)$$

1.4.3. Application pour différents modèles statistiques :

Nous proposons trois exemples simples de modèles statistiques pour lesquels on estimera le couple de constantes (a, b) par la technique du maximum de vraisemblance. Nous montrons aussi quels critères nous inciteront à choisir un modèle statistique plutôt qu'un autre, par exemple en présence d'effets de sélection.

A] MODELE STATISTIQUE [$u(M, p) = M$]
ET FONCTION DE SELECTION $\phi(m, p, \mu) = \phi(m, \mu)$

Nous nous proposons d'étudier un échantillon de galaxies soumis à des effets de sélection ne dépendant que de la magnitude apparente m et du module de distance μ . Cet exemple est intéressant puisque les échantillons de galaxies observés possèdent généralement un cut-off en magnitude apparente, que l'on modélise souvent par une distribution de Heaveside $\phi(m, p, \mu) = \Theta(m_{\text{lim}} - m)$, m_{lim} étant la magnitude apparente

limite de coupure.

D'après la discussion en I.3.3, la liberté nous est offerte d'utiliser n'importe quel modèle statistique pour traiter un tel échantillon. Nous choisissons donc le modèle statistique défini par la fonction $[u(M, p) = M]$ (régression horizontale ou relation inverse) puisqu'il simplifie notamment l'obtention du couple de constantes (a, b) . De plus, il rend les valeurs estimées de a et b indépendantes de la forme de la fonction de luminosité $f_M(M)$ et de la fonction de sélection.

Le facteur de normalisation $A_{observée}$ s'écrit :

$$A_{observée} = -a \cdot \int \phi(m, \mu) f_M(M) h(\mu) dM d\mu$$

Puisque $\mathcal{J} = 1$, les équations de vraisemblance deviennent :

$$\frac{1}{N} \sum_{k=1}^N a \cdot p_k \cdot (a \cdot p_k + b - (m_k - \mu_k)) = \sigma_{\zeta}^2$$

$$\frac{1}{N} \sum_{k=1}^N (a \cdot p_k + b - (m_k - \mu_k)) = 0$$

$$\frac{1}{N} \sum_{k=1}^N (a \cdot p_k + b - (m_k - \mu_k))^2 = \sigma_{\zeta}^2$$

Et en posant :

$$\bar{p} = \frac{1}{N} \sum_{k=1}^N p_k \qquad \overline{p^2} = \frac{1}{N} \sum_{k=1}^N (p_k - \bar{p})^2 \qquad (B.22 a, B.22 b)$$

$$\overline{M} = \frac{1}{N} \sum_{k=1}^N (m_k - \mu_k) \qquad \overline{M^2} = \frac{1}{N} \sum_{k=1}^N ((m_k - \mu_k) - \overline{M})^2 \qquad (B.23 a, B.23 b)$$

$$\overline{p \cdot M} = \frac{1}{N} \sum_{k=1}^N (p_k - \bar{p}) \cdot ((m_k - \mu_k) - \overline{M}) \qquad (B.24)$$

Les valeurs estimées des constantes a et b sont :

$$a = \frac{\overline{M^2}}{p \cdot \overline{M}} \quad (\text{B.25})$$

$$b = \overline{M} - \overline{p} \cdot \frac{\overline{M^2}}{p \cdot \overline{M}} \quad (\text{B.26})$$

Nous avons donc confirmé qu'en l'absence d'effets de sélection en p , le modèle statistique [$u = M$] est particulièrement bien adapté puisqu'il permet d'estimer les constantes a et b indépendamment de la fonction de sélection $\phi(m, \mu)$ (la relation inverse permet d'estimer a et b pour des échantillons possédant des effets de sélection en magnitude apparente m (Schechter (1980))). Il faut remarquer néanmoins que cette propriété n'est plus vraie si la fonction de sélection dépend du paramètre p .

*B) MODELE STATISTIQUE [$u(M, p) = p$]
ET FONCTION DE SELECTION $\phi(m, p, \mu) = \phi(p, \mu)$*

Les effets de sélection auxquels est soumis notre échantillon ne dépendent cette fois-ci que du paramètre p et du module de distance μ . Cet exemple est théorique puisqu'il ne tient pas compte de l'inévitable cut-off en magnitude apparente. Cependant il illustre le traitement des effets de sélection en p . Parmi les nombreuses galaxies pour lesquelles on a observé la magnitude apparente, seulement quelques unes ont une mesure du paramètre p véritablement fiable. Il semble donc que les échantillons de galaxies que nous possédons soient soumis à des effets de sélection en p actuellement mal connus (L. Bottinelli, communication privée).

Lorsque nous avons à traiter ces échantillons, c'est le modèle [$u(M, p) = p$] (relation directe ou régression verticale) qui est le mieux adapté. Comme dans l'exemple A], il ne sera alors pas nécessaire de connaître la forme de la fonction de sélection $\phi(p, \mu)$, ni celle de la fonction de distribution $f_p(p)$.

Le facteur de normalisation $A_{\text{observée}}$ s'écrit :

$$A_{\text{observée}} = \int \phi(p, \mu) f_p(p) h(\mu) dp d\mu$$

De plus, $\mathcal{J} = 1$ donc les équations de vraisemblance sont :

$$\frac{1}{N} \sum_{k=1}^N p_k \cdot (a \cdot p_k + b - (m_k - \mu_k)) = \sigma_{\zeta}^2$$

$$\frac{1}{N} \sum_{k=1}^N (a \cdot p_k + b - (m_k - \mu_k)) = 0$$

$$\frac{1}{N} \sum_{k=1}^N (a \cdot p_k + b - (m_k - \mu_k))^2 = \sigma_{\zeta}^2$$

Les valeurs estimées des constantes a et b sont donc :

$$a = \frac{\overline{p \cdot M}}{p^2} \quad (B.27)$$

$$b = \overline{M} - \overline{p} \cdot \frac{\overline{p \cdot M}}{p^2} \quad (B.28)$$

L'estimation de a et b ne dépend pas de la fonction de sélection en p et μ . Le modèle $[u = p]$ est donc bien adapté lorsqu'il s'agit de traiter des échantillons possédant des effets de sélection en p . Bien sûr, les données observationnelles semblent être soumises à des effets de sélection dépendant à la fois de la magnitude apparente m ainsi que du paramètre de distance p . Toutefois, la technique du maximum de vraisemblance nous permettra tout de même d'estimer les constantes a et b (ceci étant impossible en utilisant les méthodes usuelles de régression). Il faudra alors faire des hypothèses sur la forme de la fonction de sélection $\phi(m, p, \mu)$ ainsi que choisir le modèle statistique le plus apte à traiter les données.

*C) MODELE STATISTIQUE $[u(M, p) = p]$
ET FONCTION DE SELECTION $\phi(m, p, \mu) = \Theta(m_{lim} - m)$*

L'échantillon que l'on veut traiter possède une magnitude apparente limite de coupure m_{lim} . Nous proposons de le traiter avec le modèle statistique $[u(M, p) = p]$.

Cet exemple est intéressant pour plusieurs raisons. D'une part, il montre l'avantage de la technique du maximum de vraisemblance par rapport aux méthodes de régression

traditionnellement employées. En effet, la technique du maximum de vraisemblance va nous permettre d'estimer les valeurs des constantes a et b pour le modèle $[u = p]$ en présence d'un cut-off en magnitude apparente m . D'autre part, cet exemple illustre les différentes approches possibles lorsqu'il s'agit de calibrer la relation (B.3). Nous pourrions ainsi comparer les techniques développées dans les exemples A], B] et C].

Il s'agit donc de déterminer a et b pour le modèle statistique $[u = p]$, dans le cas où $\phi(m, p, \mu) = \Theta(m_{lim} - m)$. Cet exemple est donc différent des exemples A] et B] puisque le facteur de normalisation $A_{observée}$ va cette fois-ci dépendre de a et b . C'est une illustration typique des cas que l'on nomme "biaisés" dans la littérature (utilisation de la relation direct avec un cut-off en m). Pour estimer a et b nous serons alors obligés de faire des hypothèses sur la fonction de distribution $f_p(p)$ ainsi que sur la distribution spatiale de l'échantillon $h(\mu) d\mu$.

Hypothèse 1 : La fonction de distribution en p est une gaussienne de moyenne p_0 et de dispersion σ_p :

$$f(p) dp = g(p, p_0, \sigma_p) dp \quad (B.29)$$

Hypothèse 2 : La distribution spatiale de l'échantillon peut s'écrire sous la forme $\rho(r) dr \propto r^\beta dr$ (cas homogène $\beta = 2$) et en posant :

$$\alpha = \frac{(\beta + 1) \ln 10}{5} \quad (B.30)$$

La distribution en fonction du module de distance μ s'exprime :

$$h(\mu) d\mu \propto e^{\alpha\mu} d\mu \quad (B.31)$$

Nous allons maintenant calculer $A_{observée}$. D'après l'équation (B.11) :

$$A_{observée} = \int g(p, p_0, \sigma_p) \Theta(m_{lim} - m) dp dm g(\mu, m - a.p - b, \sigma_\zeta) \exp(\alpha\mu) d\mu$$

En intégrant sur μ :

$$A_{observée} = \int \Theta(m_{lim} - m) dm g(p, p_0, \sigma_p) \exp\left(-\alpha p + \frac{\alpha^2 \sigma_\zeta^2}{2} + \alpha(m - b)\right) dp$$

Puis sur p :

$$A_{observée} = \int \Theta(m_{lim} - m) \exp\left(\frac{\alpha^2 \sigma_\zeta^2}{2} + \frac{a^2 \alpha^2 \sigma_p^2}{2} + \alpha(m - ap_0 - b)\right) dm$$

Et enfin en intégrant sur m :

$$A_{\text{observée}} = \frac{1}{\alpha} \exp \left[\alpha (m_{\text{lim}} - ap_0 - b) + \frac{\alpha^2}{2} \sigma_{\zeta}^2 + \frac{a^2 \alpha^2}{2} \sigma_p^2 \right] \quad (\text{B.32})$$

Nous pouvons alors déduire les équations de vraisemblance d'après les équations (B.19), (B.20) et (B.21) en ajoutant la condition supplémentaire $\partial \mathcal{L} / \partial p_0 = 0$. Nous obtenons alors le système d'équations suivant :

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^N p_k \cdot (a \cdot p_k + b - (m_k + \mu_k)) &= (\alpha p_0 - a \alpha^2 \sigma_p^2) \sigma_{\zeta}^2 \\ \frac{1}{N} \sum_{k=1}^N (a \cdot p_k + b - (m_k + \mu_k)) &= \alpha \sigma_{\zeta}^2 \\ \frac{1}{N} \sum_{k=1}^N p_k &= p_0 - a \alpha \sigma_p^2 \end{aligned}$$

Ce qui nous donne avec les notations définies précédemment en (B.22), (B.23) et (B.24):

$$a = \frac{\overline{p \cdot M}}{\overline{p^2}} \quad (\text{B.33})$$

$$b = \overline{M} - a \cdot \overline{p} + \alpha \sigma_{\zeta}^2 \quad (\text{B.34})$$

L'estimation des valeurs des constantes a et b est donc différente suivant la technique utilisée afin de traiter les données.

I.4.4. Comparaison des différents modèles statistiques :

En fait, nous pouvons tirer deux conclusions des résultats obtenus dans les exemples A), B) et C) de la section I.4.3 .

A) LES VALEURS DE a ET b VARIENT AVEC LE MODELE EMPLOYE :

Nous avons simulé figure (B.4) un échantillon de 50 galaxies engendré à partir du modèle $[u(M, p) = M]$ et soumis à un cut-off en magnitude apparente $\phi(m, p, \mu) = \Theta(m_{lim} - m)$. Le processus physique sous-jacent est alors connu, puisque on impose aux variables aléatoires M et ζ d'être indépendantes. On va maintenant analyser cet échantillon avec les techniques développées dans les exemples A] et C] (section I.4.3).

L'analyse avec le modèle $[u = M]$ permet de retrouver les "vraies" valeurs des constantes a et b introduites dans la phase de simulation de l'échantillon. Ceci n'est pas étonnant puisque l'analyse $[u = M]$ est insensible aux effets de sélection dépendant de la magnitude apparente m ainsi que du module de distance μ .

L'analyse avec le modèle $[u = p]$ et en supposant un cut-off en magnitude apparente $\phi(m, p, \mu) = \Theta(m_{lim} - m)$ détermine des valeurs de a et b différentes, mais nous avons montré que ces valeurs n'en étaient pas pour autant inutilisables afin de calibrer la relation (B.3). Néanmoins, nous remarquons que l'analyse $[u = M]$ est plus apte à traiter ce genre d'échantillon puisqu'elle ne fait aucune hypothèse sur la forme de la fonction de sélection ainsi que sur la fonction de distribution en p .

La figure (B.4) confirme donc le résultat précédemment établi, à savoir que les valeurs des constantes a et b varient suivant le modèle statistique employé.

B] LE CHOIX DU MODELE EST FIXE PAR LES EFFETS DE SELECTION :

La figure (B.5) présente un échantillon de 50 galaxies cette fois-ci engendré à partir du modèle $[u(M, p) = p]$, les effets de sélection étant les mêmes que pour la figure (B.4), $\phi(m, p, \mu) = \Theta(m_{lim} - m)$. Nous allons appliquer sur cet échantillon les méthodes des exemples B] et C] (section I.4.3).

L'analyse avec le modèle $[u = p]$ et un cut-off en magnitude apparente $\phi(m, p, \mu) = \Theta(m_{lim} - m)$ permet donc bien de retrouver les valeurs des constantes a et b .

Par contre, l'analyse avec le modèle $[u = p]$ et les effets de sélection sont supposés ne pas dépendre de la magnitude apparente m ($\phi(m, p, \mu) = \phi(p, \mu)$), va nous donner une estimation du couple (a, b) biaisée. Le modèle statistique a beau être le même que celui ayant servi à engendrer l'échantillon, la technique introduite dans l'exemple B] s'avère incapable de traiter correctement les données.

Il est donc possible de conclure d'après la figure (B.5) que ceux sont les effets de sélection auxquels est soumis l'échantillon qui vont orienter le choix de la méthode à

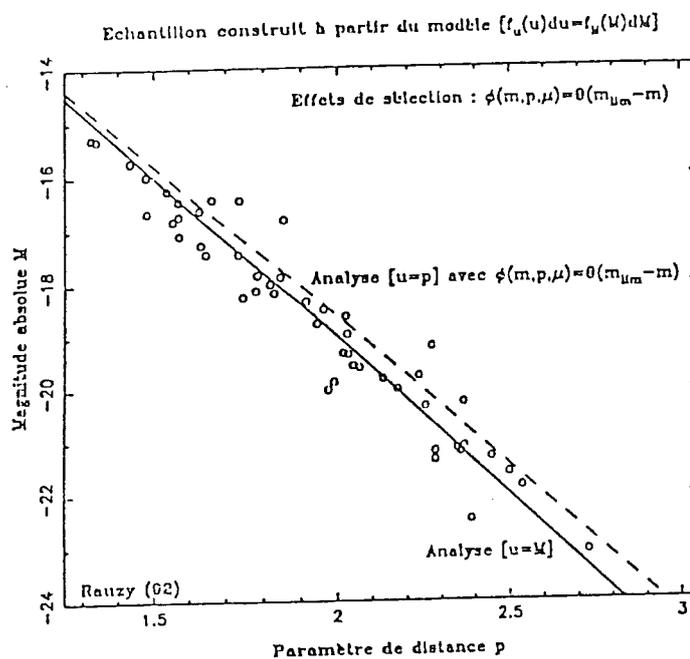


Figure B.4 : L'estimation de a et b varie suivant le modèle statistique employé.

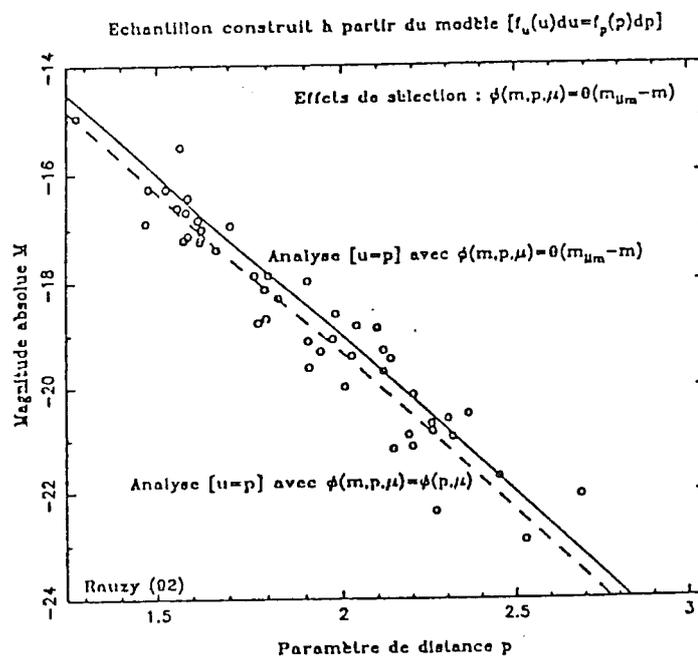


Figure B.5 : Ceux sont les effets de sélection qui vont déterminer le choix de la technique de calibration.

utiliser.

Nous avons donc montré l'existence de plusieurs couples de constantes (a, b) . Afin d'estimer la valeur de ces constantes, il est nécessaire de définir un modèle statistique décrivant la structure génératrice de l'échantillon étudié. Modulo une erreur négligeable lorsqu'il s'agit de déterminer la distance des galaxies, nous sommes libres de choisir ce modèle statistique. Les exemples A], B] et C] de la section I.4.3 nous apprennent qu'en présence d'effets de sélection particuliers, certains modèles statistiques sont privilégiés si l'on souhaite traiter correctement les données observationnelles mises à notre disposition. La technique générale du maximum de vraisemblance est sans doute la mieux appropriée pour ce genre d'étude.

I.5. Estimation de la la distance la plus probable d'une galaxie :

Nous avons vu dans la section I.4 comment calibrer la relation (B.3), $M = a.p + b + \zeta$. Le problème est maintenant de déterminer la distance la plus probable d'une galaxie après avoir adopté un modèle statistique avec lequel nous avons évalué les valeurs des constantes a et b . Nous rappelons que les paramètres effectivement mesurés sont la magnitude apparente m ainsi que le paramètre p . Il s'agit donc de calculer la valeur moyenne de la distance r ou du module de distance μ , pour m_o et p_o observés. La densité de probabilité pour m_o et p_o fixés s'écrit d'après l'équation (B.10) :

$$dP_{m_o, p_o} = \frac{1}{A_{m_o, p_o}} \delta(m - m_o) \delta(p - p_o) F(m - \mu, p) dp dm h(\mu) \phi(m, p, \mu) d\mu \quad (B.35)$$

avec $\delta(x - x_0)$ la distribution de Dirac

Et le facteur de normalisation A_{m_o, p_o} vérifie :

$$A_{m_o, p_o} = \int F(m_o - \mu, p_o) h(\mu) \phi(m_o, p_o, m_o - \mu) d\mu \quad (B.36)$$

La distance la plus probable $\langle r \rangle_{m_o, p_o}$ d'une galaxie pour laquelle on a mesuré la magnitude apparente m_o et le paramètre p_o sera alors :

$$\langle r \rangle_{m_o, p_o} = \int r dP_{m_o, p_o} \quad (B.37)$$

Nous allons calculer la distance la plus probable d'une galaxie en employant le modèle statistique [$u = M$] ainsi que le modèle [$u = p$]. Pour rester cohérent, il faut bien entendu choisir le même modèle statistique dans la phase de calibration et dans la phase d'obtention de la distance la plus probable.

Nous sommes obligés de faire des hypothèses sur la distribution spatiale de l'échantillon $h(\mu)$ ainsi que sur la fonction de sélection $\phi(m, p, \mu)$. Dès maintenant nous posons deux hypothèses communes aux modèles statistiques [$u = M$] et [$u = p$].

Hypothèse 1 : La distribution spatiale de l'échantillon peut s'écrire sous la forme (voir équations (B.30) et (B.32)) :

$$h(\mu) d\mu \propto e^{\alpha\mu} d\mu \quad \text{avec } \alpha = \frac{(\beta + 1) \ln 10}{5}$$

Hypothèse 2 : Les effets de sélection ne dépendent pas du module de distance μ .

$$\phi(m, p, \mu) = \phi(m, p)$$

On peut remarquer néanmoins que si les données observationnelles sont rarement soumises aux effets de sélection en module de distance μ , l'hypothèse sur la forme de la distribution spatiale des galaxies ne reflète pas la réalité, les galaxies étant souvent regroupées en amas par exemple.

1.5.1. Estimation pour le modèle statistique [$u = p$] :

Nous allons calculer ici le module de distance le plus probable ainsi que la distance la plus probable d'une galaxie sachant que le modèle statistique utilisé durant la phase de calibration était [$u = p$]. Nous expliciterons aussi σ_r , l'incertitude intrinsèque présente dans la détermination de la distance d'une galaxie.

D'après l'équation (B.35), la densité de probabilité s'écrit :

$$dP_{m_o, p_o} = \frac{1}{A_{m_o, p_o}} f_p(p_o) \phi(m_o, p_o) g(\mu, m_o - a.p_o - b + \alpha \sigma_\zeta^2, \sigma_\zeta) d\mu$$

On voit donc que la distribution des μ , pour m_o et p_o fixés, est une gaussienne de dispersion σ_ζ et de moyenne $m_o - a.p_o - b + \alpha \sigma_\zeta^2$. Nous pouvons maintenant estimer le module de

distance le plus probable $\langle \mu \rangle_{m_o, p_o}$, en posant :

$$\mu_e = m_o - a.p_o - b \quad (B.38)$$

On a alors :

$$\langle \mu \rangle_{m_o, p_o} = \mu_e + \alpha \sigma_\zeta^2 \quad (B.39)$$

L'estimation du module de distance le plus probable est donc affectée d'un biais, improprement appelé dans la littérature du "type biais de Malmquist". En effet ce biais n'est pas dû au cut-off en magnitude apparente et n'appartient donc pas à la même classe de biais étudiés originellement par Malmquist (voir Triay et *al.*, en préparation). Il faut souligner que la correction de ce biais est entièrement dépendante de l'hypothèse que nous faisons sur la fonction de distribution spatiale de l'échantillon (Lynden-Bell et *al.* (1988)). Ainsi, la correction effectuée dans l'équation (B.39) est plus démonstrative qu'applicable aux échantillons de galaxies réellement observés.

Il est possible d'évaluer la distance la plus probable $\langle r \rangle_{m_o, p_o}$ en posant :

$$r_e = \exp\left(\frac{\alpha}{3}(\mu_e - 25)\right) \quad (B.40)$$

L'estimation de $\langle r \rangle_{m_o, p_o}$ donne :

$$\langle r \rangle_{m_o, p_o} = r_e \exp\left(\frac{7}{18} \alpha^2 \sigma_\zeta^2\right) \quad (B.40)$$

Nous calculons enfin, l'incertitude intrinsèque présente dans ce processus d'estimation de la distance d'une galaxie. Nous appelons σ_r cette incertitude pour une galaxie dont la distance la plus probable est $\langle r \rangle$:

$$\sigma_r = \langle r \rangle \sqrt{\exp\left(\frac{1}{9} \alpha^2 \sigma_\zeta^2\right) - 1} \quad (B.41)$$

L'erreur relative commise sur la détermination de la distance d'une galaxie au moyen des indicateurs secondaires de distance est donc constante. L'estimation de cette erreur pour les relations de Tully-Fisher et de Faber-Jackson est d'environ 0.2. Ainsi, la distance d'une galaxie lointaine ne sera connue qu'à 20 % près.

I.5.2. Estimation pour le modèle statistique [$u = M$] :

La densité de probabilité s'exprime cette fois-ci :

$$dP_{m_o, p_o} = \frac{1}{A_{m_o, p_o}} \phi(m_o, p_o) f_M(m_o - \mu) g(\mu, m_o - a.p_o - b + \alpha \sigma_\zeta^2, \sigma_\zeta) d\mu$$

Il est donc nécessaire de faire une hypothèse supplémentaire sur la fonction de luminosité $f_M(M)$. Pour des raisons de simplicité, nous prendrons une gaussienne.

Hypothèse 3 : La fonction de luminosité $f_M(M)$ est une gaussienne de moyenne M_0 et de dispersion σ_M :

$$f(M) dM = g(M, M_0, \sigma_M) dM$$

En définissant :

$$\mu_0 = \mu_e \frac{\sigma_M^2}{\sigma_M^2 + \sigma_\zeta^2} + (m_o - M_0) \frac{\sigma_\zeta^2}{\sigma_M^2 + \sigma_\zeta^2} \quad \sigma_\mu = \frac{\sigma_\zeta \sigma_M}{\sqrt{\sigma_M^2 + \sigma_\zeta^2}}$$

L'estimation du module de distance le plus probable $\langle \mu \rangle_{m_o, p_o}$ donne :

$$\langle \mu \rangle_{m_o, p_o} = \mu_0 + \alpha \sigma_\mu^2 \quad (B.42)$$

Et en posant $r_0 = \exp(\frac{\alpha}{3}(\mu_0 - 25))$, la distance la plus probable $\langle r \rangle_{m_o, p_o}$ est :

$$\langle r \rangle_{m_o, p_o} = r_0 \exp\left(\frac{7}{18} \alpha^2 \sigma_\mu^2\right) \quad (B.43)$$

Si nous voulons traiter les données avec la modèle statistique [$u = M$], il faudra alors supposer une forme pour la fonction de luminosité $f_M(M)$. L'utilisation du modèle [$u = p$] semble donc moins contraignante, mais le choix du modèle statistique dépend aussi de la phase calibration. Nous illustrons dans la section suivante, les problèmes rencontrés lorsqu'il s'agit de choisir un modèle statistique plutôt qu'un autre.

I.5.3. Exemples de différentes analyses :

Les indicateurs secondaires de distance nous permettent donc d'estimer la distance des galaxies lointaines. Nous illustrons à l'aide d'une simulation quels problèmes risquent

d'apparaître lors du traitement des données. Nous confirmons d'une part le résultat établi en I.3.3, soit que la connaissance de la physique sous-jacente aux indicateurs secondaires de distance n'est pas fondamentale. D'autre part, nous montrons que le choix du modèle statistique impliqué dans la phase de calibration ainsi que dans la détermination de la distance la plus probable s'avère fondamental si l'on désire traiter correctement notre échantillon de galaxies.

Nous avons simulé un échantillon de galaxies engendré à partir du modèle $[u(M, p) = p]$ et soumis à un cut-off en magnitude apparente ($\phi(m, p, \mu) = \Theta(m_{lim} - m)$). De plus, nous avons imposé à la distribution spatiale de cet échantillon d'être homogène ($\rho(r) dr = r^2 dr$) et nous avons choisi comme fonction de distribution de la variable p une gaussienne ($f_p(p) = g(p, p_0, \sigma_p)$). Les erreurs intrinsèques intervenant dans la relation $M = a.p + b + \zeta$ pour un tel échantillon sont donc sur p . Le traitement classique de ces données simulées avec la relation direct aboutirait à une estimation biaisée de la distance de chaque galaxie, l'échantillon possédant une magnitude apparente limite de coupure m_{lim} .

Nous allons maintenant réaliser le traitement complet de ces données simulées avec trois modèles statistiques différents. C'est à dire que pour chacun des trois modèles, nous évaluerons premièrement la valeur des coefficients a et b intervenant dans la phase de calibration de la relation (B.3) et deuxièmement, nous estimerons à l'aide de ces valeurs la distance la plus probable de chaque galaxie appartenant à l'échantillon simulé. Pour chacun des trois modèles statistiques, nous serons donc en mesure de comparer la distance vraie D_{vraie} de chaque galaxie avec sa distance estimée $D_{estimée}$.

A) ANALYSE 1 : Modèle statistique $[u = p]$ et fonction de sélection $\phi(m, p, \mu) = \Theta(m_{lim} - m)$

Nous supposons donc dans l'analyse 1 que l'échantillon a été engendré par le modèle statistique $[u = p]$ et possède une magnitude apparente limite de coupure m_{lim} ($\phi(m, p, \mu) = \Theta(m_{lim} - m)$).

On retrouve donc bien lors de la phase de calibration les valeurs originales des coefficients a et b . Nous estimons ensuite le module de distance ainsi que la distance pour chaque galaxie de l'échantillon. Nous montrons figure (B.6a), le module de distance estimé

$\mu_{estimé}$ en fonction du vrai module de distance μ_{vrai} et figure (B.6b), la distance estimée $D_{estimée}$ en fonction de la vraie distance D_{vraie} . Nous vérifions d'une part, que les points sont bien répartis de part et d'autre de la bissectrice (l'évaluation de la distance la plus probable pour chaque galaxie n'est pas biaisée) et d'autre part, que l'incertitude σ_r sur l'estimation de la distance est proportionnelle à la distance.

B) ANALYSE 2 : Modèle statistique $[u = p]$ et fonction de sélection $\phi(m, p, \mu) = \phi(p)$

Comme dans l'analyse 1, nous supposons que les données ont été engendrées par le modèle statistique $[u = p]$. Par contre, nous faisons cette fois-ci l'hypothèse que l'échantillon n'est soumis qu'à des effets de sélection en p ($\phi(m, p, \mu) = \phi(p)$).

Cette analyse va donc se révéler impropre du fait que la condition imposée aux effets de sélection ne s'applique pas à notre échantillon simulé. En fait, l'estimation des constantes a et b est biaisée. Ce biais qui apparaît donc lors de la phase de calibration va induire une erreur dans la détermination de la distance la plus probable (figure (B.7a), (B.7b)).

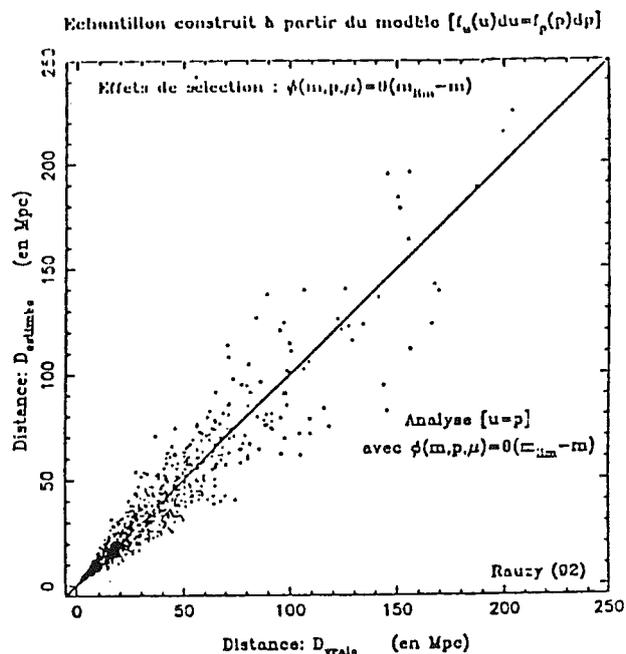
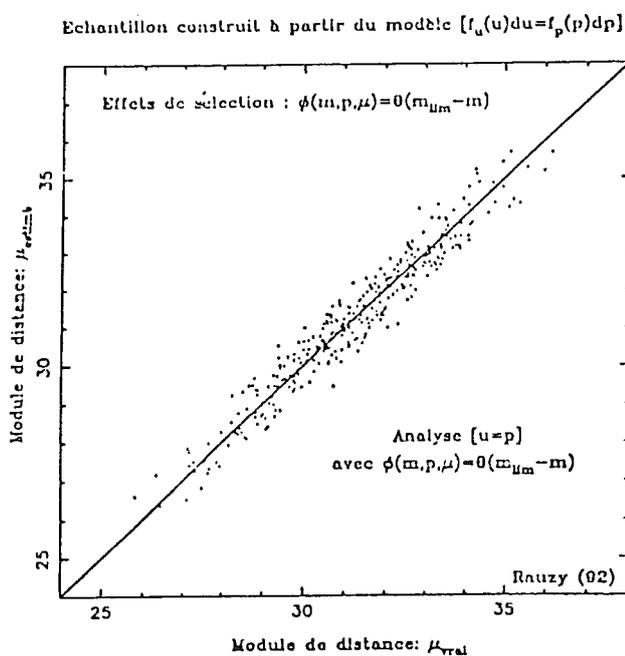
Ainsi, il est primordial d'utiliser un modèle statistique capable de rendre compte des effets de sélection réellement présents dans l'échantillon, que ce soit dans la phase de calibration de la relation (B.3), $M = a.p + b + \zeta$, ou dans la phase finale de détermination de la distance la plus probable.

C) ANALYSE 3 : Modèle statistique $[u = M]$ et fonction de sélection $\phi(m, p, \mu) = \phi(m)$

Nous supposons dans cette dernière analyse que l'échantillon a été engendré par le modèle statistique $[u = M]$ et que les effets de sélection ne dépendent que de la magnitude apparente m .

Les valeurs des coefficients a et b sont donc différentes de celles introduites dans la simulation puisque nous ne calibrons pas la relation (B.3) avec le bon modèle statistique ayant engendré les données. Néanmoins, la détermination de la distance la plus probable n'est pas biaisée pour autant (figures (B.8a) et (B.8b)). Nous voyons que les distances estimées sont comparables à celles obtenues dans l'analyse 1.

Ce résultat confirme que l'erreur commise, lorsqu'on utilise un modèle quelconque afin de traiter les données, est négligeable. Ce qui importe en fait, c'est de conserver ce modèle statistique tout au long de l'analyse, c'est à dire durant la phase de calibration ainsi que



Figures B.6a et B.6b : Diagrammes $\mu_{estimé}/\mu_{vrai}$ et $D_{estimé}/D_{vrai}$ pour l'analyse 1.

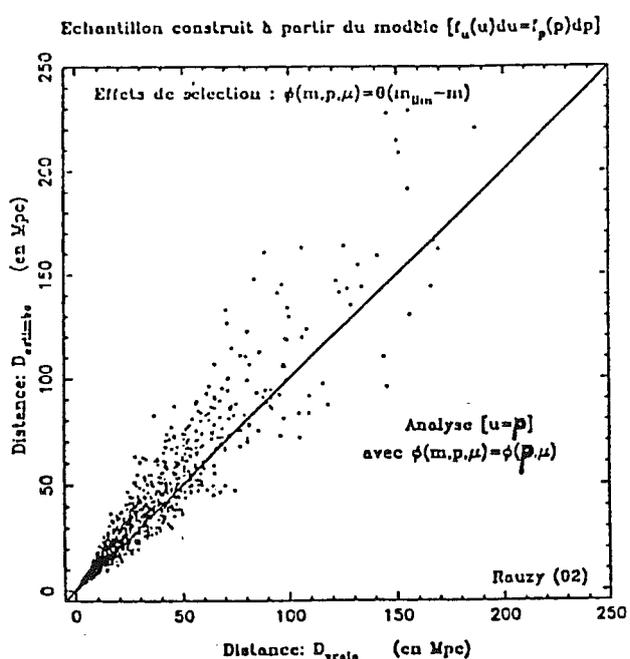
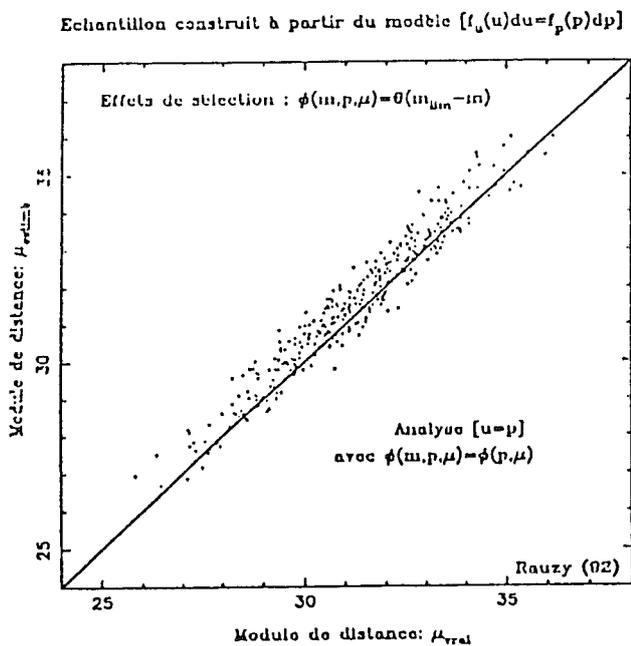


Figure B.7a et B.7b : Diagrammes $\mu_{estimé}/\mu_{vrai}$ et $D_{estimé}/D_{vrai}$ pour l'analyse 2.

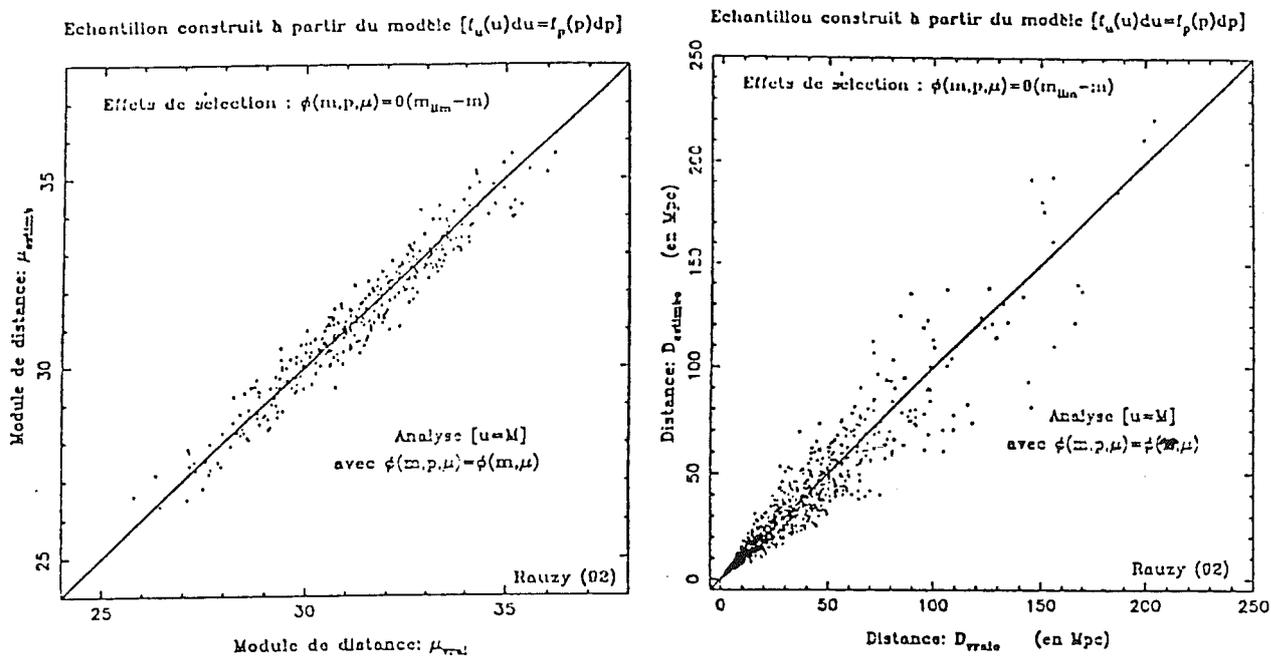


Figure B.8a et B.8b : Diagrammes μ_{estime}/μ_{vrai} et D_{estime}/D_{vrai} pour l'analyse 3.

la fonction de luminosité $f_M(M)$. Toutefois, cet avantage est perdu dans la phase de calibration, lorsqu'on analyse un échantillon soumis à des effets de sélection en magnitude apparente m , puisque une hypothèse sur la fonction de distribution $f_p(p)$ est alors nécessaire. Il semble donc qu'il n'y ait pas de modèle "miracle" permettant de traiter les données observationnelles disponibles, pour lesquelles les effets de sélection sont souvent mal connus.

I.5.4. Conclusion :

Nous avons donc vu que par l'intermédiaire de la relation (B.3), $M = a.p + b + \zeta$, il était possible d'estimer la distance la plus probable des galaxies lointaines. Nous avons montré que pour ce faire, il nous fallait choisir un modèle statistique afin de calibrer cette relation. Le choix de ce modèle statistique dépend en fait des effets de sélection auxquels est soumis l'échantillon. Après avoir estimé les valeurs des constantes a et b , nous sommes en mesure de déterminer la distance la plus probable d'une galaxie pour laquelle

montré que pour ce faire, il nous fallait choisir un modèle statistique afin de calibrer cette relation. Le choix de ce modèle statistique dépend en fait des effets de sélection auxquels est soumis l'échantillon. Après avoir estimé les valeurs des constantes a et b , nous sommes en mesure de déterminer la distance la plus probable d'une galaxie pour laquelle on a mesuré sa magnitude apparente m_o ainsi que son paramètre p_o . Si l'on souhaite rester cohérent, il est alors nécessaire d'utiliser le même modèle statistique que dans la phase de calibration. Dans tous les cas, l'estimation de la distance la plus probable d'une galaxie dépendra des hypothèses choisies afin de décrire l'échantillon observé, notamment de l'hypothèse sur la distribution spatiale des galaxies. L'incertitude intrinsèque présente dans ce processus de détermination des distances est d'environ 20% pour une galaxie isolée.

II. LE CHAMP DES VITESSES :

La mesure z du décalage des raies spectrales d'une galaxie nous donne accès à sa vitesse radiale v^r dans le repère des vitesses où notre galaxie est au repos. On décompose cette vitesse radiale en deux termes. D'une part un terme cosmologique qui rend compte de l'expansion de l'univers $v_H = H_0.D$ et qui peut être évalué si l'on a déterminé la distance D de la galaxie ainsi que la constante de Hubble H_0 . D'autre part, la composante radiale du champ des vitesses propres v_p^r , engendré par la croissance des fluctuations primordiales de la densité de matière sous l'influence de la gravitation.

II.1. Le décalage des raies spectrales d'une galaxie :

Le spectre lumineux d'une galaxie possède des raies d'absorption et d'émission bien définies. Chaque raie témoigne de la présence dans la galaxie d'un élément spécifique. En comparant la fréquence à laquelle apparaît cette raie avec la fréquence obtenue en laboratoire, on peut donc mesurer le décalage en fréquence auquel est soumis le spectre de la galaxie. On interprète alors ce décalage comme un effet Doppler-Fizeau : la galaxie est en mouvement par rapport à l'observateur. La mesure du décalage des raies spectrales d'une galaxie (le redshift z) nous permet donc d'obtenir la valeur de la vitesse radiale de la galaxie par rapport à l'observateur ($v^r = c.z$), c'est à dire la projection de la vitesse totale sur la ligne de visée joignant l'observateur et la galaxie.

La vitesse radiale d'une galaxie est souvent donnée dans le repère où notre galaxie est au repos. Il faut alors corriger le redshift mesuré du mouvement de l'observateur par rapport au centre de la terre, puis du mouvement de la terre par rapport au soleil et enfin du mouvement du soleil dans notre galaxie.

II.2. La composante radiale du champ des vitesses :

On obtient donc la composante radiale du champ des vitesses en mesurant le décalage des raies spectrales des galaxies. On a vu précédemment dans la section II.2.2 de la partie A que le champ des vitesses \mathbf{v} se décomposait dans l'approximation newtonienne de la façon suivante :

$$\mathbf{v} = \mathbf{v}_H + \mathbf{v}_p + \mathbf{v}_C \quad (\text{A.21})$$

Nous rappelons que l'on a introduit un champ de vitesses constant \mathbf{v}_C dans cette équation pour régler le problème du repère absolu (voir section A.II.2.1). Si l'on choisit maintenant un repère de référence (par exemple le repère des vitesses dans lequel notre galaxie est au repos), la vitesse radiale propre v_p^r d'une galaxie dans ce référentiel s'exprime alors :

$$v_p^r = c.z - H_0.D \quad (\text{B.44})$$

Nous voulons ici faire deux remarques importantes.

D'une part, nous n'avons accès qu'à la composante radiale du champ des vitesses propres. Si on exprime ce champ en coordonnées sphériques ($\mathbf{v}_p = (v_p^r, v_p^\theta, v_p^\phi)$), nous ne possédons aucune information sur les composantes v_p^θ et v_p^ϕ . On devra alors faire des hypothèses supplémentaires sur ces composantes. Les études du champ des vitesses propres dépendront donc fortement du choix de ces hypothèses.

D'autre part, l'équation (B.44) nous montre qu'il est nécessaire d'estimer la constante de Hubble H_0 ainsi que la distance de la galaxie D . Nous avons vu dans la section I. que l'évaluation de la distance la plus probable d'une galaxie comporte une incertitude intrinsèque d'environ 20%. Cette incertitude sur la distance D se répercute alors sur l'estimation de la vitesse radiale propre v_p^r . Ainsi, plus une galaxie est éloignée, plus la détermination de sa vitesse radiale propre est incertaine. Nous ne pouvons donc espérer qu'une connaissance locale du champ des vitesses propres.

III. ESTIMATION DE LA CONSTANTE DE HUBBLE H_0 :

La détermination de la constante de Hubble est nécessaire si l'on veut connaître le champ des vitesses propres. Plus généralement, la valeur de H_0 impose des contraintes à tous les modèles cosmologiques du type Big-Bang, par exemple sur l'âge de notre univers.

Lors de l'évaluation de cette constante, on est confronté à deux difficultés. D'une part le flot de Hubble v_H est localement contaminé par le champ de vitesses propres v_p et ce champ n'est pas statistiquement homogène et isotrope dans les catalogues de galaxies que nous possédons actuellement. Ainsi il faudra soustraire aux données observationnelles un champ de vitesses propres modélisé. D'autre part, l'estimation de la constante de Hubble H_0 nécessite la détermination des distances des galaxies. On sera donc soumis aux mêmes types de problèmes que ceux développés dans la section I .

Nous présentons dans cette section la calibration généralement adoptée. De plus, nous montrons à l'aide d'un exemple quels avantages apporte l'application du formalisme statistique présenté précédemment en I.3 lorsqu'il s'agit de déterminer la valeur de H_0 la plus probable.

III.1. La calibration généralement adoptée :

On a vu dans la section II.2 que l'expression de la vitesse radiale observée d'une galaxie fait intervenir la constante de Hubble H_0 de la façon suivante : $v^r = H_0 \cdot D + v_p^r$ (équation (B.44)). On peut alors estimer H_0 , par exemple par la méthode des droites de régression, si les deux hypothèses suivantes sont vérifiées. Les vitesses radiales propres des galaxies doivent être distribuées aléatoirement et de moyenne nulle. L'échantillon étudié doit être libre de tout effet de sélection qui biaiserait alors la détermination de la valeur de H_0 la plus probable. Notamment, les distances D des galaxies de l'échantillon ne doivent pas souffrir de biais.

Mais ces deux hypothèses ne s'appliquent pas aux catalogues de galaxies actuellement disponibles. La littérature nous offrent quelques méthodes dans le but de surmonter ces difficultés. Néanmoins l'estimation de la constante de Hubble H_0 varie suivant la méthode employée ainsi que le catalogue étudié. Typiquement, les valeurs de H_0 ainsi dérivées sont comprises entre 50 et 100 km.sec⁻¹/Mpc.

III.1.1. L'inhomogénéité du champ des vitesses propres :

L'équation (B.44) nous fournit une technique permettant l'estimation la valeur de H_0 la plus probable. On cherche en fait H_0 telle que la somme des carrés des vitesses radiales propres de toutes les galaxies de l'échantillon ($\sum_{galaxies} v_{p_i}^2 = \sum_{galaxies} (v_i^r - H_0 \cdot D_i)^2$) soit minimale. Mais cette technique n'est valable que si le champ des vitesses radiales propres n'est pas cohérent à l'échelle du catalogue de galaxies étudié. Notamment, il est nécessaire que ce champ ait une moyenne nulle et une distribution aléatoire.

Or le champ des vitesses propres réellement observé n'est pas localement homogène et isotrope statistiquement. On sait par exemple que les galaxies du super-amas local tombent toutes sur l'amas de Virgo. Et des mouvements cohérents à plus grande échelle ont été décelés récemment. Le "grand attracteur" domine le champ des vitesses propres dans un rayon d'au moins 40 Mpc (Lynden-Bell et *al.* (1988)) et peut être il existe des flots de vitesses convergents à plus grandes échelles (Rauzy et *al.* (1992)).

Il est donc nécessaire de corriger les vitesses radiales des galaxies de ces mouvements cohérents à grande échelle. Si des modèles de flot virgoцентриque sont disponibles dans la littérature (par exemple Aaronson et *al.* (1982)), nous ne possédons hélas pas encore de modèles satisfaisants capables de reproduire la cinématique régissant la région de l'espace qui nous entoure dans un rayon de 100 Mpc. Ainsi, l'inhomogénéité du champ des vitesses propres introduit des erreurs lors de la détermination de la constante de Hubble H_0 .

III.1.2. Le problème des biais statistiques :

Les méthodes classiques de calibration de la loi de Hubble nécessitent la connaissance de la distance des galaxies. Nous avons vu précédemment dans la section I que l'évaluation de la distance la plus probable d'une galaxie peut être affectée de biais introduits par la présence d'effets de sélection dans l'échantillon. Diverses techniques sont présentées dans

la littérature (par exemple Bottinelli et *al.* (1986)) qui prennent en compte ces effets de sélection. Mais la valeur de la constante de Hubble reste encore mal connue. Elle varie en effet suivant les catalogues utilisés ainsi que les méthodes employées afin de calibrer les indicateurs secondaires de distance ou d'estimer les distances les plus probables des galaxies.

III.2. Application du formalisme statistique défini en I.3 :

Nous appliquons le formalisme statistique précédemment présenté en I.3 afin d'évaluer la valeur de H_0 la plus probable (Triay et *al.*, en préparation). Nous montrons à l'aide d'un exemple simple quels sont les avantages de notre technique. D'une part nous déterminons directement H_0 sans avoir à estimer les distances les plus probables des galaxies. D'autre part, ce formalisme permet un traitement rigoureux des effets de sélection auxquels sont soumis les échantillons de galaxies observés.

Nous donnons à l'échantillon que nous voulons traiter les propriétés suivantes :

. Propriété 1 : La fonction de sélection de l'échantillon ne dépend pas du paramètre p impliqué dans les indicateurs secondaires de distance. C'est le cas par exemple si l'échantillon est soumis à un cut-off en magnitude apparente. la fonction de sélection s'écrit alors $\phi(m, p, \mu) = \phi(m, \mu)$.

. Propriété 2 : Le champ des vitesses propres v_p est négligeable. Nous avons vu dans la section III.1.1 que les échantillons réellement observés ne vérifient pas cette propriété. Pour traiter ces échantillons, il faudra faire une hypothèse différente.

Nous pouvons alors d'après la loi de Hubble $v^r = H_0.D$, exprimer le module de distance μ en fonction du redshift z :

$$\mu = 5 \log_{10} cz + 25 - 5 \log_{10} H_0 \quad (B.45)$$

Nous avons vu en I.3.3 que nous avons le choix pour le modèle statistique. De plus nous montrons (Triay et *al.*, en préparation) que la valeur estimée de H_0 possède la propriété d'être indépendante du modèle statistique. Nous traiterons donc cet exemple avec le modèle statistique $[u(M, p) = M]$ parce qu'il permet d'estimer H_0 indépendamment de la fonction de sélection $\phi(m, \mu)$ ainsi que de la fonction de luminosité $f_M(M)$. Dans

le modèle $[u(M, p) = M]$ nous calibrons donc la relation (B.3) afin d'estimer les valeurs des coefficients a et b (voir I.4.3 exemple A)].

Nous procédons maintenant à un changement de variables (m, p, μ) en (x, y, η) avec :

$$\eta = \mu + 5 \log_{10} H_0 \quad (B.46)$$

$$x = m - \eta \quad (B.47)$$

$$y = a.p + b - x \quad (B.48)$$

La densité de probabilité pour une galaxie de l'échantillon s'écrit alors en fonction des nouvelles variables (x, y, η) . D'après l'équation (B.10) et pour le modèle statistique $[u(M, p) = M]$ ($F(M, p) dM dp = f_M(M) dM g(\zeta, 0, \sigma_\zeta) d\zeta$), on a :

$$dP_{\text{observée}} = \frac{1}{A_{\text{observée}}} F_M(x + 5 \log_{10} H_0) h(\eta - 5 \log_{10} H_0) \phi(x + \eta) dx d\eta \\ \times g(y - 5 \log_{10} H_0, 0, \sigma_\zeta) dy \quad (B.49)$$

L'équation (B.49) nous montre alors que la variable aléatoire y est indépendante des deux variables aléatoires x et η . Ainsi son espérance mathématique $E(y)$ vérifie l'équation :

$$E(y) = \int y dP_{\text{observée}} = 5 \log_{10} H_0 \quad (B.50)$$

La valeur de la constante de Hubble H_0 la plus probable est donc donnée, en posant $\langle y \rangle = \sum_{i=1}^{N_{\text{gal}}} y_i$, par l'équation suivante :

$$H_0 \approx 10^{\langle y \rangle / 5} \quad (B.51)$$

On voit donc à l'aide de cet exemple que l'introduction de notre formalisme statistique permet de déterminer la valeur de la constante de Hubble H_0 sans estimer préalablement les distances des galaxies. Nous évitons ainsi de poser des hypothèses supplémentaires comme l'hypothèse d'homogénéité de la distribution spatiale des galaxies par exemple.

De plus, le choix du modèle statistique donne la possibilité de traiter des échantillons soumis à d'autres effets de sélection que ceux présents dans cet exemple. Il faudra alors trouver le modèle statistique le plus apte à traiter les données. La technique du maximum de vraisemblance nous fournit certainement un outil puissant lorsqu'il s'agit de déterminer la constante de Hubble H_0 .

IV. LE CATALOGUE DE FABER et al. (1989) :

Nous présentons dans cette section le catalogue de Faber et al. établi en 1988 (Lynden-Bell et al. (1988)) et légèrement modifié en 1989 (Faber et al. (1989)) afin de prendre en compte de nouvelles données observationnelles (Lucey et al. (1988)).

Ce catalogue comprend finalement 512 galaxies elliptiques (180 possédant des positions distinctes) pour lesquelles on a mesuré le redshift z , la dispersion des vitesses σ ainsi que le diamètre angulaire D_n .

Une relation statistique (indicateur de distance) liant ces paramètres observés permet de déterminer la distance la plus probable de chacune des galaxies et donc d'obtenir le champ des vitesses propres de la région de l'espace couverte par le catalogue.

La calibration de cette relation est effectuée à partir de 10 amas de galaxies (Dressler et al. (1987)). Une pente moyenne pour l'indicateur de distance est déduite par une méthode de régression. Le point zéro de la relation est calibré par rapport à la distance d'un amas particulier.

Les distances des galaxies sont ensuite estimées et corrigées d'un biais qui affecte les données. Ces distances sont exprimées en $\text{km}\cdot\text{sec}^{-1}$ afin d'éviter l'évaluation de la constante de Hubble H_0 .

Enfin, on calcule à l'aide du redshift z la vitesse radiale propre v_p pour chaque galaxie de l'échantillon. Le catalogue de Faber et al. donne accès au champ des vitesses propres de la région qui nous entoure dans un rayon de $6000 \text{ km}\cdot\text{sec}^{-1}$.

IV.1. Les données observationnelles :

Faber et al. ont mesuré le redshift z , la dispersion des vitesses σ ainsi qu'un nouveau paramètre photométrique (le diamètre angulaire D_n) pour 512 galaxies elliptiques.

Le paramètre photométrique D_n est défini comme le diamètre d'un cercle tel que la

brillance de surface intégrée dans ce cercle soit égale à $\Sigma = 20.75 \text{ mag.arcsec}^{-2}$ dans la bande photométrique B. Certaines corrections doivent être effectuées sur la valeur de ce paramètre. Il faut soustraire le terme de K-correction qui dépend de la latitude galactique de chaque galaxie considérée. On doit tenir compte de l'absorption interstellaire. Enfin, un terme cosmologique en $(1+z)^4$ apparaît dans l'évaluation de D_n . Mais ce terme est négligeable dans le cas présent, les redshifts mesurés étant petits devant 1.

Les effets de sélection pour ce catalogue sont mal connus. D'une part cet échantillon n'est pas complet en magnitude. Toutes les galaxies sélectionnées possèdent une magnitude apparente dans le bleu B inférieure à 13.0 mag mais la forme de la fonction de sélection en magnitude reste inconnue. Les effets de sélection en diamètre angulaire D_n existent donc puisque D_n dépend en fait de la magnitude apparente ainsi que de la distance r pour chaque galaxie. D'autre part, les galaxies ayant une dispersion des vitesses σ inférieure à 100 km.sec^{-1} ont été éliminées du catalogue. Ces galaxies introduiraient en effet une dispersion trop importante dans les indicateurs secondaires de distance.

On voit donc que la fonction de sélection de cet échantillon va dépendre simultanément de D_n , σ et r . De plus cette dépendance semble difficilement modélisable.

Sur les 512 galaxies étudiées, nombreuses sont celles qui appartiennent à un même groupe. Pour obtenir la vitesse radiale propre d'un groupe (c'est à dire en un même point de l'espace), il faut alors faire la moyenne des vitesses radiales des galaxies présentes dans ce groupe. Il ne reste en fait après cette opération plus que 180 points de coordonnées spatiales distinctes. Nous appellerons NEC ce sous-catalogue. Nous montrons figure (B.9) la répartition de ces points sur une carte bidimensionnelle en coordonnées galactiques. Un simple coup d'oeil nous permet d'affirmer que les 180 galaxies de ce sous-catalogue NEC ne sont pas réparties uniformément dans l'espace.

IV.2. Les indicateurs de distance :

La relation statistique utilisée comme indicateur secondaire de distance est une amélioration de la relation de Faber-Jackson pour les galaxies elliptiques. Si toutes les galaxies elliptiques avaient le même diamètre intrinsèque, le diamètre angulaire D_n précédemment introduit serait proportionnel à l'inverse de la distance. En fait, en notant D_i le diamètre

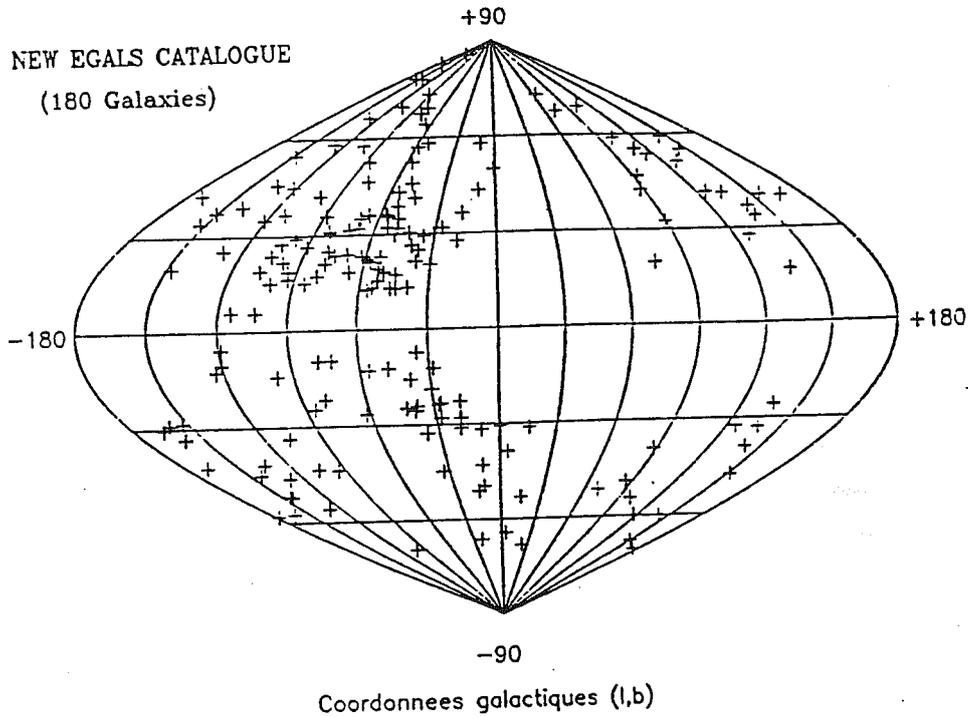


Figure B.9 : Répartition spatiale des galaxies du catalogue NEC

angulaire intrinsèque d'une galaxie, l'équation suivante est vérifiée :

$$r \propto \frac{D_i}{D_n} \quad (B.52)$$

Dressler et *al.* ont remarqué dans les amas de galaxies (Dressler et *al.* (1987)) que le diamètre angulaire intrinsèque D_i est proportionnel à la dispersion des vitesses σ élevée à une certaine puissance x ($D_i \propto \sigma^x$). Une telle dépendance implique que le rapport masse-luminosité pour les galaxies elliptiques évolue lentement en fonction de la luminosité. La relation statistique jouant le rôle d'indicateur de distance va donc s'écrire :

$$\log r = x \log \sigma - \log D_n + c + \zeta \quad (B.53)$$

avec c une constante

et ζ une variable aléatoire centrée

L'introduction du paramètre photométrique D_n permet de réduire considérablement la dispersion intrinsèque présente dans la relation de Faber-Jackson originale. La dispersion intrinsèque Δ de la variable aléatoire centrée ζ est évaluée dans les amas de galaxies ($\Delta = 0.21$). La relation (B.53) est alors comparable quantitativement avec la relation de Tully-Fisher pour les galaxies spirales.

IV.3. La phase de calibration :

Il s'agit maintenant d'estimer la pente x ainsi que le point zéro (la constante c) de la relation (B.53). Nous avons vu dans la section I qu'il est alors nécessaire de choisir une droite de régression où plus généralement un modèle statistique décrivant les données. Les auteurs expriment l'incertitude présente dans l'indicateur de distance au moyen d'une probabilité conditionnelle $p(l_e | l)$, c'est à dire la probabilité de mesurer $l_e = \log r_e$ pour une galaxie dont la vraie distance est r ($l = \log r$), le logarithme de la distance estimée l_e étant donné par la relation suivante :

$$l_e = x \log \sigma - \log D_n + c \quad (B.54)$$

Et l_e doit être distribué suivant une gaussienne centrée sur l et de dispersion Δ :

$$p(l_e | l) = (2\pi\Delta^2)^{-1/2} \exp[-1/2(l - l_e)^2 / \Delta^2] \quad (B.55)$$

On voit donc que les auteurs choisissent implicitement en faisant cette hypothèse d'appliquer une régression de $\log D_n$ sur $\log \sigma$. Les raisons de ce choix demeurent cependant assez obscures et ne peuvent certainement pas être justifiées par les effets de sélection auxquels est soumis l'échantillon.

IV.3.1. La détermination de la pente x :

Une régression de $\log D_n$ sur $\log \sigma$ est effectuée sur 10 amas de galaxies différents (Dressler et al. (1987)). La valeur de x retenue ($x = 1.2 \pm 0.1$) est alors la moyenne des valeurs de x pour chaque amas.

Nous voulons faire remarquer que la sélection des galaxies dans ces amas introduit un biais dans la détermination de la pente x . En fait, les effets de sélection vont dépendre

du diamètre angulaire D_n ainsi que de la distance r . Nous avons montré précédemment dans la section I que la régression de $\log D_n$ sur $\log \sigma$ donnait une estimation de x biaisée dans ce cas. L'utilisation de la relation inverse (c'est à dire une régression de $\log \sigma$ sur $\log D_n$) semble préférable. Hélas, les galaxies sont aussi sélectionnées suivant leur vitesse de dispersion σ . Ainsi l'estimation de x par la relation inverse est tout aussi biaisée que son estimation par la relation directe. Le formalisme statistique que nous avons introduit dans la section I.3 se montre dans un tel cas de figure (des effets de sélection non triviaux) très utile.

IV.3.2. La détermination du point zéro :

Les distances des galaxies du catalogue de Faber et *al.* sont exprimées en $\text{km}\cdot\text{sec}^{-1}$. Ceci évite en fait le calcul de la constante de Hubble H_0 . Si l'on pose :

$$R_e = H_0 r_e \quad (B.56)$$

La relation (B.54) devient alors :

$$\log R_e = x \log \sigma - \log D_n - c + \log H_0 \quad (B.57)$$

Ainsi la constante de Hubble H_0 intervient dans cette équation de la même façon que la constante c . L'estimation de H_0 sera néanmoins nécessaire si l'on désire exprimer les distances en Mpc.

Le point zéro de l'équation (B.57) (c'est à dire la valeur de l'expression $c - \log H_0$) est calculé à partir de la distance de l'amas de Coma. La vitesse radiale de l'amas de Coma est de $7200 \text{ km}\cdot\text{sec}^{-1}$ dans le repère des vitesses où le fond diffus cosmologique est au repos. Les auteurs obtiennent, d'après un modèle du champ des vitesses propres, une vitesse radiale propre pour Coma de $-260 \text{ km}\cdot\text{sec}^{-1}$. La valeur du point zéro est alors de 1.411. L'équation (B.57) devient donc une fois calibrée et en ajoutant un terme correcteur d'origine cosmologique :

$$\log R_e = 1.20 \log \sigma - \log D_n + 1.411 + \log \left(\frac{1 + 7/4 z}{1 + 7/4 z_{\text{coma}}} \right) \quad (B.58)$$

La calibration de cet indicateur de distance est donc loin d'être parfaite. D'une part l'estimation de la pente x est biaisée par les effets de sélection auxquels est soumis l'échantillon. D'autre part, la valeur de la distance de l'amas de Coma n'est pas certaine.

IV.4. L'évaluation de la distance d'une galaxie :

L'indicateur de distance permet donc de calculer la distance la plus probable d'une galaxie. Mais cette évaluation est soumise à un biais, improprement appelé dans la littérature du type biais de Malmquist. Pour corriger ce biais, il est nécessaire de faire une hypothèse sur la distribution spatiale des galaxies. Les auteurs postulent que les galaxies de l'échantillon sont réparties uniformément dans l'espace (cette hypothèse n'est clairement pas vérifiée) :

$$\rho(r) \propto r^{\alpha-1} dr \quad \text{avec } \alpha = 3 \quad (B.59)$$

La probabilité conditionnelle $p_1(l|l_e)$ s'exprime alors :

$$p_1(l|l_e) = (2\pi\Delta^2)^{-1/2} \exp[-1/2(l - (l_e + \alpha\Delta^2))^2/\Delta^2] \quad (B.60)$$

La distance la plus probable $\langle R \rangle$ d'une galaxie (exprimée en km.sec⁻¹) pour laquelle on a déterminé R_e par la mesure de D_n et σ (équation (B.58)) est alors :

$$\langle R \rangle = \int R p_1(l|l_e) dl = R_e [1 + (\alpha + 1/2)\Delta^2] \quad (B.61)$$

Il faut donc appliquer une correction de "Malmquist" égale à $3.5 \Delta^2 R_e$ à la distance estimée R_e à partir de la relation (B.58). Nous montrons figures (B.10) et (B.11) les effets de cette correction pour le catalogue de Faber et *al.* Les auteurs appliquent différemment cette correction suivant que les galaxies sont isolées ou appartiennent à un groupe.

Si les galaxies sont isolées, la correction précédente est appliquée avec une dispersion intrinsèque Δ égale à 0.21. L'incertitude dans la détermination de la distance la plus probable d'une galaxie est alors de 17 %.

Si les galaxies sont dans un groupe, la distance de ce groupe est calculée en faisant la moyenne des distances estimées pour chaque galaxie. Une correction empirique est ensuite effectuée parce que l'équation (B.58) n'est pas valable pour les galaxies appartenant à un groupe (présence d'effets de sélection). Enfin, on corrige la distance du groupe du biais de "Malmquist" en remplaçant dans l'équation (B.61) Δ par Δ/N , N étant le nombre de galaxies contenues dans le groupe. Ceci réduit en principe l'incertitude dans l'estimation de la distance d'un groupe de galaxies. Néanmoins l'évaluation de cette distance est soumise à un biais supplémentaire dont la correction demeure douteuse.

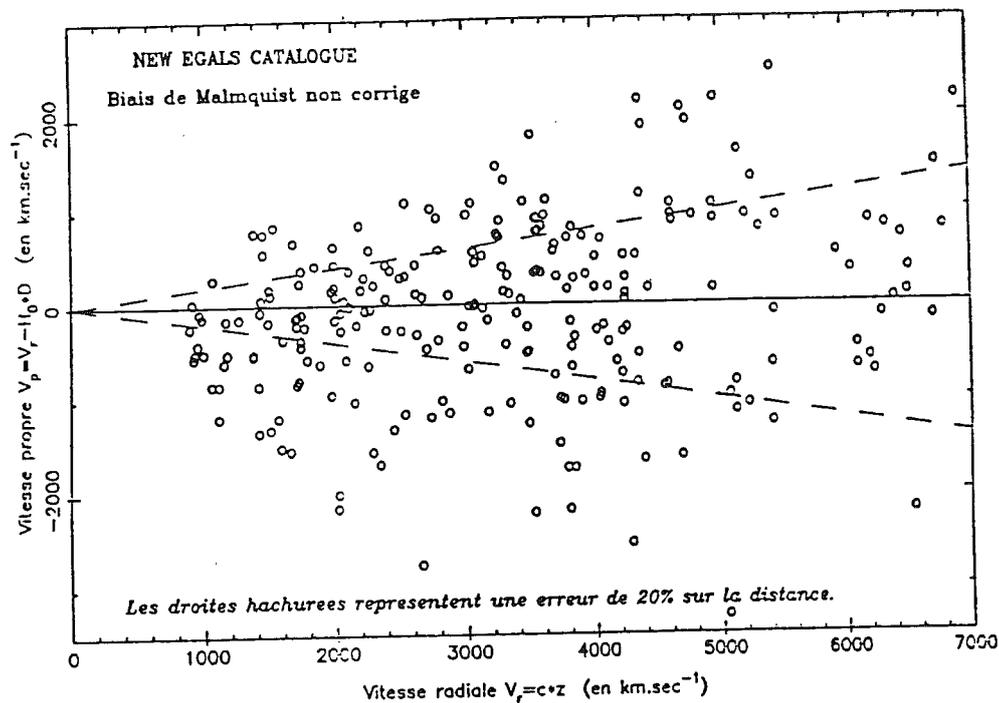


Figure B.10 : La vitesse radiale propre non corrigée du "biais de Malmquist" en fonction de la vitesse radiale pour les galaxies du catalogue NEC.

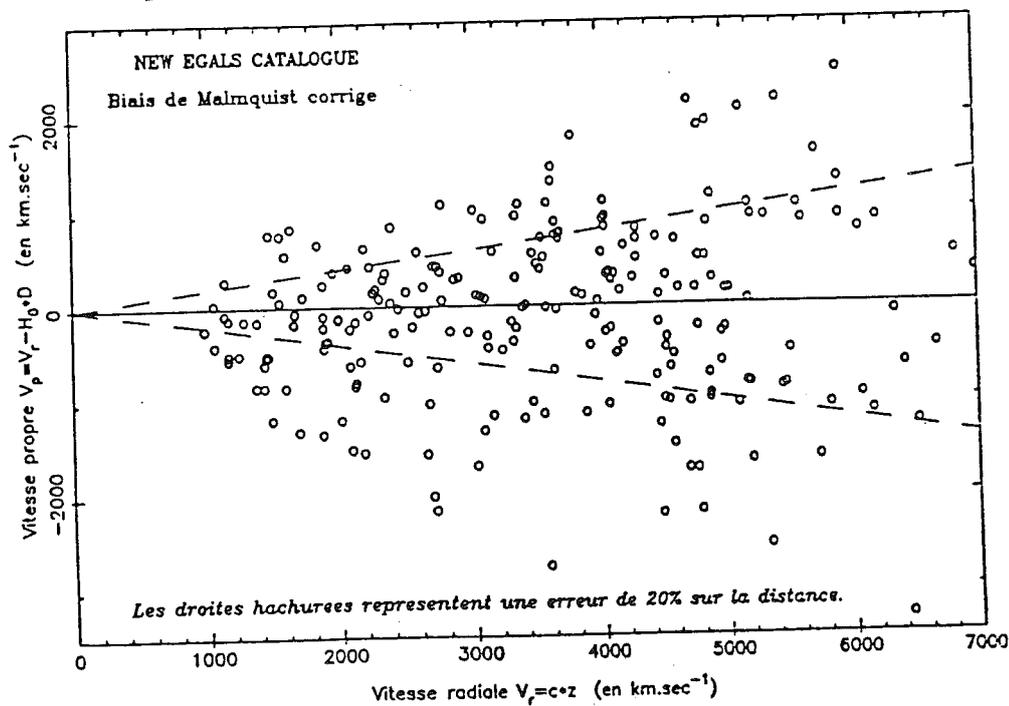


Figure B.11 : La vitesse radiale propre corrigée du "biais de Malmquist" en fonction de la vitesse radiale pour les galaxies du catalogue NEC.

IV.5. Le champ des vitesses radiales propres :

La vitesse radiale propre v_p^r pour chaque galaxie est donnée par la relation suivante :

$$v_p^r = c.z - H_0 r = v^r - R \quad (B.62)$$

Les auteurs préfèrent exprimer le champ de vitesses radiales propres dans le repère des vitesses où le fond diffus cosmologique est au repos (R_{CMB}). Il faut alors corriger toutes les vitesses du mouvement propre de notre galaxie dans ce repère. La vitesse de notre galaxie par rapport au repère R_{CMB} est calculée à partir du dipole observé dans le fond diffus cosmologique.

On a ainsi déterminé la distance et la vitesse radiale propre de 180 galaxies ou groupes de galaxies. Ces objets sont distribués de façon non homogène dans la région de l'espace qui nous entoure dans un rayon de $6000 \text{ km} \cdot \text{sec}^{-1}$. L'incertitude sur l'estimation de la distance (et donc de la vitesse) pour chaque galaxie est d'environ 20%. Mais à cette incertitude intrinsèque risquent de s'ajouter des erreurs systématiques introduites lors des différentes étapes impliquées dans le processus de détermination de la distance la plus probable d'une galaxie. La correction de ces erreurs nécessite certainement une étude plus approfondie des effets de sélection présents dans cet échantillon.

REFERENCES :

- [1] : M. Aaronson, J. Huchra, J. Mould, P. Schechter and B. Tully "The velocity field in the local supercluster", *Astrophys.J* 258, 64-76 (1982)
- [2] : M. Aaronson, G. Bothun, J. Mould, J. Huchra, R.A. Schommer, M.E. Cornell : "A distance scale from the infrared magnitude/HI velocity-width relation. V. Distance moduli to 10 Galaxy clusters, and positive detection of bulk superclusters motion toward the microwave anisotropy", *Astrophys.J* 302, 536-563 (1986)
- [3] : G. Bigot and R. Triay : "Comments on the Hubble diagram I. The Malmquist bias", *Phys. Lett. A* 150, 227 (1990a)
- [4] : G. Bigot and R. Triay : "Comments on the Hubble diagram II. The technique of fitting", *Phys. Lett. A* 150, 236 (1990b)
- [5] : L. Bottinelli, L. Gouguenheim, G. Paturel and P. Teerikorpi : "The Malmquist bias and the value of H_0 from the Tully-Fisher relation", *Astr.Astrophys.* 156, 157-171 (1986)
- [6] : A. Dressler, D. Lynden-Bell, D. Burstein, R.L. Davies, S.M. Faber, R.J. Terlevich and G. Wegner : "Spectroscopy and photometry of elliptical galaxies. I. A new distance estimator", *Astrophys.J* 313, 42-58 (1987)
- [7] : S.M. Faber and R.E. Jackson : "Velocity dispersions and mass-to-light ratios for elliptical galaxies", *Astrophys.J* 204, 668-683 (1976)
- [8] : S.M. Faber, G. Wegner, D. Burstein, R.L. Davies, A. Dressler, D. Lynden-Bell and R.J. Terlevich : "Spectroscopy and photometry of elliptical galaxies. VI. Sample selection

and data summary", *Astrophys.J Suppl.* **69**, 763-808 (1989)

[9] : P. Fouqué, L. Bottinelli, L. Gouguenheim and G. Paturel : "The extragalactic distance scale. II. The unbiased distance to the Virgo cluster from B-band Tully-Fisher relation", *Astrophys.J* **349**, 1-21 (1990)

[10] : L. Gouguenheim and L. Bottinelli : "Observational biases in the extragalactic distance scale", Vth IAP Astrophysics meeting: Astrophysical ages and dating methods

[11] : T. Isobe, E. D. Feigelson, M.G. Akritas and G.J. Babu : "Linear regression in astronomy. I.", *Astrophys.J* **364**, 104-113 (1990)

[12] : R.C. Kraan-Korteweg, L.M. Cameron and G.A. Tammann : "21 centimeter line width distances of clusters and the value of H_0 ", *Astrophys.J* **331**, 620-640 (1988)

[13] : J.R. Lucey and D. Carter : "Distances to Five nearby southern galaxy clusters and the absolute motion of the Local Group", *Mon.Not.R.astr.Soc.* (1988) **235**, 1177-1201

[14] : D.Lynden-Bell, A. Dressler, D. Burstein, R.L. Davies, S.M. Faber, R.J. Terlevich and G. Wegner : "Spectroscopy and photometry of elliptical galaxies. V. Galaxy streaming toward the new supergalactic center", *Astrophys.J* **326**, 19-49 (1988)

[15] : M.J. Pierce and R.B. Tully : "Distances to the Virgo and Ursa Major clusters and the determination of H_0 ", *Astrophys.J* **330**, 579-595 (1988)

[16] : M.J. Pierce and R.B. Tully : "Luminosity-line width relations and the extragalactic distance scale : I. Absolute calibration", *Astrophys.J* **387**, 47-55 (1992)

[17] : S. Rauzy, M. Lachièze-Rey and R.N. Henriksen : "Detecting non-Hubble velocity fields in the universe", *Astr.Astrophys.* **256**, 1-9 (1992)

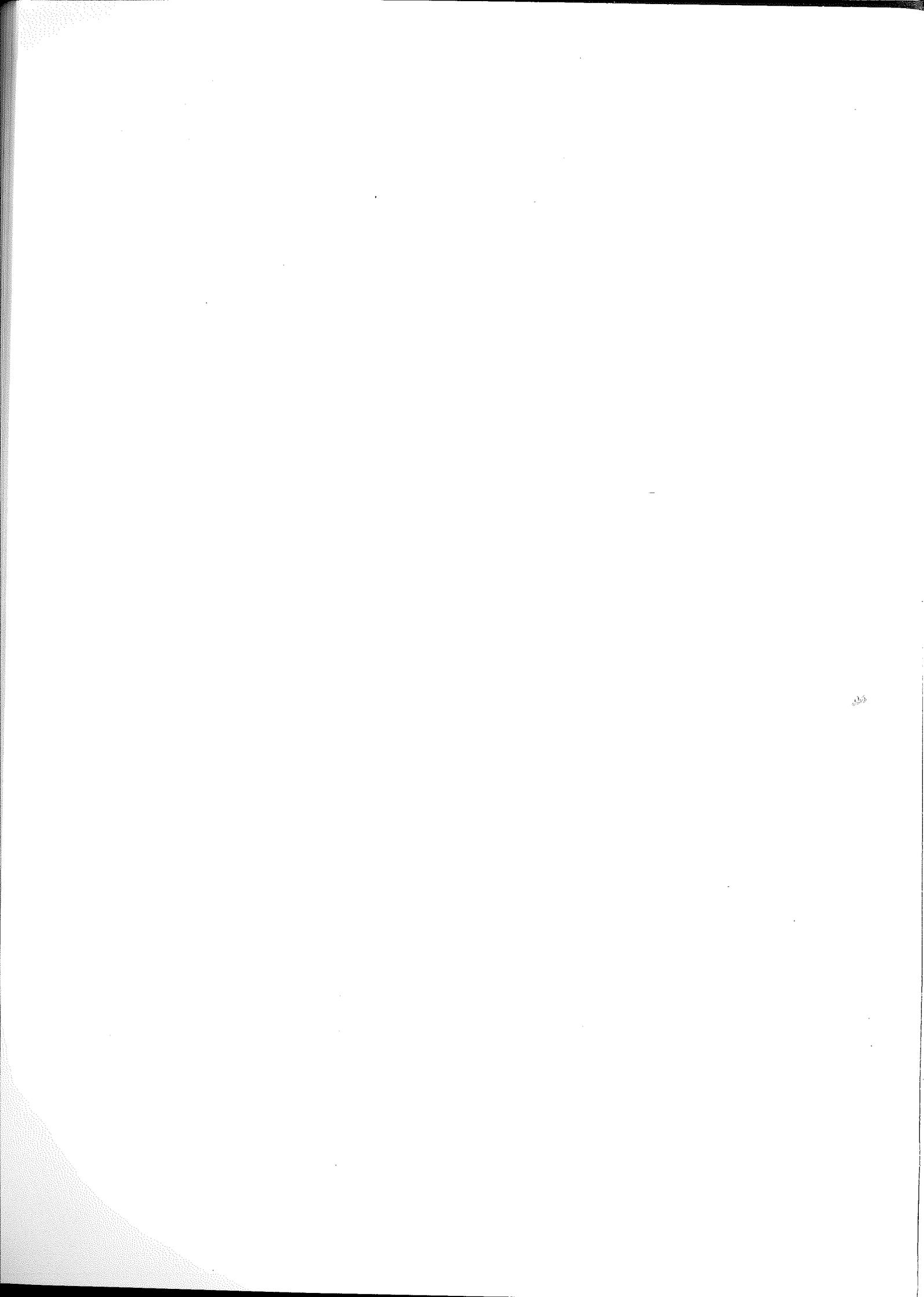
[18] : P.L. Schechter : "Mass-to-Light ratios for elliptical galaxies", *Astron.J* **85**, 801 (1980)

- [19] : P. Teerikorpi : "On the effect of the luminosity selection on the redshift-distance relationship", *Astr.Astrophys.* 45, 117-124 (1975)
- [20] : P. Teerikorpi : "Malmquist bias in relation of the form $M = a.p + b$ ", *Astr.Astrophys.* 141, 407-410 (1984)
- [21] : P. Teerikorpi : "Cluster population incompleteness bias and distances from the Tully-Fisher relation: theory and numerical examples", *Astr.Astrophys.* 173, 39-42 (1987)
- [22] : P. Teerikorpi : "Theoretical aspects in the use of the inverse Tully-Fisher relation for distance determination", *Astr.Astrophys.* 234, 1-4 (1990)
- [23] : G.A. Tammann : "The cosmic distance scale", IAU Symposium "Observational Cosmology" (1987)
- [24] : R.B. Tully and J.R. Fisher : "A new method of determining distances to galaxies", *Astr.Astrophys.* 54, 661-673 (1977)
- [25] : R.B. Tully : "Origin of the Hubble constant controversy", *Nature*, 334, 209-212 (1988)
- [26] : S. Van den Bergh : "The cosmic distance scale", *Astr.Astrophys.Review* (1989)



PARTIE C :

**LE TENSEUR
DE CORRELATION
DES VITESSES**



SOMMAIRE :

I. Définitions :

I.1. Le tenseur de corrélation des vitesses :

I.1.1. Si le champ des vitesses est homogène et isotrope : $\mathbf{v}(\mathbf{x}) = \mathbf{v}_{HI}(\mathbf{x})$

I.2. Le tenseur des différences :

I.2.1. Si le champ des vitesses s'écrit : $\mathbf{v}(\mathbf{x}) = \mathbf{v}_C + \mathbf{v}_H(\mathbf{x}) + \mathbf{v}_{HI}(\mathbf{x})$

II. Les intérêts :

II.1. Les fonctions de corrélation :

II.1.1. La fonction de corrélation spatiale des galaxies :

II.1.2. Le tenseur de corrélation des vitesses :

II.1.3. Discrimination entre différents modèles :

II.2. Le paramètre de biais :

II.3. La connaissance du champ des vitesses :

III. Le calcul du tenseur de corrélation des vitesses :

III.1. Nécessité d'une hypothèse :

III.2. L'évaluation de $B_{LL}(r)$ et $B_{NN}(r)$:

IV. Les applications :

IV.1. Les données :

IV.2. L'estimation de $B_{LL}(r)$ et $B_{NN}(r)$:

IV.2.1. Les effets de bords et d'échantillonnage :

IV.2.2. Les fonctions $B_{LL}(r)$ et $B_{NN}(r)$:

IV.2.3. Les barres d'erreurs pour la régression :

IV.3. Les incertitudes sur les distances :

L'étude du tenseur de corrélation des vitesses permet théoriquement d'accéder au spectre des fluctuations de la densité de matière totale dans l'univers. La comparaison de ce spectre avec la fonction de corrélation spatiale des galaxies nous renseigne alors sur le comportement de la masse lumineuse vis-à-vis de la masse totale. Ceci nous offre en principe la possibilité de tester la validité des différents modèles proposés afin d'expliquer la formation des grandes structures actuellement observées dans l'univers (voir section II).

Si le champ des vitesses propres de l'échantillon étudié est statistiquement homogène et isotrope (c'est-à-dire que l'échantillon est statistiquement représentatif des fluctuations du champ des vitesses), le tenseur de corrélation peut s'exprimer en fonction de ses composantes longitudinale $B_{LL}(r)$ et latérale $B_{NN}(r)$ (voir section I). De plus, il est possible d'évaluer ces fonctions à partir du champ des vitesses radiales $v_r(x)$ seulement (voir section III).

En pratique, l'estimation de ces fonctions de corrélation pour les catalogues de Faber *et al.* et de Aaronson *et al.* ne s'accorde pas avec les valeurs prédites par les théories classiques de formation de structures. Néanmoins l'amplitude des barres d'erreurs impliquées dans l'évaluation de $B_{LL}(r)$ et $B_{NN}(r)$ est considérable (voir section IV). Il semble en fait que l'hypothèse d'homogénéité et d'isotropie du champ des vitesses ne soit pas vérifiée, la cinématique de ces deux échantillons étant dominée par un flot convergent, cohérent à grande échelle (le "Grand Attracteur", voir partie D).

L'estimation des composantes du tenseur est donc impossible à partir des données observationnelles actuellement disponibles. Nous montrons de plus que l'incertitude rencontrée lors de la détermination des distances des galaxies introduit un biais et des barres d'erreurs importantes qui interdisent en pratique d'évaluer le tenseur de corrélation des vitesses.

I. DEFINITIONS :

Nous donnons dans cette section les définitions du tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ ainsi que du tenseur des différences $D_{ij}(\mathbf{r})$. Ces quantités ont été introduites dans le domaine de la mécanique des fluides (plus précisément en dynamique statistique des fluides) afin de caractériser les propriétés globales du comportement d'un fluide.

La dynamique statistique des fluides permet de décrire toute une classe de phénomènes physiques (par exemple la turbulence dans un fluide) dont les variables dynamiques locales sont soumises à des fluctuations statistiques. Les mesures en un point donné de l'espace-temps des variables dynamiques locales (comme la densité du fluide, son champ des vitesses, etc...) sont alors différentes pour deux expériences possédant le même ensemble de conditions initiales macroscopiques. Chaque variable dynamique du fluide (par exemple sa densité ρ) est ainsi considérée comme une variable aléatoire $\rho(\mathbf{x}, t)$ définie en chaque point (\mathbf{x}, t) de l'espace-temps. Chaque expérience est une réalisation de cette variable aléatoire, et pour un point donné (\mathbf{x}_0, t_0) de l'espace-temps, la mesure de la valeur de $\rho(\mathbf{x}_0, t_0)$ pour un nombre important d'expériences (de réalisations) nous permet en fait d'accéder à la distribution $dP_M = p_M(\rho) d\rho$ de la variable aléatoire ρ au point $M = (\mathbf{x}_0, t_0)$. Les quantités caractérisant globalement la dynamique du fluide sont alors définies comme des valeurs moyennes sur l'ensemble des différentes réalisations (par exemple la densité moyenne au point $M = (\mathbf{x}_0, t_0)$ s'exprime : $\overline{\rho(M)} = \int \rho(M) dP_M$).

En général, il est difficile de reproduire plusieurs fois la même expérience et ainsi d'accéder à la mesure des variables dynamiques pour différentes réalisations du même phénomène physique. Ceci est particulièrement vrai lorsqu'on désire étudier statistiquement la dynamique du fluide cosmique, puisque il existe par définition qu'une seule réalisation de notre univers. Néanmoins si le fluide cosmique possède certaines propriétés spécifiques, les théorèmes ergodiques nous permettent d'identifier les valeurs moyennes des quantités prises sur l'ensemble des différentes réalisations avec l'intégrale de ces quantités effectuée sur un certain volume de l'espace-temps. Nous exposons dans la suite (section

I.1) quelles hypothèses doit satisfaire le champ des vitesses afin que cette identification soit valable. Si ces hypothèses sont vérifiées, on pourra alors déterminer des quantités telles que le tenseur de corrélation des vitesses (ou le tenseur des différences) du fluide cosmique.

I.1. Le tenseur de corrélation des vitesses :

L'opération qui consiste à remplacer les valeurs moyennes sur l'ensemble des réalisations par une intégrale sur un volume d'espace (ou de temps) n'est valable que si les variables dynamiques du fluide possèdent certaines propriétés.

. Hypothèse 1 : Le champ des vitesses est statistiquement homogène :

Les théorèmes ergodiques nous indiquent notamment que le champ des vitesses $\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), v_2(\mathbf{x}), v_3(\mathbf{x}))$ du fluide cosmique (défini en tout point $\mathbf{x} = (x_1, x_2, x_3)$ de l'espace à 3 dimensions) doit être statistiquement homogène. C'est-à-dire qu'il existe un volume B_c tel que pour tout volume $B > B_c$, l'intégrale sur le volume B du champ des vitesses autour d'un point \mathbf{x} converge vers une constante, quelquesoit \mathbf{x} :

$$\overline{\mathbf{v}(\mathbf{x})} \approx \frac{1}{B} \int_B \mathbf{v}(x_1 + \zeta_1, x_2 + \zeta_2, x_3 + \zeta_3) d\zeta_1 d\zeta_2 d\zeta_3 = \text{constante} \quad (C.1)$$

Plus concrètement, l'hypothèse d'homogénéité statistique du champ des vitesses peut s'exprimer de la façon suivante. Le champ des vitesses du fluide cosmique est modélisé comme la superposition de diverses perturbations possédant chacune un centre, une amplitude ainsi qu'une longueur de cohérence propres. Si l'amplitude moyenne de ces perturbations décroît en fonction de leur longueur de cohérence, il existe une longueur de cohérence caractéristique l_c telle que la contribution à l'amplitude du champ des vitesses des fluctuations de longueur de cohérence supérieure à l_c est négligeable. Un volume B_c de taille l_c contient alors un nombre statistiquement significatif de perturbations de même longueur de cohérence. L'analyse des caractéristiques de l'ensemble des perturbations contenues dans un tel volume B_c permet ainsi d'étudier les propriétés globales du champ des vitesses.

Les théories classiques de formation des grandes structures prédisent que le champ

des vitesses cosmique vérifie l'hypothèse d'homogénéité statistique. Néanmoins elles ne précisent pas la valeur de l'échelle de cohérence caractéristique l_c de ce champ (c'est-à-dire la taille du volume B_c). Or notre connaissance du champ des vitesses est spatialement limitée au volume V couvert par les catalogues actuellement disponibles. L'évaluation des quantités globales (comme le tenseur de corrélation des vitesses) n'est alors possible que si le volume V du catalogue est supérieur au volume caractéristique B_c . Dans un tel cas de figure, le tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ est défini de la façon suivante (voir Monin and Yaglom, Vol.2, Chap. 7 (1975)) :

$$B_{ij}(\mathbf{r}) = \overline{v_i(\mathbf{x}) v_j(\mathbf{x} + \mathbf{r})} \approx \frac{1}{V} \int_V v_i(\mathbf{x}) v_j(\mathbf{x} + \mathbf{r}) d^3\mathbf{x} \quad (C.2)$$

Le tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ possède 9 composantes qui dépendent de la variable vectorielle $\mathbf{r} = (r_1, r_2, r_3)$. Nous voulons souligner encore une fois que cette définition du tenseur n'est valable que si le champ des vitesses est statistiquement homogène dans le volume V couvert par le catalogue. Ceci interdit par exemple la présence d'une fluctuation dominante à elle seule la cinématique de l'échantillon de galaxies étudié.

I.1.1. Si le champ des vitesses est homogène et isotrope :

L'hypothèse d'homogénéité statistique du champ des vitesses est nécessaire lorsqu'on souhaite estimer le tenseur de corrélation des vitesses. Si l'on postule de plus que le champ des vitesses est statistiquement isotrope (voir Monin and Yaglom, Vol. 1, Chap. 2 (1975)), le tenseur $B_{ij}(\mathbf{r})$ ne dépend plus que de la variable $r = \|\mathbf{r}\|$ et peut s'exprimer à l'aide de 2 fonctions scalaires $B_{LL}(r)$ et $B_{NN}(r)$ seulement.

. Hypothèse 2 : Le champ des vitesses est statistiquement isotrope :

Soit $\mathbf{v}'_R(\mathbf{x}) = \mathbf{v}(R(\mathbf{x}))$ le champ résultant d'une transformation du champ des vitesses par une rotation de l'espace R . Si le champ des vitesses est statistiquement isotrope, il doit vérifier quelquesoit R une rotation spatiale :

$$\overline{\mathbf{v}'_R(\mathbf{x})} = \overline{\mathbf{v}(\mathbf{x})} = \text{constante} \quad (C.3)$$

L'hypothèse d'isotropie statistique du champ des vitesses exclue notamment les champs de vitesses admettant une direction privilégiée (les champs constants par exemple).

On peut maintenant simplifier l'expression du tenseur de corrélation des vitesses grâce aux symétries impliquées par les deux hypothèses précédentes.

Pour chaque vecteur \mathbf{r} , on exprime le tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ dans un repère (M, x'_1, x'_2, x'_3) tel que l'axe (Mx'_1) soit parallèle à \mathbf{r} et les axes (Mx'_2) et (Mx'_3) définissent le plan perpendiculaire au vecteur \mathbf{r} . On note $B'_{ij}(\mathbf{r})$ l'expression du tenseur dans ce nouveau repère (par définition, ce tenseur ne dépend plus que de $r = \|\mathbf{r}\|$).

On utilise alors les symétries du problème. En effectuant une rotation de 180° autour de (Mx'_1) , on transforme les composantes du champ des vitesses ($v_1 \rightarrow v_1, v_2 \rightarrow -v_2, v_3 \rightarrow -v_3$). Ce champ étant statistiquement isotrope, cette transformation n'affecte pas les valeurs moyennes (équation (C.3)), ainsi :

$$B'_{12}(\mathbf{r}) = \overline{v_1(\mathbf{x})v_2(\mathbf{x} + \mathbf{r})} = -\overline{v_1(\mathbf{x})v_2(\mathbf{x} + \mathbf{r})} = -B'_{12}(\mathbf{r}) = 0$$

De même $B'_{13}(\mathbf{r}) = 0$. On peut en fait déduire à l'aide de ces symétries par rotation et réflexion que :

$$B'_{12}(\mathbf{r}) = B'_{21}(\mathbf{r}) = B'_{13}(\mathbf{r}) = B'_{31}(\mathbf{r}) = B'_{23}(\mathbf{r}) = B'_{32}(\mathbf{r}) = 0$$

$$B'_{22}(\mathbf{r}) = B'_{33}(\mathbf{r})$$

Deux fonctions sont alors nécessaires afin de caractériser ce tenseur. $B'_{11}(\mathbf{r}) = B_{LL}(\mathbf{r}) = \overline{v_L(\mathbf{x})v_L(\mathbf{x} + \mathbf{r})}$ que l'on nomme fonction de corrélation longitudinale parce qu'elle mesure la corrélation du champ des vitesses $v_L(\mathbf{x})$ dans la direction du vecteur \mathbf{r} . $B'_{22}(\mathbf{r}) = B'_{33}(\mathbf{r}) = B_{NN}(\mathbf{r}) = \overline{v_N(\mathbf{x})v_N(\mathbf{x} + \mathbf{r})}$ la fonction de corrélation latérale qui décrit les corrélations des vitesses $v_N(\mathbf{x})$ dans le plan orthogonal au vecteur \mathbf{r} . L'expression du tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ s'obtient alors par changement de repère (voir Monin and Yaglom, Vol. 2, Chap. 7, équation (12.29) (1975)) :

$$B_{ij}(\mathbf{r}) = [B_{LL}(\mathbf{r}) - B_{NN}(\mathbf{r})] \frac{r_i \cdot r_j}{r^2} + B_{NN}(\mathbf{r}) \delta_{ij} \quad (C.4)$$

$$\text{avec } \delta_{ij} = 1 \text{ si } i = j, 0 \text{ si } i \neq j.$$

Il est utile de définir la trace $\zeta_{vv}(\mathbf{r})$ du tenseur de corrélation des vitesses (Peebles (1987) ou Kaiser (1989)) dont l'expression est invariante par changement de repère :

$$\xi_{vv}(r) = B_{LL}(r) + 2B_{NN}(r) \quad (C.5)$$

On remarque alors que la valeur de la trace pour une séparation $r = 0$ est égale au carré de la dispersion du champ des vitesses :

$$\xi_{vv}(0) = \overline{\mathbf{v}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x})} \quad (C.6)$$

I.2. Le tenseur des différences :

Il est possible de former à partir du champ des vitesses $\mathbf{v}(\mathbf{x})$ une quantité tensorielle invariante par translation. Le tenseur des différences $D_{ij}(\mathbf{r})$ (ou fonctions de structure) est défini de la façon suivante (voir Monin and Yaglom, Vol.2, Chap. 7, équation (13.64') (1975)) :

$$D_{ij}(\mathbf{r}) = \overline{(v_i(\mathbf{x} + \mathbf{r}) - v_i(\mathbf{x})) \cdot (v_j(\mathbf{x} + \mathbf{r}) - v_j(\mathbf{x}))} \quad (C.7)$$

Contrairement au tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$, l'expression du tenseur des différences $D_{ij}(\mathbf{r})$ est donc la même dans tout repère galiléen.

I.2.1. Si le champ des vitesses s'écrit : $\mathbf{v}(\mathbf{x}) = \mathbf{v}_C + \mathbf{v}_H(\mathbf{x}) + \mathbf{v}_{HI}(\mathbf{x})$

Nous avons vu dans la partie A que le champ des vitesses des galaxies se décomposait dans l'approximation Newtonienne en trois termes, un champ des vitesses constant \mathbf{v}_C dépendant du choix du repère absolu des vitesses, le champ de Hubble $\mathbf{v}_H(\mathbf{x})$ caractérisant l'expansion de l'univers et le champ des vitesses propres $\mathbf{v}_p(\mathbf{x})$ engendré par la croissance des fluctuations primordiales de la densité de matière sous l'influence de la gravitation :

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}_C + \mathbf{v}_H(\mathbf{x}) + \mathbf{v}_p(\mathbf{x}) \quad (A.21)$$

. Hypothèse 1 : Le champ des vitesses propres est homogène et isotrope

On suppose que le volume de l'espace dans lequel on étudie le champ des vitesses est assez grand pour que le champ des vitesses propres soit homogène et isotrope $\mathbf{v}_p(\mathbf{x}) = \mathbf{v}_{HI}(\mathbf{x})$ (voir section I.1).

De la même manière que pour le tenseur de corrélation des vitesses, on va montrer que lorsque le champ des vitesses vérifie les hypothèses précédentes, le tenseur des différences est dépendant de $r = || \mathbf{r} ||$ seulement et s'exprime à partir des deux fonctions scalaires $D_{LL}(r)$ et $D_{NN}(r)$.

. Champ des vitesses constant : v_C

Par définition, toutes les composantes $D_{ij}^C(\mathbf{r})$ du tenseur des différences pour un champ de vitesses constant pur sont nulles :

$$D_{ij}^C(\mathbf{r}) = 0 \quad \text{pour tout couple } (i, j) \quad (C.8)$$

. Champ des vitesses homogène et isotrope : $v_{HI}(\mathbf{x})$

Pour un champ homogène et isotrope, le tenseur des différences est soumis aux mêmes symétries que le tenseur de corrélation des vitesses $B_{ij}^{HI}(\mathbf{r})$ (voir section I.1.1). On définit donc les fonctions de structure longitudinale $D_{LL}^{HI}(r)$ et latérale $D_{NN}^{HI}(r)$ (voir Monin and Yaglom, Vol. 2, Chap. 7, équation (13.70) (1975)) :

$$D_{LL}^{HI}(r) = \overline{(v_L^{HI}(\mathbf{x} + \mathbf{r}) - v_L^{HI}(\mathbf{x}))^2} \quad (C.9)$$

$$D_{NN}^{HI}(r) = \overline{(v_N^{HI}(\mathbf{x} + \mathbf{r}) - v_N^{HI}(\mathbf{x}))^2} \quad (C.10)$$

Le tenseur des différences pour un champ des vitesses statistiquement homogène et isotrope s'exprime alors :

$$D_{ij}^{HI}(\mathbf{r}) = [D_{LL}^{HI}(r) - D_{NN}^{HI}(r)] \frac{r_i r_j}{r^2} + D_{NN}^{HI}(r) \delta_{ij} \quad (C.11)$$

On remarque qu'il existe une relation liant le tenseur des différences avec le tenseur de corrélation des vitesses :

$$D_{ij}^{HI}(\mathbf{r}) = 2[B_{ij}^{HI}(0) - B_{ij}^{HI}(\mathbf{r})] \quad (C.12)$$

. Champ de Hubble : $v_H(\mathbf{x})$

Le champ de Hubble agit sur deux points de coordonnées \mathbf{x} et $\mathbf{x} + \mathbf{r}$ comme un champ longitudinale (c'est à dire parallèle à la direction de \mathbf{r}) et d'amplitude proportionnelle à r ($v_H = H_0 \cdot r$). Ainsi en appliquant la même démarche que pour le champ des vitesses homogène et isotrope, le tenseur des différences pour un champ de Hubble pur s'exprime de la façon suivante :

$$D_{ij}^H(\mathbf{r}) = D_{LL}^H(r) \frac{r_i \cdot r_j}{r^2} \quad \text{avec} \quad D_{LL}^H(r) = H_0^2 \cdot r^2 \quad (C.13)$$

. Champ total : $\mathbf{v} = \mathbf{v}_C + \mathbf{v}_H + \mathbf{v}_{HI}$

Le tenseur des différences pour le champ de vitesses total \mathbf{v} se décompose en plusieurs termes :

$$D_{ij}(\mathbf{r}) = D_{ij}^C(\mathbf{r}) + D_{ij}^H(\mathbf{r}) + D_{ij}^{HI}(\mathbf{r}) + \text{termes croisés} \quad (C.14)$$

Des termes croisés interviennent effectivement dans l'équation (C.14) en raison du caractère non-linéaire du calcul tensoriel. Dans ce cas précis néanmoins, ces termes croisés sont tous nuls. Ainsi on a pour le tenseur des différences $D_{ij}(\mathbf{r})$ du champ total :

$$D_{ij}(\mathbf{r}) = [D_{LL}(r) - D_{NN}(r)] \frac{r_i \cdot r_j}{r^2} + D_{NN}(r) \delta_{ij} \quad (C.15)$$

$$\text{avec} \quad D_{LL}(r) = H_0^2 \cdot r^2 + D_{LL}^{HI}(r) \quad \text{et} \quad D_{NN}(r) = D_{NN}^{HI}(r)$$

II. LES INTERETS :

Les théories classiques de formation des grandes structures de l'univers permettent de relier le tenseur de corrélation des vitesses avec la fonction de corrélation spatiale de la distribution des galaxies dans l'espace, et ainsi d'accéder au spectre primordial des fluctuations de la densité de matière. L'étude du tenseur de corrélation des vitesses présente en fait de multiples intérêts.

D'une part, l'évaluation du spectre des fluctuations à l'aide du champ des vitesses est insensible au problème de complétude de l'échantillon rencontré lorsqu'on calcule la fonction de corrélation spatiale des galaxies. Ainsi les corrélations des vitesses donnent accès au spectre des fluctuations à plus grande échelle. Ceci doit permettre la discrimination entre les différents modèles de formation de structures actuellement proposés.

D'autre part, le tenseur de corrélation des vitesses décrit les corrélations de la densité de masse totale. En comparant alors ces corrélations avec la fonction de corrélation des galaxies (la masse lumineuse), une information supplémentaire nous est offerte sur le comportement de la matière lumineuse vis-à-vis de la matière totale (on pourra par exemple évaluer le paramètre de biais b).

Enfin, le tenseur de corrélation des vitesses permet de tester certaines hypothèses sur le champ des vitesses lui-même. Par exemple, on peut en théorie déterminer à quelle échelle le champ des vitesses cesse d'être irrotationnel.

II.1. Les fonctions de corrélations :

Les théories classiques de formation des grandes structures dans l'univers (comme par exemple le modèle standard de matière sombre froide (CDM)) postulent l'existence, à l'époque du découplage matière-rayonnement ($z_d \approx 1000$), de petites fluctuations de la densité de matière. Ce sont ces fluctuations primordiales, amplifiées sous l'influence de la

gravitation, qui donnent naissance à la hiérarchie de structures actuellement observées dans l'espace. Le champ de densité de matière $\rho(\mathbf{r})$ est modélisé par un champ moyen constant $\bar{\rho}$ perturbé par un champ aléatoire scalaire gaussien $\Delta(\mathbf{r})$ (voir par exemple Peebles (1980) ou Kaiser (1989)).

$$\rho(\mathbf{r}) = \bar{\rho}(1 + \Delta(\mathbf{r})) \quad (C.16)$$

Un tel processus est alors complètement caractérisé par le spectre de puissance $P(k) \equiv \langle \Delta(\mathbf{k}) \cdot \Delta^*(\mathbf{k}) \rangle$, avec $\Delta(\mathbf{k})$ la transformée de Fourier du champ des fluctuations $\Delta(\mathbf{r})$ ($\Delta(\mathbf{k}) = (2\pi)^{-1/2} \int \Delta(\mathbf{r}) \exp(-i \mathbf{k} \cdot \mathbf{r}) d^3\mathbf{r}$). La fonction de corrélation spatiale $\xi(r)$ s'exprime en fonction du spectre de puissance $P(k)$ de la façon suivante :

$$\xi(r) = (2\pi^2)^{-1} \int_0^\infty k^2 P(k) j_0(kr) dk \quad (C.17)$$

avec $j_0(x) = \frac{\sin x}{x}$

Chaque modèle prédit en fait une forme ainsi qu'une amplitude pour le spectre de puissance $P(k)$ des fluctuations de la densité de matière. L'évaluation de la fonction de corrélation spatiale $\xi(r)$ ou du tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ à partir des données observationnelles donne accès à ce spectre de puissance et donc permet de tester la fiabilité des différents modèles de formation de structures.

II.1.1. La fonction de corrélation spatiale des galaxies :

Il est possible d'évaluer la fonction de corrélation $\xi(r)$ à partir de la distribution des galaxies dans l'espace. Grossièrement, $\xi(r)$ est pour chaque séparation r , proportionnelle au nombre de paires de galaxies de l'échantillon séparées par la distance r . L'estimation de cette fonction présente néanmoins plusieurs problèmes.

D'une part la distribution des galaxies (c'est à dire de la masse lumineuse) ne reflète pas forcément la distribution de la masse totale présente dans l'univers. Ainsi, l'étude de la distribution des galaxies permet seulement de déterminer la fonction de corrélation de la masse lumineuse $\xi_L(r)$.

D'autre part, les échantillons de galaxies observées possèdent des effets de sélection (par exemple une magnitude apparente limite de coupure) qui compliquent l'évaluation de la fonction de corrélation $\xi_L(r)$. En raison de ces effets de sélection, le nombre de galaxies

observées décroît rapidement lorsqu'on dépasse une certaine distance caractéristique. Au delà de cette distance, le nombre de paires de galaxies séparées par une distance r diminue considérablement en fonction de r . En pratique donc, la statistique permettant l'obtention de $\xi_L(r)$ à partir de la distribution des galaxies observées cesse d'être fiable pour des séparations r de l'ordre de $10 h^{-1}$ Mpc (voir par exemple Davis et al. (1983)).

Enfin, les distances des galaxies sont déduites pour ce genre d'études, à partir du décalage vers le rouge mesuré ($D = cz$). On fait alors implicitement l'hypothèse que le champ des vitesses radiales propres est nul, ou du moins négligeable. Il a été montré (Davis et al. (1983), Kaiser (1987)) que cette hypothèse entraînait des erreurs systématiques dans l'estimation de la fonction de corrélation spatiale des galaxies $\xi_L(r)$. La correction de ce type d'erreurs s'avère hélas difficile.

La fonction de corrélation spatiale des galaxies a été estimée pour plusieurs échantillons de galaxies et s'accorde relativement bien dans la plage des distances considérées (typiquement de $0.1 h^{-1}$ Mpc à $10 h^{-1}$ Mpc) avec une loi de puissance (Peebles (1980), Davis et al. (1983)) :

$$\xi_L(r) = \left(\frac{r}{r_0}\right)^{-1.8} \quad \text{avec } r_0 = 5 h^{-1} \text{ Mpc} \quad (C.18)$$

Cette valeur de $\xi_L(r)$ ne remet pas en cause les modèles classiques de formation de structures (Bardeen et al. (1987)). Aux petites échelles donc, les théories comme CDM restent cohérentes avec les données observationnelles.

II.1.2. Le tenseur de corrélation des vitesses :

Les fluctuations de la densité de matière induisent en se développant un champ des vitesses propres. Dans le domaine linéaire (c'est à dire lorsque les perturbations de la densité de matière restent petites $\Delta(\mathbf{r}) \ll 1$), l'équation de continuité permet de déduire l'expression de ce champ. Le champ des vitesses propres $\mathbf{v}(\mathbf{r})$ est ainsi un champ vectoriel aléatoire gaussien relié au champ scalaire $\Delta(\mathbf{r})$ par l'équation (dans l'espace de Fourier) :

$$\mathbf{v}(\mathbf{k}) = -i H_0 f(\Omega_0) \mathbf{k}/k^2 \Delta(\mathbf{k}) \quad (C.19)$$

avec $f(\Omega_0) \approx \Omega_0^{0.6}$

On peut alors utiliser le champ des vitesses propres afin d'estimer le spectre de puissance $P(k)$ des fluctuations de la densité de matière. La méthode des multipôles par

exemple (Szalay (1989) ou Regös et *al.* (1989)) permet théoriquement d'évaluer ce spectre (en pratique néanmoins, des difficultés intrinsèques à la méthode apparaissent). Plus généralement, on préfère utiliser le tenseur de corrélation des vitesses.

Si le volume de l'espace dans lequel on étudie le champ des vitesses est assez grand, c'est à dire si l'hypothèse d'homogénéité et d'isotropie du champ s'applique à l'échantillon (voir section I.1.1), alors les équations suivantes sont vérifiées (Davis et *al.* (1985), Górski (1988), Kaiser (1989) ou Szalay (1989)) :

$$B_{LL}(\tau) = \frac{H_0^2 \Omega_0^{1.2}}{2\pi^2} \int_0^\infty P(k) \frac{j_1(kr)}{kr} dk \quad (C.20)$$

$$B_{NN}(\tau) = \frac{H_0^2 \Omega_0^{1.2}}{2\pi^2} \int_0^\infty P(k) (j_0(kr) - 2 \frac{j_1(kr)}{kr}) dk \quad (C.21)$$

$$\text{avec } j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}$$

Et la trace $\zeta_{vv}(\tau)$ du tenseur de corrélation des vitesses s'exprime en fonction du spectre de puissance $P(k)$:

$$\xi_{vv}(\tau) = \frac{H_0^2 \Omega_0^{1.2}}{2\pi^2} \int_0^\infty P(k) j_0(kr) dk \quad (C.22)$$

Ainsi l'évaluation des composantes latérale et longitudinale du tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ nous donne accès au spectre de puissance $P(k)$ de la densité de matière. De plus, l'estimation des composantes $B_{NN}(\tau)$ et $B_{LL}(\tau)$ n'est pas soumise au problème des effets de sélection rencontré lors de la détermination de la la fonction de corrélation spatiale des galaxies (voir partie A, section III.2). Le tenseur de corrélation des vitesses nous offre donc une information sur $P(k)$ à de plus grandes échelles (environ de $5 h^{-1}$ Mpc à $40 h^{-1}$ Mpc).

II.1.3. Discrimination entre différents modèles :

Le modèle CDM ou le modèle PIB (Peebles (1987)) prédisent bien la valeur de la fonction de corrélation observée pour les galaxies (Bardeen et *al.* (1987)). Néanmoins certaines

études montrent que la dispersion du champ des vitesses propres mesurées n'est pas en accord avec la théorie CDM (Clutton-Brock et al. (1981), Aaronson et al. (1982a)). De plus la fonction de corrélation spatiale amas-amas est trop importante par rapport aux prédictions classiques. Mais ces résultats sont contaminés par des incertitudes et nécessitent confirmation.

En fait, toutes les observations montrent des corrélations trop fortes aux grandes échelles. C'est donc l'analyse des corrélations de la densité de matière à grande échelle qui va permettre la discrimination entre les différents modèles de formation de structure (Kaiser (1988), Vittorio et al. (1989)). Maintenant que nous possédons des grands catalogues de vitesses propres, le tenseur de corrélation des vitesses paraît être un outil puissant lorsqu'on veut confronter les modèles théoriques avec les observations (Vittorio et al. (1986), Kaiser et al. (1989), Landy et al. (1992)). Nous verrons par la suite (section IV) qu'en pratique, il est actuellement impossible de mettre en défauts les théories existentes à l'aide du tenseur de corrélation des vitesses (Juszkiewicz et al. (1990)).

II.2. Le paramètre de biais :

Le modèle standard de matière sombre froide (CDM) a été assez vite modifié en vue d'expliquer les observations récentes comme par exemple l'amplitude de la fonction de corrélation des galaxies. Un nouveau paramètre a été introduit dans cette théorie, qui quantifie la différence entre la masse lumineuse (les galaxies) et la masse totale (voir par exemple Davis et al. (1985), Peebles et al. (1989) ou Kaiser (1989)). Le paramètre de biais b relie en fait le champ des fluctuations de masse totale $\Delta_T(\mathbf{r})$ avec le champ des fluctuations de la masse lumineuse $\Delta_L(\mathbf{r})$ suivant l'équation suivante :

$$\Delta_L(\mathbf{r}) = b \Delta_T(\mathbf{r}) \quad (C.23)$$

La détermination de b est alors possible en comparant le champ des vitesses propres (qui est induit par les fluctuations de masse totale) avec la distribution spatiale des galaxies dans l'espace. En fait on utilise la relation (C.22) qui lie la trace $\xi_{vv}(\mathbf{r})$ du tenseur de corrélation des vitesses et le spectre de puissance $P_T(k)$ ($P_L(k) = b^2 P_T(k)$). L'équation (C.22), modifiée afin de prendre en compte le paramètre de biais b s'exprime alors :

$$\xi_{vv}(r) = \frac{H_0^2 \Omega_0^{1.2}}{2\pi^2 b^2} \int_0^\infty P_L(k) j_0(kr) d^3k \quad (C.24)$$

On évalue le spectre de puissance de la matière lumineuse $P_L(k)$ à partir de la fonction de corrélation spatiale des galaxies $\zeta_L(r)$ (équation (C.17)).

Aux petites échelles (typiquement 1 Mpc), les mesures simultanées des vitesses ainsi que de la fonction de corrélation spatiale suggèrent une valeur du paramètre de biais d'environ 2.5 (Davis et al. (1985), Juskiwicz et al. (1989)). En revanche, les fortes corrélations de vitesses observées aux grandes échelles imposent une limite supérieure pour la valeur de b ($b < 2$) (Kaiser (1988), Kaiser (1989), Lahav et al. (1990)). Ces résultats tendent à montrer que le lien entre la distribution de la matière lumineuse et de la matière totale est plus complexe que celui postulé dans l'équation (C.23). Une étude approfondie sur le processus responsable de la formation des galaxies est à même d'éclaircir ce genre de problème.

II.3. La connaissance du champ des vitesses :

Les théories de formation de structures prédisent qu'aux grandes échelles, le champ des vitesses propres actuellement observé est irrotationnel ($\nabla \times \mathbf{v} = 0$). La dynamique régissant l'évolution des perturbations primordiales de la densité de matière stipule en effet que la composante rotationnelle du champ des vitesses décroît au cours du temps et seule sa composante divergente subsiste encore à notre époque (Peebles (1980)). Néanmoins, lorsque les fluctuations de la densité de matière deviennent suffisamment importantes ($\Delta(\mathbf{r}) \sim 1$), les équations d'évolution sont dominées par des termes non-linéaires qui induisent à nouveau une composante rotationnelle pour le champ des vitesses. Il apparaît ainsi dans les modèles de formation de structure une échelle privilégiée τ_c au dessous de laquelle le champ des vitesses dérivant d'un potentiel est contaminé par une composante rotationnelle. La détermination de cette échelle nous permettrait alors de tester la validité des différentes théories, certaines prédisant une faible valeur pour ce paramètre (CDM par exemple), d'autres une valeur considérablement plus importante (Henriksen et al. (1990)).

Si l'échantillon étudié occupe un volume de l'espace suffisamment grand pour que l'hypothèse d'homogénéité et d'isotropie statistique du champ des vitesses soit vérifiée,

le tenseur de corrélation des vitesses permet d'évaluer ce paramètre r_c . Si le champ des vitesses est irrotationnel par exemple ($\nabla \times \mathbf{v} = 0$), il existe en fait des relations liant les composantes latérale $B_{NN}(r)$ et longitudinale $B_{LL}(r)$ du tenseur (Monin and Yaglom (1975), équation 12.70) :

$$B_{NN}(r) = \frac{1}{r} \int_0^r B_{LL}(x) dx \quad (C.25)$$

On remarque alors que (Monin and Yaglom (1975)) :

$$\text{pour toute séparation } r, \quad B_{NN}(r) > B_{LL}(r) \quad (C.26)$$

De même si le champ des vitesses propres est purement rotationnel ($\nabla \cdot \mathbf{v} = 0$) (Monin and Yaglom (1975), équation 12.70) :

$$B_{LL}(r) = \frac{2}{r^2} \int_0^r B_{NN}(x) x dx \quad (C.27)$$

On peut donc en théorie estimer l'échelle r_c au delà de laquelle la composante rotationnelle du champ des vitesses est négligeable (l'équation (C.25) n'est plus vérifiée pour toute séparation r inférieure à r_c).

III. LE CALCUL DU TENSEUR DE CORRELATION DES VITESSES :

Le calcul du tenseur de corrélation des vitesses $B_{ij}(\mathbf{r})$ nécessite la connaissance de toutes les composantes du champ des vitesses $\mathbf{v}(\mathbf{x})$. Or les observations ne donnent accès qu'à la composante radiale $v_r(\mathbf{x})$ du champ des vitesses (voir partie B, section II). Néanmoins, si le champ des vitesses est statistiquement homogène et isotrope dans le volume de l'espace étudié, il est possible d'estimer statistiquement les composantes du tenseur de corrélation des vitesses à partir du champ des vitesses radiales seulement. Les valeurs des composantes latérale $B_{NN}(r)$ et longitudinale $B_{LL}(r)$ du tenseur sont alors évaluées par une régression multiple pour chaque séparation r .

III.1. Nécessité d'une hypothèse :

La connaissance du champ de vitesses radiales $v_r(\mathbf{x})$ d'un champ de vitesse quelconque $\mathbf{v}(\mathbf{x})$ n'est pas suffisante si l'on désire calculer les 9 composantes $B_{ij}(\mathbf{r})$ du tenseur de corrélation des vitesses. Il est alors nécessaire de faire une hypothèse sur le champ des vitesses qui lui conférera des propriétés particulières.

On sait par exemple que si le champ des vitesses étudié est irrotationnel, on peut reconstruire complètement ce champ à partir de sa composante radiale seulement (voir partie E) et ainsi accéder à toute l'information concernant le tenseur de corrélation des vitesses. Mais dans la mesure où l'on souhaite justement tester si la composante rotationnelle du champ est bien nulle à grande échelle (voir section II.3), une hypothèse moins puissante est désirable.

Les théories classiques de formation de structures stipulent que le champ des vitesses propres est actuellement un champ vectoriel aléatoire gaussien aux grandes échelles (équation C.19), certainement contaminé aux petites échelles par une composante rotationnelle apparaissant durant la phase non-linéaire de l'évolution des structures peuplant notre

univers. Si l'on étudie un tel champ de vitesses dans un volume d'espace suffisamment grand, c'est à dire si ce volume est statistiquement représentatif de la répartition globale de la matière dans l'univers, alors ce champ possède la propriété d'être statistiquement homogène et isotrope.

Nous allons donc poser l'hypothèse que l'échantillon de galaxies étudié est assez grand pour être représentatif de la population globale des galaxies dans l'univers. Le champ des vitesses doit alors être statistiquement homogène et isotrope pour cet échantillon et satisfait les relations (C.1) et (C.3). Le tenseur de corrélation des vitesses est caractérisé par les fonctions de corrélation longitudinale $B_{LL}(r)$ et latérale $B_{NN}(r)$ et vérifie l'équation (C.4). Nous montrons dans la section suivante comment évaluer ces deux fonctions à partir de la composante radiale du champ des vitesses $v_r(\mathbf{x})$.

Nous voulons faire remarquer dès à présent que l'évaluation de ces deux fonctions $B_{LL}(r)$ et $B_{NN}(r)$ repose entièrement sur la validité de l'hypothèse d'homogénéité et d'isotropie du champ des vitesses étudié. Si l'hypothèse s'avère fautive, on ne peut alors tirer aucune conclusion de l'estimation des fonctions $B_{LL}(r)$ et $B_{NN}(r)$. On montrera d'ailleurs dans la section IV que cette hypothèse ne semble pas vérifiée par les catalogues de vitesses propres actuellement disponibles, le champ des vitesses des galaxies dans ces échantillons étant dominé par la présence d'un flot de vitesse convergent cohérent à grande échelle. La partie D sera d'ailleurs dédiée à démontrer l'existence d'un tel flot (le "Grand Attracteur"). Il nous est donc impossible actuellement d'estimer sans erreurs les composantes du tenseur de corrélation des vitesses.

III.2. L'évaluation de $B_{LL}(r)$ et $B_{NN}(r)$:

Il s'agit donc de déterminer à partir de la composante radiale $v_r(\mathbf{x})$ d'un champ de vitesses statistiquement homogène et isotrope, les composantes latérale $B_{NN}(r)$ et longitudinale $B_{LL}(r)$ du tenseur de corrélation des vitesses.

Pour une séparation $r = \|\mathbf{r}\|$ fixée, on sélectionne toutes les paires de galaxies de l'échantillon telles que leur séparation soit comprise entre r et $r + \Delta r$.

On forme ensuite à partir d'un angle solide $\Delta\hat{r}$, une collection de N classes disjointes de direction $C_i(r) = [\hat{\mathbf{r}}_i, \hat{\mathbf{r}}_i + \Delta\hat{r}]$ telles que ces classes recouvrent entièrement toutes les directions de l'espace ($\Delta\hat{r}$ sera choisi en fonction des caractéristiques de l'échantillon).

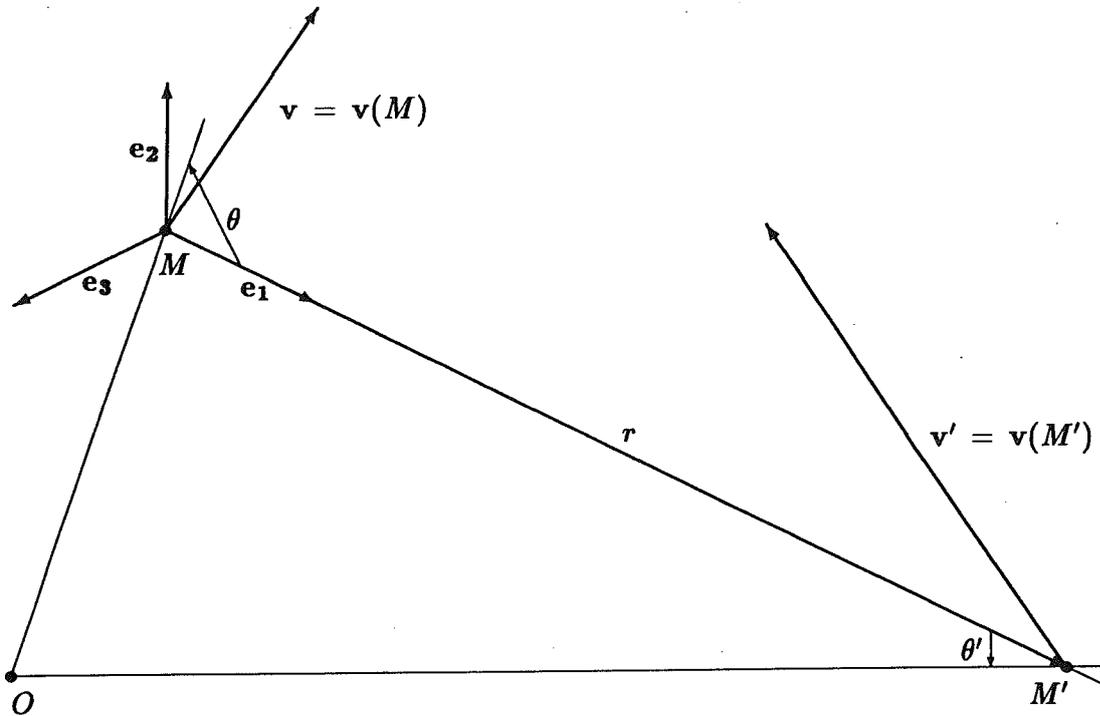


Figure C.1 : La géométrie du nouveau repère (M, e_1, e_2, e_3)

Pour chacune de ces classes de direction C_i , on sélectionne alors toutes les paires de galaxies MM' telles que la direction de leur vecteur séparation $\hat{M}\hat{M}'$ appartienne à $C_i(r)$ ($\hat{r}_i < \hat{M}\hat{M}' < \hat{r}_i + \Delta\hat{r}$). Pour chacune de ces paires MM' , on définit un nouveau repère (M, e_1, e_2, e_3) tel que e_1 soit colinéaire à r (voir figure (C.1)). On exprime dans ce nouveau repère le champ des vitesses v au point M , $v(OM) = (v_1, v_2, v_3)$ et le champ des vitesses au point M' ($OM' = OM + r$), $v(OM') = (v'_1, v'_2, v'_3)$. Par définition (voir section I.1.1) on a donc $B'_{ij}(r) = \langle v_i v'_j \rangle$, l'opération $\langle . \rangle$ indiquant une moyenne sur toutes les paires MM' de galaxies de l'échantillon appartenant à la classe de direction $C_i(r)$.

En notant les cosinus directeurs $OM/OM = (\cos\theta, \alpha, \beta)$ et $OM'/OM' = (\cos\theta', \alpha', \beta')$ (voir figure C.1), on obtient alors la relation suivante entre le produit des vitesses radiales $v_r \cdot v'_r = v_r(OM) \cdot v_r(OM')$ et les composantes $B'_{ij}(r)$ du tenseur de corrélation des vitesses exprimé dans le repère (M, e_1, e_2, e_3) :

$$\langle v_r.v'_r \rangle = B'_{11}(r) \langle \cos\theta.\cos\theta' \rangle + B'_{22}(r) \langle \alpha.\alpha' \rangle + B'_{33}(r) \langle \beta.\beta' \rangle$$

Les termes croisés n'apparaissent pas dans cette équation puisque, pour $i \neq j$, les $B'_{ij}(r)$ sont tous nuls par symétrie (voir section I.1.1). De plus la géométrie du problème impose que $\alpha.\alpha' + \beta.\beta' = \sin\theta.\sin\theta'$.

Ainsi pour chaque séparation $r = || r ||$, on possède une collection de N classes de direction $C_i(r)$ telles que :

$$\langle v_r.v'_r \rangle_i = B_{LL}(r) \langle \cos\theta.\cos\theta' \rangle_i + B_{NN}(r) \langle \sin\theta.\sin\theta' \rangle_i \quad (C.27)$$

On obtient ainsi les valeurs de $B_{LL}(r)$ et $B_{NN}(r)$ par une régression multiple. En posant :

$$Z_i = \langle v_r.v'_r \rangle_i = \frac{1}{N_C} \sum_{C_i(r)} v_r(M).v_r(M'), \quad X_i = \langle \cos\theta.\cos\theta' \rangle_i, \quad Y_i = \langle \sin\theta.\sin\theta' \rangle_i$$

L'équation (C.27) se réécrit :

$$Z_i = B_{LL}(r) X_i + B_{NN}(r) Y_i$$

Il reste à définir la fonction covariance $S(x, y)$ pour un ensemble de N réalisations (x_i, y_i) des variables x et y :

$$S(x, y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

On a alors pour chaque classe de séparation $[r, r + \Delta r]$:

$$B_{LL}(r) = \frac{S(Y, Z) S(X, Y) - S(X, Z) S(Y, Y)}{S(X, X) S(Y, Y) - S(X, Y) S(X, Y)} \quad (C.28)$$

$$B_{NN}(r) = \frac{S(X, Z) S(X, Y) - S(Y, Z) S(X, X)}{S(X, X) S(Y, Y) - S(X, Y) S(X, Y)} \quad (C.29)$$

L'estimation de $B_{LL}(r)$ et $B_{NN}(r)$ est soumise à des erreurs statistiques de variances $\sigma_{LL}(r)$ et $\sigma_{NN}(r)$ (voir par exemple Hald (1952)) :

$$\sigma_{LL}(r)^2 = \frac{S(Y, Y)}{S(X, X) S(Y, Y) - S(X, Y) S(X, Y)} \sigma(r)^2 \quad (C.30)$$

$$\sigma_{NN}(r)^2 = \frac{S(X, X)}{S(X, X) S(Y, Y) - S(X, Y) S(X, Y)} \sigma(r)^2 \quad (C.31)$$

$$\text{avec } \sigma(r)^2 = \frac{1}{N-3} \sum_{i=1}^N [Z_i - B_{LL}(r) X_i - B_{NN}(r) Y_i]^2$$

Nous avons donc déterminé l'expression des composantes latérale et longitudinale du tenseur de corrélation des vitesses sous l'hypothèse d'homogénéité et d'isotropie du champ des vitesses. Nous désirons faire plusieurs remarques.

. Remarque 1 :

La méthode précédemment exposée n'est valide que lorsque le champ de vitesses étudié est statistiquement homogène et isotrope. Si le champ des vitesses est contaminé par la présence d'un champ constant ou d'un champ de Hubble par exemple, les termes croisés du tenseur $B'_{ij}(r)$ ne s'annulent plus pour chaque classe de direction C_i . L'évaluation de $B_{NN}(r)$ et $B_{LL}(r)$ par la technique de régression multiple est alors soumise à des erreurs.

Nous avons vu dans la section I.2 que l'étude du tenseur des différences $D_{ij}(r)$ permet de traiter un champ de vitesses composé d'un champ de Hubble, d'un champ constant et d'un champ homogène et isotrope (équation (C.15)). Hélas, l'estimation des composantes $D_{NN}(r)$ et $D_{LL}(r)$ de ce tenseur est impossible à partir de la connaissance du champ des vitesses radiales $v_r(x)$ seulement. On ne peut pas en effet construire de quantité à partir de $v_r(x)$ qui soit invariante par changement de repère galiléen. Ainsi le champ des vitesses que l'on veut étudier ne doit comporter ni flot de vitesse constante, ni champ de Hubble résiduel.

. Remarque 2 :

La méthode de régression multiple est appliquée dans la littérature (Kaiser (1989), (Groth et al. (1989), Górski et al. (1989)) dans le cas limite où chaque classe C_i ne contient qu'une seule paire de galaxies. Nous voulons faire remarquer que les termes

$\langle v_i \cdot v_j' \rangle$ pour $i \neq j$ ne s'annulent alors plus (la moyenne est prise sur un élément seulement). Si ce choix n'affecte pas l'estimation des fonctions $B_{NN}(r)$ et $B_{LL}(r)$, ceci risque en revanche d'augmenter les valeurs des variances $\sigma_{NN}(r)$ et $\sigma_{LL}(r)$ impliquées dans la méthode de régression (équations (C.30) et (C.31)). Le nombre N de classes de direction C_i pour chaque classe de séparation $[r, r + \Delta r]$ doit donc vérifier les deux conditions suivantes. D'une part, il faut que ce nombre de classes N soit assez important pour permettre l'application de la technique de régression multiple. D'autre part, il est désirable que les paires de galaxies dans chaque classe C_i soit assez nombreuses afin que les composantes croisées du tenseur $B'_{ij}(r)$ s'annulent statistiquement (c'est à dire qu'il n'y ai pas trop de classes C_i). Le choix de N dépendra en fait de l'échantillon de galaxies considéré.

. Remarque 3 :

L'évaluation des composantes latérale et longitudinale du tenseur de corrélation des vitesses est en principe insensible aux effets de sélection auxquels est soumis l'échantillon. Néanmoins en pratique, la distribution spatiale très dispersée des données peut affecter des poids fantômes à certaines galaxies. On introduit alors une pondération spatiale $w(\mathbf{x})$ afin de corriger les inhomogénéités de cette distribution. La fonction de covariance est donc modifiée et s'exprime maintenant :

$$S(x, y) = \sum_{i=1}^N w_i^2 (x_i - \bar{x})(y_i - \bar{y})$$

La modélisation de cette fonction de pondération $w(\mathbf{x})$ varie largement suivant les auteurs (Kaiser (1989), Kaiser et al. (1989), Groth et al. (1989), Górski et al. (1989)) et les arguments en faveur d'une fonction plutôt qu'une autre demeurent assez obscurs. A ces problèmes de pondération viennent s'ajouter les effets de bords qui rendent impossible l'évaluation des composantes du tenseur pour des séparations r supérieures de 2/3 à la taille de l'échantillon (Groth et al. (1989)).

IV. LES APPLICATIONS :

Nous évaluons dans cette section les composantes longitudinale et latérale du tenseur de corrélation des vitesses pour le catalogue de galaxies elliptiques NEC (partie B, section IV.) ainsi que pour l'échantillon de spirales de Aaronson et *al.* (1982b).

Les estimations de $B_{LL}(r)$ et $B_{NN}(r)$ ne s'accordent pas avec les prédictions des modèles classiques de formation de structures comme CDM ou PIB, l'amplitude des corrélations observées étant trop grande. Néanmoins l'importance des barres d'erreurs impliquées dans la méthode de régression (voir section III.) indique clairement que les valeurs estimées des composantes $B_{LL}(r)$ et $B_{NN}(r)$ ne sont pas fiables. Il semble en fait que l'hypothèse d'homogénéité et d'isotropie du champ des vitesses propres ne soit pas vérifiée par les 2 catalogues. Ainsi l'évaluation du tenseur de corrélation des vitesses se révèle actuellement impossible.

De plus nous montrons que l'incertitude présente dans le processus de détermination de la distance des galaxies (voir partie B) biaise l'évaluation des fonctions $B_{LL}(r)$ et $B_{NN}(r)$. La correction d'un tel biais apparaît fort complexe.

IV.1. Les données :

Nous avons présenté le catalogue de Faber et *al.* (1989) (NEC) dans la partie B (section IV.). Cet échantillon de galaxies elliptiques possède une profondeur d'environ 6000 km.s^{-1} et permet donc d'étudier le champ des vitesses propres à grande échelle.

La distribution spatiale des galaxies de cet échantillon n'est hélas pas homogène ce qui risque de créer des problèmes lors de l'évaluation des composantes du tenseur de corrélation des vitesses. Nous avons donc simulé un échantillon de galaxies possédant la même distribution spatiale que NEC, mais pour lequel le champ des vitesses radiales propres est aléatoire, de moyenne nulle et de dispersion 300 km.s^{-1} . Ce catalogue (NECHI)

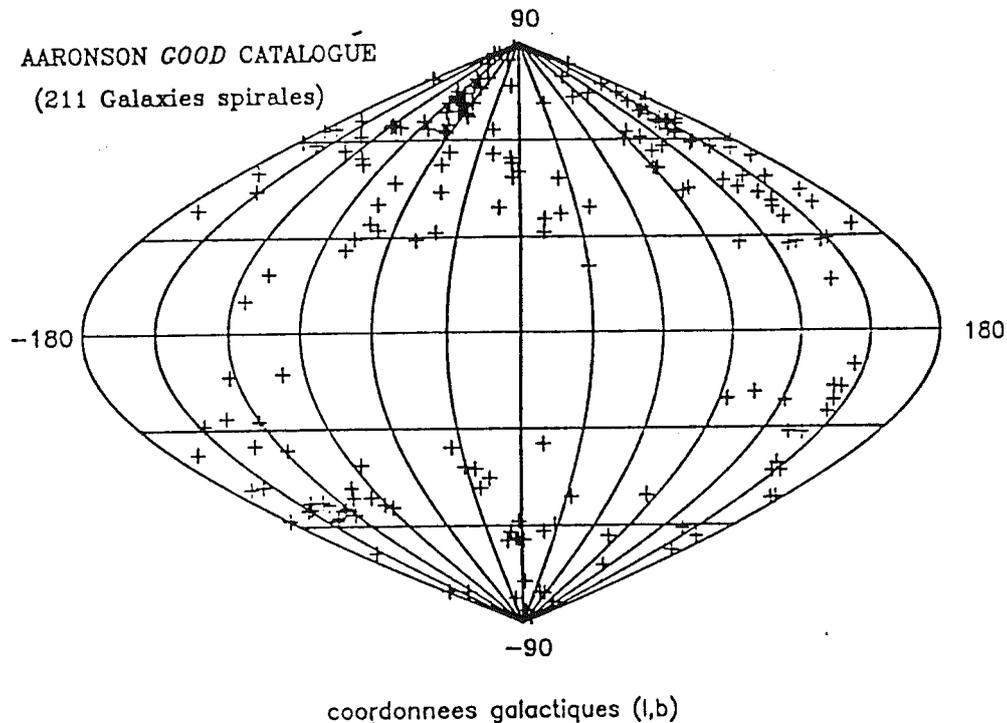


Figure C.2 : Répartition spatiale des galaxies du catalogue de Aaronson et *al.*

nous permettra alors de visualiser les erreurs dues au mauvais échantillonnage ainsi qu'aux effets de bords (voir section IV.2.1).

De plus nous analyserons le catalogue NECMGA, c'est à dire un échantillon possédant le champ des vitesses propres de NEC auquel on a soustrait un flot de vitesses convergent en un point. Un tel flot représente la possible contribution d'un "Grand Attracteur" (Lynden-Bell et *al.* (1989)) dominant la cinématique du catalogue NEC (voir partie D).

Le catalogue de Aaronson et *al.* (1982b) est un échantillon de galaxies spirales dont les distances sont évaluées par la relation de Tully-Fisher. Nous sélectionnons 211 galaxies pour lesquelles l'estimation des distances est précise (Górski et *al.* (1989)) et qui appartiennent pour la plupart au super-amas local (Aaronson et *al.* (1982a)). La profondeur de ce sous-catalogue (GAAR) est d'environ 3000 km.s^{-1} . Nous simulerons un catalogue (GAARHI) possédant la même distribution spatiale que GAAR et un champ des vitesses aléatoire (comme NECHI) afin d'examiner les effets d'échantillonnage et un catalogue

GAARMGA (comme NECMGA) . L'étude du catalogue de Aaronson et *al.* présente pour nous un double intérêt.

D'une part, il permet d'analyser la différence de comportement entre les populations de galaxies spirales et elliptiques. On sait par exemple que les galaxies elliptiques peuplent préférentiellement les amas de galaxies alors que les spirales se répartissent plus uniformément dans l'espace. Un tel comportement explique peut être la différence de corrélation des vitesses observée entre ces deux types de population (Kashlinsky (1989)).

D'autre part, l'échantillon GAAR possède une incertitude intrinsèque impliquée dans les indicateurs secondaires de distance plus faible ($\Delta \approx 15\%$) que le catalogue NEC. De plus l'estimation de la distance de ces 211 galaxies semble moins soumise aux biais statistiques que l'échantillon d'elliptiques.

IV.2. L'estimation de $B_{LL}(r)$ et $B_{NN}(r)$:

Nous appliquons donc la technique de régression afin d'évaluer les composantes $B_{LL}(r)$ et $B_{NN}(r)$ du tenseur de corrélation des vitesses. Nous avons choisi un pas $\Delta r = 500$ km.s⁻¹ et un angle solide $\Delta \hat{r}$ divisant la sphère en 400 directions différentes (voir section III.2). Nous préférons traiter chaque paire de galaxies avec le même poids ($\omega_i = 1$) (Kaiser (1989), Kaiser et *al.* (1989)), plutôt que d'utiliser certaines fonctions de pondération rencontrées dans la littérature (Groth et *al.* (1989), Górski et *al.* (1989)).

IV.2.1. Les effets de bords et d'échantillonnage :

Nous testons donc l'influence de l'échantillonnage ainsi que des effets de bords sur les catalogues NECHI et GAARHI définis précédemment section IV.1 .

Nous montrons figure (C.3) l'estimation de $B_{LL}(r)$ et $B_{NN}(r)$ ainsi que les barres d'erreurs impliquées dans la méthode de régression ($\sigma_{LL}(r)$ et $\sigma_{NN}(r)$) pour le catalogue NECHI. On voit clairement que l'évaluation des composantes du tenseur cesse d'être fiable pour des séparations r supérieures à 5500 km.s⁻¹. De plus le nombre de paires de galaxies possédant une séparation inférieure à 1500 km.s⁻¹ n'est pas suffisant. Si en théorie l'inhomogénéité spatiale de l'échantillon n'affecte pas l'évaluation de $B_{LL}(r)$ et $B_{NN}(r)$,

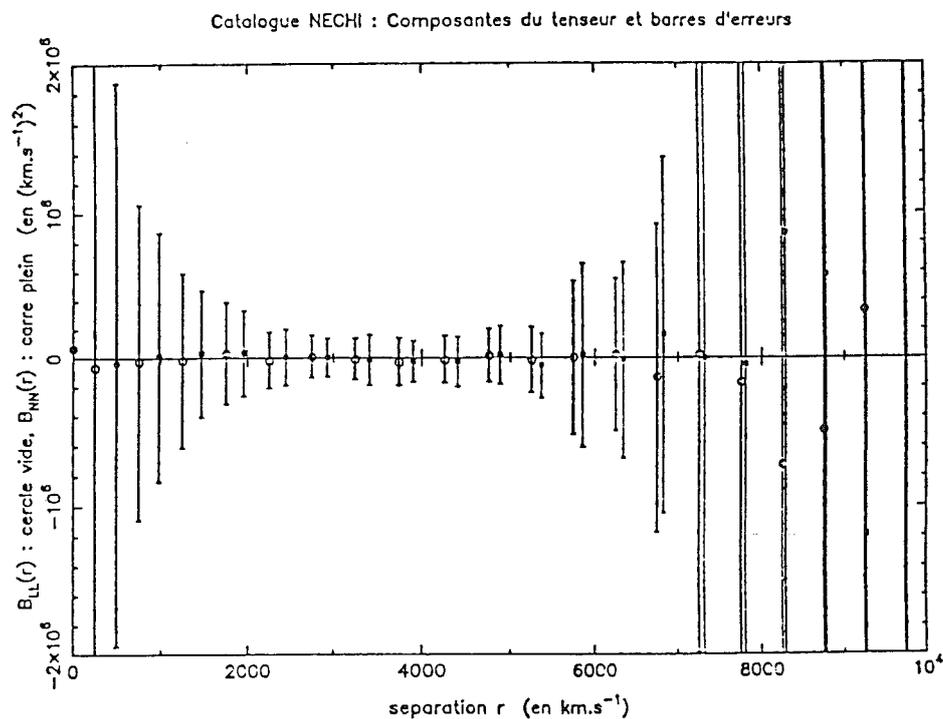


Figure C.3 : L'évaluation de $B_{LL}(r)$ et $B_{NN}(r)$ avec le catalogue NEC est possible pour des séparations comprises entre 1500 et 5500 km.s^{-1}

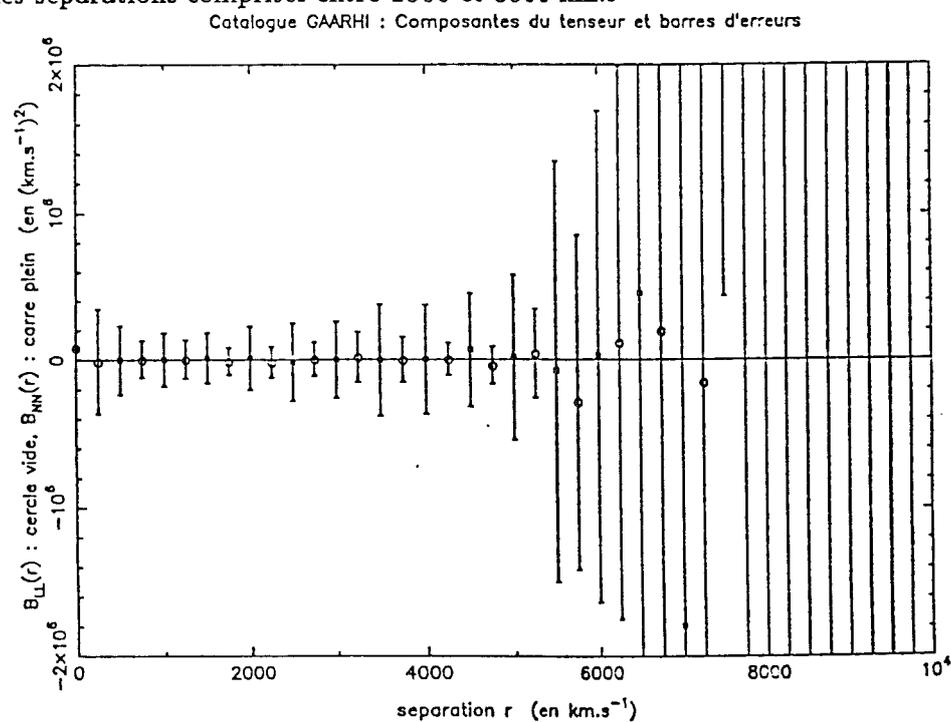


Figure C.4 : L'évaluation de $B_{LL}(r)$ et $B_{NN}(r)$ avec le catalogue GAAR est possible pour des séparations inférieures à 5000 km.s^{-1}

en pratique la statistique pour le catalogue NEC est correcte pour des séparations entre 1500 et 5500 km.s⁻¹. L'amplitude des barres d'erreurs ($\approx 3 \cdot 10^5$ (km.s⁻¹)²) nous servira dans la suite de référence pour déterminer dans quelles mesures la régression est fiable.

La figure (C.4) décrit $B_{LL}(r)$ et $B_{NN}(r)$ et les barres d'erreurs pour le catalogue GAARHI. Les barres d'erreurs deviennent importantes à partir de r supérieur à 5000 km.s⁻¹. On remarque néanmoins que le catalogue GAAR ne souffre pas d'un mauvais échantillonnage aux petites échelles comme le catalogue NEC. On peut expliquer ce comportement d'une part parce que l'échantillon GAAR est moins profond et donc accueille plus de paires de galaxies possédant une petite séparation, d'autre part parce que la distribution des galaxies de GAAR est plus uniforme que celle de NEC.

IV.2.2. Les fonctions $B_{LL}(r)$ et $B_{NN}(r)$:

Les composantes longitudinale et latérale $B_{LL}(r)$ et $B_{NN}(r)$ du tenseur de corrélation des vitesses estimées pour le catalogue d'elliptiques NEC sont exposées figure (C.5), et celles pour le catalogue de spirales GAAR figure (C.6).

Nous montrons aussi les prédictions théoriques pour deux modèles de formation de structures ($B_{LL}(r)$: ligne continue, $B_{NN}(r)$: ligne hachurée). Le modèle CDM classique assume un spectre de puissance $P(k)$ proportionnel à k ($P(k) \propto k$) alors que le modèle PIB ("Peebles Isocurvature Baryons", Peebles (1987)) possède un spectre en k^{-1} ($P(k) \propto k^{-1}$). Nous nous référons à Groth et al. (1989) pour la normalisation du spectre de puissance. Les corrélations de vitesses sont nettement plus importantes pour le modèle PIB que pour le modèle CDM dans la mesure où PIB a été créé en vue d'expliquer ces fortes corrélations observées.

On peut tirer deux conclusions des figures (C.5) et (C.6). D'une part, si l'estimation des composantes $B_{LL}(r)$ et $B_{NN}(r)$ est fiable, c'est à dire si l'hypothèse d'homogénéité et d'isotropie du champ des vitesses est vérifiée par les catalogues NEC et GAAR, alors les modèles de formation de structures du type CDM ou PIB ne prédisent pas les fonctions de corrélation observées. D'autre part les résultats ne sont pas les mêmes pour l'échantillon de galaxies spirales ou pour celui d'elliptiques. Le champ des vitesses propres d'un type de population de galaxies n'est alors pas représentatif du champ des vitesses à grande échelle dans l'univers. Kashlinsky (1989) explique par exemple que l'excès de corrélation aux

Catalogue NEC : Composantes du tenseur et prédictions théoriques

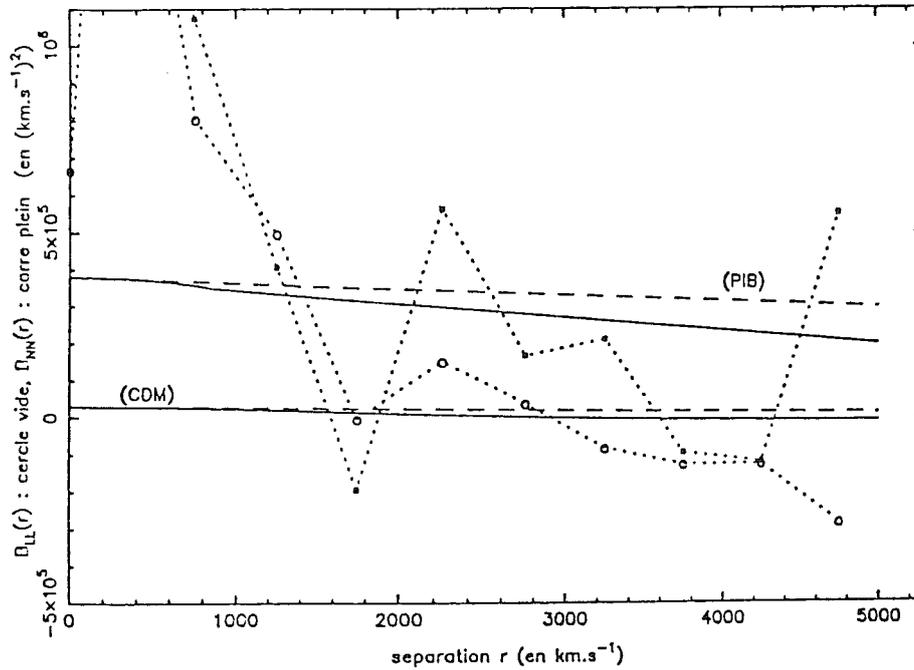


Figure C.5 : $B_{LL}(r)$ et $B_{NN}(r)$ pour le catalogue NEC et prédictions pour les modèles CDM et PIB ($B_{LL}(r)$: continue, $B_{NN}(r)$: hachurée)

Catalogue GAAR : Composantes du tenseur et prédictions théoriques

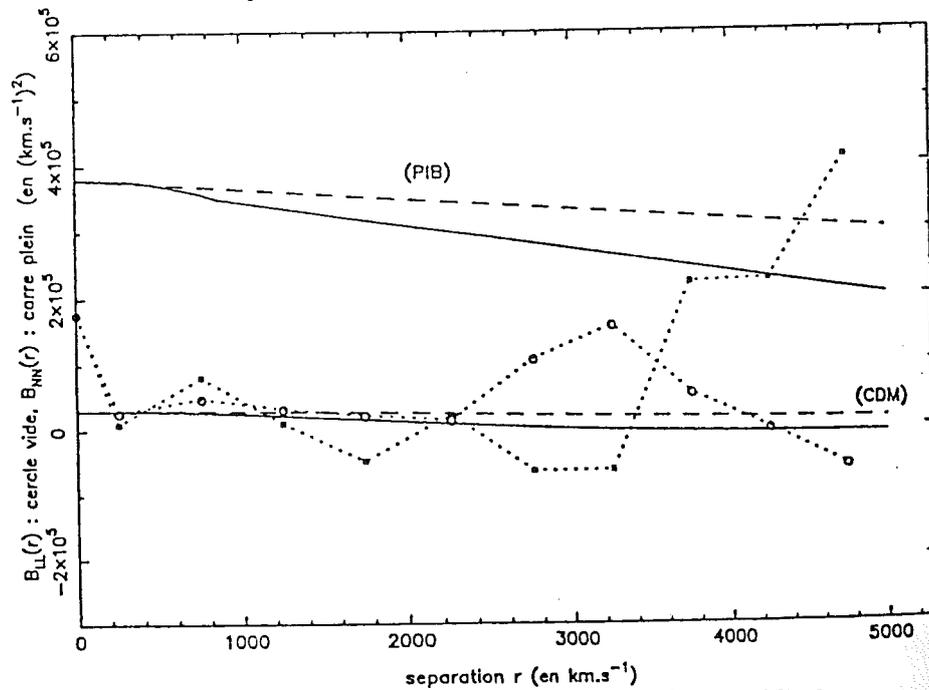


Figure C.6 : $B_{LL}(r)$ et $B_{NN}(r)$ pour le catalogue GAAR et prédictions pour les modèles CDM et PIB ($B_{LL}(r)$: continue, $B_{NN}(r)$: hachurée)

petites échelles pour la population des elliptiques est dû au fait que ces galaxies peuplent principalement les amas de galaxies.

IV.2.3. Les barres d'erreurs pour la régression :

Nous montrons figures (C.7) et (C.8) les barres d'erreurs de régression pour l'évaluation de la composante longitudinale $B_{LL}(r)$ avec les catalogues NEC et GAAR.

L'erreur dans la régression $\sigma_{LL}(r)$ pour l'échantillon NEC est 10 fois plus importante que l'erreur de référence calculée à partir du catalogue NECHI (section IV.2.1). On peut alors déduire que les données $\langle v_r \cdot v_r' \rangle_i$, $\langle \cos \theta \cdot \cos \theta' \rangle_i$ et $\langle \sin \theta \cdot \sin \theta' \rangle_i$ pour chaque classe de séparation r (voir section III.2) ne sont pas corrélées suivant l'équation (C.27). Ainsi l'importance des barres erreurs de régression pour l'estimation de $B_{LL}(r)$ nous apprend que le champ des vitesses propres des galaxies elliptiques ne semble pas statistiquement homogène et isotrope (dans le cas contraire, il vérifierait l'équation C.27)).

La situation est en revanche moins claire pour le catalogue de spirales GAAR. Les barres d'erreurs ne sont en effet que 3 fois plus grandes que les barres d'erreurs de référence (évaluées à partir du catalogue GAARHI). On peut donner deux interprétations différentes de ces résultats.

Interprétation 1 :

Soit l'échantillon GAAR est effectivement homogène et isotrope statistiquement. L'évaluation des composantes $B_{LL}(r)$ et $B_{NN}(r)$ est fiable. La figure (C.8) nous montre alors que le modèle PIB est mis en défaut. Le modèle CDM est en accord avec les données observées. Néanmoins les larges barres d'erreurs ne permettent pas de discriminer entre les différents modèles CDM (par exemple d'estimer la valeur du paramètre de biais b).

Mais si le champ des vitesses propres est homogène et isotrope à l'échelle du catalogue GAAR (3000 km.s^{-1}), il doit a fortiori vérifier cette propriété à l'échelle du catalogue NEC. La figure (C.7) nous montre que ce n'est pas le cas (les barres d'erreurs sont trop importantes). Il faut donc admettre que, pour une raison ou pour une autre, le catalogue NEC n'est pas représentatif du champ des vitesses propres réel. Cet échantillon peut par exemple souffrir de biais systématiques dans la détermination de la distance la plus probable des galaxies qui se répercutent sur l'estimation des vitesses radiales. Le champ des vitesses propres est alors perturbé par un champ fictif qui détruit l'homogénéité et

Catalogue NEC : Barres d'erreurs

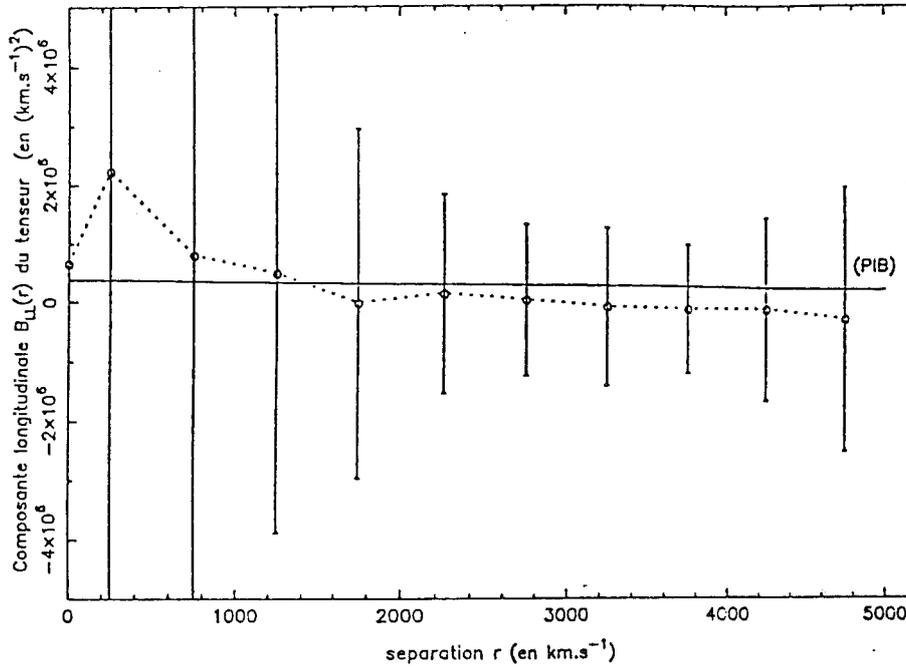


Figure C.7 : Les barres d'erreurs (régression) de $B_{LL}(r)$ pour le catalogue NEC

Catalogue GAAR : Barres d'erreurs

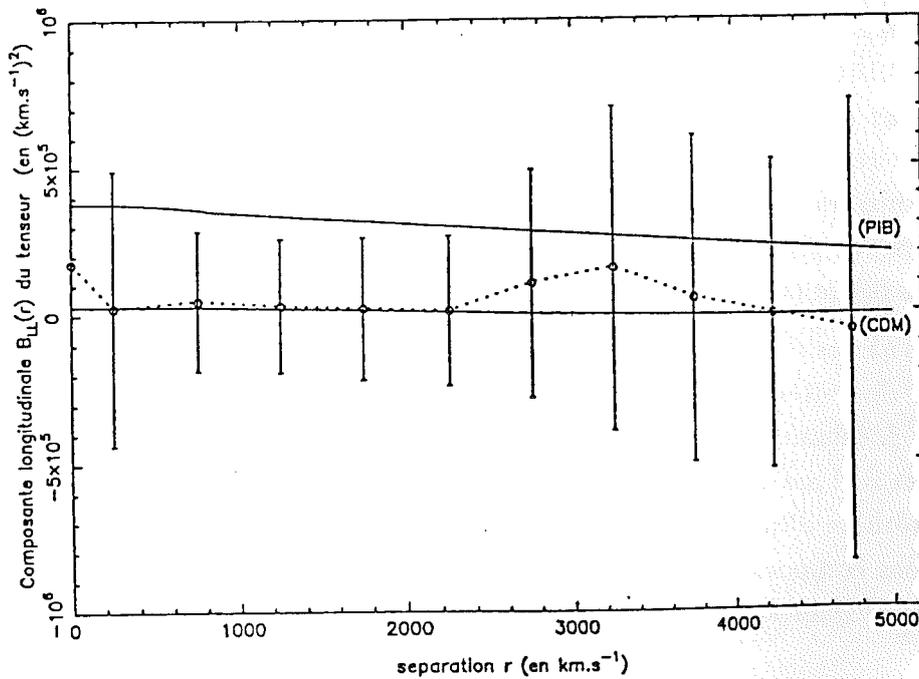


Figure C.8 : Les barres d'erreurs (régression) de $B_{LL}(r)$ pour le catalogue GAAR

l'isotropie originelle du flot. On s'attend donc à des barres d'erreurs importantes lors de l'estimation des composantes du tenseur de corrélation des vitesses à partir du catalogue NEC.

Interprétation 2 :

Mais une autre interprétation est possible. On pense en fait que le champ des vitesses propres à l'échelle du catalogue NEC est dominé par un flot convergent vers une certaine région de l'espace (le "Grand Attracteur", voir partie D). Ce champ des vitesses ne vérifie pas l'hypothèse d'homogénéité et d'isotropie statistique et les barres d'erreurs impliquées dans l'estimation de $B_{LL}(r)$ à partir du catalogue NEC sont donc importantes.

Les effets d'un tel flot convergent cohérent à grande échelle vont être moins significatifs sur un échantillon de profondeur plus petite comme le catalogue GAAR. Nous avons illustré ces effets figures (C.9) et (C.10). Les catalogues NECMGA et GAARMGA sont définis dans la section IV.1. Lorsqu'on soustrait le flot de vitesses du "Grand Attracteur" du champ de vitesses observées, on restaure l'homogénéité et l'isotropie statistique de ce champ. Le "grand attracteur" est en effet un "monstre" dont la présence interdit le traitement statistique du champ des vitesses. On voit figure (C.9) que la taille des barres d'erreurs diminue considérablement (d'un facteur 5 environ) pour le catalogue NECMGA (les barres d'erreurs sont alors 2 fois plus grandes que les erreurs de référence mesurées à partir de NECHI).

La figure (C.10) confirme l'idée qu'un échantillon moins profond comme le catalogue GAAR est moins sensible aux effets du "Grand Attracteur". Les barres d'erreurs pour NECMGA sont réduites d'un facteur 2 seulement (à comparer avec un facteur 5 pour NECMGA). Ainsi on s'attend à des barres d'erreurs plus modestes lors de l'évaluation des corrélations des vitesses à partir de l'échantillon de galaxies spirales GAAR, ce que nous observons figure (C.8).

L'étude des barres d'erreurs introduites lors de l'estimation des composantes du tenseur de corrélation des vitesses nous incite en fait à préférer la deuxième interprétation à la première : L'échantillon de galaxies elliptiques NEC est significatif du champ des vitesses propres réel mais est dominé par un flot de vitesses convergent cohérent à grande échelle. L'hypothèse d'homogénéité et d'isotropie statistique du champ n'est pas vérifiée. L'évaluation des composantes du tenseur de corrélation des vitesses est donc impossible à partir des catalogues actuellement disponibles.

Catologue NEC - GA simule : Barres d'erreurs

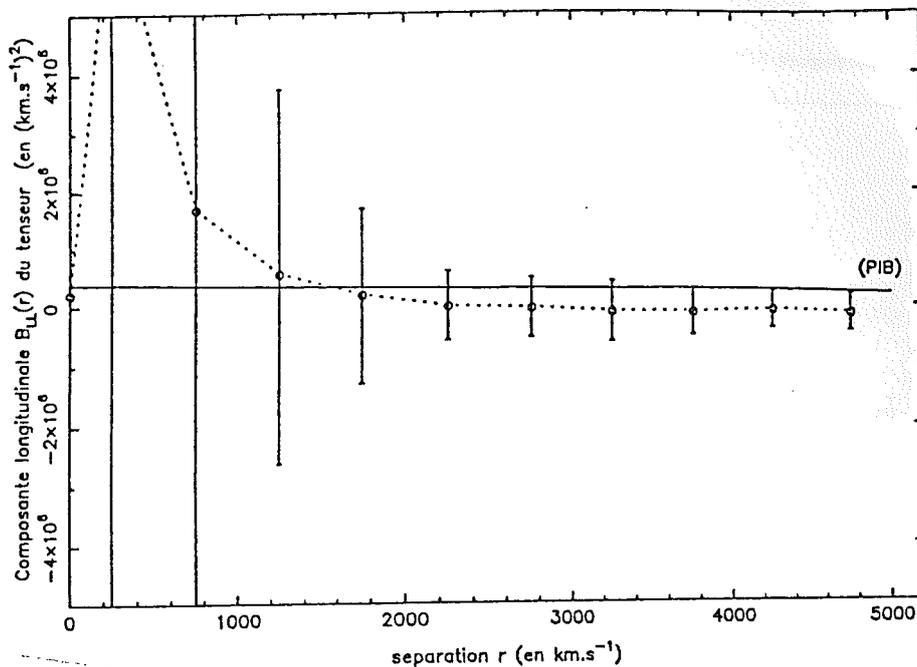


Figure C.9 : Les barres d'erreurs (régression) de $B_{LL}(r)$ pour le catalogue NEC après soustraction d'un modèle de "Grand attracteur"

Catologue GAAR - GA simule : Barres d'erreurs

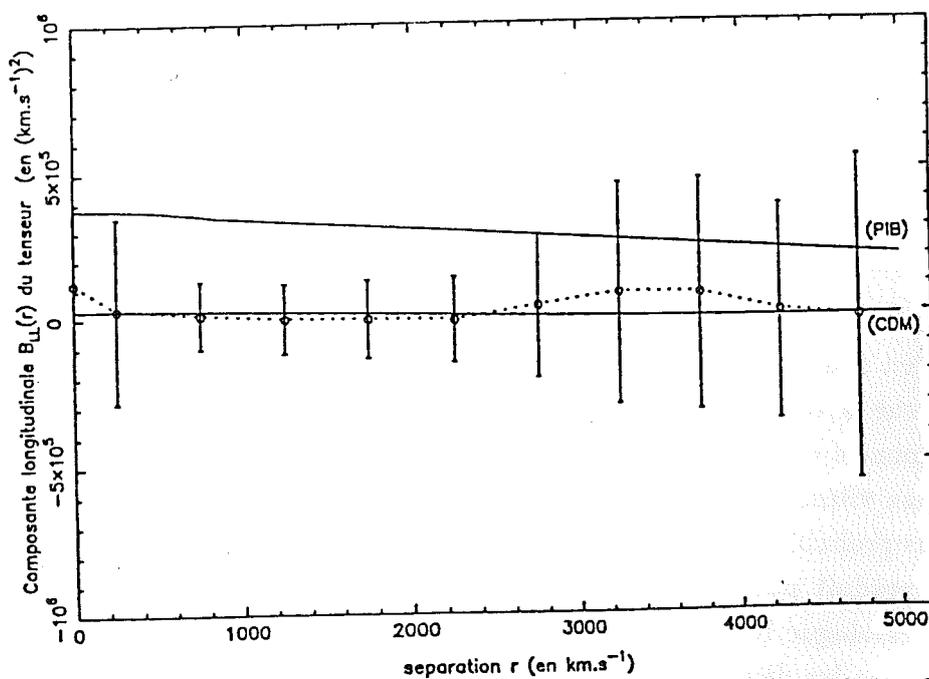


Figure C.10 : Les barres d'erreurs (régression) de $B_{LL}(r)$ pour le catalogue GAAR après soustraction d'un modèle de "Grand attracteur"

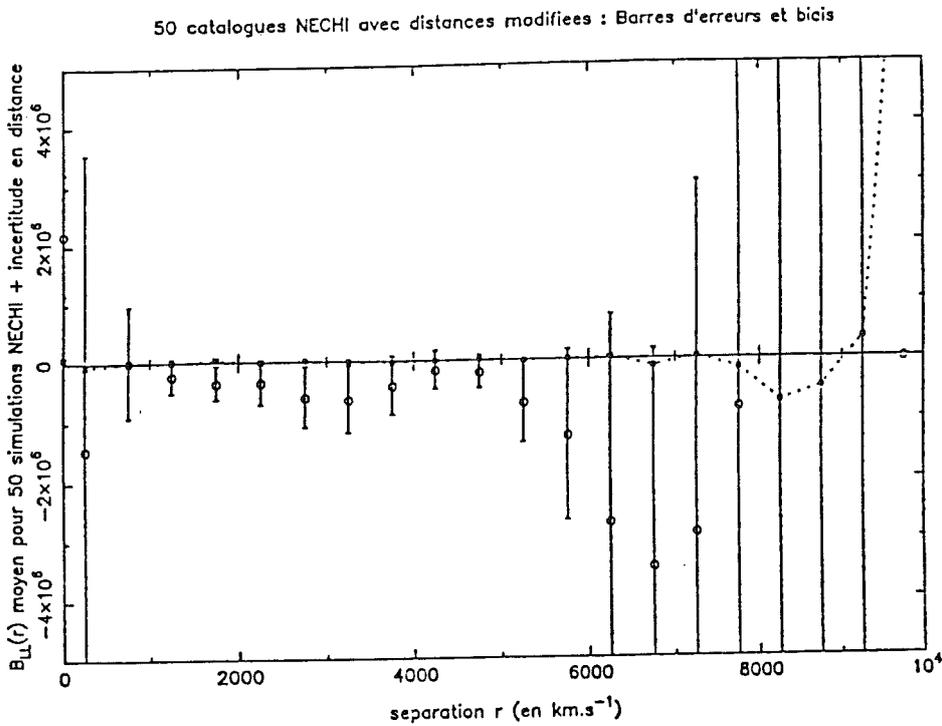


Figure C.11 : Biais et barres d'erreurs de $B_{LL}(r)$ (dûs à l'incertitude sur les distances) pour le catalogue NECHI

IV.3. Les incertitudes sur les distances :

Nous avons vu dans la section précédente que l'évaluation des fonctions de corrélation des vitesses est soumise à des erreurs de régression. A ces erreurs viennent s'ajouter les incertitudes sur la valeur du champ des vitesses radiales.

On a démontré en effet dans la partie B que l'estimation de la distance d'une galaxie (et donc de sa vitesse radiale propre) possède une distribution lognormale autour de sa valeur la plus probable et de dispersion $\Delta \approx 0.2$. Nous avons simulé, dans le but de quantifier l'effet produit par cette incertitude sur les distances des galaxies, 50 échantillons dérivant du catalogue NECHI (voir section IV.1.) et perturbés par une telle distribution lognormale pour les distances qui induit alors une nouvelle vitesse propre pour chaque

50 catalogues NEC avec distances modifiées : Barres d'erreurs et biais

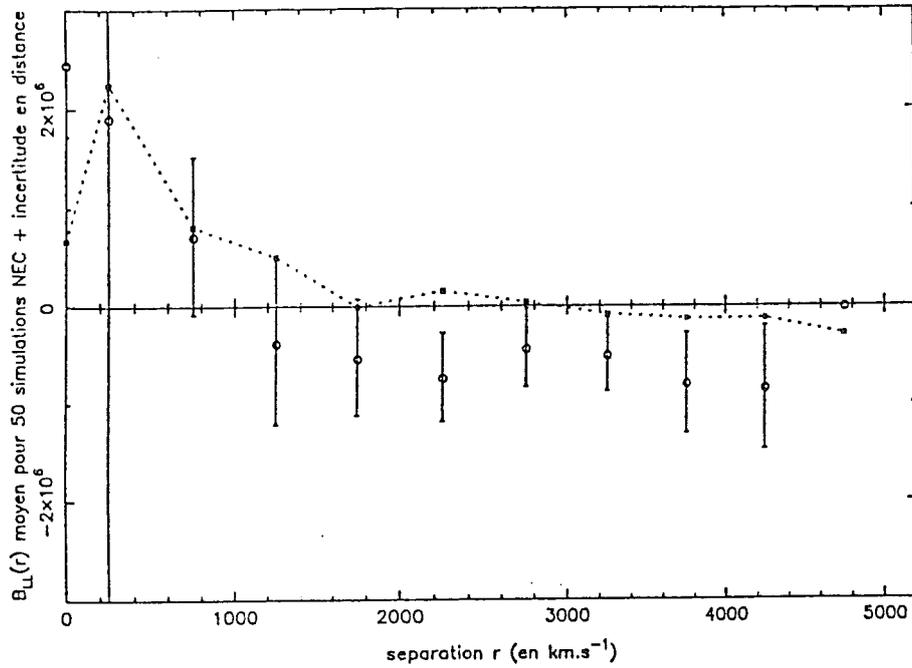


Figure C.12 : Biais et barres d'erreurs de $B_{LL}(r)$ (dûs à l'incertitude sur les distances) pour le catalogue NEC

50 catalogues NEC avec distances modifiées : Barres d'erreurs et biais

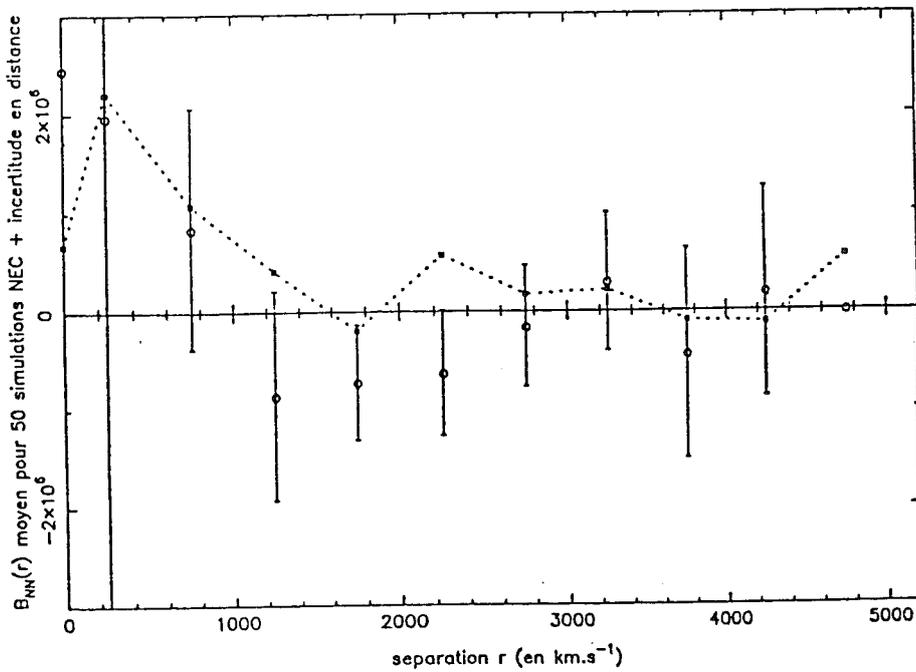


Figure C.13 : Biais et barres d'erreurs de $B_{NN}(r)$ (dûs à l'incertitude sur les distances) pour le catalogue NEC

galaxie. Pour chacun de ces échantillons, nous estimons la valeur de la composante longitudinale $B_{LL}(r)$ du tenseur de corrélation des vitesses. Nous calculons ensuite la moyenne $\overline{B_{LL}(r)}$ de $B_{LL}(r)$ sur ces simulations ($\overline{B_{LL}(r)} = 1/N \sum B_{LL}(r)$) ainsi que sa variance $\sigma_{LL}^D(r)$ ($\sigma_{LL}^D(r) = 1/N \sum (B_{LL}(r) - \overline{B_{LL}(r)})^2$). Nous montrons figure (C.11) cette valeur moyenne $\overline{B_{LL}(r)}$ et ces barres d'erreurs pour le catalogue NECHI. les fonctions $\overline{B_{LL}(r)}$ et $\overline{B_{NN}(r)}$ pour le catalogue NEC sont exposées figures (C.12) et (C.13). Nous désirons faire deux remarques.

Remarque 1 :

L'estimation de $\overline{B_{LL}(r)}$ pour le catalogue NECHI s'avère biaisée, c'est à dire que ses valeurs ne sont plus réparties uniformément autour de 0. L'origine de ce biais provient de l'inhomogénéité du champ avec lequel on perturbe les vitesses radiales de l'échantillon. En fait, le champ des vitesses perturbateur (engendré par la distribution lognormale des erreurs en distance) ne vérifie pas l'hypothèse d'homogénéité et d'isotropie statistique. La technique de régression n'est alors plus applicable à cet échantillon perturbé. Ainsi le caractère particulier des incertitudes sur les distances introduit un biais systématique dans l'évaluation des composantes longitudinale et transversale du tenseur de corrélation des vitesses. Le correction de ce biais semble délicate, compte tenu de l'aspect non-linéaire du calcul tensoriel. L'estimation non biaisée des fonctions de corrélation des vitesses apparaît donc fort compromise.

Remarque 2 :

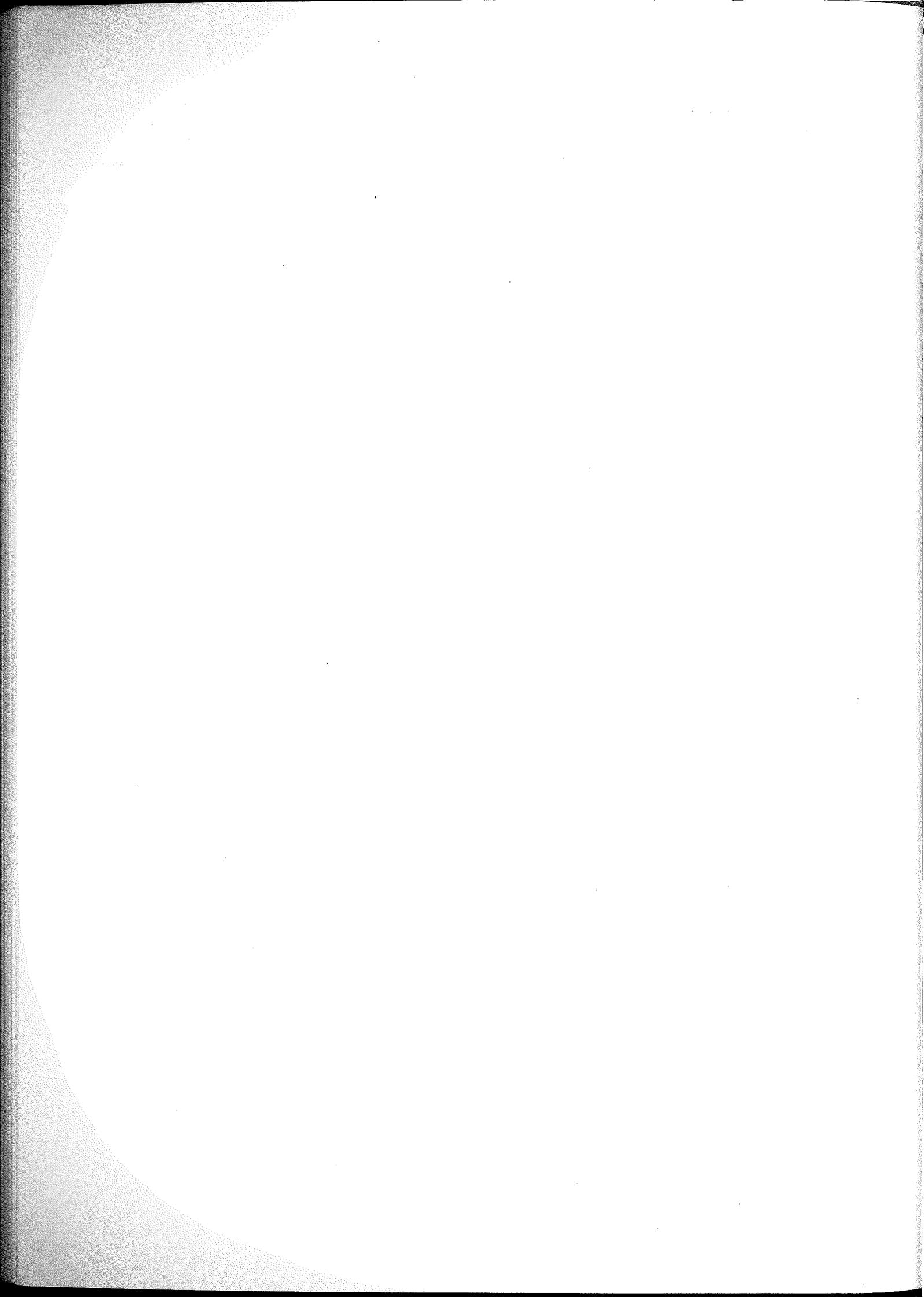
L'amplitude $\sigma_{LL}^D(r)$ des barres d'erreurs dues à l'incertitude sur les distances est environ 2 fois plus importante que les erreurs de régression. Nous avons démontré dans la section précédente que le volume occupé par l'échantillon NEC (d'une profondeur d'environ 6000 km.s^{-1}) n'était pas assez grand pour permettre l'homogénéité statistique du champ des vitesses propres et donc l'évaluation des composantes du tenseur de corrélation. Or la valeur de $\sigma_{LL}^D(r)$ augmente de manière évidente avec la profondeur de l'échantillon étudié (l'erreur sur la distance d'une galaxie est en effet proportionnelle à la distance (voir partie B, section I.5)). Il est alors à craindre que l'ampleur des valeurs $\sigma_{LL}^D(r)$, pour l'échantillon de profondeur minimum satisfaisant la contrainte d'homogénéité du champ des vitesses, interdise une quelconque estimation de $B_{LL}(r)$ et $B_{NN}(r)$. Nous avons remarqué précédemment que l'évaluation des composantes du tenseur de corrélation des vitesses était biaisée, elle est de plus fortement incertaine.

REFERENCES :

- [1] : M. Aaronson, J. Huchra, J. Mould, P. Schechter and B. Tully : "The velocity field in the local supercluster", *Astrophys.J* **258**,64-76 (1982a)
- [2] : M. Aaronson, J. Huchra, J. Mould, B. Tully, J.R. Fisher, H. Van Woerden, W.M. Gosse, P. Chamaraux, V. Mebold, B. Siegman, G. Berriman and S.E. Persson : "A catalog of infrared magnitudes and HI velocity widths for nearby galaxies", *Astrophys.J Suppl. Ser.* **50**,241-262 (1982b)
- [3] : J.M. Bardeen, J.R. Bond and G. Efstathiou : "Cosmic fluctuation spectra with large-scale power", *Astrophys.J* **321**,28-35 (1987)
- [4] : M. Clutton-Brock and P.J.E. Peebles : "Galaxy clustering and the Rubin-Ford effect", *Astron.J* **86**(8), 1115-1119 (1981)
- [5] : M. Davis and P.J.E. Peebles : "A survey of galaxy redshifts: V. The two-point position and velocity correlation", *Astrophys.J* **267**,465-482 (1983)
- [6] : M. Davis, G. Efstathiou, C.S. Frenk and S.D.M. White : "The evolution of large-scale structure in a universe dominated by Cold Dark Matter", *Astrophys.J* **292**,371-394 (1985)
- [7] : K.M. Górski : "On the pattern of perturbations of the Hubble flow", *Ap.J Letters* **332**,L7-L11 (1988)
- [8] : K.M. Górski, M. Davis, M.A. Strauss, S.D.M. White and A. Yahil : "Cosmological velocity correlations : Observations and model predictions", *Astrophys.J* **344**,1-19 (1989)

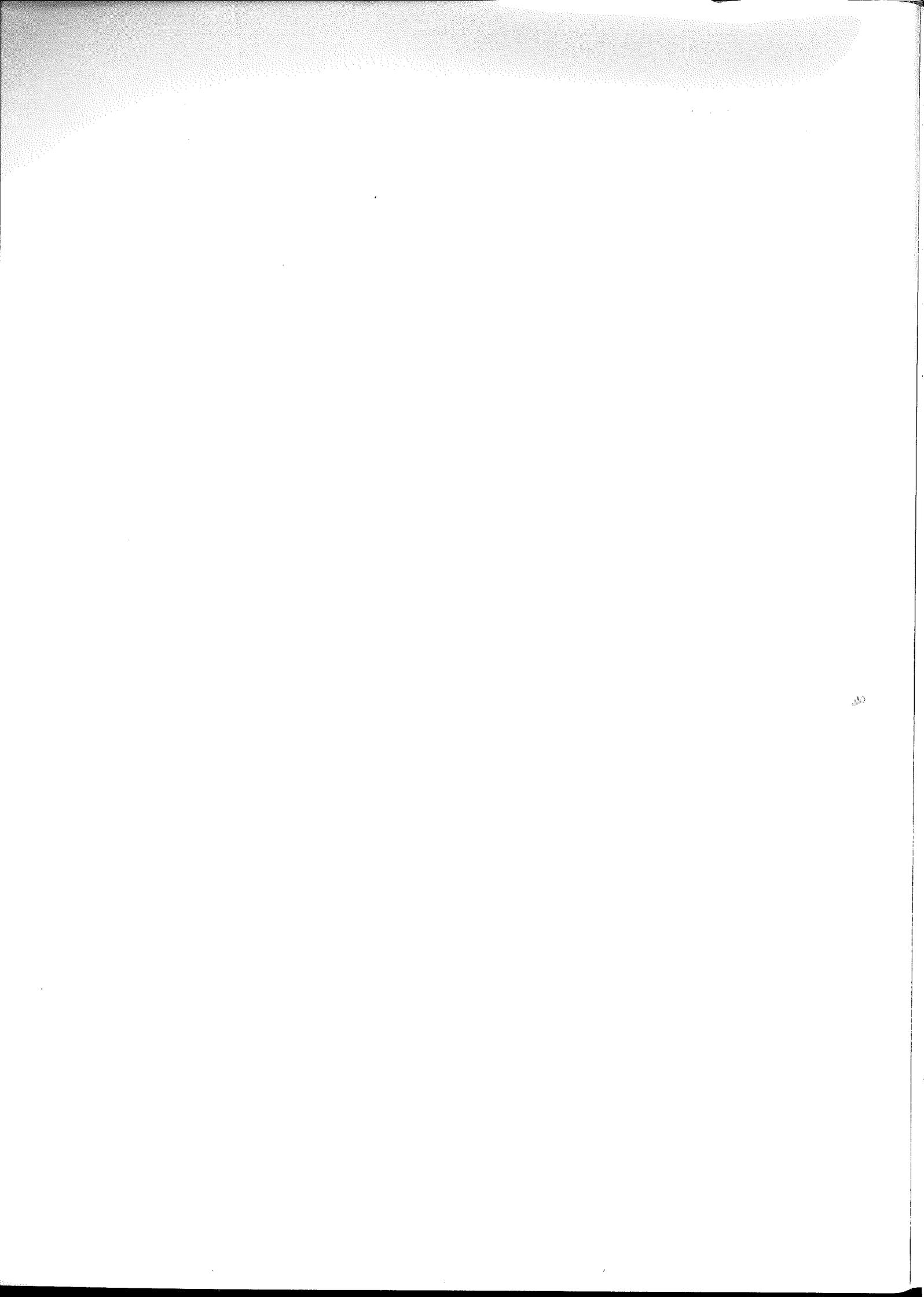
- [9] : E.J. Groth, R. Juszkiewicz and J.P. Ostriker : "An estimate of the velocity correlation tensor : Cosmological implications", *Astrophys.J* **346**,558-565 (1989)
- [10] : A. Hald : "Statistical theory with engineering applications", ed. John Wiley&Sons, Inc (1952), p. 627-638
- [11] : R.N. Henriksen and M. Lachièze-Rey : "The mass function of galaxies based on correlated velocity structure", *Mon.Not.R. astr.Soc.* (1990) **245**, 255-267
- [12] : R. Juszkiewicz and A.Yahil : "Galaxy clustering and the origin of large-scale flows", *Ap.J Letters* **346**,L49-L51 (1989)
- [13] : R. Juszkiewicz N. Vittorio and R.F.G. Wyse : "Local gravity and large-scale structure", *Astrophys.J* **349**,408-414 (1990)
- [14] : N. Kaiser : "Clustering in real space and in redshift space", *Mon.Not.R.astr.Soc.* (1987) **227**, 1-21
- [15] : N. Kaiser : "Theoretical implications of deviations from Hubble flow", *Mon.Not. R.astr.Soc.* (1988) **231**, 149-167
- [16] : N. Kaiser and O. Lahav : "Theoretical implications of cosmological dipoles", *Mon.Not.R. astr.Soc.* (1989) **237**, 129-162
- [17] : N. Kaiser : "Local large scale structure vs Cold Dark Matter" in "*Large-scale structure and motions in the Universe*", ICTP, p. 197-213, Eds. M.Mezzetti, G. Giuricin, F. Mardirassian and M.Ramella, Trieste (1989)
- [18] : A. Kashlinsky : "On the coherence length of large-scale peculiar velocities and gravitational clustering", *Ap.J Letters* **343**,L5-L7 (1989)
- [19] : O. Lahav, N. Kaiser and Y. Hoffman : "Local gravity and peculiar velocity: probes of cosmological models", *Astrophys.J* **352**,448-456 (1990)

- [20] : S.D. Landy and A.S. Szalay : "Large-scale radial velocity correlations as a test of gaussian initial conditions", *Astrophys.J* **394**,25-37 (1992)
- [21] : A.S. Monin and A.M. Yaglom : "Statistical Fluid Mechanics", Vol. 1, Chap. 2, Cambridge: MIT Press (1975)
- [22] : A.S. Monin and A.M. Yaglom : "Statistical Fluid Mechanics", Vol. 2, Chap. 7, Cambridge: MIT Press (1975)
- [23] : P.J.E. Peebles : "The large scale structure of the universe", (1980), p. 272-285, Princeton university series, N.J.
- [24] : P.J.E. Peebles : "Origine of the large-scale galaxy peculiar velocity field : a minimal isocurvature model", *Nature* **327**,210-211 (1987)
- [25] : P.J.E. Peebles, R.A. Daly and R. Juszkiewicz : "Masses of rich clusters of galaxies as a test of the biased cold dark matter theorie", *Astrophys.J* **347**,563-574 (1989)
- [26] : E. Regös and A.S. Szalay : "Multipole expansion of the large-scale velocity field: using the tensor window function", *Astrophys.J* **345**,627-636 (1989)
- [27] : A.S. Szalay : "Velocity correlations and large-scale motions" in "*Large-scale structure and motions in the Universe*", ICTP, p. 215-225, Eds. M.Mezzeti, G. Giuricin, F. Mardirassian and M.Ramella, Trieste (1989)
- [28] : N. Vittorio, R. Juszkiewicz and M. Davis : "Large-scale velocity fields as a test of cosmological models", *Nature* **323**,132-133 (1986)
- [29] : N. Vittorio and R. Juszkiewicz : "Peculiar velocity and gravity as cosmological probes" in "*Large-scale structure and motions in the Universe*", ICTP, p. 241-258, Eds. M.Mezzeti, G. Giuricin, F. Mardirassian and M.Ramella, Trieste (1989)



PARTIE D :

LE "GRAND ATTRACTEUR"



SOMMAIRE :

I. Mise en évidence du "Grand Attracteur" :

I.1. Le champ des vitesses propres local :

I.2. Le "Grand Attracteur" :

I.2.1. Analyse par la technique du maximum de vraisemblance :

I.2.2. Les faiblesses de la méthode :

I.2.3. Le flot des vitesses derrière le "Grand Attracteur" :

I.3. Les implications :

I.3.1. Identification du "Grand Attracteur" :

I.3.2. La cohérence du champ des vitesses :

I.3.3. Les anisotropies dans le fond diffus cosmologique :

II. Les tests des sphères :

II.1. Définitions :

II.1.1. Le test dipolaire des sphères :

II.1.2. Le test quadripolaire des sphères :

II.2. Application :

II.2.1. La significativité de notre analyse :

II.2.2. Les résultats :

II.2.3. Les conclusions :

III. Le test de convergence :

III.1. Définition :

III.2. Application :

II.1.1. Le flot Virgocentrique :

II.1.2. Vers une analyse par ondelettes du champ des vitesses :

IV. Annexe :

Notre connaissance de la cinématique locale des perturbations de la densité de matière totale présente dans l'univers s'améliore au fur et à mesure que des catalogues en redshift-distance de plus en plus profonds sont disponibles. La mise en évidence du mouvement de notre Groupe Local dans le repère du fond diffus cosmologique, puis de la chute des galaxies du super-amas local sur l'amas de Virgo ont dévoilé la complexité du champ des vitesses propres local. Dernièrement, une analyse du catalogue de Faber et *al.* réalisée par Lynden-Bell et *al.* (1988a) montre que ce champ est en fait dominé par la présence d'un flot de vitesses convergent, cohérent à une échelle d'au moins $60 h^{-1}$ Mpc : le "Grand Attracteur".

La présence du "Grand Attracteur" perturbe alors toutes études statistiques de la cinématique régissant notre voisinage (par exemple l'évaluation du tenseur de corrélation des vitesses), les échantillons actuellement disponibles n'étant pas assez grands pour être représentatifs de l'univers dans sa globalité. De plus l'existence d'un tel flot pose de sérieuses difficultés aux théories classiques de formation de structures : la contrepartie optique du "Grand Attracteur" n'est pas détectée, l'échelle de cohérence du flot est considérable et son potentiel gravitationnel associé doit en principe créer des fluctuations non encore détectées dans le fond diffus cosmologique. Ces difficultés ont poussées certains auteurs à douter de la fiabilité du catalogue de Faber et *al.* (1989) ainsi que de l'analyse réalisée par Lynden-Bell et *al.* . Cette étude ayant permis la mise en évidence du "Grand Attracteur" souffre en effet de quelques faiblesses (sensibilité au repère des vitesses de référence et fragilité statistique).

Nous proposons dans les sections II. et III. de cette partie des tests géométriques basés sur les propriétés d'invariance du champ des vitesses qui permettent la détection d'un flot engendré par un attracteur, ceci indépendamment du repère des vitesses de référence dans lequel est effectuée l'analyse. L'application de ces tests au catalogue de galaxies elliptiques de Faber et *al.* (1989) confirme la présence d'un attracteur d'extension spatiale importante. De plus, nos tests révèlent l'existence d'un mouvement d'ensemble du "Grand Attracteur" ainsi que de toutes les galaxies de l'échantillon par rapport au repère dans lequel le rayonnement du fond diffus cosmologique est au repos.

Ces résultats, en accords avec ceux de certains auteurs (Willick (1990), Mathewson et *al.* (1992b)) portent alors la cohérence de notre champ de vitesses local à une échelle d'au moins $100 h^{-1}$ Mpc.

I. MISE EN EVIDENCE DU "GRAND ATTRACTEUR" :

Avec la création de catalogues de vitesses propres de galaxies de plus en plus profonds, la conception de la cinématique régissant notre voisinage a été considérablement modifiée. C'est avec scepticisme que la communauté scientifique accueillit en 1988 la mise en évidence par Lynden-Bell et *al.* du "Grand Attracteur", un champ de vitesses convergent vers un point distant de $42 h^{-1}$ Mpc dans la direction ($l = 309$, $b = 18$ en coordonnées galactiques) et cohérent à une échelle d'au moins $60 h^{-1}$ Mpc.

Si l'étude de Lynden-Bell et *al.* (1988a) n'est pas sans faiblesses (sensibilité au repère de référence et fragilité statistique), la réalité d'un flot cohérent à grande échelle ne fait aujourd'hui guère de doute.

Mais l'existence du "Grand Attracteur" pose de sérieuses difficultés aux théories classiques de formation de structures. D'une part, on a du mal à identifier parmi les concentrations de matière lumineuse actuellement observées la source massive d'un tel champ de vitesses convergent. D'autre part, le potentiel gravitationnel du "Grand Attracteur" doit théoriquement créer une anisotropie dans le rayonnement du fond diffus cosmologique. Cet anisotropie n'a pas encore été détectée, compte-tenu de la limite expérimentale de détection des appareils de mesure. Néanmoins l'amplitude des fluctuations de température dans le rayonnement cosmologique engendrées par une structure du type "Grand Attracteur" s'approche dangereusement de cette limite supérieure de détection dictée par l'observation. Les prochaines années devraient être déterminantes pour le devenir des modèles décrivant le processus de formation des grandes structures dans l'univers.

I.1. Le champ des vitesses propres local :

On a longtemps cru que le champ des vitesses des galaxies, à l'image de la cinématique régissant notre Groupe Local, était bien décrit par un flot de Hubble perturbé par un champ de vitesses propres aléatoire de faible dispersion. Cette vision simpliste fut pourtant mise en défaut lorsque Rubin et *al.* (1976a) établirent un catalogue des vitesses propres de

96 galaxies spirales dont les distances sont comprises entre 3500 et 6500 km.s⁻¹. Rubin et al. (1976b) montrèrent en effet que notre Groupe Local possède, dans le repère où cet échantillon est au repos, une vitesse d'environ 500 km.s⁻¹ dans la direction ($l = 163, b = -11$). Un an plus tard, on découvrait une anisotropie en température dans le fond diffus cosmologique (le dipole, voir par exemple Smoot et al. (1979)), interprétée comme l'empreinte de notre mouvement dans le repère pour lequel le fond diffus est au repos. La création d'un catalogue de vitesses propres plus précis fut alors décidée afin de découvrir l'origine de ce mouvement (600 km.s⁻¹ dans la direction ($l = 270, b = 30$)).

L'échantillon de galaxies spirales de Aaronson et al. (1982a) permit en fait une étude approfondie de la cinématique du super-amas local, dominée par un mouvement d'ensemble de chute sur l'amas de Virgo (à une distance d'environ 1000 km.s⁻¹ dans la direction ($l = 280, b = 70$)). Aaronson et al. (1982b) déduirent que la contribution de ce flot virgocentrique à la vitesse de notre Groupe Local (250 km.s⁻¹ dans la direction ($l = 280, b = 70$)), n'expliquait que partiellement notre mouvement dans le repère du fond diffus cosmologique. Il doit alors exister des champs de vitesses cohérents à plus grande échelle, responsables de ce mouvement. La détection d'un flot quadripolaire (signature d'un centre de convergence pour le champ des vitesses) dont l'axe d'expansion pointe vers la direction ($l = 308, b = 13$) par Lilje et al. (1986), le suggère d'ailleurs fortement.

Le catalogue de galaxies elliptiques de Faber et al., environ 3 fois plus profond que celui de Aaronson et al., confirma l'existence d'un champ de vitesses cohérent à grande échelle. Dressler et al. (1987) montrèrent que l'échantillon d'elliptiques est soumis à un mouvement d'ensemble d'amplitude 600 km.s⁻¹ vers la direction ($l = 312, b = 6$). Une étude plus raffinée (Lynden-Bell et al. (1988a)) permit de détecter le caractère quadripolaire du champ des vitesses et d'introduire le concept de "Grand Attracteur", un champ de vitesses cohérent à une échelle d'au moins 60 h⁻¹ Mpc, convergent vers un centre situé à une distance de 4200 km.s⁻¹ de notre Groupe Local dans la direction ($l = 309, b = 18$). L'influence du "Grand Attracteur" sur la vitesse du Groupe Local (500 km.s⁻¹ dans la direction ($l = 309, b = 18$)) conjuguée avec celle du flot Virgocentrique, ne suffit pas néanmoins à expliquer notre mouvement dans le repère du fond diffus cosmologique. Il est alors nécessaire d'ajouter des contributions supplémentaires à ces flots, soit à petites échelles comme l'anomalie locale (Burstein et al. (1989)), soit à plus grande échelle (Willick (1990)).

I.2. Le "Grand Attracteur" :

La présence d'un flot de vitesses convergent à grande échelle vers une région de l'espace centrée en ($l = 309 \pm 10$, $b = 18 \pm 10$ et $r = 4200 \pm 300 \text{ km.s}^{-1}$) (le "Grand Attracteur") a été mise en évidence pour la première fois par Lynden-Bell et *al.* (1988a).

La méthode utilisée lors de cette étude (technique du maximum de vraisemblance) possède néanmoins quelques faiblesses. D'une part le nombre de paramètres libres invoqués dans la maximisation de la fonction de vraisemblance est important, d'autre part cette analyse est sensible au choix du repère des vitesses de référence ainsi qu'au choix du modèle de champ des vitesses engendré par l'attracteur.

La mesure des vitesses propres des galaxies demeurant derrière la région du "Grand Attracteur" devrait théoriquement permettre de confirmer la réalité de ce flot convergent à grande échelle. Néanmoins, la détermination de ces vitesses propres est soumise à des biais d'autant plus importants que les galaxies sont éloignées (Dressler et *al.* (1990) ou Bothun et *al.* (1992)) qui rendent toutes conclusions difficiles. Ainsi contrairement à Mathewson et *al.* (1992), Burstein et *al.* (1990) détectent un mouvement de chute des galaxies derrière le "Grand Attracteur".

I.2.1. Analyse par la technique du maximum de vraisemblance :

L'étude de Lynden-Bell et *al.* (1988a) repose sur la technique du maximum de vraisemblance. Il s'agit de postuler une forme particulière que le champ des vitesses $v(\mathbf{r})$ vérifiera, cette modélisation du champ des vitesses dépendant d'un certain nombre de paramètres libres. Il est alors possible d'écrire la densité de probabilité dP_G pour chaque galaxie de l'échantillon en fonction de ses variables observées. Ensuite, la maximisation de la fonction de vraisemblance permet d'évaluer les valeurs les plus probables des paramètres introduits lors de la modélisation du champ des vitesses.

Les variables observées dans l'analyse de Lynden-Bell et *al.* sont la vitesse radiale $v_r = c.z$ (exprimée dans le repère du fond diffus cosmologique) et la distance estimée R_e (voir partie B (section IV.4, équation (B.58)). Le champ des vitesses modélisé $v(\mathbf{r})$ est décomposé en 3 termes : $\mathbf{v} = \mathbf{v}_I + \mathbf{v}_V + \mathbf{u}$, avec \mathbf{v}_I un champ aléatoire de distribution gaussienne et de dispersion σ_I , \mathbf{v}_V un modèle de flot Virgocentrique et \mathbf{u} un champ des

vitesse qui prendra par la suite plusieurs formes. En notant $v'_r = v_r - \mathbf{v}_V \cdot \hat{\mathbf{r}}$ (v'_r est la vitesse radiale observée corrigée du modèle de flot Virgocentrique), l'équation (B.60) (voir partie B, section IV.4) permet d'exprimer, avec une bonne approximation, la densité de probabilité dP_G pour une galaxie de l'échantillon de la façon suivante :

$$dP_G(v'_r, R_e) = (2\pi\sigma_e^2)^{-1/2} \exp\left[-\frac{1}{2}(v'_r - \mathbf{u}(\mathbf{r}) \cdot \hat{\mathbf{r}})^2 / \sigma_e^2\right] dv'_r dR_e \quad (D.1)$$

$$\text{avec } \sigma_e^2 = \sigma_f^2 + R^2 (\exp \Delta^2 - 1) , \quad \Delta = 0.21$$

$$\text{et } R = R_e (\exp[(\alpha + 1/2) \Delta^2]) , \quad \alpha = 3$$

La fonction de vraisemblance L de l'échantillon de galaxies s'exprime alors en fonction des variables mesurées v'_r (ou v_r) et R_e ($L = \prod dP_G(v'_r, R_e) / dv'_r dR_e$) et dépend des paramètres libres introduits lors de la modélisation du champ des vitesses $\mathbf{u}(\mathbf{r})$ ainsi que de la dispersion σ_f du champ des vitesses aléatoire \mathbf{v}_I . Les auteurs testent plusieurs formes pour le champ des vitesses $\mathbf{u}(\mathbf{r})$ (un champ constant, un flot quadripolaire et un flot engendré par une concentration de masse) afin de déterminer lequel de ces modèles décrit le mieux la réalité.

. Un champ constant :

Le champ des vitesses $\mathbf{u}(\mathbf{r})$ est modélisé par un champ de Hubble $H \mathbf{r}$ plus un flot de vitesses constant \mathbf{u}_0 : $\mathbf{u}(\mathbf{r}) = \mathbf{u}_0 + H \mathbf{r}$. La maximisation de la fonction de vraisemblance permet d'évaluer les paramètres libres du modèle, c'est-à-dire la dispersion des vitesses $\sigma_f = 365 \text{ km.s}^{-1}$ ainsi que les composantes du champ constant (d'amplitude $521 \pm 89 \text{ km.s}^{-1}$ dans la direction $l = 307 \pm 11$, $b = 9 \pm 8$). Le χ^2 est alors réduit de 30 unités par rapport au χ^2 évalué pour un champ de Hubble seulement. Les données sont donc bien mieux décrites après l'introduction de ce flot constant, ce qui confirme les résultats précédemment établis (Dressler et al. (1987)).

. Un champ quadripolaire :

Les auteurs introduisent une anisotropie quadripolaire dans le champ de Hubble qui rend compte d'un mouvement plus complexe qu'une simple vitesse de groupe. On s'attend notamment à détecter un tel flot quadripolaire si le champ des vitesses est engendré

par une concentration de masse localisée au bord de l'échantillon. La constante H précédemment définie est alors remplacée par la quantité tensorielle \mathbf{H} et le champ des vitesses $\mathbf{u}(\mathbf{r})$ prend la forme suivante : $\mathbf{u}(\mathbf{r}) = \mathbf{u}_0 + \mathbf{H} \cdot \mathbf{r}$.

Une anisotropie est bien détectée dont l'axe d'expansion (d'amplitude 1.25 ± 0.06) pointe vers la direction ($l = 336, b = 18$). Ce résultat confirme les conclusions d'une étude réalisée par Lilje et *al.* (1986) sur l'échantillon de galaxies spirales de Aaronson et *al.*. L'existence de ce flot quadripolaire pratiquement aligné avec le champ constant \mathbf{u}_0 est en fait la signature d'un mouvement plus complexe qu'un déplacement "en bloc" des galaxies par rapport au fond diffus cosmologique. L'échantillon doit être soumis à un champ de vitesses convergent à grande échelle engendré par une forte concentration de masse : le "Grand Attracteur".

. Un champ engendré par une concentration de matière :

Les auteurs testent un modèle de champ de vitesses engendré par une concentration de masse centrée au point de coordonnées ($R_m, l = 307, b = 9$). Le profil des vitesses est fixé à partir du centre de l'attracteur par l'équation suivante : $v(r) = V_m R_m / r$ avec V_m la vitesse induite par l'attracteur pour notre Groupe Local. De plus un flot constant supplémentaire est autorisé.

Les résultats de la maximisation de la fonction de vraisemblance sont très positifs. D'une part le χ^2 est réduit de 56 unités par rapport au χ^2 obtenu en modélisant le champ des vitesses par un flot de Hubble pur (à comparer avec une réduction du χ^2 de 30 pour la modélisation avec flot constant). D'autre part la dispersion des vitesses σ_f diminue significativement ($\sigma_f = 245 \text{ km.s}^{-1}$). De plus l'amplitude du flot constant chute sensiblement ($106 \pm 78 \text{ km.s}^{-1}$ dans la direction $l = 308 \pm 49, b = 13 \pm 31$). La cinématique de l'échantillon d'elliptiques semble donc bien expliquée par la présence d'un "Grand Attracteur".

Les auteurs retiennent finalement pour caractériser le "Grand Attracteur" une solution sans vitesse de groupe (peut-être dans le but de réduire le nombre de paramètres libres du modèle). La concentration de masse est alors centrée à une distance de $4350 \pm 350 \text{ km.s}^{-1}$ dans la direction ($l = 307, b = 9$) et induit une vitesse pour notre Groupe Local $V_m = 570 \pm 60 \text{ km.s}^{-1}$. Ces résultats seront légèrement transformés par la suite (Burstein et *al.* (1989), $l_{GA} = 309 \pm 10, b_{GA} = 18 \pm 10, r_{GA} = 4200 \pm 350 \text{ km.s}^{-1}$ et $V_m = 535 \pm 60 \text{ km.s}^{-1}$) afin de prendre en compte les modifications apportées au catalogue initial (par

exemple Lynden-Bell et *al.* (1988b)).

1.2.2. Les faiblesses de la méthode :

La mise en évidence en 1988 du "Grand Attracteur" a été accueillie sceptiquement par la communauté scientifique et continue d'ailleurs aujourd'hui d'être un sujet de controverse. Les difficultés rencontrées par les modèles classiques de formation de structures afin de rendre compte d'un champ de vitesses cohérent à une échelle d'au moins $60 h^{-1}$ Mpc, ont poussées certains auteurs (par exemple Silk (1989) ou Rowan-Robinson et *al.* (1990)) à remettre en cause la validité du catalogue d'elliptiques de Faber et *al.* ainsi que l'analyse de Lynden-Bell et *al.* (1988).

Nous avons exposé dans la section IV. de la partie B quels étaient les problèmes rencontrés lors de la détermination des vitesses radiales propres des galaxies du catalogue de Faber et *al.* . La calibration de la relation donnant accès aux distances des galaxies est peut-être biaisée (voir section IV.3, partie B) et l'estimation de la distance la plus probable d'une galaxie est effectuée sous l'hypothèse nulle que les galaxies de l'échantillon possèdent une distribution spatiale uniforme (voir section IV.4, partie B). De plus, il est possible que le champ des vitesses des galaxies elliptiques ne soit pas représentatif du champ des vitesses à grande échelle présent dans l'univers, ces galaxies peuplant préférentiellement les amas de galaxies (voir section IV.2.2, partie C). On voit donc que les doutes exprimés par certains auteurs sur la fiabilité du catalogue de Faber et *al.* ne sont pas sans fondements, et une étude plus approfondie des problèmes de biais semble nécessaire. Néanmoins, un catalogue de 1355 galaxies spirales couvrant une bonne partie de l'hémisphère Sud (Mathewson et *al.* (1992a)) confirme l'idée d'un flot cohérent à grande échelle, même si ce catalogue ne permet pas la mise en évidence du "Grand Attracteur".

Si l'on crédite la validité des données de Faber et *al.*, la méthode employée par Lynden-Bell et *al.* (technique du maximum de vraisemblance) présente cependant 3 principaux défauts. D'une part le nombre de paramètres libres impliqués dans la maximisation de la fonction de vraisemblance est considérable (13 paramètres libres pour 200 points données indépendants). D'autre part, la détection du "Grand Attracteur" dépend fortement du repère des vitesses dans lequel est effectuée l'analyse (Groth et *al.* (1989)). Enfin, la position ainsi que la distance du "Grand Attracteur" varient sensiblement suivant les

caractéristiques choisies afin de modéliser le champ des vitesses dû à l'attracteur.

. Le nombre de paramètres libres impliqués :

Le modèle de champ des vitesses finalement retenu par Lynden-Bell et *al.* possède 13 paramètres libres (Burstein et *al.* (1989)) dont les valeurs sont déterminées par la technique du maximum de vraisemblance. Compte tenu du faible nombre de points données indépendants (environ 200), on peut à juste titre s'interroger sur la fiabilité statistique de la méthode. La dispersion des vitesses après soustraction du champ des vitesses modélisant le "Grand Attracteur" reste d'ailleurs importante (640 km.s^{-1} incluant la dispersion due aux erreurs en distance plus la dispersion σ_f du champ des vitesses aléatoire). Les valeurs estimées des paramètres du modèle (par exemple la position du "Grand Attracteur") nécessitent donc une certaine prudence.

. Influence du repère de référence pour les vitesses :

Lynden-Bell et *al.* postulent implicitement dans leur analyse que la vitesse observée de chaque galaxie de l'échantillon est engendrée par le "Grand Attracteur", au repos dans le repère des vitesses défini par le fond diffus cosmologique. Les caractéristiques de l'attracteur sont alors déduites par la technique du maximum de vraisemblance. Il est toutefois fort possible que le "Grand Attracteur" possède un mouvement propre par rapport au repère du fond diffus cosmologique, charriant avec lui toutes les galaxies de l'échantillon. Dans un tel cas de figure, l'analyse de Lynden-Bell et *al.* ne permet plus de dériver les valeurs correctes des caractéristiques de l'attracteur, la technique du maximum de vraisemblance n'étant effectivement applicable que dans le repère des vitesses dans lequel le "Grand Attracteur" est au repos. En d'autres termes, la mise en évidence du "Grand Attracteur" ainsi que l'estimation de ces caractéristiques par la méthode du maximum de vraisemblance dépend fortement du repère de référence dans lequel est effectuée l'analyse (Gunn (1988)). Groth et *al.* (1989) remarquent par exemple que l'échelle de cohérence du "grand Attracteur" diminue considérablement lorsqu'on exprime le champ des vitesses observées dans le repère où notre Groupe Local est au repos.

Or diverses études indépendantes tendent à montrer que le "Grand Attracteur" possède en effet un mouvement propre par rapport au repère du fond diffus cosmologique.

Willick (1990) (puis Han et *al.* (1992)) démontre que les vitesses des galaxies du super-

amas Perseus-Pisces ($l \approx 120$, $b \approx -30$ et $r \approx 7000 \text{ km.s}^{-1}$) d'environ 450 km.s^{-1} dans la direction du "Grand Attracteur" ne peuvent être inférées par le "Grand Attracteur" uniquement (contribution d'environ 100 km.s^{-1} seulement). La vitesse propre de Perseus-Pisces ne peut alors être expliquée que si l'on suppose l'existence d'un champ de vitesses cohérent à une échelle d'au moins $100 \text{ h}^{-1} \text{ Mpc}$ dont la source serait localisée dans les régions situées derrière le "Grand Attracteur". Un tel champ de vitesses doit alors communiquer au "Grand Attracteur" un mouvement propre dans le repère du fond diffus cosmologique.

L'utilisation de tests géométriques invariants par changement de repère (Rauzy et al. (1992), voir section D.II.) nous permet de corroborer cette hypothèse.

Des études plus récentes confirment d'ailleurs la présence d'un flot de vitesses à très grande échelle (Shaya et al. (1992)) dont la direction est grossièrement parallèle à la direction du "Grand Attracteur". Il est alors fondamental d'échantillonner le champ des vitesses derrière la région du "Grand Attracteur". Si le "Grand Attracteur" est au repos dans le repère du fond diffus cosmologique, les galaxies derrière le "Grand Attracteur" doivent posséder une vitesse de chute importante (Dressler et al. (1990) ou Burstein et al. (1990)). En revanche, cette vitesse de chute sur l'attracteur est atténuée dans le cas de figure où l'attracteur possède une vitesse propre qui compense le mouvement de chute des galaxies situées derrière le "Grand Attracteur". Mathewson et al. (1992b) ainsi que Bothun et al. (1992) ne détectent qu'une faible chute de ces galaxies situées derrière le "Grand Attracteur". Plus significativement encore, il semble que les galaxies loin derrière le "Grand Attracteur" ($r \approx 7000 \text{ km.s}^{-1}$) possèdent des vitesses similaires en intensité ainsi qu'en direction aux vitesses des galaxies peuplant le super-amas Perseus-Pisces (Mathewson et al. (1992b)). Ce comportement est bien caractéristique de la présence d'un attracteur (peut-être d'amplitude moins importante que celle déduite par Lynden-Bell et al. (1988a), transporté par un flot de vitesses cohérent à très grande échelle (au moins 15000 km.s^{-1}) dont la direction est approximativement parallèle à celle de l'attracteur.

L'origine d'un tel flot réclame alors l'existence de concentrations de matière gigantesques (peut-être détectées par Scaramella et al. (1989) au alentour de $r = 14000 \text{ km.s}^{-1}$ grossièrement dans la direction du "Grand Attracteur" et 3 fois plus éloignées). Il est néanmoins difficile d'expliquer avec les théories classiques de formation de structures la cohérence du champ des vitesses à de telles échelles.

Il semble donc que les galaxies qui nous entourent dans un rayon de 100 Mpc en-

viron adhérent à un mouvement d'ensemble par rapport au repère défini par le fond diffus cosmologique. Ce mouvement d'ensemble disparaît quasiment si l'on choisit comme référentiel des vitesses le repère dans lequel notre galaxie est au repos. Gunn (1988) remarque que si le dipole en température détecté n'est pas la signature du mouvement propre de notre galaxie par rapport au fond diffus cosmologique (soit que notre galaxie serait grossièrement au repos par rapport à ce rayonnement), on élimine les problèmes liés à la présence d'un flot de vitesses cohérent à très grande échelle. Néanmoins, il est alors nécessaire d'expliquer l'origine de cette anisotropie dipolaire en température détectée. Des cosmologies du type Tolman-Bondi impliquant un univers inhomogène à symétrie sphérique prévoient par exemple l'existence d'une anisotropie dipolaire dans le fond diffus cosmologique (Paczynski et al. (1990)).

. Influence des caractéristiques de l'attracteur :

L'analyse de la cinématique de l'échantillon de Faber et al. par la technique du maximum de vraisemblance nécessite la modélisation du champ des vitesses. Le concept de "Grand Attracteur" stipule qu'il existe une concentration de masse, à symétrie sphérique autour d'un centre, responsable du champ des vitesses régissant la cinématique de notre voisinage dans un large volume. D'après les théories classiques de formation de structures, ce champ des vitesses à grande échelle reste parallèle au champ d'accélération, et donc converge vers le centre de la concentration de masse. La géométrie du flot étant ainsi fixée, on doit choisir une forme pour le profil des vitesses engendrées par l'attracteur.

Initialement, Lynden-Bell et al. (1988a) modélisèrent ce profil des vitesses par l'équation $v(r) = V_m R_m / r$ avec r la distance séparant le point du centre de l'attracteur, R_m la distance du Groupe Local à l'attracteur et V_m la vitesse induite par l'attracteur à notre Groupe Local. On s'attend à un tel profil lorsque la densité de matière varie en $1/r^2$ à partir du centre de l'attracteur.

Des modifications ont été apportées plus tard à cette modélisation (voir Bertschinger et al. (1988) par exemple). Le nouvel attracteur, qui possède un cœur de rayon c_A , une densité de matière $\rho_A(r) \propto r^{n_A+1}$ et $d_A = R_m$, $v_A = V_m$, est décrit par l'équation suivante :

$$v(r) = -v_A \left(\frac{r}{d_A} \right) \frac{[d_A^2 + c_A^2]^{(n_A+1)/2}}{[r^2 + c_A^2]^{(n_A+1)/2}} \quad (D.2)$$

Après avoir maximisé la fonction de vraisemblance, on obtient les valeurs : $v_A = 535 \text{ km.s}^{-1}$, $r_A = d_A = 4200 \text{ km.s}^{-1}$, $n_A = 1.7$ et $c_A = 0.34 d_A$. Si les caractéristiques de ce "Grand Attracteur" n'ont pas énormément variées par rapport au modèle précédent (150 km.s^{-1} en distance et 15° en direction), il est pourtant à craindre que le choix du modèle influence la détermination des paramètres définissant le "Grand Attracteur". En fait, il n'existe aucune raison théorique nous incitant à choisir un profil de vitesses plutôt qu'un autre, et ceci va introduire un arbitraire difficilement quantifiable dans l'évaluation des paramètres. C'est l'un des points faibles de la technique du maximum de vraisemblance.

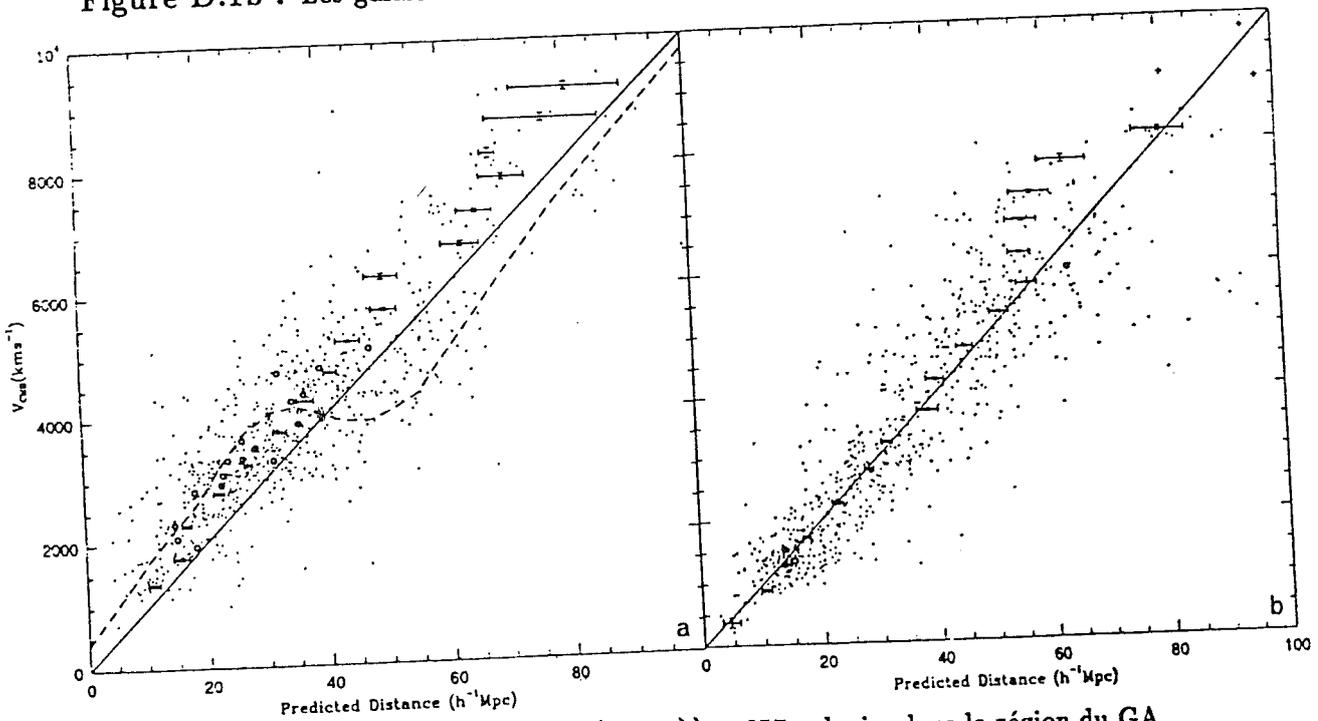
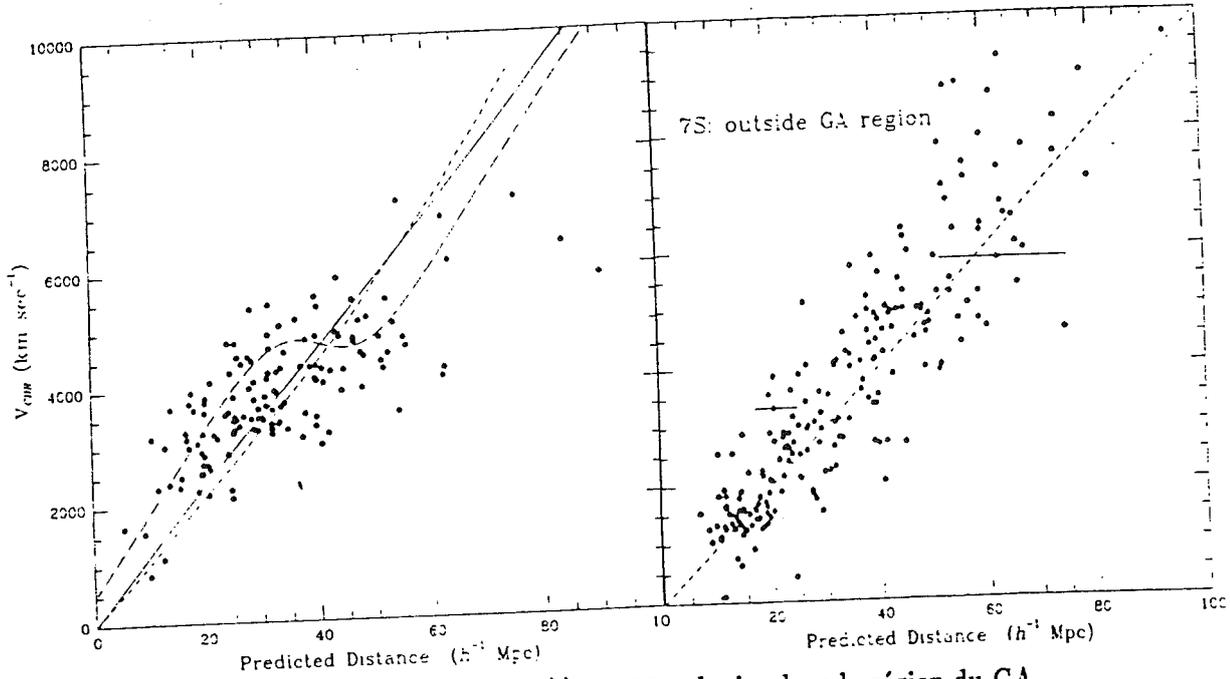
1.2.3. Le flot de vitesses derrière le "Grand Attracteur" :

Les nombreuses polémiques soulevées par l'existence du "Grand Attracteur" ont poussées Dressler et *al.* (1990) à mesurer les vitesses propres des galaxies derrière la région du "Grand Attracteur". Si le flot de vitesses convergent est vraiment dominé par le "Grand Attracteur", les vitesses radiales propres des galaxies observées dans la direction du centre du flot possèdent en effet des propriétés particulières. Les vitesses radiales propres des galaxies résidant entre l'observateur et le centre du "Grand Attracteur" doivent être positives et celles derrière l'attracteur négatives.

Dressler et *al.* (1990) montrent figure (D.1a) le diagramme de Hubble (les vitesses radiales des galaxies en fonction de leurs distances estimées) pour 136 galaxies appartenant à un cône pointé vers la direction ($l = 310$ et $b = 20$) et d'ouverture 40° . La courbe hachurée est calculée à partir du modèle de "Grand Attracteur" de Faber et *al.* (1989) (centré en $r = 42 h^{-1} \text{ Mpc}$) et représente les effets d'un attracteur sur le champ de vitesses radiales des galaxies. La figure (D.1b) présente le même diagramme pour les galaxies à l'extérieur de la région du "Grand Attracteur".

Les vitesses radiales propres dans la région du "Grand Attracteur" sont nettement positives jusqu'à une distance d'environ 3000 km.s^{-1} et semblent devenir négatives vers 4500 km.s^{-1} . Ce comportement n'est pas observé pour les galaxies à l'extérieur de la région du "Grand Attracteur". Les auteurs concluent de ces mesures la vraisemblable existence du "Grand Attracteur".

La détection de la chute des galaxies derrière Le "Grand Attracteur" ne peut pourtant pas être affirmée avec certitude (Dressler et *al.* (1990), Burstein et *al.* (1990)). Certains



biais, du type "biais de Malmquist", apparaissent en effet lors de l'évaluation des distances les plus probables des galaxies. Ces erreurs, qui se répercutent sur l'estimation du champ des vitesses propres, sont hélas difficiles à corriger compte-tenu de l'inhomogénéité spatiale de la distribution des galaxies par exemple. Ainsi, le mouvement de chute détecté pour l'échantillon de Dressler et *al.* (1990) n'est pas statistiquement significatif.

Plus récemment Mathewson et *al.* (1992b) ont réalisé la même étude pour un échantillon de 1355 galaxies elliptiques situées dans l'hémisphère Sud (Mathewson et *al.* (1992a)). La figure (D.2a) présente le diagramme de Hubble pour 657 galaxies dont les coordonnées galactiques sont comprises entre $l \in [260, 360]$ et $b \in [-40, 45]$ (leurs directions pointent donc vers la région du "Grand Attracteur") et la figure (D.2b) montre le même diagramme pour les galaxies à l'extérieur de cette région. On constate que ces données n'indiquent nullement un mouvement de chute derrière la région du "Grand Attracteur". Au contraire, les vitesses radiales des galaxies restent positives jusqu'aux limites du catalogue. Mathewson et *al.* remarquent que ce comportement est la signature d'un flot de vitesses constant transportant avec lui toutes les galaxies de l'échantillon, y compris un éventuel attracteur.

Ainsi la cinématique dans la région du "Grand Attracteur" est dominée pour les uns par les effets du "Grand Attracteur" (Dressler et *al.* (1990), Burstein et *al.* (1990)), et par la présence d'un champ de vitesses constant (du moins jusqu'aux limites de l'échantillon) pour les autres (Mathewson et *al.* (1992)). Nous proposons dans la section suivante (Rauzy et *al.* (1992)) une interprétation réconciliant ces deux effets : Le champ des vitesses est engendré par un attracteur en mouvement par rapport au fond diffus cosmologique. Cette interprétation possède le mérite de rendre compte d'une part de la détection du flot quadripolaire (Lilje et *al.* (1986), Lynden-Bell et *al.* (1988)) et d'autre part des vitesses radiales positives observées derrière le "Grand Attracteur" par Mathewson et *al.* . Une étude approfondie du catalogue de Mathewson et *al.* devrait nous permettre de tester cette hypothèse.

I.3. Les implications :

Dans la section précédente, nous avons montré que la cinématique de notre voisinage dans un large volume semble gouvernée par la présence du "Grand Attracteur" et peut-être même par un champ de vitesses cohérent à plus grande échelle. Mais l'existence de tels flots représente un "challenge" pour les théories classiques de formation de structures.

D'une part le champ de densité de masse totale (reconstruit à partir du champ des vitesses) et le champ de densité de matière lumineuse possèdent des différences remarquables. D'autre part, la présence actuelle de structures du type "Grand Attracteur" doit créer des fluctuations dans le rayonnement du fond diffus cosmologique (par diffusion Compton inverse ou par effet Sachs-Wolfe pour des structures actuellement non-linéaires), non encore détectées. Le seuil expérimental de détection de ces fluctuations permet alors d'obtenir des limites supérieures pour la longueur de cohérence ainsi que l'amplitude du champ des vitesses engendré par l'attracteur.

I.3.1. Identification du "Grand Attracteur" :

Si le "Grand Attracteur" de Lynden-Bell et *al.* est réellement présent, il doit alors exister une forte concentration de masse approximativement localisée autour de ($r = 4200$ km.s⁻¹, $l = 310$ et $b = 20$) et responsable de ce champ des vitesses cohérent à grande échelle.

En fait, on détecte dans la direction du "Grand Attracteur" deux principales concentrations de galaxies : l'amas de Centaurus (à une distance d'environ 3000 km.s⁻¹) et une concentration lumineuse à 4500 km.s⁻¹ qui pourrait être associée au "Grand Attracteur" (Lynden-Bell et *al.* (1988b)). Dressler (1991) remarque néanmoins que, du fait de l'absorption galactique, l'identification optique du "Grand Attracteur" est difficile (le "Grand Attracteur" est approximativement situé dans le plan galactique). En pratique donc, les observations dans le domaine du visible ne permettent pas une étude quantitative de la distribution de matière dans la direction du "Grand Attracteur".

En revanche, le catalogue de redshifts mesurés par le satellite IRAS (Infrared Astronomical Satellite) nous fournit la distribution des galaxies jusqu'à une profondeur d'environ $140 h^{-1}$ Mpc (Saunders et *al.* (1991)). Cet échantillon est particulièrement bien

adapté à l'étude de la densité de matière lumineuse puisque les magnitudes collectées dans l'infrarouge sont moins concernées par l'absorption galactique que leurs équivalents dans le domaine du visible (les données IRAS couvrent tout le ciel pour $|b| > 10^\circ$). Des études préliminaires ont montré que les galaxies IRAS ne possédaient pas les mêmes propriétés que les échantillons établis dans le domaine du visible (Yahil et al. (1986)). Il semble cependant que les données IRAS actuelles (environ 3000 galaxies) soient représentatives du champ de densité de la matière lumineuse pour des échelles supérieures à $5 - 10 h^{-1}$ Mpc (Strauss et al. (1992)).

En supposant que la distribution spatiale des galaxies du catalogue IRAS est représentative de la distribution de la masse totale (soit que la masse cachée possède le même comportement que la masse visible), il est possible de déterminer le champ des vitesses propres engendré par cette distribution de matière (Yahil et al. (1991)). Notamment, on peut déduire de cette densité de masse lumineuse la vitesse de notre Groupe Local par rapport au repère défini par le fond diffus cosmologique (Rowan-Robinson et al. (1990) ou Lynden-Bell et al. (1989) dans le domaine du visible). Cette vitesse converge vers celle évaluée indépendamment par le dipole observé dans le fond diffus cosmologique lorsque la contribution massique de toutes les galaxies dans un rayon de $100 h^{-1}$ Mpc est prise en compte. Autrement dit, si la matière lumineuse (les galaxies) trace la matière totale, notre mouvement propre dans le repère du fond diffus cosmologique est engendré gravitationnellement par la distribution de masse présente dans une sphère de rayon $100 h^{-1}$ Mpc.

Rowan-Robinson et al. (1990) concluent de ce résultat qu'il n'est pas nécessaire d'introduire un "hypothétique Grand Attracteur". La présence de 12 amas de galaxies identifiés optiquement suffirait en effet à expliquer le mouvement de notre Groupe Local. Une conclusion un peu hâtive semble-t-il. D'une part le concept de "Grand Attracteur" ne signifie pas forcément une concentration de masse fortement localisée (voir par exemple Heavens (1991)). D'autre part, il faut que le champ des vitesses déduit à partir des données IRAS coïncide en tout point de l'espace avec le champ de vitesses réellement observé (et non pas seulement à la position de notre Groupe Local).

En fait, deux approches permettent de dériver les propriétés dynamiques locales de notre univers, comme le champ de densité de masse totale ou son champ des vitesses propres associé :

- (1) En supposant que le champ des vitesses propres est irrotationnel à grande échelle

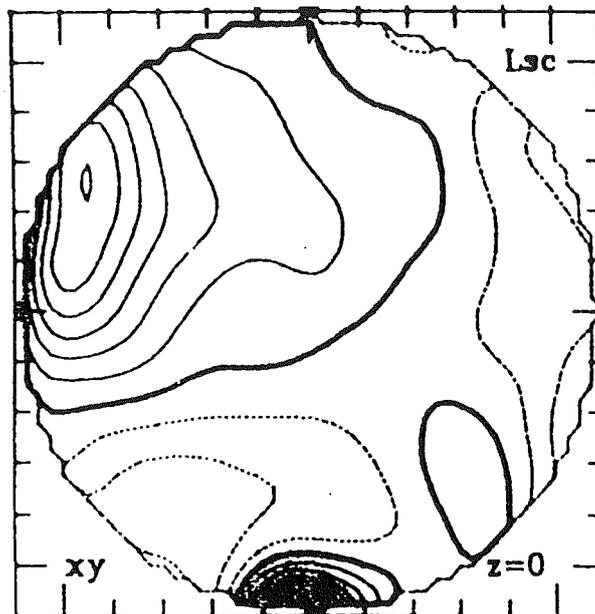


Figure D.3 (Bertschinger et al. (1990b)) : Champ de densité (dérivé des vitesses). Plan supergalactique. lignes pleines = surdensité, hachurées = sous-densité

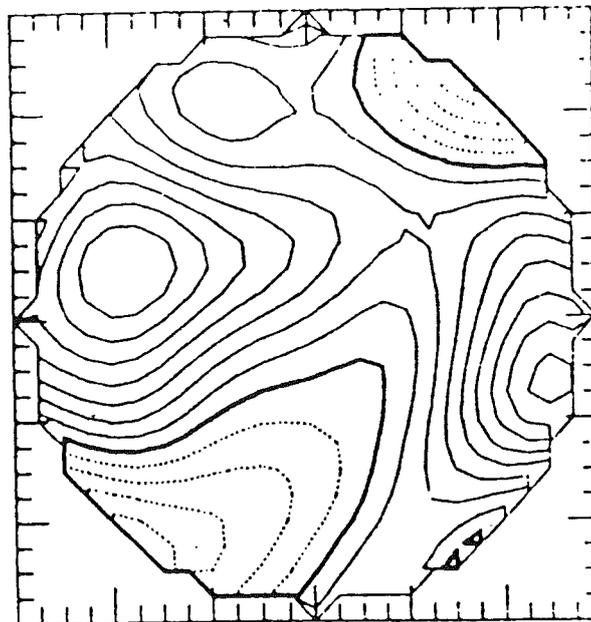


Figure D.4 (Saunders et al. (1991)) : Champ de densité (dérivé des données IRAS). Plan supergalactique. lignes pleines = surdensité, hachurées = sous-densité

(typiquement supérieure à 5 Mpc), il est possible de reconstruire le champ de densité total à partir du champ des vitesses radiales observées (méthode POTENT (Bertschinger et *al.* (1989), Dekel et *al.* (1990) et Bertschinger et *al.* (1990)).

(2) En supposant que la masse cachée possède le même comportement que la masse lumineuse (la distribution massique des galaxies dans l'espace est représentative de la distribution de la masse totale), des méthodes permettent de reconstruire le champ de densité total à partir de la distribution observée de la population des galaxies (Yahil et *al.*, Saunders et *al.* (1991), Strauss et *al.* (1992)).

Nous montrons figures (D.3) et (D.4) les isocontours du champ de densité (les courbes de niveau en traits pleins indiquent les surdensités, celles en traits hachurés les sous-densités) reconstruit dans le plan super-galactique à partir de : (D.3) La méthode POTENT appliquée au catalogue de vitesses radiales propres de Faber et *al.* (1989) (Bertschinger et *al.* (1990)). (D.4) La distribution massique des galaxies de l'échantillon IRAS dans l'hypothèse que celle-ci trace la distribution de masse totale (Saunders et *al.* (1991)). La différence est marquante. Le pic de densité correspondant au "Grand Attracteur" (à gauche sur les figures) est présent pour les deux reconstructions. Mais le pic identifié au super-amas Perseus-Pisces dans la reconstruction IRAS (à droite sur la figure) est complètement absent de l'analyse à partir du champ des vitesses. Plusieurs explications peuvent être avancées afin d'expliquer cette différence :

La distribution de la matière lumineuse n'est pas représentative de la densité totale de masse présente dans l'univers (Dressler (1991)) et nous devons abandonner tout espoir d'étudier la dynamique de notre univers à partir de la distribution des galaxies (Gunn (1988)).

L'échantillon de galaxies IRAS n'est pas suffisamment grand de sorte à contenir toute la masse responsable des propriétés dynamiques de notre voisinage dans un rayon de 50 Mpc environ.

La reconstruction du champ de densité par la méthode POTENT n'est pas fiable dans la région du super-amas Perseus-Pisces (Bertschinger et *al.* (1990)).

Le catalogue de vitesses radiales de Faber et *al.* est soumis à des biais.

La différence remarquable entre ces deux reconstructions du champ de densité de matière totale reste actuellement une énigme à élucider.

I.3.2. La cohérence du champ des vitesses :

Nous avons déjà présenté dans la partie C (section II.1.3) les difficultés rencontrées par les théories classiques de formation de structures afin de rendre compte de la cohérence du champ des vitesses à grande échelle. Il ne fait aucun doute maintenant que des flots de vitesses cohérents à une échelle d'au moins $60 h^{-1}$ Mpc régissent la cinématique de notre voisinage.

Dans les scénari de formation de structures comme CDM, la probabilité d'observer de tels flots est assez faible. Néanmoins ces prédictions varient considérablement suivant les auteurs. Si le "Grand Attracteur" signe l'arrêt de mort des modèles CDM pour certains (par exemple Bertschinger et *al.* (1988)), Kaiser (1991) démontre que l'existence du "Grand Attracteur" ne permet pas d'éliminer les modèles CDM mais tout au plus pose certaines contraintes sur la valeur des paramètres impliqués dans ces modèles.

On peut toutefois remarquer qu'on a dû progressivement ajouter des nouveaux paramètres aux modèles classiques de formation de structures afin d'expliquer les observations du champ des vitesses (comme le paramètre de biais pour le modèle CDM). La multiplication de ces paramètres altère quelque peu la simplicité originelle de ces modèles, les rendant de ce fait moins attrayants.

I.3.3. Les anisotropies dans le fond diffus cosmologique :

Des structures telles que le "Grand Attracteur" perturbent en principe l'homogénéité du rayonnement du fond diffus cosmologique. Les photons qui nous parviennent cèdent en effet une partie de leur énergie lorsqu'ils traversent la région de l'espace occupée par un attracteur. Il est alors possible, connaissant l'ampleur des fluctuations du potentiel gravitationnel engendrées par la présence du "Grand Attracteur", de prédire l'amplitude de l'anisotropie résultante pour la température du fond diffus cosmologique.

La valeur de l'amplitude de cette anisotropie varie néanmoins sensiblement suivant les processus et le choix des paramètres pris en compte durant la phase de calcul (Vittorio et *al.* (1984), Martinez et *al.* (1989)). L'amplitude maximum de l'anisotropie engendrée par la présence du "Grand Attracteur" est atteinte pour un angle d'environ 1° , ce qui correspond actuellement à une région de l'espace d'un diamètre de 100 Mpc. Typiquement, la

valeur prédite des fluctuations en température est de l'ordre de $\Delta T/T \approx 1.7 \cdot 10^{-5}$ pour un angle de 1° (Bertschinger et al. (1990)), $\Delta T/T > 10^{-5}$ pour un angle de 2° (Juszkiewicz et al. (1987)) ou une anisotropie quadripolaire d'environ $7 \cdot 10^{-6}$ (Górski (1991)).

La limite expérimentale de détection des fluctuations de température du rayonnement du fond diffus cosmologique est actuellement au dessus de toutes ces prédictions théoriques. En fait, aucune anisotropie n'est détectée dans le fond diffus cosmologique, ce qui impose une limite supérieure à l'amplitude des perturbations du rayonnement engendrées par le "Grand Attracteur" ($\Delta T/T < 3 \cdot 10^{-5}$ pour un angle de 1° (Górski (1991))). Ce seuil expérimental de détection devrait être amélioré ces prochaines années pour passer sous la barre des $\Delta T/T = 10^{-5}$. Les résultats récents obtenus par le satellite COBE (Smoot et al. (1992)) annoncent d'ailleurs la détection d'une possible anisotropie dans le rayonnement cosmique pour un angle de 7° (ce qui correspond à une échelle d'environ $800 h^{-1}$ Mpc actuellement). Ces mesures ne nous apprennent malheureusement rien sur des anisotropies à plus petites échelles, directement comparables aux prédictions théoriques établies à partir des flots de vitesses observés (Wright et al. (1992)). De plus, il est nécessaire d'attendre la confirmation de ces résultats, le signal observé étant à la limite du seuil de détection du radiomètre embarqué à bord du satellite COBE.

Les observations simultanées des flots de vitesses cohérents à grande échelle et des fluctuations du rayonnement du fond diffus cosmologique représentent un test fondamental pour les théories classiques de formation de structures.

L'obtention de catalogues de vitesses de galaxies de plus en plus profonds nous permet d'étudier la cinématique de l'univers à grande échelle, dominée par des champs de vitesses cohérents dans un volume d'au moins $100 h^{-1}$ Mpc de diamètre. Le potentiel gravitationnel associé à de tels flots de vitesses crée alors des fluctuations dans le rayonnement du fond diffus cosmologique d'amplitude d'autant plus grande que le pic du potentiel est prononcé. Ainsi l'amplitude des champs de vitesses observés (comme le "Grand Attracteur") impose une limite inférieure pour l'anisotropie en température du rayonnement cosmologique.

Les tentatives de mesure de fluctuations de température dans le fond diffus cosmologique pour des angles de quelques degrés ont été jusqu'à présent infructueuses. Si des anisotropies existent dans le rayonnement cosmologique, leurs amplitudes est alors inférieures au seuil expérimental de détection des appareils de mesure.

Les prochaines années vont être déterminantes pour le devenir des théories classiques

de formation de structures, puisque le seuil de détection des perturbations du fond diffus devrait tomber sous les limites inférieures imposées par la présence de structures telles que le "Grand Attracteur". Si des fluctuations en température du rayonnement cosmique ne sont pas détectées dans la direction du "Grand Attracteur", alors c'est la fin des modèles classiques décrivant le processus de formation des structures à grande échelle actuellement observées.

II. LES TESTS DES SPHERES :

Nous avons vu dans la section précédente que l'analyse de Lynden-Bell et *al.* ayant permis la mise en évidence du "Grand Attracteur" souffrait de quelques faiblesses. D'une part, l'estimation des valeurs des paramètres libres du modèle n'est pas indépendante du repère de référence choisi afin d'exprimer le champ des vitesses propres observé (voir section I.2.2). D'autre part, le nombre important de paramètres libres intervenant lors de la maximisation de la fonction de vraisemblance rend cette étude statistique fragile.

Nous présentons dans cette section une méthode impliquant un minimum de paramètres libres qui permet la détection de champs des vitesses du type "Grand Attracteur", ceci indépendamment du repère des vitesses de référence. Notre analyse est basée sur la propriété géométrique suivante. A chaque flot de vitesses convergent vers un centre G , il est possible d'associer une sphère ($ZRVS$) qui passe par G ainsi que par l'observateur O telle que OG est l'un de ses diamètres. Le champ des vitesses radiales observé est alors nul sur cette sphère, positif à l'intérieur et négatif à l'extérieur. L'utilisation de cette propriété ainsi que de certaines symétries inhérentes au problème nous permet en fait de distinguer un champ de vitesses engendré par un attracteur d'avec un flot de vitesses constant. De plus nos tests géométriques sont théoriquement insensibles à la présence d'un champ de Hubble résiduel affectant les données (dû par exemple à une soustraction incomplète du champ de Hubble) et ne nécessitent pas la modélisation du champ des vitesses.

L'application de ces tests à l'échantillon de galaxies elliptiques de Faber et *al.* (1989) révèle la présence d'un attracteur d'extension spatiale importante centré à une distance $r = 3800 \text{ km.s}^{-1}$ dans la direction ($l = 280, b = 10$ (en coordonnées galactiques)). Si le catalogue de Faber et *al.* est fiable, les résultats de notre analyse nous permettent de plus de conclure que cet attracteur ainsi que toutes les galaxies de l'échantillon ne sont pas au repos dans le repère défini par le fond diffus cosmologique.

II.1. Définitions :

Nous désirons, avant de présenter les définitions de nos différents tests, caractériser les propriétés géométriques du champ des vitesses à analyser. On a vu précédemment dans la partie A (section II.2.2) que le champ des vitesses des galaxies se décomposait dans l'approximation Newtonienne de la façon suivante :

$$\mathbf{v} = \mathbf{v}_H + \mathbf{v}_C + \mathbf{v}_p \quad (\text{A.21})$$

On peut théoriquement soustraire le champ de Hubble \mathbf{v}_H afin d'obtenir le champ des vitesses propres \mathbf{v}_p . Néanmoins cette opération nécessite l'estimation de la constante de Hubble H_0 ou la détermination de la distance réelle d'un amas de galaxies (voir partie B, section IV.3.2). Nous avons vu que H_0 est en fait mal connue (voir partie B, section III). Il est donc fort possible qu'après soustraction du flot de Hubble, le champ des vitesses propres \mathbf{v}_p demeure contaminé par un champ de Hubble résiduel. Plus généralement, une mauvaise correction du biais de "Malmquist" intervenant dans l'évaluation de la distance d'une galaxie perturbe l'estimation du champ des vitesses propres réel par un champ radial \mathbf{V}_H dont l'intensité varie seulement en fonction de la distance à l'observateur.

Le flot constant \mathbf{v}_C apparaît naturellement dans l'expression du champ des vitesses total. Il varie selon le choix du repère de référence pour les vitesses. En général, on fait l'hypothèse que ce champ constant est nul dans le repère pour lequel le rayonnement du fond diffus cosmologique est au repos (soit que le repère défini par le fond diffus est absolu (voir partie A, section II.2.1)). Toutefois il peut exister des flots à grande échelle qui provoquent un mouvement d'ensemble de toutes les galaxies de l'échantillon par rapport à ce repère absolu. Nous retenons donc à priori une possible composante constante \mathbf{V}_C pour le champ des vitesses propres.

Enfin, nous voulons mettre en évidence des champs de vitesses convergents vers un centre du type "Grand Attracteur". De tels flots \mathbf{V}_A possèdent une symétrie sphérique autour du centre de l'attracteur. Afin d'introduire un minimum de paramètres dans notre analyse, nous ne modélisons pas le profil des vitesses engendrées par l'attracteur. Nous permettons de plus la présence d'un champ statistiquement homogène et isotrope \mathbf{V}_{HI} , de moyenne nulle pour un volume assez grand. Nous faisons donc l'hypothèse que le champ des vitesses propres des galaxies de l'échantillon à analyser s'exprime de la façon suivante

:

$$\mathbf{V} = \mathbf{V}_H + \mathbf{V}_C + \mathbf{V}_A + \mathbf{V}_{HI} \quad (D.3)$$

Nos tests sont basés sur les propriétés géométriques vérifiées par le champ des vitesses \mathbf{V}_A convergent vers un centre GA . Nous associons à ce flot \mathbf{V}_A la sphère (*ZRVS*, Zero Radial Velocity Sphere) passant par le point centre GA et l'observateur O telle que le segment d'extrémités O et GA soit un diamètre de cette sphère. En chaque point de cette sphère (*ZRVS*) la vitesse radiale observée ($V_r = \hat{\mathbf{r}} \cdot \mathbf{V}_A$) est alors nulle, le vecteur \mathbf{V}_A formant un angle droit avec la ligne de visée (voir figure 1 de l'article en annexe). On constate de plus que le champ des vitesses radiales est positif à l'intérieur de la sphère et négatif à l'extérieur. La détection de cette sphère nous permet en fait de déterminer la direction ainsi que la distance du centre du flot convergent qui lui est associé.

Nous proposons deux tests complémentaires qui permettent d'exploiter au maximum les symétries présentes dans la géométrie du champ des vitesses. Nous définissons deux boules B_+ et B_- de diamètre D et tangentes en O , telles que la droite passant par les centres des deux boules ainsi que par l'observateur O ait pour direction $\mathbf{d} = (l, b)$ (orientée vers la boule B_+) (voir figure 2 en annexe). Nous formons les deux fonctions linéaires F_+ et F_- suivantes :

$$F_{\pm}(\mathbf{d}, D; \mathbf{V}) = \int_{B_+} \hat{\mathbf{r}} \cdot \mathbf{V} dV \pm \int_{B_-} \hat{\mathbf{r}} \cdot \mathbf{V} dV \quad (D.4)$$

Ces deux quantités F_+ et F_- sont donc fonctions de 3 paramètres qui fixent le diamètre D ainsi que la direction \mathbf{d} des boules B_+ et B_- .

II.1.1. Le test dipolaire des sphères :

Le test dipolaire des sphères (*DTS*) implique la quantité F_- . La linéarité de cette fonction nous permet de traiter indépendamment les différentes composantes du champ des vitesses total \mathbf{V} .

Pour un volume d'espace assez grand, l'intégrale dans les boules B_+ et B_- de la composante radiale d'un champ des vitesses statistiquement homogène et isotrope s'annule. Ainsi, quelquesoient la direction \mathbf{d} et le diamètre D des boules :

$$F_-(\mathbf{d}, D; \mathbf{V}_{HI}) = 0 \quad (D.5)$$

On peut de plus montrer facilement que pour un champ à symétrie sphérique autour de l'observateur O du type champ de Hubble V_H , la fonction F_- est nulle, l'intégrale du champ radial dans la boule B_+ étant égale à l'intégrale dans la boule B_- . On a donc :

$$F_-(d, D; V_H) = 0 \quad (D.6)$$

La quantité F_- ne dépend plus que du champ engendré par l'attracteur V_A ainsi que du champ constant V_C :

$$F_-(d, D; V) = F_-(d, D; V_A) + F_-(d, D; V_C) \quad (D.7)$$

On voit clairement dans le cas du champ convergent V_A que F_- est maximum lorsque la boule B_+ est superposée à la sphère $ZRVS$ associée à l'attracteur. De même, la valeur maximale de F_- pour un champ constant V_C est atteinte lorsque d est alignée avec la direction du flot constant et le diamètre de B_+ et B_- tend vers l'infini.

Le test dipolaire des sphères nous permet donc de déterminer la direction d'un flot constant V_C ou les coordonnées du centre d'un flot convergent V_A . Il faut néanmoins remarquer que si ces deux flots sont simultanément présents, le test dipolaire des sphères n'est pas apte à séparer les contributions des deux champs.

II.1.2. Le test quadripolaire des sphères :

Nous nous intéressons cette fois à la fonction F_+ . De la même façon que pour F_- , la contribution d'un champ statistiquement homogène et isotrope à F_+ est nulle :

$$F_+(d, D; V_{HI}) = 0 \quad (D.8)$$

De plus, le test quadripolaire des sphères est insensible à la présence d'un champ constant, la contribution à l'intégrale de chaque point dans la boule B_+ s'annulant avec la contribution de son symétrique par rapport à O dans la boule B_- :

$$F_+(d, D; V_C) = 0 \quad (D.9)$$

L'expression de la quantité F_+ devient alors :

$$F_+(d, D; V) = F_+(d, D; V_A) + F_+(d, D; V_H) \quad (D.10)$$

On peut aussi remarquer que le diamètre des boules étant fixé, la contribution à F_+ d'un champ de Hubble V_H ne change pas suivant la direction :

$$F_+(d, D_{\text{fixé}}; V_H) = \text{constante} \quad (D.11)$$

Le maximum de F_+ pour un flot convergent V_A étant atteint lorsque la boule B_+ est superposée à la sphère associée à l'attracteur, le test quadripolaire des sphères nous permet donc de détecter un champ de vitesses du type "Grand Attracteur" indépendamment du repère de référence.

Ainsi nous pouvons utiliser les tests dipolaire et quadripolaire de façon complémentaire. Premièrement, nous localisons le "Grand Attracteur" à l'aide du test quadripolaire des sphères. Deuxièmement, nous comparons la position des maxima des fonctions F_+ et F_- . Si les galaxies de l'échantillon sont soumises à un champ constant V_C , alors la présence de ce flot modifie la position du maximum de la quantité F_- . Ainsi la détection d'une différence entre la position des maxima des tests dipolaire et quadripolaire des sphères indique la présence d'un mouvement d'ensemble de toutes les galaxies de l'échantillon par rapport au repère des vitesses dans lequel est effectuée l'analyse.

II.2. Application :

Afin de quantifier la significativité de notre analyse, nous avons d'abord appliqué les tests dipolaire et quadripolaire des sphères à des données simulées. L'inhomogénéité spatiale de la distribution réelle des galaxies altère sensiblement les propriétés d'invariance de nos tests. Néanmoins, nous introduisons un processus de pondération qui corrige ces effets indésirables. Ceci nous permet en fait de localiser la présence d'un attracteur simulé avec une incertitude d'environ 10° en direction et 200 km.s^{-1} en distance.

L'analyse de l'échantillon de galaxies elliptiques de Faber et *al.* (1989) dans le repère du fond diffus cosmologique nous permet de tirer les deux conclusions suivantes.

Le test quadripolaire des sphères converge vers une région de l'espace centrée en ($r = 3800 \text{ km.s}^{-1}$, $l = 280$ et $b = 10$) d'extension spatiale importante (entre $r = 3400 \text{ km.s}^{-1}$ et $r = 4200 \text{ km.s}^{-1}$). Ce résultat, indépendant du repère des vitesses dans lequel est effectuée l'analyse, confirme l'existence du "Grand Attracteur".

A une distance de $r = 4000 \text{ km.s}^{-1}$, le test dipolaire des sphères converge vers la

direction ($l = 330$ et $b = 30$). Cette direction diffère de plus de 50° par rapport à la position du maximum du test quadripolaire des sphères. Le "Grand Attracteur" ainsi que toutes les galaxies de l'échantillon doivent donc posséder une vitesse de groupe dans le repère pour lequel le rayonnement du fond diffus cosmologique est au repos.

II.2.1. La significativité de notre analyse :

Dans le but de tester l'efficacité de notre méthode, nous avons simulé un échantillon de galaxies possédant la même distribution spatiale que le catalogue de Faber et *al.* (1989) et un champ de vitesses engendré par un attracteur modélisé par Bertschinger et *al.* (1988). Nous avons placé le centre de cet attracteur à une distance de $r = 4200 \text{ km.s}^{-1}$ dans la direction ($l = 310$ et $b = 20$). L'application des tests dipolaire et quadripolaire à ces données simulées nous permet en fait de quantifier l'influence de l'inhomogénéité spatiale de la distribution des galaxies du catalogue de Faber et *al.* sur les résultats de notre analyse.

Ce mauvais échantillonnage détruit les propriétés d'invariance que possèdent nos tests. On voit par exemple que pour le test quadripolaire des sphères, le nombre de galaxies dans les boules B_+ et B_- doit être identique si l'on veut que la contribution d'un flot constant V_C s'annule effectivement. Nous avons donc introduit dans l'évaluation des quantités F_+ et F_- une procédure de pondération destinée à éliminer ces effets indésirables.

Nous avons divisé chaque boule B_+ et B_- en une famille de croissants emboîtés $\{C_i\}$ d'épaisseur ΔD qui recouvrent tout le volume de la boule. A chaque croissant C_i , nous avons affecté le poids P_i égal au volume du croissant C_i . On remplace alors la valeur de l'intégrale dans les boules B_+ et B_- par la somme pondérée sur tous les croissants de la valeur moyenne des vitesses radiales des galaxies appartenant à chaque croissant. En notant $W_i = P_i/N_i$, avec N_i le nombre de galaxies appartenant au croissant C_i , les fonctions F_+ et F_- s'expriment :

$$F_{\pm} = \sum_{C_i \in B_+} W_i \sum_{j \in C_i} \hat{r}_j \cdot \mathbf{V}_j \pm \sum_{C_i \in B_-} W_i \sum_{j \in C_i} \hat{r}_j \cdot \mathbf{V}_j \quad (D.12)$$

Une telle procédure donne un poids égal aux boules B_+ et B_- et ainsi restaure les symétries vérifiées par les tests dipolaire et quadripolaire des sphères.

Pour une distance D de 4000 km.s^{-1} , le test dipolaire converge vers la direction ($l = 320$ et $b = 20$) et le test quadripolaire vers ($l = 310$ et $b = 20$) (voir figures 3 et 5

dans l'article en annexe). Nos tests détectent donc le "Grand Attracteur" simulé avec une précision en direction d'environ 10° . La détermination de la distance de l'attracteur à l'aide du test quadripolaire des sphères possède une incertitude intrinsèque de 200 km.s^{-1} (voir figure 5c). Ces erreurs ne dépendent que légèrement des caractéristiques de l'attracteur simulé et proviennent en grande partie des effets du mauvais échantillonnage non complètement corrigés par notre procédure de pondération. Nous avons de plus vérifié que le test quadripolaire des sphères est effectivement invariant par changement de repère galiléen.

II.2.2. Les résultats :

Nous avons appliqué le test quadripolaire des sphères au catalogue de galaxies elliptiques de Faber et *al.* (1989). Nous détectons, pour une distance D de 4000 km.s^{-1} un maximum dans la direction ($l = 280$ et $b = 10$) (voir figure 6 dans l'article en annexe). Nous montrons figure 6c la valeur de F_+ dans cette direction en fonction de la distance D . Nous observons un maximum centré en 3800 km.s^{-1} et d'extension spatiale importante (entre 3400 km.s^{-1} et 4200 km.s^{-1}). De plus nous avons effectué la même analyse dans le repère *LSR* (Local Standard of Rest) dans lequel notre Groupe Local est au repos. On retrouve bien la même direction de convergence pour le test quadripolaire des sphères. Le test quadripolaire nous permet donc de détecter la présence du "Grand Attracteur" indépendamment du repère des vitesses dans lequel est effectuée l'analyse.

Nous appliquons ensuite le test dipolaire des sphères dans le repère *LSR* pour une distance de 4000 km.s^{-1} . La valeur maximale de F_- (dans la direction ($l = 0$ et $b = 60$)) ne coïncide pas avec le maximum du test quadripolaire. Cette différence angulaire d'environ 140° n'est pas étonnante et nous permet de conclure que l'échantillon de galaxies n'est pas au repos dans le repère *LSR*.

Les résultats obtenus pour le test dipolaire des sphères dans le repère des vitesses défini par le fond diffus cosmologique sont en revanche plus intéressants. Pour une distance de 4000 km.s^{-1} , le maximum de ce test pointe vers la direction ($l = 330$ et $b = 30$) (voir figure 4) et diffère donc d'un angle de plus de 50° par rapport au test quadripolaire. De plus la fonction F_- évaluée dans la direction ($l = 280$ et $b = 10$) continue de croître avec la distance jusqu'aux confins de l'échantillon, contrairement à la fonction F_+ (figure

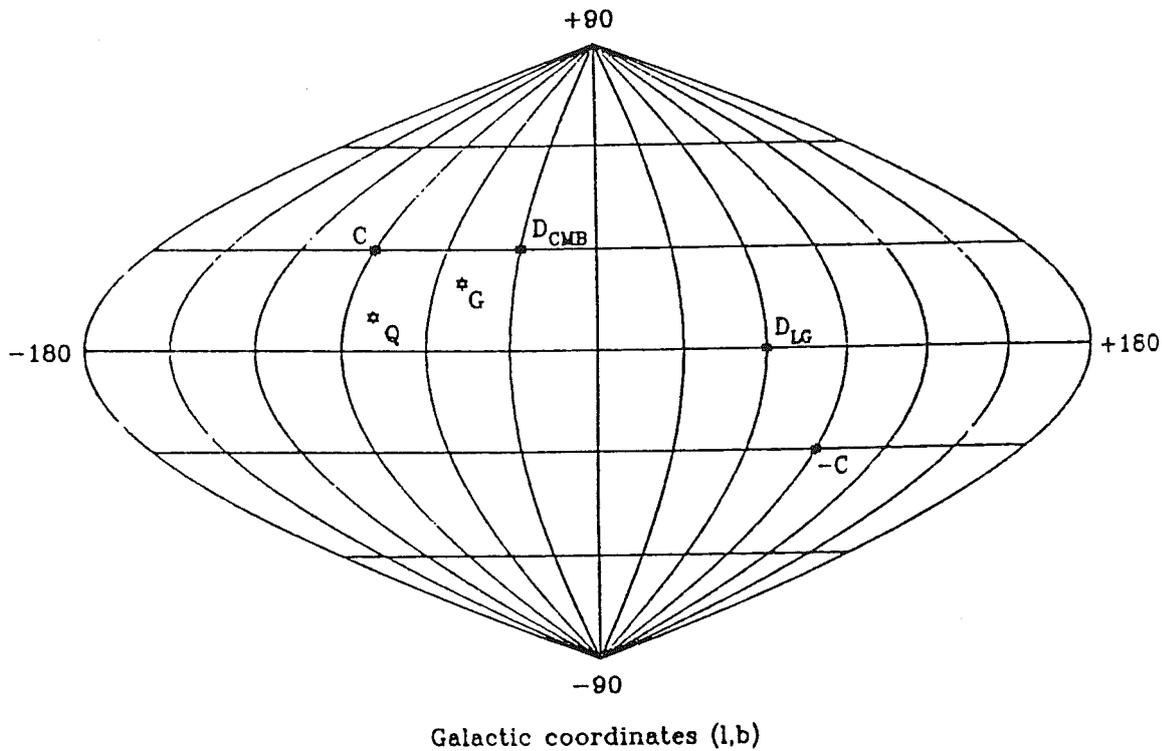


Figure D.5 : Récapitulatif des résultats des tests des sphères.

6c). Ce comportement est fortement significatif de la présence d'un champ de vitesses constant V_C .

II.2.3. Les conclusions :

Des résultats des tests dipolaire et quadripolaire des sphères appliqués au catalogue de Faber et *al.*, nous tirons les deux principales conclusions suivantes.

Premièrement, le test quadripolaire nous permet de détecter la présence d'un attracteur d'extension spatiale importante (environ 800 km.s^{-1}) distant de 3800 km.s^{-1} dans la direction Q ($l = 280$ et $b = 10$, voir figure (D.5)). La direction de notre attracteur diffère de 30° par rapport à la direction standard G du "Grand Attracteur". Nous expliquons cette différence entre les directions G et Q du fait que, contrairement à notre étude, l'analyse de Lynden-Bell et *al.* (1988a) est dépendante du choix du repère

des vitesses de référence.

Deuxièmement, la direction de convergence du test dipolaire des sphères ne correspond pas avec la direction Q du test quadripolaire des sphères, du moins lorsque le repère dans lequel notre groupe local est au repos (LSR) ou le repère défini par le fond diffus cosmologique (CMB) sont choisis comme référentiel des vitesses. On peut remarquer que la différence en direction entre D_{CMB} et Q est moins importante que celle entre D_{LG} et Q , soit que l'échantillon de Faber et *al.* possède une vitesse de groupe plus importante dans le LSR que dans le CMB . Le repère du fond diffus cosmologique est ainsi meilleur candidat que le repère LSR pour occuper la place de référentiel absolu des vitesses.

Néanmoins nous interprétons l'écart angulaire d'environ 50° entre Q et D_{CMB} comme la signature d'un mouvement d'ensemble de toutes les galaxies de l'échantillon par rapport au repère CMB . Hélas, nos tests ne nous permettent pas d'évaluer l'amplitude ainsi que la direction de ce flot.

III. LE TEST DE CONVERGENCE :

Les tests des sphères nous offrent donc la possibilité de détecter tout flot de vitesses convergent vers un point du type "Grand Attracteur". Nous avons appliqué les tests dipolaire et quadripolaire pour une distance D de 1000 km.s^{-1} dans le but d'identifier le flot Virgocentrique. Mais nos tests ne montrent aucune convergence vers la direction de l'amas de Virgo. Il semble que le nombre réduit de galaxies présentes dans les boules B_+ et B_- (une vingtaine environ) nous interdise une quelconque analyse à cette distance. En fait, c'est l'un des points faibles de nos tests dipolaire et quadripolaire. La position d'un attracteur étant spécifiée, ces tests n'utilisent pas toute l'information disponible contenue dans le catalogue.

Nous présentons dans cette section un nouveau test de convergence qui nous permet de détecter le flot Virgocentrique. Cette analyse prend en compte toute l'information de l'échantillon. Néanmoins, elle perd les propriétés d'invariance rencontrées précédemment pour les tests des sphères.

Ce test de convergence peut être vu comme une décomposition imparfaite du champ des vitesses sur une certaine famille de fonctions de l'espace. Le test de convergence nous conduit naturellement vers une analyse par ondelettes du champ des vitesses (partie E).

III.1. Définition :

Pour un point M de position \mathbf{r} et un attracteur A centré en \mathbf{r}_A , nous notons le vecteur séparation $\mathbf{R} = \mathbf{AM} = \mathbf{r} - \mathbf{r}_A$. Il est alors clair que le cosinus de l'angle entre le vecteur \mathbf{R} et la ligne de visée ($\cos \theta = [\mathbf{R} \cdot \mathbf{r}] / [Rr]$) est nul sur la sphère associée à l'attracteur, négatif à l'intérieur et positif à l'extérieur. On définit alors la fonction suivante :

$$G_+(d, D; \mathbf{V}) = \int_V -\cos \theta V_r dV \quad (D.13)$$

Lorsque les paramètres fixant la distance D ainsi que la direction \mathbf{d} de la sphère coïncident avec la position \mathbf{r}_A de l'attracteur, le produit $-\cos \theta \cdot V_r$ est maximum pour chaque galaxie contribuant à l'intégrale. Ainsi le test de convergence défini par la quantité G_+ atteint sa valeur maximum lorsque $D \cdot \mathbf{d} = \mathbf{r}_A$ et donc permet de détecter un attracteur. Nous remarquons d'une part qu'en évaluant l'intégrale sur tout l'espace, toute l'information du catalogue est utilisée. D'autre part, le test de convergence n'est pas invariant par changement de référentiel. A ce titre, il peut être comparé au test dipolaire des sphères.

Afin de tenir compte des inhomogénéités dans la distribution réelle de l'échantillon de galaxies, il est nécessaire d'introduire une pondération lors de l'évaluation de la fonction G_+ . Des considérations théoriques nous pousseraient à choisir une pondération en $-\cos \theta / R^2$ (voir article en annexe). En pratique néanmoins, cette procédure accorde trop de poids aux galaxies demeurant à proximité du point de convergence de la quantité G_+ . Nous avons finalement adopté l'expression suivante pour G_+ :

$$G_+ = \frac{1}{\mu} \sum_i -\cos \theta_i V_r^i \quad \text{avec} \quad \mu = \sum_i |\cos \theta_i| \quad (D.14)$$

III.2. Application :

Le test de convergence appliqué à l'échantillon des données simulées nous permet de retrouver les coordonnées du "Grand Attracteur" avec une très bonne précision, ce qui valide ainsi notre choix pour la procédure de pondération.

Pour l'échantillon réel et à une distance D de $4000 \text{ km} \cdot \text{s}^{-1}$, le maximum de la fonction G_+ est nettement observé pour la direction ($l = 310$ et $b = 20$) (voir figure 7 de l'article en annexe). Ces résultats confirment l'analyse réalisée précédemment avec le test dipolaire des sphères. L'écart angulaire entre la direction du test quadripolaire et du test de convergence (sensible au repère de référence) est significatif de la présence d'un flot constant V_C . De plus, la recherche en distance pour une direction fixée (voir figure 9) nous montre que la fonction G_+ ne décroît pas derrière la région du "Grand Attracteur".

III.2.1. Le flot Virgo-centrique :

Nous avons enfin appliqué le test de convergence pour une distance D de 1000 km.s^{-1} dans le but de détecter le flot Virgo-centrique. On voit figure 8 que le test converge effectivement vers la direction de l'amas de Virgo ($l = 280$ et $b = 70$). Le pic du maximum est néanmoins considérablement perturbé par la présence de champ des vitesses à plus grande échelle comme le flot constant ou le "Grand Attracteur". Toutefois, l'utilisation de toute l'information contenue dans le catalogue de Faber et *al.* nous permet de mettre en évidence la chute des galaxies sur l'amas de Virgo.

III.2.2. Vers une analyse par ondelettes du champ des vitesses :

On remarque que le test de convergence fait correspondre au champ des vitesses radiales $V_r(\mathbf{x})$ défini en tout point \mathbf{x} de l'espace, la quantité scalaire $G_+(\mathbf{y})$ par l'équation suivante :

$$G_+(\mathbf{y}) = \int_V V_r(\mathbf{x}) H(\mathbf{x}, \mathbf{y}) d^3\mathbf{x} \quad (D.15)$$

$$\text{avec } H(\mathbf{x}, \mathbf{y}) = - \frac{\mathbf{y} - \mathbf{x}}{\|\mathbf{y} - \mathbf{x}\|} \cdot \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad (D.16)$$

Nous introduisons ici la famille de fonctions vectorielles $\mathbf{G}_y(\mathbf{x})$ indicée par la variable \mathbf{y} et que l'on exprime à partir de la fonction vectorielle de base $\mathbf{g}(\mathbf{r}) = -\mathbf{r}/r$ de la façon suivante : $\mathbf{G}_y(\mathbf{x}) = \mathbf{g}(\mathbf{x} - \mathbf{y})$. Les $\mathbf{G}_y(\mathbf{x})$ sont en fait la famille des fonctions convergentes vers un centre \mathbf{y} et de profil constant.

Si le champ des vitesses $\mathbf{V}(\mathbf{x})$ de notre échantillon est irrotationnel (soit qu'il dérive d'un potentiel) et que nous travaillons sur un volume limité V de l'espace afin d'éviter les problèmes aux infinis, on peut décomposer le champ des vitesses sur la famille de fonctions $\mathbf{G}_y(\mathbf{x})$:

$$\mathbf{V}(\mathbf{x}) = \int_V A(\mathbf{y}) \mathbf{G}_y(\mathbf{x}) d^3\mathbf{y} \quad (D.17)$$

La fonction scalaire $A(\mathbf{y})$ doit être vue ici comme les coefficients de la décomposition du champ des vitesses sur la famille des fonctions convergentes vers \mathbf{y} et de profil constant.

Cette expression devient pour le champ des vitesses radiales :

$$V_r(\mathbf{x}) = \int_V A(\mathbf{y}) H(\mathbf{y}, \mathbf{x}) d^3\mathbf{y} \quad (D.18)$$

On remarque alors qu'il serait possible d'associer les fonctions $A(\mathbf{x})$ et $G_+(\mathbf{y})$ (définies en (D.17) et (D.15)), et donc d'établir une correspondance entre $V_r(\mathbf{x})$ et $G_+(\mathbf{x})$ si l'équation suivante était vérifiée :

$$K(\mathbf{x}, \mathbf{y}) = \int_V H(\mathbf{z}, \mathbf{x}) H(\mathbf{y}, \mathbf{z}) d^3\mathbf{z} = \delta(\mathbf{y} - \mathbf{x}) \quad (D.19)$$

avec δ la distribution de Dirac

Mais cette équation (D.19) n'est pas vérifiée. En fait, la famille de fonctions $G_{\mathbf{y}}(\mathbf{x})$ ne satisfait pas la condition nécessaire stipulant que la famille sur laquelle on décompose le champ des vitesses doit être une base de l'espace des fonctions vectorielles à 3 dimensions. Le test dipolaire peut néanmoins être interprété comme une décomposition "imparfaite" du champ des vitesses sur la famille des fonctions convergentes vers un centre et de profil constant.

La recherche de familles de fonctions vérifiant l'équation (D.19) nous apparaît par contre prioritaire. De telles bases de fonctions permettent en effet d'évaluer toutes sortes de quantités intéressantes (comme le divergence ou le potentiel du champ des vitesses) en correspondance avec le champ des vitesses radiales. La transformée en ondelettes s'avère un outil idéal si l'on veut construire des bases de fonctions possédant ces propriétés. Nous exposons dans la partie E les avantages de l'analyse par ondelettes du champ des vitesses.

REFERENCES :

- [1] : M. Aaronson, J. Huchra, J. Mould, B. Tully, J.R. Fisher, H. Van Woerden, W.M. Gosse, P. Chamaraux, V. Mebold, B. Siegman, G. Berriman and S.E. Persson : "A catalog of infrared magnitudes and HI velocity widths for nearby galaxies", *Astrophys.J Suppl. Series* **50**,241-262 (1982a)
- [2] : M. Aaronson, J. Huchra, J. Mould, P. Schechter and B. Tully : "The velocity field in the local supercluster", *Astrophys.J* **258**,64-76 (1982b)
- [3] : E. Bertschinger and R. Juskiewicz : "Searching for the Great Attractor", *Ap.J Letters* **334**,L59-L62 (1988)
- [4] : E. Bertschinger, K.M. Górski and A. Dekel : "Effect of the Great Attractor on the cosmic microwave background radiation", *Nature (Letters)* **345**,507-509 (1990)
- [5] : E. Bertschinger, A. Dekel, S.M. Faber, A. Dressler and D. Burstein : "Potential, velocity, and density fields from redshift-distances samples. Applications : Cosmography within 6000 kilometers per second", *Astrophys.J* **364**,370-391 (1990)
- [6] : G.D. Bothun, R.A. Schommer, T.B. Williams, J.R. Mould and J.P. Huchra : "Peculiar velocities of field spiral galaxies near and beyond the Great Attractor", *Astrophys.J* **388**,253-267 (1992)
- [7] : D. Burstein, R.L. Davies, A. Dressler, S.M. Faber, D.Lynden-Bell, R.J. Terlevich and G. Wegner : "Large-scale motions in the nearby Universe" in "*Large-scale structure and motions in the Universe*", M. Mezzetti et al. (eds.) 179-196 (1989)

- [8] : D. Burstein, S.M. Faber and A. Dressler : "Evidence from the motions of galaxies for a large-scale, large-amplitude flow toward the Great Attractor", *Astrophys.J* **354**,18-32 (1990)
- [9] : A. Dekel, E. Bertschinger and S.M. Faber : "Potential, velocity, and density fields from sparse and noisy redshift-distances samples : Method ", *Astrophys.J* **364**,349-369 (1990)
- [10] : A. Dressler, S.M. Faber, D. Burstein, R.L. Davies, D.Lynden-Bell, R.J. Terlevich and G. Wegner : "Spectroscopy and photometry of elliptical galaxies : A large-scale streaming motion in the local Universe", *Ap.J. Letters* **313**,L37-L42 (1987)
- [11] : A. Dressler and S.M. Faber : "Confirmation of a large-scale, large-amplitude flow in the direction of the Great Attractor", *Astrophys.J* **354**,13-17 (1990)
- [12] : A. Dressler : "The Great Attractor: do galaxies trace the large-scale mass distribution ?", *Nature* **350**,391-397 (1991)
- [13] : K.M. Górski : "Will *COBE* challenge the inflationary paradigm ? Cosmic microwave background anisotropies versus large-scale streaming motions revisited", *Ap.J Letters* **370**,L5-L9 (1991)
- [14] : E.J. Groth, R. Juszkiewicz and J.P. Ostriker : "An estimate of the velocity correlation tensor : Cosmological implications", *Astrophys.J* **346**,558-565 (1989)
- [15] : J.E. Gunn : "Deviation from pure Hubble flow : a review", in "The extragalactic distance scale", A.S.P. Conference Series, Volume 4, p. 344-359 (1988)
- [16] : M. Han and J.R. Mould : "Peculiar velocities of clusters in the Perseus-Pisces supercluster", *Astrophys.J* **396**,453-459 (1992)
- [17] : A.F. Heavens : "Great Attractors and the value of Ω ", *Mon.Not.R.astr.Soc.* (1991) **251**, 267-280

- [18] : R. Juszkiewicz, K. Górski and J. Silk : "Minimal cosmic background fluctuations implied by streaming motions", *Ap.J Letters* **323**, L1-L6 (1987)
- [19] : N. Kaiser : "Bulk flows, shear, and the Great Attractor", *Astrophys.J* **366**,288-392 (1991)
- [20] : B. Lilje, A. Yahil and B.J.T. Jones : "The tidal velocity field in the local supercluster", *Astrophys.J* **307**,91-96 (1986)
- [21] : D. Lynden-Bell, A. Dressler, D. Burstein, R.L. Davies, S.M. Faber, R.J. Terlevich and G. Wegner : "Spectroscopy and photometry of elliptical galaxies. V. Galaxy streaming toward the new supergalactic center", *Astrophys.J* **326**,19-49 (1988a)
- [22] : D. Lynden-Bell, D. Burstein, R.L. Davies, A. Dressler, S.M. Faber, R.J. Terlevich and G. Wegner : "On the best distance estimators and galaxy streaming", in "The extragalactic distance scale", *A.S.P. Conference Series*, Volume 4, p. 307-316 (1988b)
- [23] : D. Lynden-Bell, O. Lahav and D. Burstein : "Cosmological deductions from the alignment of local gravity and motion", *Mon.Not.R.astr.Soc.* (1989) **241**, 325-345
- [24] : E. Martínez-González and J.L. Sanz : "Large-scale streaming motions and microwave background anisotropies", *Astrophys.J* **347**,11-15 (1989)
- [25] : D.S Mathewson, V.L. Ford and M. Buchhorn : "A southern sky survey of the peculiar velocities of 1355 spiral galaxies", *Ap.J Suppl. Series* **81** 413-659 (1992a)
- [26] : D.S Mathewson, V.L. Ford and M. Buchhorn : "No back-side infall into the Great Attractor", *Ap.J Letters* **389**,L5-L8 (1992b)
- [27] : B. Paczyński and T. Piran : "A dipole moment of the microwave background as a cosmological effect", *Astrophys.J* **364**,341-348 (1990)
- [28] : M. Rowan-Robinson, A. Lawrence, W. Saunders, J. Crawford, R. Ellis, C.S. Frenk,

I. Parry, X. Xiaoyang, J. Allington-Smith, G. Efstathiou and N. Kaiser : "A sparse-sampled redshift survey of *IRAS* galaxies: I. The convergence of the *IRAS* dipole and the origin of our motion with respect to the microwave background", *Mon.Not.R. astr.Soc.* (1990) 247, 1-18

[29] : V.C. Rubin, W.K. Ford, N. Thonnard, M.S. Roberts and J.A. Graham : "Motion of the Galaxy and the Local group determined from the velocity anisotropy of distant Sc I galaxies. I. The data", *Astron. J* 81,687-718 (1976a)

[30] : V.C. Rubin, N. Thonnard and W.K. Ford : "Motion of the Galaxy and the Local group determined from the velocity anisotropy of distant Sc I galaxies. II. The analysis for the motion", *Astron. J* 81,719-737 (1976b)

[31] : R.K. Sachs and A.M. Wolfe : "Perturbations of a cosmological model and angular variations of the microwave background", *Astrophys.J* 147,73-90 (1967)

[32] : W. Saunders, M. Rowan-Robinson, A. Lawrence, J. Crawford, R. Ellis, C.S. Frenk, I. Parry, X. Xiaoyang, J. Allington-Smith, G. Efstathiou and N. Kaiser : "The density field of the local universe", *Nature* 349,32-38 (1991)

[33] : R. Scaramella, G. Baiesi-Pillastrini, G. Chincarini, G. Vettoiani and G. Zamorani : "A marked concentration of galaxy clusters: is this origin of large-scale motions ?", *Nature (Letters)* 338,562-564 (1989)

[34] : E. Shaya, R.B. Tully and M.J. Pierce : "Nearby galaxy flows modeled by the light distribution", *Astrophys.J* 391,16-33 (1992)

[35] : J. Silk : "Is cosmic drift a cosmic myth ?", *Ap.J Letters* 345,L1-L4 (1989)

[36] : G.F. Smoot and P.M. Lubin : "Southern anisotropy in the cosmic microwave background radiation", *Ap.J Letters* 324,L83-L86 (1979)

[37] : G.F. Smoot, C.L. Bennett, A. Kogut, E.L. Wright, J. Aymon, N.W. Boggess, E.S.

IV. ANNEXE :

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

Detecting non-Hubble velocity fields in the universe

Stéphane Rauzy, Marc Lachièze-Rey, and R.N. Henriksen*

Service d'Astrophysique, C.E.N. Saclay, F-91191 Gif sur Yvette Cedex, France

Received July 16, accepted October 29, 1991

Abstract. In this paper we present several simple geometrical methods for analyzing the Cosmic velocity field using a sample of catalogued galaxies. These methods combined have the merit of allowing us to distinguish an attractor velocity field from a velocity field that does not converge within the confines of the sample, that is which is effectively 'constant'. We do not model the velocity field but rather work directly with the data so that our tests are parameter free. The key idea is to transform the search for a 'great attractor' into a search for the associated geometric quantity, the Zero Radial Velocity Sphere. We have tested our methods on simulations, and we have applied them to the elliptical galaxy catalogue of Faber *et al.* (1989). We conclude that a dominant attractor region is present, largely extended around a center at galactic coordinates $[280^\circ, 10^\circ]$ and at a distance of 3800 km sec^{-1} . Moreover, we conclude that the velocity field contains also a constant component in a direction different from that of the attractor region.

Key words: Cosmology - Galaxies:elliptical - Reference systems

1. Introduction

It is by now apparent that much information concerning the distribution of mass in the universe can be extracted from the velocity structure that remains after subtracting a mean Hubble flow from the observed radial velocities. Indeed the earliest results (Rubin *et al.* 1976a, 1976b) appeared some 15 years ago and sought to establish an 'absolute' streaming motion of the local group (lg) with respect to a distant shell of galaxies. At present an absolute frame of reference is generally assumed to be provided by the last scattering surface which generated the cosmic microwave background (cmb), so that the observed dipolar anisotropy of the cmb gives (after correction for the earth's motion in the lg) our absolute proper motion.

Now as Gunn (1988) has observed, this assumption plays an important role in the recent work (Lynden-Bell *et al.* 1988; Burstein *et al.* 1990; Dressler and Faber, 1990) interpreting the large scale coherent velocity structure as a convergent, spherically symmetric motion towards a localized region of space, the Great Attractor (GA). For the velocity, at the earth, due to this inferred component of the velocity field is roughly equal and opposite to our 'absolute' motion as deduced from the cmb anisotropy. Thus

if this anisotropy were to be due to other causes not related to the velocity field, one would have created a large scale convergent field on converting to the 'absolute' frame. Although such a bulk flow differs from a locally convergent flow, there is a danger of confusion and it has not always been clear how efficient the previous analyses were in detecting such a situation. Since other interpretations of the cmb and its dipole anisotropy have been in fact proposed (Paczinsky and Piran 1990), and since this hypothesis has also been questioned recently on observational grounds (Matthewson, Ford, and Buchhorn, 1991); it is worth trying to find a clear evidence of the presence of the GA field without correcting to the absolute frame of reference, that is independently of the presumed cosmological origin of the cmb. If this can be accomplished and the same conclusions are reached, this would be an additional argument favouring the usual interpretation of the cmb, and so indirectly, big bang cosmology.

A related complication arises if one wishes to demonstrate the existence of a GA region which may be in motion relative to the absolute frame. For it is likely that such a motion would carry the galaxies subject to the GA with it, and such a coherent motion can render the GA field much less evident. *One of the principal results of this paper is to provide evidence of just such an effect.*

Thus we attempt to establish the existence of GA velocity fields in a very simple way based directly on the observed redshifts as far as possible, and employing the minimum number of physical assumptions and free parameters. The resulting procedure is the fundamental content of this paper. Moreover since there is a prejudice (founded securely on the notion of the formation of structure by gravitational instability) that large scale velocity structure should be purely divergent (i.e. curl-free) as it remains in the linear regime, we concentrate here on the detection of such large scale, coherent, divergent flows.

2. Decomposition of the velocity field

The field of proper velocities is obtained from measured redshifts (reduced to the frame of the microwave background) from which an estimation of the Hubble flow has been subtracted. A bad estimation of the Hubble constant (which may occur if the calibration were made on a catalogue itself dominated by large scale flows) or a bias in the estimation of distances (for instance of the Malmquist type) would create errors which may be seen as an undesirable radial field, whose intensity varies almost only with distance (non-uniform sampling complicates this simple picture). We will refer to such fields as residual Hubble type fields

Send offprint requests to: Marc Lachièze-Rey

* Dept. of Physics, Queen's University, Kingston, Ontario, Canada K7L 3N6

V_H (even if the radial dependance is non linear). A method independent of the presence of such residual Hubble flows would allow us to work directly with the redshifts without introducing systematic error, and would give results independently of the presence of Malmquist biases. We provide such a method below.

Now let us call V_c the constant velocity flow to which the sample under study may be subjected. Its value is evidently not invariant under Galilean velocity transformations. This is why calculations are generally performed in the cmb frame, which is assumed to be an absolute frame so that V_c is zero. This would not be true as noted above for the galaxies under the influence of a great attractor which has itself a constant velocity in the absolute frame. To dispense with the hypothesis of the cmb determined absolute frame, or to allow for the uniform motion of an attractor's region we must then allow a non-zero V_c .

Furthermore, assuming that there is no large scale vorticity, we will decompose the remaining velocity field into a contribution from the hypothetical large scale attractor flow, V_A , plus a homogeneous isotropic component, V_{HI} . Such a decomposition makes sense only after some appropriate intermediate scale in the data has been chosen, for which the spatial average of V_{HI} reduces to zero.

Thus we write the decomposition of the velocity field as

$$V = V_H + V_c + V_A + V_{HI}. \quad (1)$$

We introduce in the next section the essential geometrical idea of the *Zero Radial Velocity Sphere* (ZRVS) and we propose a series of tests based on it to determine different components of the velocity field. They have all proved to be promising when applied to numerical simulations and to a real catalogue of Galaxies. We are convinced that they are the minimum treatment of the data by which one can reveal a great attractor's flow. We are thus able to present some new results in this paper. However we can not of course address the validity of the field of proper velocities itself, which may contain large and systematic errors if the proper distances suffer from biases. Moreover such catalogues always suffer from irregular sampling.

3. The zero radial velocity sphere

A ZRVS is associated with each GA-flow, so that the search for a GA-region can be identified to a certain extent with the search for its associated ZRVS. The ZRVS associated with a given GA-region centered at $r_A \equiv (r_A, \theta_A, \phi_A) \equiv G$ is defined as the spherical surface passing through our position O and that of the center of the GA-region G , in such a way that $OG \equiv r_A$ is a diameter. Any galaxy lying on this surface has the property that the radial (as seen by observer O) component of its velocity induced by the attractor G is exactly zero, provided that we assume that velocities remain parallel to accelerations (that is that the flow is perfectly divergent). This was already pointed out by Lynden-bell et al. (1988) although the idea was not extensively employed in their paper. It is moreover also evident from the same geometry that the GA's velocity field contributes positively to the radial velocity of a galaxy that is inside the ZRVS, while it contributes negatively when the galaxy lies outside (figure 1). These observations depend neither on the spatial extension of the GA (provided it is distant) nor on the shape of the radial velocity profile that it induces. These simple remarks have motivated the following analyses of the cosmic radial velocity field.

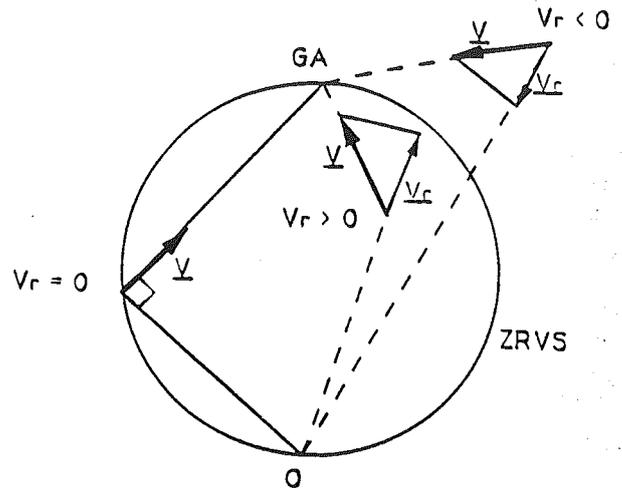


Fig. 1. The drawing shows the contribution to the observed radial velocity of the velocity field due to a point attractor: outside, inside, and on, the ZRVS associated with this attractor and this observer

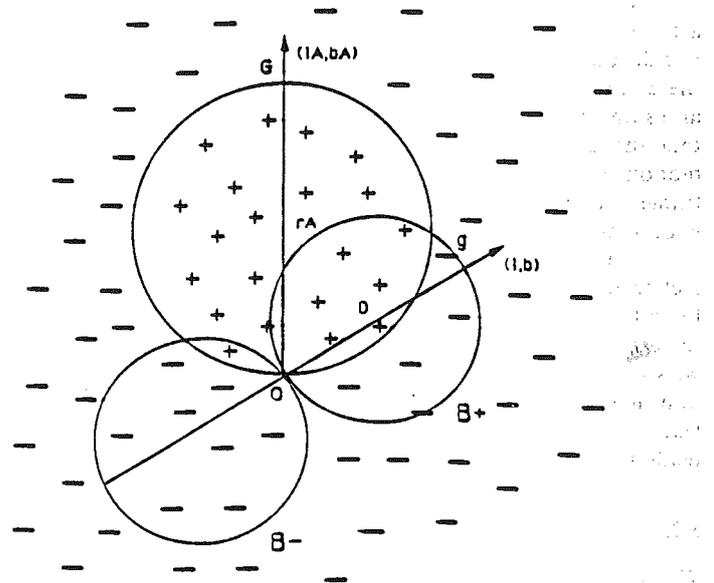


Fig. 2. The drawing emphasizes the sign relations implicit in figure 1, and indicates the search procedure using the balls B_+ and B_- . The largest number of positive contributions to the radial velocity field lies within the ZRVS

3.1. Dipolar Test using Spheres (DTS)

We refer to the sketch of figure 2. We introduce two balls B_+ and B_- of equal diameter D that are tangent at the position O of the observer. The common direction, d , of the diameters through O , as well as the common value D of the diameter are freely variable. For a velocity field V we define the two linear functions

$$F_{\pm}(d, D; V) \equiv \int_{B_+} V \cdot \frac{r}{r} dV \pm \int_{B_-} V \cdot \frac{r}{r} dV. \quad (2)$$

and we refer to F_+ and F_- as the dipolar and quadrupolar functions respectively. In this subsection we concentrate on F_+ .

For a homogeneous and isotropic velocity field, the radial velocity component will be a random variable so that the integrals in the 2 balls take a zero value. Hence one sees readily that

$$F_{\pm}(\mathbf{d}, D; \mathbf{V}_{HI}) \equiv 0. \quad (3)$$

over sufficiently large volumes.

Moreover one readily sees that for F_{\pm} , one has the important condition that

$$F_{\pm}(\mathbf{d}, D; \mathbf{V}_H) \equiv 0. \quad (4)$$

since the contributions from the two spheres cancel exactly equal for a Hubble-type flow. Thus F_{\pm} is insensitive to the presence of an artificial velocity field created for example by a Malmquist Bias. Consequently we see that

$$F_{\pm}(\mathbf{d}, D; \mathbf{V}) = F_{\pm}(\mathbf{d}, D; \mathbf{V}_c) + F_{\pm}(\mathbf{d}, D; \mathbf{V}_A). \quad (5)$$

Finally it is clear from figure (2) that F_{\pm} will adopt its extremal values (maximum for \mathbf{d} parallel \mathbf{r}_A and minimum when these vectors are anti-parallel) when $\mathbf{D} \equiv D\mathbf{d} = \mathbf{r}_A$. A similar result is evident for the constant component \mathbf{V}_c when \mathbf{d} is parallel to \mathbf{V}_c .

Thus the dipolar test using spheres (DTS) will reveal a GA in a frame in which no constant streaming motion is present. The test then consists in evaluating F_{\pm} as a function of \mathbf{d} , and D until the extrema are obtained. The values of \mathbf{d} , and D then provide an estimate of \mathbf{r}_A . This test requires us to adopt the hypothesis that the anisotropy in the cmb is due to our own streaming motion in the Universe so that we can correct to the absolute frame of reference. As such it can not alone be used to check this hypothesis.

Should \mathbf{V}_c be really not zero due either to errors in the correction to the absolute frame, or to the incorrectness of the basic hypothesis, or to some convoy motion of the galaxies subject to the attractor or in general to a combination of these cases; then we see from equation (5) that F_{\pm} should still converge to a direction that lies in the plane defined by \mathbf{V}_c and \mathbf{V}_A . Complementary tests could then allow us to discriminate between some types of fields \mathbf{V}_c and \mathbf{V}_A .

3.2. The Quadrupolar Test using Spheres (QTS)

This test uses the function F_{\pm} defined in equation (2). It is easy to see that as in the previous test

$$F_{\pm}(\mathbf{d}, D; \mathbf{V}_{HI}) = 0. \quad (6)$$

but that now in addition

$$F_{\pm}(\mathbf{d}, D; \mathbf{V}_c) = 0. \quad (7)$$

Moreover it is also clear that although $F_{\pm}(\mathbf{d}, D; \mathbf{V}_H) = 0$, it will vary only as a function of D but *not* with the direction \mathbf{d} .

Thus the key property of F_{\pm} may be summarized as

$$F_{\pm}(\mathbf{d}, D; \mathbf{V}) = F_{\pm}(\mathbf{d}, D; \mathbf{V}_A) - F_{\pm}(\mathbf{d}, D; \mathbf{V}_H), \quad (8)$$

so that we have found a measure on the data that is sensitive to the attractor field but independent of a constant component in this field. Moreover any field \mathbf{V}_H will not interfere with the determination of the direction to the center of the GA-region provided it remains a function of r only, as remarked already above. Of course to the extent that this field is small, then the QTS (which is the process of finding the extremal values of \mathbf{d} , and D for F_{\pm}) will

also provide a good measure of the distance to the center of the GA. In fact it is this test which proves to be the more reliable in the present case.

3.3. A procedure to characterize attractor flows

A procedure valid in principle is to use the DTS and the QTS in a complementary fashion. That is, the invariance properties of the QTS allow us to use it directly to find the direction to the center of any GA-region, independent of any assumed absolute frame. Subsequently, the application of the DTS as a function of D along this and nearby directions until the extremum is located for that direction will give an estimate for the distance D even though there will not in general be an extremum in angle simultaneously.

However if the catalogue may be reasonably supposed to be free of a residual Hubble flow, then QTS can also be used for the distance determination, as we have in fact done below. For in practice one finds that DTS is not a reliable distance indicator on the present sample because of its sensitivity to the sample inhomogeneity, and that QTS is in this instance the more reliable for both purposes.

We should also remark here that because of the angular dependence of the DTS on \mathbf{V}_c (see equation (5)) the difference in the angles to which the QTS and the DTS separately converge is an indication of the presence of a component \mathbf{V}_c in the catalogue. In this sense also the DTS is a complementary test to the QTS, and it gives a very striking monotonic signal for a flow that does not converge within the limits of the sample (i.e. \mathbf{V}_c).

Although neither DTS nor QTS uses all of the data simultaneously, it should be noted that in fact no information is wasted since the complete procedure requires displacing the balls over the whole catalogue. In fact our results on both the real catalogue and on the simulations show that the tests are sufficiently sensitive.

3.4. Corrections due to irregular measures on the sampled volumes

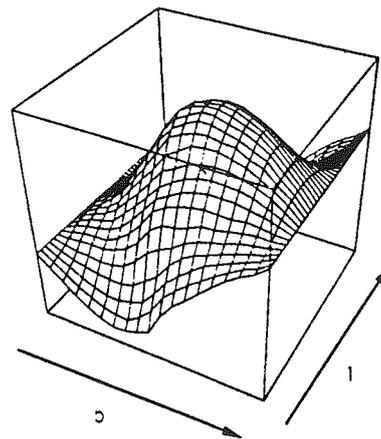
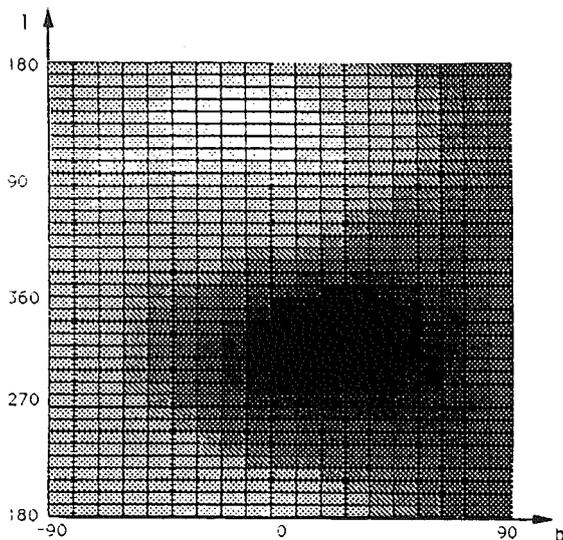
We have checked with simulations that the tests work as desired when the radial velocity field is well sampled. But existing galaxy catalogues provide us with a highly irregular sampling of the cosmic velocity field rather than one which is homogeneous and isotropic. It is thus necessary to introduce some appropriate normalization in order to restore the symmetry properties of the integrals. For example, a regular measure would imply an equal number of points in the two balls B_{\pm} and B_{\mp} . Any deviation from this equality, which may occur because of edge effects and irregular sampling, will distort the tests. Thus estimators for the functions used in DTS and QTS are given by

$$F_{\pm} \equiv \left(\frac{1}{N_{\pm}} \right) \sum_{i=1}^{N_{\pm}} \frac{\mathbf{r}_i \cdot \mathbf{V}_i}{r_i} \pm \left(\frac{1}{N_{\mp}} \right) \sum_{j=1}^{N_{\mp}} \frac{\mathbf{r}_j \cdot \mathbf{V}_j}{r_j}, \quad (9)$$

where N_{\pm} is the number of galaxies in the ball B_{\pm} .

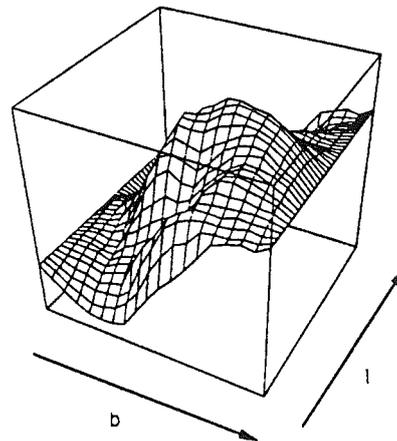
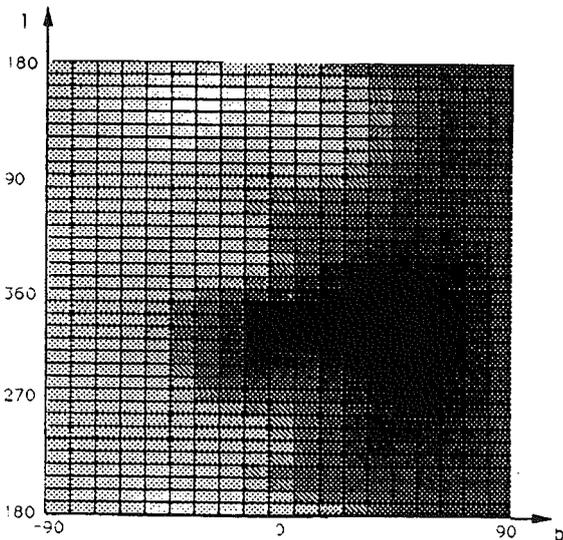
Unfortunately, simulations have shown that this normalization does not rectify completely the poor estimate of F_{\pm} , due to the uneven sampling. Thus the zero values for F_{\pm} on the appropriate components will only be approximate.

We have therefore also used a more subtle weighting scheme based on the following remark. In each direction \mathbf{d} the ball $B_{\pm}(D)$ may be considered as the ensemble of crescents



DTS on NECGA, $r_A=4000$ km/sec
Maximum at $l=320$, $b=20$

Fig. 3. The result of the search ($320^\circ, 20^\circ$) in angle for the GA using DTS at an assumed distance of 4000 km sec^{-1} on the simulated catalogue. The correct direction to the GA is seen to coincide with the peak in both representations



DTS on NEC, $r_A=4000$ km/sec
Maximum at $l=330$, $b=30$

Fig. 4. The result ($330^\circ, 30^\circ$) of the search in angle using DTS on the real catalogue at a distance of 4000 km sec^{-1} . The direction is again well defined

$C \equiv (d B_{\perp} / d D) \Delta D$ encompassing the smallest ball in the family. Here ΔD is the bin size taken for the quantity D . Thus one can accomplish a weighting by crescent in each of our tests by summing the contributions from the galaxies inside each crescent, and then multiplying by the average volume per galaxy W calculated for each crescent. Finally one adds all of the crescents weighted in this way comprising each ball. Thus explicitly the estimators for DTS and QTS become

$$F_{\pm} = \sum_{C \in B_{\pm}} W_i \sum_{C, j} \frac{r_j \cdot V_j}{r_j} \pm \sum_{C \in B_{\pm}} W_i \sum_{C, j} \frac{r_j \cdot V_j}{r_j} \quad (10)$$

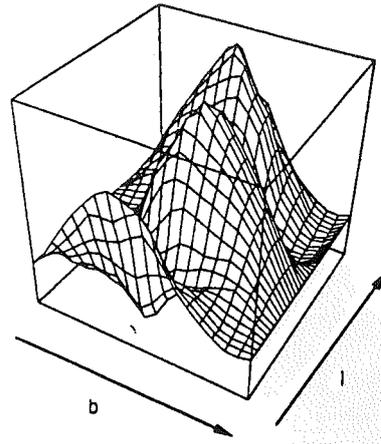
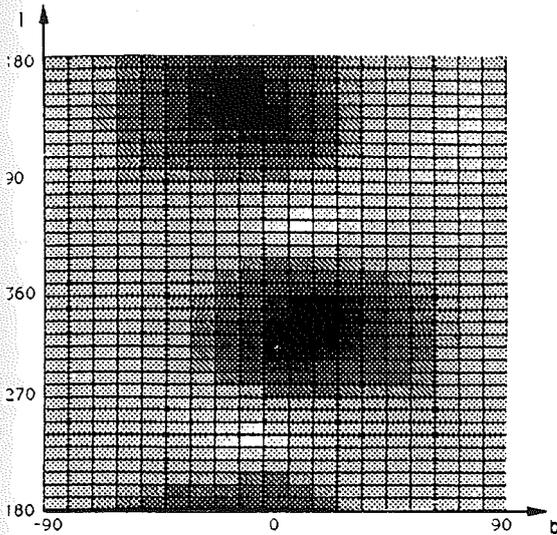
This weighting scheme gives equal weight to the B_{\perp} and B_{\parallel} balls, and thus tends to render the determination of the GA-center's distance independent of the velocity profile of the hypothetical GA-flow, just as would be the case for the ideal test on a set of perfect measure.

4. Results

We have used the various tests described above to analyse both simulated catalogues and a real sample.

4.1. The data

We work on the New Elliptical Galaxies catalogue of Faber et al. (1989). Predicted distances are corrected for the Malmquist bias in the same way as in Lynden-Bell et al. (1988). From these data we have constructed a sub-catalogue (NEC) by removing redundant information concerning galaxies which lie in clusters: all galaxies belonging to the same clusters are taken as only one data point. NEC contains finally galactic coordinates (l, b), distance r (in km sec^{-1}) and radial peculiar velocities, for 180 galaxies (in fact, data points).



QTS on NECGA, $r_A=4000 \text{ km/sec}$
Maximum at $l=320, b=10$ ($l=130, b=-10$)

Fig. 5a and b. The result ($320^\circ, 10^\circ$) of the search in angle on the simulation using QTS at a distance of 4000 km sec^{-1} . There is a small error in angle due presumably to irregular sampling

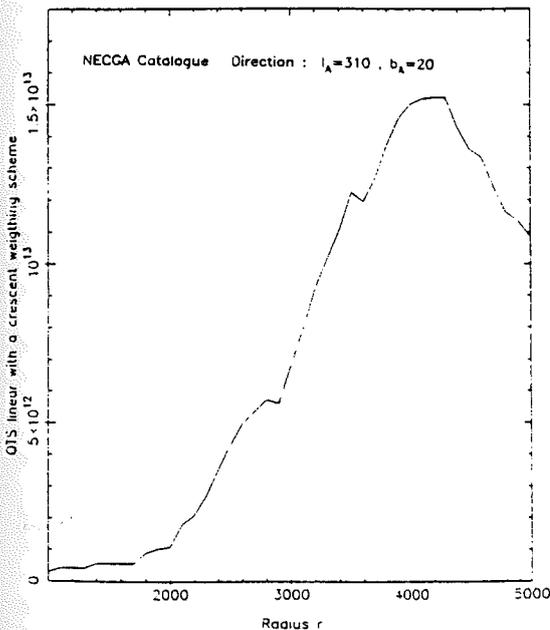


Fig. 5c. The result of the same search in distance with a crescent weighting scheme

Similarly, we have simulated a catalogue (NECGA) which has the same set of galaxy spatial coordinates as NEC, but where the velocity field is synthesized according to a great attractor model (we have chosen the one of Bertschinger et al., 1988) wherein $r_A = 4200 \text{ km sec}^{-1}$, $l_A = 310^\circ$, and $b_A = 20^\circ$. In fact the simulated attractor is somewhat extended. Thus, since the velocity field is perfectly known by construction, and since the spatial distribution is the same as that in the real catalogue, we are able to evaluate the effectiveness of our various tests under realistic conditions.

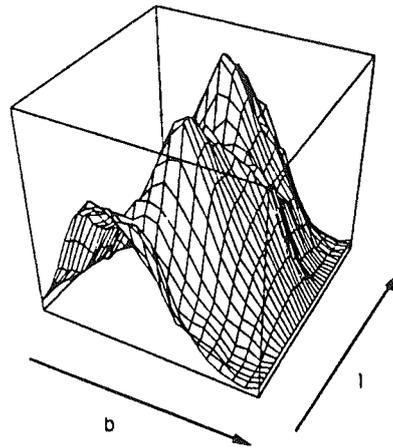
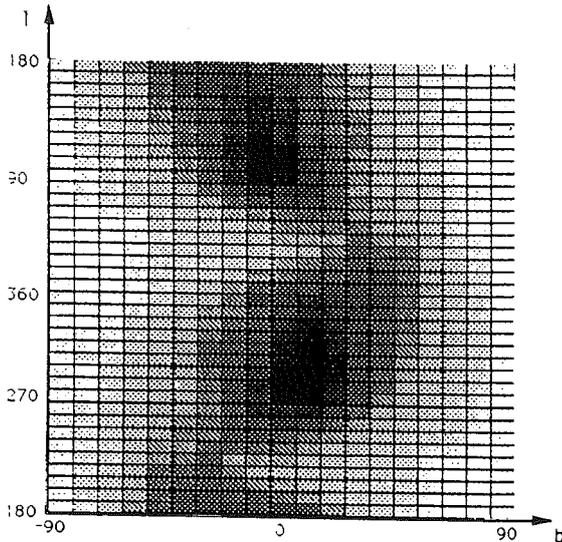
It should also be remarked here that we have repeated our tests on the real galaxy catalogue after having removed the fa-

mous 11 galaxies which have been claimed (Lucey et al., 1988; Gorski et al., 1989) to represent an undue weight in previous determinations of the GA coordinates. This did not alter the results as reported below.

4.2. The analysis

In this section we use three different kinds of graphical representation. To illustrate the search in direction, we display a two dimensional surface which gives each test's value as a function of direction (l, b), for a given distance r . A grid measuring 18×36 is used to cover the range $(-90^\circ, -90^\circ)$ in b and $(-180^\circ, -180^\circ)$ in l with each grid element covering the range $10^\circ \times 10^\circ$. In addition, we choose occasionally to display the same information using a more visually striking contrast luminosity plot. Here low values of the test are white and high values are dark, and the grid is the same as described in the previous paragraph. Finally the search in distance for an appropriate test is shown in a two dimensional plot showing the value of the test versus the distance r , for a given direction (l, b).

We begin with a discussion of the DTS-QTS (properly weighted) sequence as outlined above. Discussing first the results of the tests on the simulation, we see that both DTS (figs 3a, 3b) and QTS (figs 5a, 5b, 5c) succeed in locating properly the synthetic GA flow, with an angular precision of about 10° , and to within 200 km sec^{-1} . We attribute the limited (although rather good) precision to the fact that bad sampling and edge effects are not exactly compensated by the normalization. These discrepancies between the real location and that found by the tests may be used as an estimate of our intrinsic errors, since the sampling of the simulation is the same as for the real catalogue by construction. It is not impossible however that this error is systematic for a given attractor type and a given sampling of the velocity field. In this case we should correct the direction using an offset determined from the simulations. This will require further extensive investigation to be quantified however, and we proceed here by treating the error as random.



QTS on NEC, $r_A=4000 \text{ km/sec}$
Maximum at $l=280, b=10$ ($l=100, b=-10$)

Fig. 6a and b. The result ($280^\circ, 10^\circ$) of the search in angle on the real catalogue using QTS at a distance of 4000 km sec^{-1}

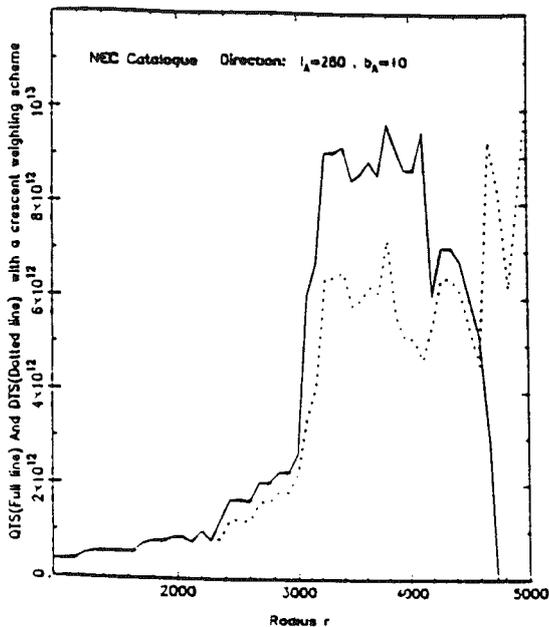


Fig. 6c. The search in distance with QTS (solid line) and with DTS (dotted line). The direction is well defined by QTS but the distance indicates an extended GA

For the real velocity field, the QTS (figs 6a,6b) leads to the conclusion that an GA-flow is indeed present, towards the direction ($280^\circ, 10^\circ$). We emphasize that it has not been assumed that the cmb provides an absolute frame and that, due to the nature of QTS *this convergence cannot be the result of a constant velocity field.*

On the other hand, a search with the DTS (figs 4a,4b) converges towards a different direction ($330^\circ, 30^\circ$). Since the disagreement between the two tests (about 50°) appears somewhat larger than the errors estimated from the simulations (about 15°), we interpret it as real and we must conclude that if, as anticipated, the QTS indicated the distance of the center of a real attractor, then *the DTS result indicates the net effect of an attractor field*

combined with a constant velocity field. This conclusion is reinforced by figure (6c) which shows the results of the searches in distance along the direction ($280^\circ, 10^\circ$) in the real catalogue using both DTS and QTS. The DTS continues to increase to the limits of the catalogue after the QTS has shown the expected decline. *This behaviour of DTS is very characteristic in the presence of a velocity field that does not converge within the limits of the sample.*

This is true even though we find that DTS does not work particularly well even on the simulation when used along the known direction. For there the discrepancy is due to poor sampling in the catalogue, and although it leads to an overestimate in the distance for the present sampling a *decline is seen within the limits of the catalogue.*

We note that the "standard" direction usually invoked for the GA center ($[309^\circ, 18^\circ]$; Dressler and Faber 1990; Faber and Burstein 1988) is very near the direction to which our DTS converges. This DTS direction moreover lies between the standard direction and the direction found from the QTS. This is compatible with the hypothesis that the standard effect is due to the combined influence of an attractor plus a constant velocity field (Lynden-Bell et al. 1989). These latter authors deduce from the galactic light distribution that only a part of the local systematic motion is due to the GA region, and that the rest is due to regions which are closer to us. Moreover the associated convoy motion is not in the same direction as the GA convergence.

To find the distances of the attractors found by these tests we have applied the (crescent-weighted) QTS as a function of distance, in the direction ($280^\circ, 10^\circ$) resulting from the search in angle with the (crescent weighted) QTS. This latter result (Fig 6c) shows very clearly that a GA 'region' (by definition a broad peak in these zero parameter tests: a distinct peak indicates a localised GA region as in figure 5c) exists at a distance lying between 3400 and 4200 km sec^{-1} . This is rather nearer than estimates produced by other methods $\approx 4500 \text{ km sec}^{-1}$ (Dressler and Faber 1990), but is in reasonable agreement with the extended nature of the light dipole (Lynden-Bell et al. 1989). Thus our results support the idea that the light traces the mass in at least some average way (Dressler, 1991).

We would then have to decide in favour of the coordinates ($280^\circ, 10^\circ$) for the GA-center, i.e. in a direction pointing inside

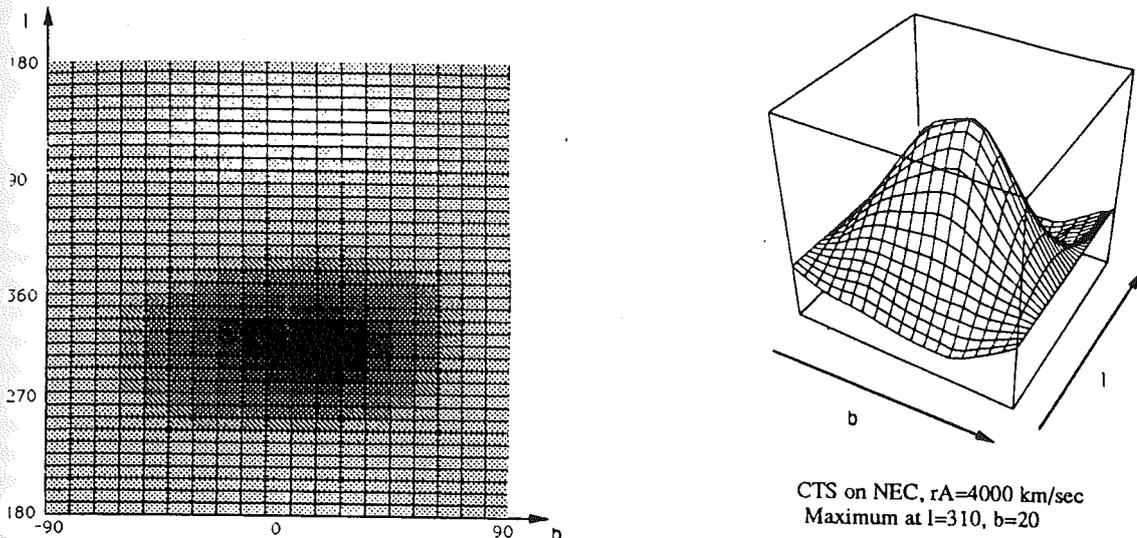


Fig. 7. The result ($310^\circ, 20^\circ$) of the search in angle on the real catalogue again at a distance of 4000 km sec^{-1} using CTS

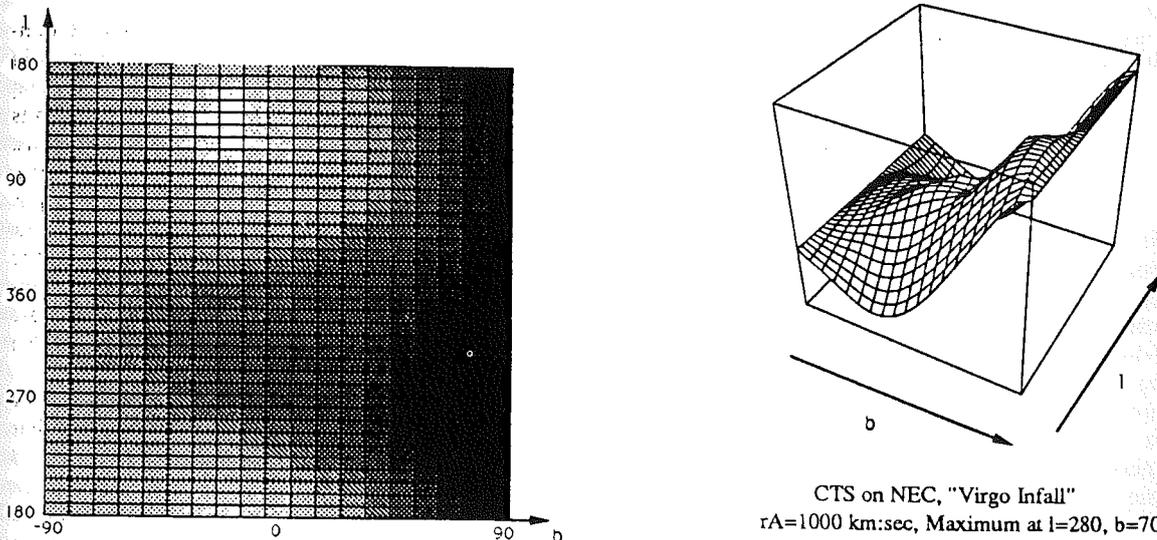


Fig. 8. The detection of the Virgo infall in the real catalogue is shown here as a result of a search in angle using CTS at a distance of 1000 km sec^{-1} . The maximum ($280^\circ, 70^\circ$) is however strongly moderated relative to the background at other angles

the galactic plane, and extended in distance between 3400 and 4200 km sec^{-1} . However based on the evidence of our own tests alone we can not exclude the possibility of a constant (i.e. not convergent at the scale of the catalog) velocity component that has not been removed from the catalogue. It would lie somewhere in the plane containing the directions ($280^\circ, 10^\circ$) and ($330^\circ, 30^\circ$).

Exploration of this hypothesis will require a more careful modelling of the data but it is important to realize how this very simple test can address this question. Such a result, if confirmed, could call into question the reduction to the absolute reference frame using the conventional interpretation of the cmb, and thus it must be checked on more elaborate data sets.

5. The virgocentric flow

One obvious question to put to these tests is whether they can detect reliably the well known nearer attractor convergences such

as the Virgo infall. The direct answer is that neither DTS nor QTS succeed in the present form, fundamentally because the catalogue is not rich enough out to these distances for the tests to be sensitive. Thus we found it necessary for these relatively nearby searches to invent a method which makes use of the whole catalogue for each assumed position of the attractor. This test we call CTS and we present it below.

5.1. Convergence Test using Spheres (CTS)

This is again a test that will only work well for a catalogue accurately corrected to an absolute reference frame so that Hubble, bias and constant components are absent from the velocity field. Thus once again, differences in r_A as determined using this method and the QTS procedure are an indication of the presence of either V_H or V in the field of proper velocities.

One defines

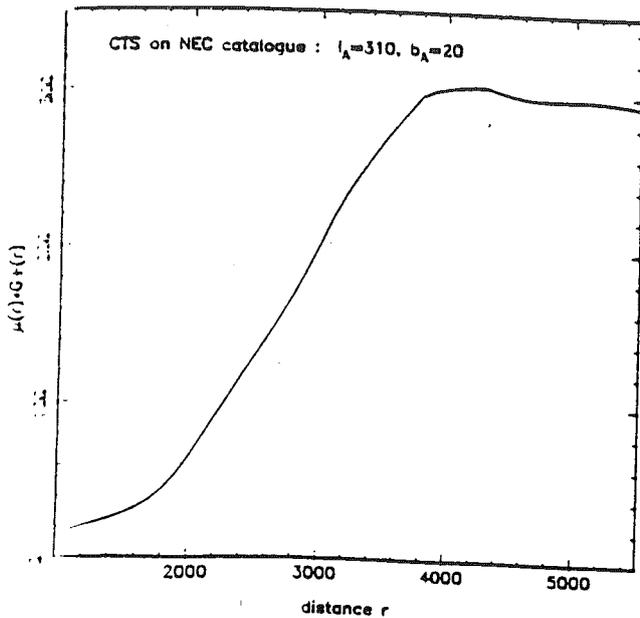


Fig. 9 The search in distance for the GA using CTS on the real catalogue in the direction found in figure 7. There is no clear indication of a point attractor, but rather of an extended region

$$G_-(\mathbf{d}, D; \mathbf{V}) \equiv \int_{B_-} -\frac{\mathbf{R} \cdot \mathbf{r}}{R \cdot r} V_r dV, \quad (11)$$

where $\mathbf{R} \equiv |\mathbf{r} - \mathbf{r}_A|$ is the radius vector to a galaxy drawn from the GA-center. The integral in equation (11) is to be extended over the entire sample. Since however for the real ZRVS associated with an hypothetical attractor, each point inside the ZRVS will contribute positively while each exterior point will contribute negatively: it is clear that the location of the algebraic maximum as a function of (\mathbf{d}, D) gives in principle an estimate of r_A .

However the cosine factor which weights each value of the radial velocity is not perhaps an obvious choice. To motivate this choice we recall that an irrotational field in the absence of boundaries can always be written as an integral over its source (the divergence) multiplied by the appropriate Greens function. When the radial component relative to a given observer is chosen this becomes

$$v_r = \frac{1}{4\pi} \int_{V'} \frac{\mathbf{R} \cdot \mathbf{r}}{R \cdot r} \frac{\nabla' \cdot \mathbf{v}'}{R^2} dV', \quad (12)$$

where in this integral $\mathbf{R} = \mathbf{r} - \mathbf{r}'$. For a point attractor at r_A the divergence is a delta function at r_A so that one obtains the trivial Newtonian result, but for an extended attractor the integral must be evaluated over all space. The Greens function multiplying the divergence gives the natural weighting for each point in space. Thus in principle we should have again used this weighting in the definition of CTS to best suit the divergent peculiar velocity field whose source we are seeking.

In practice however it was found best to use the cosine factor alone as in equation (11) and to weight with the measure

$$\mu = \sum_i \cos \theta_i, \quad (13)$$

where $\cos \theta_i = -(\mathbf{R} \cdot \mathbf{r} / r R)_i$. Thus the estimator becomes

$$G_- = \left(\frac{1}{\mu} \right) \sum_{i=1}^{N_-} -\frac{(\mathbf{r}_i \cdot \mathbf{R}_i)}{r_i R_i} V_{ri}. \quad (14)$$

The reason for taking the absolute values is to avoid biasing the location of the grand attractor towards the middle of the sample. For it is clear (see e.g. figure 2) that assuming a nearby attractor will produce mostly negative cosines in the integral, while choosing a location at the edge of the sample will produce mostly positive cosines. A middle distance will tend in a uniform sample to produce comparable numbers of positive and negative values. Thus a weighting by the inverse algebraic sum would bias r_A towards the middle of the sample. In addition both the irregular sampling with distance and the increasing errors in the measurements with distance have required us not to give the points closest to the hypothetical attractor their full weight. Thus we have not used the inverse square weighting.

Figure 7 shows the results of using CTS on the real catalogue (the method works extremely well on the simulation and recovers both in distance and direction the simulated attractor). One sees that the peak in direction (figures 7a and 7b) is exceptionally cleanly defined at an assumed distance of 4000 km sec^{-1} to be $(310^\circ, 20^\circ)$. This is in better agreement with the DTS result than with the QTS result as one would expect, since DTS and CTS are both sensitive to bias and constant component velocity fields. Thus we find a certain consistency in all of our tests regarding the hypothesis of a large scale bulk velocity component. Figure 9 shows moreover that the distance given by this test is not well defined, failing to decline markedly in the attractor region as found by QTS as might be expected in the presence of a constant velocity component.

Finally in order to demonstrate that one can be sensitive to GA flows on smaller scales in the data, we show in figure 8 the result of applying CTS to the real data at a distance of 1000 km sec^{-1} which is appropriate for the VIRGO infall. The results of the search in direction with CTS give a peak that is rather diffuse (which may of course be real) but which is certainly consistent with $(280^\circ, 70^\circ)$ as required. Thus we certainly concur with the existence of the VIRGO infall. The correct value is consistent with the presence of a large scale velocity field since at small distances it does not dominate the convergent field. However it is evident from figure 8 that the Virgo feature is superimposed on a background that increases steadily over a range of angles.

6. Conclusions

We have presented here a very simple geometric idea (the ZRVS) which allows us to draw extremely direct conclusions concerning the presence of attractors in a field of radial velocities. We have shown with simulations that the associated tests allow us to derive the presence and the localization of an attractor region with reasonable precision *without any assumptions concerning the modelling of the associated flow, and without the usual assumptions regarding an absolute frame used in such a search.*

Our tests lead to the conclusion that the kinematics of the sample by Faber et al. (1989) is dominated by attractor regions that we are able to localize. We conclude that the velocity field is dominated by a GA, independently of any assumptions concerning the cmb. Since the velocity at the earth produced by this GA is close to that of the cmb dipole, this confirms in general terms the standard interpretation of this dipole as due to our motion relative to the absolute frame of the cmb.

However our tests converge toward a direction which differs by about 50 degrees from the standard direction and we are reasonably confident that this discrepancy is not due to a systematic Malmquist type bias where the error depends only on distance. However we can not exclude the possibility that it is due to a more complicated bias involving either the irregular clustering in the sample - an hypothesis that neither our method nor the standard analysis already applied can check - or a rotational component of the velocity field.

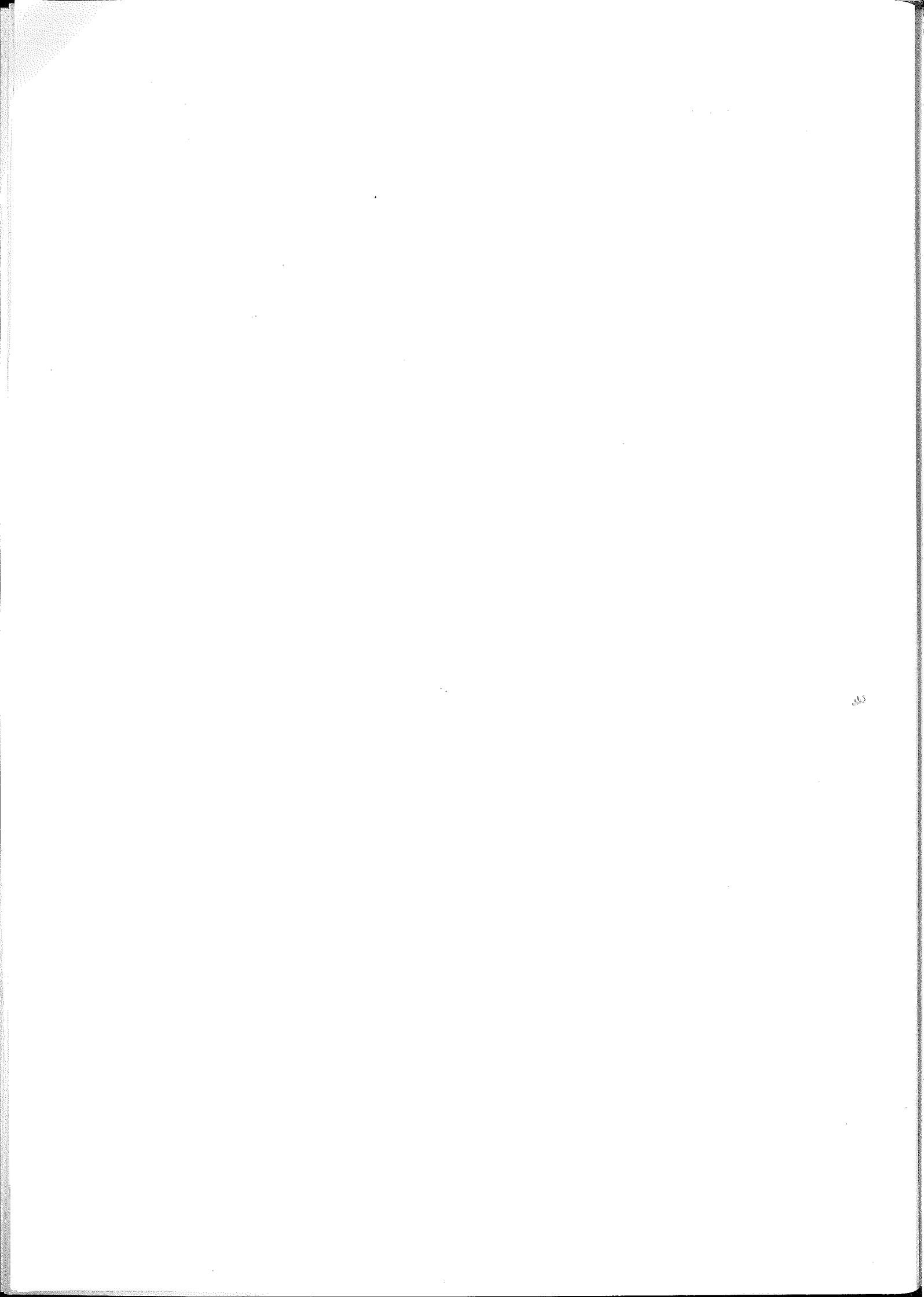
However our tests suggest another possible interpretation : that a dominant attractor is present, approximately centered at $(280^\circ, 10^\circ, 3800 \text{ km sec}^{-1})$, that is in fact an extended attractor region lying at distances between 3400 and 4200 km sec^{-1} in the same direction; but that the field is also affected by a constant (i.e. not convergent at the scale of the catalog) field in a different direction (in the plane containing the directions $(280^\circ, 10^\circ)$ and $(330^\circ, 30^\circ)$). This hypothesis needs more work on more extensive catalogues to be validated but it seems compatible with results previously published. For example Groth, Juszkiewicz and Ostriker (1989) find that much of the observed large-scale velocity correlation can be explained in terms of a constant velocity component. It is also consistent with the suggestion made recently and independently in a preprint from Mathewson, Ford and Buchhorn (1991), which follows up the work of Willick (1990). The importance of the cosmological consequences to which it might lead justifies further analyses of its validity. The simplest explanation is that the attractor's region is itself moving, and thus so is the convoy of the galaxies it dominates.

Acknowledgements. RNH wishes to acknowledge the continuing hospitality of the Service d'Astrophysique of the C.E.N. Saclay.

References

- Bertschinger E., and Juszkiewicz R., 1988, ApJ 334, L59
 Burstein D., Faber S. M., Dressler A., 1990, ApJ 354, 18
 Dressler A., 1988, ApJ 329, 519
 Dressler A., Faber S.M., 1990, ApJ 354, 13
 Dressler A., 1991, Nature 350, 391
 Faber S., Burstein D., 1988, Large-Scale Motions in the Universe ed. V.C. Rubin and G.V. Coyne, Princeton U. Press, p116
 Faber S., Wegner G., Burstein D., Davies R.L., Dressler A., Lynden- Bell D. and Terlevich R.J., 1989, ApJ, Suppl. 69 763
 Górski K., Davis M., Strauss M.A., White S.D., Yahil A., 1989, ApJ 344, 1
 Groth E.J., Juszkiewicz R., Ostriker J.P., 1989, ApJ 346, 558
 Gunn J.E., 1988, PASP Conf. Series 4, 344
 Lucey J.R., Carter D., 1988, MNRAS 235, 1177
 Lynden-Bell D., Faber S. M., Burstein D., Davies R., Dressler A., Terlevich R., Wagner G., 1988, ApJ 326, 19
 Lynden-Bell D., Lahav O., Burstein, D., 1989, MNRAS 241, 325
 Mathewson D.S., Ford V.L., Buchhorn M., 1991, ApJ Let. preprint
 Paczynski B., Piran T., 1990, ApJ 364, 341
 Rubin V.C., Ford W.K., Thonnard N., Roberts M.S., and Graham J.A., 1976a, AJ 81, 719
 Rubin V.C., Thonnard N., Ford W.K., Roberts M.S., 1976b, AJ 81, 687
 Willick J.A., 1990, ApJ 351, L5

This article was processed by the author using Springer-Verlag T_EX A&A macro package 1991.



SOMMAIRE :

I. Transformée en ondelettes:

I.1. Transformée en ondelettes en 1 dimension :

I.1.1. Transformée en ondelettes continue :

I.1.2. Transformée en ondelettes discrète :

I.2. Généralisation à N dimensions :

II. Reconstruction par ondelettes du champ des vitesses :

II.1. L'hypothèse d'irrotationnalité du champ des vitesses :

II.2. Reconstruction du champ des vitesses par transformée en ondelettes :

II.2.1. Reconstruction du potentiel des vitesses :

II.2.2. Reconstruction d'autres quantités :

II.2.3. Séparation des échelles :

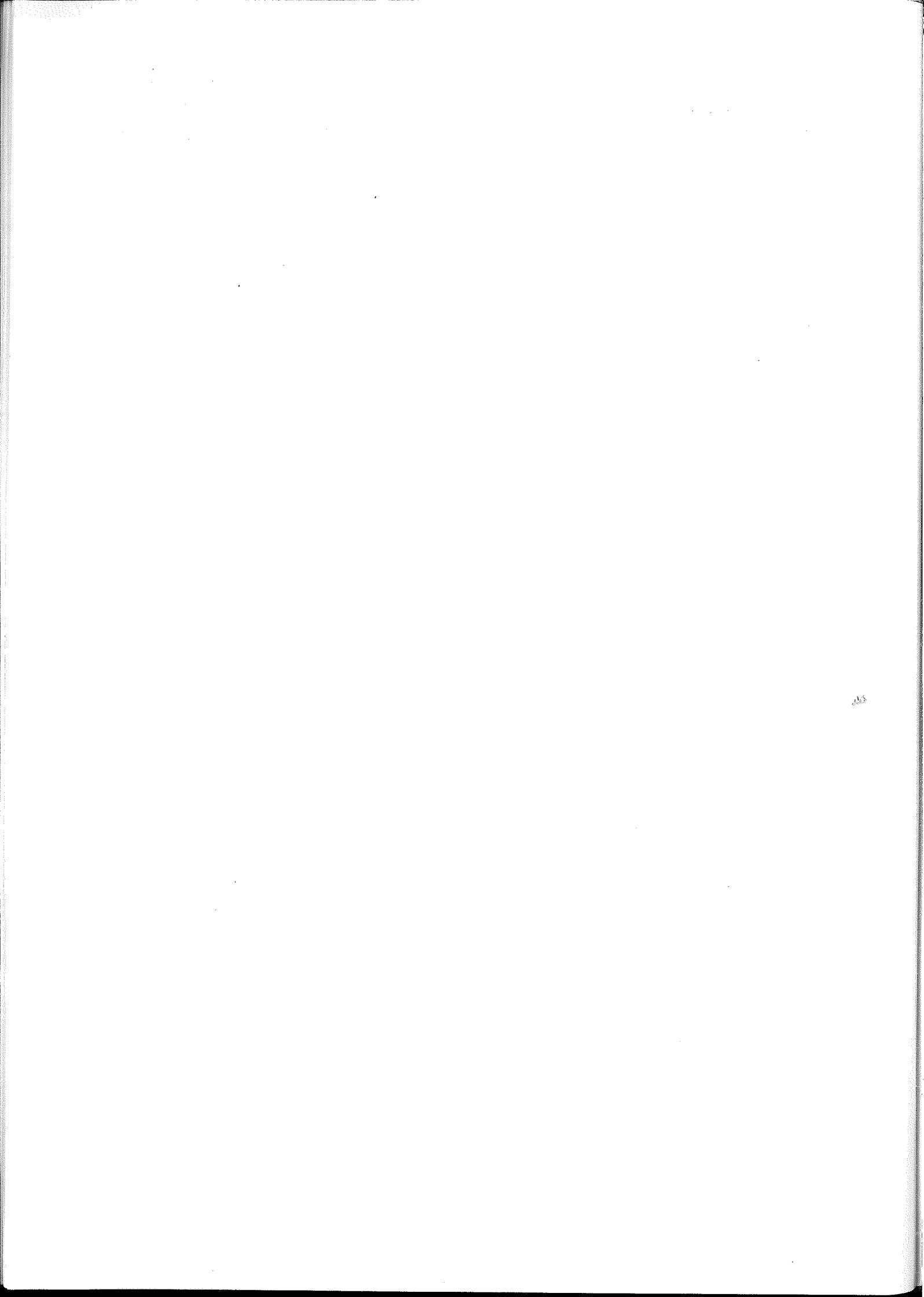
III. Notre méthode comparée à POTENT :

III.1. Le lissage des données :

III.2. L'influence des erreurs en distance :

III.3. L'inhomogénéité spatiale des données :

IV. Annexe :



Les petites fluctuations initiales de densité de matière, amplifiées sous l'influence de la gravitation, ont engendré le champ des vitesses propres actuellement observé. Les théories classiques de formation des structures prédisent qu'à suffisamment grande échelle, ce champ des vitesses est irrotationnel. Ce champ dérive alors d'un potentiel qu'il est possible d'estimer à partir du champ des vitesses radiales seulement en intégrant celui-ci le long de la ligne de visée. Cette remarque est à la base de la méthode POTENT (Bertschinger et Dekel (1989)) qui permet la reconstruction du potentiel des vitesses à partir du champ des vitesses radiales préalablement lissé.

Dans cette partie, nous proposons une méthode, basée sur l'analyse par ondelettes, alternative à la méthode POTENT. La transformée en ondelettes est une transformation linéaire qui permet d'obtenir le comportement spectral d'une fonction tout en conservant ses caractéristiques spatiales. Ces propriétés confèrent à notre méthode deux principaux atouts.

D'une part, notre procédure de reconstruction est linéaire vis-à-vis du champ des vitesses. C'est une propriété appréciable, spécialement lorsqu'il s'agit de traiter le champ des vitesses propres observé, bruité par des incertitudes inhérentes au processus de détermination de la distance la plus probable des galaxies.

D'autre part, notre méthode nous offre un contrôle naturel du lissage des données intervenant obligatoirement lors de la reconstruction du potentiel des vitesses ou d'autres quantités.

Nous avons appliqué notre méthode à des données simulées afin de déterminer sa fiabilité. Notre procédure de reconstruction et la méthode POTENT s'avèrent en fait complémentaires, le traitement des données réelles nécessitant l'introduction de certaines corrections effectivement propres à chacune des deux méthodes. L'analyse des catalogues de vitesses propres disponibles est actuellement en cours.

I. TRANSFORMÉE EN ONDELETTES :

Nous présentons succinctement dans cette section la transformée en ondelettes, un outil mathématique puissant permettant une analyse à la fois spatiale et par échelles d'une fonction.

Nous introduisons le concept d'ondelette mère, de noyau reproducteur ainsi que le théorème de reconstruction pour la transformée en ondelettes continue en 1 dimension. Nous montrons ensuite qu'il est possible de construire une version discrète de la transformée en ondelettes et nous rappelons certaines de ses propriétés.

Enfin nous généralisons ces définitions lorsque les fonctions à analyser agissent sur un espace à N dimensions. Nous introduisons de plus quelques notations propres qui nous seront utiles par la suite.

I.1. Transformée en ondelettes en 1 dimension :

Nous sommes régulièrement amenés en physique à transformer l'expression d'une fonction (par exemple les résultats d'une mesure) afin de visualiser plus clairement ses caractéristiques propres. Mathématiquement, cette opération correspond à définir une fonctionnelle T qui à chaque fonction d'un espace E associe la fonction Tf dans l'espace F , généralement différent de E .

$$T : \begin{cases} E \rightarrow F \\ f \mapsto Tf \end{cases} \quad (E.1)$$

Parmi ces transformations, on s'intéresse particulièrement aux fonctionnelles, que l'on appellera par la suite *transformées*, vérifiant les deux propriétés suivantes.

D'une part on désire que T soit une transformation linéaire ($T(f + \alpha g) = Tf + \alpha Tg$) et que la fonctionnelle T conserve les propriétés métriques des espaces E et F dans le sens que, si on note \langle, \rangle_E et \langle, \rangle_F les produits scalaires des espaces E et F , la relation

$\langle f, (g + \alpha h) \rangle_E = \langle f, g \rangle_E + \alpha \langle f, h \rangle_E$ vérifiée pour tout f, g et h appartenant à E , soit également vérifiée par les transformées dans l'espace F ($\langle Tf, T(g + \alpha h) \rangle_F = \langle Tf, Tg \rangle_F + \alpha \langle Tf, Th \rangle_F$).

D'autre part, on exige que la transformation T possède une transformation inverse T^{-1} ($T^{-1} \circ T = \mathbf{1}_E$). Cette condition stipule simplement que l'on ne souhaite perdre aucune information lors de la transformation (si $f \neq g$ alors $Tf \neq Tg$), c'est-à-dire que l'on peut toujours reconstruire la fonction f à partir de l'expression de sa transformée Tf .

La transformée de Fourier est un exemple de transformation vérifiant ces propriétés de linéarité et d'inversibilité. Soit E l'espace des fonctions de la droite réelle \mathbb{R} , on engendre à partir de la fonction de base $g : g(x) = \frac{1}{\sqrt{2\pi}} e^{ix}$ une famille de fonctions $\{g_w\}$ indicée par la variable $w \in \mathbb{R}$ telle que :

$$g_w(x) = g(wx) \quad (E.2)$$

La transformée de Fourier est alors la fonctionnelle F de E dans E , définie de la façon suivante :

$$F : \begin{cases} E \longrightarrow E \\ f \longmapsto Ff : Ff(w) = \langle g_w, f \rangle_E \end{cases} \quad (E.3)$$

avec $\langle g_w, f \rangle_E = \int_{\mathbb{R}} dx \overline{g_w(x)} f(x)$ le produit scalaire de g_w et f dans E . La fonctionnelle F est évidemment linéaire et possède un inverse lorsqu'on réduit l'espace E à $L^2(\mathbb{R})$, l'espace de Hilbert des fonctions agissant sur \mathbb{R} de carré sommable ($L^2(\mathbb{R}) = \{f \in E \mid \|f\| = \langle f, f \rangle^{1/2} < \infty\}$).

La transformée de Fourier est largement utilisée en physique afin de visualiser les caractéristiques spectrales d'une fonction. C'est en fait la structure de la famille des fonctions $\{g_w\}$ qui permet cette analyse par échelles. Les $\{g_w\}$ sont en effet engendrées à partir de la fonction de base g par l'action du groupe des dilatations (homothéties de rapport w) de l'espace. Lorsqu'on applique une dilatation de rapport w à l'espace ($x \mapsto wx$), la fonction g est effectivement transformée en g_w ($g(x) \mapsto g(wx) = g_w(x)$). La transformée de Fourier est donc caractérisée par les symétries que possède la famille des fonctions $\{g_w\}$ et permet une analyse par échelle d'une fonction f , à comparer avec

l'expression naturelle de f ($f : x \mapsto f(x)$) qui nous renseigne sur son comportement spatial.

Car l'expression naturelle d'une fonction f peut très grossièrement être interprétée comme la décomposition de f sur la base de l'espace des distributions de Dirac $\delta_{x_0}(x) = \delta(x - x_0)$ ($f(x) = \int_{\mathbb{R}} f(x_0) \delta(x - x_0) dx_0$). La famille des $\{\delta_{x_0}\}$ est alors engendrée par l'action du groupe des translations de l'espace à partir de la distribution de base δ .

$$x \mapsto x - x_0, \delta(x) \mapsto \delta(x - x_0) = \delta_{x_0}(x)$$

On voit donc que cette procédure générale permet d'établir une correspondance entre analyse spatiale et groupe des translations de l'espace ainsi qu'entre analyse par échelles et groupe des dilatations. Nous savons néanmoins que ces deux analyses s'excluent mutuellement. La transformée de Fourier efface en effet toute l'information spatiale contenue dans une fonction et l'expression naturelle de cette fonction ne permet en rien de déduire son comportement spectral. Mais si il est possible d'engendrer une famille de fonctions par l'action du groupe des dilatations-translations de l'espace à partir d'une fonction de base g , alors on doit pouvoir définir une transformation fournissant simultanément une vision spatiale et par échelles d'une fonction. La transformée en ondelettes explore justement une telle éventualité.

I.1.1. Transformée en ondelettes continue :

On engendre à partir d'une fonction de base $g(x)$ la famille des fonctions $\{g_{s,u}\}$ indicée par les deux paramètres $s \in \mathbb{R}^{*+}$ (s le facteur d'échelle) et $u \in \mathbb{R}$ (u la position spatiale). La construction de cette famille est effectuée par l'action du groupe des dilatations-translations de l'espace sur la fonction de base g .

$$x \mapsto \frac{x - u}{s}, g(x) \mapsto s^\gamma g\left(\frac{x - u}{s}\right) = g_{s,u}(x) \quad (E.4)$$

Le paramètre γ qui intervient dans l'expression des $g_{s,u}(x)$ pourra prendre les valeurs $\gamma = -1$ ou $\gamma = -\frac{1}{2}$, suivant le choix de la représentation du groupe des dilatations-translations sur l'espace des fonctions dans \mathbb{R} (voir Grossmann et al. (1985)). On définit maintenant la fonctionnelle W_g de E , espace des fonctions sur \mathbb{R} , dans F espace des fonctions de $\mathbb{R}^{*+} \times \mathbb{R}$ (\langle, \rangle_E est le produit scalaire dans E) :

$$W_g : \begin{cases} E \rightarrow F \\ f \mapsto W_g f : W_g f(s, u) = \langle g_{s,u}, f \rangle_E \end{cases} \quad (E.5)$$

. La condition d'admissibilité :

On peut alors montrer que si E est réduit à $L^2(\mathbb{R})$, l'ensemble des fonctions sur \mathbb{R} de carré sommable, il existe certaines fonctions de base $g(x)$ telles que la transformation W_g associée à g soit inversible (Grossmann et al. (1985)). Ces fonctions g appelées ondelettes "mères" ou ondelettes d'analyse doivent vérifier la *condition d'admissibilité* suivante :

$$C_g = \|g\|^{-2} \int_{\mathbb{R}^{**} \times \mathbb{R}} d\mu_\gamma |\langle g_{s,u}, g \rangle_E|^2 < \infty \quad (E.6)$$

avec $\|g\| = \langle g, g \rangle^{1/2}$ la norme de g dans $E = L^2(\mathbb{R})$ et $d\mu_\gamma = s^{-(2\gamma+3)} ds du$ la mesure que l'on associe à l'espace des fonctions $L^2(\mathbb{R}^{**} \times \mathbb{R})$ et qui dépend du choix du paramètre de normalisation γ précédemment introduit en (E.4). Nous présentons figure (E.1) et (E.2) deux exemples d'ondelette mère vérifiant la condition d'admissibilité (E.6).

L'ondelette type "chapeau haut-de-forme" est à support compact et est définie de la façon suivante :

$$g : \begin{cases} g(x) = \frac{1}{\sqrt{2}}, & \text{si } 0 \leq |x| < 1; \\ g(x) = -\frac{1}{2\sqrt{2}}, & \text{si } 1 \leq |x| < 3; \\ g(x) = 0, & \text{si } 3 \leq |x|. \end{cases}$$

Nous montrons figure (E.1) la fonction $g_{s,u}(x)$ pour différentes valeurs des paramètres s et u .

L'ondelette type "chapeau mexicain" (figure (E.2)) est un exemple d'ondelette mère à support infini :

$$g(x) = \frac{2}{\sqrt{3}} \pi^{-1/4} (1 - x^2) \exp\left(-\frac{x^2}{2}\right)$$

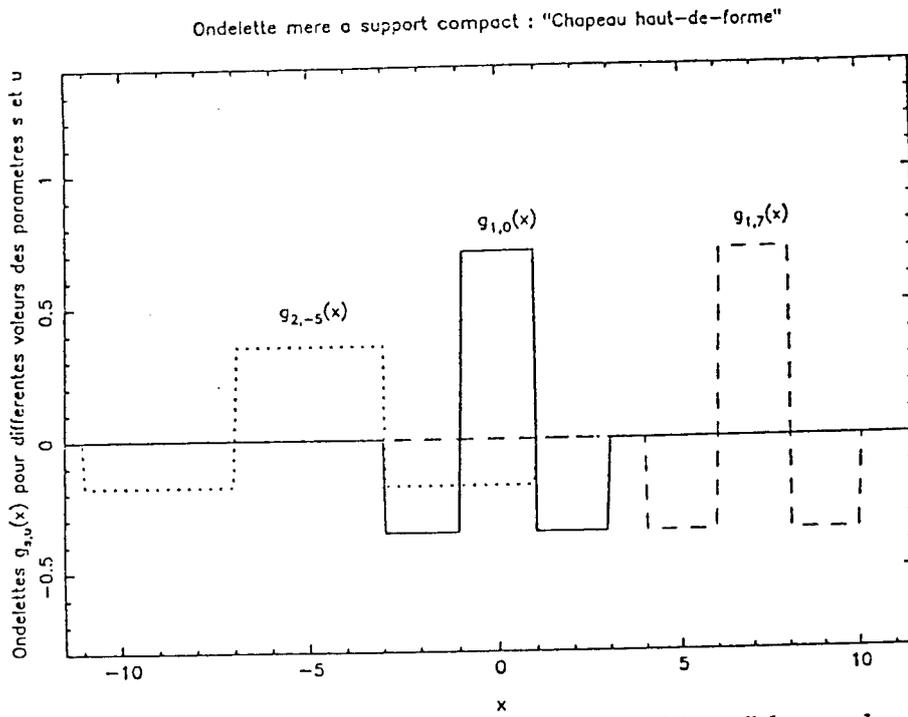


Figure E.1 : Exemple d'ondelette à support compact : ondelette "chapeau haut-de-forme"

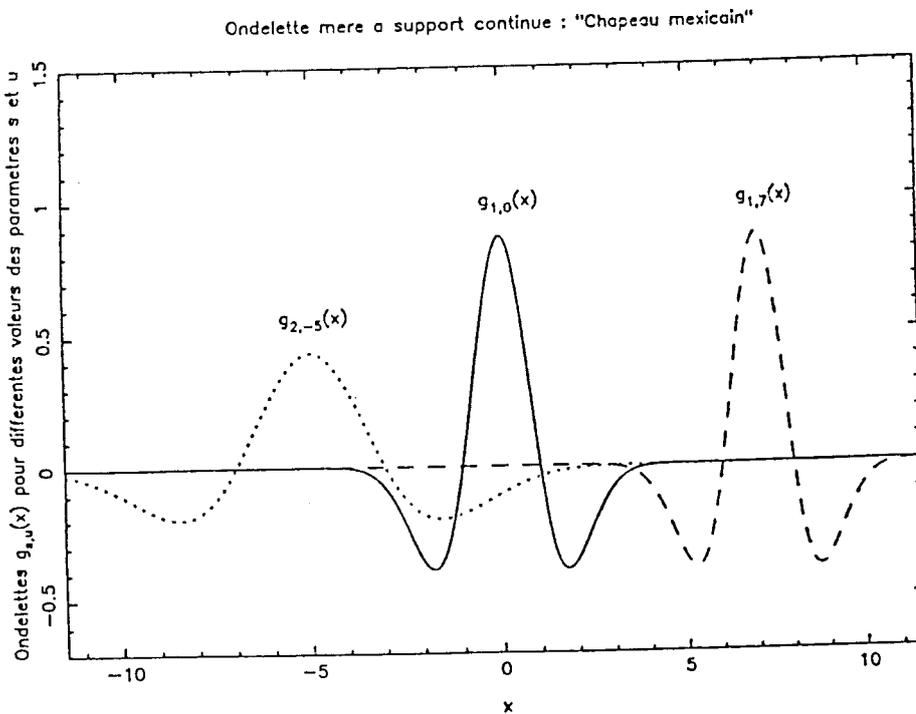


Figure E.2 : Exemple d'ondelette à support infini : ondelette "chapeau mexicain"

. La transformée en ondelettes :

La transformée en ondelettes W_g , d'ondelette mère g vérifiant la condition d'admissibilité (E.6), est finalement définie de la façon suivante :

$$W_g : \begin{cases} L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R}^{*+} \times \mathbb{R}) \\ f \longmapsto W_g f : W_g f(s, u) = \frac{1}{\sqrt{C_g}} \int_{\mathbb{R}} dx s^\gamma \overline{g\left(\frac{x-u}{s}\right)} f(x) \end{cases} \quad (E.7)$$

On peut alors remarquer que W_g est une isométrie de $L^2(\mathbb{R})$ dans $L^2(\mathbb{R}^{*+} \times \mathbb{R})$ muni de la mesure $d\mu_\gamma = s^{-(2\gamma+3)} ds du$, la propriété suivante étant en effet vérifiée quelques soient f_1 et f_2 appartenant à $L^2(\mathbb{R})$:

$$\int_{\mathbb{R}^{*+} \times \mathbb{R}} d\mu_\gamma \overline{W_g f_1(s, u)} W_g f_2(s, u) = \int_{\mathbb{R}} dx \overline{f_1(x)} f_2(x) \quad (E.8)$$

Et la fonctionnelle W_g étant une isométrie, l'expression de sa transformée inverse W_g^{-1} est alors déduite facilement :

$$W_g^{-1} : \begin{cases} L^2(\mathbb{R}^{*+} \times \mathbb{R}) \longrightarrow L^2(\mathbb{R}) \\ h \longmapsto W_g^{-1} h : W_g^{-1} h(x) = \frac{1}{\sqrt{C_g}} \int_{\mathbb{R}^{*+} \times \mathbb{R}} d\mu_\gamma h(s, u) s^\gamma g\left(\frac{x-u}{s}\right) \end{cases} \quad (E.9)$$

On peut montrer que, si la fonction de base $g(x)$ vérifie la condition d'admissibilité (E.4), l'identité $W_g^{-1} \circ W_g = \mathbb{1}_E$ est vraie pour toute fonction sur \mathbb{R} de carré sommable ($E = L^2(\mathbb{R})$).

. Le théorème de reconstruction :

Cette propriété peut s'exprimer par le *théorème de reconstruction* qui stipule que, quelquesoit f appartenant à $L^2(\mathbb{R})$:

$$f(x) = \frac{1}{\sqrt{C_g}} \int_{\mathbb{R}^{*+} \times \mathbb{R}} s^{-(2\gamma+3)} ds du W_g f(s, u) g_{s,u}(x) \quad (E.10)$$

Aucune information n'est alors perdue lors de l'application de la transformée en ondelettes W_g puisqu'il est possible de reconstruire la fonction f à partir de ses coefficients

$W_g f(s, u)$. La fonction $W_g f(s, u)$ nous permet de plus de visualiser le comportement de f sous deux aspects différents.

D'une part, pour s fixé, $W_g f(s, u)$ nous fournit les caractéristiques spatiales de f à l'échelle s .

D'autre part, pour u fixé, l'étude de la variation de $W_g f(s, u)$ avec l'échelle s dévoile les caractéristiques spectrales de f au point u . Dans ce sens, la transformée en ondelettes permet une analyse spectrale locale d'une fonction (B. Torrèsani (1989)).

Une redondance d'information :

La transformée en ondelettes possède néanmoins une propriété particulière dont est dépourvue la transformée de Fourier par exemple. Nous avons vu précédemment que la famille des fonctions $\{g_{s,u}(x)\}$ était engendrée à partir de la fonction de base $g(x)$ par l'action du groupe des dilatations-translations. Or ce groupe est non-commutatif, contrairement au groupe des dilations impliqué dans la transformée de Fourier ou au groupe des translations pour l'expression naturelle d'une fonction. Cette propriété de non-commutativité implique notamment qu'il existe des couples de fonctions $(g_{s,u}(x), g_{s',u'}(x))$, avec $(s, u) \neq (s', u')$, tels que :

$$\langle g_{s,u}, g_{s',u'} \rangle_E \neq 0 \quad (E.11)$$

La famille de fonctions $\{g_{s,u}\}$ n'est donc pas une famille orthogonale de l'espace des fonctions $L^2(\mathbb{R}^{*+} \times \mathbb{R})$. En fait, la famille des $\{g_{s,u}\}$ engendre seulement un sous-espace vectoriel F' de l'espace $L^2(\mathbb{R}^{*+} \times \mathbb{R})$, qui est l'image de $L^2(\mathbb{R})$ par la transformée en ondelettes W_g ($F' = W_g(L^2(\mathbb{R}))$). Afin de caractériser ce sous-espace, on introduit une fonction $K(s, u, s', u')$ appelée le *noyau reproducteur* définie pour tout (s, u) et (s', u') appartenant à $(\mathbb{R}^{*+} \times \mathbb{R})$ de la façon suivante :

$$K(s, u, s', u') = \frac{1}{C_g} \langle g_{s,u}, g_{s',u'} \rangle_E \quad (E.12)$$

D'après la remarque précédente (équation E.11), on sait que le noyau reproducteur $K(s, u, s', u')$ est une fonction plus complexe que la fonction identité pour $(s, u) = (s', u')$. Afin d'appartenir à $F' = W_g(L^2(\mathbb{R}))$, la fonction $h(s, u)$ devra alors vérifier la condition :

$$h(s, u) = \int_{\mathbb{R}^{++} \times \mathbb{R}} s'^{-(2\gamma+3)} ds' du' h(s', u') K(s, u, s', u') \quad (E.13)$$

La transformée en ondelettes $W_g f(s, u)$ de toute fonction $f(x)$ appartenant à $L^2(\mathbb{R})$ vérifie donc l'équation (E.13). Ainsi, les coefficients $W_g f(s, u)$ ne sont pas indépendants les uns des autres. La transformée en ondelettes présente donc une redondance d'information (voir Daubechies (1988)), dans le sens qu'un nombre réduit de coefficients $W_g f(s, u)$ est suffisant en pratique afin de reconstruire la fonction $f(x)$.

I.1.2. Transformée en ondelettes discrète :

La transformée en ondelettes continue est donc une fonctionnelle qui à toute fonction f de $L^2(\mathbb{R})$ associe sa décomposition $W_g f(s, u)$ sur la famille d'ondelettes $\{g_{s,u}\}$ indicée par le couple de paramètres (s, u) variant continuellement sur $\mathbb{R}^{++} \times \mathbb{R}$. Il est parfois souhaitable de réduire cet espace des indices à un sous-espace dénombrable, c'est-à-dire de considérer une sous-famille d'ondelettes $\{g_{m,n}\}$ cette fois-ci indicée par le couple $(m, n) \in \mathbb{Z}^2$, \mathbb{Z} l'ensemble des entiers relatifs. On parle alors de transformée en ondelettes discrète ou de séries d'ondelettes comme on a défini les séries de Fourier par exemple. Il faut néanmoins souligner que la transformée en ondelettes discrète s'applique à toute fonction de $L^2(\mathbb{R})$, contrairement aux séries de Fourier qui nécessitent la périodicité de la fonction à étudier (voir par exemple Meyer (1989)).

La transformée en ondelettes discrète est définie de la façon suivante. A partir des constantes $s_0 > 1$ et $u_0 \neq 0$, on engendre la famille d'ondelettes $\{g_{m,n}\}$ telle que :

$$g_{m,n}(x) = (s_0^m)^\gamma g\left(\frac{x - nu_0 s_0^m}{s_0^m}\right) \quad \text{avec } m, n \in \mathbb{Z} \quad (E.14)$$

On remarque alors que la famille des $\{g_{m,n}\}$ s'identifie à la famille d'ondelettes continue $\{g_{s,u}\}$ lorsque $s_0 \mapsto 1$ et $u_0 \mapsto 0$. Si la fonction de base $g(x)$ vérifie la condition d'admissibilité (E.6), la transformée en ondelettes discrète s'exprime (voir Daubechies et al. (1986)) :

$$W_g : \begin{cases} L^2(\mathbb{R}) \longrightarrow l^2(\mathbb{Z}^2) \\ f \longmapsto W_g f : W_g f(m, n) = \int_{\mathbb{R}} dx \overline{g_{m,n}(x)} f(x) \end{cases} \quad (E.15)$$

Cette procédure de discrétisation de la transformée en ondelettes nécessite néanmoins quelques précautions. En effet, si dans le cas continu W_g possède toujours un inverse, la condition d'admissibilité (E.6) n'implique pas obligatoirement l'existence de W_g^{-1} dans le cas discret. Il faudra alors imposer une contrainte supplémentaire à l'ondelette mère g afin de permettre la reconstruction de la fonction f à partir des coefficients $W_g f(m, n)$. Cette contrainte stipule que si les familles de coefficients $\{W_g f_1(m, n)\}$ et $\{W_g f_2(m, n)\}$ sont "proches", il en va de même pour les fonctions f_1 et f_2 . Il existe alors $A > 0$ et $B < \infty$ tels que, quelquesoit $f \in L^2(\mathbb{R})$ (voir Daubechies (1988)) :

$$A \|f\|^2 < \sum_{m,n} \|W_g f(m, n)\|^2 < B \|f\|^2 \quad (E.16)$$

La condition (E.16) assure la stabilité de la reconstruction de f . Des exemples de conditions nécessaires et suffisantes pour les fonctions de base g sont donnés dans Daubechies et al. (1986). Il est alors possible de reconstruire la fonction f à partir de ses coefficients $W_g f(m, n)$ (voir Daubechies (1989)) :

$$f(x) = \frac{2}{(A+B)} \sum_{m,n} g_{m,n}(x) W_g f(m, n) + Rf(x) \quad (E.17)$$

$$\text{avec } \|Rf\|^2 < \frac{BA^{-1} - 1}{BA^{-1} + 1} \|f\|^2$$

Il faut remarquer que la fonction résiduelle $Rf(x)$ peut être rendue arbitrairement petite par itérations successives de la formule de reconstruction (E.17) (le facteur $\frac{BA^{-1}-1}{BA^{-1}+1}$ est en effet strictement inférieur à 1). Ainsi une approximation de la valeur de f est obtenue quelquesoit la précision désirée. Le choix des paramètres s_0 et u_0 permet d'ailleurs de contrôler ce degré de précision (voir par exemple Grossmann et al. (1989)).

Parmi les ondelettes mères g vérifiant les propriétés (E.6) et (E.16), certaines possèdent une particularité intéressante, soit que la borne inférieure A est égale à la borne supérieure B intervenant dans l'équation (E.16). On a alors :

$$\sum_{m,n} \| W_g f(m,n) \|^2 = A \| f \|^2 \quad (E.18)$$

Pour de telles fonctions de bases g , la formule de reconstruction (E.17) se réécrit exactement de la façon suivante :

$$f(x) = \frac{1}{A} \sum_{m,n} g_{m,n}(x) W_g f(m,n) \quad (E.19)$$

La transformée en ondelettes discrète W_g devient alors dans ce cas précis un multiple d'une isométrie de $L^2(\mathbb{R})$ dans $l^2(\mathbb{Z}^2)$. Il faut noter que la famille des $\{g_{m,n}\}$ n'est pas nécessairement orthonormale, soit que la transformée en ondelettes discrète W_g , g vérifiant (E.6) et (E.18), peut présenter une redondance d'information comme la transformée en ondelettes continue définie précédemment en I.1.1. Cette redondance est parfois utile pour certains types d'étude et peut être ajustée à volonté par le choix des paramètres s_0 et u_0 intervenant dans la définition de la famille d'ondelettes $\{g_{m,n}\}$ (équation (E.14)). Lorsque s_0 est voisin de 1 et u_0 voisin de 0, la transformée en ondelettes montre une redondance importante que l'on peut éliminer en choisissant s_0 et u_0 tels que la famille des $\{g_{m,n}\}$ devienne orthonormale (voir par exemple Daubechies et al. (1986) ou Daubechies (1988)).

Car la discrétisation apporte une propriété fondamentale à la transformée en ondelettes. Il est en effet possible de s'affranchir de la redondance d'information obligatoirement présente dans le cas continu, en travaillant avec une famille d'ondelettes discrète $\{g_{m,n}\}$ orthonormale. Cette propriété est particulièrement importante lorsqu'on désire stocker dans un minimum de place l'information présente dans une fonction. De plus, le choix d'une base d'ondelettes orthonormale permet d'obtenir très facilement les caractéristiques spectrales d'une fonction f à partir de ses coefficients $W_g f(m,n)$.

Le problème de la construction de familles d'ondelettes $\{g_{m,n}\}$ a été abordé par divers auteurs (par exemple Daubechies et al. (1986), Daubechies (1988), Meyer (1989) ou Mallat (1989)). L'algorithme de S. Mallat nous fournit notamment une méthode élégante permettant d'engendrer des bases d'ondelettes discrètes orthonormales $\{g_{m,n}\}$ ainsi que de calculer rapidement les coefficients $W_g f(m,n)$ associés à une fonction f de $L^2(\mathbb{R})$ (voir Mallat (1989) ou Daubechies (1988)).

I.2. Généralisation à N dimensions :

La transformée en ondelettes peut se généraliser sans difficulté pour des fonctions agissant sur \mathbb{R}^N . Nous donnons dans cette section les expressions en N dimensions des définitions précédemment exposées dans la section I.1 ainsi que leurs équivalents exprimés dans l'espace de Fourier. De plus, nous introduisons quelques notations propres qui nous seront utiles par la suite.

Nous travaillons donc avec des fonctions définies sur l'espace \mathbb{R}^N et à valeurs dans le corps des complexes \mathbb{C} , un élément de \mathbb{R}^N étant noté $\mathbf{x} = (x_1, x_2, \dots, x_N)$. Nous définissons $\langle f_1, f_2 \rangle$ le produit scalaire de deux fonctions f_1 et f_2 sur \mathbb{R}^N et en posant $d^N \mathbf{x} = dx_1 dx_2 \dots dx_N$, on a :

$$\langle f_1, f_2 \rangle = \int_{\mathbb{R}^N} d^N \mathbf{x} \overline{f_1(\mathbf{x})} f_2(\mathbf{x})$$

On engendre à partir de la fonction de base $g(\mathbf{x})$ la famille de fonctions $\{g_{s,\mathbf{u}}\}$ indicée par $N + 1$ paramètres : s , le facteur d'échelle et \mathbf{u} la position spatiale dans \mathbb{R}^N . On construit cette famille par l'action du sous-groupe du groupe des dilations-translations de l'espace \mathbb{R}^N sur la fonction de base g :

$$\mathbf{x} \mapsto \frac{\mathbf{x} - \mathbf{u}}{s}, \quad g(\mathbf{x}) \mapsto s^\gamma g\left(\frac{\mathbf{x} - \mathbf{u}}{s}\right) = g_{s,\mathbf{u}}(\mathbf{x}) \quad (E.20)$$

Comme dans le cas à une dimension, le paramètre γ qui intervient dans l'expression des $g_{s,\mathbf{u}}(\mathbf{x})$ pourra prendre les valeurs $\gamma = -1$ ou $\gamma = -\frac{1}{2}$. Le choix de γ dépend de l'utilisation que l'on fera par la suite de la transformée en ondelettes. En fait, on travaille généralement en physique avec des variables dimensionnées, la valeur de γ fixera alors la dimension de l'espace des transformées en ondelettes.

De plus, on s'intéresse uniquement à un sous-groupe du groupe des dilations-translations sur \mathbb{R}^N qui ne contient que les dilations isotropes de l'espace. On pourrait en effet construire une famille de fonctions $\{g_{s,\mathbf{u}}\}$ indicée par $2N$ paramètres $(s, \mathbf{u}) \in (\mathbb{R}^{*+N} \times \mathbb{R}^N)$ avec $s = (s_1, s_2, \dots, s_N)$. Pour des raisons de simplicité, nous ne traitons ici que le cas isotrope.

La condition d'admissibilité introduite en (E.6) se généralise en N dimensions de la façon suivante :

$$C_g = \|g\|^{-2} \int_{\mathbb{R}^{*+} \times \mathbb{R}^N} d\mu_\gamma | \langle g_{s,\mathbf{u}}, g \rangle |^2 < \infty \quad (E.21)$$

$$\text{et } d\mu_\gamma = k_s ds d^N \mathbf{u} \quad \text{avec } k_s = s^{-(1+2N(\gamma+1))} \quad (E.22)$$

On exprime souvent cette condition d'admissibilité dans l'espace de Fourier. Si l'on note Ff la transformée de Fourier de la fonction f :

$$Ff(\mathbf{w}) = (2\pi)^{-N/2} \int_{\mathbb{R}^N} d^N \mathbf{x} e^{-i\mathbf{w} \cdot \mathbf{x}} f(\mathbf{x}) \quad (E.23)$$

La condition d'admissibilité pour la fonction de base $g(\mathbf{x})$ devient alors dans la représentation Fourier :

$$C_g = \int_{\mathbb{R}^N} d^N \mathbf{w} \|\mathbf{w}\|^{-1} |\hat{g}(\mathbf{w})|^2 < \infty \quad (E.24)$$

Notamment, cette condition implique que $Ff(\mathbf{0}) = 0$, soit que l'intégrale de l'ondelette mère g sur l'espace soit nulle :

$$\int_{\mathbb{R}^N} d^N \mathbf{x} g(\mathbf{x}) = 0 \quad (E.25)$$

Nous pouvons alors définir la transformée en ondelettes W_g en N dimensions, d'ondelette mère g vérifiant la condition d'admissibilité (E.21), qui à toute fonction f de $L^2(\mathbb{R}^N)$ ($L^2(\mathbb{R}^N) = \{f, \|f\| = \langle f | f \rangle^{1/2} < \infty\}$) associe la fonction $W_g f$ appartenant à $L^2(\mathbb{R}^{*+} \times \mathbb{R}^N)$ telle que :

$$W_g : \begin{cases} L^2(\mathbb{R}^N) \longrightarrow L^2(\mathbb{R}^{*+} \times \mathbb{R}^N) \\ f \longmapsto W_g f : W_g f(s, \mathbf{u}) = \frac{1}{\sqrt{C_g}} \int_{\mathbb{R}^N} d^N \mathbf{x} f(\mathbf{x}) s^{\gamma N} \overline{g((\mathbf{x} - \mathbf{u})/s)} \end{cases} \quad (E.26)$$

Il est parfois utile d'exprimer la transformée en ondelettes $W_g f(s, \mathbf{u})$ dans l'espace de Fourier :

$$W_g f(s, \mathbf{u}) = \frac{1}{\sqrt{C_g}} s^{(\gamma+1)N} \int_{\mathbb{R}^N} d^N \mathbf{w} Ff(\mathbf{w}) e^{i\mathbf{w} \cdot \mathbf{u}} Fg(s\mathbf{w}) \quad (E.27)$$

W_g est une isométrie de $L^2(\mathbb{R}^N)$ dans $L^2(\mathbb{R}^{++} \times \mathbb{R}^N)$ muni de la mesure $d\mu_\gamma$ précédemment définie en (E.22). Son adjoint W_g^\dagger s'identifie alors avec sa transformée inverse W_g^{-1} :

$$W_g^{-1} : \begin{cases} L^2(\mathbb{R}^{++} \times \mathbb{R}^N) \longrightarrow L^2(\mathbb{R}^N) \\ h \longmapsto W_g^{-1}h : W_g^{-1}h(\mathbf{x}) = \frac{1}{\sqrt{C_g}} \int_{\mathbb{R}^{++} \times \mathbb{R}^N} d\mu_\gamma h(s, \mathbf{u}) g_{s, \mathbf{u}}(\mathbf{x}) \end{cases} \quad (E.28)$$

Et si l'on introduit les fonctions $f^{(s)}(\mathbf{x})$, les composantes de la fonction $f(\mathbf{x})$ à l'échelle s (dépendant du choix de l'ondelette mère g) :

$$f^{(s)}(\mathbf{x}) = \frac{k_s}{\sqrt{C_g}} \int_{\mathbb{R}^N} d^N \mathbf{u} W_g f(s, \mathbf{u}) s^{\gamma N} g((\mathbf{x} - \mathbf{u})/s) \quad (E.29)$$

Le théorème de reconstruction pour toute fonction $f(\mathbf{x}) \in L^2(\mathbb{R}^N)$ prend alors la forme :

$$f(\mathbf{x}) = \int_0^\infty ds f^{(s)}(\mathbf{x}) \quad (E.30)$$

Enfin, le noyau reproducteur $K(s, \mathbf{u}, s', \mathbf{u}')$ qui caractérise la redondance d'information obligatoirement présente dans la transformée en ondelettes continue s'exprime de la façon suivante :

$$K(s, \mathbf{u}, s', \mathbf{u}') = \frac{1}{C_g} s^{\gamma N} s'^{\gamma N} \int_{\mathbb{R}^N} d^N \mathbf{x} \overline{g((\mathbf{x} - \mathbf{u})/s)} g((\mathbf{x} - \mathbf{u}')/s') \quad (E.31)$$

Ou dans l'espace de Fourier :

$$K(s, \mathbf{u}, s', \mathbf{u}') = \frac{1}{C_g} s^{(\gamma+1)N} s'^{(\gamma+1)N} \int_{\mathbb{R}^N} d^N \mathbf{w} e^{i\mathbf{w} \cdot (\mathbf{u}' - \mathbf{u})} \overline{g(s\mathbf{w})} g(s'\mathbf{w}) \quad (E.32)$$

On remarque alors qu'en notant $K_s(\mathbf{x}, \mathbf{y}) = K(s, \mathbf{x}, s, \mathbf{y})$, la fonction $f^{(s)}(\mathbf{x})$ définie précédemment en (E.29) s'écrit aussi :

$$f^{(s)}(\mathbf{x}) = k_s \int_{\mathbb{R}^N} d^N \mathbf{y} f(\mathbf{y}) K_s(\mathbf{x}, \mathbf{y}) \quad (E.33)$$

La discrétisation de la transformée en ondelettes est aussi possible lorsqu'on travaille en dimension N . L'algorithme de S. Mallat est par exemple appliqué en dimension 2 pour

le traitement d'image (Mallat (1989)) et peut se généraliser sans difficulté à 3 dimensions (même si en pratique les calculs impliqués dans la phase calculatoire deviennent vite assez pesants). La discrétisation de la transformée en ondelettes en plusieurs dimensions possède néanmoins des faiblesses, notamment si l'on veut travailler avec une base d'ondelettes orthogonale. On doit dans ce cas faire intervenir un maillage (en 2D) ou un pavage (en 3D) qui recouvre l'espace et qui impose une forme bien précise à l'ondelette mère. On ne peut alors généralement plus conserver les symétries originelles auxquelles est soumis l'espace. En dimension 3 par exemple, on doit abandonner la symétrie sphérique pour l'ondelette mère si l'on désire utiliser une famille d'ondelettes orthonormale.

II. RECONSTRUCTION PAR ONDELETTES DU CHAMP DES VITESSES :

Nous proposons dans cette section une méthode basée sur l'analyse par ondelettes qui permet de reconstruire totalement un champ de vitesses irrotationnel (c'est-à-dire dérivant d'un potentiel) à partir de sa composante radiale seulement.

D'après les théories classiques de formation des structures présentes dans l'univers, le champ des vitesses propres des galaxies possède à suffisamment grande échelle cette propriété d'irrotationnalité. Ainsi notre méthode de reconstruction doit s'appliquer au champ des vitesses radiales propres actuellement observées.

Notre méthode apporte une alternative à la méthode POTENT (reconstruction du potentiel des vitesses à partir du champ des vitesses radiales) proposée en 1989 par E. Bertschinger et A. Dekel. L'utilisation de la transformée en ondelettes présente en fait de nombreux avantages. D'une part, notre analyse est linéaire vis-à-vis du champ des vitesses. D'autre part, nous pouvons reconstruire diverses quantités (par exemple les composantes transverses de la vitesse) sans évaluer préalablement le potentiel des vitesses. Enfin, la transformée en ondelettes nous permet une analyse par échelle du champ des vitesses ainsi qu'un contrôle naturel des effets indésirables intervenant lors de l'application de notre méthode à des données discrètes (par exemple le problème de lissage des données).

Nous présentons dans la section suivante une comparaison détaillée entre la méthode POTENT et notre méthode.

II.1. L'hypothèse d'irrotationnalité du champ des vitesses :

Dans l'approximation Newtonienne et lorsque la dynamique demeure encore linéaire, le champ des vitesses $V(x, t)$, engendré par des petites fluctuations de la densité de matière amplifiées au cours du temps sous l'influence de la gravitation, s'exprime de la façon suivante (voir Peebles (1980), chapitre 14 et 15) :

$$\mathbf{V}(\mathbf{x}, t) = -a(t) \left[\frac{\partial D_1}{\partial t}(t) \nabla \Phi_1(\mathbf{x}) + \frac{\partial D_2}{\partial t}(t) \nabla \Phi_2(\mathbf{x}) \right] + a^{-1}(t) \mathbf{F}(\mathbf{x}) \quad (E.34)$$

$$\text{avec } \nabla \cdot \mathbf{F} = 0$$

où \mathbf{x} est la position (au sens Eulérien) dans l'espace comobile, $a(t)$ le facteur d'échelle croissant au cours du temps, $D_1(t)$ et $D_2(t)$ donnent la dépendance temporelle des modes linéaires croissant et décroissant des potentiels gravitationnels proportionnels à $\Phi_1(\mathbf{x})$ et $\Phi_2(\mathbf{x})$. Enfin, $\mathbf{F}(\mathbf{x})$ représente la contribution au champ de vitesses à l'instant t d'un flot possédant une divergence nulle, c'est à dire purement rotationnel. On remarque donc qu'aux échelles où les équations décrivant la dynamique du fluide cosmique restent linéaires, la composante rotationnelle ainsi que le mode décroissant du champ des vitesses vont être supplantés progressivement durant l'expansion de l'univers par le mode croissant des perturbations de densité. Ainsi, les théories classiques de formation des grandes structures prédisent qu'à suffisamment grande échelle, le champ des vitesses actuellement observé $\mathbf{V}(\mathbf{x})$ dérive d'un potentiel $\Phi(\mathbf{x})$:

$$\mathbf{V}(\mathbf{x}) = \nabla \Phi(\mathbf{x}) \quad (E.35)$$

On sait bien sûr qu'aux petites échelles (typiquement à l'échelle des galaxies ou des amas de galaxies), les termes non-linéaires présents dans les équations régissant la dynamique de ces structures ne sont pas négligeables. Le champ des vitesses propres des galaxies cesse alors d'être purement irrotationnel ($\nabla \times \mathbf{V}(\mathbf{x}) \neq \mathbf{0}$) et l'équation (E.35) n'est plus vérifiée. Néanmoins il est possible théoriquement, soit en lissant le champ des vitesses (méthode POTENT), soit en définissant une échelle inférieure de coupure (notre méthode), de faire disparaître ce champ des vitesses turbulent agissant à petite échelle. Toutefois, nous montrons dans la section III.2 que la présence de ce champ turbulent risque en pratique de créer des difficultés lors de la reconstruction du champ des vitesses $\mathbf{V}(\mathbf{x})$.

Si le champ des vitesses $\mathbf{V}(\mathbf{x})$ dérive effectivement d'un potentiel $\Phi(\mathbf{x})$, alors il est possible de déduire ce potentiel à partir de la composante radiale $V_r(\mathbf{x})$ du champ des vitesses en intégrant cette composante le long de la ligne de visée :

$$\Phi(\mathbf{x}) = \int_0^1 dl V_r(l\mathbf{x}) \quad (E.36)$$

L'équation (E.36) est à la base de la méthode POTENT (voir Bertschinger et al. (1989), Bertschinger et al. (1990) et Dekel et al. (1990)) et de notre méthode. Il est alors possible de reconstruire le potentiel des vitesses $\Phi(\mathbf{x})$ et donc d'après l'équation (E.35), le champ des vitesses total en 3 dimensions $\mathbf{V}(\mathbf{x})$.

Nous voulons souligner que cette hypothèse de travail (le champ des vitesses est purement irrotationnel : $\nabla \times \mathbf{V}(\mathbf{x}) = \mathbf{0}$) ne peut pas être démontrée dans le cadre de cette étude. Certains tests sont théoriquement à même de vérifier cette hypothèse (voir la discussion partie C, section II.3), mais ne sont pas applicable en pratique compte-tenu des données actuellement disponibles. La valeur du potentiel des vitesses reconstruit ne sera donc d'aucune utilité si le champ des vitesses réellement observé ne vérifie pas notre hypothèse de travail.

II.2. Reconstruction du champ des vitesses par transformée en ondelettes :

Nous faisons donc l'hypothèse que le champ des vitesses $\mathbf{V}(\mathbf{x})$ à étudier dérive d'un potentiel $\Phi(\mathbf{x})$, soit que les équations (E.35) et (E.36) sont satisfaites. De plus, seul le champ des vitesses radiales $V_r(\mathbf{x})$ est connu. Nous décomposons ce champ par échelles suivant les notations introduites dans le chapitre précédent (équations (E.29) et (E.30) pour un espace à 3 dimensions, avec $k_s = s^{-7-6\gamma}$, voir équation (E.22)) :

$$V_r(\mathbf{x}) = \int_0^\infty ds V_r^{(s)}(\mathbf{x}) \quad (E.37)$$

$$\text{avec } V_r^{(s)}(\mathbf{x}) = \frac{k_s}{\sqrt{C_g}} \int_{\mathbf{R}^3} d^3\mathbf{u} W_g V_r(s, \mathbf{u}) s^{3\gamma} g((\mathbf{x} - \mathbf{u})/s) \quad (E.38)$$

Et si nous décomposons suivant la même procédure le potentiel des vitesses $\Phi(\mathbf{x})$:

$$\Phi(\mathbf{x}) = \int_0^\infty ds \Phi^{(s)}(\mathbf{x}) \quad (E.39)$$

L'équation (E.36) nous assure que la relation suivante entre les composantes $\Phi^{(s)}(\mathbf{x})$ et $V_r^{(s)}(\mathbf{x})$ est vérifiée :

$$\Phi^{(s)}(\mathbf{x}) = \int_0^1 dl V_r^{(s)}(l\mathbf{x}) \quad (E.40)$$

On peut alors définir une familles de fonctions $\{G_{s,\mathbf{u}}(\mathbf{x})\}$ à partir de la famille d'ondelettes $\{g_{s,\mathbf{u}}(\mathbf{x})\}$ et que nous appellerons *ondelettes généralisées* :

$$G_{s,\mathbf{u}}(\mathbf{x}) = \int_0^1 dl g_{s,\mathbf{u}}(l\mathbf{x}) \quad (E.41)$$

Et on a alors :

$$\Phi^{(s)}(\mathbf{x}) = \frac{k_s}{\sqrt{C_g}} \int_{\mathbf{R}^3} d^3\mathbf{u} W_g V_r(s, \mathbf{u}) G_{s,\mathbf{u}}(\mathbf{x}) \quad (E.42)$$

Les $\{G_{s,\mathbf{u}}(\mathbf{x})\}$ ne sont pas des ondelettes dans le sens qu'elles ne vérifient pas généralement la condition d'admissibilité (E.24) par exemple. De plus elles perdent la symétrie spatiale à laquelle sont soumises les ondelettes $\{g_{s,\mathbf{u}}(\mathbf{x})\}$. Néanmoins, il est possible de définir (par analogie avec l'équation (E.28)) la fonctionnelle P_G^{-1} de $L^2(\mathbf{R}^{**} \times \mathbf{R}^3)$ dans $L^2(\mathbf{R}^3)$ (qui n'admet pas forcément un inverse P_G) telle que :

$$P_G^{-1} : \begin{cases} L^2(\mathbf{R}^{**} \times \mathbf{R}^3) \longrightarrow L^2(\mathbf{R}^3) \\ h \longmapsto P_G^{-1}h : P_G^{-1}h(\mathbf{x}) = \frac{1}{\sqrt{C_g}} \int_{\mathbf{R}^{**} \times \mathbf{R}^3} k_s ds d^3\mathbf{u} h(s, \mathbf{u}) G_{s,\mathbf{u}}(\mathbf{x}) \end{cases} \quad (E.43)$$

L'introduction de cette fonctionnelle P_G^{-1} nous offre alors une procédure exacte afin de reconstruire le potentiel des vitesses $\Phi(\mathbf{x})$ à partir de la connaissance du champ des vitesses radiales $V_r(\mathbf{x})$. On peut en effet facilement montrer d'après les équations (E.37), (E.39) et (E.42) que la relation suivante est toujours satisfaite :

$$\Phi = [P_G^{-1} \circ W_g] V_r \quad (E.44)$$

En pratique donc, la reconstruction du potentiel des vitesses est effectuée en deux étapes. Premièrement nous calculons les coefficients d'ondelettes $W_g V_r(s, \mathbf{u})$ (dans l'espace

des (s, \mathbf{u}) du champ des vitesses radiales $V_r(\mathbf{x})$. Deuxièmement, nous reconstruisons le potentiel des vitesses $\Phi(\mathbf{x})$ par l'application de la transformation P_G^{-1} sur ces coefficients $W_g V_r(s, \mathbf{u})$. Cette procédure peut paraître quelque peu imposante, notamment lorsqu'il s'agit de stocker les valeurs des coefficients d'ondelettes $W_g V_r(s, \mathbf{u})$. Toutefois, la transformée en ondelettes continue présente une redondance d'information caractérisée par son noyau reproducteur (voir section I.1.1) qui nous permet de simplifier considérablement cette procédure.

Par analogie avec la définition du noyau reproducteur (équation (E.31)), nous introduisons un *noyau généralisé* $N(s, \mathbf{u}, s', \mathbf{u}')$ propre à la transformation linéaire $P_G^{-1} \circ W_g$ définie précédemment en (E.44) :

$$N(s, \mathbf{u}, s', \mathbf{u}') = \frac{1}{C_g} \int_{\mathbb{R}^3} d^3 \mathbf{x} \overline{g_{s, \mathbf{u}}(\mathbf{x})} G_{s', \mathbf{u}'}(\mathbf{x}) \quad (E.45)$$

En notant $N_s(\mathbf{x}, \mathbf{y}) = N(s, \mathbf{x}, s, \mathbf{y})$, on peut aisément démontrer la validité de la relation suivante (voir équation (E.42)) :

$$\Phi^{(s)}(\mathbf{x}) = k_s \int_{\mathbb{R}^3} d^3 \mathbf{y} V_r(\mathbf{y}) N_s(\mathbf{x}, \mathbf{y}) \quad (E.46)$$

Notre procédure de reconstruction du potentiel des vitesses se limite alors à une convolution entre le champ des vitesses propres V_r et le noyau généralisé N_s suivie d'une intégration sur les échelles. De plus, nous verrons que cette intégrale sur les échelles peut être en pratique remplacée par une somme sur un nombre réduit d'échelles, compte-tenu de la redondance d'information présente dans la transformée en ondelettes.

II.2.1. Reconstruction du potentiel des vitesses :

Dans le but de tester la validité de notre méthode, nous avons appliqué la procédure précédemment décrite à des données simulées. Dans cette section, nous présentons en détails les différentes phases de cette application.

Les données simulées :

Nous avons simulé sur une grille (en 3 dimensions) un champ de vitesses radiales dérivant d'un potentiel des vitesses $\Psi = \Psi_1 + \Psi_2$. Nous avons choisi pour les potentiels

Ψ_1 et Ψ_2 un profil gaussien de dispersion (ou longueur de cohérence) respectivement σ_1 et σ_2 et de centre M_1 et M_2 . La grille possède $50 \times 50 \times 50$ points et un pas de 200 km.s^{-1} , soit que notre grille définit un cube centré sur l'observateur d'extension $([-5000, 5000], [-5000, 5000], [-5000, 5000])$ (en km.s^{-1}). Le potentiel Ψ_1 est centré en M_1 de coordonnées $(4500, 0, 0)$ et possède une dispersion σ_1 de 4500 km.s^{-1} , le potentiel Ψ_2 en M_2 $(0, 2000, 0)$ avec $\sigma_2 = 2000 \text{ km.s}^{-1}$. Nous avons choisi ces potentiels afin de reproduire très grossièrement le champ des vitesses propres observé dans le catalogue de Faber et *al.* (1989) par exemple. Le potentiel Ψ_1 peut alors être associé au "Grand Attracteur", un flot de vitesses convergent vers un centre situé au bord de l'échantillon et cohérent à une échelle importante (voir partie D). Nous avons introduit le potentiel Ψ_2 afin de reproduire un attracteur local (la chute sur l'amas de Virgo par exemple), proche de l'observateur et de longueur de cohérence moins importante. De plus la présence de 2 attracteurs d'échelles de cohérence différentes nous permettra d'illustrer que notre méthode nous offre la possibilité de séparer les échelles (voir section II.2.3). Finalement, les données en entrée seront la vitesse radiale, en chaque point de la grille, induite par le potentiel des vitesses Ψ .

Afin de visualiser les résultats obtenus, nous présenterons les valeurs du potentiel simulé et reconstruit sur le plan P_1 (une coupe) passant par les centres des attracteurs M_1 et M_2 ainsi que l'observateur O situé au centre de la grille. Nous montrons dès à présent figure (E.3) l'amplitude du potentiel simulé $\Psi(\mathbf{x})$ sur la coupe P_1 . L'observateur O est donc au centre du carré, le point M_1 au fond et M_2 sur la droite.

Le choix de l'ondelette mère :

Nous utilisons dans cette étude l'ondelette d'analyse dite "chapeau haut-de-forme" généralisée à 3 dimensions. Cette ondelette possède une symétrie sphérique $g(\mathbf{x}) = g'(r)$ avec $r = \|\mathbf{x}\|$ et $g'(r)$ s'exprimant de la façon suivante :

$$g' : \begin{cases} g'(r) = \frac{1}{C}, & \text{si } 0 \leq r < 1; \\ g'(r) = -\frac{1}{26} \frac{1}{C}, & \text{si } 1 \leq r < 3; \\ g'(r) = 0, & \text{si } 3 \leq r. \end{cases} \quad (E.47)$$

$$\text{avec } C = \sqrt{\frac{4}{3}\pi \left(1 + \frac{1}{26}\right)}$$

Nous présentons figures (E.4) et (E.5) une projection en 2 dimensions de notre on-

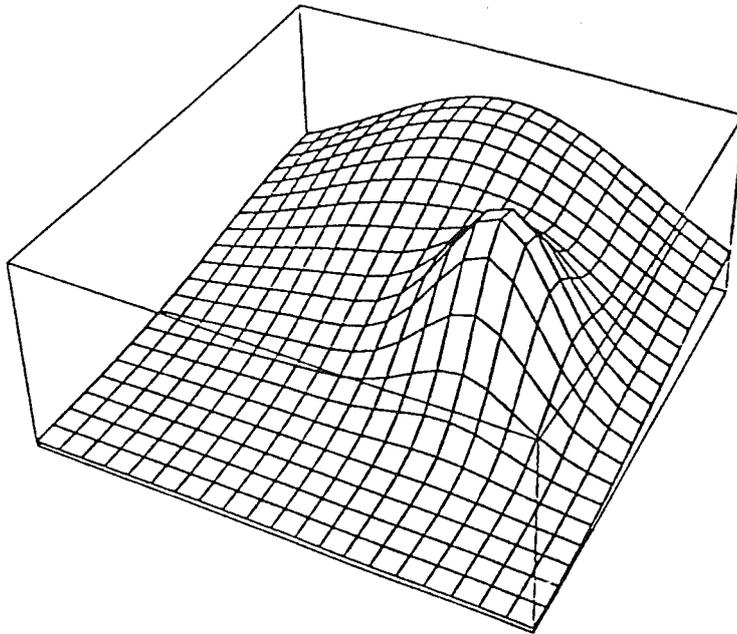


Figure E.3 : Amplitude du potentiel des vitesses simulé Ψ sur la coupe P_1 .

ondelette d'analyse "chapeau haut-de-forme" $g(x)$ ainsi que de son noyau reproducteur associé $K_s(x, y)$.

L'utilisation d'une ondelette mère possédant des discontinuités peut en principe créer des problèmes, notamment nuire à la convergence des coefficients d'ondelettes $W_g V_r(s, u)$ (voir par exemple Daubechies (1988)). Néanmoins, ces problèmes de convergence ne nous concernent pas dans cette étude puisque nous appliquons, afin de reconstruire le potentiel des vitesses, la procédure simplifiée décrite précédemment (équation (E.46)). Seuls les noyaux reproducteur $K_s(x, y)$ ou généralisé interviennent alors, et ceux-ci ne montrent pas de discontinuités (voir figure (E.5) ou figures (E.9a) et (E.9b)).

De plus notre ondelette d'analyse possède un support compact (d'extension sphérique $3s$ autour de x pour l'ondelette $g_{s,u}(x)$), ce qui permet un meilleur contrôle des effets de bords que pour une ondelette à support infini par exemple.

Enfin, l'expression simple de notre ondelette "chapeau haut-de-forme" nous offre la possibilité de calculer analytiquement les fonctions intervenant dans la reconstruction du

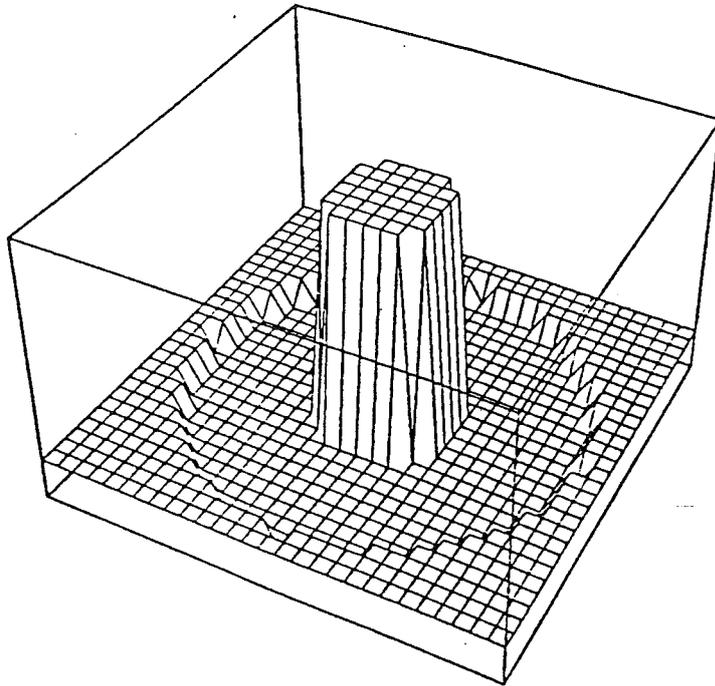


Figure E.4 : Projection 2D de notre ondelette d'analyse "chapeau haut-de-forme".

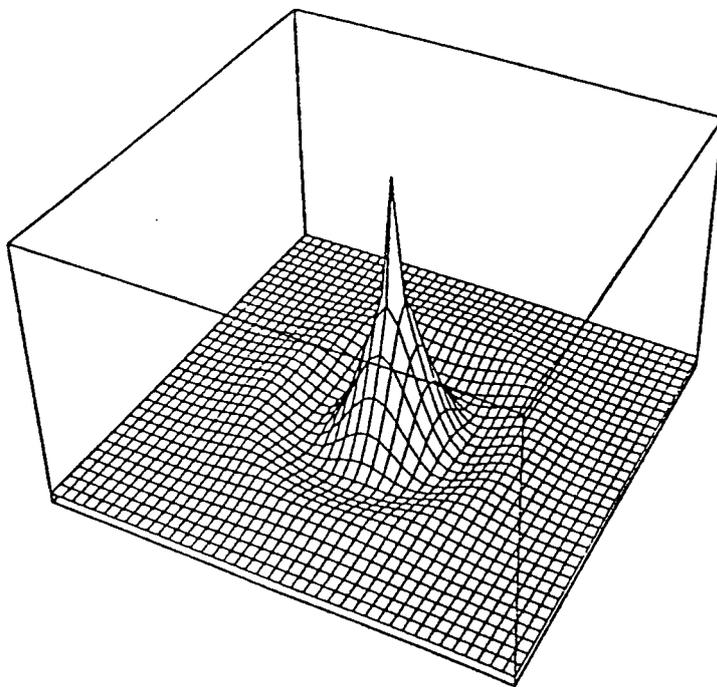


Figure E.5 : Projection 2D du noyau $K_s(x, y)$ de notre ondelette mère (centré sur x et y variant).

potentiel des vitesses comme le noyau reproducteur $K_s(x, y)$ ou l'ondelette généralisée $G_{s,u}(x)$ ainsi que le noyau généralisé $N_s(x, y)$ associés à la transformation linéaire $P_G^{-1} \circ W_g$.

La somme sur les échelles :

La transformée en ondelettes continue nécessite deux opérations d'intégrations. L'une est une intégration spatiale (dans l'espace des u) que nous évitons en utilisant directement les noyaux reproducteur $K_s(x, y)$ ou généralisé $N_s(x, y)$ dans notre procédure simplifiée (équation (E.46)). L'autre est une intégration sur les échelles que nous devons effectuer en phase finale de notre étude afin de reconstruire le potentiel Φ à partir de ses composantes $\Phi^{(s)}$ (équation (E.39)). Néanmoins dans le cas étudié, cette intégration peut être en pratique remplacée par une somme sur un nombre réduit d'échelles, compte-tenu de la redondance d'information présente dans la transformée en ondelettes. Nos données simulées possèdent en effet, comme les échantillons de galaxies réellement observées, une échelle supérieure et inférieure de coupure.

D'une part la grille des points données a une résolution minimum s_{min} (dans notre simulation, $s_{min} = 200 \text{ km.s}^{-1}$). L'intégration sur les échelles impliquée dans l'équation (E.39) doit alors être stoppée à partir de cette échelle inférieure s_{min} . Dans le cas contraire, les résultats obtenus pour la reconstruction du potentiel Ψ seraient plus caractéristiques du support de la fonction Ψ (c'est à dire la grille de points) que des valeurs de Ψ en ces points. On se heurte ici au problème du lissage obligatoire inhérent à l'origine discrète des données physiques (résultats d'une mesure). Lorsqu'on fixe une échelle minimum de coupure s_{min} pour l'intégration intervenant dans l'équation (E.39), on effectue en fait une interpolation Ψ_I de la fonction Ψ aux points où l'information n'est pas disponible (c'est à dire les points extérieurs au support). Il est clair que Ψ_I va être dépendant de la forme de l'ondelette d'analyse $g(x)$ ainsi que de l'échelle minimum de coupure s_{min} . Néanmoins si le potentiel Φ est suffisamment lisse à cette échelle, la reconstruction de Ψ ne dépendra pas en pratique du choix de l'ondelette mère. Nous illustrerons dans la section III.1 que la transformée en ondelettes nous permet, par l'intermédiaire du choix de s_{min} , un contrôle naturel de ces problèmes de lissage.

D'autre part, une échelle supérieure s_{max} est fixée par la taille de l'échantillon (pour notre grille, $s_{max} = 10000 \text{ km.s}^{-1}$). L'information présente dans les composantes $\Psi^{(s)}$ du potentiel au dessus de cette échelle supérieure s_{max} est alors complètement contenue dans

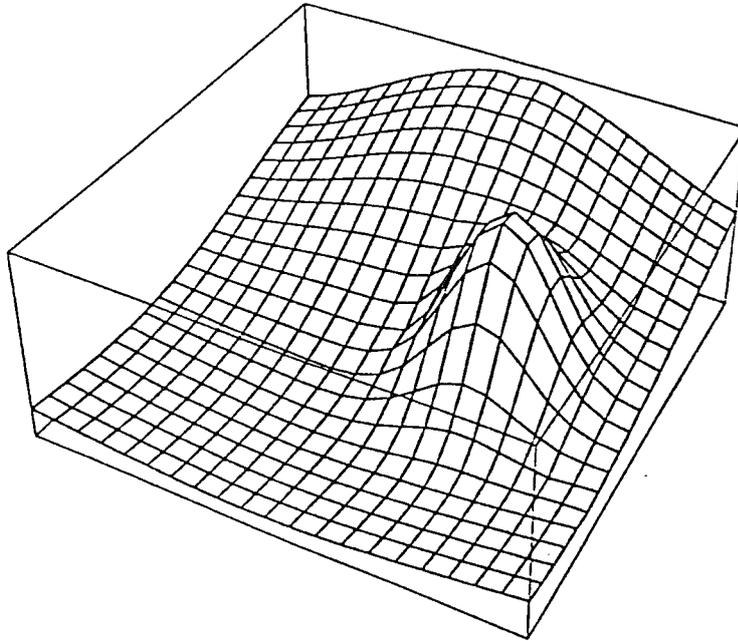


Figure E.6 : Les valeurs du potentiel des vitesses reconstruit Ψ_{rec} sur la coupe P_1 .

les $\Psi^{(s)}$ des échelles inférieures.

En pratique donc, nos données possèdent une échelle inférieure s_{min} et supérieure s_{max} de coupure. La forte redondance d'information présente dans la transformée en ondelettes continue nous permet de transformer l'intégrale en somme dans l'équation (E.39) avec une très bonne approximation. En fait, nous montrons dans notre étude que 6 échelles distribuées logarithmiquement entre $s_{inf} = 200 \text{ km.s}^{-1}$ et $s_{sup} = 2000 \text{ km.s}^{-1}$ nous fournissent une bonne reconstruction du potentiel Ψ .

Reconstruction de $\Psi(x)$:

Nous présentons donc figure (E.6) les valeurs du potentiel des vitesses reconstruit Ψ_{rec} sur la coupe P_1 . L'intégrale de l'équation (E.39) a donc été remplacée par une somme sur les échelles échantillonnées logarithmiquement entre $s_{inf} = 200 \text{ km.s}^{-1}$ et $s_{sup} = 2000 \text{ km.s}^{-1}$. Qualitativement, nous voyons que cette reconstruction est bonne (à comparer

avec le potentiel simulé Ψ (figure (E.3)). On retrouve en effet correctement les positions des 2 attracteurs ainsi que leurs échelles de cohérence. Des distorsions apparaissent aux limites de la grille, dues aux effets de bords.

L'étude quantitative de notre reconstruction s'avère néanmoins plus délicate puisqu'il faut définir une quantité *ad hoc* représentative de la validité de notre méthode. Nous avons choisi la fonction $S(\Psi, \Psi_{rec})$ s'exprimant de la façon suivante (la somme étant prise ici sur les points du plan P_1) :

$$S(\Psi, \Psi_{rec}) = \left(\frac{\sum [(\Psi^i - \langle \Psi \rangle) - (\Psi_{rec}^i - \langle \Psi_{rec} \rangle)]^2}{\sum [\Psi^i - \langle \Psi \rangle]^2} \right)^{1/2} \quad (E.48)$$

$$\text{avec } \langle \Psi \rangle = 1/N \sum_{i=1}^N \Psi^i$$

Le dénominateur de l'équation (E.48) est représentatif de l'amplitude du signal sur le plan P_1 alors que le numérateur quantifie la différence entre les potentiels simulé Ψ et reconstruit Ψ_{rec} . Notre indicateur $S(\Psi, \Psi_{rec})$ possède la propriété souhaitable d'être invariant si le potentiel Ψ est multiplié par une constante ou si l'on modifie sa valeur moyenne. La valeur de la quantité $S(\Psi, \Psi_{rec})$ sera donc d'autant plus proche de 0 que la procédure de reconstruction du potentiel des vitesses est valable.

Nous avons calculé $S(\Psi, \Psi_{rec})$ pour les différents $\Psi_{rec}(s_{lim}) = \sum_{s_{lim}}^{s_{sup}} \Psi_{rec}^{(s)}$ intervenant dans les étapes du calcul du potentiel reconstruit Ψ_{rec} . $\Psi_{rec}(200)$ correspond en fait à notre potentiel reconstruit final Ψ_{rec} (sommé sur les 6 échelles), $\Psi_{rec}(502)$ n'est plus sommé que sur 4 échelles (entre $s_{lim} = 502 \text{ km.s}^{-1}$ et $s_{sup} = 2000 \text{ km.s}^{-1}$) et $\Psi_{rec}(2000)$ ne contient plus que l'information à l'échelle $s_{sup} = 2000 \text{ km.s}^{-1}$ (un terme seulement intervient dans la somme). Nous présentons les valeurs de $S(\Psi, \Psi_{rec})$ pour ces différents Ψ_{rec} dépendant de s_{lim} dans le tableau suivant :

Nombre d'échelles	6	5	4	3	2	1
s_{lim} en km.s^{-1}	200	317	502	796	1262	2000
$S(\Psi, \Psi_{rec})$	0.14	0.12	0.14	0.43	0.71	0.86

Nous pouvons tirer 2 conclusions de ces résultats.

D'une part, il est rassurant de constater que la quantité $S(\Psi, \Psi_{rec})$ diminue lorsqu'on

ajoute progressivement les contributions $\Psi_{rec}^{(s)}$ du potentiel des vitesses aux petites échelles. Ceci confirme la stabilité de notre procédure de reconstruction. Il faut néanmoins remarquer que l'adjonction de la dernière composante $\Psi_{rec}^{(200)}$ rehausse légèrement la valeur du test. Nous y voyons l'influence de la résolution finie de la grille. A partir de cette échelle $s = 200 \text{ km.s}^{-1}$ donc, l'opération classique de remplacer les intégrales par des sommes dans l'équation (E.46) devient dangereuse. On voit cependant que la transformée en ondelettes nous fournit un moyen naturel de contrôler les problèmes dus au lissage obligatoire des données. En effet, on stoppera en pratique la reconstruction à l'échelle $s = 317 \text{ km.s}^{-1}$.

D'autre part, la valeur de 0.14 pour $S(\Psi, \Psi_{rec}^{(200)})$ (Ψ_{rec} sommé sur les 6 échelles) est satisfaisante. Ceci nous montre que les effets de bords n'influencent que modestement notre procédure de reconstruction. De plus, nous voyons qu'un nombre réduit d'échelles est en pratique suffisant afin de reconstruire le potentiel des vitesses avec une bonne approximation. Nous voulons ajouter que le test $S(\Psi, \Psi_{rec})$ doit normalement s'appliquer sur tous les points de la grille (et non pas qu'aux points du plan P_1). En fait, nous avons appliqué le test sur un autre plan P_2 (passant par les centre des 2 attracteurs et perpendiculaire à P_1 (voir article en annexe)). Pour cette coupe P_2 , $S(\Psi, \Psi_{rec}^{(200)}) = 0.07$ ce qui est légèrement meilleur que pour le plan P_1 . Cette différence est semble-t-il due aux symétries présentes dans le plan P_2 qui tendent à compenser les erreurs produites par les effets de bords.

Nous concluons donc d'après ces résultats que notre procédure de reconstruction du potentiel des vitesses à partir du champ des vitesses radiales est fiable. Les erreurs introduites par les effets de bords ainsi que par la discrétisation de l'intégrale sur les échelles (équation (E.39)) sont en tout cas négligeables comparées aux erreurs dues à l'incertitude sur l'évaluation des distances (et donc des vitesses) des galaxies (voir section III.3).

La soustraction du cône :

Notre méthode ne nous permet en pratique que d'analyser des champ de vitesses radiales de moyenne nulle sur l'espace. Or le champ des vitesses radiales réellement observé n'est généralement pas nul en moyenne et il en va de même pour nos données simulées (on notera $\langle V_r \rangle$ cette valeur moyenne). Il est alors nécessaire de corriger la valeur du potentiel reconstruit par une fonction conique représentant la contribution au potentiel de l'intégrale d'un champ de vitesses radiales constant le long de la ligne de visée

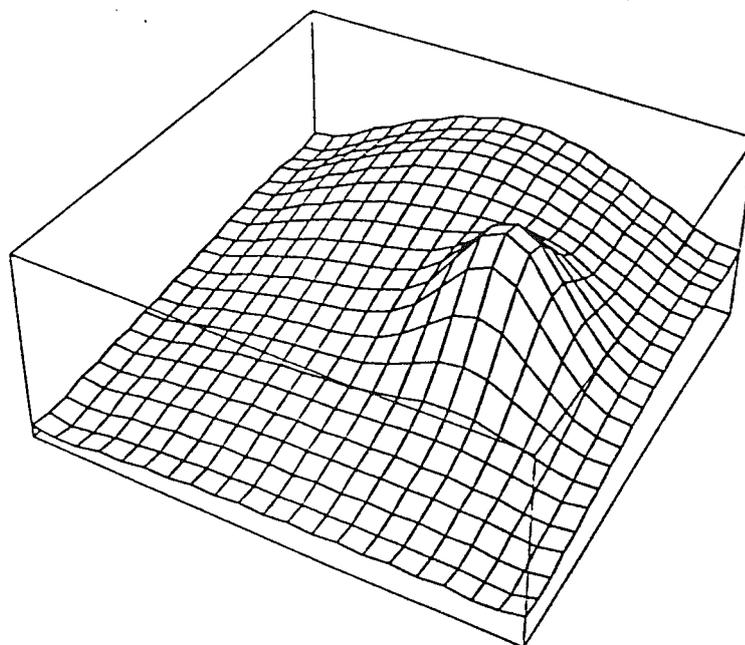


Figure E.7 : Le potentiel $[\Psi_{rec} - \Psi_C]$ sur la coupe P_1 .

(voir équation (E.36)). Il faut ainsi soustraire à la valeur du potentiel reconstruit $\Psi_{rec}(\mathbf{x})$ un "potentiel" $\Psi_C(\mathbf{x})$ qui prend la forme suivante :

$$\Psi_C(\mathbf{x}) = \langle V_r \rangle \|\mathbf{x}\| \quad (E.49)$$

Nous présentons figure (E.7) les effets de cette correction sur le potentiel des vitesses reconstruit à partir de nos données simulées (voir figure (E.6)). On remarque que la soustraction de ce "potentiel" conique améliore légèrement les distorsions aux bords présentes sur la figure (E.6). L'application du test $S(\Psi, \Psi_{rec})$ au potentiel reconstruit corrigé nous confirme d'ailleurs cette impression ($S(\Psi, \Psi_{rec}^{(200)} - \Psi_C) = 0.1$).

II.2.2. Reconstruction d'autres quantités :

Nous avons vu dans la section précédente que notre méthode nous permet de recon-

truire avec une précision satisfaisante le potentiel des vitesses $\Phi(\mathbf{x})$ à partir du champ des vitesses radiales $V_r(\mathbf{x})$. Pour se faire, nous avons défini une ondelette généralisée $G_{s,u}(\mathbf{x})$ (équation (E.41)) ainsi que la fonctionnelle P_G^{-1} (équation (E.43)) qui lui est associée. Cette opération nous permet d'introduire la transformation linéaire $P_G^{-1} \circ W_g$, caractérisée par son noyau généralisé $N_s(\mathbf{x}, \mathbf{y})$ (équation (E.45)), qui au champ des vitesses radiales $V_r(\mathbf{x})$ fait correspondre le potentiel des vitesses $\Phi(\mathbf{x})$ (équation (E.44)).

Nous présentons figures (E.8a) et (E.8b) la projection sur un plan de notre ondelette généralisée $G_{s,u}(\mathbf{x})$ pour deux valeurs différentes du paramètre d'échelle s (et le noyau généralisé $N_s(\mathbf{x}, \mathbf{y})$ figures (E.9a) et (E.9b)). La symétrie sphérique de l'ondelette d'analyse $g_{s,u}(\mathbf{x})$ (voir figure (E.4)) est remplacée ici par une symétrie plus complexe faisant intervenir l'observateur O ainsi que le point \mathbf{x} où l'on reconstruit le potentiel. La fonction $G_{s,u}(\mathbf{x})$ n'est plus forcément une ondelette (elle ne vérifie pas généralement la condition d'admissibilité (E.24)). La forme de $G_{s,u}(\mathbf{x})$ va donc être fixée d'une part par l'ondelette mère $g_{s,u}(\mathbf{x})$ et d'autre part par l'opération associée à l'obtention de Φ à partir de V_r , soit une intégration sur la ligne de visée.

On voit alors que notre méthode nous permet de reconstruire d'autres quantités que le potentiel des vitesses. En fait, on peut associer à toute quantité Q dérivant linéairement du potentiel Φ (et donc du champ V_r dans l'hypothèse d'irrotationnalité) une ondelette généralisée $\mathcal{G}_{s,u}(\mathbf{x})$ et sa fonctionnelle T_G^{-1} telle que $Q = [T_G^{-1} \circ W_g] V_r$. La procédure à suivre est alors la suivante. Soit \mathbf{Q}_x l'opérateur linéaire reliant la quantité $Q(\mathbf{x})$ à reconstruire avec le potentiel des vitesses $\Phi(\mathbf{x})$ (par exemple $\mathbf{Q}_x = \Delta_x$ pour la divergence du champ des vitesses, $\mathbf{Q}_x = \nabla_x$ pour le champ lui-même) :

$$Q(\mathbf{x}) = [\mathbf{Q}_x \Phi](\mathbf{x}) \quad (E.50)$$

On associe alors à cette quantité Q l'ondelette généralisée $\mathcal{G}_{s,u}(\mathbf{x})$ suivante :

$$\mathcal{G}_{s,u}(\mathbf{x}) = \mathbf{Q}_x G_{s,u}(\mathbf{x}) = \int_0^1 dl \mathbf{Q}_x g_{s,u}(l \mathbf{x}) \quad (E.51)$$

Et la transformation T_G^{-1} associée :

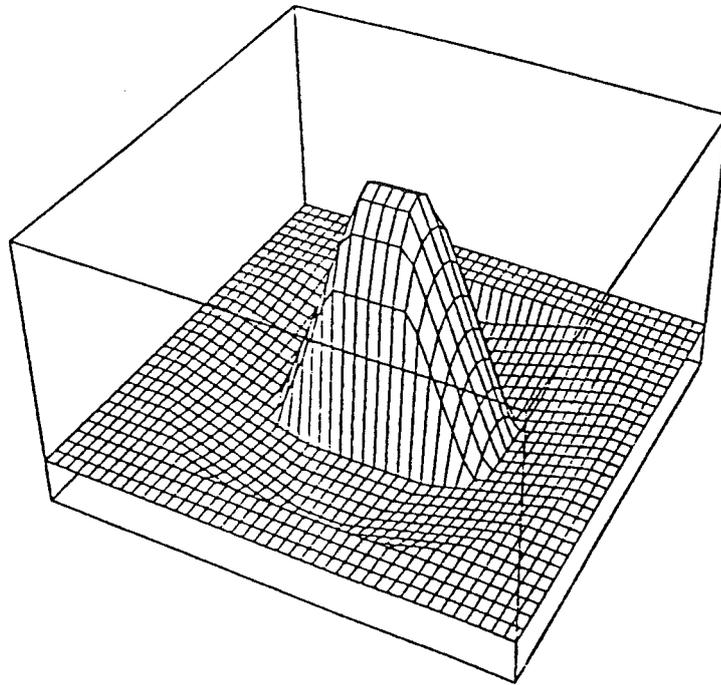


Figure E.8a : Projection 2D de l'ondelette généralisée $G_{s,u}(x)$ (x fixé et u variant pour une échelle donnée s_1).

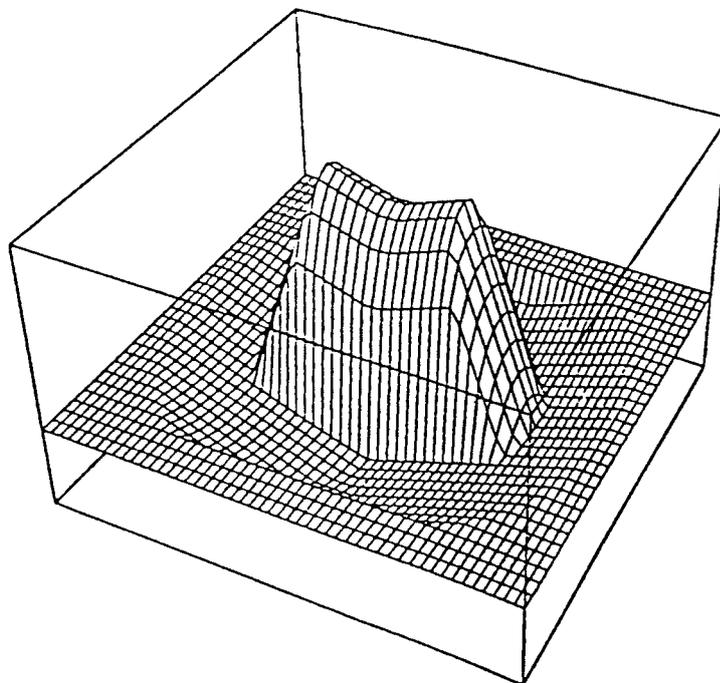


Figure E.8b : Projection 2D de l'ondelette généralisée $G_{s,u}(x)$ (x fixé et u variant pour une échelle donnée $s_2, s_2 > s_1$).

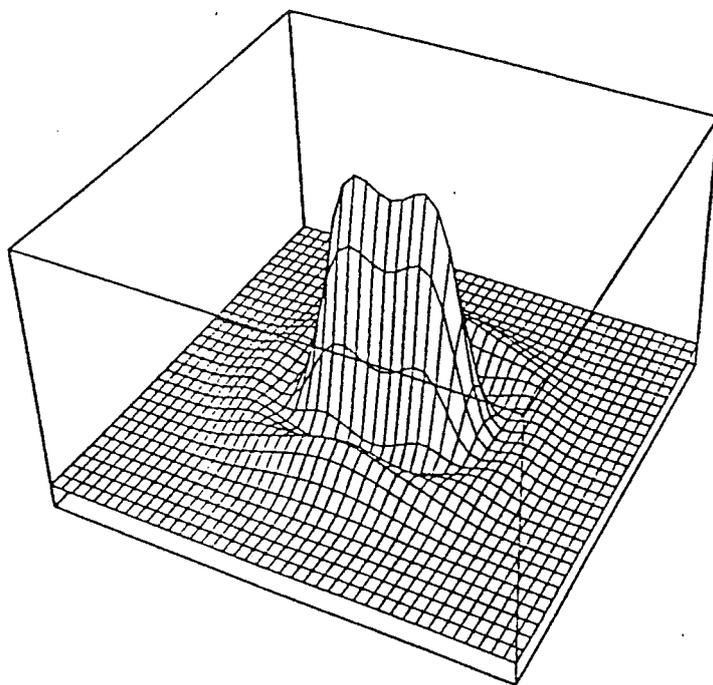


Figure E.9a : Projection 2D du noyau généralisé $N_s(x, y)$ (x fixé et y variant pour une échelle donnée s_1).

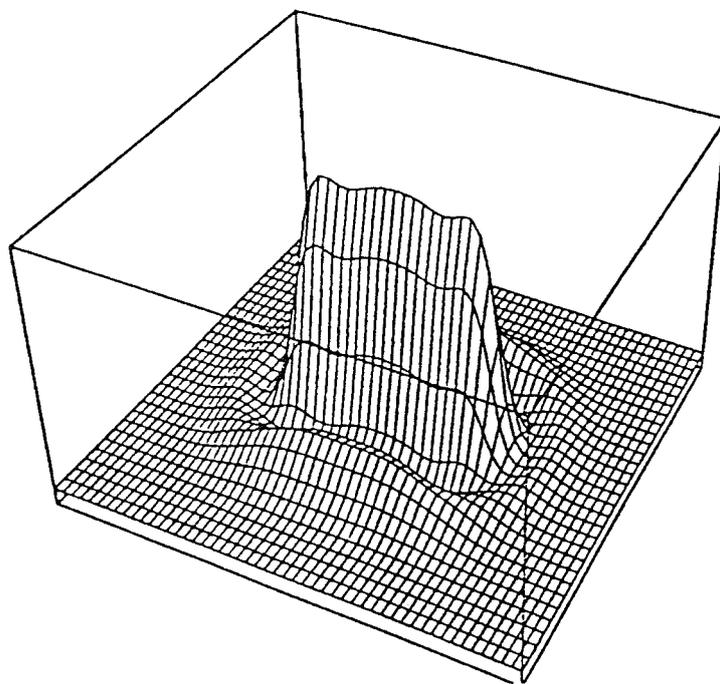


Figure E.9b : Projection 2D du noyau généralisé $N_s(x, y)$ (x fixé et u variant pour une échelle donnée s_2 , $s_2 > s_1$).

$$T_G^{-1} : \begin{cases} L^2(\mathbb{R}^{*+} \times \mathbb{R}^3) \longrightarrow L^2(\mathbb{R}^3) \\ h \longmapsto T_G^{-1}h : T_G^{-1}h(\mathbf{x}) = \frac{1}{\sqrt{C_g}} \int_{\mathbb{R}^{*+} \times \mathbb{R}^N} k_s ds d^3\mathbf{u} h(s, \mathbf{u}) \mathcal{G}_{s, \mathbf{u}}(\mathbf{x}) \end{cases} \quad (E.52)$$

Nous pouvons alors reconstruire la quantité $Q(\mathbf{x})$ à partir du champ des vitesses radiales $V_r(\mathbf{x})$ ($Q = [T_G^{-1} \circ W_g] V_r$). Et nous appliquerons en pratique la procédure simplifiée en associant à la transformation linéaire $T_G^{-1} \circ W_g$ le noyau généralisé $\mathcal{N}_s(\mathbf{u}, \mathbf{u}')$:

$$\mathcal{N}_s(\mathbf{u}, \mathbf{u}') = \frac{1}{C_g} \int_{\mathbb{R}^3} d^3\mathbf{x} \overline{g_{s, \mathbf{u}}(\mathbf{x})} g_{s, \mathbf{u}'}(\mathbf{x}) \quad (E.53)$$

Chaque quantité Q possédera alors son ondelette généralisée $\mathcal{G}_{s, \mathbf{u}}(\mathbf{x})$ ainsi que son noyau $\mathcal{N}_s(\mathbf{x}, \mathbf{y})$ généralisé que l'on pourra calculer numériquement ou analytiquement lorsque ceci est possible.

Cette procédure de reconstruction des quantités reliées linéairement avec le potentiel des vitesses présente un intérêt particulier. En fait, nous n'avons pas à évaluer préalablement le potentiel Φ puisque nous dérivons ces quantités directement des données du champ des vitesses radiales propres V_r .

Nous avons vu précédemment (section II.2.1) que le caractère discret des données d'entrées introduisait un lissage obligatoire de la fonction à reconstruire. Imaginons qu'à partir du potentiel reconstruit, nous désirions évaluer une quantité comme la divergence du champ par exemple. Il serait alors nécessaire de dériver 2 fois le potentiel reconstruit. On sait que ces opérations de dérivation consistent à extraire du potentiel reconstruit de l'information fortement localisée, justement aux échelles où le lissage obligatoire des données intervient. Il est alors clair que la valeur estimée de la divergence va dépendre fortement du choix de la technique de lissage.

Nous évitons complètement ce problème lors de notre procédure de reconstruction. Toutes ces opérations de dérivation, intégration et autres sont en effet appliquées sur les fonctions ondelettes. Le problème est alors proprement éliminé puisque toutes les opérations sont maintenant accomplies sur des fonctions que l'on peut choisir continues et dérivables à notre guise. C'est un des points forts de notre méthode, directement lié aux puissantes propriétés que possède la transformée en ondelettes. Bien entendu, nous n'éliminons pas le lissage obligatoire dû au phénomène de discrétisation des données

d'entrée, mais ce lissage est adapté à chaque quantité à reconstruire et possède les mêmes facilités de contrôle que dans le cas de la reconstruction du potentiel des vitesses par exemple. On arrête en pratique la reconstruction d'une quantité à l'échelle inférieure pour laquelle l'influence de l'échantillonnage commence à se faire sentir.

II.2.3. Séparation des échelles :

Notre méthode nous offre de plus la possibilité de séparer les contributions au champ des vitesses pour chaque échelle. Les composantes $\Phi^{(s)}$ intervenant dans l'équation (E.39) peuvent en effet être interprétées comme l'expression du potentiel Φ à l'échelle s .

Nous présentons figures (E.10) et (E.11) les fonctions $\Psi_{rec}^{(s)}$ pour les échelles $s = 2000$ km.s⁻¹ et $s = 502$ km.s⁻¹. On remarque que notre procédure permet effectivement de distinguer le potentiel Ψ_1 qui agit à grande échelle ($\sigma_1 = 4500$ km.s⁻¹) du potentiel Ψ_2 de longueur de cohérence plus modeste ($\sigma_2 = 2000$ km.s⁻¹). Cette propriété est très intéressante en pratique si l'on veut séparer des flots de vitesses cohérents à grande échelle comme le "Grand Attracteur" (qui domine la cinématique des catalogues actuellement connus) d'avec des attracteurs locaux (le mouvement de chute sur l'amas de Virgo par exemple) que l'on doit pouvoir traiter en principe d'un point de vue statistique.

Notre procédure de reconstruction ne nous permet néanmoins pas de séparer rigoureusement les échelles puisque nous utilisons en fait la transformée en ondelettes continue qui présente une redondance d'information inévitable. Nous pouvons d'ailleurs visualiser cette propriété de redondance sur les figures (E.10) et (E.11) (sur le côté gauche) où la cavité présente à l'échelle $s = 2000$ km.s⁻¹ est compensée par l'excroissance apparaissant aux échelles plus petites. De plus, les composantes $\Phi^{(s)}$ dépendent de la forme de l'ondelette d'analyse. Si notre décomposition par échelles n'est donc pas parfaite, les figures (E.10) et (E.11) nous apprennent que l'analyse des composantes individuelles $\Phi^{(s)}$ n'est toutefois pas sans intérêt.

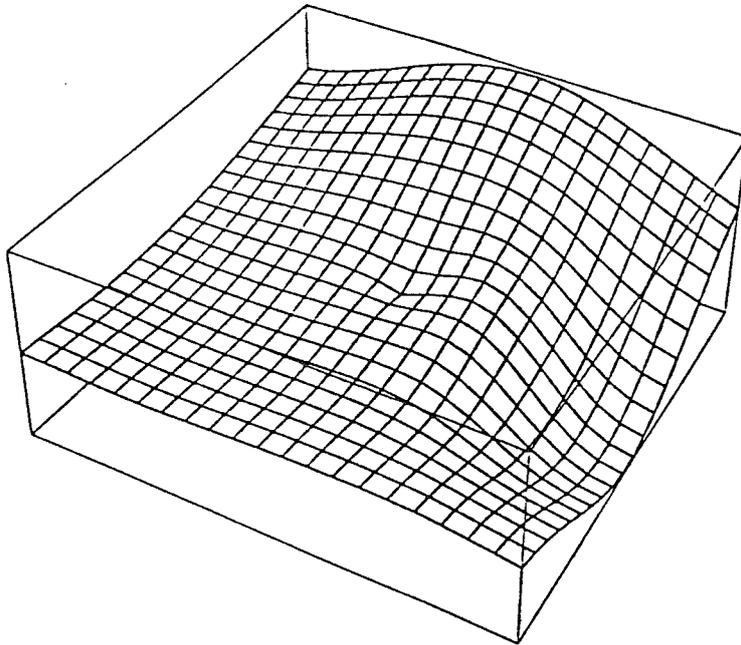


Figure E.10: La composante $\Psi_{rec}^{(s)}$ du potentiel des vitesses à l'échelle $s = 2000 \text{ km.s}^{-1}$.

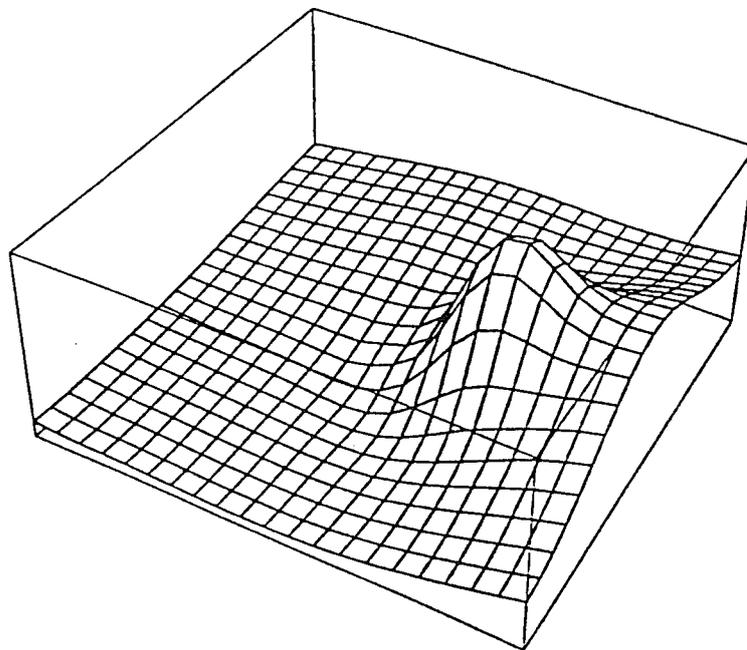


Figure E.11 : La composante $\Psi_{rec}^{(s)}$ du potentiel des vitesses à l'échelle $s = 502 \text{ km.s}^{-1}$.

III. NOTRE METHODE COMPAREE A "POTENT" :

Nous proposons dans cette section une comparaison entre notre méthode et la méthode POTENT de E. Bertschinger et A. Dekel (1989). Dans la méthode POTENT, il est nécessaire de lisser le champ des vitesses radiales observées afin de reconstruire le potentiel des vitesses, les autres quantités étant ensuite dérivées de ce potentiel. Le lissage du champ des vitesses est effectué par l'intermédiaire d'une fenêtre tensorielle, ce qui rend la procédure de reconstruction non linéaire. En fait, notre approche est toute autre. Notre méthode et POTENT sont ainsi complémentaires, dans le sens que les techniques que nous employons afin de surmonter les divers problèmes présents lors du traitement des données observées sont complètement différentes.

D'une part, la méthode POTENT effectue un lissage du champ des vitesses afin d'éliminer la composante rotationnelle du champ présente à petite échelle. Ce lissage nécessite alors l'introduction de paramètres arbitraires comme la forme de la fenêtre de lissage ainsi que sa taille. Notre procédure nous fournit un moyen naturel de contrôler l'influence de ce genre de problème en agissant sur l'échelle inférieure de reconstruction des quantités désirées.

D'autre part, les données réelles possèdent des erreurs intrinsèques (comme les erreurs sur les distances des galaxies (voir partie B)). Ces erreurs peuvent alors faire apparaître des biais dans l'estimation du potentiel des vitesses. Le caractère tensoriel propre à la méthode POTENT rend difficile le traitement de ces biais. L'un des points forts de notre méthode, c'est que notre procédure de reconstruction par l'application de la transformation générale $T_G^{-1} \circ W_G$ est totalement linéaire. La prise en compte des erreurs s'en verra ainsi notamment simplifiée.

Enfin, les catalogues de vitesses actuellement disponibles sont soumis à des inhomogénéités spatiales importantes, c'est-à-dire que nous ne possédons pas d'information sur le champ des vitesses pour certaines régions de l'espace. La méthode POTENT propose une approche statistique afin de régler ce problème. Nous y préférons une approche géométrique basée sur l'analyse du support spatial du champ des vitesses propres. Cette étude est actuellement en cours.

III.1. Le lissage des données :

Sous l'hypothèse d'irrotationnalité, l'équation (E.36) permet de calculer le potentiel des vitesses $\Phi(\mathbf{x})$ à partir du champ des vitesses radiales $V_r(\mathbf{x})$. Néanmoins il est nécessaire de connaître les valeurs du champ aux points situés le long de la ligne de visée. La méthode POTENT dérive ces valeurs en lissant le champ des vitesses. Soient $\{\mathbf{x}_i\}$ et $\{V_r^i\}$ la position et la vitesse radiale des galaxies du catalogue, le champ des vitesses radiales propres lissé $U(\mathbf{x})$ au point \mathbf{x} s'exprime de la façon suivante (voir Dekel et al. (1990)) :

$$U(\mathbf{x}) = \sum_i \tilde{W}(\mathbf{x}, \mathbf{x}_i) V_r^i \quad (E.54)$$

$$\text{avec } \tilde{W}(\mathbf{x}, \mathbf{x}_i) \equiv [\hat{\mathbf{x}} \cdot \mathbf{A}^{-1}(\mathbf{x}) \cdot \hat{\mathbf{x}}_i] W(\mathbf{x}, \mathbf{x}_i), \quad \mathbf{A} \equiv \sum_i W(\mathbf{x}, \mathbf{x}_i) \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i$$

$W(\mathbf{x}, \mathbf{x}_i)$ est la fenêtre de lissage et \mathbf{A} sa quantité tensorielle associée, la fonction lissée étant en effet la projection du champ des vitesses total sur la ligne de visée ($V_r(\mathbf{x}) = \hat{\mathbf{x}} \cdot \mathbf{V}(\mathbf{x})$). On remarque dès à présent que cette procédure de lissage n'est pas linéaire vis-à-vis des données d'entrées (les V_r^i). Il faut maintenant faire un choix pour la fenêtre de lissage. Les auteurs testent plusieurs possibilités pour la fonction $W(\mathbf{x}, \mathbf{x}_i)$. L'une d'entre elles possède la forme suivante :

$$W_F(\mathbf{x}, \mathbf{x}_i) = \frac{V_i}{\sigma_i^2} \exp \left[-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2 R_w^2} \right] \quad (E.55)$$

Avec $V_i = R_4^3(\mathbf{x}_i)$ où $R_4(\mathbf{x}_i)$ est la distance moyenne de l'objet i par rapport à ces 4 plus proches voisins. Le facteur V_i donne donc plus de poids aux galaxies dans les vides. La valeur σ_i quantifie la fiabilité de la valeur mesurée V_r^i pour l'objet i ($\sigma_i \propto r_i$). On affecte alors moins de poids aux galaxies éloignées dont les distances estimées r_i sont plus incertaines. Enfin, la fonction fenêtre a la forme d'une gaussienne de dispersion R_w . La valeur du champ des vitesses lissé $U(\mathbf{x})$ dépend donc du champ observé V_r autour de \mathbf{x} dans un rayon de l'ordre de R_w . On remarque que cette procédure de lissage nécessite le choix arbitraire des quantités comme la taille de la fonction fenêtre R_w ou de la fonction $R_4(\mathbf{x}_i)$.

Notre méthode est sur ce point nettement avantagée. En effet, l'utilisation de la transformée par ondelettes nous fournit un moyen élégant et naturel de lisser les données (voir section II.2.1 et II.2.2). Nous donnons un exemple de la capacité de notre méthode à résoudre ces problèmes de lissage.

On a vu dans la section II.1 que l'hypothèse d'irrotationnalité cessait d'être vérifiée à petite échelle, lorsque les termes non-linéaires présents dans les équations régissant la dynamique des structures dominaient de nouveau. Le champ des vitesses observé possède ainsi une composante turbulente de faible longueur de corrélation. Cette composante peut être éliminée théoriquement si l'on effectue la moyenne du champ sur un petit volume (par exemple en prenant la moyenne des vitesses radiales des galaxies regroupées dans un amas). Néanmoins cette procédure n'est pas réalisable en pratique pour les galaxies isolées. La présence de la composante turbulente à petite échelle va alors perturber le champ des vitesses radiales purement irrotationnel que l'on désire étudié.

Nous avons introduit dans nos simulations une telle perturbation, modélisée par un bruit blanc en chaque point de dispersion σ_B telle que ($\sigma_B^2 = 1/V \int_V V_r^2 dV$, V le volume défini par notre grille de données). Nous montrons figure (E.12) la reconstruction du potentiel des vitesses Ψ_{rec}^B jusqu'à l'échelle inférieure limite $s = 200 \text{ km.s}^{-1}$ (sommée sur 6 échelles). La reconstruction du potentiel est effectivement détériorée par la présence du bruit blanc. Le test S (voir section II.1) appliqué entre le potentiel initial non bruité Ψ^{NB} et le potentiel reconstruit à partir du champ des vitesses bruité Ψ_{rec}^B nous donne la valeur $S(\Psi^{NB}, \Psi_{rec}^B) = 0.2$, soit 0.6 de plus que la valeur du test appliqué au potentiel reconstruit à partir du champ des vitesses non bruité Ψ_{rec}^{NB} . Toutefois, nous présentons figure (E.13) les valeurs de Ψ_{rec}^B reconstruit cette fois-ci jusqu'à l'échelle $s = 317 \text{ km.s}^{-1}$ (somme sur 5 échelles seulement). On peut voir que le bruit disparaît complètement, ce que confirme le test S ($S(\Psi^{NB}, \Psi_{rec}^B) = 0.13$) à comparer à $S(\Psi^{NB}, \Psi_{rec}^{NB}) = 0.12$). Ainsi le choix de notre échelle inférieure de coupure nous permet en pratique de gérer très efficacement ces problèmes de lissage.

III.2. L'influence des erreurs en distance :

Nous avons montré dans la partie B que la détermination de la distance des galaxies était soumise à une incertitude intrinsèque, qui se répercutait sur l'estimation de la vitesse

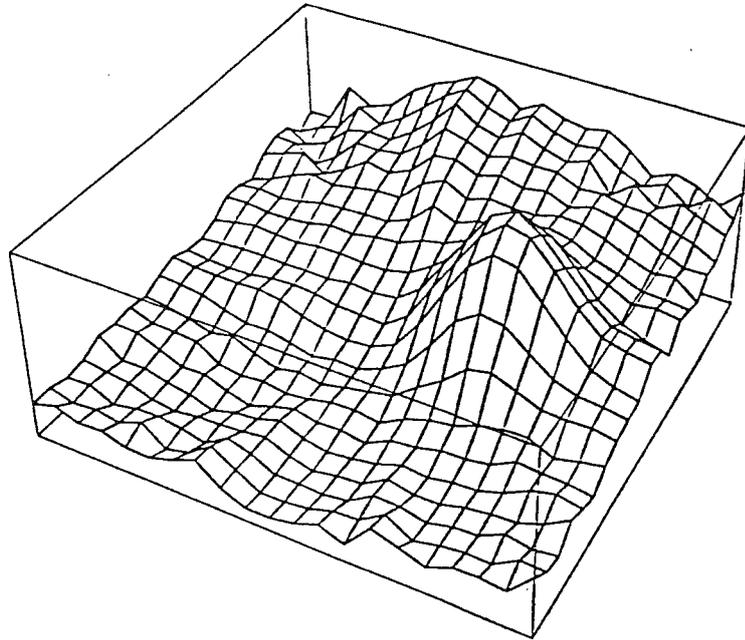


Figure E.12: Le potentiel Ψ_{rec}^B (les vitesses étant perturbées par un bruit blanc).
(Reconstruction jusqu'à l'échelle inférieure $s = 200 \text{ km.s}^{-1}$)

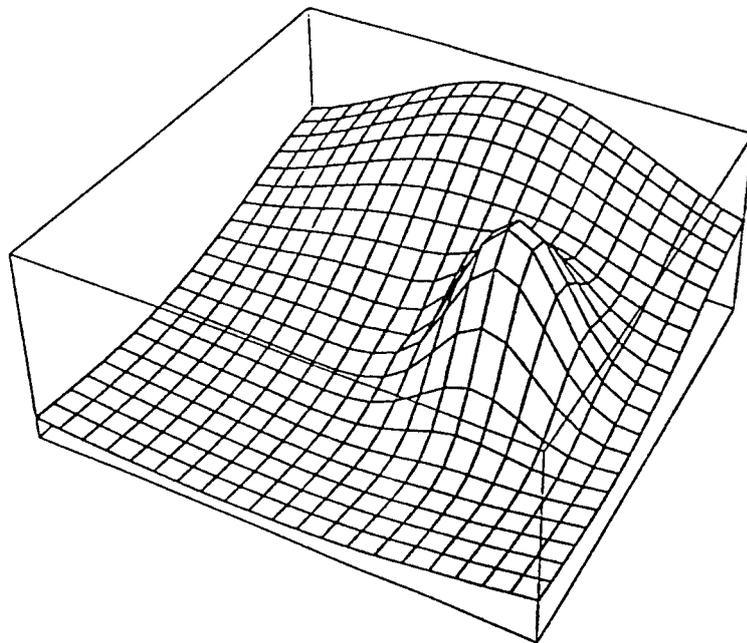


Figure E.13 : Le potentiel Ψ_{rec}^B (les vitesses étant perturbées par un bruit blanc).
(Reconstruction jusqu'à l'échelle inférieure $s = 317 \text{ km.s}^{-1}$)

radiale propre V_r . Ces incertitudes sur les distances risquent alors d'introduire des biais lors de la reconstruction du potentiel des vitesses à partir du champ des vitesses radiales. De plus, ces erreurs sur la distance des galaxies sont responsables d'une incertitude sur l'estimation des valeurs du potentiel, de variance importante en pratique.

L'analyse des effets de ces erreurs en distance sur la méthode POTENT est complexe. On a vu précédemment que la méthode nécessitait préalablement le lissage du champ des vitesses radiales observé. Ce lissage présente un caractère tensoriel qui rend effectivement difficile l'estimation des biais et de la variance dus aux erreurs en distance. En fait, les auteurs utilisent une méthode perturbative sur les quantités tensorielles impliquées lors du lissage des données (voir Dekel et *al.* (1990)). Il est alors possible d'estimer théoriquement l'amplitude du biais qui est négligeable, si le champ des vitesses vérifie certaines hypothèses. L'estimation analytique de la variance s'avère être un problème complexe et les auteurs ont recours à des simulations afin de quantifier les effets des erreurs en distance sur la reconstruction du potentiel des vitesses. En fait, les résultats montrent que l'incertitude sur les distances des galaxies est l'une des majeures sources d'erreurs pour la méthode POTENT.

L'un des points forts de notre méthode, c'est sa linéarité par rapport au champ des vitesses radiales. Le traitement des effets dus aux incertitudes présentes dans la détermination des distances s'en voit donc notamment simplifié. Si le champ des vitesses varie lentement, on voit que le champ des vitesses radiales observé peut être décomposé entre le champ vrai plus le champ des vitesses radiales perturbateur induit par les erreurs en distance (un champ aléatoire à symétrie sphérique autour de l'observateur et dont la variance est proportionnelle à la distance à l'observateur). La linéarité de notre procédure de reconstruction permet de traiter séparément ces deux champs (le champ vrai et le champ perturbateur). La valeur reconstruite du potentiel des vitesses ne souffrira donc pas de biais.

Néanmoins les distances des galaxies vont elles aussi varier. Sur nos données simulées par exemple, il en résulte une distorsion de la grille des points données (le support). Cette distorsion peut créer des problèmes lors de la reconstruction du potentiel des vitesses. Nous avons donc simulé quelques échantillons perturbés par des erreurs en distance (de dispersion $\Delta = 0.15$) à partir de notre potentiel Ψ (voir section II.2.1). De plus, nous avons réglé l'amplitude du potentiel Ψ pour que le champ des vitesses radiales engendré soit comparable en magnitude au champ des vitesses radiales réellement observé. Ainsi

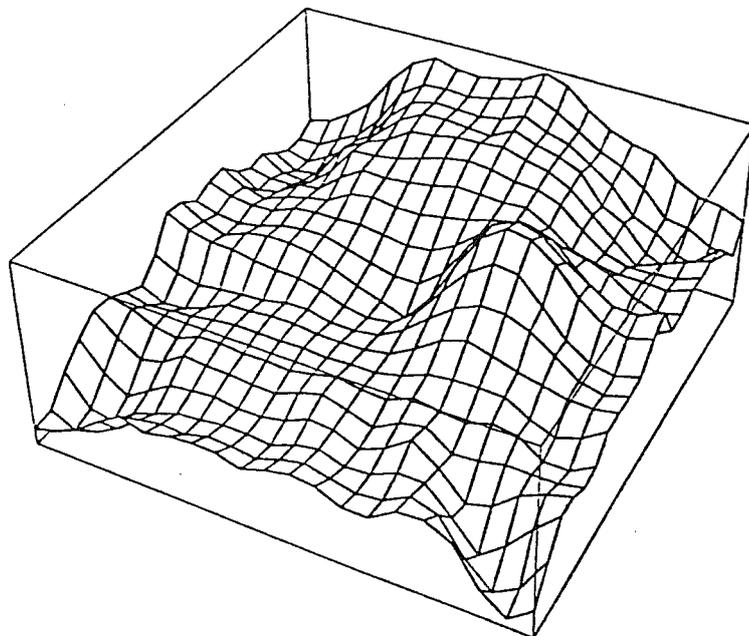


Figure E.14 : Le potentiel Ψ_{rec} pour un échantillon bruité par des erreurs en distance

nos simulations nous permettent de quantifier quelle est l'influence des erreurs en distance pour la reconstruction du potentiel des vitesses réel.

Nous avons vérifié sur nos simulations (en faisant une moyenne) que les erreurs en distances n'introduisent pas de biais dans notre procédure de reconstruction. La propriété de linéarité dont jouie notre méthode est donc appréciable. D'autre part, nous montrons figure (E.14) la reconstruction du potentiel Ψ_{rec} pour une de nos simulations bruitées. On voit que les incertitudes dans la détermination des distances des galaxies vont être une des principales sources d'erreurs impliquées lors de l'évaluation du potentiel des vitesses ou d'autres quantités.

III.3. L'inhomogénéité spatiale des catalogues :

Les données réellement observées ne sont pas généralement réparties sur une grille. En fait, certaines régions de l'espace sont vides de galaxies. Il est alors impossible de déduire le champ des vitesses dans ces régions. L'opération qui consiste à remplacer les intégrales par des sommes sur les points de l'échantillon dans l'équation (E.46) devient alors incorrecte. Il faudra donc d'une part corriger l'influence de ces inhomogénéités spatiales sur la valeur des intégrales intervenant dans l'équation (E.46) et d'autre part, évaluer quelles sont les incertitudes impliquées par ce manque d'information dans certaines régions de l'espace sur l'estimation du potentiel des vitesses.

La méthode POTENT inclue le terme en $R_4(x_i)$ dans la fenêtre de lissage afin de restaurer l'homogénéité spatiale des données. Plus une galaxie est dans une zone de basse densité, plus son poids est important. La quantification des incertitudes sur la valeur du champ lissé dues au manque d'information dans les régions vides de l'espace est traitée par la méthode POTENT de façon probabiliste. Une fonction de sélection est modélisée afin de rendre compte de la distribution des galaxies dans l'échantillon réellement observé. Ensuite, les auteurs simulent d'autres échantillons possédant cette fonction de sélection et le même champ des vitesses que le catalogue réel. L'analyse de la variance du champ lissé de ces échantillons (comparé au champ lissé pour l'échantillon réel) est alors un indicateur de l'erreur commise sur la valeur du champ des vitesses lissé.

Cette approche probabiliste du traitement des erreurs dues à l'inhomogénéité spatiale du catalogue ne nous satisfait pas. D'une part, la modélisation de la fonction de sélection des échantillons disponibles est un problème délicat encore très mal connu. D'autre part, la méthode POTENT crée un champ de vitesses lissé artificiel dans les régions de l'espace où l'information est physiquement manquante. Nous préférons utiliser une approche plus géométrique du problème basée sur l'analyse du support du champ des vitesses (les positions spatiales des galaxies).

Notre méthode est bien sûr sensible aux inhomogénéités spatiales présentes dans les échantillons actuellement disponibles. Nous montrons figure (E.15) les résultats de la reconstruction du potentiel Ψ_{rec} (jusqu'à l'échelle $s = 502 \text{ km.s}^{-1}$) lorsqu'on remplace notre grille de données par les positions réelles des points du catalogue de Faber et *al.* (1989) (voir partie B). Il est clair que nous n'analysons pas seulement le champ des vitesses. En

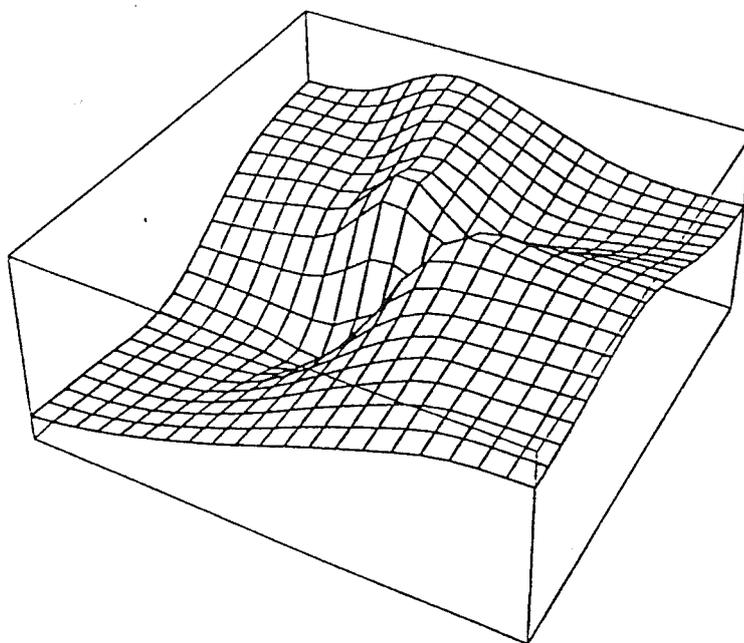


Figure E.15 : Le potentiel Ψ_{rec} (les points donnés sont le support du catalogue de Faber et *al.*)

fait, nos valeurs reconstruites pour le potentiel sont aussi caractéristiques du support des points. Nous distinguons deux sources d'erreurs responsables de ces effets indésirables.

D'une part, notre méthode est sensible à l'inhomogénéité spatiale du support, et ceci indépendamment du problème du manque d'information. En effet, imaginons que l'information sur le champ des vitesses soit présente mais que les points donnés ne soient pas répartis uniformément dans l'espace (il existe par exemple des surdensités). Notre méthode ne nous permet pas telle quelle de reconstruire le potentiel des vitesses, par exemple parce que l'intégrale (remplacée par la sommes sur les points du support) des ondelettes ne sera pas nulle comme le réclame la condition d'admissibilité (E.24). Il est possible de corriger cet effet en pondérant les points donnés afin de restaurer l'homogénéité de la distribution spatiale comme ceci est fait dans la méthode POTENT (le facteur $R_1(x_i)$). Néanmoins, nous souhaitons effectuer cette correction en restant dans le cadre des ondelettes afin de conserver les avantages de notre méthode (linéarité et contrôle naturel du

lissage). Nous sommes actuellement en train de mettre au point différentes techniques permettant cette correction à base d'ondelettes de l'influence des inhomogénéités spatiales sur notre procédure de reconstruction.

D'autre part, une deuxième source d'erreurs vient du manque d'information sur le champ des vitesses dans certaines zones de l'espace. La méthode POTENT introduit une information artificielle dans ces vides. Nous préférons aborder le problème différemment. En fait, ce sont les bonnes propriétés de localisation à la fois spatiale et spectrale que possède la transformée en ondelettes qui vont guider notre approche.

Aux grandes échelles, l'information sur le champ des vitesses radiales est accessible, même si le support de l'information (c'est-à-dire les positions spatiales des galaxies du catalogue) n'est pas reparti uniformément dans l'espace. En appliquant la technique mentionnée dans le paragraphe précédent, on pourra alors estimer les composantes du potentiel des vitesses $\Phi^{(s)}$ aux grandes échelles (la technique en cours d'élaboration permettra en effet de corriger la reconstruction du potentiel dans le cas d'une distribution non-uniforme des données).

Plus nous descendons vers les petites échelles maintenant, plus le manque d'information va se faire sentir dans les zones de basse densité du support du catalogue. Ainsi pour chacun de ces vides, il existe une échelle de "coupure" s_c sous laquelle nous ne possédons plus d'information sur le champ des vitesses. Cette échelle de coupure s_c est approximativement égale à la taille du vide considéré. Plutôt que de reconstruire une valeur artificielle du potentiel dans ces régions pauvrement échantillonnées du catalogue, nous préférons stopper la reconstruction du potentiel au dessous de l'échelle s_c . On pourra de plus continuer la reconstruction à des échelles inférieures dans les régions bien échantillonnées. Les caractéristiques concernant les échelles de coupure $s_c(\mathbf{x})$ pour les différentes régions de l'espace nous seront données par l'analyse du support des données.

Ainsi notre approche n'ajoute pas d'information artificielle au champ des vitesses radiales observées. L'analyse par ondelettes du champ des vitesses nous permet en fait de définir les composantes du potentiel $\Phi^{(s)}(\mathbf{x})$ pour chaque échelle s . Et l'analyse du support spatial spécifie alors les échelles de coupure $s_c(\mathbf{x})$ au dessus desquelles il est possible d'estimer les composantes $\Phi^{(s)}(\mathbf{x})$, compte-tenu de l'information présente dans le catalogue étudié. Nous sommes convaincus que notre approche nous permet de tirer un maximum d'information des données. Nous appliquerons cette procédure de reconstruction aux données réelles dès que notre technique de correction sera au point.

REFERENCES :

- [1] : E. Bertschinger and A. Dekel : "Recovering the full velocity and density fields from large-scale redshift-distance samples", *Astrophys.J Letters* **336**,L5-L8 (1989)
- [2] : E. Bertschinger, A. Dekel, S.M. Faber, A. Dressler and D. Burstein : "Potential, velocity, and density fields from redshift-distances samples. Applications : Cosmography within 6000 kilometers per second", *Astrophys.J* **364**,370-391 (1990)
- [3] : I. Daubechies, A. Grossmann and Y. Meyer : "Painless nonorthogonal expansions", *J. Math. Phys.* **27**, 1271 (1986)
- [4] : I. Daubechies : "Orthonormal bases of compactly supported wavelets", *Communications on Pure and Applied Mathematics*, Vol. XLI 909-996 (1988)
- [5] : I. Daubechies : "Orthonormal bases of wavelets with finite support : Connection with discrete filters", p. 38-66 in *Wavelets* eds. J.M. Combes, A. Grossmann and P. Tchamitchian, Springer-Verlag (1989)
- [6] : A. Dekel, E. Bertschinger and S.M. Faber : "Potential, velocity, and density fields from sparse and noisy redshift-distances samples : Method ", *Astrophys.J* **364**,349-369 (1990)
- [7] : A. Grossmann, J. Morlet and T. Paul : "Transforms associated to square integrable group representations. I. General results", *J. Math. Phys.* **26**, 2473 (1985)
- [8] : A. Grossmann, R. Kroland-Martinet and J. Morlet : "Reading and understanding continuous wavelet transforms", p. 2-20 in *Wavelets* eds. J.M. Combes, A. Grossmann

and P. Tchamitchian, Springer-Verlag (1989)

[9] : S. Mallat : "A theory for multiresolution signal decomposition: The wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. II No. 7 p. 674 (July 1989)

[10] : Y. Meyer : "Orthonormal wavelets", p. 21-37 in *Wavelets* eds. J.M. Combes, A. Grossmann and P. Tchamitchian, Springer-Verlag (1989)

[11] : P.J.E. Peebles : "The large scale structure of the universe", (1980), Princeton university series, N.J.

[12] : B. Torrèsani : "Wavelet analysis: A tool for signal processing mathematics and physics", Préprint CPT (1989)



IV. ANNEXE :

Wavelet Analysis of Cosmic Velocity Fields

S. Rauzy^{1,2*}, M. Lachièze-Rey¹, and R.N. Henriksen³

¹ Service d'Astrophysique, C.E.N. Saclay, F-91191 Gif sur Yvette Cedex, France

² Centre de Physique Théorique - C.N.R.S., Luminy Case 907, F-13288 Marseille Cedex 9, France

³ Astronomy Group; Dept of Physics, Queen's University at Kingston, Ontario K7L 3N6 Canada

Received ?, 1992; accepted ?, 199?

Abstract. We present a method, based on the properties of wavelet transforms, for inferring an irrotational velocity field from its observed radial component. The method is comparable in its objective to POTENT but it offers in addition a natural smoothing that does not lose information, and a separation of the scales. This has required us to conduct for the first time a wavelet analysis based essentially on the redundancy of the transform components in frequency-displacement space, and in three spatial dimensions.

We test the method, in this paper, on simulated data, which reproduce, with and without noise, the cosmic velocity fields. The results validate the principles of the method, and illustrate its properties. An application to the observed cosmic velocity fields is in progress.

Key words: velocity fields.....

1. Introduction

The study of cosmic velocity fields may provide valuable information concerning the distribution of the mass in the universe, whether visible or not, and of the dynamics which led to the formation of this pattern. Compared to the direct analysis of the position of cosmic objects (*e.g.* galaxies) the analysis of the velocity field offers at least two advantages. Firstly it depends on the distribution of the mass directly, so that 'biasing' of the luminous mass relative to the dark mass does not enter our considerations. Secondly the relevant information can in principle be extracted even without comprehensive knowledge of the spatial distribution of the objects under study. In particular the sample need not be complete.

There are however difficulties peculiar to the use of observed velocities. Proper radial velocities are obtained by subtracting from the observed redshift an estimated distance (multiplied by the Hubble constant H_0). This procedure incurs errors of order 15 – 20%, in addition to possible systematic biases. However in this paper we will not treat errors in the estimation of the proper radial velocities but will assume rather that they are known.

Send offprint requests to: M. Lachièze-Rey

*on leave from C.P.T. Marseille and Boursier MRT at Univ. Aix-Marseille II

A more fundamental difficulty is the fact that only the radial component V_r of the velocity field can be measured. Bertschinger and Dekel (1989) were among the first to recognize that the assumption of purely irrotational velocity fields allows a complete reconstruction. There are good reasons, in the standard models of galaxy and structure formation, to believe that the velocity field is irrotational at large (linear) scales, and even throughout the quasilinear regime described by the Zel'dovitch approximation. Although some doubt about this assumption is possible (*see, e.g.*, Henriksen and Lachièze-Rey 1990), here we accept that it applies at least above some (ill defined) 'non-linear' scale without further justification.

Dekel, Bertschinger and Faber (1990) devised the POTENT method for the reconstruction of such a velocity field, which has already given impressive results (Bertschinger *et al.* 1990). Our confidence in these results is *a posteriori* strengthened by convergent results derived from independent approaches (Yahil 1988; Shaya, Tully and Pierce 1992). However although the principle of POTENT is not in question, there are limitations in practice, and it is difficult to know with precision what confidence one can have in the detailed results. These limitations are all consequences of the fact that the data themselves are far from perfect, together with the difficulty in assessing the manner in which errors in the data propagate into the results.

One difficulty arises from the necessity, in POTENT, of smoothing the data. Smoothing is a dangerous process since it adds and suppresses information in a way difficult to control. Thus a shape for the window function, and a scale have to be assumed, which introduces some arbitrariness. This is particularly critical when a vector (not scalar) field is to be smoothed.

Another particularity of POTENT comes from its manner of reconstructing the potential (the primary quantity to be calculated) step by step, starting from the position of the observer. This propagates any error from nearby positions to the whole reconstructed field. Moreover, since all other quantities are derived from the reconstructed potential by the application of differential operators, errors are amplified for each derived field.

For these reasons we have judged it interesting to devise another method for the reconstruction of a potential velocity field. We certainly do not claim that it is necessarily better than POTENT. We analyse its drawbacks in sections 3.2 and

4.2, as well as in a forthcoming paper (Rausy, Lachièse-Rey and Henriksen, in preparation). But its defects are different than those of POTENT, so that the two methods are complementary.

Our method, based on the use of wavelets, does not require any preliminary smoothing of the data. It is often remarked that wavelet analysis is merely a form of smoothing, but we point out that (i) no information is lost or added during the process; and (ii) the results depend neither on the shape of the wavelet chosen nor on the choice of scales. Moreover, the wavelet transform is not applied here to a vector, but to a scalar field.

More importantly, as it will appear, *the method has a non-local character*: thus the potential is not derived radial step after radial step, but rather by analysing data in a large region (with comparable weights). In fact many other quantities of interest (e.g. the divergence, other components of the velocity field, components of the deformation tensor, etc.) can also be calculated directly from the data without first calculating the potential. This allows an economy of numerical operations, and hence of possible sources of error.

Finally our methods also allow us naturally to isolate the structure on different spatial scales. This is of particular interest in cosmology when the contribution of local attractors (e.g. the Virgo flow), or more distant great attractor type flows, or of very large scale attractors, are to be distinguished. This may have a very important consequence when one is interested in reconstructing the initial conditions of the velocity field (see e.g. Nusser and Dekel 1992, or Lachièse-Rey 1993). Such a reconstruction can only be performed, given the present limitations on our understanding of dynamical and galaxy forming processes, in the linear regime. Our scale sensitive method provides a natural way to select the velocity fluctuations which fall in this regime and thus to better define the subsequent reconstruction of the initial conditions.

In this paper we first recall in section 2 some general properties and definitions concerning wavelets, including their generalization to three dimensions. In section 3 we show how wavelets apply to vector fields, and we concentrate on the case of potential fields. We present the fundamentals of our method and the formulae which allow us to infer the potential from the radial component of the velocity. The construction process is reduced to a single convolution, followed by a sum over a small number of different scales. Section 4 details how the method is to be implemented and applied to discrete data. We test it by applying it to "synthetic data" which simulate a flow similar to the cosmic velocity field. In a forthcoming paper (in preparation) we apply this method to real data and so infer the complete cosmic velocity field.

2. The philosophy of wavelets

The main characteristic of wavelets is to allow a simultaneous study of both the spatial and spectral (in the sense of spatial frequencies) properties of spatial fields, i.e. general functions of the real vectorial variable x . The spectral analysis may be compared with Fourier analysis but, unlike it, positional information is retained.

Essentially, the wavelet transform of a field is a family of fields on the same support constructed at different characteristic scales s . Each member of this family of fields is a component of the wavelet transform at the scale s . It resembles the

original field smoothed to this scale, but the smoothing function has very special properties discussed below, which in fact makes it a 'wavelet'. A reconstruction theorem (Grossmann, Kronland-Martinet and Morlet, 1990), asserts that the original field f , if square integrable, can be written exactly as a sum of component-fields $f^{(s)}$.

Three kinds of studies can thus be performed: (1) when a scale s is given, each component may be considered as the original field examined on this scale. Here the wavelet transform acts as a (natural) smoothing; (2) The variation of the properties with s provide spectral information; (3) Summing the components allows the reconstruction of the field. This last property, although not much used in standard wavelet analysis (except for image compression) is in fact the basis of our analysis. We will show below that this allows the construction of unknown fields implied by the data. Thus we consider *constructions* rather than *reconstructions*.

2.1. The wavelet transform in 1-dimension

The wavelet transform in one dimension is a mathematical transformation which relates the real, square-integrable, function $f(x)$ of the real variable x , to its "wavelet transform"

$$Wf(s, u) = s^\gamma \int f(x) g[(x-u)/s] dx, \quad (1)$$

where $g(x)$ is a real function called the "analysing" wavelet, which has to satisfy certain conditions (see below) [remark: the usual normalisation sets $\gamma = -1/2$. But other normalisations, such as $\gamma = -1$, may be used for special purposes]. For convenience we will denote the normalised dilation of g : $g_s(x) = s^\gamma g(x/s)$, and the wavelet transform appears as a decomposition over the family of functions $g_s(x)$. The previous formula make Wf appear as a filtering of f with the band pass filter $g_s(x)$.

To be a wavelet (and especially to allow the "reconstruction" described below), a function g must obey the "admissibility condition":

$$C_g = \int_0^\infty |Fg(\omega)|^2 / \omega d\omega < \infty, \quad (2)$$

where Fg denotes its Fourier transform. This implies (if Fg is differentiable):

$$Fg(0) = 0, \quad \text{or} \quad \int_{-\infty}^\infty g(x) dx = 0. \quad (3)$$

A wavelet g must also have a finite "energy" (defined as the norm in the space of square-integrable functions L^2):

$$\|g\|^2 = \int_{-\infty}^\infty |g(x)|^2 dx < \infty.$$

For normalisation purpose, we only consider wavelets such that $\|g\|^2 = 1$, which also implies that $\|g_s\|^2 = 1$ (with our choice of γ). Moreover we restrict ourselves to even wavelets i.e. such that $g(-x) = g(x)$.

A dominant feature of wavelets is the reconstruction theorem:

$$f(x) = (1/C_g) \int_0^\infty ds k_s \int_{-\infty}^\infty du W f(s, u) g_s(x - u), \quad (4)$$

with $k_s = s^{-3-2\gamma}$. This applies to any function f , as far as its 'energy', i.e. the integral of its square, is finite.

It is essential to our arguments that we note that a continuous wavelet transform is redundant, in the sense that the various components are not orthogonal in s space. Thus, in many cases, the wavelet transform can be totally reconstructed from the knowledge of a part of it only. This redundancy can be expressed thanks to the "reconstruction kernel" (see fig. (1c)) which depends only on the selection of the analysing wavelet as

$$K(s, u, s', u') = (1/C_g) \int_{-\infty}^\infty dx g_s(x - u) g_{s'}(x - u'). \quad (5)$$

We have then the redundancy expressed clearly as

$$W f(s', u') = \int_0^\infty ds k_s \int_{-\infty}^\infty du W f(s, u) K(s, u, s', u').$$

Moreover using this definition, the reconstruction theorem may be written (using even wavelets) :

$$f(x) = \int_0^\infty ds k_s \int_{-\infty}^\infty dy f(y) K(s, x, s, y) \quad (6)$$

and it is useful to introduce the notation

$$f^{(s)}(x) = k_s \int_{-\infty}^\infty dy f(y) K(s, x, s, y),$$

such that

$$f(x) = \int_0^\infty ds f^{(s)}(x). \quad (7)$$

This redundancy allow the complete determination (without any loss of information) of the wavelet transform from the knowledge of its values on a suitable grid in the (u, s) half-plane. The grid depends only on the shape of the particular wavelet chosen (Mallat 1988).

2.2. The wavelet transform in 3-dimensions

Wavelets were originally defined in 1 dimension, mainly for the analysis of temporal signals. The formalism was extended to 2 dimensions, for the analysis of images (see, e.g., Mallat 1988 and Mallat and Zhong, 1992). Here we extend it to 3 dimensions, without any particular difficulty. Although it is possible to construct a more general theory, we restrict ourselves to the case of wavelets which possess spherical symmetry. Although wavelets which do not have this simple property (that is no preferred orientation) may be helpful for cosmology, we do not consider such cases in this paper.

We consider an analysing wavelet in 3-dimensions as a function of the radial coordinate $r = |\mathbf{x}|$ only : $g(\mathbf{x}) = G(r)$. The dilated wavelet is written therefore

$$g_s(\mathbf{x}) = s^{3\gamma} g(\mathbf{x}/s)$$

and it is easy to generalize the definitions of the previous section in 3 dimensions. An illustration is given in fig.(1a). Thus we write the wavelet transform of a scalar field $f(\mathbf{x})$ as

$$W f(s, \mathbf{u}) = \int_{-\infty}^\infty d^3 \mathbf{x} f(\mathbf{x}) g_s(\mathbf{x} - \mathbf{u}); \quad (8)$$

and the reproducing kernel :

$$K(s, \mathbf{u}, s', \mathbf{u}') = (1/C_g) \int_{-\infty}^\infty d^3 \mathbf{x} g_s(\mathbf{x} - \mathbf{u}) g_{s'}(\mathbf{x} - \mathbf{u}'),$$

so that

$$W f(s', \mathbf{u}') = \int_0^\infty ds k_s \int_{-\infty}^\infty d^3 \mathbf{u} W f(s, \mathbf{u}) K(s, \mathbf{u}, s', \mathbf{u}'),$$

with $k_s = s^{-7-6\gamma}$, in three dimensions.

Moreover the reconstruction theorem may be written (using even wavelets) :

$$f(\mathbf{x}) = \int_0^\infty ds f^{(s)}(\mathbf{x}) \quad (9)$$

where

$$\begin{aligned} f^{(s)}(\mathbf{x}) &= (k_s/C_g) \int_{-\infty}^\infty d^3 \mathbf{u} W f(s, \mathbf{u}) g_s(\mathbf{x} - \mathbf{u}) \\ &= k_s \int_{-\infty}^\infty d^3 \mathbf{y} f(\mathbf{y}) K_s(\mathbf{x}, \mathbf{y}), \end{aligned}$$

where we defined $K_s(\mathbf{x}, \mathbf{y}) = K(s, \mathbf{x}, s, \mathbf{y})$.

Note that, as in the 1-dimensional case, the wavelet must obey the condition :

$$\int d^3 \mathbf{x} g(\mathbf{x}) = 4\pi \int_0^\infty r^2 dr G(r) = 0, \quad (10)$$

and also we define

$$|g|^2 = 4\pi \int_0^\infty r^2 dr |G(r)|^2,$$

so that g acts as a band pass filter with no preferential orientation.

3. Wavelets and vector fields

3.1. Velocity fields

Wavelets have not been applied to our knowledge previously, to vector fields. A general vector field may be considered as a collection of three scalar fields, corresponding for instance to its 3 components in some rectangular system of coordinates. Although the wavelet transform can indeed be applied to each of these 3 components, this approach offers only a limited interest since it fails to represent the spatial structure that is sought. There is however no general theorem concerning wavelet transforms of vector fields.

As a first step towards such an analysis for the study of cosmic velocity fields, we proceed here by restricting ourselves to the case of purely curl-free fields. This case is very important in cosmology for 2 complementary reasons. First, there is

a prejudice that the process of gravitational instability responsible for the apparition of cosmic velocity fields, only generates divergent fields (this is well established for the linear development of fluctuations). Secondly, only one component (radial with respect to the observer) of the cosmic velocity field is observable so that a complete knowledge of the field is not possible without additional hypotheses, such as purely divergent or 'potential' fields. This latter hypothesis was used by Dekel, Bertschinger and Faber (1990) as the basis of the POTENT method and our analysis applies under the same conditions. This will allow to compare the two methods, and to use them in complementarity.

3.2. Wavelet analysis of a potential field

A potential field V may be written $V(\mathbf{x}) = \nabla\Phi(\mathbf{x})$ where $\Phi(\mathbf{x})$ is the kinematical potential. This implies $\text{curl } V = 0$. Let us first develop the potential in 3-dimensional wavelets as :

$$\Phi(\mathbf{x}) = \int_0^\infty ds \Phi^{(s)}(\mathbf{x}), \quad (11)$$

where

$$\Phi^{(s)}(\mathbf{x}) = (k_s/C_g) \int_{-\infty}^\infty d^3u W\Phi(s, u) g_s(\mathbf{x} - u).$$

Each component $\Phi^{(s)}$ represent the potential field examined at the scale s ; the set of the spatially averaged (rms) values of the $\Phi^{(s)}$ as a function of s is an indication of the power spectrum of the kinematical potential (see below).

Applying to Φ the gradient operator we derive :

$$V(\mathbf{x}) = \int_0^\infty ds V^{(s)}(\mathbf{x}), \quad (12)$$

with

$$V^{(s)}(\mathbf{x}) = (k_s/C_g) \int_{-\infty}^\infty d^3u W\Phi(s, u) \nabla_{\mathbf{x}} g_s(\mathbf{x} - u).$$

In particular, defining $\hat{r} = \mathbf{r}/r$, we can write the radial (with respect to the observer) component $V_r = \hat{r} \cdot \nabla\Phi$ of V :

$$V_r(\mathbf{x}) = \int_0^\infty ds V_r^{(s)}(\mathbf{x}),$$

with

$$V_r^{(s)}(\mathbf{x}) = (k_s/C_g) \int_{-\infty}^\infty d^3u W\Phi(s, u) \frac{dg_s(\mathbf{x} - u)}{dr}.$$

However from the point of view of the data analysis the inverse formulae are more interesting. Let us start from

$$\Phi(\mathbf{x}) = \int_0^1 dl V_r(l, \mathbf{x}),$$

which is also the basis of POTENT. The integral operator used here is the reciprocal of d/dr . We can apply it also to the wavelet and define the function

$$G_s(\mathbf{x}, u) = \int_0^1 dl g_s(l, \mathbf{x} - u)$$

which is not a wavelet (fig. (1b)). It is however very easy to prove that the component of the kinematical potential at scale s is related to the wavelet transform of V_r at scale s through the relation :

$$\Phi^{(s)}(\mathbf{x}) = (k_s/C_g) \int_{-\infty}^\infty d^3u WV_r(s, u) G_s(\mathbf{x}, u).$$

Here we find the essence of the method proposed for the analysis of the cosmic velocity field in this paper. For this relation offers a *practical, exact*, procedure by which we reconstruct the kinematical potential from V_r : first we calculate the wavelet transform WV_r of the data. Then we apply to this field the G-transform defined for general f as

$$Gf(s, u) = (k_s/C_g) \int_{-\infty}^\infty d^3u f(u) G_s(\mathbf{x}, u),$$

which according to the theory above gives directly the component $\Phi^{(s)}$ of the potential resolved at the scale s .

It may be important to recall that the reconstruction theorem, and consequently our reconstruction method, applies only if the square of the potential has a finite integral over space. Thus a component of the potential which is constant over space is automatically excluded. This is a advantage of our method since such a potential has no physical meaning. We discuss this point more deeply in section (4.4). This also excludes from the analysis a component of the potential which would be a linear function of the coordinates x, y or z , corresponding to a constant component of the velocity. Thus our method is insensitive to a constant component of the velocity and, consequently, does not allow to recover it. On the other hand, it reconstructs correctly a possible component which would be proportional to r . Of course, the detection of such a component would be hard to interpret and would probably reflect a bias rather than a physical field.

We can immediately see some advantages of this method of constructing the velocity potential. First of all, no preliminary smoothing of the velocity field is necessary. It is highly desirable to avoid smoothing data which adds or suppresses some information and is difficult to control. In fact as noted above the wavelet transform itself is sometimes compared to a smoothing procedure. But the following points are important to emphasise once again. (1) This process does not suppress any information since the original field may be recovered from the transform. (2) Here the reconstruction procedure does not depend at all - although an usual smoothing does - of the shape of the wavelet. This is another way to express the existence of the reconstruction theorem. (3) A particular difficulty for velocity fields comes from having to smooth a vector field for which only the radial component is known. In our procedure, the wavelet transform is applied to a scalar field - the kinematical potential. (4) When a normal smoothing is applied, there is always an arbitrariness in the choice of the smoothing scale. Here the analogue of the smoothing scale, namely the scale of the analysing wavelet, is adapted continuously to each component. Thus, at each scale the "smoothing" introduced by our procedure is the minimal one necessary.

From a practical point of view, this analysis may however appear comparatively cumbersome : it involves two successive operations of convolution ; plus an additional summation over scales for the final reconstruction of the potential. Thus as it stands, this operational complexity may appear to be a serious disadvantage compared to POTENT which requires only one convolution (the smoothing), followed by a second step, the radial integration. However we proceed to show that our procedure can be considerably simplified thanks to the powerful theorems applying to wavelets.

3.3. A practical reconstruction procedure for the potential

In order to reduce the complexity of the procedure, we define a new operator L - a generalized kernel - in the same spirit as the kernel introduced above (in particular which depends only on the wavelet and its support) as:

$$L(s, u, s', u') = (1/C_g) \int_{-\infty}^{\infty} d^3 u G_s(\mathbf{x}, u) g_{s'}(u - u'). \quad (13)$$

In the following we will only consider the case when the 2 scales s and s' are equal so that we simplify the notation and write $L(s, u, s, u') = L_s(u, u')$. It then follows that:

$$\Phi^{(s)} = k_s \int_{-\infty}^{\infty} d^3 \mathbf{x}' L_s(\mathbf{x}, \mathbf{x}') V_r(\mathbf{x}').$$

Our procedure is now much simplified since a unique convolution of V_r by the generalized kernel L_s directly gives the $\Phi^{(s)}$. The kernel, at a given scale s , is a function of the 2 variables and may be calculated analytically, or tabulated, depending on the particular shape of the wavelet chosen (an illustration is given in fig. (1d)). In principle this can be done in an error free manner.

3.4. Reconstructing other quantities

For a potential field, the knowledge of the kinetic potential is sufficient, and the operator L was constructed specifically to provide it. Other quantities may however be desired, such as the other components of the velocity field, its divergence, the deformation tensor, etc. ; or even dynamical quantities if we devise a dynamical theory that gives access to them. The important idea is that for each such quantity which depends linearly on the velocity field, a specific operator similar to L for the potential can be constructed.

Assume for example that the quantity q of interest depends on the potential Φ through a linear operator $O_{\mathbf{x}}$:

$$q(\mathbf{x}) = O_{\mathbf{x}} \Phi(\mathbf{x})$$

(here the index indicates that the operator acts on the \mathbf{x} variable). Then to q is associated the generalised kernel :

$$D_s(u, u') = (1/C_g) \int_{-\infty}^{\infty} d^3 u H_s(\mathbf{x}, u) g_{s'}(u - u'),$$

where we have defined

$$H_s(\mathbf{x}, u) = (k_s/C_g) O_{\mathbf{x}}^{-1} g_s(\mathbf{x} - u),$$

$O_{\mathbf{x}}^{-1}$ being the inverse of the operator $O_{\mathbf{x}}$. The operator $O_{\mathbf{x}}$ has the same scalar, vectorial, or tensorial character as the quantity q . Thus, for the divergence,

$$O_{\mathbf{x}} = \frac{d^2}{d\mathbf{x}.d\mathbf{x}}$$

while for the velocity itself,

$$O_{\mathbf{x}} = \frac{d}{d\mathbf{x}}$$

and for the deformation tensor,

$$O_{\mathbf{x}} = \frac{d}{d\mathbf{x}} \frac{d}{d\mathbf{y}'},$$

etc.

So, a method similar to that described for the potential, provides any of these quantities directly, merely by replacing the quantity G of equation (13) by H as defined generally here.

4. Application of the method

4.1. Scales and spatial resolution

Hence we may calculate in an economical way the wavelet components of the potential, and other quantities, from the radial velocities. The formulae above involve however, for the complete reconstruction, an integral over the scales. Fortunately, available mathematical theorems (see e.g. Grossmann, Kronland-Martinet and Morlet, 1990) state that the replacement of these integrals by sums, leads to exactly the same results. This is due to the fact that the set of all wavelet transforms contains redundant information and there is an optimum procedure to minimize this redundancy without any loss of information. In the cosmological case, the presence of a small scale cut-off (due to the sampling of our sample) and an large-scale cut-off (the limited size of the sample) yields always a finite sum. We will see in the next section that the number of terms to consider is not very large.

The spatial resolution of the wavelet-transform of a field at the scale s depends on s . The wavelet becomes wider, in physical space, as s increases. But the basic property of a wavelet is to offer finite resolution both in real space and in Fourier space, that is the (spatial) frequency space. The standard deviation of an analysing wavelet g in real and Fourier spaces are denoted σ_u and σ_ω respectively. The "energy" of a wavelet $g_s(x - x_0)$ is concentrated in space around x_0 , with a standard deviation $s \sigma_u$. In frequency space, energy is concentrated around ω_0/s , with a standard deviation σ_ω/s . In the phase-space, the wavelet is associated with a resolution-cell of surface $4 \sigma_u \sigma_\omega$ which does not depend on scale. Cells corresponding to the Fourier transform also have a constant resolution, but their shape remain constant. For a wavelet, the shape changes with scale. This is their essential feature.

Our goal is to reconstruct an original field without loss in information. A continuous integral over the scales does the job, but with a considerable redundancy, expressed by the fact that the cells in phase space recover each other. Mallat (1988) has shown that the replacement of this integral by a discrete sum is possible without any loss of information, as far as the phase space remains totally covered by the reunion of the resolution cells. He showed that an exponential sampling is sufficient : with a correct choice of the elementary dilation step α , it is sufficient to consider a countable sequence of scales α^n , (corresponding to spatial frequencies α^{-n}) where n is an whole number. Then, at the scale α^n , the signal is spatially sampled with a step $\beta \alpha^n$. The volume of the resolution cell remains constant and equals β and the phase space is completely covered. The limits to the sum will result from the sampling and the size of the sample.

In one dimension, discrete wavelets allow to make an ideal choice, with no redundancy and no loss of information. This is however impossible in 3 dimensions without abandon the spherical symmetry of the wavelet (since this would imply the

tessellation of space with cells representing the support of the wavelet). Here however we keep the spherical symmetry and thus renounce orthogonal wavelets.

4.2. Choice of the "analyzing" wavelet

In this paper, we perform our analysis with the following 3-Dimensional "Top Hat" wavelet $g(\mathbf{x}) = G(r)$, where $r = \|\mathbf{x}\|$ and

$$\begin{aligned} G(r) &= 1/C \text{ if } 0 \leq r < 1 \\ G(r) &= -1/(26 C) \text{ if } 1 \leq r < 3, \\ G(r) &= 0 \text{ if } 3 \leq r, \end{aligned} \quad (14)$$

with $C = \sqrt{4/3\pi(1+1/26)}$. It is well-known that the discontinuity of this particular Haar type analysing wavelet is sometimes responsible for undesirable effects, such as a poor convergence of the coefficients of the wavelet transform. But we avoid this difficulty by using the generalised kernel as a reconstruction procedure. Moreover, this wavelet has compact support and its simple expression allow us to calculate the generalised kernel analytically.

We give in figures (1a) and (1b) a representation, on a 2-dimensional plane, of our analysing wavelet $g_s(\mathbf{x})$, and its integral along the line of sight $G_s(\mathbf{x}, u)$, for some arbitrary selected position u . (1c) and (1d) present the kernel $K_s(\mathbf{x}, \mathbf{x}')$, which is also a wavelet, and its integral along the line of sight $L_s(\mathbf{x}, u)$, which is our generalised kernel, associated with the potential defined in equation (13). We emphasise that, for a reconstruction point \mathbf{x} , the domain where the values of generalised kernel do not vanish, is confined to the region at a distance lower than $6s$ from E the line joining the origin O to \mathbf{x} . The choice of a compactly supported analysing wavelet allows us in some respects, to minimize the edge effects.

It is also worth noticing that the integral over the scales can be replaced by a sum over a set of few scales logarithmically distributed. In our case, we find that 3 scales per decade are already sufficient to accurately reconstruct the velocity potential. To reduce the noise, however, we chose to reconstruct with 6 different scales.

4.3. Synthetic data

Application to real astronomical data involves some additional practical problems which will be presented in forthcoming paper. Here we check the validity of our method, and illustrate its capacities by an application to synthetic data. However by adding random noise to the synthesised velocity field so as to degrade the effective signal to noise ratio to unity, we are able to demonstrate that the physical features on the larger scales are left relatively unaffected, as anticipated by the above theory.

We have claimed that our method has the property to separate the contributions of velocity substructure that is coherent on different scales. In order to illustrate this feature, we synthesise on a 3-dimensional grid ($[-5000, 5000], [-5000, 5000], [-5000, 5000]$), a radial velocity field induced by a potential $\Psi = \Psi_1 + \Psi_2$ (length unit can be thought as $km.s^{-1}$). Ψ_1 and Ψ_2 have a gaussian shape, with the same amplitude but different dispersions (i.e different coherent scales) and different central locations. We choose Ψ_1 to be centered at $M_1(4500, 0, 0)$ and to have a dispersion

$\sigma_1 = 4500$, which is intended to mimic a large scale flow similar to the Great Attractor flow in our real universe. Then, we add Ψ_2 as an local attractor flow on a smaller scale (Virgo infall, for example) located at $M_2(0, 2000, 0)$ and having a dispersion $\sigma_2 = 2000$. Thus, the centers of the potentials Ψ_1 and Ψ_2 lie on perpendicular axes in the plane $z = 0$ as in figure (2). Our synthesised data are then the values of the induced radial velocity field $V_r(\mathbf{x}_i)$ at each point \mathbf{x}_i of the grid defined above.

In order to visualize the results of our method, we find it convenient to present the values of both the synthesised and the reconstructed potential on 2 cuts through the initial cubic grid. We choose as the plane P_1 the cut with the equation $z = 0$, and which therefore contains the observer O as well as the centers of the potentials Ψ_1 and Ψ_2 , as discussed above. It is also of interest to show the results on a plane which doesn't contain the origin O . We therefore choose also the plane P_2 perpendicular to P_1 and crossing P_2 at the line of intersection ($M_1 M_2$) (figure 2).

4.4. Reconstruction of the potential

The mean value of the radial velocity field in a sample generally differs from zero. Since our method allows us only to treat square-integrable functions, it reconstructs in fact the potential Ψ_{rec} which yields the radial velocity field $V_r - \langle V_r \rangle$, where $\langle V_r \rangle$ is the mean value of $V_r(\mathbf{x})$. In order to recover the real synthesised potential $\Psi_{real}(\mathbf{x})$ corresponding to V_r , we are obliged to add the conic (in $r-s$ space) function $\Psi_{conic}(\mathbf{x}) = r \langle V_r \rangle$. Similarly, our method does not reconstruct a constant component for the potential, or a component proportional to cartesian coordinates, which would be artificial.

Finally we have to choose the set of scales over which the potential will be reconstructed. We take 6 steps in a decade of scale ($[200, 2000]$) logarithmically sampled. We expressly keep the large scale in order to visualize the edge effects and also the smallest one for the same reason since this is affected by the finite resolution of the data grid.

4.5. The results

We illustrate qualitatively the validity of our method by showing figures of the original (Ψ) and reconstructed (Ψ_{rec}) potentials. In order to quantify the discrepancy between the values of these two potentials, we define the following test :

$$S(\Psi, \Psi_{rec}) = \left(\frac{\sum [(\Psi^i - \langle \Psi \rangle) - (\Psi_{rec}^i - \langle \Psi_{rec} \rangle)]^2}{\sum [(\Psi^i - \langle \Psi \rangle)]^2} \right)^{1/2} \quad (15)$$

with the sums performed on the set of points belonging to the surface where the potential is reconstructed (the plane P_1 for example) and $\langle f \rangle$ denotes the mean value of the function f on this surface. Our choice for the S test is motivated by its invariant properties with respect to a multiplication of Ψ by a constant or a shift of its mean value. Finally, the quantity $S(\Psi, \Psi_{rec})$ should be close to 0 if our reconstruction procedure is valid.

In figure 3 we present both the original (a) and the reconstructed (b) potential. The level of the surface represents the intensity of the potential on the "horizontal" plane passing through observer, and the centers of the two attractors, that is the plane P_1 . Figure (3b) represents the reconstruction in the absence of noise. Only weak distortions due to edge

effects are detectable. The value of the S test is satisfyingly small ($S(\Psi, \Psi_{rec}) = 0.14$). Let us emphasize that we have arbitrarily allocated the value $s = 200 \text{ km.s}^{-1}$ to the small scale cut-off involved in the summation over the scales. The reconstruction is better if we choose a larger scale ($s = 317 \text{ km.s}^{-1}$) for this small scale cut-off ($S(\Psi, \Psi_{rec}) = 0.12$), since the component $\Psi_{rec}^{(200)}$ of the reconstructed potential is affected by the finite sampling of the 3-dimensional data grid. The powerful properties of the wavelet analysis thus provide us with a natural procedure to deal with the "inevitable" smoothing due to the discreteness of the data. Practically, one must simply stop the reconstruction at the inferior scale ($s = 317 \text{ km.s}^{-1}$ for our example) for which the effects of the discrete sampling are negligible.

Figure (3c) represents the result of a more rigorous test of the method wherein Gaussian noise with zero mean and a dispersion $\sigma = (1/V) \int_V V_r^2 dV$ (V is the sample volume), were added to the velocity field of (3a). The mean signal to noise over the sample is thereby adjusted to unity. A comparison with (b) shows that nevertheless the large scale features remain reliable ($S(\Psi, \Psi_{rec}) = 0.2$). Moreover, the influence of this noise on our reconstruction procedure can be eliminated by stopping the reconstruction of the potential at larger scale. In fact the noise appears only for the small scale components $\Psi^{(s)}$ of the potential and is smeared out at large scale. We have reconstructed the noisy potential without including the last component $\Psi_{rec}^{(200)}$ in the sum over the scales. The value of the S test is noticeably reduced ($S(\Psi, \Psi_{rec}) = 0.13$). Our method thus allows us to control naturally the influence of shot noise by adjusting the inferior limiting scale for the reconstruction of the potential.

Fig 4 presents a different cut through the simulated data, namely in the "vertical" plane P2, orthogonal to the previous one and passing through the two centers. Again figure (4a) shows the original simulated data in this plane, while figure (4b) shows the (almost perfect) reconstruction ($S(\Psi, \Psi_{rec}) = 0.07$). Thus the reconstruction works very well in both planes. Correction for the constant component of the radial velocity has been included in these reconstructions as indicated above.

It is remarkable that, although the data and the analysis are effected on a finite volume as opposed to the infinite support assumed in the mathematical treatment above, no, or at worst extremely weak, edge effects appear. The original and reconstructed potential agree well, even near the edges of the sample.

4.6. Decomposition by scales

This procedure gives by construction the separate components of the potential at the different scales. This is an indication of the power spectrum, which is very often desired in order to compare the kinematics with predictions of dynamical theories. More generally it is very interesting that the reconstruction of the potential at some scale is not influenced by the other scales: this minimises the effects of the noise. It also allows us to have a direct knowledge of the characteristic scale of any feature (e.g. an "attractor") which is discovered in the data.

In fig 5 we present the components of the reconstructed potential in the plane P1 at the two different scales $s = 500$ and $s = 2000$, respectively in (a) and (b), while in figure 6 the same information is presented for the plane P2. These should be compared with the exact velocity fields in figure (3a) and (4a)

respectively. As expected, we check that the two components of the potential at different scales, Ψ_1 and Ψ_2 are well separated.

One sees in these figures an interesting illustration of the redundancy between the scales, where for example in figures (5a) and (5b) distortions of opposite sign along the right hand edge of the figure must add to give something much more nearly flat, as is found in (3b). The strength of these effects is at least partly due to the fact that the integral over a constant mean radial velocity does not disappear scale by scale. There is moreover no unique procedure for subtracting these mean fields as a function of scale. Thus although this information is 'real', it shows that one must be careful in deducing a 'power spectrum' at different, even well-separated, scales from a non-orthogonal wavelet analysis on data with a non-zero mean.

Despite this difficulty we see by comparing the decompositions with the original data in the two planes, that the method allows us to describe the velocity field by scales: even if a power spectrum is probably uniquely meaningful only over an orthogonal (in the limit a well-separated in scale) basis on a data set of zero mean.

5. Conclusion

We have devised a method which allows the complete reconstruction of a vectorial field, assumed to be curl-free, from only the knowledge of its radial component. This applies to cosmic velocity fields, allowing one to derive the potential or, similarly, any desired quantity that is a linear function on the vectorial field.

This method requires no preliminary smoothing of the data, and is performed by a convolution of the data with a known kernel. It gives directly the potential or any desired quantity as decomposed over a range of scales (less than 6 scales is sufficient to deal with the problem of cosmic velocity fields). The summation over these scales gives with very good precision the reconstructed quantity. In addition the method allows also a study of the dependence on spatial scale and is indicative of the power spectrum.

We have checked the validity of the method on simulated data which represents a field not greatly different from what the cosmic velocity field might be: i.e. dominated by a "Grand Attractor", with additional smaller scale flows and noise. We have checked that the method yields a proper reconstruction of the total field; and also allows an isolation of the contributions from the flows at different scales. We estimate that it is very well adapted to the study of cosmic fields. A forthcoming paper presents its application to real data.

References

- Bertschinger E. and Dekel A., 1989, Ap. J., 336, L5-L8
- Bertschinger E., Dekel A., Faber S.M., Dressler A. and Burstein D., 1990, Ap. J., 364, 370
- Daubechies I., 1990, in *Wavelets*, Proc. of the International Conference, Combes, Grossman and Tchamitchian ed., Springer-Verlag
- Dekel A., Bertschinger E. and Faber S.M., 1990, Ap. J., 364, 349
- Grossmann, Kronland-Martinet and Morlet, 1990, in *Wavelets*, Proc. of the International Conference, and Tchamitchian ed., Springer-Verlag
- Henriksen R.N. and Lachièze-Rey M., 1990, MNRAS, 245, 255
- Lachièze-Rey 1993, ApJ, in press

- Mallat S.G., 1988, Robotics report 178, New York University, Dept. of Computer Science
- Mallat S.G. and Zhong S., 1992, IEEE Transactions on pattern analysis and machine intelligence, 14, 7
- Meyer Y., 1990, in *Wavelets*, Proc. of the International Conference, Springer-Verlag
- Nusser T. A., and Dekel A., 1992 Ap. J., 391, 443-452
- Shaya E.J., Tully R.B. and Pierce M.J. 1992 Ap. J., 391, 16
- Yahil A., 1988, in Vatican Study Week : Large Scale Motion in the Universe, ed. V.C. Rubin and G.V. Coyne (Princeton University Press), 219

FIGURE CAPTIONS

1.
 - (a) The level of the wavelet $g_s(x)$ as a function of x in a plane passing through $x = 0$
 - (b) The integral $G_s(x, u)$ of g_s along the line of sight, for a fixed value of u .
 - (c) The kernel K ,
 - (d) The generalised kernel $L_s(x, u)$, integral of K , along the line of sight, for a fixed value of u .
2. This is an illustration of the simulated volume showing the orientation of the axes and the location of the two attracting centres at M1 and M2. The planes P1 and P2 are indicated.
3.
 - (a) The original simulated velocity potential represented as an amplitude over the plane P1.
 - (b) The reconstructed velocity potential using the method presented in this paper, and using the simulated radial velocity field with zero noise.
 - (c) The reconstructed velocity potential when noise has been added to the simulated radial velocity field so as to produce a mean signal to noise ratio over the sample of unity.
4.
 - (a) The original simulated velocity potential represented as an amplitude over the plane P2.
 - (b) The reconstructed velocity potential represented over the plane P2 based on the simulated velocity field with zero noise.
5.
 - (a) The velocity potential $\Phi^{(s)}$ represented as an amplitude over the plane P1 for $s = 500$.
 - (b) Same as (a) but for $s = 2000$.
6.
 - (a) The velocity potential $\Phi^{(s)}$ represented as an amplitude over the plane P2 for $s = 500$.
 - (b) Same as in (a) but for $s = 2000$.

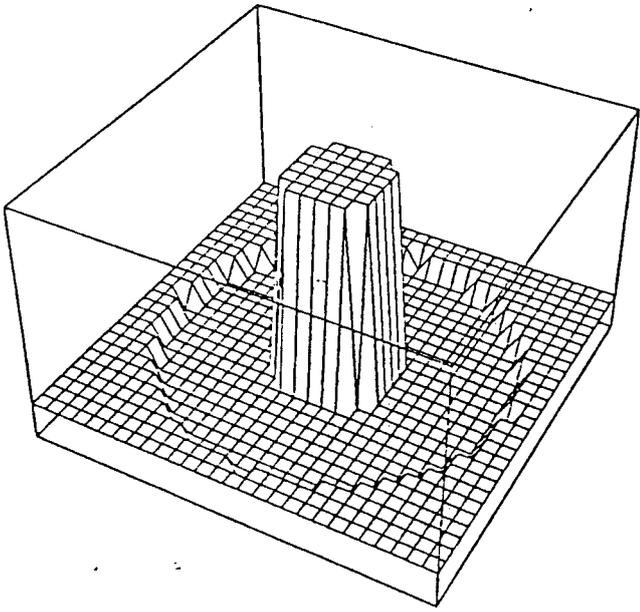


Figure 1a

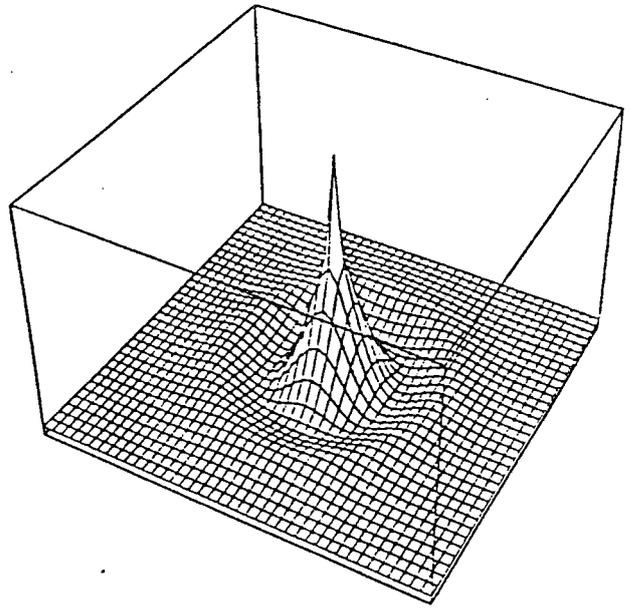


Figure 1b

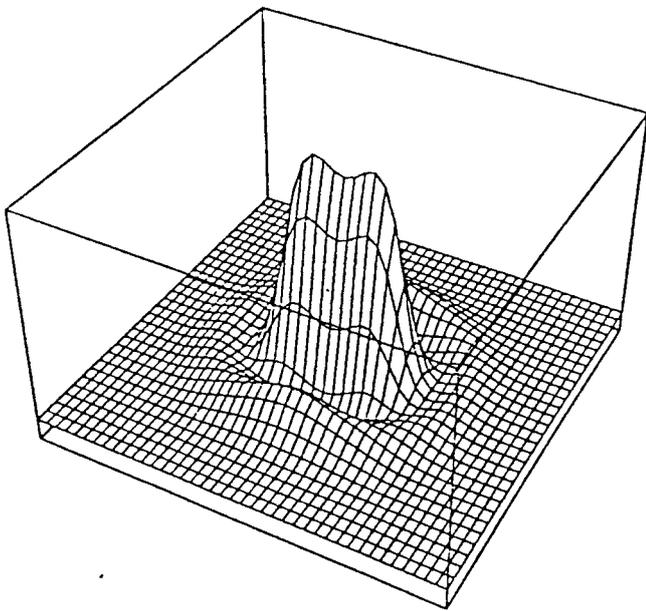


Figure 1d

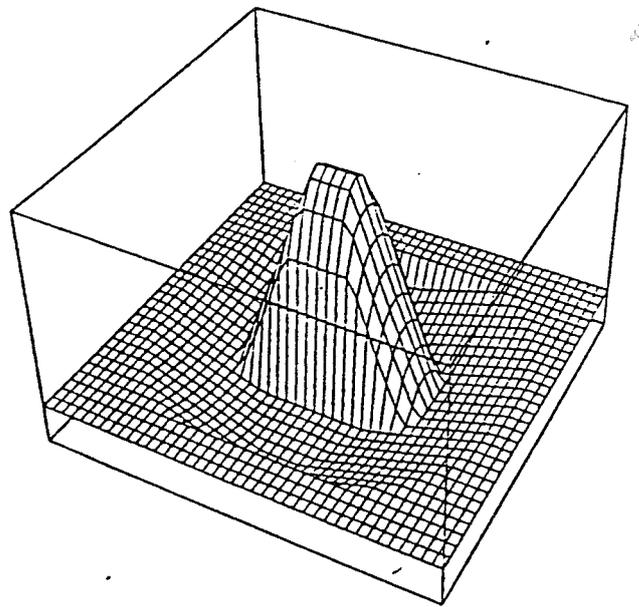


Figure 1c

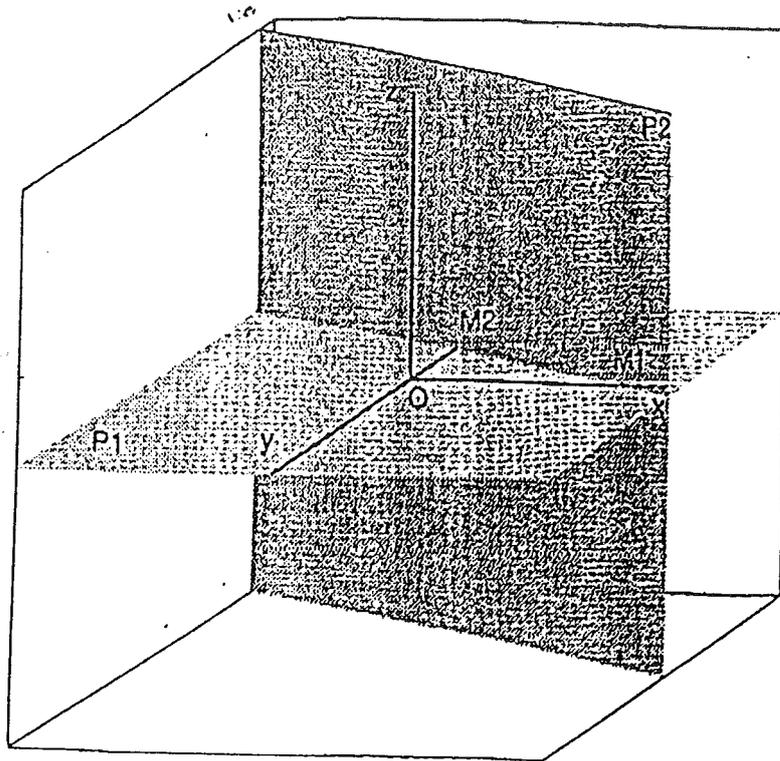


Figure 2

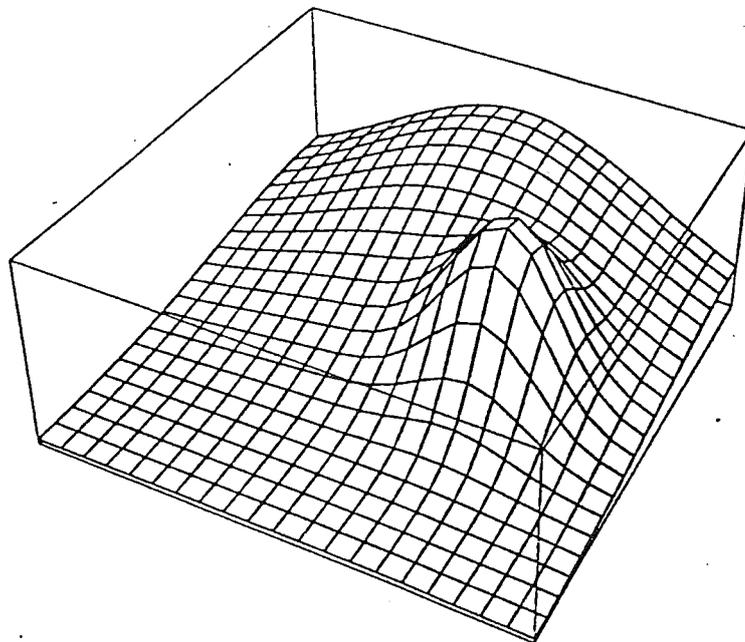


Figure 3a

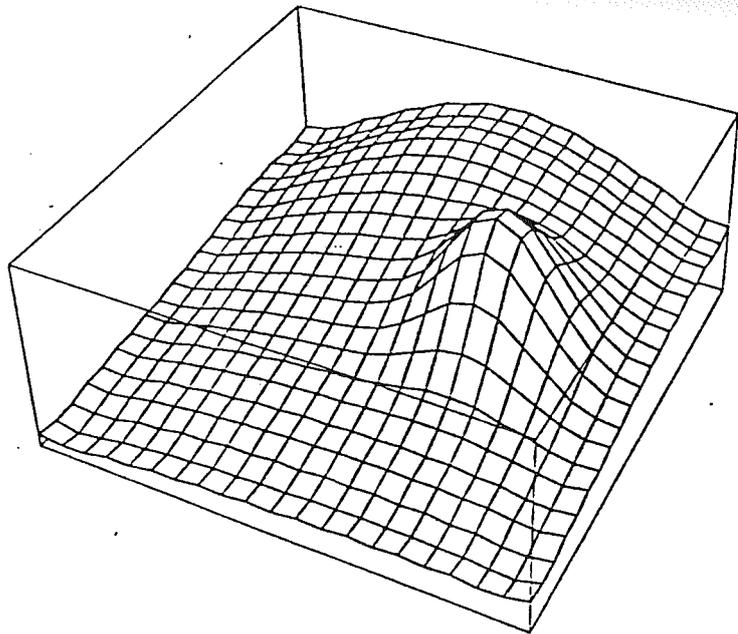


Figure 3b

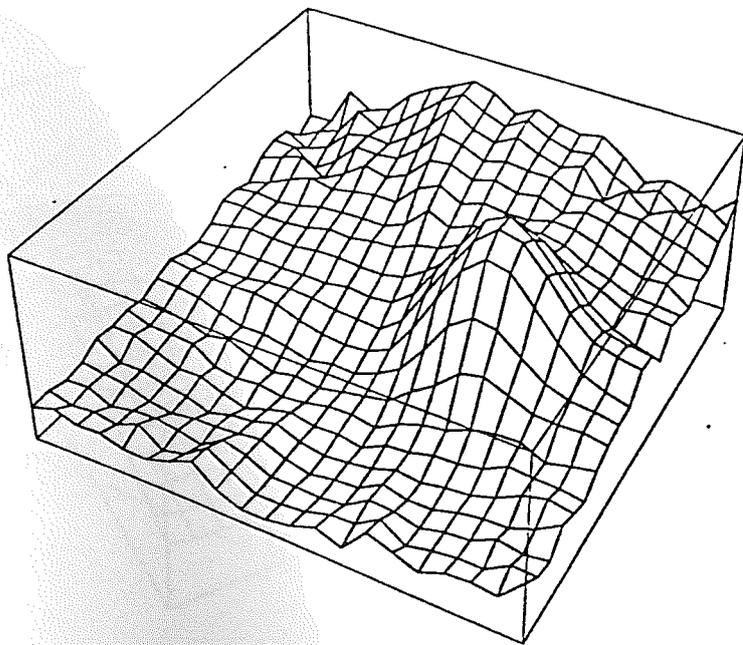


Figure 3c

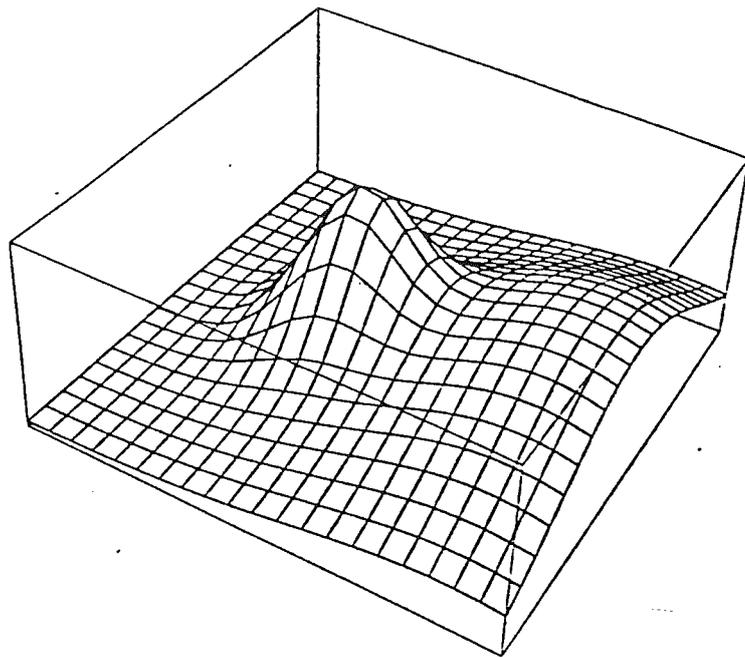


Figure 4a

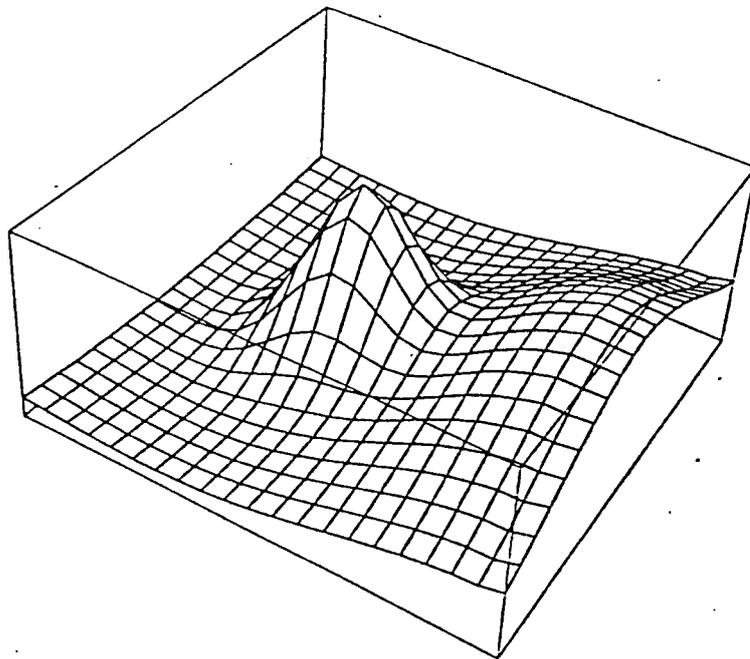


Figure 4b

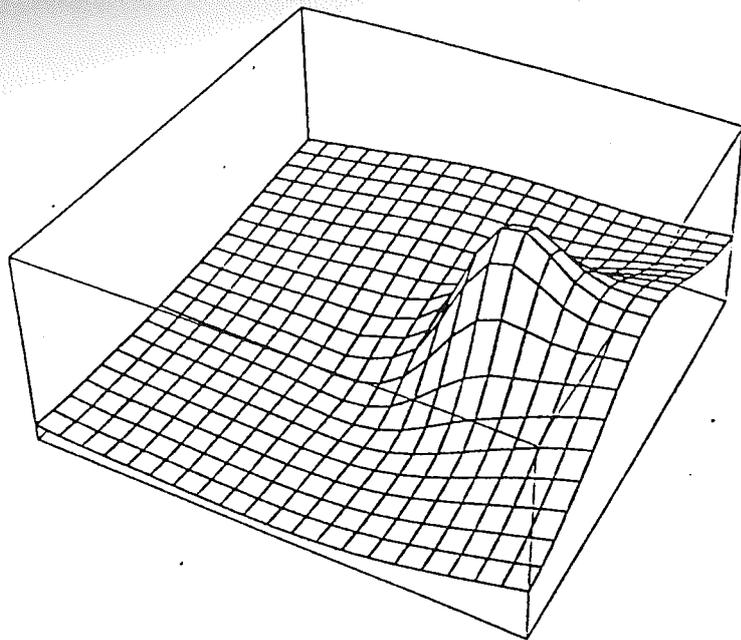


Figure 5a

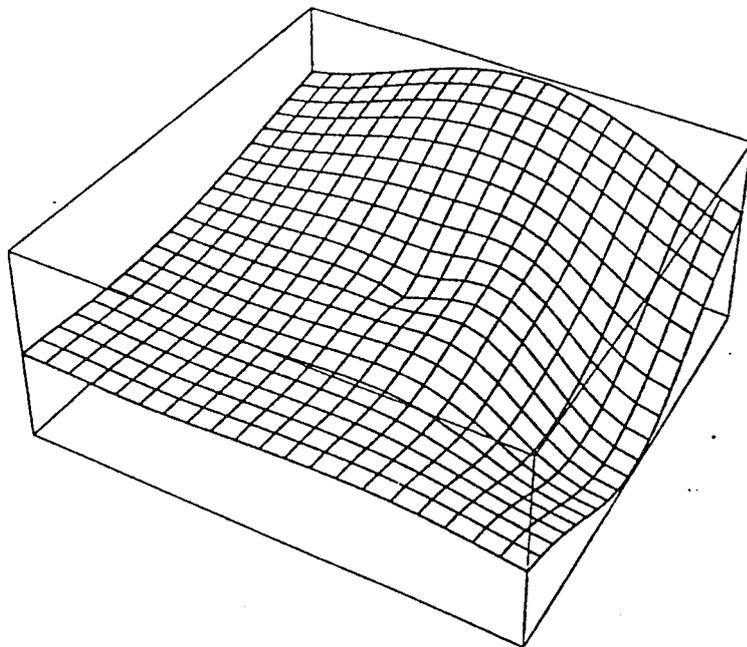


Figure 5b

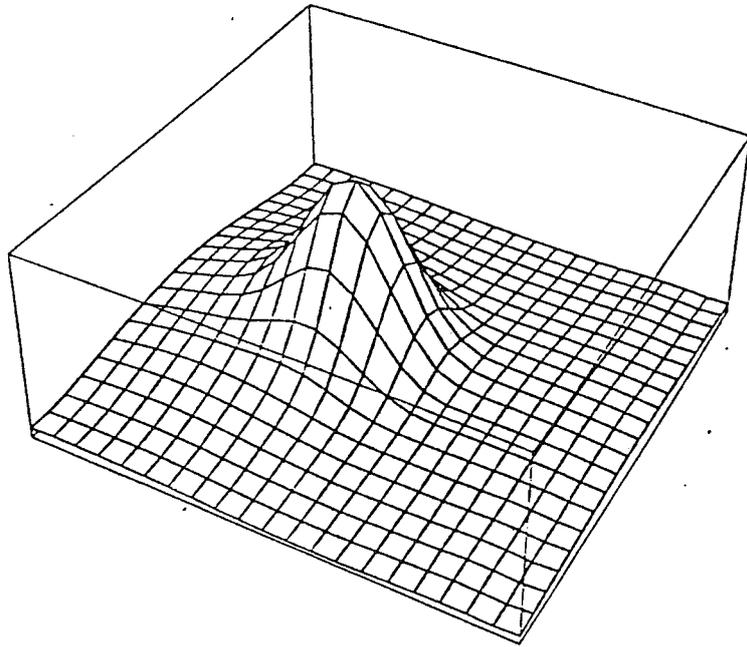


Figure 6a.

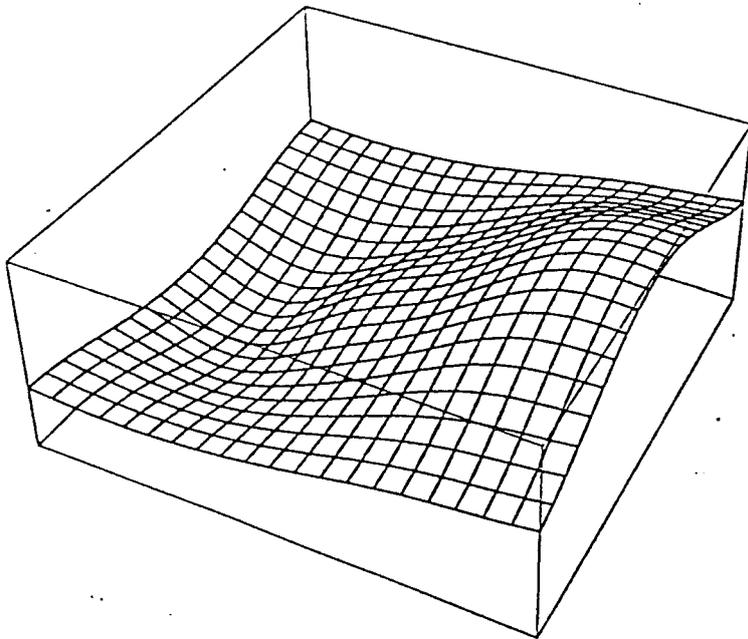


Figure 6b.