Mapping the Large Scale Structure of the Universe

by

Andrew M. Newsam BSc.

Thesis
submitted to the
Department of Physics and Astronomy,
University of Glasgow
for the degree of
Ph.D.

Astronomy and Astrophysics Group,

Department of Physics and Astronomy,

University of Glasgow,

Glasgow G12 8QQ

November 1994

ProQuest Number: 11007897

ProQuest 11007897

For Mother, for wearing purple!

Preface

The work described in this thesis was performed while the author was a research student in the Department of Physics and Astronomy in the University of Glasgow and in receipt of a PPARC (formerly SERC) research studentship. The work concerns the study of the large scale structure of the universe and in particular, the use of distance estimates to galaxies to derive information about the motion of the galaxies and hence the processes that led to their formation.

The POTENT analysis of Bertschinger and Dekel (1989) is implemented and adapted to minimise the systematic errors introduced by inaccurate distance estimates. Also, the estimates themselves are improved by a new calibration technique. Finally, and most importantly, the various techniques developed are applied to a set of actual data and the results analysed in the light of the careful and detailed testing performed as the methods were devised and compared. In this way, a methodology is developed for reconstructing maps of the local velocity and density fields which should lead to better results that can be confidently used to constrain and inform theories of structure formation and the origin of the universe.

There are, of course, far too many people to thank, and I hope that all those I cannot thank here will realise that it is not because I do not appreciate all their help and support but because they are less likely to beat me up than those I *do* mention. Foremost among the dangerous is Keith, whose constant stream of abuse has been an education in itself – I can only hope I gave as good as I got. Also my thanks to everyone else who has been forced to share an office with me – Giota for her patience and enthusiasm and Aidan for trying not to apologise too often. I must also thank Norman for, well, for helping with LaTeX and wearing a hat and enthusing about Macs and just being generally *taller* than everybody

else. Looking back, I cannot pass by Lyndsay Fletcher and David Alexander, who welcomed me into the group with archetypal Glasgow hospitality made the first few months here more pleasant than I could possibly have hoped.

However, in spite of all the pleasure and help that everyone in the department has given me (especially Daphne – I will never look at a photocopier in quite the same way again) none of this would have been possible without constant pestering and advice from three people. First is Gavin Starks. He is without a doubt the friendliest musician turned astronomer turned musician turned computer administrator turned astronomer that I have ever met and without the constant distraction of his music and his (slightly bizarre) sense of humour, the last three years would have been much less interesting (albeit spent with slightly more sleep).

Then there is Martin Hendry. It seems slightly odd that the person whose desk I stole when I arrived here should be so central to my work but his enthusiasm, experience, contacts, help and above all, belief in the work and the joy that can come of it, have given spice to everything we have done together.

Finally, no acknowledgements would be complete without my supervisor, John "F.L." Simmons. His amazingly broad knowledge, enthusiasm and patience have been an inspiration and his home-baked bread has been uniformly delicious. I can only say that I consider myself extremely lucky to have been under his wing, even if it has meant learning some statistics, and his influence has extended far beyond astrophysics and into the realms of film, art and literature. Thank you for everything.

My three years here have been very good ones. I have made a number of friends (and surprisingly few enemies) and I have learnt some valuable lessons. If I have taken nothing else from Glasgow, you can all be sure that I will carry my passport with me, whenever I travel!

Summary

In this thesis I will be examining the study of the large scale structure of the universe. In particular, I will be looking at the rôle of peculiar motions of galaxies – ie motions that deviate from the uniform Hubble expansion of the universe. The study of these motions holds much promise for cosmology, but there are considerable problems with measuring and mapping them, a number of which I will be addressing in the course of this work.

However, I will start in chapter 1 by giving a brief introduction to the theory behind the formation of structure in the universe and the current state of our knowledge about its form and history. The framework provided by the Big Bang theory of the origins of the universe, has withstood a number of observational tests with remarkable success and this has enabled cosmologists to extend and refine the theory. This enlarged model can also be tested by observations and in its success or failure, improve our understanding and narrow the limits of future research.

In chapter 2, therefore, I will describe one of the most powerful set of observational tools – the measurement of the local velocity and density fields. Inherent in the Big Bang theory, and almost any other reasonable explanation of the universe, is the idea of formation and evolution of structure. The velocity and density fields tell us about the current state of that evolution and, therefore, will set very strong constraints on any theory. However, as I will show, the measurement and analysis of such fields is a complicated and difficult business. In particular, many methods rely on estimation of the distances to galaxies and this estimation is subject to very large errors. In the latter half of the chapter, I will be considering this problem and looking at some of the most popular distance estimators and some of the systematic errors or biases that they can introduce to field recoveries.

The most commonly used estimators (Tully-Fisher and D_n-σ techniques) rely on a strong

correlation between two observable properties of each galaxy – one that varies with distance and another that is distance-independent. Therefore, by using the distance-independent quantity to estimate the *absolute* values of the dependent observable, the distance can be estimated. However, one of the major problems with such distance estimation techniques is calibration – exactly how are the two parameters of the chosen relation correlated? In chapter 3 I will describe a new method of calibration that attempts to combine a number of clusters of galaxies together into one large cluster that can then be used as a calibration yardstick. This is an extension of the standard technique where a single cluster is used. While describing this technique, I will also demonstrate its efficacy with several carefully devised tests that will show clearly how it improves over the single-cluster approach.

Given a set of distance estimates, we now wish to derive some information about the local velocity field. One very successful method for doing this is described in chapter 4 - the POTENT method (Bertschinger and Dekel, 1989). With this technique, expansion of the universe is removed from the velocity field by comparing the recession velocities of galaxies with their estimated distances and the resultant smoothed *peculiar* velocity field recovered under the assumption of potential flow. I will describe my implementation of the method and go on to test it with a variety of different forms of distance estimator thereby demonstrating the large biases that can result (especially the so-called Malmquist bias). Although a number of "corrections" for this bias already exist, I will show that none are ideal and when applied in the wrong situation or without a full understanding of the properties of the chosen distance estimator, the results can be far worse than with no correction at all. Chapter 5, therefore, is concerned with a number of techniques I have developed during the course of this thesis to improve on this situation.

The first of these attempts to minimise the problems by performing as much of the analysis as possible in redshift-space, thereby avoiding much of the use of distance estimates. However, although successful for simple tests, this proves to be inadequate when confronted with a realistically complicated situation. More successful is an iterative technique based around Monte Carlo error estimates that gradually adjusts an estimate of the velocity field until its recovery (with biases) matches the recovery with the actual data. This method has the particular advantage of making no assumptions about the causes of the various biases, but simply tries to estimate their effect and remove them. The results are noticeably better

than any other method in all the tests performed. Finally in this chapter, an attack is made upon the random errors in POTENT recoveries. The method is adapted so that areas particularly sparse in galaxies are avoided except where absolutely necessary. The results, however, are inconclusive although the method clearly warrants further study.

After this, in chapter 6 we come to the core of the thesis. Here the techniques described and tested during the previous chapters are applied to a collection of actual data and the results analysed. The results compare well to other published data, and the confidence gained from the rigorous testing performed enables a higher level of confidence in the results than has been possible up to now. Therefore, even though a full analysis is beyond the scope of this thesis, I am able to use a simple technique to place limits on the density parameter of the universe, Ω_0. Namely, Ω_0 is seen to be greater than 0.3 at a 2σ confidence level.

Finally, in chapter 7, I will examine the future directions of the work – in particular how the improvements to POTENT and distance calibration can be refined and applied to larger data sets to get much more stringent information about the nature of the universe.

Contents

List of Figures

Chapter 1

The Large Scale Structure of the Universe

> If you can look into the seeds of time,
>
> And say which grain will grow and which will not,
>
> Speak then to me, who neither beg nor fear
>
> Your favours nor your hate.
>
> *Macbeth I:iii*

1.1 Thesis Outline

In this thesis, I will be examing the role of peculiar velocity and density fields of the local universe in the study of Large Scale Structure. In particular, I will concentrated on the use of redshift-independent distance estimates in methods such as POTENT (Bertschinger et al., 1990) to rederive these fields. An appreciation of the statistical properties of the distance estimates and how they interact with the reconstruction method is very important if large, unrecognised systematic errors are to be avoided, and I will examine such errors closely.

In chapter 4, this will be shown to be particularly important when using the various forms of "Malmquist Correction" (eg Lynden-Bell et al. (1988) and Landy and Szalay (1992)). These corrections attempt to modify the distance estimates in such a way that the errors

introduced by smoothing data in *estimated* distance space are removed. However, the corrections will obviously depend on the distribution of galaxies and some assumption must be made. For the Homogeneous Malmquist correction (HMC), the assumption is that the *underlying* distribution of galaxies is homogeneous whereas for the General or Inhomogeneous correction (IMC), the radial distribution of *observed* galaxies is modelled by the radial distribution of estimated distances (ie $\hat{p}(\omega) = p(\hat{\omega})$ where ω is the log of the distance).

Clearly, there is a fundamental difference in these assumptions – one relates to the distribution of objects in the real universe, whereas the other has this distribution convolved with some selection function. This difference is the main motivation behind the choice of 'raw' distance estimate. It has been shown (eg Hendry and Simmons, 1994) that when the selection of galaxies is not considered, the estimator should be 'directly' calibrated such as the DTF estimator. On the other hand, corrections based on observed distributions should be applied to 'inverse' estimators (eg ITF). Simple tests in this thesis on the POTENT method show the basic validity of this approach, but for realistic distributions of galaxies, the assumption in all the corrections considered here that the main effects are purely radial is no longer valid and the corrections begin to loose effectiveness. Also, other biases in POTENT become important such as sampling gradient bias which is related to the actual velocity field as well as the distance errors and galaxy distribution. Given these problems, I propose a number of alternatives to the orthodox POTENT procedure which are developed using a carefully designed Monte Carlo procedure to test the effect of distance errors and galaxy distributions on the recoveries (chapter 5).

This first of these methods involves removing some of the dependence on distance estimates by performing the smoothing in redshift space. There is some evidence from the tests performed in this chapter that this approach does significantly reduce Malmquist-like biases but, unfortunately, the transition from redshift to real space introduces noise into the recoveries for all except the most trivial velocity fields.

The second technique, however, shows much more promise. The aim with this is to move away from the details of the cause of the biases and the exact form they take, and instead simply treat them as some unknown systematic errors to be removed. This is done by imposing a known velocity field onto the galaxy catalogue and estimating the bias of this field using a Monte Carlo run. The imposed velocity field is then changed to a best-guess

estimate of the actual velocity field and the process iterated towards a solution. This rather unimaginatively titled 'Iterative Monte Carlo Correction Technique' will be seen to be a considerable improvement on any other in all the tests performed, largely because, since it makes no assumptions about the form of the bias, it can happily cope with non-radial Malmquist effects and sampling gradient biases. It should even correct for unknown biases. This is clearly a technique with considerable potential.

Finally, an attempt is made to reduce the random errors in the POTENT field recoveries. This is done by utilising a little-appreciated aspect of the POTENT smoothing method. Although POTENT only requires the radial component of the peculiar velocity field from the smoothing, it can actually provide estimates of the non-radial components as well since it effectively fits to a three-dimensional bulk flow within the smoothing window. The idea behind the Max-flow method (sect. 5.3) is to perform the POTENT integral over mildly non-radial paths that avoid areas particularly sparse in galaxies. Unfortunately, because of the rather arbitrary weighting given to each path and the need in this crude implementation to impose a regular grid on the field, the net results, rather than improving the recoveries actually make the velocity and density fields *more* noisy. Nevertheless, the basic applicability of the technique is clearly demonstrated and in sect. 7.1 I will outline some possible improvements.

Of course, no matter how good the technique adopted is, it relies heavily on the quality of the distance estimates and, equally importantly, on accurate knowledge of that quality. In chapter 3, therefore, I consider the calibration of distance estimators to get both relative and absolute distances. It is clear that calibration using a single cluster is prone to large errors in the slope and zero-point of the relationship, probably enough to make it impossible to distinguish between 'direct' and 'inverse' estimators. Also, the tempting approach of choosing a particularly 'tight' cluster is a dangerous one since this is obviously a selection effect which could have large, unseen consequences. I, therefore, explore and extend the approach proposed by Lattimer (1993) where several clusters are carefully combined to form a single large cluster with a subsequent decrease in the slope and zero-point errors. Tests showed that this method could substantially improve 'inverse' estimates, but improvements in 'direct' calibration were less clear-cut. Nevertheless, since the iterative Monte Carlo correction

method can, in theory, be applied to any sensible distance estimates, 'inverse' estimates are perfectly adequate and have, indeed, some aesthetic advantages (see sect. 5.2.4).

Absolute distances are then derived by assuming that there is no systematic deviation from Hubble flow when averaged over the entire survey volume. The distance scale can then be set to minimise the scatter about a Hubble recession velocity.

The final stage of the thesis, therefore, is to apply all the techniques described and tested in chapters 2 to 5 to a set of genuine galaxy catalogues. This is done in chapter 6. The Mathewson et al. (1992) and Burstein (1990) catalogues were used with a large part of the Mathewson data set replaced with improved data provided by Salucci (1994). Also, a number of the Burstein galaxies are removed for a variety of reasons. The velocity and density fields are then reconstructed from this carefully selected and calibrated catalogue using most of the POTENT-like methods described previously. The results are, in general, consistent with the tests and also with other results obtained by other groups. Given the improvement displayed by the iterative corrections previously, this is taken as the "best guess". When compared to the most recent results from the POTENT collaboration (Dekel, 1994) a number of significant differences are seen. In particular, the "Great Attractor" is far less prominent in my result. The interpretation of this is complicated by the different data sets used and a detailed analysis will have to wait until the Burstein Mark III data set is available, but this is a clear indication of the dangers of the biases in POTENT.

Given these reservations, however, it is still possible to get useful information even from the results of the limited catalogue assembled here. As an example, in section 6.5, a simple technique devised by Dekel and Rees (1993) is used to get a lower bound on the density factor Ω_0. Treating the errors conservatively and working as close to the origin as possible to minimise the problems, this analysis gives the limit of $\Omega > 0.3$ at the 2σ level.

However, before getting down to the nitty-gritty of velocity and density fields, in the rest of this chapter I will give a introduction to the study of large scale structure.

1.2   Observing the Universe

The study of the large scale structure of the universe (LSS) forms the backbone of our attempts to examine the origins and nature of the universe around us. However, despite the preoccupation with cosmology throughout the recorded history of humanity, it is only in the last century that we have been able to confidently observe beyond the narrow confines of our own galaxy.

The study of extragalactic astronomy really began in the 1920s when Edwin Hubble was observing the distribution of "nebulae" (other galaxies) both over the sky and as a function of their apparent magnitude. He was able, somewhat tentatively, to conclude that the distribution of objects did not seem to fall off rapidly and was at least broadly consistent with a homogeneous distribution of objects out to the limits of his survey (Hubble, 1926). Two years later, Hubble published further results relating the estimated distances to clusters of galaxies (using the apparent magnitude of the fifth brightest member) to their recession velocity or redshift. The simple linear relationship that he discovered has become known as Hubble's Law and it was soon realised that, although seemingly indicative of some cataclysmic event centred on our humble galaxy throwing matter away from us, because of its linear nature, it could also have a much less egocentric interpretation. If the universe were expanding uniformly, then *any* observer would perceive the same recession.

The stage was set, therefore, for the study of LSS. Hubble's observations gave the first indications of a homogeneous, expanding universe and the earlier appearance of General Relativity provided the mathematical framework to attempt to model this. I will return to this mathematical description in sect. 1.3, but first there are several more constraints that observations have been able to impose.

1.2.1   The Homogeneous Universe

Hubble's original results extended only fraction of the distance available with modern telescopes and recording devices. However, the most extended surveys now available such as those of high redshift QSOs and radio galaxies show a high degree of isotropy. Perhaps even more significant is the discovery of the Cosmic Microwave Background Radiation (CMBR). The CMBR was first detected by Penzias and Wilson in 1965. They were using a radio
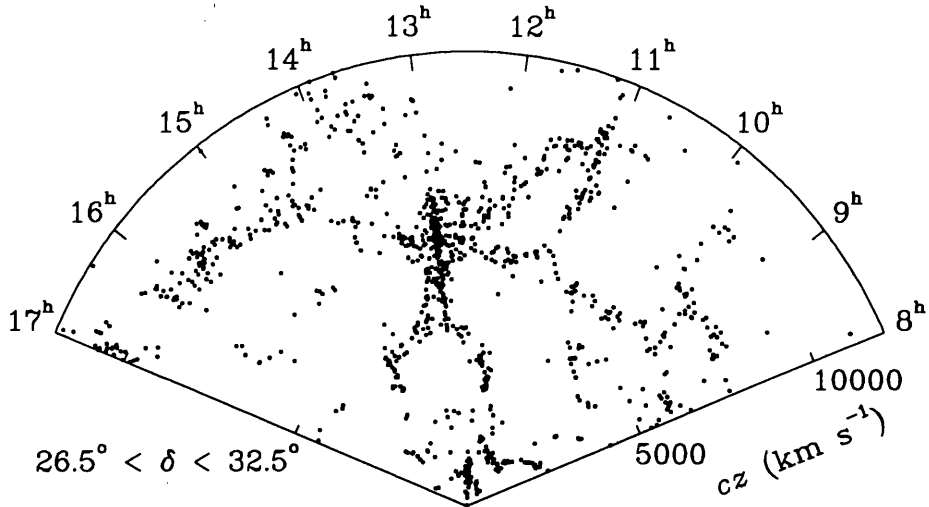
Figure 1.1: A section the local universe from de Lapparent et al. (1991). A thin slice of the universe was considered centred on Coma with the distances given by the redshifts of the galaxies.

horn at Bell Laboratories in Holmdel, New Jersey and developing techniques to calibrate the antenna to previously unknown levels of accuracy. However, they discovered a constant background noise in all directions on the sky which would not disappear no matter how well they prepared the horn. Its isotropy seemed to remove the possibility of a solar or galactic origin, so it was clear that here was an important cosmological measurement. In fact, at the same time, a group at Princeton led by Robert Dicke were studying possible observational consequences of the Big Bang theory and predicted the presence of just such background which should form a black-body spectrum of a few Kelvin (see sect. 1.3.1 for more details). Further experiments from balloon based instruments and, notably, the COBE satellite confirmed the black-body nature of the spectrum to extraordinary accuracy and fixed the temperature to 2.735 K (Mather et al., 1990). In addition, the isotropy of the radiation remains secure down to 1 part in 10^5 (Wright et al., 1992). However, not all observations are so smooth.

1.2.2   Observed Structure

Figure 1.1 shows one of the first clear measurements of large scale clustering of galaxies. A magnitude limited sample of galaxies from a thin slice of the sky is shown with the distance

being given by the redshift. As can be seen, there is considerable 'clumping' including some very large structure such as the Great Wall stretching off both sides of the slice and the "CfA Man" in the centre. More recent redshift surveys including IRAS (Strauss et al., 1990) and QDOT (Efstathiou et al., 1990) have spread over the entire sky and to much greater depths and the scale of structure grows with the surveys to encompass clusters and superclusters of galaxies and extensive voids.

There is also some much more immediate evidence for structure in the universe. The density within a galaxy is about 10^5 times the average density of the universe and on *very* small scales we have humans with a density of about 10^{25} times more than that! We therefore have the contrast between high levels of structure on small to medium scales giving way to an extraordinary level of homogeneity on very large scales (greater than about 100 Mpc). Reconciling these two extremes is the major task in the study of LSS today.

1.3   The Standard Cosmological Model

Before we can look at the theory for structure formation, we need to set up a mathematical basis – a cosmological model. To do so, we take three basic ingredients:

1. An assumption of Homogeneity and Isotropy (the Cosmological Principle).

2. The observed expansion of the universe.

3. Einsteins field equations (General Relativity).

The model is created by solving 3 with the constraints imposed by 1 and 2. Einstein's field equations can be written in the form

$$R_{\alpha\beta} - \frac{1}{2} R \; g_{\alpha\beta} \equiv G_{\alpha\beta} = 8\pi \; G \; T_{\alpha\beta} + \Lambda \; g_{\alpha\beta} \qquad (1.1)$$

where $R^{\alpha\beta}$ is the Ricci tensor[1] $G_{\alpha\beta}$ is the Einstein tensor, $T_{\alpha\beta}$ is the stress-energy tensor and Λ is the cosmological constant which will be taken to be zero from here onwards unless otherwise stated. Also, $g_{\alpha\beta}$ is the metric for the spacetime and must have spatial sections that are homogeneous and isotropic if the cosmological principle is to be satisfied. The

---

[1] I will be using the convention that greek indices (α, β etc) will vary from 0 to 3 whereas roman indices (i, j etc) will cover only the spatial indices 1 to 3.

most general metric with these properties is the Robertson-Walker (RW) metric with the line element

$$ds^2 = dt^2 - \text{R}^2(t) \left\{ \frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right\} \qquad (1.2)$$

with $c = 1$ and $\mathbf{x} = \{r, \theta, \phi\}$ is a comoving coordinate and the metric convention is $g_{00} = 1$, $\text{sign}(g_{ii}) = -1$. R is the scale factor of the universe and incorporates the expansion. k takes the values -1, 0 or 1 and is used to denote the form of the curvature of the spacetime. From this metric the non-zero components of the Ricci tensor can be derived (eg Narlikar (1993) chapter 4), then we just need to calculate $T_{\alpha\beta}$.

From the symmetries of the metric we know that $T_{\alpha\beta}$ must be diagonal and from the assumption of isotropy, all the spatial parts must be equal. One simple solution of this form is the *perfect fluid*

$$T^{\mu\nu} = (p + \rho)u^\mu u^\nu - p \, g^{\mu\nu} \qquad (1.3)$$

where ρ is the energy density, p the pressure and u^μ is the four-velocity of a fluid element. It follows from the Bianchi identities, $G^{\mu\nu}{}_{;\nu} \equiv 0$, that.

$$T^{\mu\nu}{}_{;\nu} = 0 \qquad (1.4)$$

For the perfect fluid in (1.3), this gives

$$d(\rho\text{R}^3) = -p \, d(\text{R}^3) \qquad (1.5)$$

the first law of thermodynamics. Of course, ρ and p can always be related by some value ω (ie $p = \omega\rho$), and assuming that this value is time independent we have

$$\rho \propto \text{R}^{-3(1+\omega)} \qquad (1.6)$$

For example, for radiation we know that $\rho = 3p$ so $\rho \propto \text{R}^{-4}$ and for matter, $p \approx 0$ giving $\rho \propto \text{R}^{-3}$. Another interesting case is when $\rho = -p$. This occurs when the universe is dominated by vacuum energy and $\rho \propto$ const. In general, we shall see that in the early stages of the universe, radiation dominates whereas at later times, matter domination takes over. An inflationary regime occurs whenever vacuum energy dominates. I will not be considering inflation in any detail, but a brief outline is given in sect. 1.3.2.

We now have all of the tools we need to solve Einstein's equations. The $0 - 0$ component gives the Friedmann equation.

$$\frac{\dot{\text{R}}^2}{\text{R}^2} + \frac{k}{\text{R}^2} = \frac{8\pi}{3}G\rho \qquad (1.7)$$

and each $i - i$ component gives

$$2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = -8\pi G p \qquad (1.8)$$

From these two equations, an expression for the acceleration of the scale factor $\ddot{R}$ can be obtained

$$\frac{\ddot{R}}{R} = -\frac{4\pi}{3}G(\rho + 3p). \qquad (1.9)$$

We know that currently, the universe is expanding (ie $\dot{R}_0 > 0$ where the subscript 0 indicates the *present* value so ρ_0 is the present energy density etc.). Therefore, if in the past, $\rho + 3p$ was positive, then at some time (usually taken as $t = 0$) R must have been zero. This clearly corresponds to the Big Bang and is an unavoidable singularity in the model. Close to the moment of the singularity, classical GR cannot describe the state of the universe, but at sufficiently late times (which are not that late at all!) we can describe much of what should have been happening.

To do this in a useful way, we need to have a limited number of parameters to look at. Two such values which are very useful and should be fairly readily measurable in the current epoch are the Hubble parameter H and the density parameter Ω (together with their present values H_0 and Ω_0).

$$H \equiv \frac{\dot{R}}{R}, \qquad H_0 = 100 \; h_0 \; \text{km s}^{-1} \; \text{Mpc}^{-1} \qquad (1.10)$$

$$\Omega \equiv \frac{\rho}{\rho_c}, \qquad \rho_c \equiv \frac{3H^2}{8\pi G} \qquad (1.11)$$

where ρ_c is the *critical density* such that

$$k = +1 \quad \Longrightarrow \quad \Omega > 1 \qquad \text{Closed Universe}$$
$$k = 0 \quad \Longrightarrow \quad \Omega = 1 \qquad \text{Flat Universe}$$
$$k = -1 \quad \Longrightarrow \quad \Omega < 1 \qquad \text{Open Universe} \qquad (1.12)$$

1.3.1 A Brief History of the Universe

Figure 1.2 gives a schematic outline of the various stages in the evolution of the universe from moments after the Big Bang up to the present epoch.

Time (seconds)

10^{18}

10^{12}

10^{6}

10^{0}

10^{-6}

10^{-12}

10^{-18}

10^{-24}

10^{-30}

10^{-36}

10^{-42}

Big
Bang

Quantum
Gravity

GUT
Era

Electroweak
Unification

Nucleosynthesis

Matter Domination

Formation of atoms
and Decoupling

NOW
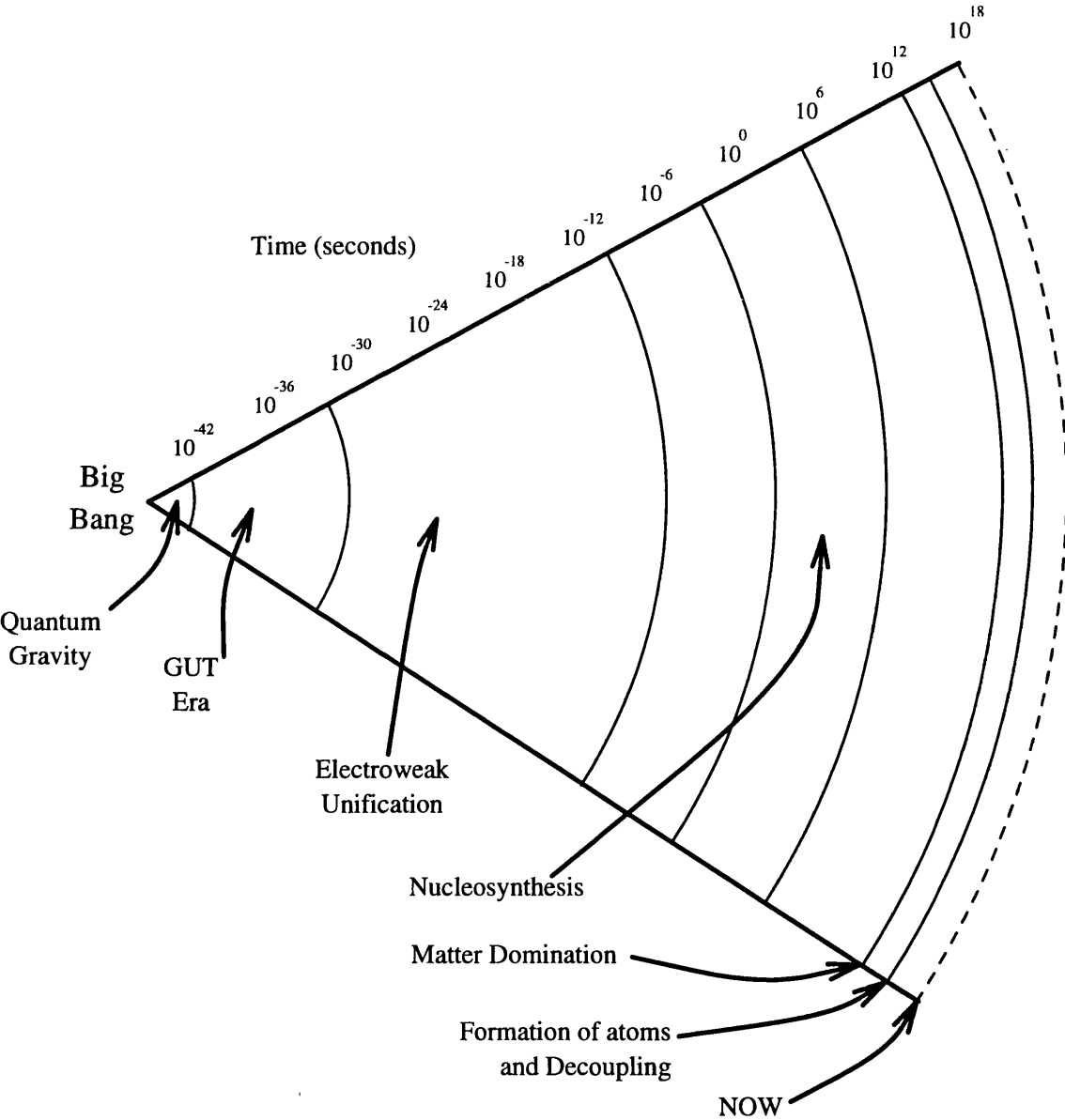
Figure 1.2: A schematic history of the Universe as given by Big Bang theory.

The Planck Epoch: $t < 10^{-43}$ sec, $T > 10^{19}$ GeV

This is the limiting epoch for classical GR. For times earlier than this, a quantum theory of gravity is needed.

The GUT Era: up to $t \sim 10^{-34}$ sec, $T \sim 10^{15}$ GeV

After the Planck epoch, we enter the realm of the Grand Unified Theories (GUTs) of particle physics. Here, the universe is a hot soup of relativistic particles such as quarks and leptons and at sufficient temperature for the fundamental forces excluding gravity to be indistinguishable (hence unification theories). However, as the temperature drops below $\sim 10^{15}$ GeV, this unification breaks and there is a *phase transition*. It is at this point that an inflationary stage might occur (see sect. 1.3.2).

The Electroweak Era: up to $t \sim 10^{-12}$ sec, $T \sim 10^3$ GeV

Another phase transition occurs at $\sim 10^3$ GeV when the electromagnetic and nuclear weak forces separate. Soon after this quarks begin to bind together to form baryons and mesons. This is known as the *Quark-Hadron transition.*

Primordial Nucleosynthesis: $10^{-2} < t < 10^2$ sec, $T \sim 10$ to 0.1 MeV

The era of *Primordial Nucleosynthesis* provides the earliest direct test of the standard Big Bang theory. As the temperature falls below the nuclear binding energy, atoms (or rather nuclei) begin to form. The details of the process are beyond the scope of this thesis, but the theory predicts specific relative abundances of the light elements and imposes an upper limit on Ω_B – the fraction of the critical density made up from baryonic matter.

Matter Domination: $t \sim 10^{12}$ sec, $T \sim 10$ eV

We saw in eqn. (1.6) that for our model,

$$\rho_{\text{rad}} \quad \propto \quad R^{-4}$$

$$\rho_{\text{matter}} \quad \propto \quad R^{-3}$$

so, even with radiation dominating the energy density at early times, there will eventually come an epoch when matter dominates. This regime marks the beginning of the actual formation of structure.

Decoupling: $t \sim 10^{13}$ sec, $T \sim 1$ eV

Very soon after matter domination, the temperature finally falls below the atomic binding energy of the light elements created during the nucleosynthesis phase. However, as the atoms form, the interaction between photons and the previously charged particles is effectively cut off and the radiation *decouples* from the matter. This leaves two distinct 'soups': one of light elements such as H, D (Deuterium), ^{3}He, ^{4}He and ^{7}Li, and another of free flowing photons. The photons continue unhindered and gradually lose energy by gravitational redshift in the expanding universe. It is these photons that form the CMBR having 'cooled' from their decoupling temperature of about 3000K to the 2.7K observed today. Note that this epoch is very important in the process of structure formation and its apparently coincidental position so close to matter dominance is still a source of some speculation.

Up to the present: $t \sim 10^{18}$ sec, $T = 2.735$K

After decoupling, the business of structure formation really gets under way with galaxies forming at about $t = 10^9$ years.

So, that is a basic outline of Big Bang theory. In general it has proved to be remarkably resilient, with the discovery of the CMBR setting it up firmly as the theory to beat. However, there have been problems associated with some aspects of the theory and a number of variants and adjustments have been proposed. Again, the details of all these are beyond the scope of this introduction, but it is interesting and instructive to glance briefly at one of them to get some feel for the sort of changes that are envisaged.

1.3.2 Inflation

Three of the main questions raised by the standard Big Bang model are the *Horizon*, *Flatness* and *Monopole* problems. The horizon problem arises from the very property that

enabled us to develop the standard model – the isotropy and homogeneity of the large scale universe. Why is it so smooth? In particular, how can this be consistent with the finite velocity of light? Ignoring the perhaps somewhat contrived solution that the universe was created homogeneous, let us assume that some physical process acted to smooth out the cosmos at some early time. As this time, the maximum scale of any physical interaction will be the particle horizon which is limited by the finite age of the universe and velocity of light. When brought forward to the current epoch, this region of the universe will be very much smaller than the current horizon. In fact, it could be as little as 1 metre! (eg Narlikar (1993) chapter 6 or Kolb and Turner (1990) chapter 8). This would mean that we could not really expect to see homogeneity on *any* reasonable cosmological scale.

The flatness problem arises because of the evolution of the curvature of the universe. At the present time, the curvature is such that Ω_0 is pretty close to 1 (certainly not as much as 5 or less than 0.05 anyway). However, in order have such values at our epoch, Ω must have been *much* closer to 1 at any earlier epoch (for example, the GUT era). To demonstrate this, recall the Friedmann equation (1.7). If this is rewritten in terms of Ω, we can obtain the following:

$$\frac{k}{R^2} = (\Omega - 1)\,\frac{\dot{R}^2}{R^2} \tag{1.13}$$

Taking $R \propto t^{1/2} \propto T^{-1}$ and substituting for values at the current epoch where appropriate, we see

$$(\Omega - 1) = 4\,(\Omega_0 - 1)\,H_0^2\,t_0^2\,\frac{T^2}{T_0^2} \quad \text{for } k = \pm 1 \tag{1.14}$$

So, for typical values for the GUT era,

$$\Omega_{GUT} - 1 \approx 10^{-53}\,(\Omega_0 - 1) \tag{1.15}$$

Therefore, an $\Omega_0 \approx 1$ universe is a surprising coincidence that requires *very* little deviation from $\Omega = 1$ at early epochs. This level of "fine tuning" is uncomfortable when not backed up by any strong reasoning and requires some addition to the standard theory.

The final of the three problems involves the Grand Unified Theories. It is an unavoidable result of the symmetry breaking in these theories that at least one magnetic monopole is produced within each horizon. However, not only are they not currently detected, but given their high stability and large mass (typically $\sim 10^{16}$ GeV), their combined density should be far in excess of ρ_c.

So, we have three apparent discrepancies with the standard theory – how do we go about solving them? It is clear that, to some extent, the problems share a similarity in their dependence on *scale*. Therefore, some suitable adjustment of the scale of the universe might be a way forward. Such an approach is *Inflation* proposed by Alan Guth (1981).

Inflation occurs at the breakdown of the GUT symmetry. At this point, the state of lowest energy (or vacuum state) changes and if the true vacuum energy is not immediately obtained, we have a situation with a finite, positive vacuum energy density. This will dominate the total energy density and lead to the situation mentioned earlier where ρ is independent of the scale factor R. We then get a *de Sitter phase* with the universe expanding exponentially, $R(t) \propto \exp(Ht)$, until the true vacuum is reached.

The net result of this massive increase in R is highly fortuitous for the three problems outlined above. After inflation, the observable horizon is a very small fraction of the horizon pre-inflation so even at the present epoch, there is a time when the entire observable universe was causally connected. The monopole problem is dealt with in the same way. There should only be a few monopoles per pre-inflation horizon, so there would be at most a handful in the entire observable universe today. Finally, the flatness problem is solved since no matter what the spatial curvature before inflation, the radius R after inflation will be much greater than the horizon scale, giving an extremely flat universe. In fact, inflation can be considered as *predicting* $\Omega_0 = 1$, and any other result will be a considerable hurdle to overcome.

Thus we see how the concept of inflation helps to explain some of the problems with the standard Big Bang model. For a *much* more detailed description of inflation in a number of its flavours, see Narlikar and Padmanabhan (1991).

1.3.3 From Homogeneity to Structure

On the very large scale, therefore, we have a remarkably smooth universe and a highly successful model. The presence of the CMBR and its smooth, black-body nature give enough weight to the model to ensure longevity and many of the problems with the details can be overcome with testable 'fixes' such as inflation.

However, as it stands, the Big Bang model does very little to explain the smaller scale distribution of mass - how and when do galaxies form? and why does the particular hierarchy

of clusters and superclusters that we observe come about? To answer these questions, or at least to attempt to, we need another extension to the theory.

1.4 Structure Formation in the Universe

The problem of the formation of galaxies and other large scale structures in the universe at the current epoch is one of the most vexed problems in cosmology today. Although the observational evidence is greater than in most areas, or perhaps because of this, there is little or no consensus about the processes involved. Even if one restricts oneself to the big bang cosmology outlined above, there are many unanswered fundamental questions. However, by making some reasonable assumptions and some perhaps slightly less reasonable but essential ones, it is possible to derive a sequence of events that could be expected to lead to structures similar to those we see around us, and also predict some observable properties that can be used to refine or if necessary rewrite or discard the theory.

One such theory, usually considered the *Standard Model* of structure formation, is the evolution of density fluctuations by gravitational instability (GI). Here, small "early" fluctuations are amplified by self-gravity to form individual structures. The basis, therefore, is the evolution of such fluctuations with time in the expanding universe.

1.4.1 Jeans Instability in an Expanding Fluid

The development of Jeans, or gravitational, instabilities involves solving the Eulerian equations of motion for small perturbations of the underlying state of the matter. So, given the usual Newtonian equations

$$\frac{\partial \rho}{\partial t} + \nabla.(\rho \mathbf{v}) = 0$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}.\nabla)\mathbf{v} + \frac{1}{\rho}\nabla p + \nabla \Phi = 0$$

$$\nabla^2 \Phi = 4\pi G \rho \tag{1.16}$$

where Φ is the gravitational potential, we can choose a suitable base solution and perturb it. The simplest base is static ($\mathbf{v}_0 = 0$) and spatially uniform ($\rho_0 = $ const, $p_0 = $ const) with perturbations:

$$\mathcal{X} = \mathcal{X}_0 + \mathcal{X}_1 \tag{1.17}$$

where $\mathcal{X} \equiv \{\rho, \ p, \ \mathbf{v}, \ \Phi\}$

(Note that the '0' subscript now refers to the unperturbed value rather than the value at the current epoch). We will assume, for simplicity, that the equation of state relating the pressure and density has no spatial variation implying a sound speed v_s given by

$$v_s^2 = \frac{p_1}{\rho_1} \tag{1.18}$$

Using this simplifying term, the solutions of the equations of motion to first order in the perturbations are

$$\frac{\partial \rho_1}{\partial t} + \rho_0 \nabla.\mathbf{v}_1 = 0$$

$$\frac{\partial \mathbf{v}_1}{\partial t} + \frac{v_s^2}{\rho_0}\nabla \rho_1 + \nabla \Phi_1 = 0$$

$$\nabla^2 \Phi_1 = 4\pi G \rho_1 \tag{1.19}$$

However, the simple static base we have assumed is not suitable for an expanding universe. For a suitable expanding fluid, our unperturbed base is given by

$$\rho_0 = \rho_0(t_0)\mathrm{R}^{-3}, \qquad \mathbf{v}_0 = \frac{\dot{\mathrm{R}}}{\mathrm{R}}\,\mathbf{r}, \qquad \nabla \Phi_0 = \frac{4}{3}\pi G \rho_0\,\mathbf{r} \tag{1.20}$$

Again solving to first order, we now obtain

$$\frac{\partial \rho_1}{\partial t} + 3\frac{\dot{\mathrm{R}}}{\mathrm{R}}\rho_1 + \frac{\dot{\mathrm{R}}}{\mathrm{R}}(\mathbf{r}.\nabla)\rho_1 + \rho_0 \nabla.\mathbf{v}_1 = 0$$

$$\frac{\partial \mathbf{v}_1}{\partial t} + \frac{\dot{\mathrm{R}}}{\mathrm{R}}\mathbf{v}_1 + \frac{\dot{\mathrm{R}}}{\mathrm{R}}(\mathbf{r}.\nabla)\mathbf{v}_1 + \frac{v_s^2}{\rho_0}\nabla \rho_1 + \nabla \Phi_1 = 0$$

$$\nabla^2 \Phi_1 = 4\pi G \rho_1 \tag{1.21}$$

It is convenient to express the density fluctuation as a contrast to the underlying density such that $\delta = \rho_1/\rho_0$. This, together with $\mathbf{v}_1$ and Φ_1 can then be expressed in fourier components

$$\mathcal{X}(\mathbf{r}, t) = (2\pi)^{-3} \int \mathcal{X}_{\mathbf{k}}(t) \exp\left[-i\frac{\mathbf{k}.\mathbf{r}}{\mathrm{R}(t)}\right] d^3 r \tag{1.22}$$

where $\mathcal{X} = \{\delta, \ \mathbf{v}_1, \ \Phi_1\}$.

One further useful adjustment is to divide the perturbed velocity field into its rotational ($\mathbf{v}_\perp$) and irrotational components ($\mathbf{v}_\parallel$) components such that

$$\mathbf{v}_1 = \mathbf{v}_\perp + \mathbf{v}_\parallel \tag{1.23}$$

The equations (1.21), expressed in terms of the fourier components of the field then show that

$$\mathbf{v}_\perp \propto \mathrm{R}^{-1}(t) \qquad (1.24)$$

In other words, the rotational modes *decay* with time so it is reasonable to assume that, if no more rotational modes are created, the velocity field will eventually become entirely irrotational. This will have significant and useful consequences for the velocity field recovery methods described in this thesis.

But to return to the evolution of the density fluctuations. One of the perturbed equations of motion that can be derived from (1.21) describes the time development of $\delta_\mathbf{k}$

$$\ddot{\delta}_\mathbf{k} + 2\frac{\dot{\mathrm{R}}}{\mathrm{R}}\dot{\delta}_\mathbf{k} + \left(\frac{v_s^2 k^2}{\mathrm{R}^2} - 4\pi G\rho_0\right)\delta_\mathbf{k} = 0 \qquad (1.25)$$

It is clear from this that an important factor in the evolution of density perturbations is whether $v_s^2 k^2/\mathrm{R}^2$ is greater or less than $4\pi G\rho_0$. We therefore define the *Jeans Wavenumber*

$$\mathrm{k}_J^2 = 4\pi G\rho_0 \mathrm{R}^2/v_s^2 \qquad (1.26)$$

where $\mathrm{k} = |\mathbf{k}|$. When $\mathrm{k} \gg \mathrm{k}_J$ the perturbations oscillate as a sound wave with a gradually decreasing amplitude. However, when $\mathrm{k} \ll \mathrm{k}_J$, there are growing modes to the solution and the density fluctuation can grow and lead on to the formation of structures.

The physical basis for this is simple. The development of the fluctuation is a contest between the self-gravity of the system trying to collapse into a dense region and the pressure and expansion of the universe holding the collapse back. Below a certain scale, the fluctuation does not contain sufficient mass to overcome the pressure and expansion forces, but above the scale defined by the Jeans wavenumber, the self-gravity begins to dominate and the density contrast increases.

What we need to consider, therefore, is the evolution of the Jeans scale. (This is often expressed as a mass (the Jeans mass) which is defined as the mass contained within a sphere of of radius π/k_J). As we follow this parameter through the various epochs of the big bang history, we should be able to get a handle on how and when the structure formed and also see how the model can be tested and constrained by observations.

1.4.2 Forming the Perturbations

The analysis so far has considered the evolution of some small fluctuations that were created at some earlier time. However, there is no consensus exactly when these could be formed in the real universe. One possibility is that they are formed before the Planck era, but as this is exactly the region where our current understanding of physics breaks down, it seems a little convenient to stuff all the problems into this period, and it is also difficult to see why *any* particular set of initial conditions that deviate from homogeneity should be preferred over any other. Another, more comfortable option is during the early phase transitions (for example inflation or cosmic string theories). However, whenever they are formed exactly, it is sufficient for our purposes to say that it is very early (certainly less that $t \sim 10^{-34}$ seconds).

We also need to be able to describe the form of the perturbations, both at this initial epoch and at later stages. The normal way of doing this is through the power spectrum of the fluctuations $P(\mathbf{k})$. In doing this, it is implicitly assumed that the nature of the fluctuations is not a function of position, but only of scale, however, this is a perfectly reasonable assumption and is certainly preferable to trying to follow the growth of individual perturbations over all space! Using this spectrum, we can set the initial conditions at some early epoch to be a particular form. Given that the processes that form this are unknown, it is necessary to parametrise the relation, for example as a power law

$$P(\mathbf{k}) = A\mathbf{k}^n \qquad (1.27)$$

where A can be fixed from observations. Obviously, for a general spectrum, n should be a function of k, but without any reason to do this, it is convenient to keep it constant. In fact, the most commonly used spectrum has $n = 1$ (the *Harrison-Zel'dovich* spectrum) which corresponds to having the gravitational energy equally distributed over all scales. Inflationary models can also favour a spectrum of this form.

1.4.3 Fluctuations before Decoupling

When we look at very early times, we see that the universe is radiation dominated and that matter is coupled to the radiation. This means that the sound speed is just $v_s^2 = \frac{1}{3}$. The

corresponding Jeans wavenumber is

$$k_J^2 = 3 \times 4\pi G \rho R^2 \qquad (1.28)$$

Which expressed in physical coordinates is

$$k_{J_{phys}} = \frac{k_J}{R} = (12\pi G \rho)^{1/2} \qquad (1.29)$$

The mass in baryonic matter contained within this is

$$M_{J_B} = \frac{4\pi}{3} \rho_B \left(\frac{\pi}{k_{J_{phys}}} \right)^3 \qquad (1.30)$$

However, the total baryonic mass within the horizon is

$$M_{HOR_B} = \frac{4\pi}{3} \rho_B t^3 \qquad (1.31)$$

$$= \left(\frac{1}{\pi} t\, k_{J_{phys}} \right)^3 M_{J_B} \qquad (1.32)$$

Clearly, the all fluctuations sufficiently large to grow are larger than the horizon. In order to examine such fluctuations properly, the Newtonian analysis given above is insufficient and we need to employ a fully general relativistic theory. However, for our qualitative survey, a simpler approach will suffice.

First, it is important to realise that there are two types of perturbations to consider. The first – '*Adiabatic*' or curvature fluctuations – are straightforward fluctuations in the density $\delta\rho \neq 0$. However, it is also possible to have fluctuations where $\delta\rho = 0$, but the equation of state varies spatially, for example by varying the relative contributions to ρ from photons and matter. These are called *Isocurvature* fluctuations. When a fluctuation is sub-horizon sized, the isocurvature form becomes an adiabatic fluctuation as the matter redistributes itself, but when they are super-horizon sized, causality prevents this from happening and the two types must be considered separately.

For adiabatic fluctuations, consider the simple case of a flat FRW universe. The Friedmann equation (1.7) gives

$$H^2 = \frac{8}{3} \pi G \rho_0 \qquad (1.33)$$

For another region with the same expansion rate, but a slightly higher density ρ_1 we have

$$H^2 = \frac{8}{3} \pi G \rho_1 - \frac{k}{R^2} \qquad (1.34)$$

The density contrast between the two models is, therefore,

$$\delta \propto \frac{R^{-2}}{\rho} \tag{1.35}$$

which means that

$$\delta \propto \begin{cases} R^2 & \propto t & \text{Radiation Dominated} \\ R & \propto t^{2/3} & \text{Matter Dominated} \end{cases}$$

Isocurvature fluctuations do not evolve when super-horizon sized.

1.4.4 Sub-horizon Sized Perturbations

Eventually, the perturbations enter the horizon and cease growing. Isocurvature fluctuations evolve into simple density contrasts and everything happily sits there and expands with the universe. However, as we approach decoupling, things begin to change. The first thing that happens is that the fluctuations begin to be damped.

Collisional or Silk Damping

As the universe nears decoupling, the photon mean free path λ_γ becomes very long and photons can 'drag' the charged baryons out of overdense regions into underdense ones, thus damping the fluctuations. A crude analysis of can show the scale of this effect.

Consider the random walk of a photon

$$(\Delta r)^2 \approx \frac{\Delta t \, \lambda_\gamma(t)}{R^2(t)} \tag{1.36}$$

In the total time up to decoupling, the diffusion length of the photon is given by

$$\lambda_S^2 = \int_0^{t_{dec}} \frac{\lambda_\gamma}{R^2} dt \tag{1.37}$$

Also, since λ_γ is until quite close to decoupling, we can assume that universe is matter dominated and $R \sim (t \times 10^{-17})^{2/3}$ where t is in seconds (Kolb and Turner, 1990 page 354). Therefore, λ_S^2 is of the order of the photon diffusion length at the epoch of recombination, and this defines a scale (the *Silk mass*) below which fluctuations are erased at decoupling.

So far, structures are not forming particularly well. Although some super-horizon sized fluctuations have grown for a period, they grow at best $\propto t$ and small scale fluctuations are wiped out at recombination. However, after this epoch, there is a dramatic change.

1.4.5 Post-decoupling Structure Evolution

The big change occurs because, since matter is no longer tied to the photons, $v_s^2 \neq 1/3$. This means that the pressure drops precipitously and M_{J_B} plunges inside the horizon. We can, therefore, use a Newtonian analysis to describe the further evolution of the perturbations. Recall from eqn. (1.25) that

$$\ddot{\delta}_{\mathrm{k}} + 2\frac{\dot{\mathrm{R}}}{\mathrm{R}}\dot{\delta}_{\mathrm{k}} + \left(\frac{v_s^2 k^2}{\mathrm{R}^2} - 4\pi G\rho_0\right)\delta_{\mathrm{k}} = 0 \qquad (1.38)$$

If we assume that $\mathrm{k} \ll \mathrm{k}_J$ and a $k = 0$ FRW universe, we have

$$\ddot{\delta}_{\mathrm{k}} + \frac{4}{3t}\dot{\delta}_{\mathrm{k}} - \frac{2}{3t^2}\delta_{\mathrm{k}} = 0 \qquad (1.39)$$

which has a growing mode solution of the form

$$\delta_{\mathrm{k}+}(t) \propto t^{2/3} \qquad (1.40)$$

so all perturbations greater than the Jeans mass, which now lies comfortably within the horizon, continue to grow, albeit slowly. However, our analysis so far has assumed *small* perturbations. What happens as δ grows, particularly when it reaches ~ 1? Then the fluctuation begins to break away from the expansion of the universe and we need a new, non-linear approach.

1.4.6 Non-linear Evolution

Unfortunately, the non-linear regime of structure formation, when $\delta \geq 1$ and the fluctuations break away from the expansion of the universe completely is very difficult to treat analytically. Indeed, to produce anything reasonably like real structures such as galaxies can only be achieved numerically such as through large N-body simulations. However, some techniques do exist from probing some way into the non-linear arena, the foremost being the *Zel'dovich Approximation*.

The Zel'dovich Approximation

The Zel'dovich approximation (Zel'dovich, 1970) is a way of following the growing mode of fluctuations some way beyond the $\delta = 1$ barrier without the need for an impossibly complex

non-linear approach. The approximation starts by phrasing the problem in Lagrangian coordinates **q** where **q** labels some fluid element by its position at some 'initial' time. This is related to the more customary *Eulerian* coordinates **r** at some time t by a displacement function $\mathcal{D}$

$$\mathbf{r} = \mathrm{R}(t)[\mathbf{q} + \mathcal{D}(\mathbf{q}, t)] \tag{1.41}$$

The approximation comes in expressing $\mathcal{D}$ as a combination of a purely spatial function and a purely time dependent one:

$$\mathbf{x} = \frac{\mathbf{r}}{\mathrm{R}} = \mathbf{q} + D(t)\psi(\mathbf{q}) \tag{1.42}$$

Then, from simple mass conservation,

$$\rho(\mathbf{r}, t) = \frac{\rho_0}{\mathrm{R}^3} \left\| \frac{\partial \mathbf{r}}{\partial \mathbf{q}} \right\|^{-1} \tag{1.43}$$

where $\| \dots \|$ represents the determinant of the Jacobian transformation. Given the eigenvalues of the matrix in the Jacobian λ_1, λ_2 and λ_3, the density can be written as

$$\rho(\mathbf{r}, t) = \frac{\rho_0}{\mathrm{R}^3} (1 - D\lambda_1)^{-1} (1 - D\lambda_2)^{-1} (1 - D\lambda_3)^{-1} \tag{1.44}$$

Taking, without loss of generality $\lambda_1 \geq \lambda_2 \geq \lambda_3$, we see that as $D(t)$ grows, the density becomes infinite as $D\lambda_1 \to \infty$. The density fluctuation is, therefore, forming a 2-dimensional sheet called a Zel'dovich pancake. However, when the approximation reaches this stage, it breaks down as orbits begin to cross. At this point, the Lagrangian coordinates are no longer valid as they do not uniquely determine a particular fluid element. The approximation fails to account for the interaction between the particles crossing and they continue apparently undisturbed (see fig 1.3). One very successful extension to the Zel'dovich approximation is also shown in the figure. In the *Adhesion approximation*, a viscosity is introduced which has the effect of "sticking" particles together as they cross. However, sufficiently outside the pancake region, the viscosity is negligible, so the pure Zel'dovich result is obtained.

Between them, the Zel'dovich approximation and its adhesion extension provide a highly successful quasi-linear treatment of fluctuation evolution up to a δ of several. In addition, it can be used along with other techniques (such as the *truncated Zel'dovich*, see Sathyaprakash et al. (1994)) as an input to N-body simulations to significantly increase efficiency. However, beyond $\delta \sim 5$ to 10, the approximation is too dogged by orbit crossings on small scales to be useful and it becomes the realm of the number-crunchers.

Actual orbits   Zeldovich   Adhesion
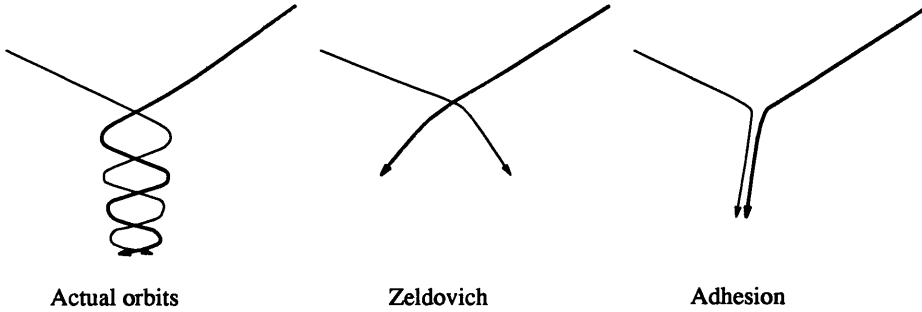
Figure 1.3: The effect of the Zel'dovich and adhesion approximations on two particles crossing orbits. The actual behaviour is shown on the left with the Zel'dovich approximation all but ignoring the orbit crossing and the adhesion approach effectively replacing the two particles with one large particle.

1.4.7 Dark Matter

Throughout the analysis in this section, we have been assuming that the universe is composed exclusively of photons and baryonic matter. However, primordial nucleosynthesis puts severe constraints on Ω_B (the fraction of the critical density in baryonic form). In fact, $\Omega_B < 0.15$. However, we have also seen that inflation predicts $\Omega = 1$, so how can we reconcile this seeming disparity? The only way would seem to be *Dark Matter*. This is, in general, non-baryonic matter which interacts weakly or not at all with other constituents except by gravity (hence dark). There are a number of possible candidates, no one of which has any overwhelming case for priority, but as far as our argument is concerned, they can be divided into two camps – Hot and Cold dark matter where the distinction is simply that hot matter is relativistic.

Hot Dark Matter (HDM)

The main hot dark matter candidates are massive neutrinos. These have a severe effect on short wavelength fluctuations because, being collisionless, they can readily diffuse from overdense into underdense regions, thus damping the fluctuations (note the distinction between this *collisionless* damping and Silk damping). The effect of this is to cut off all perturbations below $\lambda_C \sim 40(m_\mu/30\text{eV})^{-1}$ Mpc where m_μ is the mass of the neutrino.

The first structures to form, therefore, are of this size – roughly a supercluster – and the

final non-linear collapse of these objects must have occurred relatively recently ($z < 4$). The good news is that numerical simulations of HDM universes show structure remarkably like observed (eg fig. 1.1). However, galaxies form far too late to be consistent with the observation of radio galaxies out to $z > 1$ to 2.

HDM, therefore, has problems with its top-down nature of structure formation, and is likely to drop out of contention. However, the discovery of a 30 eV neutrino species could easily revive its chances!

Cold Dark Matter (CDM)

Cold dark matter candidates are the products of particle theories such as Axions, and none have been detected as yet. However, their effect on structure formation is convenient for the standard model and CDM has, until recently, been the fore-runner in the dark matter race. Because they are not collisionless like HDM, CDM particles do not damp out the small scale fluctuations. In addition, not being charged particles, fluctuations in the dark matter can survive the Silk damping effect as the photons do not drag the non-baryons around. Finally, quite large fluctuations can exist at decoupling and not be imprinted on the CMBR which the baryons can then 'fall into'. This, in contrast to the HDM top-down approach, gives a bottom-up scenario with galaxies forming quite early and grouping together to form the larger clusters and then superclusters. Unfortunately, the amount of galaxy clustering in CDM simulations can only be made to match observations with $\Omega_0 h \sim 0.2$, which stretches H_0 uncomfortably for an $\Omega = 1$ inflationary universe.

Mixed Dark Matter (MDM)

The most recent attempts to solve these problems involve a judicious combination of Hot and Cold matter – Mixed Dark Matter. The idea is to combine the attractive properties of both scenarios while avoiding their shortfalls. So far, the results seem to be attractive (cf Pogosyan (1993) and Davis et al. (1992)), but the extra parameter of the relative abundances of HDM and CDM perhaps leaves things a little too unconstrained for testing what is, fundamentally, a model based on a number of assumptions with no direct corroboration. This is something that may well be resolved over the next few years as there are more

observational constraints such as measurements of detailed, small scale fluctuations on the CMBR and advances in experimental particle physics, but at present is still very much an open issue.

The study of LSS, therefore, is an active and exciting one with many things uncertain, but the possibility of significant advances in the near future. One area of research that has borne fruit over the past decade and promises to go on doing so is the study of the velocity fields produced by the flow of galaxies and other matter in the universe at the current epoch. This area forms the core of my work for this thesis.

Chapter 2

Velocity Field Reconstruction

If the dull substance of my flesh were thought,

Injurious distance should not stop my way.

Sonnet XLIV

The observational study of LSS is, naturally, limited to objects and events very near to time t_0. Even the study of the CMBR is strongly influenced by relatively recent events such as possible distortions produced by gravitational waves rather than initial fluctuations, and the main sources of information – galaxies and quasars – are by their very nature from the very late, highly non-linear phase of structure formation. Therefore, the problem facing cosmologists is one of *initial conditions.* To understand the universe we need to know the quantity and form of any dark matter, we need to know something of the very early spectrum of fluctuations and so on. These early parameters will remain forever unobservable directly, and it is unlikely that we could reverse the flow of structure formation and recover them from the present structure due to the quantity of filtering that has occurred. However, particle and cosmological theory have combined together to provide a wealth of possibilities all of which can, or at least should, make predictions about the state of observable systems. In this way, the theoretical models can be refined or rejected as appropriate, and our knowledge of the origins and nature of the universe improved.

2.1   The Peculiar Velocity Field and its Measurement

The most obvious probe of the current state of structure in the universe is the distribution of galaxies, both on the sky (2-dimensional surveys) and through space (3-dimensional surveys such as that shown in fig. 1.1). However, such positional surveys do not exhaust the available information. As well as positional information forming a density field, the motions of the galaxies give rise to a *velocity* field. The most obvious aspect of this is the universal expansion, but when this is removed there can remain significant motions. These motions form the *Peculiar Velocity Field.*

One important reason for wishing to measure this field is because the peculiar velocities act to distort the distribution of galaxies in *redshift* space when compared to real space. For example, the redshift map of a cluster of galaxies will be elongated along the line of sight due to random motions within the cluster leading to "Finger of God" effects and on a large scale it is not unreasonable to suppose that large mass concentrations associated with superclusters, for example, might lead to some form of infall from neighbouring regions, further distorting the redshift picture. However, peculiar velocities are far more useful as direct probes of the dynamical state of the universe.

One particular problem that plagues the analysis of redshift surveys is *cosmological biasing.* From a survey, one is observing not the density contrast δ as one would wish, but the distribution of *luminous* matter. To resolve this ambiguity, some knowledge of how the luminous matter is related to the dark (or "how light traces mass") is needed, but this is still totally unknown. The normal assumption, more for its simplicity in the lack of any further information than any other reason, is that there is a universal relationship governed by a single parameter b such that

$$\left[\frac{\delta_n}{n}\right]_{lum} = b\,\delta \tag{2.1}$$

where $[\delta_n/n]_{lum}$ is the galaxy number density contrast. However, even if the biasing para-meter is fit to during some recovery process, the rather *ad hoc* form of the relationship is far from satisfactory. However, the *motions* of the galaxies are gravitationally induced and dependent only on the mass field. Therefore, if some way can be found to map the velocity field, this will help to understand the density field and the biasing itself.

2.1.1   An Observational History of Peculiar Velocities

The earliest firm measurements of systematic deviations from quiet Hubble flow were by
Rubin et al. (1973). They used apparent magnitude to estimate the distances to a selection
of spiral galaxies in the redshift range 3500 $km\,s^{-1}$ to 6500 $km\,s^{-1}$. These distances were
then compared to the redshifts to estimate the radial peculiar velocities of the galaxies
and the variation across the sky considered. They found a systematic variation which was
consistent with a Local Group motion of 450 ± 125 $km\,s^{-1}$ towards $l = 163°$, $b = -11°$.
This large value was considerably above expectations and several other groups attempted
to reproduce the results with very little success (see, for example, Hart and Davies, 1982 or
de Vaucouleurs and Peters, 1984). However, studies of the microwave background indicated
a considerable dipole motion even when corrected for Earth's solar system and internal
galactic motions. Although the magnitude of this dipole was of a similar scale to the Rubin
result, the direction was significantly different: $l = 269°$, $b = 28°$. Therefore, given that
both results are correct, a fairly complex model of local motions starts to emerge.

The first major attempt to *map* the local velocity field was performed by Lynden-Bell et al.
(1988) in a highly influential paper. The team prepared a survey of more than 400 elliptical
galaxies out to $z \sim 8000$ $km\,s^{-1}$ with redshift and photometric data were available. They
estimated the distances to the galaxies using the D_n-σ relation (see sect. 2.2.1 for more
details of such techniques) and from these results estimated the radial component of the
peculiar velocity of each galaxy.

To analyse this data, they constructed a parametric velocity field model that incorporated
ideas from all the previous studies and attempted to fit the data using maximum-likelihood
techniques. This 'best-fit' model involved a large mass concentration at about 4000 $km\,s^{-1}$
in the $l = 307°$, $b = 9°$ direction which is known as the Great Attractor. In order to
reproduce the large 500+ $km\,s^{-1}$ motion of the local group the total mass of this attractor
must be in the region of 5×10^{16} $M_\odot$.

These results, and in particular the Lynden-Bell et al. survey, were the beginnings of a
thriving field of study. However, the reliance of the conclusions on the velocity model
gave the impetus to search for other techniques as free as possible from initial assumptions
whether it be through more complicated models or, ideally, from a model-free standpoint.

This approach has been made possible by the huge increase in the size of redshift surveys due to improved observational techniques and equipment, and has led to a dramatic increase in the use (and, no doubt, abuse) of velocity fields.

2.1.2   Velocity Field Reconstruction From Redshift Surveys

IRAS techniques

The most generally used technique for a relatively model-independent velocity field recovery from redshift data was devised for surveys based on the IRAS catalogue of infra-red sources. There were two initial surveys. The Strauss et al., 1990 catalogue consists of all extra-galactic sources in the catalogue with a 60μm flux greater than 1.2 Jansky – a total of more than 2500 objects up to ~ 3000 km s^{-1}. The QDOT survey of Efstathiou et al., 1990 also contains of the order of 2000 galaxies, but extends down to 0.6 Jansky (about 7000 km s^{-1}) with one in six galaxies selected at random. Since then, catalogues have increased dramatically in size but the techniques for analysing velocity fields remain largely unaltered.

That technique is iterative in nature. Under the assumption of a quiet Hubble regime where the redshift and real space distributions are identical, the number density contrast is calculated smoothed on some suitable scale (on the order of 1000 km s^{-1}). The scale is chosen so that the density contrast, δ, can be assumed to be not much more than 1 and linear theory can be used to calculate the velocity field. In this regime, the velocity field is simply proportional to the "gravitational acceleration" $\mathbf{g}$,

$$\mathbf{v} = -\frac{f(\Omega)}{3H\Omega}\mathbf{g} = \frac{f(\Omega)}{3H\Omega}\frac{\nabla\Phi}{\mathbf{R}} \tag{2.2}$$

where $f(\Omega)$ is some function of Ω that can be well approximated by (see Peebles (1980)).

$$f(\Omega) = \Omega^{0.6} \tag{2.3}$$

Solving the Eulerian equations of motion (1.21) it can be found that

$$\mathbf{v}(\mathbf{r}) = \frac{2}{3}f(\Omega)H^{-1}G\,\mathbf{R}\,\rho_c\int\frac{(\mathbf{r}'-\mathbf{r})}{|\mathbf{r}'-\mathbf{r}|^3}\delta(\mathbf{r})\,d^3r' \tag{2.4}$$

The peculiar velocities recovered by this method can then be used to correct the distance estimates from the redshifts and the process repeated iteratively.

There are, however, two problems with this process. The integral in (2.4) should be over all space but is, naturally, limited to the survey volume and secondly, the cosmological bias parameter b enters into the expression for δ. It is, therefore, necessary to determine it, or rather the factor $\beta = b/f(\Omega)$, as part of the iteration process. This is a useful way of estimating the value of β, but it would perhaps be preferable to have a method that was not based on the assumption of b as a universal parameter, so that that assumption could be checked. I will return to this later.

Spherical Harmonic Reconstruction

Another approach to velocity and density reconstruction is described in Fisher et al. (1994) (see also Hoffman, 1993 and Scharf et al., 1992). Here, the distribution of galaxies is expanded as a set of spherical harmonics and radial Bessel functions. Given a value of β and applying linear theory they attempt to correct for redshift distortions due to peculiar motions. This is an inversion procedure and the application of a "Wiener Filter" helps to deal with incomplete sky coverage and noise giving a minimum variance estimate of the harmonics of the real space distribution. From this, if wished, the peculiar velocities can be calculated.[1]

Wavelet Analysis

A related approach is proposed by Rauzy et al., 1993. They represent the velocity field as a wavelet transform, thereby providing a natural form of smoothing on any chosen scale and also allowing easy analysis of the scale dependence of structure. Unlike the other methods in this section, they require distance estimates to the galaxies (see below) and there are problems moving from the uneven distribution of galaxies to the regular 'support' required by the wavelet form (Rauzy et al., 1994b), but the elegant and useful result is likely to more than compensate for the effort required to overcome these problems.

---

[1] Heavens and Taylor, 1994 also apply spherical harmonics to the study of redshift distortions, but here the aim is to estimate β and not to map the fields.

2.2 Galaxy Distance Estimation

Most of the methods described above infer the peculiar velocity field given only information about the radial motion of the galaxies involved. However, if we can determine the actual distance to a galaxy in some redshift independent way, we can use this to find at least the radial component of the velocity v_r directly

$$v_r = cz - H_0 r \qquad (2.5)$$

This is information that can clearly help to considerably enhance our understanding of the velocity field, but how do we go about estimating distances to galaxies?

Let us consider one very simple distance estimator. One readily determined observational property of a galaxy is its apparent magnitude m which is related to the *absolute* magnitude M of the galaxy by[2]

$$\log_{10} r = 0.2(m - M - 25) \qquad (2.6)$$

where r is in Mpc. However, if we assume that the galaxy is drawn from some population of galaxies with their absolute magnitudes randomly distributed about some mean M_0 we can *estimate* the distance to the galaxy:

$$\hat{\omega} \equiv \widehat{\log \mathbf{r}} = 0.2(\mathbf{m} - M_0 - 25) \qquad (2.7)$$

Before continuing, we should note that this equation introduces a number of notational points. Firstly, up to now we have been using a bold typeface to indicate vectors (eg $\mathbf{r}$). However, it is convenient when discussing these statistical problems to use the standard statistical notation where a bold character indicates a statistical variable – one drawn from a distribution. I will, therefore, subsequently use the $\vec{r}$ notation to distinguish vectors. In addition, estimators of some quantity or parameter will have a 'hat' (eg $\hat{\omega}$).

Since $\mathbf{M}$ is sampled from the luminosity function, both it and $\mathbf{m}$ should be considered statistical variables. In the limiting case where the entire population of galaxies is observable, it is clear that

$$E(\hat{\omega}|\omega_0, M_0) = \omega_0 \qquad (2.8)$$

---

[2]From here I will always be using logarithms in base 10 unless otherwise stated, so the subscript will be dropped.

where E denotes the expectation value averaged over the population. However, in reality there is a definite limit on the magnitude level that can be observed and galaxies whose apparent magnitude fall foul of this limit will be selected out of the sample. With the average now taken over this reduced population, we see that $E(M|r_0, M_0) \neq M_0$, and hence that

$$E(\hat{r}|r_0, M_0) \neq r_0 \text{ and } E(\hat{\omega}|\omega_0, M_0) \neq \omega_0 \tag{2.9}$$

Thus these estimators of both r_0 and ω_0 are, in the statistical sense of the word, biased. This problem plagues estimators based around only magnitude (see Hendry and Simmons, 1990) but the judicious use of a second observable can significantly improve the situation.

2.2.1 Estimators Using Two Observables

We will again consider a simple case. Let us take one observable that is distance dependent and subject to selection problems, **m**, and its absolute counterpart **M**. In addition, a second observable, **P**, is available but this has no selection problems and is not distance dependent (for example, line width). If M and P are highly correlated, the observed value **P** can be used to estimate a value for **M**, ie $\hat{M}$. This can then be used in place of M_0 in (2.7) to give an improved estimate of ω_0. An example of such an estimator is the Tully–Fisher (TF) estimator which uses magnitude and line width or the D_n-σ relation between apparent diameter and velocity dispersion.

With these relations, the correlation between **M** and **P** is usually found by some kind of linear regression. For example, for log distance ω_0,

$$m - M = 5\omega_0 - 25 \tag{2.10}$$

M can be estimated by assuming a linear relationship between **M** and **P** such that

$$\hat{M} = E(M|P) = aP - b \tag{2.11}$$

where a and b are constants. Then, $\hat{\omega}$ for a TF estimator takes the form

$$\hat{\omega} = 0.2(m - aP - b - 25) \tag{2.12}$$

In general $\hat{\omega}$ is still a biased estimator. In Hendry and Simmons (1994), however, it is shown that if the selection function is only dependent on apparent magnitude and not on P, then
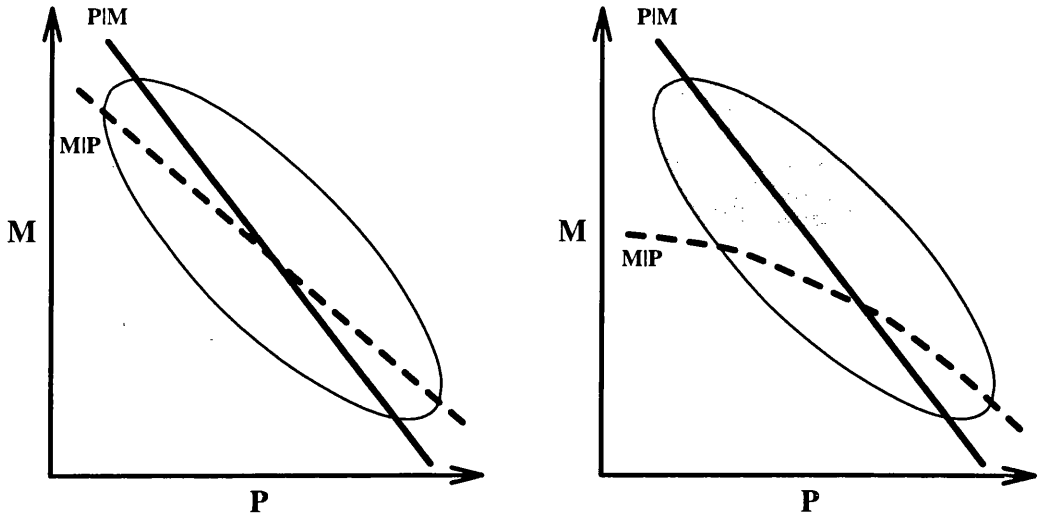
Figure 2.1: These two diagrams demonstrate the effect of selection on two choices of regression line through the parameter distribution. Both diagrams show an example distribution of galaxies taken, for simplicity, to all be at the same distance. On the left, there is no selection and the two regression lines **M|P** corresponding to $\hat{\omega}_{DTF}$ and **P|M** for $\hat{\omega}_{ITF}$ are clear. On the right, however, a sharp cutoff in apparent magnitude has been imposed, removing the shaded region from the distribution. Now, although the **P|M** line is unchanged, the **M|P** line is strongly effected.

$\hat{\omega}$ is unbiased for all ω_0, provided that a and b in eqns. (2.11) and (2.12) are derived from a regression of **P** on **M**, the so-called 'Inverse Tully Fisher' (ITF) relation. The estimator, $\hat{\omega}_{DTF}$, corresponding to a 'Direct' regression of **M** on **P** is, on the other hand, biased for all ω_0. This can be seen for a simple case in fig. 2.1. Here, a sharp magnitude cutoff has been imposed (such as one would find for a magnitude limited survey) on a group of galaxies at the same distance (eg a cluster). The ITF estimates are unaffected by the magnitude selection, but the DTF estimates become biased. This is because, although in both cases the parameters $\hat{\sigma}_M$, $\hat{\sigma}_P$ and $\hat{\rho}$ differ from the intrinsic values, the slope of the ITF regression is $\hat{\sigma}_M/\hat{\sigma}_P\hat{\rho}$ which is unaffected. Note that, although in the unselected case the regressions are different and will give different estimates of distance for a given galaxy, they are *both* unbiased. The unbiased property of the ITF relation under selection was first pointed out by Schechter (Schechter, 1980), and has been generally recognised in the literature, although few discussions have approached the subject in a fully rigorous manner.

2.2.2 Calibrating and Using Distance Estimators

We can, therefore, devise a variety of distance estimators to galaxies each with its own identifiable statistical properties. As we shall see later, which properties are desirable will vary from situation to situation, but we will almost always want to reduce the bias and variance as much as possible. Therefore, we will only be considering estimators that make use of at least *two* observables.

The next stage in the process, having chosen an estimator, is to calibrate the relation for a given galaxy survey. To do this, we need to know the values of a number of parameters. For simplicity and clarity we will choose the intrinsic joint distribution of $\mathbf{M}$ and $\mathbf{P}$ to be a bivariate normal. We will then need M_0 and P_0 which mark the 'centre' of the distribution shown in fig. 2.1 and σ_M and σ_P, the dispersions. Finally we need to know ρ, the correlation coefficient of the distribution. From these we can calculate the parameters of the two estimators outlined above (as well as any of a number of others – see Hendry (1992) for some examples).

Assuming a sharp magnitude cut-off, both estimators take the form

$$5\hat{\omega} = (\mathbf{m} - m_L) - \Delta(\mathbf{P} - P_0) \tag{2.13}$$

where m_L is the magnitude limit and Δ is the slope of the regression given by

$$
\begin{aligned}
\Delta_{\text{DTF}} &= \frac{\sigma_M \rho}{\sigma_P} \\
\Delta_{\text{ITF}} &= \frac{\sigma_M}{\sigma_P \rho}
\end{aligned}
\tag{2.14}
$$

Given these parameters, therefore, we can proceed with estimating the distances of all the galaxies in our sample. However, we obviously do not know them *a priori*, so they too must be *estimated* from the catalogue.

This is not a trivial task as the observable $\mathbf{m}$ is distance dependent and we would need to know the distance to all the galaxies in advance in order to get the distribution of $\mathbf{M}$ and $\mathbf{P}$! However, one way out of this is to find a subset of the galaxies that can be safely assumed to be at the same distance – a cluster. For these galaxies, the distribution of $p(\mathbf{m}, \mathbf{P})$ will have the same form as $p(\mathbf{M}, \mathbf{P})$ and $\hat{\mathbf{P}}_0$, $\hat{\sigma}_M$, $\hat{\sigma}_P$ and $\hat{\rho}$ can easily be determined. From these we can calculate *relative* distance between galaxies, for example expressing all distances in

terms of the Virgo distance, but to get *absolute* distance we need to find a value for $\hat{M}_0$. This is one of the weakest links in the chain for distance estimation, but if a reasonable absolute distance can be determined to just one cluster – for example by cepheid distances or redshifts corrected for peculiar motions by some velocity model – then this can be used to set all the others. Note that an incorrect zero-point derived in this way will lead to an effective error on the Hubble constant in eqn. (2.5): in other words a spurious outflow or inflow.

I mentioned above that different forms of distance estimator can be chosen to provide different selections of properties. Clearly if an estimator is available with very small errors of any type (systematic or random), then this is the best one to choose for the majority of situations! However, even with the best data, TF estimators typically have errors of 15% to 20% and the exact form of estimator can become very significant. Again, a simple example will serve to demonstrate this.

I have shown that the ITF estimator is unbiased in that it has the property $E(\hat{\omega}_{\mathrm{ITF}}|\omega_0) = \omega_0$. However, if we wish to determine the value of H_0 (a very common sport among cosmologists) we might assume that peculiar velocities are negligible and simply use the expression $H_0 = cz/\hat{r}$ and average over some catalogue of galaxies. But clearly, just because $\hat{\omega}_{\mathrm{ITF}}$ is unbiased, there is no reason to suppose that $\hat{r} = 10^{\hat{\omega}_{\mathrm{ITF}}}$ will be. Therefore, in order to obtain an unbiased estimate of H_0, we need to correct for the bias introduced into $\hat{r}$. Assuming that the distribution $p(\mathbf{M}, \mathbf{P})$ is a bivariate normal, $\hat{\omega}_{\mathrm{ITF}}$ will have gaussian errors with some variance σ^2 and it can easily be shown that distance estimates of the form

$$\hat{r} = 10^{-\frac{1}{2}(\log_e 10)\,\sigma^2}\,10^{\hat{\omega}_{\mathrm{ITF}}} \qquad (2.15)$$

will be unbiased and give an unbiased estimate of H_0.

Clearly, considerable care is needed when using estimators with such large errors. The above effect would be negligible for errors of $\sim 1\%$, but becomes quite important up at the 10% level when the bias in H_0 can become comparable to the random errors for any reasonable catalogue of galaxies. Unfortunately, not all problems are as easy to solve as this example.

2.2.3 Malmquist Bias

Consider a further example. We wish to study the radial components of the peculiar velocity in a particular direction on the sky. In fact, these velocities are negligible, but the distribution of galaxies is highly clumped at about 5000 $km\,s^{-1}$. An ITF like distance estimator is used to give unbiased distance estimates to each galaxy and $\hat{v}_r$ is calculated using a relation like eqn. (2.5). To further simplify the problem, we will assume that there is no selection and *all* galaxies out to 10000 $km\,s^{-1}$ are observed. The distribution of estim-
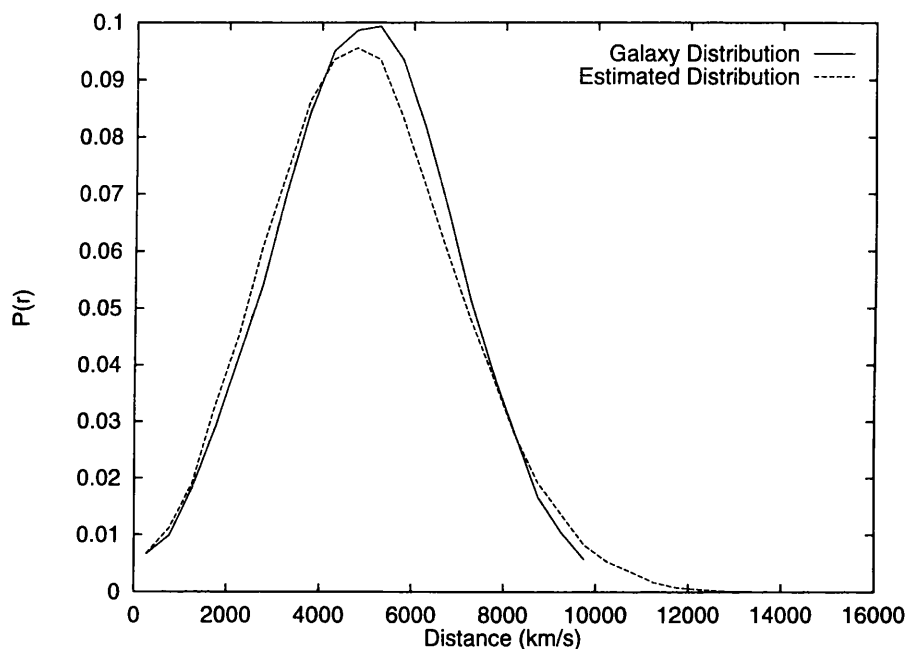
Figure 2.2: A comparison of the distribution of estimated distances (error 15%) and the intrinsic distribution of galaxies. Galaxies are divided among bins of 500 $km\,s^{-1}$ width. The two distributions show very good agreement.

ated distances is compared to the underlying distribution in fig. 2.2. As would be expected from unbiased distances, there is good agreement: a slight increase in the variance of the distribution due to the distance errors, but no appreciable shift in the mean. However we are not particularly interested in the distances themselves, only in what they can tell us about the peculiar velocity field. Given the large errors in the distance estimates (about 15%), it is necessary to perform some averaging on $\hat{v}_r$: this is done by binning the galaxies into a series of radial bins. The results of this process is shown in fig. 2.3. Here we can see
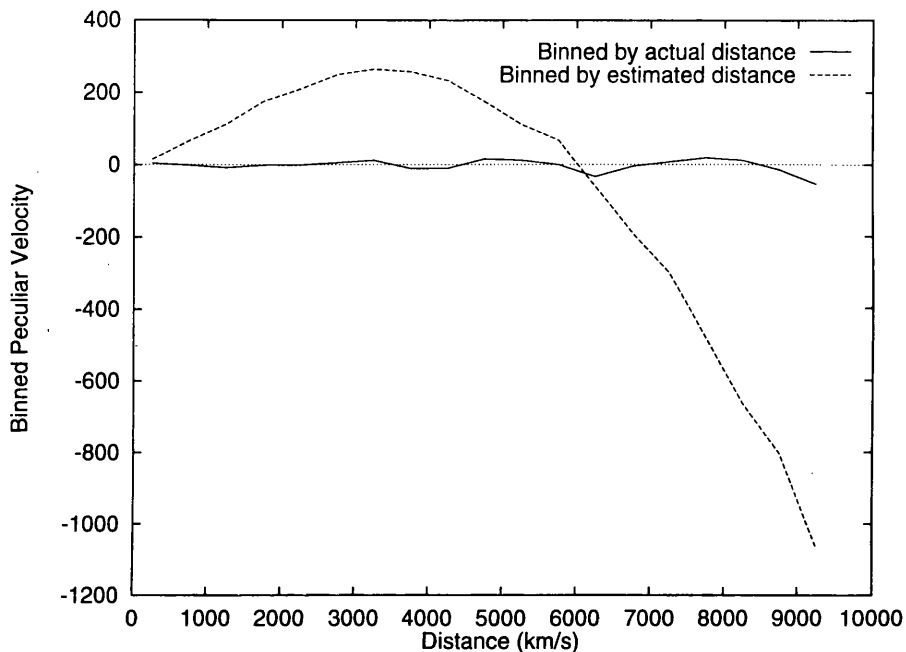
Figure 2.3: The estimated peculiar velocities of the galaxies in fig. 2.2 are averaged in radial bins. There is a marked difference between the derived peculiar velocities when the binning is by estimated or real distance.

the insidious effect of *Malmquist bias* (cf. Hendry et al., 1993b). Although when we bin by the true distances we recover the expected Hubble flow, when the galaxies are grouped according to *apparent* position, there is a large infall seen towards the centre of the galaxy clump. For $r < 5000$ km s^{-1}, the majority of galaxies have been scattered *down* from the clump and have an average $\hat{v}_r$ towards the centroid, and the opposite for $r > 5000$ km s^{-1}.

We therefore have large systematic errors introduced that are strongly dependent not just on the properties of the distance estimator, but also the underlying distribution of galaxies that the sample is drawn from, which is obviously unknown. One approach to dealing with these biases is to attempt to derive a distance estimator with properties that counteract those of the Malmquist bias – Malmquist Corrected distance estimates.

In interpolating a peculiar velocity from galaxies appearing in the catalogue to a given spatial point with radial coordinate s, the *essential* effect of the windowing is to pick out the galaxy whose *estimated* position is nearest to the prescribed point. This galaxy's actual distance could be radically different, and will, as we have seen, depend on the true spatial

distribution of galaxies. By requiring that on average the actual radial coordinate of the galaxy deemed to be closest to the grid point equals s one would ensure also that on average the correct peculiar velocity would be ascribed to s. Expressed mathematically we require

$$E(\mathbf{r}_0|\hat{\mathbf{r}} = s) = s \qquad (2.16)$$

or for work involving log distances

$$E(\omega_0|\hat{\omega} = \omega_s) = \omega_s \qquad (2.17)$$

Note that this definition is *not* the definition of an unbiased estimator ($E(\hat{\mathbf{r}}|r_0) = r_0$) and the frequent use of the term "unbiased" in connection with this property in the literature has led to considerable confusion and even acrimony. However, in the situation described above and many others that frequently occur in the study of Large Scale Structure, it is preferable to unbiased estimates. The important point, as ever, is to be consistent and to ensure that the choice of estimator matches the demands of the problem.

But enough sermonising. How do we derive an estimator with this property? The expectation value in (2.17) is calculated as

$$E(\omega_0|\hat{\omega}) = \int \omega_0\, p(\omega_0|\hat{\omega})\, d\omega_0 \qquad (2.18)$$

where $p(\omega_0|\hat{\omega})$ is the probability density function (pdf) of ω_0 given the estimated value $\hat{\omega}$. This is an unknown, of course, but we can use *Bayes' theorem* to derive an expression for it,

$$p(\omega_0|\hat{\omega}) = \frac{p(\hat{\omega}|\omega_0)\, p(\omega_0)}{\int p(\hat{\omega}|\omega_0)\, p(\omega_0)\, d\omega_0} \qquad (2.19)$$

where $p(\hat{\omega}|\omega_0)$ is the distribution function of some distance estimator and $p(\omega_0)$ is the actual distribution of those galaxies that are observed (ie the actual distribution of galaxies convolved with a selection function). Note that implicit in this definition is the assumption that the observed galaxy distribution is a function of *distance* only and has no angular dependence. I will return to this assumption later.

Therefore, if we have some "raw" distance estimator of a known form we have $p(\hat{\omega}|\omega_0)$ and if we can estimate the distribution function $p(\omega_0)$ sufficiently well, we can calculate a correction for the distance estimates

$$\hat{\omega}_{corr} = \int \omega_0\, p(\omega_0|\hat{\omega}_{raw})\, d\omega_0 \qquad (2.20)$$

In general, it would seem to be sensible to keep the form of the uncorrected distance estimator as simple as possible, so the ITF estimator with its unbiased nature and simple gaussian form (when the joint distribution of **M** and **P** is bivariate normal) would seem to be a good choice.

The function $p(\omega_0)$ is known in Bayesian statistics as the *prior*, and the results of the analysis can be very sensitive to it – a poor estimate of it will give poor results. However, we are in danger of entering a circular problem as this distribution, in a sense, is what we need our distance estimates for!

One approach that is frequently used was first applied during the velocity field work of Lynden-Bell et al. (1988) mentioned above. They proposed approximating the distribution of galaxies as uniform. The form of $p(\omega_0)$ is then a simple power law because of the increase in the volume within a unit solid angle with distance. However, the form of the selection function $S(\omega_0)$ is still unknown and we really require the true distribution of *observed* objects, ie

$$p_{obs}(\omega_0) = p_{universe}(\omega_0)\, S(\omega_0) \qquad (2.21)$$

Here, the different properties of the DTF estimator come to the rescue. It can be shown (Hendry et al., 1994) that when using this biased estimator, the selection function effectively cancels with the bias. Therefore, the correction can be applied to $\hat{\omega}_{\mathrm{DTF}}$ with the underlying distribution of galaxies $p_{universe}(\omega_0)$ or to $\hat{\omega}_{\mathrm{ITF}}$ with the selected function $p_{obs}(\omega_0)$.

This correction is know as the Homogeneous Malmquist Correction (HMC) and has been much applied. However, the assumption of a homogeneous universe is not really a good one on these scales (if it were, there would be no peculiar velocities to measure) so Landy and Szalay (1992) proposed a different approach which they termed the General Malmquist Correction (now more frequently known as the Inhomogeneous Malmquist Correction or IMC).

With this correction, the prior is approximated by the distribution of *estimated* distances

$$\hat{\mathbf{p}}(\omega_0) = p(\hat{\omega}) \qquad (2.22)$$

This assumption will clearly only be reasonable if $\hat{\omega}$ is unbiased and has random errors of a size less than the scale of any features you wish to include. In fact this is not as restrictive as it may seem as the distribution is smoothed on quite a large scale before the correction

is applied to prevent small random fluctuations giving large spurious corrections, so the $\hat{\omega}_{\mathrm{ITF}}$ is in fact quite reasonable. In addition, the derivation of the correction is based on the assumption of gaussian errors in $\hat{\omega}$, so the ITF is again the best candidate (cf. Simmons, 1993).

A more significant problem is the lack of any form angular dependence, but the noise in the distribution $\hat{p}(\omega_0)$ becomes severe if the catalogue is split up into lots of "cones". This effect will be examined further in sect. 4.2.1. Note also that if the universal distribution of galaxies were known, for example from the analysis of a large redshift survey, the IMC formalism could easily be adapted to use the DTF estimator and another correction obtained. Again, I will return to this later in the thesis.

Chapter 3

Calibration of Distance Estimators

Linger your patience on; and we'll digest

The abuse of distance.

Henry V^{th} II:Prologue

Thus far, I have generally considered only the problems of distance estimators caused by the intrinsic scatter in whatever relation is used. However, the calibration of that relation will also be subject to errors which must be taken into account.

3.1   A Simple Calibration Method

The purpose of calibration is to determine the slope and zero point of the distance estimator. This is done by estimating the parameters of the $p(\mathbf{M}, \mathbf{P})$ distribution (ie σ_M, ρ etc). However, because we only know the apparent value of any distance dependent quantities and not its absolute counterpart, we need to use a cluster of galaxies to perform the calibration on since, as the galaxies can all be considered to be at the same distance,

$$\mathbf{m} - \langle \mathbf{m} \rangle \equiv \mathbf{M} - M_0 \qquad (3.1)$$

where $\langle \dots \rangle$ denotes the average over the cluster. Of course, this is not the only method. Han and Mould (1990) and (1992) simultaneously fit to the distance relation and a model velocity field which is used to correct the redshifts of the galaxies. However, by using a cluster the need for a model field is removed. Given estimates of the distribution parameters,
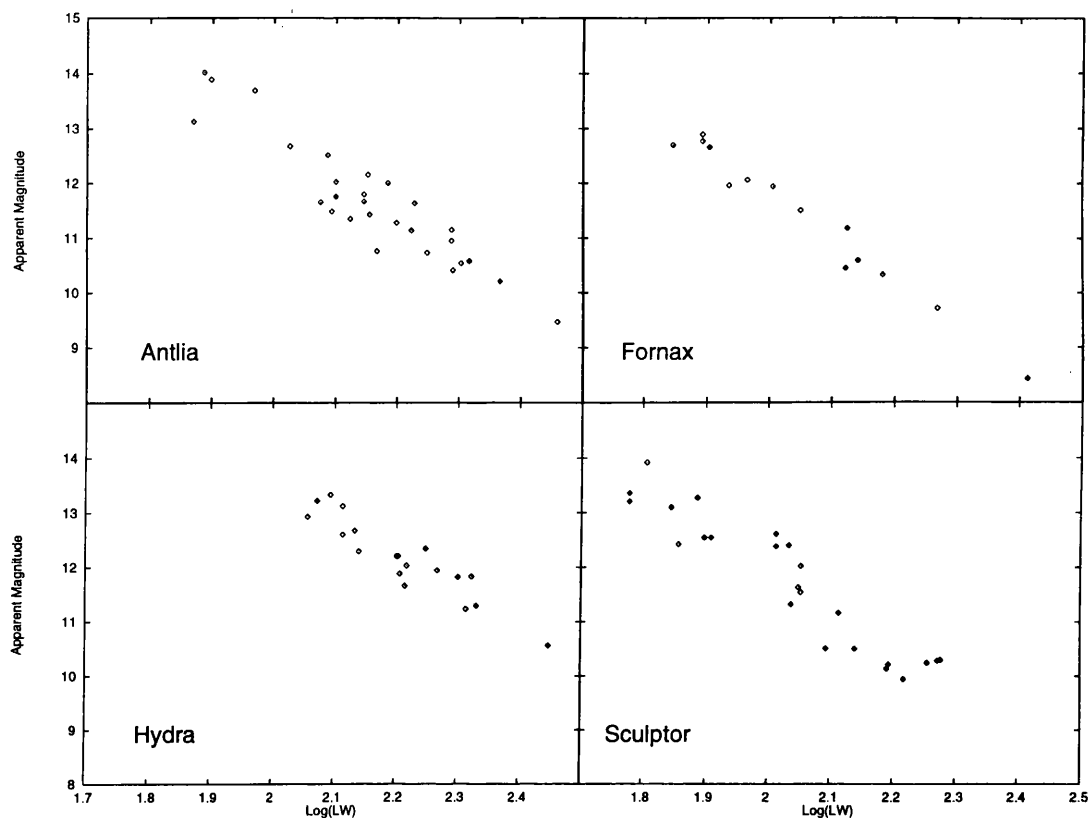
Figure 3.1: The distribution of **M** and **P** for four of the larger clusters in the Mathewson et al. (1992) catalogue. **M** is the magnitude (given here as apparent magnitude since all the galaxies are assumed to be at the same distance) and **P** is the log of the linewidth.

relative distances between all of the galaxies can be estimated, but for absolute distances, a calibration distance must be chosen. This is frequently done by choosing a reasonably nearby cluster where observational errors can be expected to be relatively small, and using a velocity field model in conjunction with the redshift to fix an absolute distance. Clearly, there are problems with this form of approach, particularly since we wish to avoid velocity models, but before considering the absolute calibration, I would like to look at the parameter estimation to see if there can be some way in which it can be improved.

3.1.1   Calibration from Composite Clusters

Figure 3.1 shows four possible candidates for a calibrating cluster. They are four of the larger clusters from the Mathewson et al. (1992) catalogue. The relationship between **M** and **P** is very clear for all four clusters, but none of them contain enough observed galaxies

to really determine the parameters of the distribution with any confidence. However, of the four, the Fornax relationship has the highest correlation coefficient, and therefore looks the tightest, and this is indeed the cluster chosen for the original calibration. The formal errors in the distances using this form of calibration are quite good, but it is dangerous to rely on them. By choosing the cluster because of the tightness of the distribution one is clearly introducing an unpleasant selection effect. As the tightest, Fornax *must* be, in some way, unusual and even if this does not strongly effect the slope of the distance estimator, it will lead to a significant under-estimate of the resultant distance errors since

$$\sigma_{\mathbf{M}|\mathbf{P}} = \sigma_{\mathbf{M}}\sqrt{1 - \rho^2} \qquad (3.2)$$

Perhaps even more importantly, with only 14 galaxies, the calibration will be very uncertain and sampling errors on the calibration slope and zero point introduced in this way would be very difficult to determine (Hendry and Simmons, 1994).

What is needed, therefore, is a calibration technique that can bring together the information from a number of clusters thereby both increasing the number of galaxies under consideration and avoiding any dangerous ad hoc choices.

The technique I have adopted was first developed by Lattimer and Hendry (see Lattimer, 1993 and Lattimer et al., 1994) and I have since joined them in extending it. (A similar approach is considered by Willick et al. (1994) and Young et al. (1993)). The basic idea of the method is to choose a set of clusters (for example all those with greater than, say, 10 galaxies) and 'slide' them together to form one larger cluster. Without selection, this would be a trivial task since one would just need to match up the mean apparent magnitudes, $\langle \mathbf{m} \rangle$, of each cluster. But selection complicates the procedure since the observed mean is biased with respect to the underlying mean, ie:

$$E(\langle \mathbf{m} \rangle) \neq m_0 \qquad (3.3)$$

where m_0 is the apparent magnitude of a galaxy with absolute magnitude M_0 at the distance of the cluster, ie $m_0 = M_0 - 5\,\omega_{\text{clust}} - 25$.

Correcting $\langle \mathbf{m} \rangle$ for Magnitude Selection

If the distribution $p(\mathbf{M}|\mathbf{P})$ is assumed to be gaussian, it can be seen (Hendry, 1994) that for a cluster,

$$\langle \mathbf{m} \rangle = m_0 - \frac{\exp\left[-\frac{1}{2}\left(\frac{m_L - m_0}{\sigma_M}\right)^2\right]}{\sqrt{2\pi}\,\Phi\left(\frac{m_L - m_0}{\sigma_M}\right)} \tag{3.4}$$

where $\Phi(x)$ is the cumulative normal distribution. Therefore, an iterative procedure can be devised to get a good estimate $\hat{m}_0$. Starting with the mean of the actual cluster distribution $\langle \mathbf{m} \rangle_0$, we obtain, after n iterations,

$$\langle \mathbf{m} \rangle_{n+1} = \langle \mathbf{m} \rangle_0 + \frac{\exp\left[-\frac{1}{2}\left(\frac{m_L - \langle \mathbf{m} \rangle_n}{\sigma_M}\right)^2\right]}{\sqrt{2\pi}\,\Phi\left(\frac{m_L - \langle \mathbf{m} \rangle_n}{\sigma_M}\right)} \tag{3.5}$$

which will, hopefully, converge to a good estimate of m_0.

Unfortunately, σ_M is also biased by magnitude selection. It is possible to incorporate a second iterative component to allow for determination of this value as well, but this is considerably more complicated. An alternative is to use the *maximum* value from all of the clusters. This will probably be the closest to the true value since selection will always decrease the estimates and is likely to be a good approximation, particularly if some relatively nearby clusters are included. This is the procedure I have chosen to use.

Creating the Composite Cluster

We now have, in effect, four ways of combining a set of clusters together to form a single composite. We can match all the $\langle \mathbf{m} \rangle$'s either with or without applying the iterative correction above, and we can also choose to match all the $\langle \mathbf{P} \rangle$'s if we wish. These four options are shown schematically in fig. 3.2. The obvious choice is to leave the distance independent $\mathbf{P}$ unchanged and match the corrected $\langle \mathbf{m} \rangle$'s, and indeed for DTF-style calibration, this is the only sensible choice. However, for an ITF-like $\mathbf{P}$ on $\mathbf{M}$ regression, it can be seen from the diagrams that matching both uncorrected $\langle \mathbf{m} \rangle$ and $\langle \mathbf{P} \rangle$ should also produce an unbiased regression. Indeed, this might give some improvement. For the fully corrected situation (bottom left), there will be a tendency for the majority of the data to be cramped into the bottom of the distribution. This should not affect the $\mathbf{M}$ on $\mathbf{P}$ regression unduly, but the $\mathbf{P}$ on $\mathbf{M}$ slope, although not subject to any systematic errors, will be badly constrained
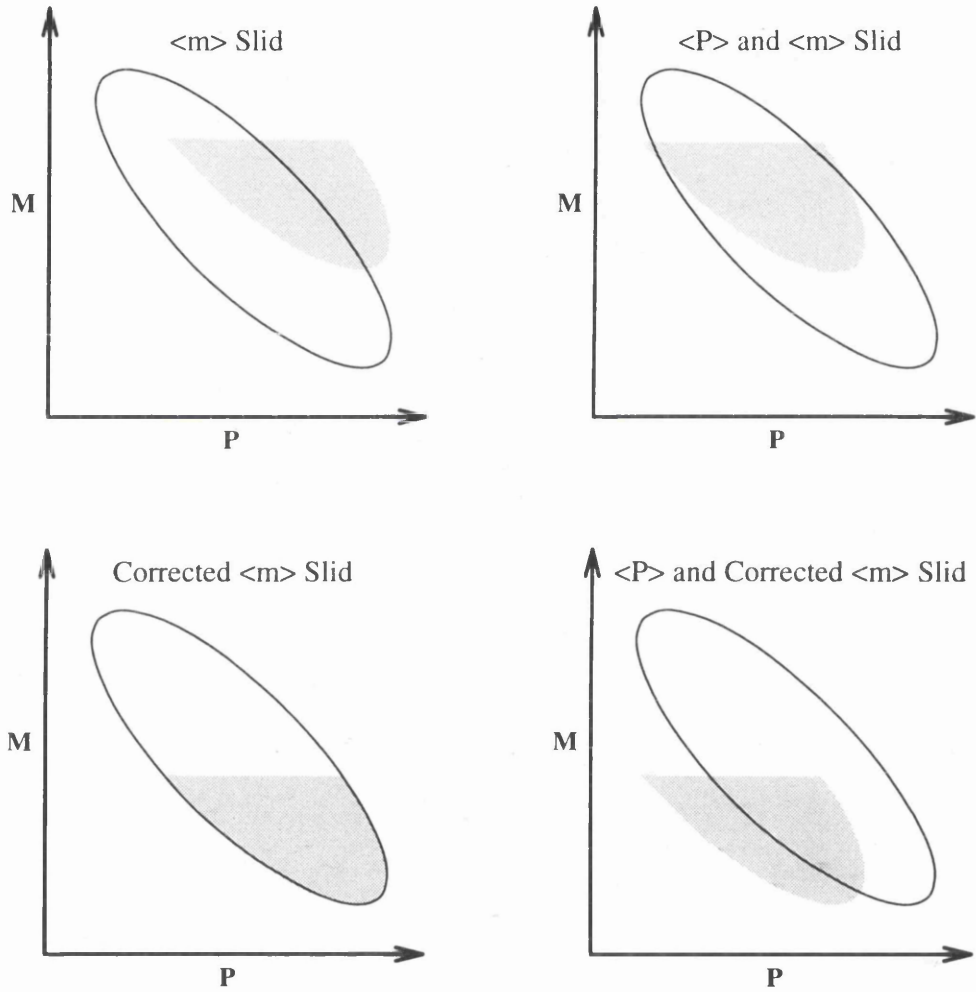
Figure 3.2: A schematic representation of the various ways in which an individual cluster can be added to the composite. The ellipse represents the actual parameter distribution and the shaded area, the contribution from some cluster subject to magnitude selection. On the left hand side, both diagrams have just $\langle m \rangle$ matched to the mean whereas on the right both $\langle m \rangle$ and $\langle P \rangle$ are slid. The difference between the upper and lower diagrams is that in the latter, $\langle m \rangle$ is corrected for magnitude selection.

and perhaps quite noisy. It is, therefore, worth considering the ostensibly inaccurate but potentially useful uncorrected $\langle m \rangle$ and $\langle P \rangle$ match for the ITF regression.

3.1.2 Testing the Composite Cluster Calibration Procedure

There are a number of motivations behind the testing of this calibration technique. Obviously, we need to ensure that the method works, but there are a number of other reasons. First of all, how large an improvement can be expected in comparison to calibration from an individual cluster and what kind of balance should be struck between the number of clusters used and the size each cluster? Also, it is interesting to look at the trend as magnitude selection is decreased and, finally, we wish to determine which composition procedure is most effective for ITF-like estimates.

With these aims in mind, I will employ a Monte Carlo method. Using an actual catalogue (I chose the Mathewson data set) I obtain estimates for the distribution parameters. Then, assuming that all galaxies in a cluster are indeed at exactly the same distance and, for simplicity, using the average redshift to each cluster as that distance, I create a series of mock clusters like the actual ones but with **M** and **P** drawn from a known distribution with a chosen level of magnitude selection imposed as normal. The regression information for a single cluster is then calculated and the various techniques used to create and calibrate composite clusters. The procedure is then repeated a number of times and the results averaged. Note that the slopes and zero points should be calculated for each individual realisation and then averaged, not derived from the averaged distribution parameters.

Figure 3.3 shows results for calibration from a single cluster ('Fornax'). Although there is no apparent bias in either the slope or the zero point of the determinations, the random errors are larger than the distinction between DTF and ITF. In the next chapter, I will show how sensitive to that distinction methods (such as POTENT) for reconstructing the peculiar velocity and density fields are. So, these errors are clearly far too large for the distance estimates to be profitably used in such methods. By contrast, fig. 3.4 shows much smaller deviations, but a very large bias in the determination of DTF characteristics. This is not surprising since this graph is for clusters created by matching uncorrected $\langle m \rangle$ – top left in fig. 3.2. (For this and all the following figures I have taken the situation where all clusters
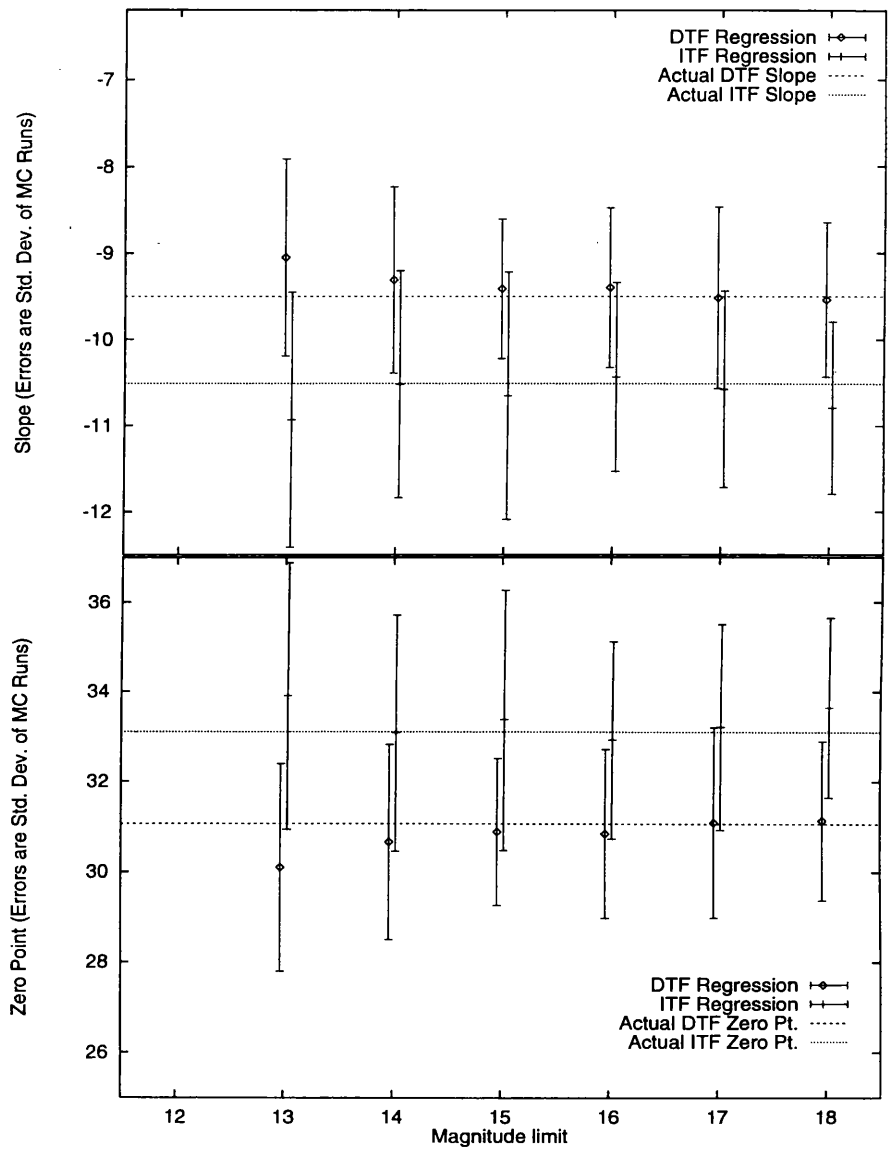
Figure 3.3: The results of the Monte Carlo tests of the calibration of both DTF and ITF style distance estimators using a single cluster. The chosen cluster was the 'Fornax' equivalent for each mock catalogue. The upper graph shows the average slope and the lower graph the average zero point for a variety of different magnitude selections. The actual values are shown as horizontal lines. The error bars are one standard deviation from the Monte Carlos.
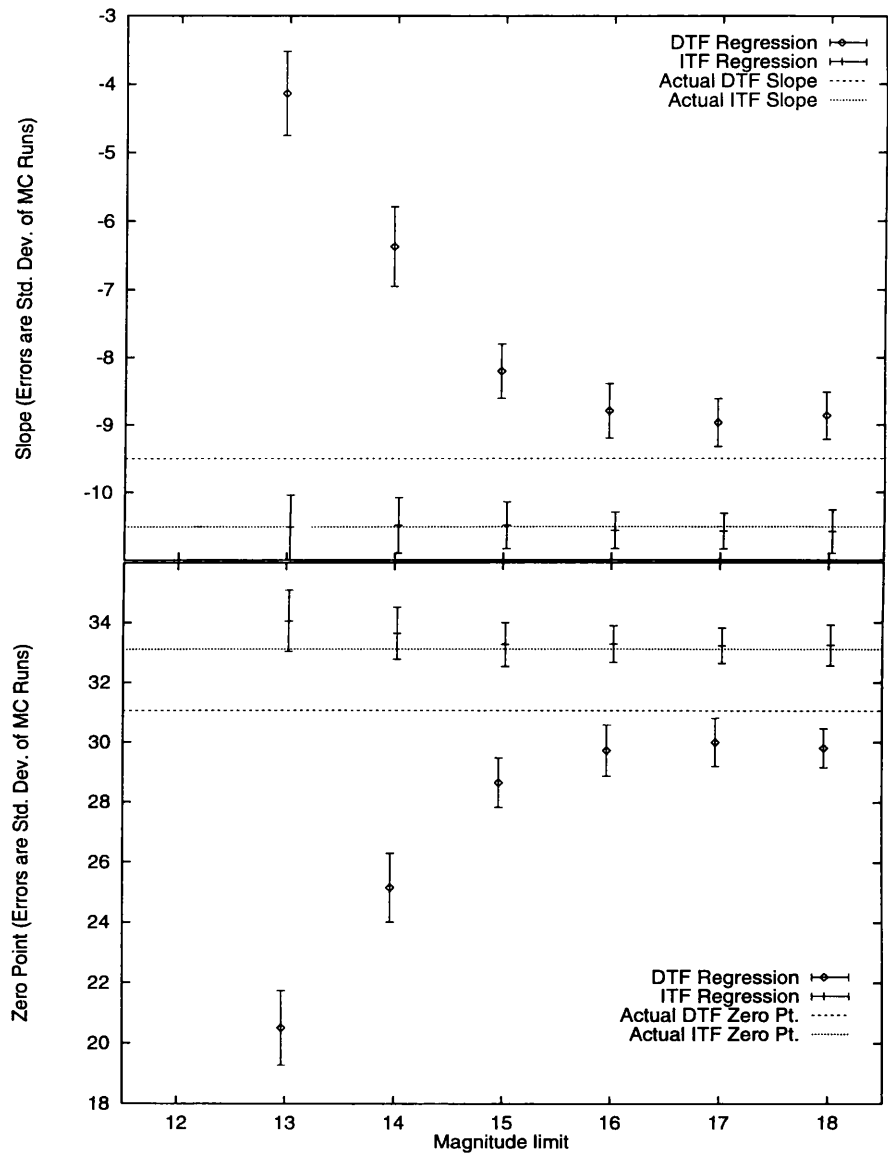
Figure 3.4: Test results for composite clusters created by matching $\langle \mathbf{m} \rangle$ without correction for magnitude selection.

with more than 10 galaxies in the Mathewson catalogue are included in the calibration). The ITF calibration is much better with no apparent problems with the slope at all although there is evidence of a slight bias in the zero points for more extreme magnitude limits. Again, this is exactly as expected. Since uncorrected $\langle m \rangle$'s are used, the DTF calibration is severely affected by magnitude selection. However, the ITF slope is unaffected by this (sect. 2.2.1) and the only systematic effect is on the zero point due to the displacement of $\langle P \rangle$ relative to its true value (fig. 3.2).

Moving on to fig. 3.5 we have a very pleasing result. This is for the upper right case in fig. 3.2 – $\langle m \rangle$ and $\langle P \rangle$ matching. As hoped, the ITF regression is unbiased even up to extreme magnitude limits and the noise is dramatically better than the single cluster case.

One would expect the determination from matching $\langle P \rangle$ and a corrected $\langle m \rangle$ to be biased for both cases, and indeed it is (although this uninteresting case is not included here). However, we have yet to find a suitable calibration for DTF estimators and we also need to check that matching by corrected $\langle m \rangle$ is indeed noisy for ITF as expected. This is demonstrated in fig. 3.6. The random errors are even more unpleasant than expected, but there is another, more severe problem. In spite of the correction for magnitude selection, the DTF determination still shows a definite, if relatively small, bias. This is mainly because the correction is imperfect due to errors in $\hat{\sigma}_M$ and the small sample size of each individual cluster. However, even for extreme magnitude limits of about 13 (remember that the actual Mathewson selection limit is nearer 15) the bias is comparable to or less than the random errors in the single cluster determination. Nevertheless, it is a significant problem as can been seen in fig. 3.7. The single cluster calibration is contrasted with the composite. Although the single cluster determinations are biased, this is to be expected from DTF estimates with magnitude selection. The nature of this bias is such that it makes DTF estimates suitable for Malmquist corrections when the intrinsic distribution of galaxies is used to define the correction. However, the "improved" calibration estimates are biased *differently*, thereby losing the properties needed of a useful DTF estimator.

For comparison, fig. 3.8 shows the equivalent situation for ITF estimates, but using the $\langle m \rangle$ and $\langle P \rangle$ composite cluster calibration. Here, both determinations are unbiased, but improved estimates have marginally smaller errors. Even more importantly, although an improvement of even a few percent is highly significant, the formal errors on the distance
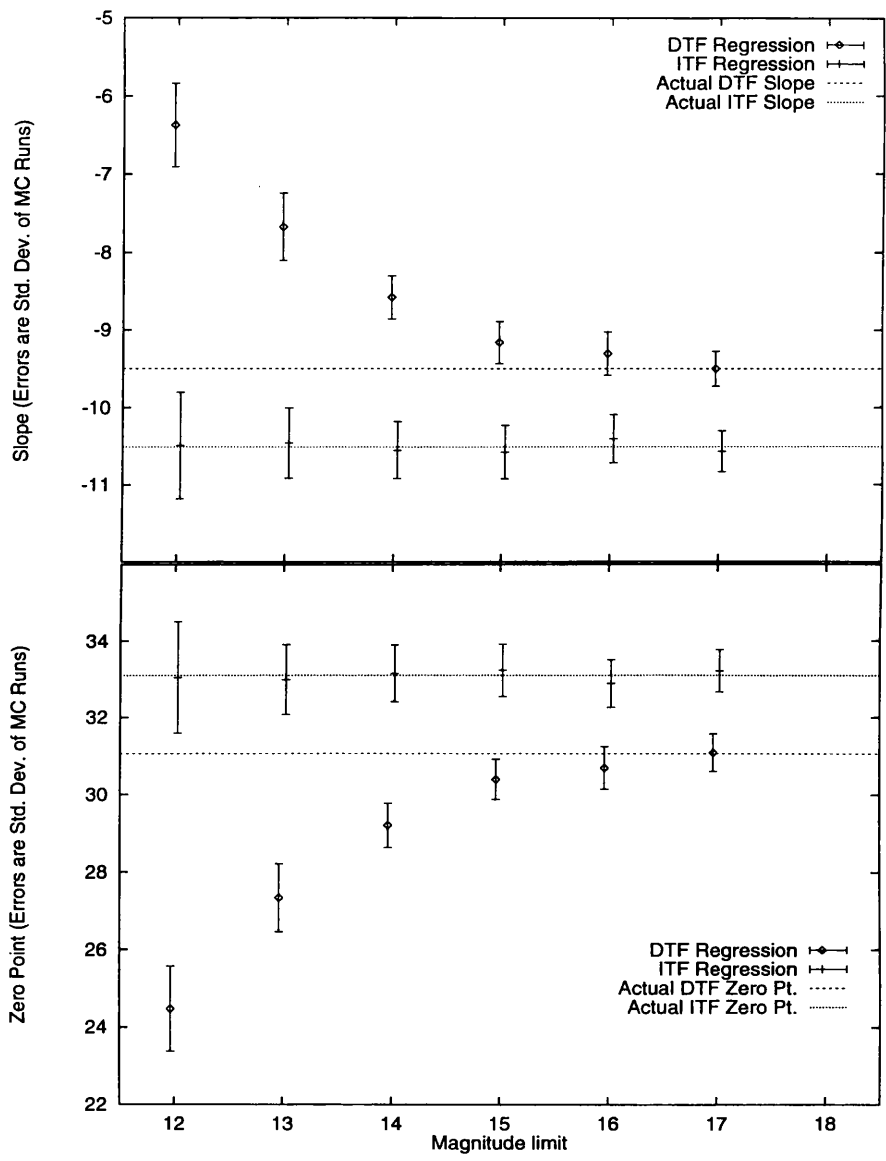
Figure 3.5: Test results for composite clusters created by matching $\langle \mathbf{m} \rangle$ without correction and $\langle \mathbf{P} \rangle$.
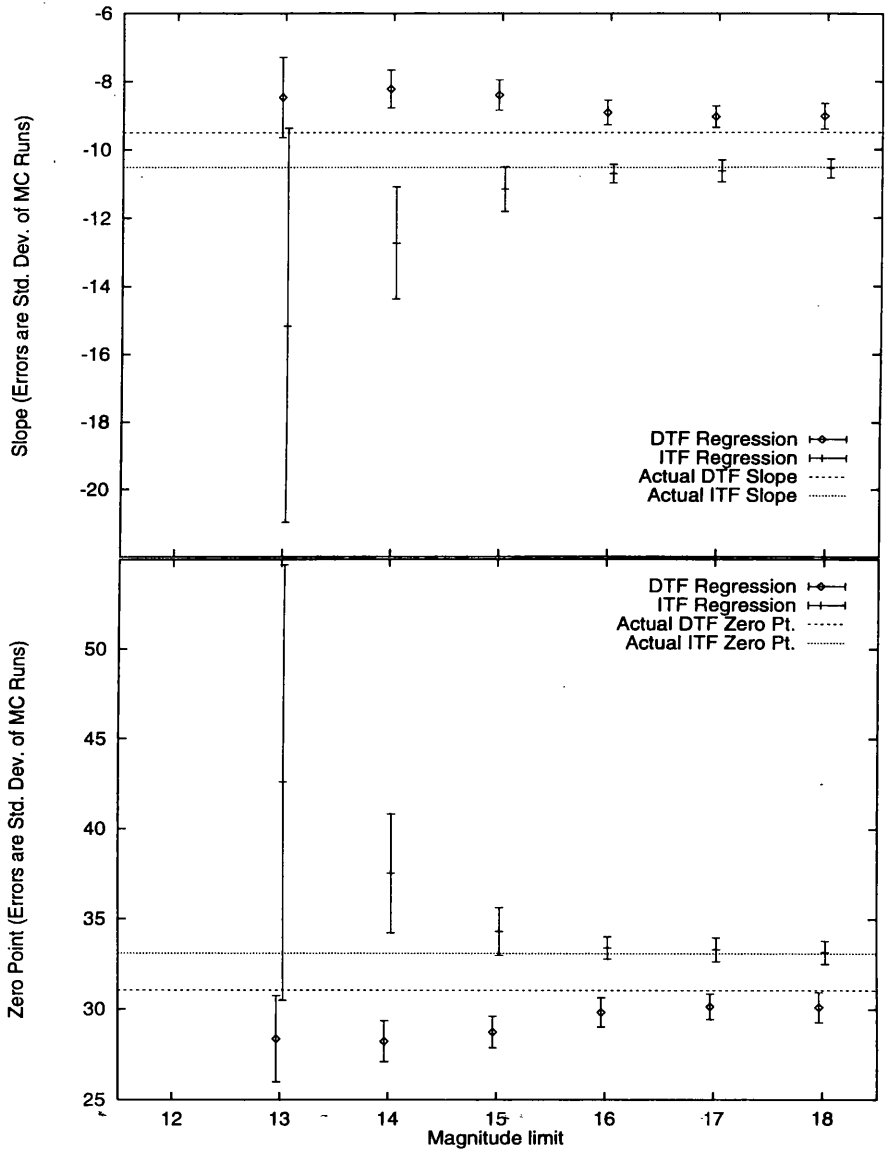
Figure 3.6: Test results for composite clusters created by matching ⟨m⟩ corrected for magnitude selection.
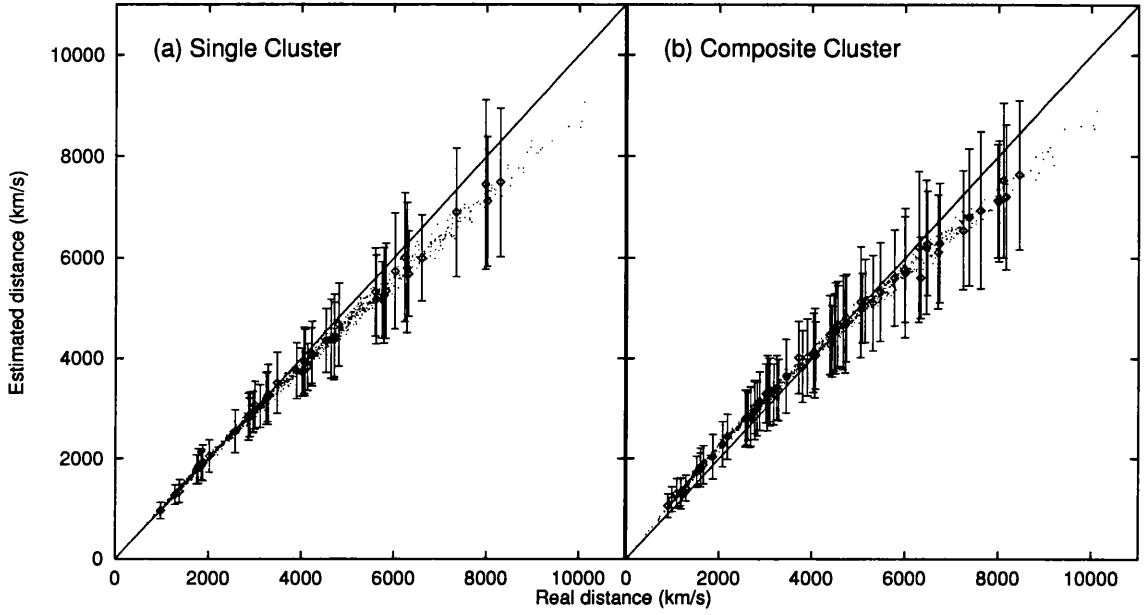
Figure 3.7: Average estimated distances are compared to true distances for DTF estimation in the Monte Carlo tests with a magnitude selection limit of $m_L = 14$. On the left are the results for calibration using a single cluster and on the right for a composite cluster created by matching selection corrected $\langle m \rangle$. One in twenty randomly selected galaxies have one std. dev. error bars and the $y = x$ line is drawn in to 'guide the eye'.

estimate will be more accurate since the larger cluster will be less prone to underestimates of $\hat{\sigma}_M$, $\hat{\sigma}_P$ and $\hat{\rho}$.

In conclusion, it can be confidently stated that the composite cluster calibration method shows significant improvement when used appropriately for ITF-like distance estimation techniques. However, DTF estimates are far more sensitive to magnitude selection and the improvement, if any, is less significant. However, there are possibilities for improvement. A large part of the problem arises from the sparseness of each individual cluster making the determination of the selection correction very noisy. Therefore, by limiting the procedure to only the very richest clusters, some of the bias would be lost – traded for a slightly noisier determination. This kind of fine tuning will depend on the details of the application, and will obviously vary from catalogue to catalogue, but tests such as the one given here could help to determine the 'best' compromise. However, for the purpose of this thesis, I do not intend to use DTF estimators to any great extent since, as I shall show, the best determinations are to be found using methods that are based around ITF estimators. Therefore, I will not
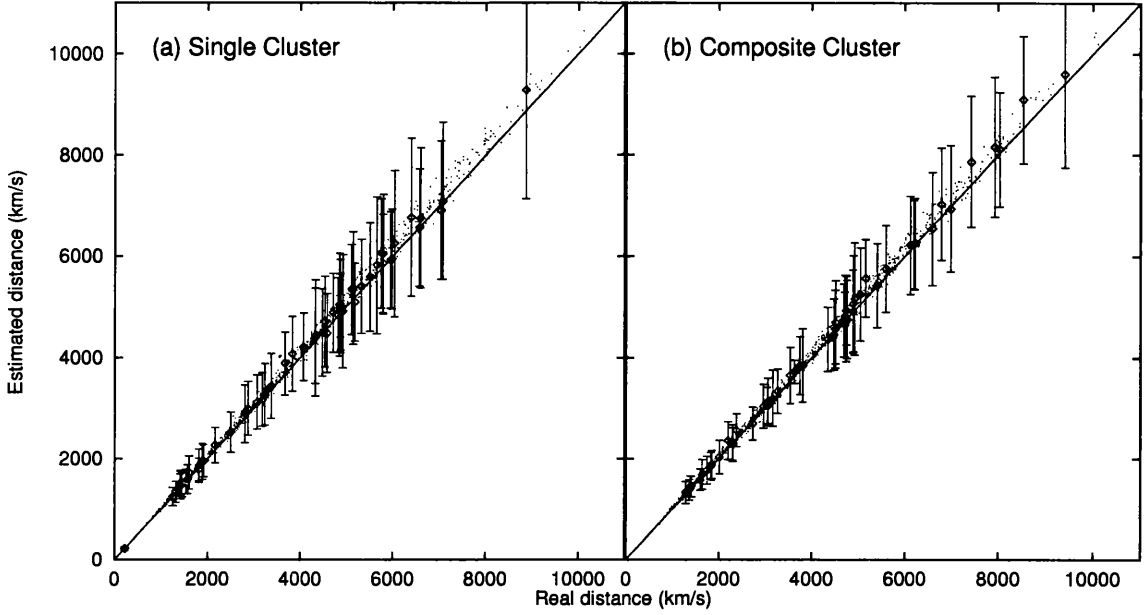
Figure 3.8: As fig. 3.7, but for ITF estimates. Here, the composite cluster is produced by matching $\langle \mathbf{m} \rangle$ and $\langle \mathbf{P} \rangle$.

pursue this further at this stage. It is also very important to remember that the chosen method for combining cluster is, strictly speaking, wrong. It is purely a convenient effect of the particular distribution of $\mathbf{M}$ and $\mathbf{P}$ chosen and has little justification beyond the success of the numerical tests. Any other distribution would have to be similarly tested and may not be so successful.

Another point that should really be taken into consideration is the *radial* extent of clusters. All of the work in this chapter has assumed that all the galaxies in each cluster are at the same distance, but given the angular extent of some clusters on the sky, this is perhaps an unreasonable assumption. This problem has been approached by Willick et al. (1994) and will be tackled in some detail in a forthcoming paper (Rauzy et al., 1994a), but for the purposes of this thesis, I shall assume any effect is small, at least for reasonably distant clusters.

3.2 Absolute Distances

Given the result of the calibration, we can estimate relative distances between galaxies, but for peculiar velocity applications, absolute distances are essential. All that is really needed for this is a reliable distance to any galaxy or cluster, preferably one close enough so that the error in its relative distance is small. However, such a measure is difficult to come by. One frequently used method is to correct the redshift of a cluster for some peculiar velocity model (eg Mathewson et al., 1992), but even aside from the limitations of a model based approach, any model will be for *averaged* or smoothed flows and it is unreliable to apply it to an individual tracer. Also, if that one cluster should have a particularly poor relative distance determination, the error will feed through to affect the entire sample.

However, if a particular catalogue has sufficient sky coverage, there is an alternative. If we assume that, over the entire survey volume, peculiar velocities can be considered as forming a random distribution with zero mean about the Hubble expansion, we can set the absolute calibration to be that which minimises the residuals about the observed redshifts. In other words, fit a line through the distribution of $\hat{r}$ and z and adjust the absolute scaling until the slope of the line is 1. Of course, if the survey covers only a small part of the sky, it is unreasonable to suppose there will be no overall systematic deviations from Hubble flow, but averaged over a significant proportion of the sky and to a reasonable depth, the assumption is likely to be valid.

One advantage of the method is that we are free to choose the weighting of the fit in such a way as to take into account everything that we know about the distance estimates. For example, as in the weighting of points in the POTENT smoothing procedure (sec. 4.1.3, eqn. 4.14) each galaxy or cluster can be weighted its estimated error with the addition of some 'field variance' to prevent over-weighting of nearby points. Indeed, this is the weighting procedure I adopt to give maximum consistency between the calibration procedure and POTENT.

Of course, given that DTF estimates should in fact be biased, this technique is perhaps not entirely suitable. However, the same is true for the standard calibration technique since fixing to the true distance of a galaxy or cluster is not really what is needed. However, the weighting procedure gives the most emphasis to relatively local galaxies and large clusters

where the bias should be negligible anyway and certainly any errors introduced by inaccurate absolute calibration because of failing to take any DTF bias into account will be small compared to the intrinsic errors of the distance estimates and, of course, the DTF bias itself. Again, if one wishes to use DTF estimators for any serious work, these points must be considered, but I shall not pursue them further in this thesis.

By using this technique, I avoid the problems associated with velocity models and also do not give excessive emphasis to any one relative distance determination. In addition, since the line-fitting has two parameters (slope and zero point) but the absolute calibration is achieved by adjusting the slope, the zero point can be used as a crude check on the method. Since we would, obviously, expect the relation to pass very close to the origin, a significantly non-zero intercept is a clear indicator that something is not working – perhaps because of a particularly bad calibration or perhaps because of significant systematic deviations from Hubble flow even when averaged over the entire survey. In this sort of situation, the procedure can be carefully re-examined and cause of the problem found, or if necessary, the catalogue discarded. This may seem rather harsh, but I will show how prone many methods are to systematic errors even with idealised distance estimates – the addition of highly unreliable distances would probably be more than they could take!

It is clear, therefore, that distance estimation of galaxies is a delicate problem with the large uncertainties in the estimates making their analysis and use tricky and fraught with unseen perils. However, with care these problems can be overcome, or at least their impact kept to a minimum, and we can use the estimates to considerably enhance our knowledge of the peculiar velocity and density fields. We have already seen how the model fitting of Lynden-Bell et al. (1988) opened the door to mapping the field, but more recently another, model independent method has been developed that makes full use of distance estimation. This is the POTENT method and it forms the basis of much of my work for this thesis.

Chapter 4

The Potent Method

The potent poison quite o'er-crows my spirit

Hamlet V:ii

The POTENT method for peculiar velocity field reconstruction was devised by Bertschinger and Dekel (1989) and has subsequently been developed and applied to galaxy catalogues (Bertschinger et al., 1990, Dekel et al., 1990, Nusser et al., 1991, Dekel et al., 1993). However, the basic method has remained unchanged, with only details of the implementation adjusted.

4.1   Orthodox Potent

As mentioned in chapter 2, the method makes use of extra information provided by redshift independent distance estimates to galaxies to get an estimate of the radial component of the peculiar velocity of each galaxy:

$$\hat{\mathbf{v}}_{\mathrm{r}} = cz - H_0 \hat{\mathbf{r}} \tag{4.1}$$

(From here onwards, I will take redshifts, distances and velocities all in units of $\mathrm{km\,s^{-1}}$, giving $c = H_0 = 1$). However, unlike most previous methods, these estimates are not used to fit some parametric model velocity field, but a 'model-independent' recovery is found. In fact, there is a model inherent in the POTENT procedure, but it is much less restrictive than any reasonable parametric velocity field. The basis of this flexible model is

the assumption that, above some unknown but fairly small scale, the velocity field can be accurately represented by a potential $\Phi_V(\vec{r})$, ie

$$\vec{v}(\vec{r}) = -\nabla\Phi_V(\vec{r}) \tag{4.2}$$

What is the basis for this assumption? We have seen in sect. 1.4.1 that in an expanding FRW universe, the growing mode is irrotational in linear theory (see eqn. (1.24)). Therefore, any rotational components will have been lost by the present epoch and, from the Kelvin circulation theorem we know that an initially irrotational field will remain that way until orbit crossing occurs. However, if the field is averaged or smoothed on some suitably large scale, we need only consider linear or quasi-linear evolution and can assume that orbit crossing has not occurred. The value of this scale is not known, but is likely to be less than ~ 1000 km s^{-1}, so smoothing on scales greater than this should be sufficient.

This potential Φ_V can, therefore, be calculated at any point $\vec{r}$ by a path integral from the origin to $\vec{r}$

$$\Phi_V(\vec{r}) = -\int_0^{\vec{r}} \vec{v}(\vec{s}) \cdot d\vec{s} \tag{4.3}$$

Of course, the choice of path is arbitrary and by selecting a *radial* path, we obtain the following:

$$\Phi_V(\vec{r}) = -\int_0^r v_r(r', \theta, \phi) dr' \tag{4.4}$$

which involves only the radial component of the peculiar velocity. Using this equation, the potential can be derived at any point and the full, three-dimensional peculiar velocity field recovered by (4.2). However, there is an obvious complication. The integration in eqn. (4.4) requires v_r at *all points* along the radial path. However, the estimates $\hat{v}_r$ are only known at the positions of galaxies — some form of smoothing is clearly needed.

4.1.1 The Smoothing Procedure

The purpose of the smoothing in POTENT is twofold. Of course, it is necessary in order to perform the integration, but some level of smoothing is also required to maintain the potential flow condition over regions where mixing has occurred. In practice, it is the former motivation which dominates as catalogues typically have separations between nearest neighbour galaxies of up to 2000 km s^{-1}. In addition, the errors on the estimates of the

peculiar velocities are very large (frequently greater than 100% for galaxies more distant than ~ 3000 km s^{-1}) and in order to obtain a reasonable signal–to–noise, we really require several galaxies within each smoothing volume. A typical smoothing radius, therefore, is > 1000 km s^{-1}, sufficient to ensure an irrotational field.

The smoothing procedure itself, is quite a complicated one. The complication arises from the fact that the averaging is over radial components at different angular positions. The solution fixed upon by Dekel et al. (1990) is to use a *Tensor window function* that uses a maximum likelihood technique to fit to a bulk flow solution within each window.

From Dekel et al. (1990), equation 9 gives

$$L(\vec{r}) \;=\; -\frac{1}{2}\sum_{i}^{N_{gal}} W(\vec{r},\hat{\vec{r}}_i)\Big[\hat{\mathbf{v}}_{\mathbf{r}i} - \hat{\vec{r}}_i \cdot \vec{v}(\vec{r})\Big]^2 \tag{4.5}$$

where:

$L(\vec{r})$ is the 'Likelihood function',

$W(\vec{r},\hat{\vec{r}}_i)$ is a window function,

$\vec{r}$ is some position vector,

$\hat{\vec{r}}_i$ is the estimated position vector of galaxy i out of N_{gal},

$\hat{\mathbf{v}}_{\mathbf{r}i}$ is the estimated radial peculiar velocity of galaxy i.

Writing $W_i = W(\vec{r},\hat{\vec{r}}_i)$ and $\vec{v}(\vec{r}) = \sum_{\alpha} v^{\alpha}\vec{e}_{\alpha}$ where the $\vec{e}_{\alpha}$ form a coordinate basis,

$$L(\vec{r}) \;=\; -\frac{1}{2}\sum_{i,\alpha} W_i\Big[\hat{\mathbf{v}}_{\mathbf{r}i} - \hat{\vec{r}}_i \cdot v^{\alpha}\vec{e}_{\alpha}\Big]^2 \tag{4.6}$$

$$\;=\; -\frac{1}{2}\sum_{i,\alpha} W_i\Big[\hat{\mathbf{v}}_{\mathbf{r}i} - \hat{\vec{r}}_{i\alpha}v^{\alpha}\Big]^2 \tag{4.7}$$

where $\hat{\vec{r}}_{i\alpha}$ is the α component of $\hat{\vec{r}}_i$ at $\vec{r}$. We will now assume summation over repeated greek indices.

Maximising $L(\vec{r})$ with respect to v^{α} gives

$$\frac{\partial L}{\partial v^{\beta}} \;=\; \sum_{i} W_i\Big[\hat{\mathbf{v}}_{\mathbf{r}i} - \hat{\vec{r}}_{i\alpha}v^{\alpha}\Big]\hat{\vec{r}}_{i\gamma}\frac{\partial v^{\gamma}}{\partial v^{\beta}} \tag{4.8}$$

$$\;=\; \sum_{i} W_i\Big[\hat{\mathbf{v}}_{\mathbf{r}i} - \hat{\vec{r}}_{i\alpha}v^{\alpha}\Big]\hat{\vec{r}}_{i\gamma}\delta_{\gamma\beta} \tag{4.9}$$

$$\;=\; \sum_{i} W_i\Big[\hat{\mathbf{v}}_{\mathbf{r}i} - \hat{\vec{r}}_{i\alpha}v^{\alpha}\Big]\hat{\vec{r}}_{i\beta} \;=\; 0 \tag{4.10}$$

therefore

$$\sum_i W_i \hat{\mathbf{v}}_{\mathrm{r}i} \hat{\tilde{\mathbf{r}}}_{i\beta} = \sum_i W_i \hat{\tilde{\mathbf{r}}}_{i\alpha} v^\alpha \hat{\tilde{\mathbf{r}}}_{i\beta} \tag{4.11}$$

putting $\sum_i W_i \hat{\mathbf{v}}_{\mathrm{r}i} \hat{\tilde{\mathbf{r}}}_{i\beta} = d_\beta$ and $\sum_i W_i \hat{\tilde{\mathbf{r}}}_{i\alpha} \hat{\tilde{\mathbf{r}}}_{i\beta} = E_{\alpha\beta}$, we have

$$v^\alpha = d_\beta (E^{-1})_{\alpha\beta} \tag{4.12}$$

Therefore, we have v^α, which is the maximum likelihood bulk flow velocity within the window $W(\vec{r}, \hat{\tilde{\mathbf{r}}}_i)$. It should be noted that this is, in fact, all three components of the peculiar velocity at this point. However, in the orthodox POTENT method, only the radial component of v^α is used. We will introduce an alternative that makes use of all three components in sect. 5.3.

The next stage in the procedure is to choose a form for the window function W_i. However, before this can be done, we need some knowledge of the errors in the POTENT method.

4.1.2 Errors in Potent

There are two main sources of error in POTENT. Firstly, as we have already seen, the estimates of the distances to galaxies and, by eqn. (4.1), $\hat{\tilde{\mathbf{v}}}_{\mathrm{r}}$ are subject to very large errors. In addition, the galaxies in the catalogue sample space in a sparse and non-uniform way, both because of intrinsic fluctuations in the number density of galaxies in the Universe (eg. the Great Wall) and selection effects in the survey (eg. the Zone of Avoidance). Because of the nature of the smoothing procedure these two problems combine together to produce very complicated errors, both systematic and random. The main goal of the implementation of the POTENT method, and of the other techniques in this thesis, is to minimise these errors.

The Treatment of Distance Estimators in Potent

Whether or not a distance estimator is biased is not the crucial question when attempting to correct for bias in POTENT. What is important is to construct an unbiased smoothed peculiar velocity field. POTENT attempts to construct an unbiased peculiar velocity field in the following sense:

One assumes an underlying smoothed peculiar velocity field (taken to be potential) and effective density distribution that is determined by some selection function and underlying

density distribution of galaxies.

1. This field is sampled at n points and the galaxies taken to be at these points corresponding to the actual distances $(r_{10}, r_{20}, r_{30}, ..., r_{n0})$

2. Errors are added to these distances. A smoothed initial radial peculiar velocity field is derived using the tensor window function.

3. Hence, one obtains a potential velocity field by radial integration.

If this smoothed recovered potential velocity field is the same as the input potential velocity field when it is averaged over all realisations of $(r_{10}, r_{20}, r_{30}..r_{n0})$ and of the distance errors, it is unbiased.

In the Appendix of Dekel et al. (1990) an attempt is made to prove that if one applies a homogeneous Malmquist correction to the raw distance estimates then one does obtain a peculiar velocity field which is almost unbiased. Essentially their analysis proceeds by expressing the bias of the recovered velocity field as a function of the errors, ϵ_i, in the galaxy distance estimates, and depends upon making several Taylor expansions in ϵ_i and discarding terms of order 3 and above. If the distance errors are large, as they will be at large true distances, this procedure will break down. Dekel et al. employ Monte Carlo simulations to back up their analytic treatment.

It is instructive to examine why the application of Malmquist corrections appears to work for the POTENT analysis. In this respect, the important factor is the window function. In interpolating a peculiar velocity from galaxies appearing in the catalogue to a given spatial point with radial coordinate s, one is effectively binning by estimated distance. As was seen in sect. 2.2.3, this is exactly the situation where Malmquist corrections can be expected to be useful. Of course this will strictly only be valid if galaxies are not too sparse, and if the gradient of velocity field is not too large, or the effective radius of the window function is not too wide.

Assumptions for Malmquist Corrections

Both the homogeneous and inhomogeneous Malmquist corrections are computed by applying Bayes' theorem to obtain $\hat{\omega}$ and both derivations assume that the probability density

function, $p(\hat{\omega}|\omega_0)$, is a normal distribution, and that $\hat{\omega}$ is unbiased. Lynden-Bell et al. (1988) assume the prior distribution, $p(\omega_0)$, of true log distance to correspond to a homogeneous distribution of galaxies. Landy and Szalay (1992), on the other hand, estimate $p(\omega_0)$ by constructing a spline fit to the histogram of log distance *estimates* for the galaxies in the survey, thus in principle taking into account inhomogeneities in the galaxy distribution. Due to the sparseness of surveys, however, it is usually necessary to average the distribution of galaxies over large solid angles, if not all, of the sky. Therefore, the effects of clustering may still go largely unaccounted for (c.f. Newsam et al. (1994) and Newsam et al. (1992)).

A more serious problem with the IMC, however, stems from the use of $\hat{\omega}_{\mathrm{DTF}}$. In Hendry and Simmons, 1994 it is shown that this will result in an *incorrect* Malmquist Correction due to the bias of $\hat{\omega}_{\mathrm{DTF}}$. In general, if $p(\omega_0)$ is constructed from the observed distribution of log distance estimates then one must apply the IMC to the ITF estimator. See also the discussion in Teerikorpi (1993) and Feast (1994), where the same conclusion is reached. Malmquist corrections derived from $\hat{\omega}_{\mathrm{DTF}}$ will be valid only when $p(\omega_0)$ is equal to the *intrinsic* distribution of true log distance – an approximation to which one might obtain from, for example, a deeper redshift survey (c.f. Hudson, 1993; Dekel, 1994). As a special case of this result, note that the homogeneous Malmquist correction (HMC) applied to $\hat{\omega}_{\mathrm{DTF}}$ will be valid provided that the intrinsic distribution of galaxies is homogeneous. This will frequently be a reasonable assumption but is difficult to test. What is certainly clear, however, is that applying the Inhomogeneous Malmquist Correction (IMC) as derived by Landy and Szalay to $\hat{\omega}_{\mathrm{DTF}}$ will be completely inappropriate, since the prior distribution obtained from a histogram of raw log distance estimates will *not* correspond to the intrinsic, but rather to the observed, true log distance distribution.

Sampling Gradient Biases

Malmquist bias is not, unfortunately, the last of the bias problems in POTENT. Another significant error comes from *Sampling Gradient Biases*. These are caused by information from areas with a high density of observed galaxies 'leaching' into areas with a lower density. This is a problem because the assumption of a potential for the velocity field applies to the underlying velocity field smoothed by a spherical window weighted by the *mass*. However, in the POTENT procedure, the smoothed field is weighted by the number density of objects

$n_i \equiv n(\vec{r}_i)$ and the selection function $S_i \equiv S(\vec{r}_i)$. This can be partly corrected for by eliminating the nS weighting entirely to give a *volume weighted* solution, but full correction would depend on knowledge of the underlying mass field which is the final aim of the entire procedure.[1]

4.1.3 Choice of Potent Window Function

The aim of the selection of a window function is to minimise the bias and variance of the recovered radial peculiar velocity field whilst keeping the *amount* of smoothing as low as possible. In practice, there is something of a conflict between these goals and the final choice must be a compromise. Dekel et al. (1990) propose three different forms of window, and it is these that I have been using. The simplest form is the 'biased' window W_B:

$$W_{\mathrm{B}}(\vec{r}, \hat{\vec{r}}_i) = \frac{1}{\sigma_i^2} \exp\left[-\frac{|\vec{r} - \hat{\vec{r}}_i|}{2R_{\mathrm{w}}^2}\right] \tag{4.13}$$

where R_{w} is the window radius and σ_i^2 is the weighting for the ith galaxy. It might seem preferable to have constant weights in order to maintain a uniform average and keep the sampling gradient bias low, but in order to minise the variance, we would wish to give less weight to galaxies with larger errors. For this window, variance minimisation is the prime requirement, so for a galaxy with an error $r_i\Delta$, the weighting function has the form:

$$\sigma_i^2 = r_i^2\Delta^2 + \sigma_f^2 \tag{4.14}$$

with σ_f^2 included as a 'field variance' to exclude the effect of intrinsic fluctuations in the velocity field and, in practice, to prevent excessive weight being given to nearby galaxies. A value of $\sigma_f = 150\mathrm{km\,s}^{-1}$ is adopted as recommended by Dekel et al. (1990).

Although simple and quite effective in keeping down the noise in recovered radial velocities, the 'biased' window function is exactly that — it is very prone to sampling gradient biases. These can be minimised by weighting by $(n_iS_i)^{-1}$ if the slight increase in variance is a good trade-off. In practice, $(n_iS_i)^{-1}$ is unknown, but it can be approximated by a suitable 'volume' for each galaxy V_i, for example

$$V_i = R_4^3(\hat{\vec{r}}_i) \tag{4.15}$$

---

[1]This is another example of problems incurred because of a poorly known prior. However, in this case the sensitivity to the prior in not too acute and the errors introduced can be kept under control.

where $R_4(\hat{\vec{r}}_i)$ is the distance from i to the fourth nearest neighbour galaxy. This weighting gives rise to the 'fixed' window W_{F}:

$$W_{\mathrm{F}}(\vec{r}, \hat{\vec{r}}_i) = \frac{V_i}{\sigma_i^2} \exp\left[-\frac{|\vec{r} - \hat{\vec{r}}_i|}{2R_{\mathrm{w}}^2}\right] \tag{4.16}$$

This window is, in fact, the one used in most of the POTENT analyses and will be the one I shall be concentrating on. However, there is one other window proposed by Dekel et al. (1990) that has some interesting properties.

This window, the 'variable' window function W_{V}, unlike the other functions, does not have one fixed window radius but varies in size to give high resolution in areas where the catalogue is well-sampled, but smoothing on much larger scales in regions with few galaxies in order to keep the variance within reasonable limits.

$$W_{\mathrm{V}}(\vec{r}, \hat{\vec{r}}_i) = \frac{V_i}{\sigma_i^2} \exp\left[-\frac{|\vec{r} - \hat{\vec{r}}_i|}{2R_{\mathrm{w}}^2(\vec{r})}\right] \tag{4.17}$$

The choice of $R_{\mathrm{w}}^2(\vec{r})$ is somewhat arbitrary, but from "trial and error" with Monte Carlo simulations, Dekel et al. (1990) chose $R_{\mathrm{w}}(\vec{r}) = R_5(\vec{r})$, the distance from the position $\vec{r}$ to the fifth nearest galaxy in the sample. In this way, the quality of data in well sampled regions can be considerably improved without significant loss in the less abundant areas. However, the variable radius and subsequent increase in the complexity of the error analysis make results obtained this window very difficult to analyse. In particular, comparison with theoretical models or results from another technique becomes very tricky since the smoothing effects must be as similar as possible to enable sensible comparison. In the majority of situations, therefore, the fixed window is the most appropriate.

4.1.4   Density Field recovery

As a result of the smoothing and line integrals, we have a smoothed velocity potential Φ_{V} from which we can derive the full three-dimensional peculiar velocity field using eqn. (4.2). However, what we are really interested in is the distribution of matter. This is particularly true with POTENT since the velocity field is dynamical in origin and makes no assumptions about the way in which light traces mass. Comparison with other studies can, therefore, help to give constraints on the cosmological bias parameter (see sect. 2.1). We would therefore like a method that enables us to calculate a density field from the peculiar velocity data.

Such a method is described in Nusser et al. (1991) and is outlined below. This is the method adopted by Dekel et al. (1993) and I have, therefore, also implemented it.

Consider the evolution of the universe as a pressureless gravitating fluid against the standard (FRW) cosmological background (refer back to chapter 1 for details). The pressure can be neglected since our result from POTENT is for the case where particle orbits have not crossed. Using the peculiar velocity $\vec{v}$, the peculiar *gravitational* potential Φ_G and the density contrast δ in comoving coordinates we obtain the three usual equations of state — the continuity equation,

$$\dot{\delta} + \nabla \cdot \vec{v} + \nabla \cdot (\delta \vec{v}) = 0 \qquad (4.18)$$

the Euler equation,

$$\dot{\vec{v}} + 2H\vec{v} + (\vec{v} \cdot \nabla)\vec{v} = -\nabla \Phi_G \qquad (4.19)$$

and the Poisson field equation,

$$\nabla^2 \Phi_G = \frac{3}{2} H^2 \Omega \delta \qquad (4.20)$$

Note that both H, the Hubble parameter, and Ω, the cosmological density parameter, are time varying. Using (4.20) to remove the potential from the divergence of (4.19) and recalling that, since the velocity field is assumed to be vorticity–free,

$$(\vec{v} \cdot \nabla)\vec{v} \equiv \frac{1}{2}\nabla v^2 \qquad (4.21)$$

we obtain

$$\frac{\partial}{\partial t}(\nabla \cdot \vec{v}) + 2H\nabla \cdot \vec{v} + \frac{1}{2}\nabla^2 v^2 = -\frac{3}{2}H^2 \Omega \delta \qquad (4.22)$$

For simplicity, let us first consider the linear approximation neglecting all terms involving $\delta \vec{v}$ or v^2.

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}H^2 \Omega \delta = 0 \qquad (4.23)$$

We are only interested in the growing mode of this $\delta \propto D_+(t)$ which is usually expressed in terms of the function $f(\Omega)$

$$f(\Omega) \equiv \frac{\dot{D}_+}{HD_+} \qquad (4.24)$$

Equation (4.23) can then be written as

$$\delta = -(Hf)^{-1}\nabla \cdot \vec{v} \qquad (4.25)$$

which depends only on the *local* velocity.

However, when we move to the second order solution, not only does the solution become much more complicated, it also loses its local nature and requires integrals over a large volume (Nusser et al. (1991), Peebles (1980)). This is unfortunate as POTENT is essentially local in nature and limited by the size of the catalogue, but it still requires something beyond the linear regime as fluctuations with $\delta > 1$ are not uncommon even smoothed on $1000+$ km s^{-1} scales. Nusser et al. (1991) therefore proposed the use of the *quasi-linear* Zel'dovich approximation (Zel'dovich, 1970). This approximation is normally considered in *Lagrangian* space, where the coordinates $\vec{q}$ are the initial comoving positions of particles (ie a coordinate is 'labelled' by a particle). The displacement of a particle from its initial position after some time t is then given by $\vec{x}(\vec{q}, t) - \vec{q}$ where $\vec{x}$ is the *Eulerian* coordinate of the particle. The Zel'dovich approximation consists of writing this displacement as a product of a purely spatial perturbation function $\vec{\psi}(\vec{q})$ and a separate, universal time–dependent function $D(t)$.

$$\vec{x}(\vec{q}, t) - \vec{q} = D(t)\vec{\psi}(\vec{q}) \qquad (4.26)$$

Because this is only an approximation, care must be taken over its use. In particular, Nusser et al. (1991) found that when employed in eqn. (4.22), equivalent to conserving momentum, the final result was identical to the linear approximation. However, when used in the continuity equation (4.18) — mass conservation — the following result was obtained

$$\delta(\vec{x}) = \left\| I - (Hf)^{-1}\frac{\partial \vec{v}}{\partial \vec{x}} \right\| - 1 \qquad (4.27)$$

where I is the identity matrix and $\| \ldots \|$ denotes the Jacobian determinant. Tests showed that this approximation was remarkably good (Nusser et al., 1991), and has since been used in POTENT.

4.1.5 Numerical Implementation

So far I have considered POTENT in terms of continuous fields. However, for implementation on a computer, it is necessary to discretise onto grids etc. In this brief section, I will describe the details of the actual implementation of POTENT.

Three forms of grid are used within POTENT. The simplest one is a spherical grid with N_R shells centred on the origin and equally spaced out to some maximum distance. These

shells are sub-divided into N_θ latitude circles and N_ϕ longitude circles. This grid simplifies the calculation of $\Phi_V(\vec{r})$ since all points on the grid lie on one of the $N_\theta N_\phi$ radial lines and peculiar velocities can be calculated along the grid lines to give three orthogonal components at each point. However, there is an excess of points towards the poles of the grid and it is easier to perform analysis on a cubic, cartesian grid. More recently, therefore, this simple grid has been replaced by one with $N_\phi \sin\theta$ longitude circles and the potential $\Phi_V(\vec{r})$ is interpolated onto a cubic grid using Cloud–in–Cell interpolation (Hockney and Eastwood, 1981) and all other calculations are performed on this grid (Dekel et al., 1993).

The integration of eqn. (4.4) is approximated by fitting a Chebyshev polynomial to the smoothed radial velocity field along each radial line and integrating the entire line in one go (Press et al., 1992).[2] The resultant potential field is interpolated onto a spherical grid as described above and the derivatives for the calculation of both $\vec{v}(\vec{r})$ and $\delta(\vec{r})$ are evaluated on cubic splines fit through the grid points (Press et al., 1992).

4.2 Testing Potent

In this section, I will be examining the extent to which the biases outlined above are still a problem after the correction procedures have been applied. To do this I use a Monte Carlo technique that differs from that in the original POTENT analysis (Dekel et al., 1990). There, they chose a two-level process with averaging performed over the distribution of objects *and* the distribution of distance errors. However, I am interested in the effect of a *particular* distribution of objects (for example, a galaxy catalogue) and so the first level of averaging is dropped. As well as simplifying the procedure this also removes the Monte Carlo bias that complicates the analysis in Dekel et al., 1990. Other than this, however, my analysis is kept as similar to theirs as possible, for example by choosing the same window size (1200 km s^{-1}) and using the volume weighted, fixed window function $W_F(\vec{r}, \hat{\vec{r}}_i)$.

I perform two sets of tests. Both involve Monte Carlo realisations of POTENT velocity field recoveries. The first set of results are for an idealised situation. The underlying velocity field is quiet Hubble flow, and galaxies are drawn randomly from an homogeneous universe with

---

[2]This approach was compared to a more rigorous, but much more intensive integration method using Simpsons rule, but no significant difference in the results was found.

complete sky coverage. Since it is impossible to create a numerical sample with truly infinite depth, galaxies are created homogeneously within a sphere. This sphere is centered around the Milky Way and its radius is such that a galaxy whose absolute magnitude is $M_0 + 3\sigma_m$ is just visible where σ_m is the standard deviation of the intrinsic luminosity function. To generate estimated distances to these galaxies, $\mathbf{M}$ and $\mathbf{P}$ are sampled from a bivariate normal distribution and subjected to magnitude selection. If the galaxy is unobservable, it is completely discarded. For each observable galaxy, $\mathbf{M}$ and $\mathbf{P}$ are then used to estimate the distance using DTF and ITF estimators and a variety of Malmquist corrections. We take typical values of the distribution parameters obtained for the D_n-σ and Tully-Fisher relation that give a log distance error of about 15%. Note that by doing this I am assuming that the calibration procedure has provided the *exact* slope and zero point and that all the errors are due to the intrinsic scatter in the relation. This is partly for simplicity, but mainly because I wish to isolate these effects from those produced by calibration errors. A number of POTENT realisations for each method are calculated and the average used to show the biases. This test is designed to show the effect of the various distance estimation and correction techniques in what is, in some sense, a 'best case' situation. The use of quiet Hubble flow will ensure that no sampling gradient biases are introduced and the homogeneous universe is ideal for testing the assumptions of the Malmquist corrections.

The second set of tests are for a more realistic situation. Quiet Hubble flow is replaced with a more complex peculiar velocity field involving a large void and an attractor region. Also, the galaxies are positioned much more realistically.

This realistic distribution is generated as follows. I position a galaxy at each point in the combined data of Mathewson et al. (1992) and Burstein (1990)[3] (see fig. 4.1). For simplicity, I neglect the improvement in the distance estimates that can be found for clusters by combining estimates from all the galaxies in the cluster. However, in order to avoid "finger of God" type effects, one galaxy is chosen at random from each cluster. Again, $\mathbf{M}$ and $\mathbf{P}$ are drawn from the same bivariate normal and the same selection applied. However, if the galaxy is found to be unobservable, $\mathbf{M}$ and $\mathbf{P}$ are regenerated and the process repeated

---

[3]The Burstein Mark II data is a combination of a number of other catalogues with some otherwise unpublished data. The catalogues include elliptical galaxies from Lynden-Bell et al. (1988), Lucey and Carter (1988) and Dressler and Faber (1990) and spirals from Aaronson et al. (1982), (1986), (1989), Bothun et al. (1984) and de Vaucouleurs and Peters (1984). The Mark II catalogue replaces the Mark I catalogue (Burstein et al., 1987) which was found to contain a number of errors.
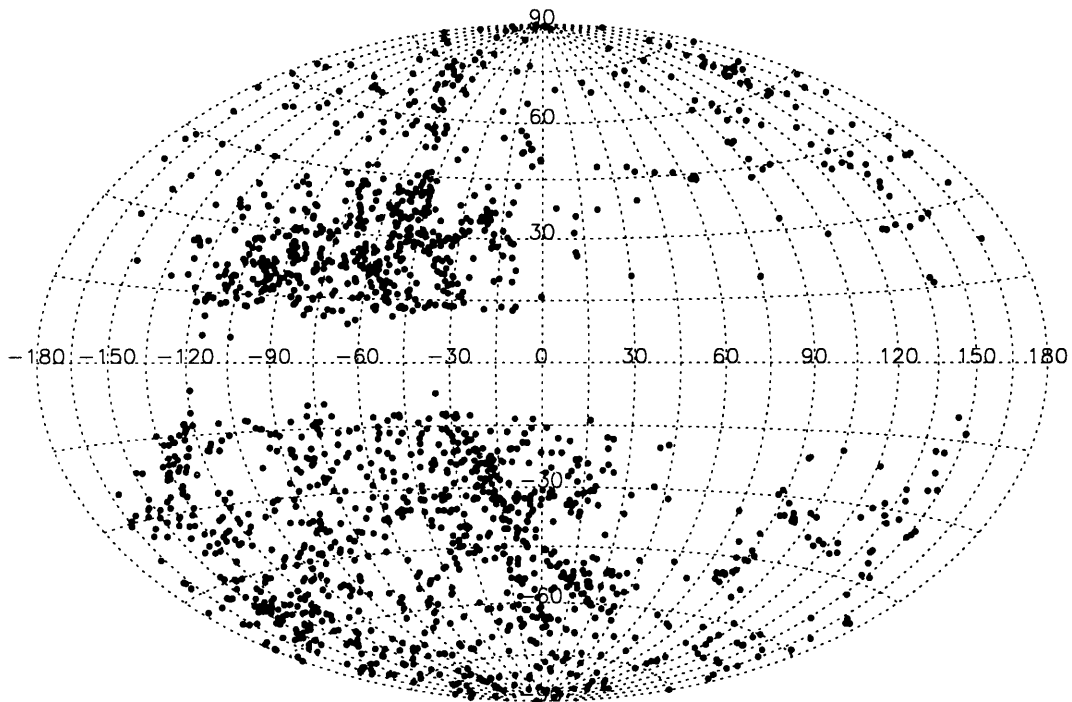
Figure 4.1: An Aitoff projection of the galaxies used as the sample for the more realistic Monte Carlo realisations. Coordinates are galactic longitude and latitude. In particular, note the large void centered around longitude 90° where the Mathewson sample is particularly sparse.

until the galaxy is 'observed'. In this way, given that the distribution parameters ($\hat{\sigma}_M$, $\hat{\rho}$ etc) are reasonable, I can create a series of mock catalogues that have a spatial distribution somewhat like that of a genuine catalogue and with each galaxy sampled from an appropriate region of the luminosity function.

The field I have chosen to use for these tests is formed by creating a potential consisting of two large spherically symmetric gaussian fluctuations, one positive, centered around (4000, 0), and one negative at (-2000, 2000). The velocities are then constructed as in eqn. (4.2) and these velocities used to assign redshifts to the galaxies. When comparisons are made to POTENT results, it is necessary to smooth the field with a gaussian window of 1200 km s^{-1}. This is the radius of the window used in the POTENT realisations (Dekel et al., 1993) and enables easier and more accurate comparison. Figure 4.2 shows a slice through this smoothed field.
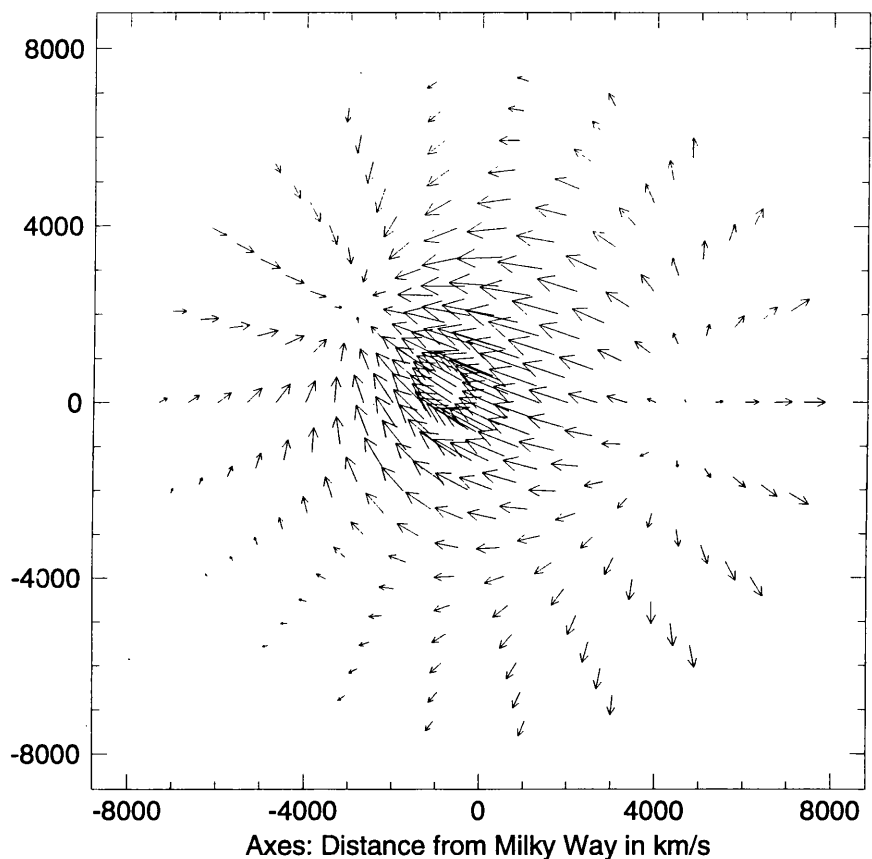
Figure 4.2: A slice through the centre of the velocity field used. The void and attractor regions are modelled by deriving the field from a potential which contains a large gaussian well and a wider gaussian peak. Note that this graph actually shows the field smoothed with a gaussian window of 1200 km s^{-1} radius to enable direct comparison to the POTENT results.
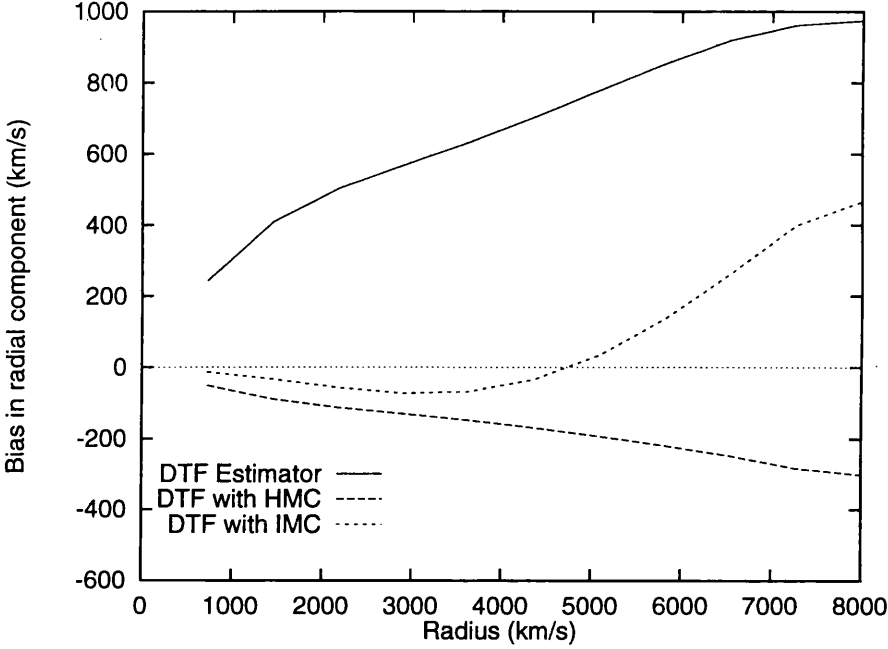
Figure 4.3: Bias as a function of distance for POTENT recoveries of quiet Hubble flow with galaxies drawn from an homogeneous universe. The DTF estimator has been used on its own, with an homogeneous Malmquist correction and with an inhomogeneous correction. The distance estimator has an error in log distance of about 15%.

4.2.1 Monte Carlo Potent velocity field recoveries

The results of the first, simpler test are shown in Figures 4.3 and 4.4. Since I am dealing with such a simplified galaxy sample and velocity field, all the bias is in the radial direction and is the same along all radial lines. These graphs show this bias as a function of distance. The need for some form of correction is clear, particularly for the DTF estimator, as are the dangers of an inappropriate correction – for example, the homogeneous Malmquist correction applied to the ITF estimator which is far worse than using the uncorrected distances. Slightly surprising is the good recovery from the inhomogeneous correction when applied to DTF type estimates. However, we know that the biased, non-gaussian nature of the distance estimator violates some basic assumptions of the correction and such a complex form of bias from such a simple sample distribution does not augur well for more realistic galaxy surveys. Figure 4.5 highlights this. Here the M and P distribution parameters have been changed to give only 10% log distance errors. Suddenly the inhomogeneously corrected recovery is dramatically worsened showing that the good results with 15% errors were just
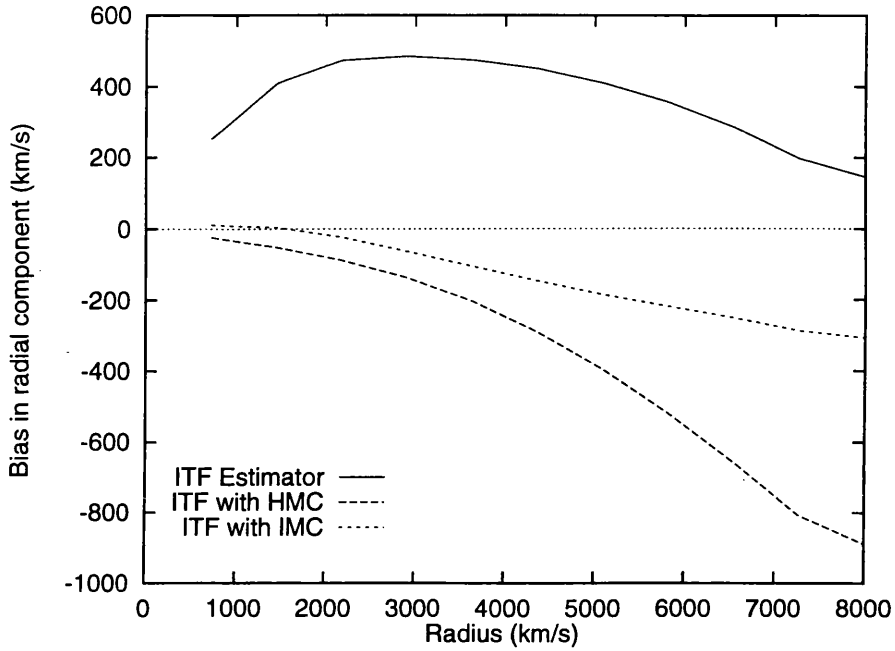
Figure 4.4: Bias as a function of distance for POTENT recoveries as in Fig. 4.3. This time, the ITF estimator has been used, both 'raw' and with the corrections.
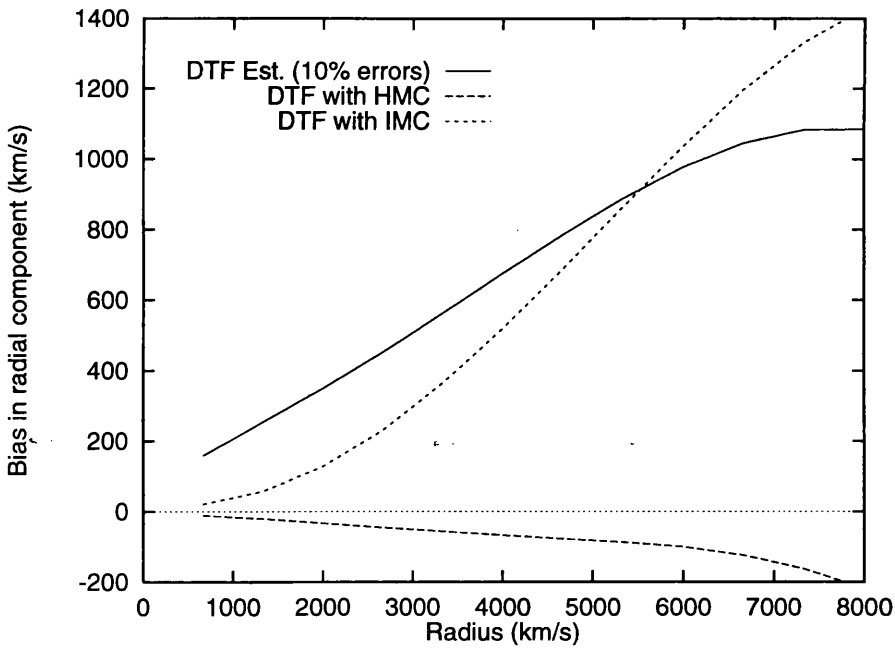

Figure 4.5: Bias as a function of distance for POTENT recoveries as in fig. 4.3. Again, DTF estimators are used, but this time they have only 10% log distance errors.

a lucky coincidence.

I have found that such cancellations are quite common in my Monte Carlo tests. Because of the many sources of systematic errors in POTENT, circumstances are bound to occur where two or more biases approximately cancel out. It is only by rigorously performing tests for a large variety of distances estimators and errors, galaxy distributions, velocity fields and so on, that such coincidences can be recognised and the true advantages and problems of any method brought to light. Too few tests may well make inappropriate methods seem attractive. Considerable care needs to be taken to avoid falling into traps of this sort.

The best recoveries, therefore, are as expected. Since we are drawing from an homogeneous universe, the homogeneous correction is indeed effective when used with the direct regression line estimator and the recovery from an ITF type estimator with an inhomogeneous correction is equally good (fig. 4.4).

So, the theory holds out well for this situation and the corrections, when properly applied, seem to be adequate. However, we need to consider the effect of inhomogeneities in the universe, incomplete samples and more complex velocity fields.

I perform the second, more realistic test on the same distance estimation techniques described above. The results of these tests are shown in fig. 4.6. For clarity, in fig. 4.7 the actual smoothed field has been subtracted from the Monte Carlo recoveries to give just the bias. As expected, the inhomogeneous correction provides an improvement over the 'raw' ITF recovery, although not a particularly substantial one. However, again, the inhomogeneously corrected DTF is quite good and surprisingly, the 'raw' DTF recovery is the best of the lot and the homogeneous correction makes it far worse. Since we know that in the idealised case exactly the opposite is true, the DTF recovery must be another "lucky" coincidence. These coincidences come about because of the complexity of the biases in POTENT: Malmquist type biases come out of the smoothing, the distance estimates themselves can be intrinsically biased, inhomogeneous galaxy distributions lead to pollution or sampling gradient biases and so on. The interaction between all the errors is complicated and unpredictable. Although the chances of a fortunate cancellation exist, they cannot be relied upon and the chance of an unrecognised bias getting through the correction procedure is high. One such bias can be seen in all four recoveries. In the lower right hand quadrant of the figures there is a large inward bias. This is caused by a large gap in the coverage of the
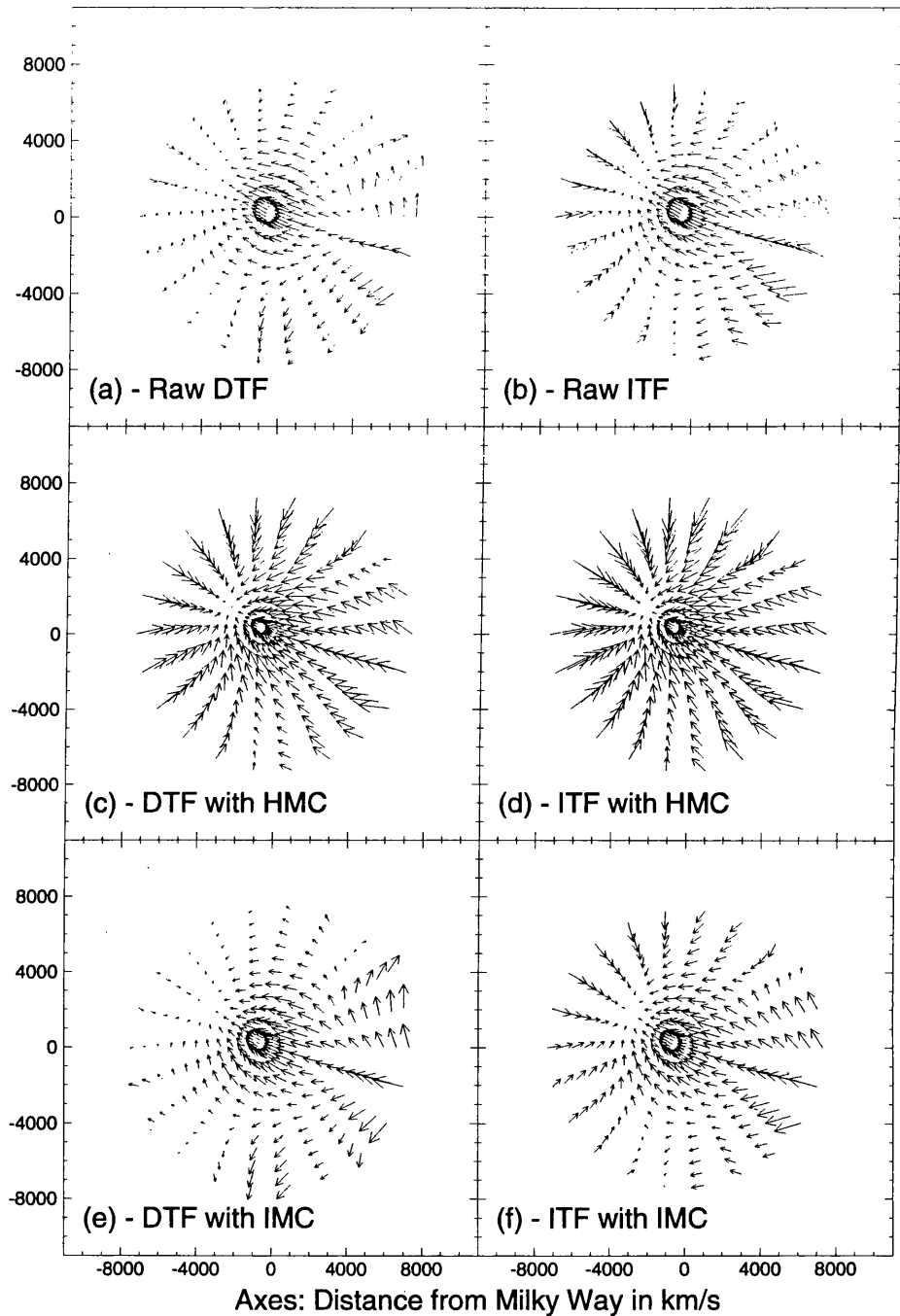
Figure 4.6: Monte Carlo POTENT recoveries of test velocity field shown in Fig. 4.2 using four different methods of distance estimation. In graphs (a) and (b), 'raw' DTF and ITF distance estimates are used. Graphs (c) and (d) have homogeneous corrections applied to the two 'raw' estimators and in (e) and (f), the inhomogeneous corrections have been used.
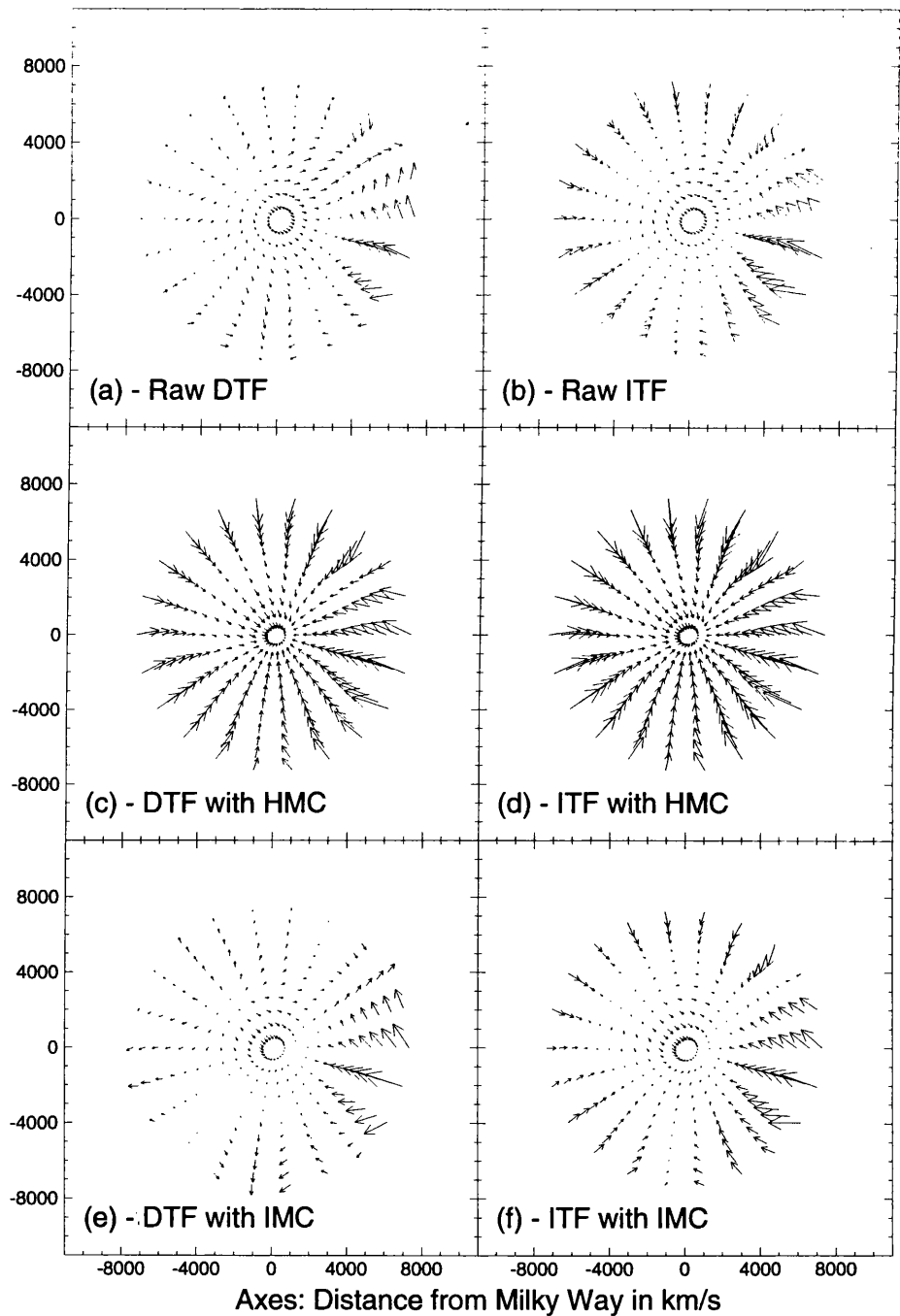
Figure 4.7: Biases of the four recoveries shown in Fig. 4.6. The actual smoothed velocity field has been subtracted from the Monte Carlo POTENT recoveries.

Mathewson et al. (1992) data set and it appears to shift the void to considerable greater distances in all cases. In fact, the situation is not quite as bad as it may seem as areas such as these have relatively large random errors and are usually discarded from the recovery before any analysis is done, but the danger remains.

The problems with the homogeneous correction with both 'raw' estimators are twofold. Firstly, the assumption of an homogeneous sample implied when ITF estimates are used is clearly not valid (see fig. 4.1) and the universe is probably not homogeneous enough on these scales to enable us to use the DTF. However, these effects are not sufficient to account for the large spurious inflows seen here. These are due to the modelling of the selection effects in the Monte Carlo simulations. Since I have a combination of two data sets and they are being used solely to provide realistic spatial distributions rather than actual data to be processed, the choice of magnitude cut-off and the sharpness of that cut-off will not exactly match the real data. Therefore, even if the real universe were homogeneous on these scales, the universe of the fake surveys will not be, since they are forced to reproduce the same sample distribution with different selection criteria. This is, therefore, an artifact of the Monte Carlo technique rather than the Malmquist corrections used, but the magnitude of the effect in POTENT recoveries should be taken as a warning both for the design of Monte Carlo type tests and for the use of the homogeneous correction when the homogeneity requirement is suspect.

Coned Inhomogeneous Malmquist Corrections

One possible improvement to the standard IMC is a coned version where the sky is divided up into a number of solid angle cones and the correction calculated separately for each one. However, with the current size of galaxy catalogues with redshift independent distance estimates, this is not very effective. For example, in fig. 4.8 another set of Monte Carlos have been performed on the same catalogue as the previous tests but with quiet Hubble flow imposed to remove and sampling gradient biases. The graph on the left shows the effect of an unconed IMC correction applied to ITF estimates. On the right, a separate IMC has been calculated for each galaxy by averaging the radial distribution over all the galaxies in a cone centred on the chosen galaxy and wide enough to contain 200 others. Given the sample contains less than 2000 galaxies, this would seem to be a reasonable sub-sample
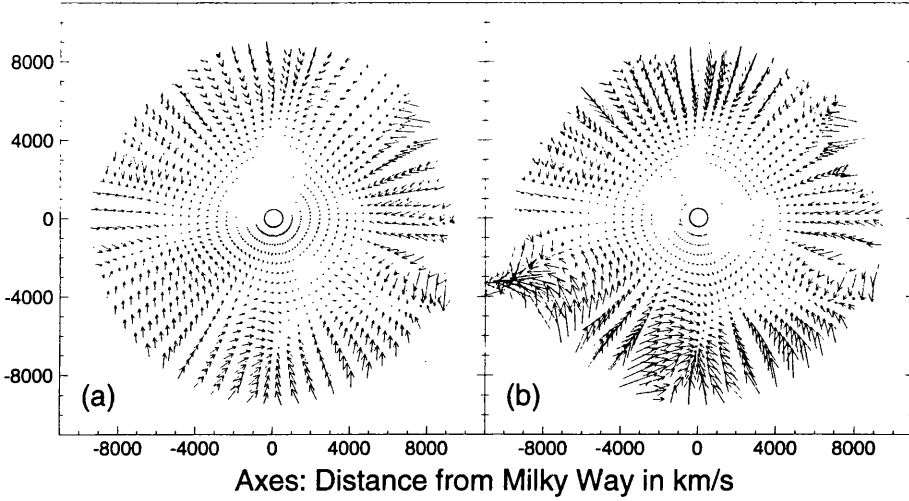
Axes: Distance from Milky Way in km/s

Figure 4.8: Two slices through Monte Carlo recoveries of quiet Hubble flow using POTENT on realisticly inhomogeneous data. Graph (a) shows the effect of applying IMC to ITF estimates. Graph (b) has an inhomogeneous correction calculated separately for each galaxy using a cone centered around it wide enough to contain 200 other galaxies.

to use, but as can be seen, the biases increase because of noise in the estimate $\hat{p}(\omega_0)$. If the number of galaxies in each cone is increased, the correction does not differ significantly from the unconed correction, if decreased, the noise because worse. Therefore, without a significant advance in the size of catalogues, this approach is unlikely to yield particularly useful results.

It is clear from these tests that even with careful choice of window function and corrected distance estimates, the POTENT method is susceptible to serious systematic errors which can go unnoticed and lead to dangerously erroneous analysis. The aim of the work in this thesis is to try to develop alternative methods of velocity field recovery or improved correction techniques. Simultaneously, I will be attempting to devise careful tests for the various methods to ensure that their regions of validity are well defined and all areas of any recovery can be used or discarded in confidence.

Chapter 5

Adapting and Improving Potent

Sure, all's effectless; yet nothing we'll omit

That bears recovery's name.

Pericles – Prince of Tyre V:i

The main area of potential improvement in POTENT is clearly in the correction of Malmquist bias. In particular, there is the need for some form of correction or procedure that can properly take into account the full three-dimensional distribution of galaxies.

5.1 Redshift Potent

It is important to remember that as well as distance estimates, POTENT obviously requires redshifts of galaxies. So far we have only used them to determine $\hat{v}_r$, but we have already seen that, even without any independent galaxy distance estimates, the redshift distribution of galaxies can provide considerable information about the peculiar velocity field. Also, in comparison to even the very best distance estimation techniques, redshifts are effectively error free. Given this, any technique that can replace some of the reliance on $\hat{r}$ with redshift information is likely to provide scope for improvement.

First of all, where exactly is $\hat{r}$ used in POTENT? Its main use is to estimate $\hat{v}_r$ as in eqn. (4.1). However, it is also needed to position galaxies during the smoothing procedure and it is at this stage that Malmquist biases arise. The idea behind *Redshift Potent* (or

Z-POTENT) is to perform as much of the analysis as possible in redshift space rather than estimated distance space.

It would be very convenient to perform the entire process without ever having to leave redshift space, but the assumption of a potential velocity field will not be valid, so the integration over the smoothed radial velocity field (4.4) must be done in real space. However, prior to this, the assumption of a potential is not utilised. We can, therefore, proceed as follows (a similar procedure has been proposed by Stebbins, 1993).

5.1.1 The Z-Potent Procedure

The first stage is unchanged. A radial peculiar velocity is estimated for each galaxy i by

$$\hat{\mathbf{v}}_{ri} = z_i - \hat{\mathbf{r}}_i \tag{5.1}$$

However, instead of associating the galaxy with a particular position in "estimate space", $\hat{\vec{r}} = \{\hat{\mathbf{r}}, \theta, \phi\}$, I use the position of the galaxy in redshift space, $\vec{z} = \{z, \theta, \phi\}$. (Note that, although z, θ and ϕ are all actually estimates, their errors are negligibly small). These velocities are then smoothed in redshift space by minimising the likelihood function:

$$L(\vec{z}) = -\frac{1}{2} \sum_i^{N_{gal}} W(\vec{z}, \vec{z}_i)[\hat{\mathbf{v}}_{ri} - \vec{z}_i \cdot \vec{v}(\vec{z})]^2 \tag{5.2}$$

(cf eqn. (4.5)). The result of this procedure is as before as smoothed peculiar velocity field but of the form

$$\hat{\mathbf{v}}_r(\vec{z}) \equiv \hat{\mathbf{v}}_r(\vec{r}(\vec{z})) \tag{5.3}$$

$$\text{where } r(\vec{z}) = z - \hat{\mathbf{v}}_r \tag{5.4}$$

Each radial line can, therefore, be converted from redshift space to real space before integration of eqn. (4.4) and the POTENT procedure continue as normal.

There are two possible problems with this approach. Firstly, each radial line in redshift space must be extended below the origin and beyond the maximum extent of the recovery so that (5.4) is defined all along the radial path in real space. This is easily done. The second problem is more significant.

Take some region along a radial line z_a to z_b where

$$\frac{d\hat{\mathbf{v}}_r}{dz} < 1, \quad z < z_a$$

$$\frac{d\hat{\mathbf{v}}_r}{dz} > 1, \quad z_a < z < z_b$$

There will, therefore, be a region in real space between $r_a = z_a - \hat{\mathbf{v}}_r(z_a)$ and $r_b = z_b - \hat{\mathbf{v}}_r(z_b)$ where more than one $\hat{\mathbf{v}}_r$ is associated with each point. This is clearly un-physical and comes about because there is no longer a one-to-one correspondence between real and redshift spaces (so-called triple-value regions). Unfortunately, there is no simple solution to this problem and the radial integral must be truncated when such a situation occurs. However, this occurrence is also an indication that we are moving out of the linear regime of structure formation and into a area where we might expect the assumption of a potential to fall down, so such regions should perhaps not be included in a POTENT analysis anyway.

5.1.2 Testing Z-Potent

As usual, I shall start with a simple test by imposing quiet Hubble flow on an homogeneous universe. The results of this test are shown in fig. 5.1. There is a clear improvement in the results with the smoothing performed in redshift space, both in the systematic errors and the standard deviation, particularly within $\sim 5000 \, \mathrm{km \, s^{-1}}$. However, in fig. 5.2, there is a different story.

Here, the Hubble flow field has been kept, but the Mathewson and Burstein catalogues are again used to give a realistic distribution of galaxies. Over most of the field, the two results are comparable with perhaps a slight preference towards the Z-POTENT results. However, for two of the shells, the errors in the redshift based recovery become enormous. This is because of the troublesome void in the catalogue that keeps recurring. As can be seen from fig. 5.3 (a), on the edge of this area there is a large region with the results undefined and some large, clearly erroneous velocities at $\sim (2000, 500)$. However, when the final field is smoothed slightly to remove this ugly area, the results are very different (fig. 5.3 b). The averaged errors for this case are shown in fig. 5.4 and again, the Z-POTENT recovery is the best result (albeit only slightly).

Of course, such ad-hoc smoothing because the field does not look attractive is far from satisfactory! However, in cases of this type, where a particular region is clearly invalid, it
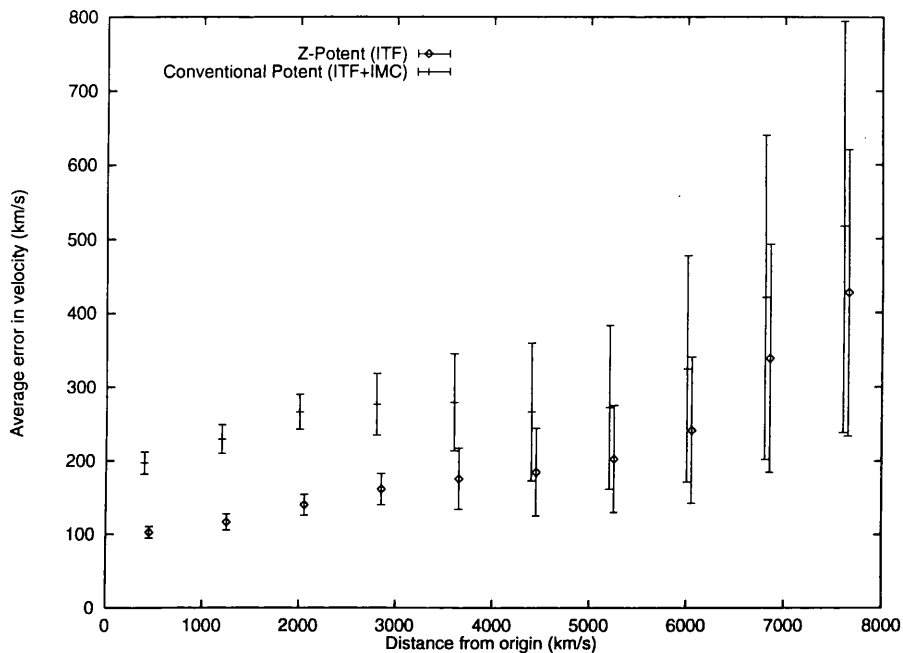
Figure 5.1: Average error and standard deviations are shown for the recovered velocities of quiet Hubble flow in an homogeneous universe for conventional POTENT applied to inhomogeneously Malmquist corrected ITF estimates and Z-POTENT applied to 'raw' ITF estimates, both techniques designed to minimise Malmquist biases. The velocities have been binned into ten radial shells. Note that the Z-POTENT results have been shifted slightly to the right for clarity.

can alternatively be discarded, also yielding similar results. And, again, the question needs to be posed that, if such an error did not show up in a conventional POTENT analysis, would our assumption of a potential still hold and the result still be valid?

5.2   Monte Carlo Bias Correction

Although attractive, the Z-POTENT method is still not an ideal solution, partly because of the areas where it is invalid, but also because it only addresses Malmquist-like biases, and not any other forms of systematic error such as sampling gradient biases. One approach I have adopted for dealing with this type of bias is to take a step back from the details of how particular errors arise, and treat them all jointly as a systematic error that needs to be calculated and corrected. I therefore wish to use POTENT itself to find these errors, for example using Monte Carlo recoveries similar to the ones shown in fig. 4.7, and then simply

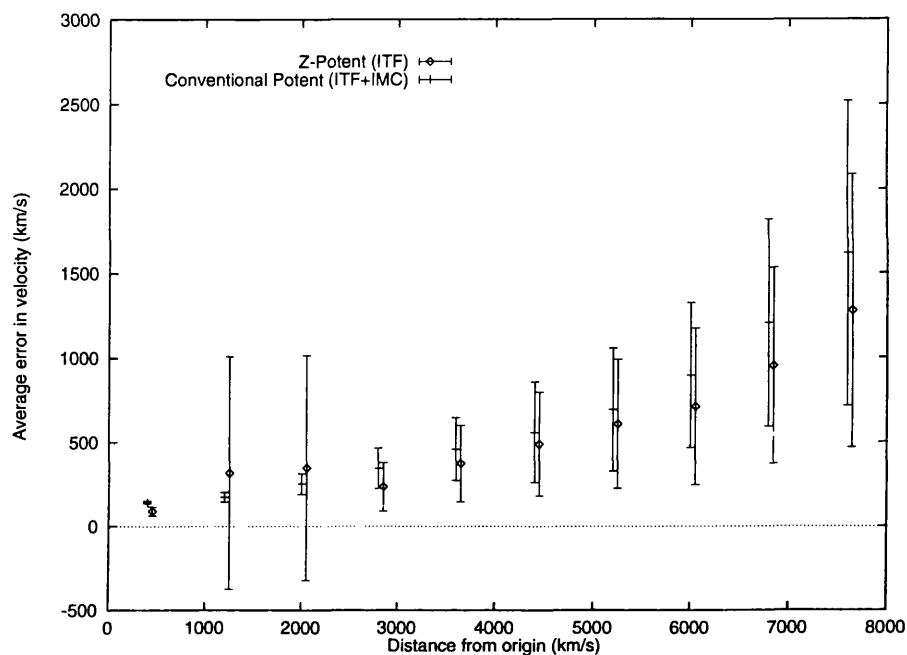Figure 5.2: As figure 5.1 but with QHF imposed on the positions of the Mathewson et al. (1992) and Burstein (1990) galaxy data sets.
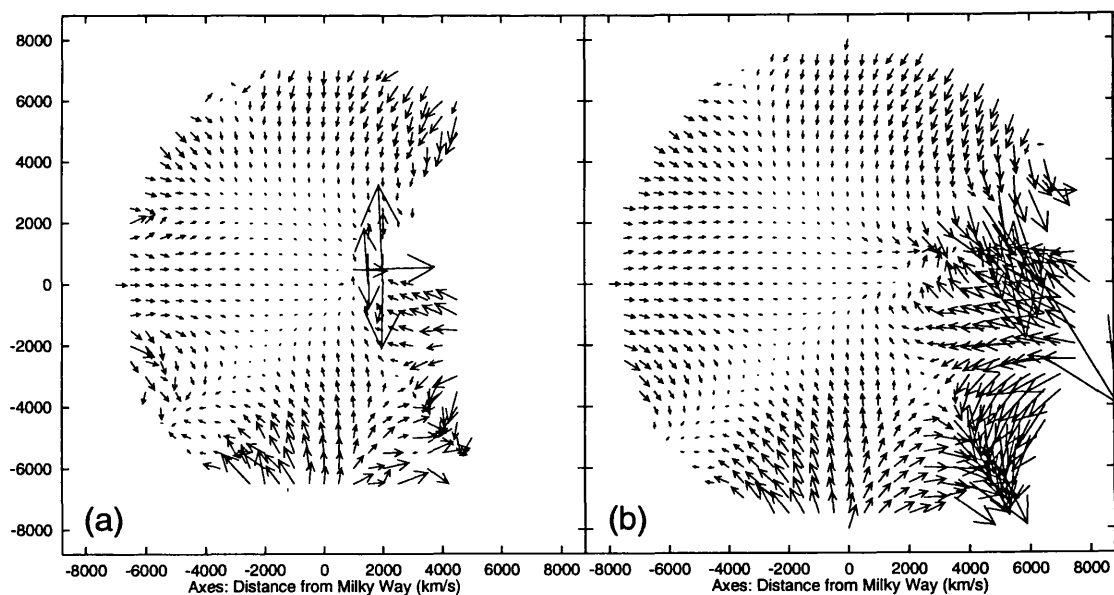

Figure 5.3: Slices through two example Z-POTENT recoveries. Both have QHF imposed on realistic galaxy positions, but in (a), the field is the simple recovery, and in (b), the recovered field has been smoothed slightly to remove the unrealistic velocities at about (2000, 500).
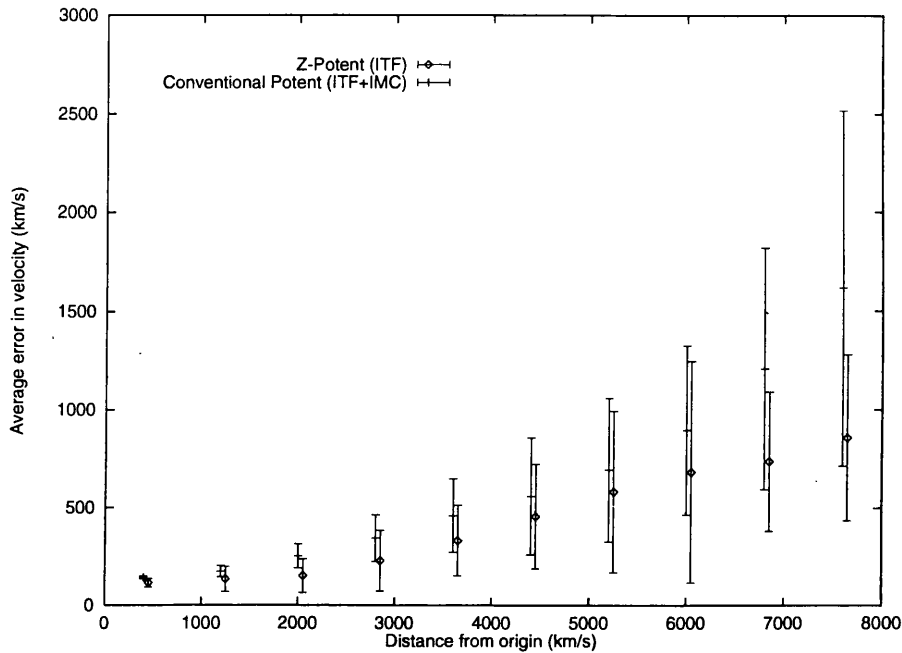
Figure 5.4: As figure 5.2 but with the Z-POTENT recovery smoothed slightly (see fig. 5.3).

subtract them from the actual recovery. This technique was first described in Newsam et al. (1993a) and was extended in Newsam et al. (1993b).

To do this, I divide the systematic errors into two classes; those that do not depend on the underlying velocity field and those that do. For example, the Malmquist bias is independent of the velocity field, but sampling gradient biases have a strong dependence. The former can be dealt with under the assumption of any velocity field (I take quiet Hubble flow for simplicity), but the latter requires some means of iterating towards a solution. Before I get to this, however, a Monte Carlo method must be carefully chosen.

5.2.1 Monte Carlo procedure

I need a procedure that uses the distribution of galaxies given by the data set but can impose some chosen velocity field onto them. I also need to be sure that the details of the distance estimator and selection of galaxies are consistent with those that went into forming the original sample. Finally, I need to be certain that no systematic errors are introduced by the procedure itself, so that the results can be confidently treated as a correction.

I therefore use the following method.

1. Create a mock universe of observable galaxies by assigning some definite position $\vec{r}_f$ to each galaxy in a catalogue observed at some $\hat{\vec{r}}$. (For simplicity, this is done by $\vec{r}_f = \hat{\vec{r}}$ for all galaxies. However, it may be better in some cases to use the redshift as a position for more distant galaxies).

2. Imposing some velocity field $\vec{v}_f(\vec{r}_f)$, assign a redshift z_f to each galaxy such that

$$z_f = v_{rf} - r_f \tag{5.5}$$

3. For the example of an estimator based on two observables **M** and **P**, assign an **M** and a **P** to each galaxy. These will be randomly sampled from a distribution whose parameters are estimated from the original data set.

4. Impose selection on each galaxy and, if unobservable, go back to step 3.

5. Get estimates of distances to galaxies $\hat{r}_f$.

6. Use $\hat{r}_f$ and z_f in POTENT to get a recovered velocity field.

7. Repeat steps 3 to 6 a suitable number of times and average the resulting velocity field.

This will result in a good approximation to the systematic errors introduced by POTENT when acting on a particular velocity field with a particular galaxy catalogue. The iterative process involves modifying the imposed velocity field $\vec{v}_f(\vec{r}_f)$ at each stage and, hopefully, converging it towards the underlying smoothed field.

5.2.2   Application of iterative corrections

The basis of the iteration is to use the Monte Carlo results of some 'guess' velocity field to give a next guess that is a little bit closer to the underlying $\vec{v}$.

I shall call the first guess field $\vec{v}_f^{(0)}$. I then apply the Monte Carlo procedure to obtain an estimate of the systematic errors $\vec{b}_r(\vec{r})$ this field produces. For simplicity of notation, I will call this Monte Carlo velocity field $\bar{\vec{v}}_f^{(0)}$ where the bar denotes the average including all biases. Thus

$$\vec{b}_f^{(0)}(\vec{r}) = \bar{\vec{v}}_f^{(0)}(\vec{r}) - \vec{v}_f^{(0)}(\vec{r}) \tag{5.6}$$

The next iteration could then be found by removing these biases from the single POTENT recovery of the original catalogue. However, this recovery is noisy and using it as the basis of an underlying velocity field in the iterative scheme will try and force the method to fit the noise. In order to avoid this, it is necessary to perform a Monte Carlo on the 'raw' data by scattering the distances as above, but using the actual redshifts from the catalogue to define the imposed velocity field. This will give us an averaged field $\bar{\vec{v}}_{raw}(\vec{r})$ which well approximates the combination of systematic errors from POTENT and the real smoothed velocity field.

One then obtains the next iteration velocity field:

$$\vec{v}_f^{(1)} = \bar{\vec{v}}_{raw} - \vec{b}_f^{(0)} \tag{5.7}$$

The Monte Carlo procedure can then be repeated so that, in general, to obtain the velocity field for iteration n,

$$\vec{v}_f^{(n)} = \bar{\vec{v}}_{raw} - \vec{b}_f^{(n-1)} \tag{5.8}$$

5.2.3 Convergence criteria

Convergence will have occurred when, within some tolerance,

$$\vec{v}_f^{(n)} \approx \bar{\vec{v}}_{raw} \tag{5.9}$$

However, exactly how to define the comparison of the two fields is not simple. The simplest approach would be to do a point by point comparison of the two grids and average all the values obtained. Unfortunately, this approach has two problems. Firstly, some areas of the POTENT recovery will be extremely noisy (particularly at large distances) and to get the residual noise in the Monte Carlos down to really acceptable levels would be prohibitively time consuming. In addition, the large gaps in galaxy surveys, whether they are due to actual voids or incomplete sky coverage, give areas where there is little or no information about the velocity field. Since the correction method acts by adjusting the redshifts of the galaxies it is using, where these galaxies are sparse very little effect can be achieved. However, if the convergence criteria gives too much weight to these regions, they may prevent the tolerance level being reached and the corrections will 'over-shoot'. In practice, areas which are subject to these problems are discarded before any analysis of the POTENT

results is performed, so we would wish to minimise their effect on the convergence criteria without making any ad hoc decisions about 'good' and 'bad' areas of the field.

This is done in two simple ways. Firstly, instead of comparing at grid points, I interpolate the fields to the 'fixed' positions of the galaxies and perform the comparison there. In this way, most weight is put on densely sampled regions where the recovery is likely to be most useful. Also, the error of the $\vec{v}_{\mathrm{raw}}$ recovery is calculated and used to weight the comparison at every point, again making the correction best where it is needed. This weighted average is used as the value of the current iteration, and when it drops below some tolerance (usually in the tens of $\mathrm{km\,s^{-1}}$), the process is stopped.

5.2.4 Tests of Iterative Monte Carlo Corrections

The tests in this section are for direct comparison to the second, more realistic set of results in sec. 4.2. Therefore, the underlying velocity field is the same, as are the distribution of galaxies and the parameters used to create the data for the distance estimates. Although the method can, in theory, be used with any number of different distance estimators, I have chosen to use an ITF estimator. There is no overwhelming reason why this is the best estimator to use, but since it is unbiased, it will not introduce any systematic errors of its own, thereby perhaps speeding up convergence. In addition, its gaussian nature makes it easy to perform analysis on.

For the tests shown in this section, I used a convergence tolerance of 50 $\mathrm{km\,s^{-1}}$ where the convergence criteria is as described in sec. 5.2.3. In order to ensure that residual noise in the Monte Carlo averages was not significant at this level, each Monte Carlo recovery used 200 POTENT realisations of the velocity field. This is almost certainly over-kill, but it is important to be sure that the tests are as free from confusion as possible.

All that remains to be decided, therefore, is the choice of the initial guess field $\vec{v}_{\mathrm{f}}^{(0)}(\vec{r})$. I choose quiet Hubble flow partly because it has no features that might pose unreasonable constraints on the recovery method, and also, since all peculiar velocities are zero, there are no sampling gradient biases in the recovery of quiet Hubble flow. This means that the bias of the initial recovery $\vec{b}_{\mathrm{f}}^{(0)}(\vec{r})$ is due solely to the effects of smoothing over the galaxy distribution distorted by distance errors. Therefore, this bias can be used as a form
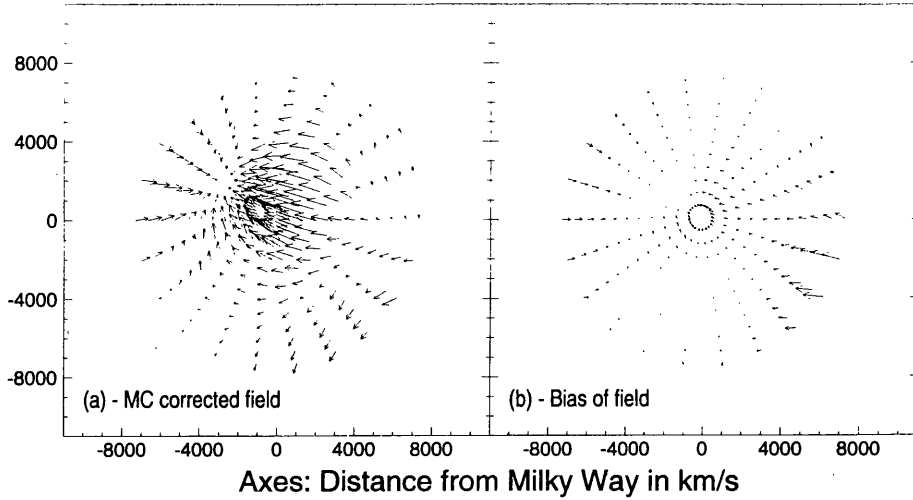
Figure 5.5: Average POTENT recovery of field given in fig. 4.2 using Monte Carlo iterative correction method. Graph (a) is from the final corrected recovery and (b) is its bias.

of Malmquist correction that takes into account *all* the inhomogeneities in the sample and makes no assumptions about the form of the distance estimator. If this is all that is required, the process can be stopped there. (In Hendry et al. (1993a), the effect of this correction on the problem of the "Zone of Avoidance" is considered and a significant improvement is shown).

The usual slice through the final velocity field is shown in fig. 5.5 together with its bias. The recovery is considerably better than any of the previous results in sec. 4.2. Just how much better can be seen in fig. 5.6. Here, the various averaged recovered velocity fields have been interpolated onto a cubic grid truncated to the 8000 $\mathrm{km\,s^{-1}}$ sphere. The grid spacing is chosen to be 500 $\mathrm{km\,s^{-1}}$ for comparison with the POTENT90 results of Dekel et al. (1993). Then, I simply compare the velocity at each grid point with the actual smoothed velocity field and average all the errors. The weighted comparison uses the variance of the POTENT recovery to weight each point. With both comparisons the Monte Carlo correction is better, considerably so without the weighting.

In fig. 5.7, I perform the same comparisons with each iteration of the Monte Carlo correction. The fourth iteration is the one I used since this is the one where the convergence level dropped below 50 $\mathrm{km\,s^{-1}}$ (shown by the thick line in the figure). However, it is clear that the best iteration is the second one, if only by a small margin. This is caused in part by
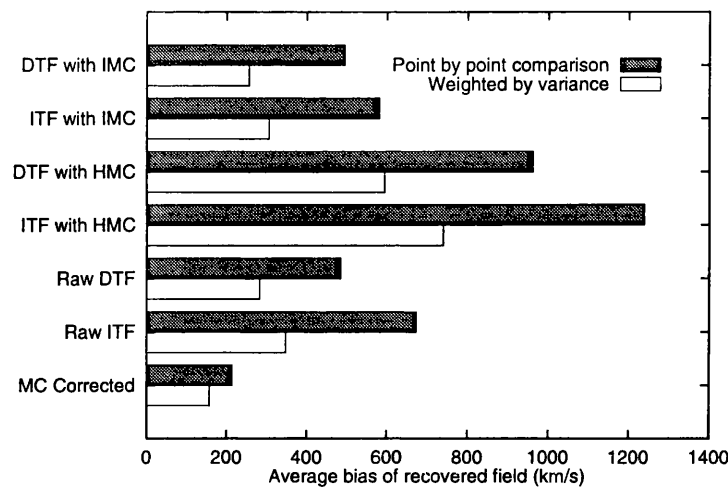
Figure 5.6: A comparison of the various methods. For the purposes of the comparison, the Monte Carlo recoveries shown in previous figures are interpolated on to cubic grids with grid spacing of 500 $km s^{-1}$. The shaded bars show the average errors of a simple point by point comparison of these interpolated fields with the actual smoothed velocity field. The hollow bars are the same comparison with each point weighted by the variance of the POTENT recovery at that point.
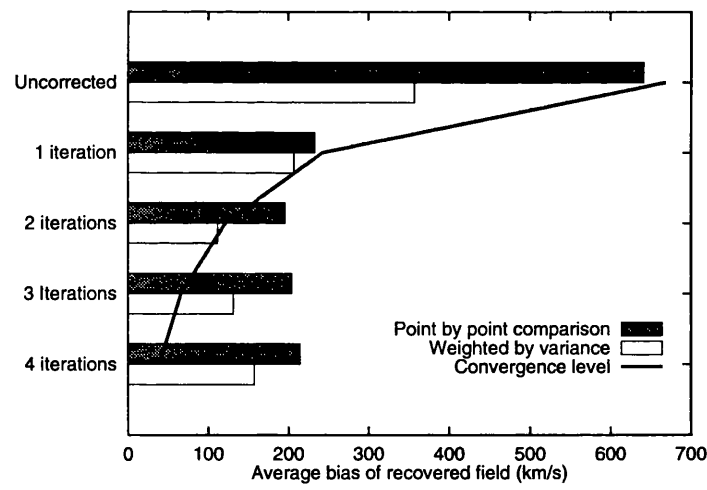
Figure 5.7: The average biases of the Monte Carlo correction technique at each iteration. As with fig. 5.6, the shaded bars show the average of a point by point comparison of the recovered average field with the smoothed real field on a cubic grid. For the hollow bars, the comparison was weighted at each point by the variance of the POTENT recovery. The thick line shows the level of the convergence criteria value as described in sec. 5.2.3.

corrections in the sparse regions over-shooting since the limited number of galaxies gives the correction very little to work with and so it continually tries to remove the same bias. But the small residual bias in the final iteration (fig. 5.5) even in the very sparse region mention earlier (fig. 4.1) shows that this is a relatively small effect, and something else is needed to account for the seeming divergence. To understand what, we need to recall that the convergence level was calculated using a comparison involving $\bar{v}_{\mathrm{raw}}(\vec{r})$. This is only an estimate of the combination of systematic errors from POTENT and the real smoothed velocity field. The fact that the method is as successful as it is, is a testament to the fact that it is a fairly good estimate, but down at the levels of tens of $\mathrm{km\,s}^{-1}$ which we are considering, this assumption must fall down. Therefore, if convergence levels this low are really required, then some means of improving this estimate will need to be found. However, most of the work is clearly done in the first two or three iterations, so convergence levels of $\sim 100\ \mathrm{km\,s}^{-1}$ would really be more useful, and would considerably reduce the computational overhead. Also, there is no reason to expect to method to converge at all. All that can be realistically hoped for is asymptotic convergence, so trying to force the method to do too much may well cause the recovery to diverge (as seen in fig. 5.7).

5.3 Potent and Max-flow Algorithms

So far in this chapter, I have only considered the systematic errors from POTENT because of their insidious effect. However, the random errors in the solution are in many ways an even bigger problem. Unlike biases, they can clearly not be 'corrected for', and where they are large, there is no option but to discard large slabs of the recovered velocity field.

The main area of attack on random errors is the smoothing procedure (sec. 4.1.3) where the errors in $\hat{v}_r$ are averaged out over large volumes. However, in sparser regions, the smoothing windows have to become uncomfortably large and for the sparse and non-uniform distribution of galaxies in a real catalogue, this can be a major consideration and lead either to unreasonably large smoothing windows that mask any useful information, or very noisy recovered fields. Recall, though, that the smoothing procedure of eqn. (4.5) is actually fitting to a *bulk flow* and although the conventional POTENT method only makes use of the radial component of this fit, both the transverse components are also estimated (Simmons

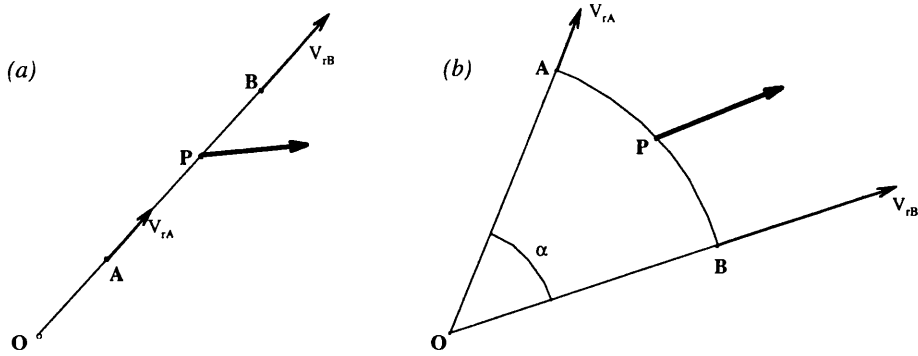Figure 5.8: Best fit vectorial velocity at given point P. In (a), galaxies A and B lie on a ray. Redshifts provide no information on transverse component of velocity at P. In (b), galaxies lie on an arc. Both transverse and radial components are determined.
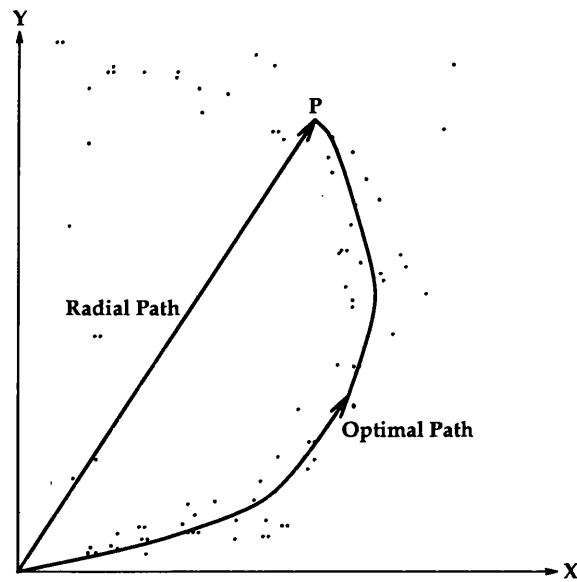

Figure 5.9: The optimal path minimises the error of the velocity potential, and will be pulled towards regions of high galaxy number density.

et al. (1993a), Simmons et al. (1993b)). Of course there are reasons to suppose the radial component will be more accurately determined than the transverse (θ, ϕ) components since the redshifts are essentially telling us the radial components. This can be easily seen from simplified picture given in fig. 5.8 where only two galaxies are considered. In case (a) where the observed galaxies are positioned along a ray it is evident that the radial component of the initial smoothed velocity field is well determined, but there is no information regarding the transverse component. Hence the error on the transverse component will be large. In case (b) where the galaxies are transversally positioned, the transverse as well as radial component will be reasonably well determined. The larger the angle, α, between the lines of sight to the two galaxies, the better determined will be the transverse component, v_θ. However, even if the error on the radial component were a factor of ten smaller than on the transverse, it is still possible to gain advantage by taking non-radial paths. Figure 5.9 schematically depicts such a situation.

5.3.1 Errors on the Radial and Transverse Components

Dekel et al. (1990) obtain the initial smoothed velocity field, $\vec{v}(\vec{r})$, at arbitrary spatial point $\vec{r}$, by minimising

$$\sum_{i=1}^{n} (\vec{v}.\vec{e}_r(\vec{r}_i) - u_i)^2 W(\vec{r}, \vec{r}_i) \tag{5.10}$$

where $\vec{r}_i$ is the position vector of the ith galaxy, and there are a total of n galaxies in the survey. $\vec{e}_r(\vec{r}_i)$ is the unit vector in the radial direction to the ith galaxy at position $\vec{r}_i$. $W(\vec{r}, \vec{r}_i)$ is the weighting or window function that determines the relative importance of the ith galaxy as in sec. 4.1.3.

Writing

$$\vec{v}(\vec{r}) = v_r \vec{e}_r(\vec{r}) + v_\theta \vec{e}_\theta(\vec{r}) + v_\phi \vec{e}_\phi(\vec{r}) \tag{5.11}$$

and

$$
\begin{aligned}
s_{rr}^i &= \vec{e}_r(\vec{r}).\vec{e}_r(\vec{r}_i) \\
s_{\phi r}^i &= \vec{e}_\phi(\vec{r}).\vec{e}_r(\vec{r}_i) \\
s_{\theta r}^i &= \vec{e}_\theta(\vec{r}).\vec{e}_r(\vec{r}_i)
\end{aligned}
\tag{5.12}
$$

and for expediency $W^i = W(\vec{r}, \vec{r}_i)$, minimisation yields

$$\sum_i W^i \begin{pmatrix} s_{rr}^i s_{rr}^i & s_{rr}^i s_{\theta r}^i & s_{rr}^i s_{\phi r}^i \\ s_{rr}^i s_{\theta r}^i & s_{\theta r}^i s_{\theta r}^i & s_{\theta r}^i s_{\phi r}^i \\ s_{rr}^i s_{\phi r}^i & s_{\theta r}^i s_{\phi r}^i & s_{\phi r}^i s_{\phi r}^i \end{pmatrix} \begin{pmatrix} v_r \\ v_\theta \\ v_\phi \end{pmatrix}$$

$$= \sum_i \begin{pmatrix} u_i s_{rr}^i W^i \\ u_i s_{\theta r}^i W^i \\ u_i s_{\phi r}^i W^i \end{pmatrix} \tag{5.13}$$

which can be written

$$\mathbf{AV} = \mathbf{b} \tag{5.14}$$

$\mathbf{V}$ is simply the vector of components of the smoothed velocity. (Compare this to the simplified approach given in sec. 4.1.1). Evidently the inversion of this equation yields all three components of the initial smoothed velocity field.

Since the estimated distances of galaxies are subject to error there will be a corresponding error on $\vec{v}(\vec{r})$. Let us write the initial smooth peculiar velocity field obtained from one realisation as

$$\hat{\vec{v}}(\vec{r}) = \vec{v}(\vec{r}) + \delta\hat{\vec{v}}(\vec{r}) \tag{5.15}$$

Here, for simplicity, we shall assume that all biases in the estimated velocity have been removed, so that on average, over many realisations, the estimated velocity yields the smoothed velocity field achieved when no noise is present, but using the same window function and redshifts.

In their analysis, Dekel et al. (1990) (see their appendix A) carry out a linear error analysis in which they derive a bias and variance on the radial component v_r of $\vec{v}(\vec{r})$ at an arbitrary spatial point. Of course the errors on the estimated components $\hat{\mathbf{v}}_r(\vec{r})$, $\hat{\mathbf{v}}_\phi(\vec{r})$, and $\hat{\mathbf{v}}_\theta(\vec{r})$ are not statistically independent.

Furthermore, the errors on the estimated velocity at different spatial points will also be correlated, typically over a distance scale determined by the diameter of the window function. This analysis can easily be generalised to the vectorial case. Since we have assumed that the bias has been removed from $\hat{\mathbf{v}}_r(\vec{r})$ so that $E(\delta\hat{\mathbf{v}}_r(\vec{r})) = 0$, we can write the covariance as

$$E(\delta\hat{\mathbf{v}}_i(\vec{r})\,\delta\hat{\mathbf{v}}_j(\vec{s})) = R_{ij}(\vec{r}, \vec{s}) \tag{5.16}$$

where E denotes expected value. The velocity error autocorrelation function, $R_{ij}(\vec{r}, \vec{s})$, will depend on the window function, number density, the dispersion of the distance estimator and also to some extent on the input peculiar velocities. My main purpose during this thesis is to demonstrate the viability of the method and I shall not attempt to model accurately $R_{ij}(\vec{r}, \vec{s})$.

5.3.2 Optimal Paths

The estimated potential, $\Phi(\vec{r})$, of the peculiar velocity field is given by the path integral

$$\hat{\Phi}(\vec{r}) = - \oint_0^{\vec{r}} \vec{v}(\vec{s}).d\vec{s} \tag{5.17}$$

It is important to note that the error, $\delta\hat{\Phi}(\vec{r})$ arises only from the error, $\delta\hat{\vec{v}}(\vec{s})$ in the estimated initial peculiar velocity. The radius vector of the path is not a statistical variable, but is, once the path has been chosen, strictly determined. The error on the potential is therefore given by

$$\delta\hat{\Phi}(\vec{r}) = - \oint_0^{\vec{r}} \delta\hat{\vec{v}}(\vec{s}).d\vec{s} \tag{5.18}$$

I shall take the optimal path for obtaining the potential of the velocity field at position $\vec{r}$ to be that path for which the variance of the line integral is minimised.

If we assume that $\hat{\vec{v}}$ is unbiased so that $E(\delta\hat{\vec{v}}) = 0$ at every spatial point, $\vec{s}$, then evidently

$$E(\delta\hat{\Phi}) = 0 \tag{5.19}$$

The variance of $\delta\hat{\Phi}$ is given by

$$\begin{aligned}
E(\delta\hat{\Phi}(\vec{r}))^2 &= E\left(\oint_0^{\vec{r}} \delta\hat{\vec{v}}(\vec{s}).d\vec{s} \oint_0^{\vec{r}} \delta\hat{\vec{v}}(\vec{t}).d\vec{t} \right) \\
&= \int_0^\alpha \int_0^\alpha R_{ij}(\vec{s}(\mu), \vec{t}(\nu))\, x^{i'}(\mu)\, x^{j'}(\nu)\, d\mu d\nu
\end{aligned} \tag{5.20}$$

where the path has been parametrised

$$x^i = x^i(\mu), \quad 0 \le \mu \le \alpha \tag{5.21}$$

and

$$x^{i'}(\mu) = \frac{dx^i}{d\mu} \tag{5.22}$$

Thus the optimal path will be given by

$$\delta \int_0^\alpha \int_0^\alpha R_{ij}(\vec{s}(\mu), \vec{t}(\nu))\, x^{i'}(\mu)\, x^{j'}(\nu)\, d\mu d\nu = 0 \tag{5.23}$$

In the case where the autocorrelation function is a delta function, ie

$$R_{ij}(\vec{s}, \vec{t}) = \delta^3(\vec{s} - \vec{t})\, \sigma_{ij}(\vec{s}) \tag{5.24}$$

Equation (5.23) simply defines a geodesic on a Riemannian space with σ_{ij} as its metric tensor. Generally, however, we can expect components of the initial smoothed field at different spatial points to be correlated, and the correlation length to be of the same order of magnitude as the effective radius of the window function.

Although eqn. (5.23) is very interesting from a mathematical viewpoint, I have not proceeded further along those lines. In practice we do not know the form of the autocorrelation function, and it could be best approximated by numerical simulations. A more natural way to proceed is to use finite element methods.

5.3.3 Dijkstra's Algorithm

We wish to calculate the 'best' velocity potential, from which the peculiar velocity field may be obtained by taking the gradient. This requires determining the potential Φ at regular grid points, at least in regions of space where the galaxies are sufficiently dense for the reconstruction to be meaningful.

Suppose we have N gridpoints at which we wish to evaluate the potential of the velocity field. Let us assume that the error in moving between the ath and the bth gridpoint is well defined and known for all a and b. This 'error length', which I shall call an arcweight, should depend on the number density of galaxies in the joint neighbourhood of both gridpoints, the distance between the gridpoints and the distance of both from the origin. I take the potential to be zero at the first gridpoint (our galaxy), and so wish to find the path along which the total error is least. At first sight it might appear that this is an *NP-complete* problem (i.e. only soluble exponentially and therefore impractical on even the largest computers. For a fuller description NP-complete and related problems, see the reference below). Luckily this is not the case. Dijkstra's algorithm (cf Papadimitriou and Steiglitz, 1982) finds the exact
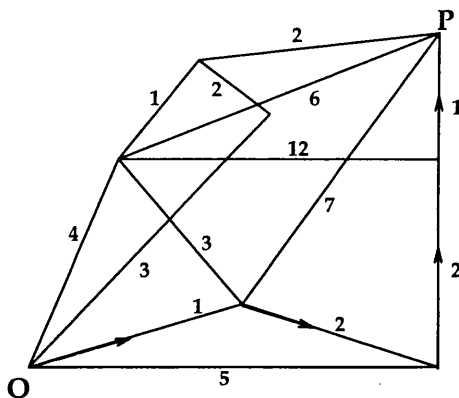
Figure 5.10: Shortest path from O to P, indicated by arrows, can be obtained using Dijkstra's algorithm.

and global solution to this problem in a number of steps that is bounded above by N^2. This algorithm is one of a class of *Max-flow* algorithms used in network theory.

Description of Dijkstra's Algorithm

Figure 5.10 gives a schematic representations of a network of nodes (gridpoints). We write the set of nodes as

$$\mathbf{N} = \{n_1, n_2, n_3, \ldots, n_N\} \tag{5.25}$$

The arcweight between each pair of nodes has been calculated according to some prescription. Let the arcweight between node n_a and node n_b be c_{ab}. We shall assume that $c_{ab} = c_{ba}$, although this is not strictly necessary for what follows. A path will be determined by its nodes. Thus the path $n_{a_1} n_{a_2} n_{a_3} n_{a_4} \ldots n_{a_M}$ has M nodes and total arcweight or pathlength

$$c_{a_1 a_2 a_3 \ldots a_M} = c_{a_1 a_2} + c_{a_2 a_3} + \ldots + c_{a_{M-1} a_M} \tag{5.26}$$

Clearly we may also write

$$c_{a_1 a_2 a_3 \ldots a_M} = c_{a_1 a_2 a_3 \ldots a_{M-1}} + c_{a_{M-1} a_M} \tag{5.27}$$

The problem is to find the path from node n_1 to every other node n_a that minimises the sum of arcweights.

Suppose that at some stage we have a set $\mathbf{M}$ of nodes to which we have established the minimum pathlength. Let the remaining nodes form the set $\mathbf{P}$ (see Fig. 5.11). Define $s_{\mathbf{M}}(p)$
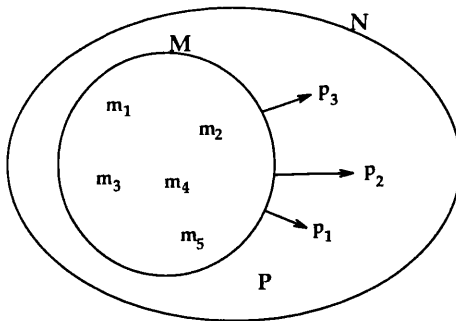
Figure 5.11: Illustration of Dijkstra's algorithm. **M** is set of nodes for which optimal paths from m_1 have been found. **N** is the set of all nodes and **P = N − M**.

to be the shortest pathlength from n_1 to p *that uses only intermediate nodes* in **M**. This must also be the shortest path from n_1 to p. Now choose from the set of nodes **P** the node p for which the path length is shortest, and add it to the set **M** and repeat until until the set **M** contains all the nodes. This algorithm is similar, both in principle and in implementation, to the *Minimal Spanning Tree*, already used in statistical analyses of large scale structure (Barrow et al., 1985).

5.3.4   Application of Dijkstra's Algorithm to Potent

The main problem in applying this method to POTENT is how to establish the arcweights between nodes. The arguments I present here are largely heuristic. Errors between two nodes (gridpoints) can be expected to be determined predominantly by the number density of galaxies in the mutual neighbourhood of the two nodes. The higher the galaxy number density the smaller the error. Radial components of the initial peculiar velocity field will probably have lower errors than the transverse components, but we might expect the two transverse components to have the same errors. If we take the gridpoints n and m to be separated by more than the correlation length of the autocorrelation function, then we can assume that $\delta\vec{v}(m)$ and $\delta\vec{v}(n)$ are uncorrelated. Hence it would be reasonable to take

$$
\begin{aligned}
E(\delta\hat{\Phi}(\vec{r}))^2 &= \sum E(\delta\vec{v}_{ab}\cdot\Delta\vec{x}_{ab})^2 \\
&= \sum E(\delta v_{iab}\,\delta v_{jab})\,\Delta x^i_{ab}\,\Delta x^j_{ab} \qquad (5.28)
\end{aligned}
$$

where n_a and n_b are consecutive nodes along the path of integration, $\Delta\vec{x}_{\rm ab}$ the separation between these two gridpoints, and $\delta v_{i\rm ab}$ is the error in the ith component of the initial smoothed peculiar velocity evaluated at the midpoint of the segment. Since the variance of the distance estimator increases with radial distance squared, we shall take the arcweight to also scale with r^2. To simplify, we disallow arcs between gridpoints more than three gridlengths away, and assume that

$$\sigma_{r\theta} = \sigma_{r\phi} = \sigma_{\phi\theta} = 0 \text{ and } \sigma_{rr} = k^2\sigma_{\theta\theta} = k^2\sigma_{\phi\phi} \qquad (5.29)$$

where k is some parameter.

Thus we shall take $c_{\rm ab}$ to be of the form

$$c_{\rm ab} = n^{\alpha}r^2((\Delta r)^2 + k^2 r^2(\sin^2\theta(\Delta\phi)^2 + (\Delta\theta)^2)) \qquad (5.30)$$

The value of k in the above equation essentially tells one the errors on the transverse components of the initial field compared with the radial. This will obviously depend on the window function. It will also depend on the actual peculiar velocity field. Very rough simulations indicate that for simple peculiar velocity fields the transverse components will be poorly recovered, and consequently k will be large, typically between 5 and 10. However, as the quality of the data is improved, k could drop as low as 2 or 3.

A map of some values of k for a typical recovery are shown in fig. 5.12. This map is obtained by imposing quiet Hubble flow onto a catalogue of galaxies (see sect. 4.2) and applying the POTENT smoothing procedure on its own to recover the velocity field. Each grid point is then assigned a value of k according to the magnitude of the transverse errors with respect to the radial ones and the k-field smoothed slightly (using a spherical gaussian window of $300\,\rm km\,s^{-1}$ radius). Some areas clearly have very large k, but in general, values of around 10 are common, and for much of the slice – particularly in regions with higher galaxy density such as the lower left corner, k is considerably less than 10.

n^{α} is used to used to control the number of steps in any path. In general, α can be varied from 0 (no preference) to about -9 (large numbers of steps strongly preferred).

As before, I have taken a spatial distribution of galaxies to be given by the Mathewson survey (Mathewson et al., 1992) combined with the Burstein Mark II compilation (Burstein, 1990) – see fig. 4.1. Two peculiar velocity fields are taken corresponding to quiet Hubble flow and
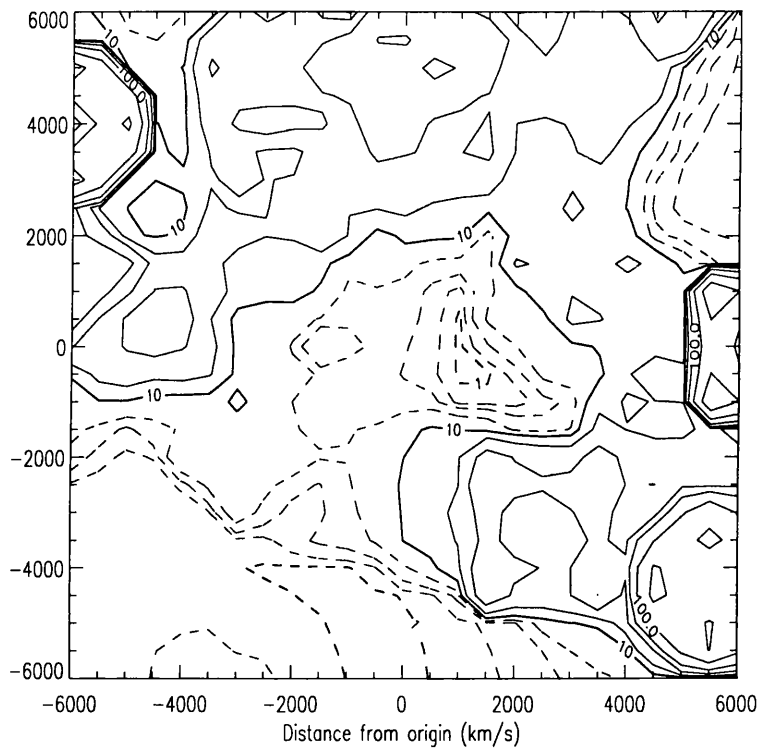
Figure 5.12: The graph show a slice though a typical set of k values. POTENT smoothing is performed on a mock catalogue based on the Mathewson and Burstein data sets but with quiet Hubble flow imposed. The contours (logarithmically spaced) show the value of k (relative errors in radial to transverse components - see eqn (5.30)) at a number of gridpoints. Contours with $k < 10$ are dashed and the $k = 1$ and $k = 10$ contours are drawn with thicker lines.
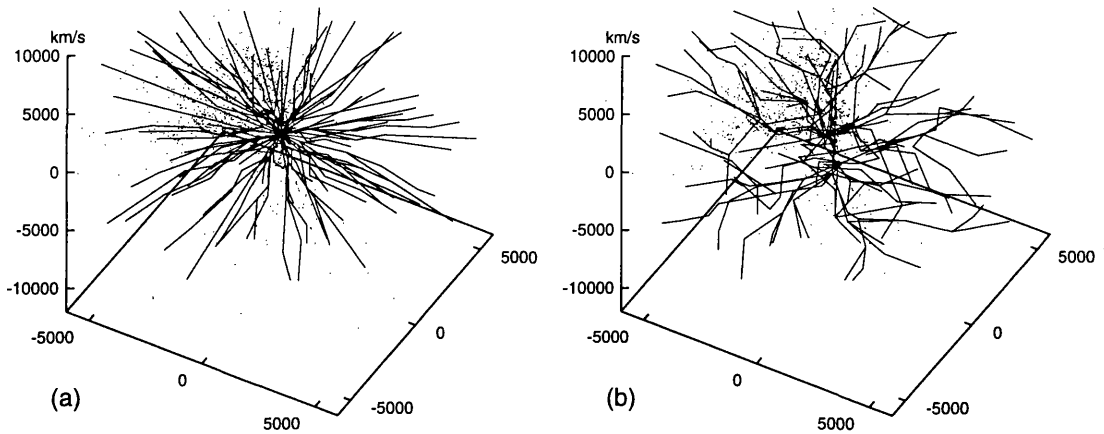
Figure 5.13: Some optimal paths with positions of galaxies indicated by dots. Paths in figure (a) have low density weighting and most are almost radial. In figure (b), paths have high density weighting and are non-radial.

uniform streaming. Galaxy distances are subjected to distance errors, and the minimum length paths found using arcweights of the form in eqn. (5.30) to define the line integral in eqn. (5.17), and hence the velocity potential. Figures 5.13 and 5.14 shows the optimal paths for two different heuristic arcweight functions, and their corresponding rederived velocity fields.

For both fields it turns out that the optimal paths are almost radial. This is not very surprising since the smoothing uses such a wide window function that variations in galaxy number density have little effect on causing the optimal path to deviate from the radial. By insisting on arcweights that are heavily dependent on density, and for which $k \sim 1$ one can achieve significantly non-radial paths. However, in these cases the recovered potential velocity field is noisy and bears little resemblance to the input field.

Nevertheless, the method does offer some hope. Because k is, in general, lower in high number density regions and close to the origin – areas which are given higher weighting by the procedure anyway – a value of $k \sim 4$ will be reasonable and should produce some improvement in the low density areas without affecting the regions with an already adequate recovery. In this way, although there may be very little change, one can be confident that any changes are for the better.
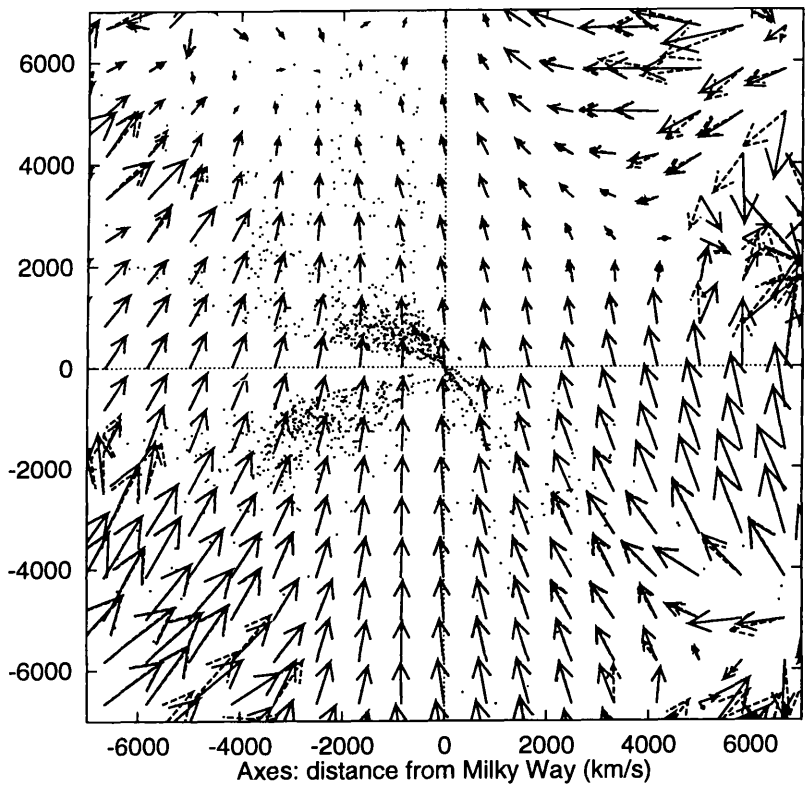
Figure 5.14: Potential velocities derived from optimal paths obtained for different density weightings (α) and ratios of transverse to radial errors (k). The solid arrows are the velocities for $\alpha = 0$ and $k = 9$. Dashed arrows for $\alpha = -4$ and $k = 5$. Galaxy number density is projected onto the plane.

5.4   Comparison of Density Field Recoveries

So far, I have been considering only the recovery of velocity fields using POTENT. However, one of the main aims of the method is to produce a map of the density field (see sec. 4.1.4). It might be reasonable to assume that methods that produce good (unbiased) velocity fields will also lead to good density recoveries, but this is not necessarily the case and it is important to test the assumption and, if necessary, develop alternative procedures.

I wish to develop a test that examines the different methods without interference from the question of validity of the various assumptions inherent in the POTENT procedure. In particular, I want to be sure that assumptions of a potential velocity field and no significant non-linear density evolution are always valid. I therefore start from a velocity potential and derive the velocity field to impose directly from this. The density field can then be calculated using the quasi-linear scheme of Nusser et al. (1991) that is adopted by POTENT and this can be compared directly to the results obtained by the various methods.

The final necessity is a selection of galaxies to appear in a mock catalogue. I am going to be performing Monte Carlo tests, and as before, I wish to use a set of observable galaxies as a basis with the distances scattered for each iteration of the Monte Carlo procedure. However, I now need to produce an underlying distribution of galaxies that is consistent with the density field of the "universe". This is done as follows

1. The cumulative density field, ρ_{cum}, is calculated for a large volume centred on the origin and scaled such that the maximum cumulative density is 1 and the minimum 0.

2. 'Galaxies' are dropped uniformly and randomly through out the volume and each assigned a random value τ between 0 and 1. Only those with $\tau(\vec{r}_i) \leq \rho_{\text{cum}}(\vec{r}_i)$ are retained, thus giving an underlying distribution of galaxies consistent with the density field.

3. Finally, the catalogue of galaxies is created by assigning an **M** and **P** to each galaxy and imposing selection.

In this way, the real positions of a catalogue of galaxies can be produced consistent with the underlying density field. Note that the catalogue will be *considerably* less clustered than an actual survey because of the non-linear evolution that has led to the galaxies in the real

universe. Nevertheless, it should be sufficient to distinguish between the various methods given a suitable varied velocity potential.

5.4.1 The Test Model

I choose a simple but quite demanding model field. The potential consists of two large gaussian fluctuations – one at $(-4000, 0, 0)$ km s^{-1} producing infall and a larger 'void' at $(5000, -2000, 0)$ km s^{-1}. The velocity and density fields produced by this potential are shown in fig. 5.21 as three perpendicular slices through the origin ($X = 0$, $Y = 0$ and $Z = 0$). As before, this figure shows not the underlying fields but the field smoothed on the same scale as the POTENT recoveries (ie 1200 km s^{-1}). However, the density field is not derived from the unsmoothed velocity field and then smoothed, but derived directly from the smoothed velocities. This is analogous to the POTENT procedure since the smoothing there is also over velocities and, as shown in Rauzy et al. (1994b), it is important to compare with the correct version as they can be significantly different[1].

5.4.2 Test Results

The Monte Carlo procedure is that same as that used before and described in sec. 4.2 with the luminosity parameters and imposed velocity field, obviously, being those used to create the galaxy sample. Before looking at the density recoveries, I shall examine the velocity fields produced by a variety of methods to check that the conclusions of the earlier tests are still valid.

Figure 5.15 shows the bias of the recoveries in radial shells for 'raw' ITF distances and homogeneously and inhomogeneously Malmquist corrected distances. Each data-point is obtained by averaging the *biases* of all the grid points that fall into a particular spherical shell. The error bars, therefore, are not a measure of the *random* errors, but an indication of the variation of *systematic* errors within each shell.

Clearly, the HMC is inappropriate, particularly at large distances as expected. However, there is very little to choose between the 'raw' and inhomogeneously corrected recoveries

---

[1]This problem will be particularly acute when comparing kinematically derived density maps to those based on smoothed number densities such as the POTENT90 analysis of Dekel et al. (1993). However, in my simple case, consistency is all that is required.
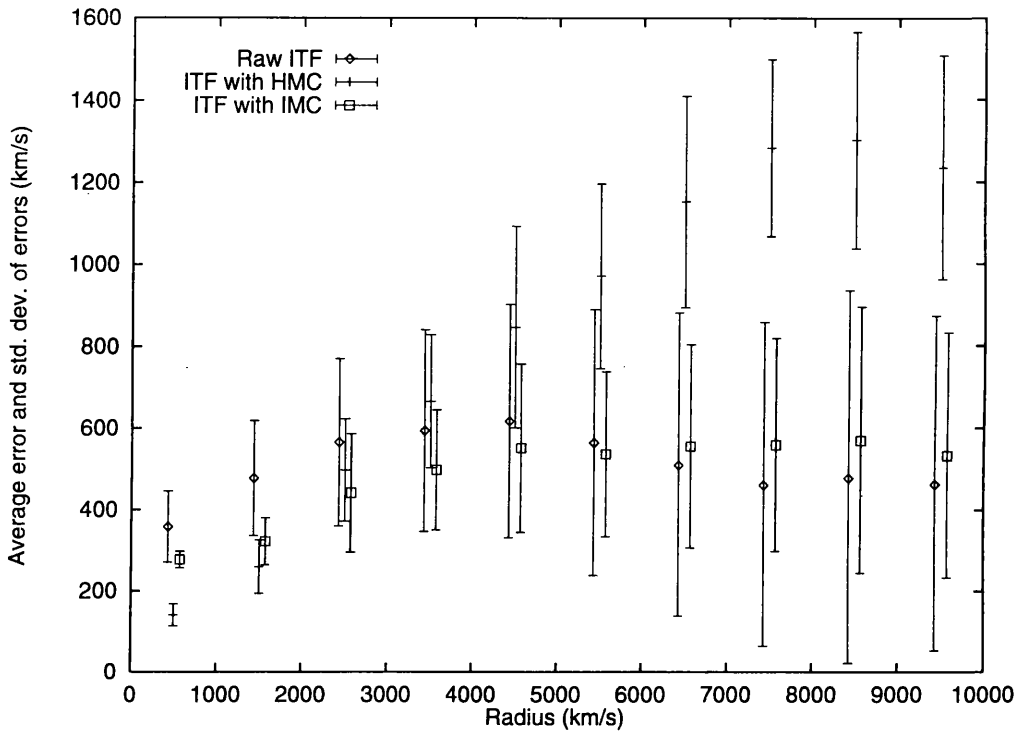
Figure 5.15: The biases in the velocity field recovery for ITF estimates with no correction, HMC and IMC. The biases are found from the Monte Carlo average and binned into radial shells. Note that the error bars do not indicate the size of the random errors of the recoveries, but one standard deviation of the *biases* within the shell.

– only the reduced range of biases makes the corrected distances more appropriate and neither recovery would seem to be particularly good, at least from this simple analysis.

One possible correction that I have not considered so far is a Malmquist correction based on some independent determination of the number density of galaxies (for example from an IRAS survey). Of course, since the selection function of such a survey will be very different from the catalogue of distances estimates, and perhaps neither of them very well known, one would hope to be able to estimate the *underlying* smoothed galaxy density and apply it to DTF-like estimates (see sec. 2.2.3). Such a procedure has a number of advantages – the increased number of galaxies and much deeper surveying of redshift-only catalogues will reduce the noise on the radial distribution function, and non-radial inhomogeneities can be included – but there are other problems since the density field is not trivial to determine and the population of objects may well be drawn from a significantly different distribution
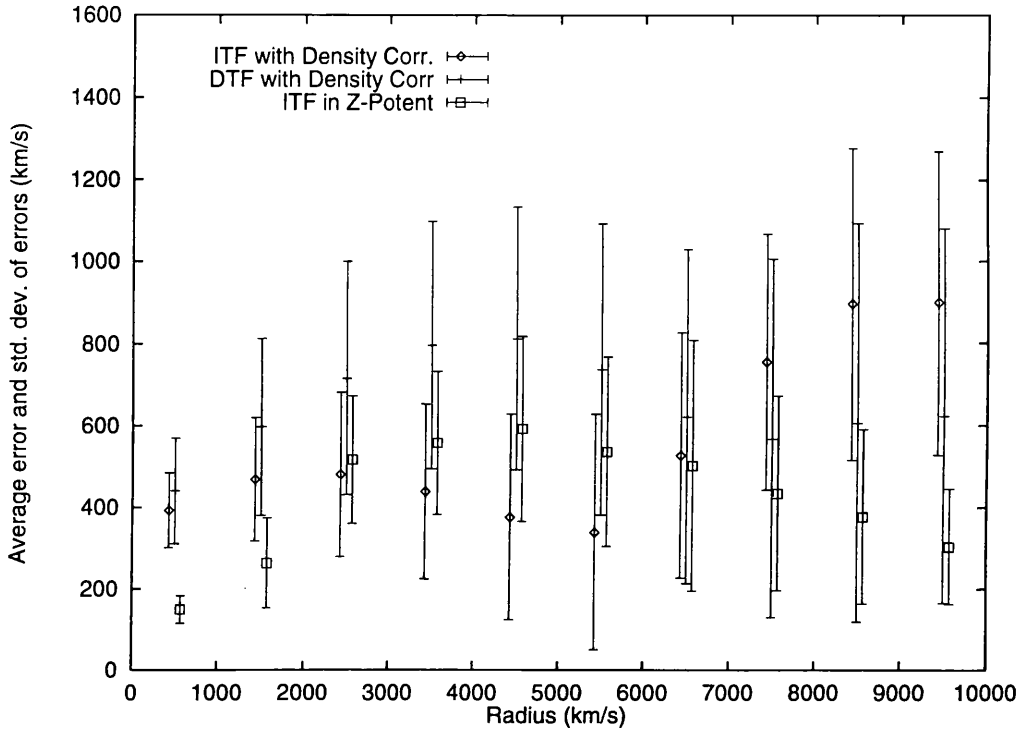
Figure 5.16: As fig. 5.15 but for Z-POTENT and Malmquist corrections based on the true underlying density.

(eg IRAS galaxies are, naturally, infra-red selected). I will test the method here by using the actual smoothed density field, thereby bypassing these problems and getting to the heart of the method.

The results for this are shown in fig. 5.16. For comparison, the density Malmquist corrections have been applied to DTF and (inappropriately) ITF estimates. Surprisingly, the results are not that good. Again, there is little to choose between appropriate and inappropriate corrections and the Z-POTENT recovery also shown on the graph is in many ways preferable to both. (Although this is partly because the areas towards the centres of the void and 'attractor' regions are discarded from the recovery since $d\hat{v}_r/dz > 1$ – see the maps in sec. 5.4.4). This poor correction will be partly connected with the fact that I am using the mass density rather than the galaxy number density, but because of the absence of non-linear evolution in the model and the fact that no cosmological biasing has been introduced, this is not enough to explain all of the problem. As significant is the fact that many of the assumptions behind the applicability of Malmquist corrections are being stretched. Errors
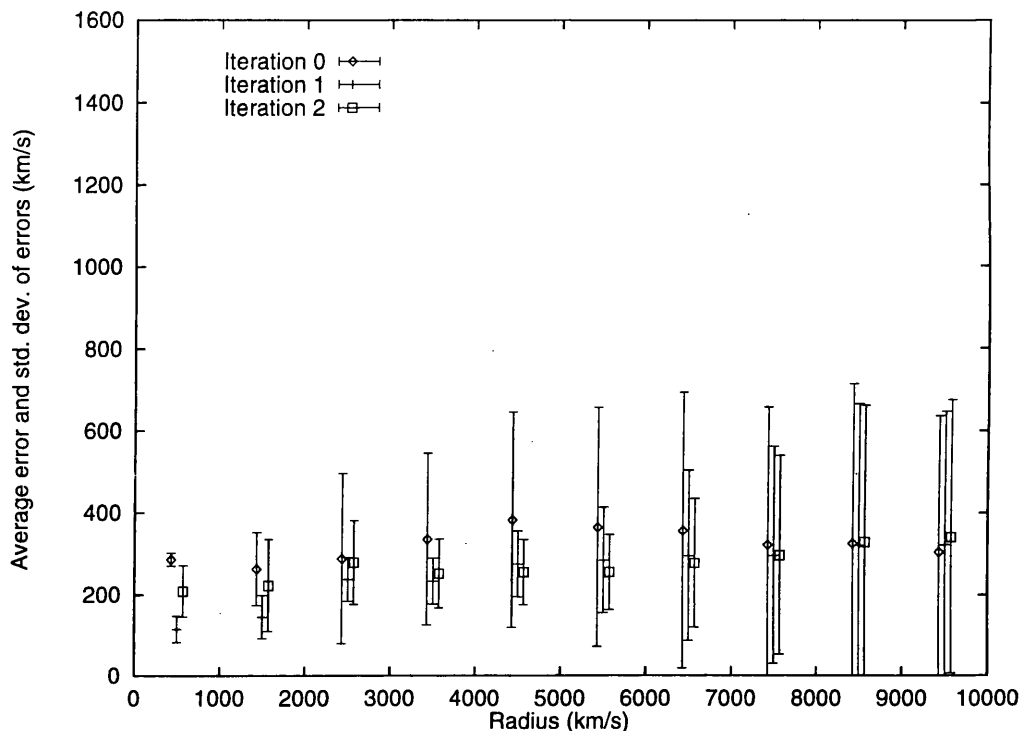
Figure 5.17: As fig. 5.15 but for various stages of the iterative correction procedure.

in the distance estimates are *not* small, and the smoothing scale is significant even compared to the entire recovery volume.

Nevertheless, for a more clumped distribution of galaxies with significant non-radial inhomogeneities, the density correction method is likely to be more applicable than a simple IMC, as long as the problems mentioned above are carefully addressed. However, as long as smoothing volumes remain large and errors are significantly above 10%, Malmquist corrections in POTENT of any type must be treated with considerable caution.

The final figure in this section shows the results of the Iterative Monte Carlo correction procedure (5.17). The results are shown at three stages. After iteration 0 (when quiet Hubble flow is imposed), no sampling gradient bias has been addressed, but a correction has been found for Malmquist-like biases. Even at this stage, the improvement over the previous methods is considerable. Iteration 1, the first with the actual redshifts used to define the velocity field shows a slight improvement again and it is at this stage that the convergence tolerance of 200 km s^{-1} was reached. However, for demonstration purposes,
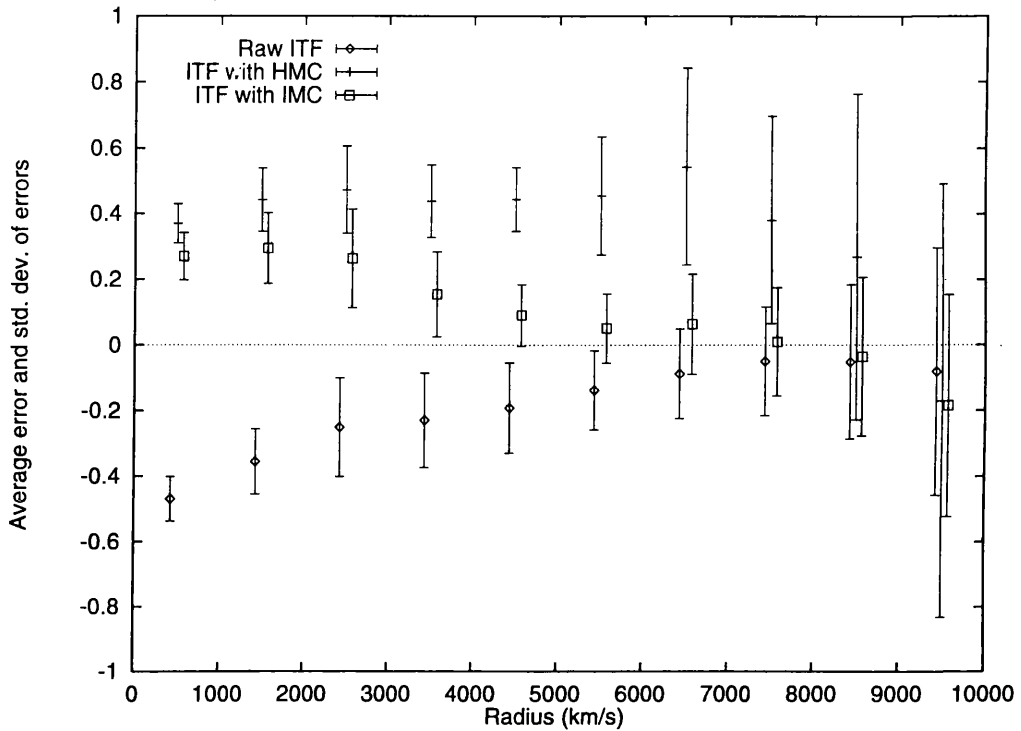
Figure 5.18: The biases in the density field recovery for ITF estimates with no correction, HMC and IMC as in fig. 5.18.

the process was continued for another iteration. There is very little distinction between the three stages implying that Malmquist-like effects account for the majority of the systematic errors. This is not surprising given the relatively homogeneous distribution of galaxies which will give little cause for sampling gradient bias, but even the little there is, is tackled in iteration 1. In iteration 2, one might expect the procedure to over-correct in some regions, but again because of the small sampling gradient biases, this is not a large effect.

5.4.3 Test Density Recoveries

The velocity recoveries, therefore, are consistent with the results already obtained with IMC applied to ITF a marginal improvement over the other simple corrections, but the iterative process providing by far the best results. Is the same true of the reconstructed density fields?

Figures 5.18, 5.19 and 5.20 show the biases for the density recoveries in the same form as
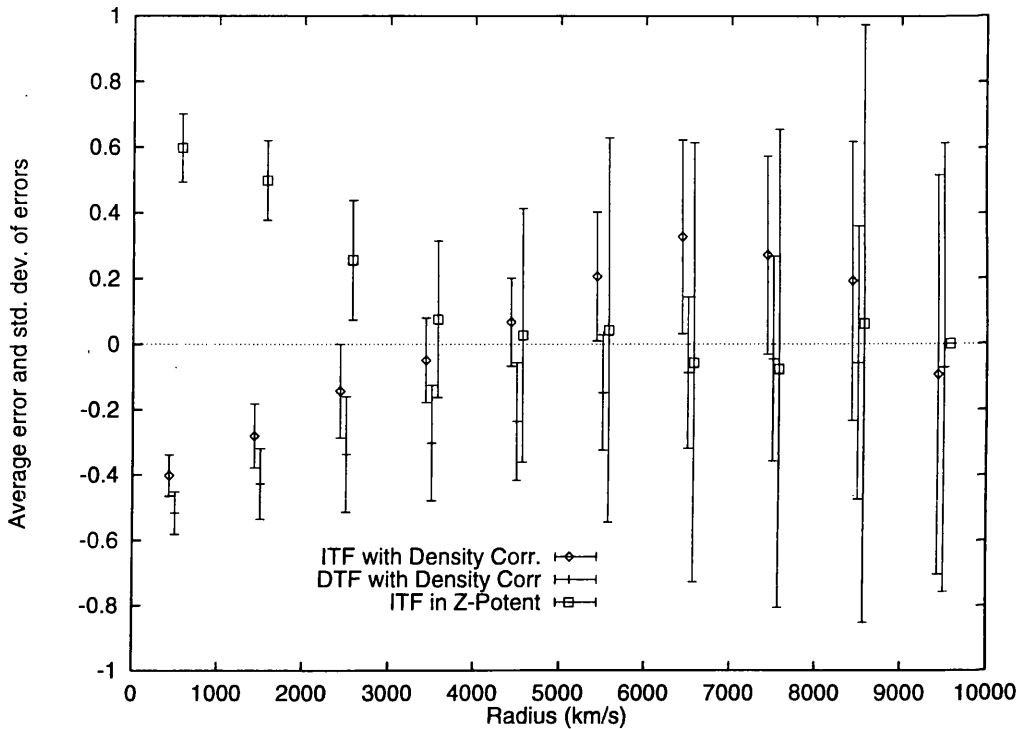
Figure 5.19: As fig. 5.16 but for the density recovery.

the velocity biases. Again the inhomogeneously corrected ITF estimates give marginally the best recovery of the 'simple' corrections, although there is a large bias near the origin. The density model based Malmquist corrections are poor although they now show some improvement over the Z-POTENT recovery – again because of the large invalid areas in the resultant Z-POTENT fields. Finally, the iteratively corrected recoveries are excellent with the biases kept small over the entire range and the majority of the correction again being done in the first iteration.

So, the assumption of a 'good' velocity field recovery giving a 'good' density field appears to be justified. However, before continuing on to look at the problems associated with real data in the next chapter, it is interesting to look at some of the maps of the velocity and density fields recovered during these tests.

5.4.4 Recovered Velocity and Density Maps

As well as the statistical properties of the fields, we want POTENT to provide accurate

Figure 5.20: As fig. 5.17 but for the density recovery.
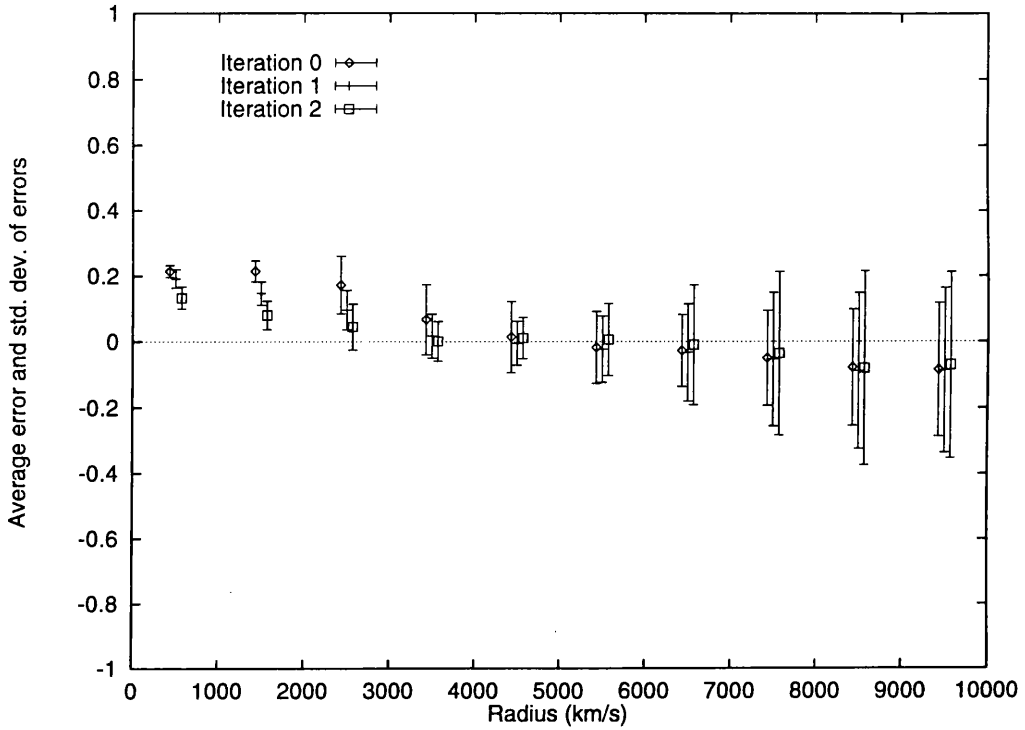
maps of the local velocity and density structure. Three slices though such a map of the imposed model are shown in fig. 5.21. Recall that this is, in fact, the model with the velocities smoothed on the same scale as the POTENT recoveries. The void and attractor regions can be clearly see and since the top slice is $Z = 0$, it includes the centres of both features. The first two sets of recoveries (remember that these are Monte Carlo averages) in figures 5.22 and 5.23 show the equivalent slices for the 'raw' ITF estimates and the inhomogeneously corrected ones respectively. The universal radial outflow from the origin in the ITF recovery is clear producing the origin centred void in the $X = 0$ density field and shifting and distorting the genuine features. The IMC corrected recovery is, however, better with the centres of the void and attractor well determined and only small spurious velocities in the $X = 0$ slice. However, away from the features, there are still considerable flows and the density fluctuations of the void and attractor lack the tight, compact form of the model. This is characteristic of sampling gradient bias as well as uncorrected Malmquist-like biases. However, neither recovery is unrecognisable!
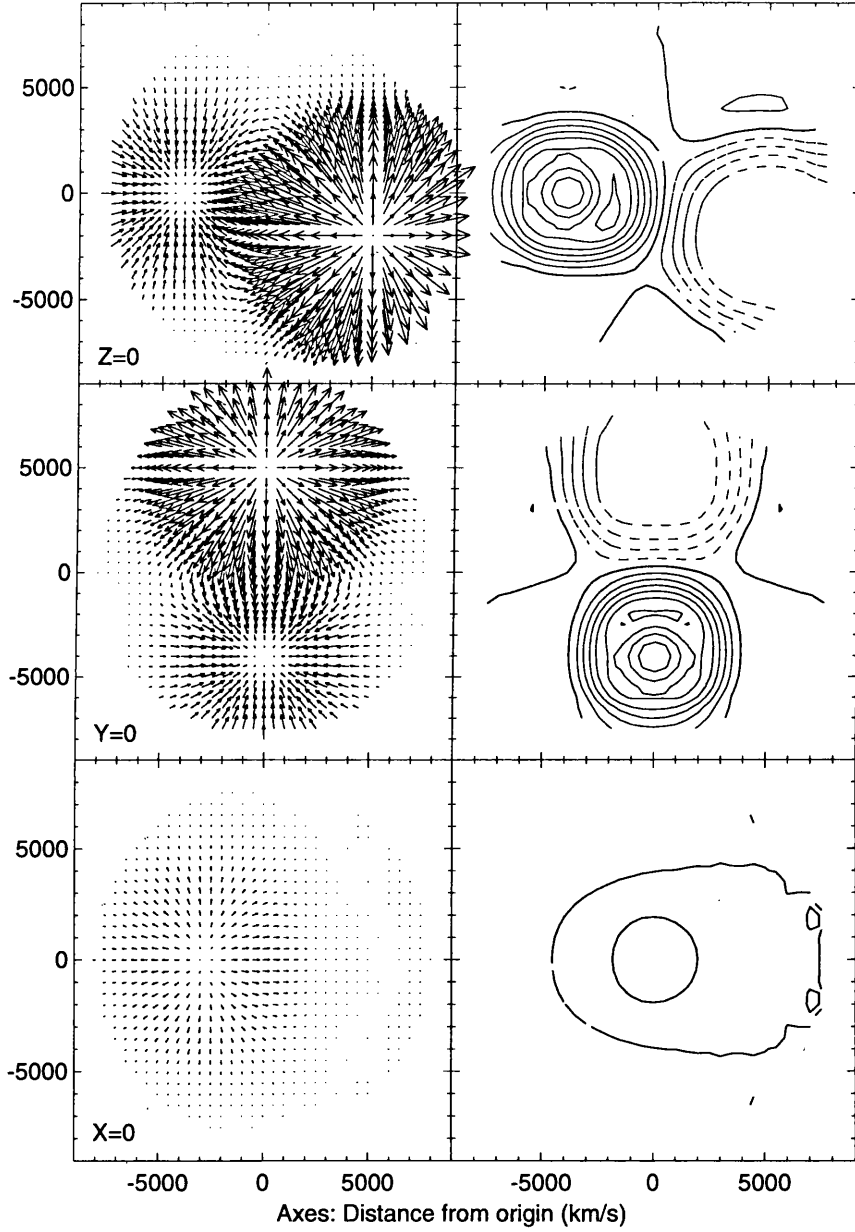
Figure 5.21: The smoothed velocity and density fields on three perpendicular slices through the model for density reconstruction tests. The velocity field has been smoothed by a gaussian window of radius 1200 $km\,s^{-1}$ and the density derived from this using the quasi-linear method of Nusser et al. (1991). Negative density contours are dashed, positive solid and the $\delta = 0$ contour thicker. Contour spacing is $\delta = 0.2$.
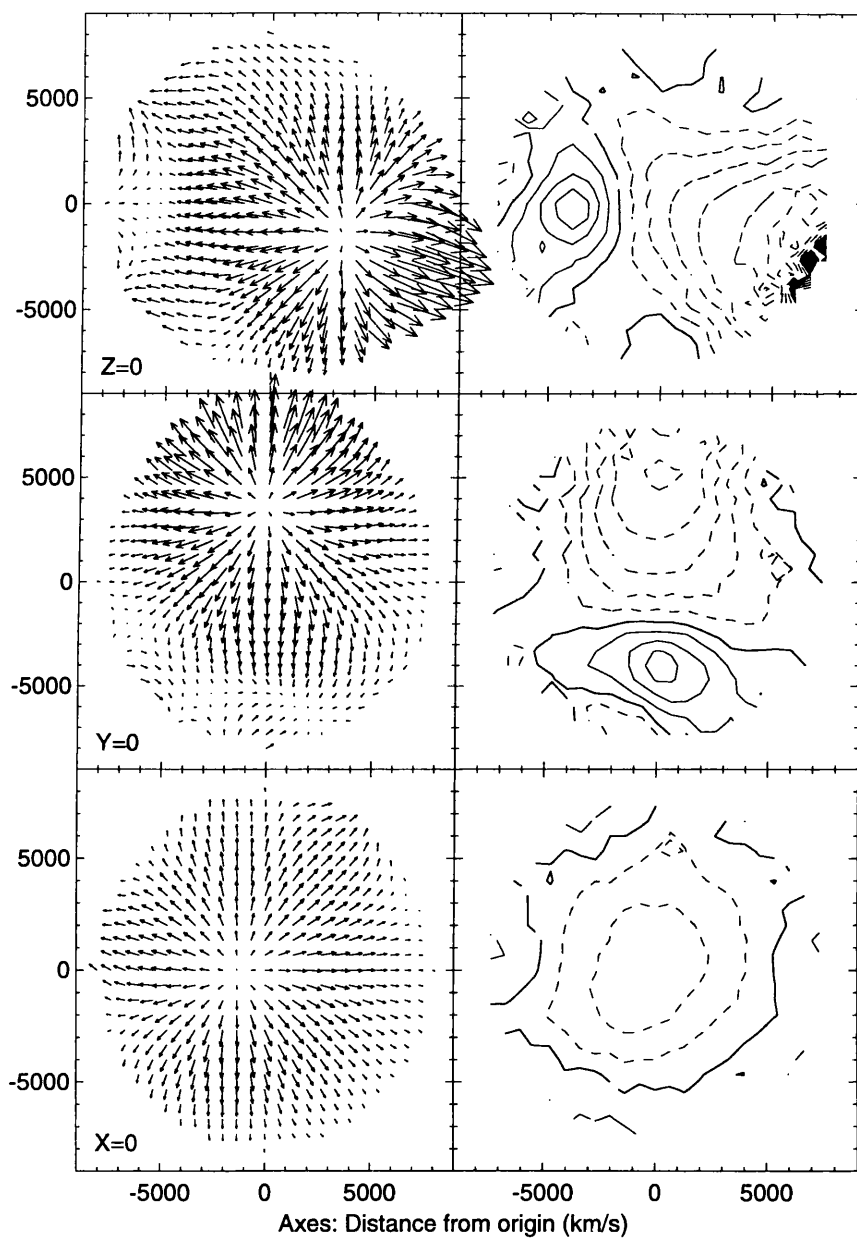
Figure 5.22: The recovery of the model field in fig. 5.21 using uncorrected ITF distance estimates.
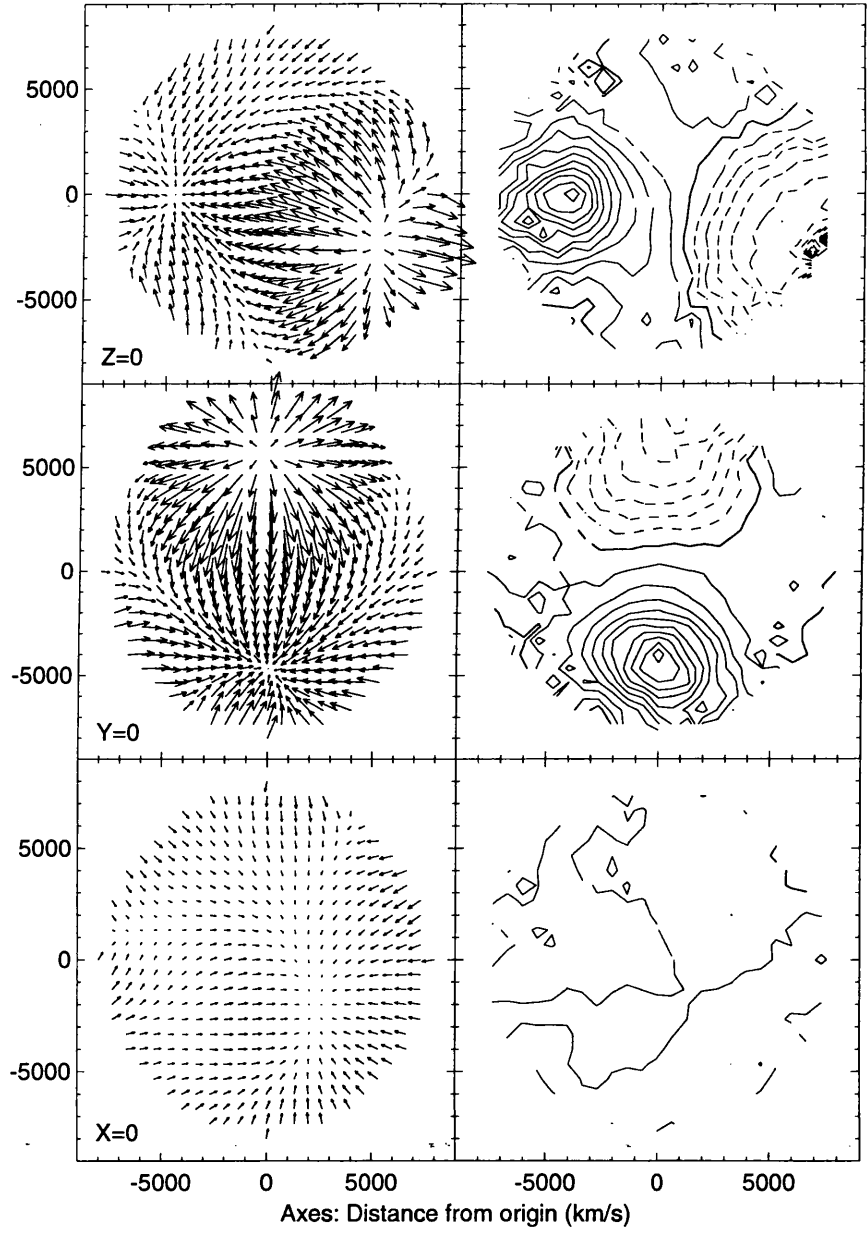
Figure 5.23: The recovery of the model field in fig. 5.21 using ITF distance estimates with an IMC applied.
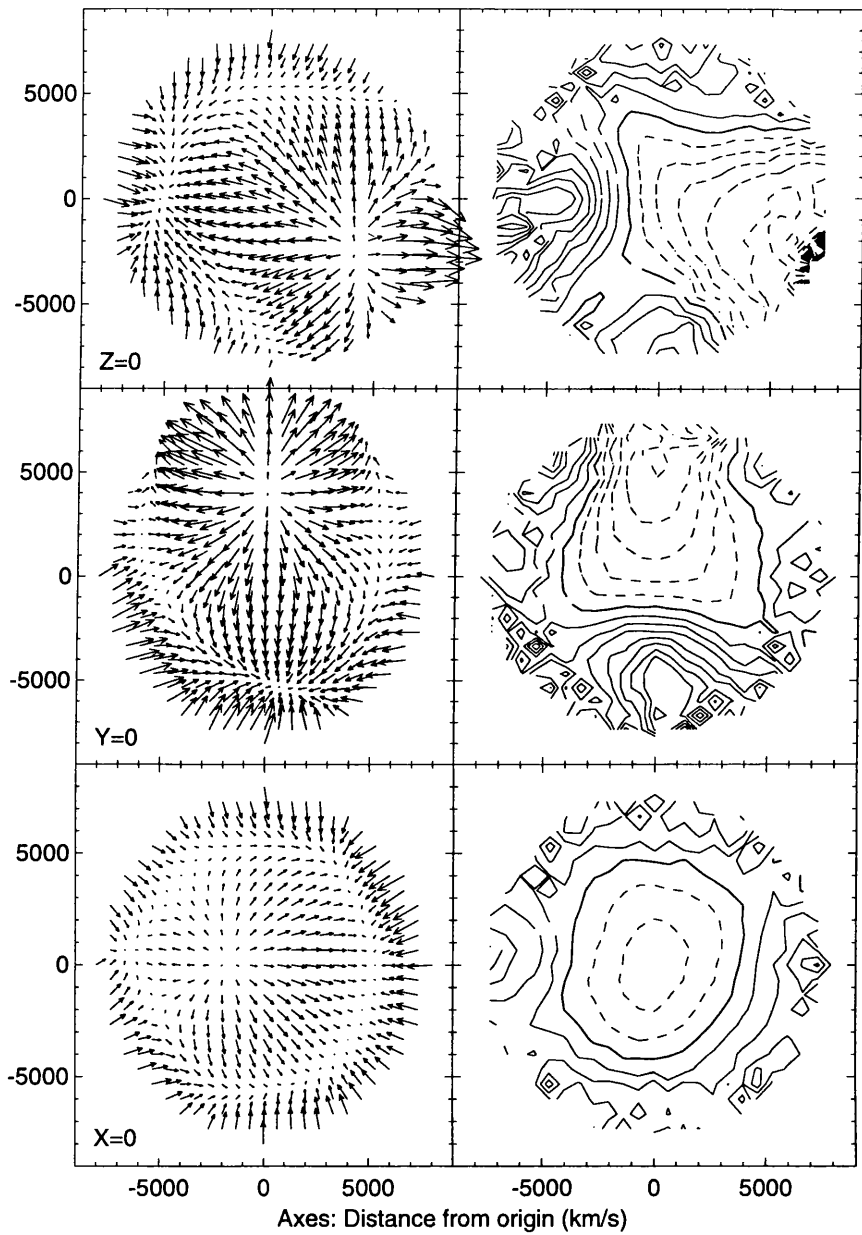
Figure 5.24: The recovery of the model field in fig. 5.21 using ITF distance estimates with a Malmquist correction based on the true density field.

Figure 5.25: The recovery of the model field in fig. 5.21 using DTF distance estimates with a Malmquist correction based on the true density field.

By contrast, fig. 5.24 has considerable distortion. These are the maps for ITF estimates corrected according to the underlying density and the inapplicability of such a correction to ITF estimates is clear. The correction is better when applied to DTF estimates (fig. 5.25), but there are still major problems with the shape of the features and the radial distortion.
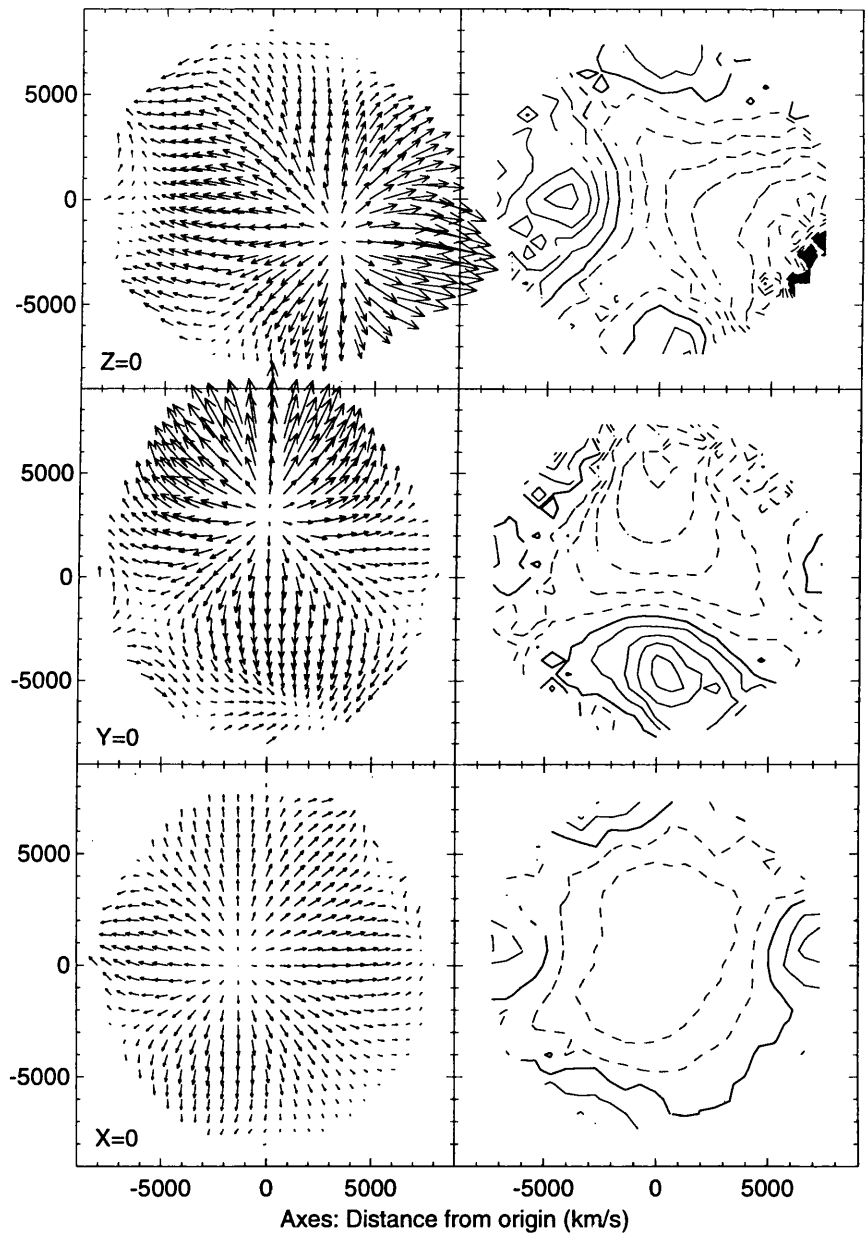
Figure 5.26 is for the Z-POTENT recovery. This is not really ideal for direct comparison to the other recoveries as intricacies in the Monte Carlo method mean that the smoothing process is slightly different, leaving the rather noisy edged density field. However, the problems caused by such large features can easily be seen in the undefined areas towards the centres of the void and attractor. Also, these areas tend to lead to distortion in the remaining field stretching them out in one direction and shifting the centres. Obviously, such large features would not be likely to be found in reality, particularly so close to the origin, but the Z-POTENT method shows some serious shortcomings and should only be used with care and any undefined areas of the recovery carefully considered.

Finally, figures 5.27 and 5.28 show the state of the Monte Carlo corrections after the zeroth (quiet Hubble flow imposed) and first iterations respectively. At the first stage, the maps are already very like the imposed field with very few residual biases and very tight features unlike the IMC recovery. However, the height of the density peak falls a little short and the void is not as deep as it should be. This is corrected by the next iteration which is strikingly like the original.

I hope I have now made it clear in this chapter that errors, and particularly systematic errors in POTENT-like techniques, are a serious but not insurmountable problem. Obviously some technique like the iterative Monte Carlo corrections is extremely effective and when added to the slight improvement in random errors that the Max-flow procedures can bring should provide an extremely effective tool. However, the computational intensity of the method is a drawback, and for simpler analyses, a carefully applied Malmquist correction, perhaps in comparison with a Z-POTENT recovery, can also yield useful results. The important point here is to understand exactly when the corrections are applicable and how they should be applied.

The final stage of this thesis, therefore, is to apply these techniques to genuine galaxy observations and so, hopefully, draw some conclusions about the distribution of matter in
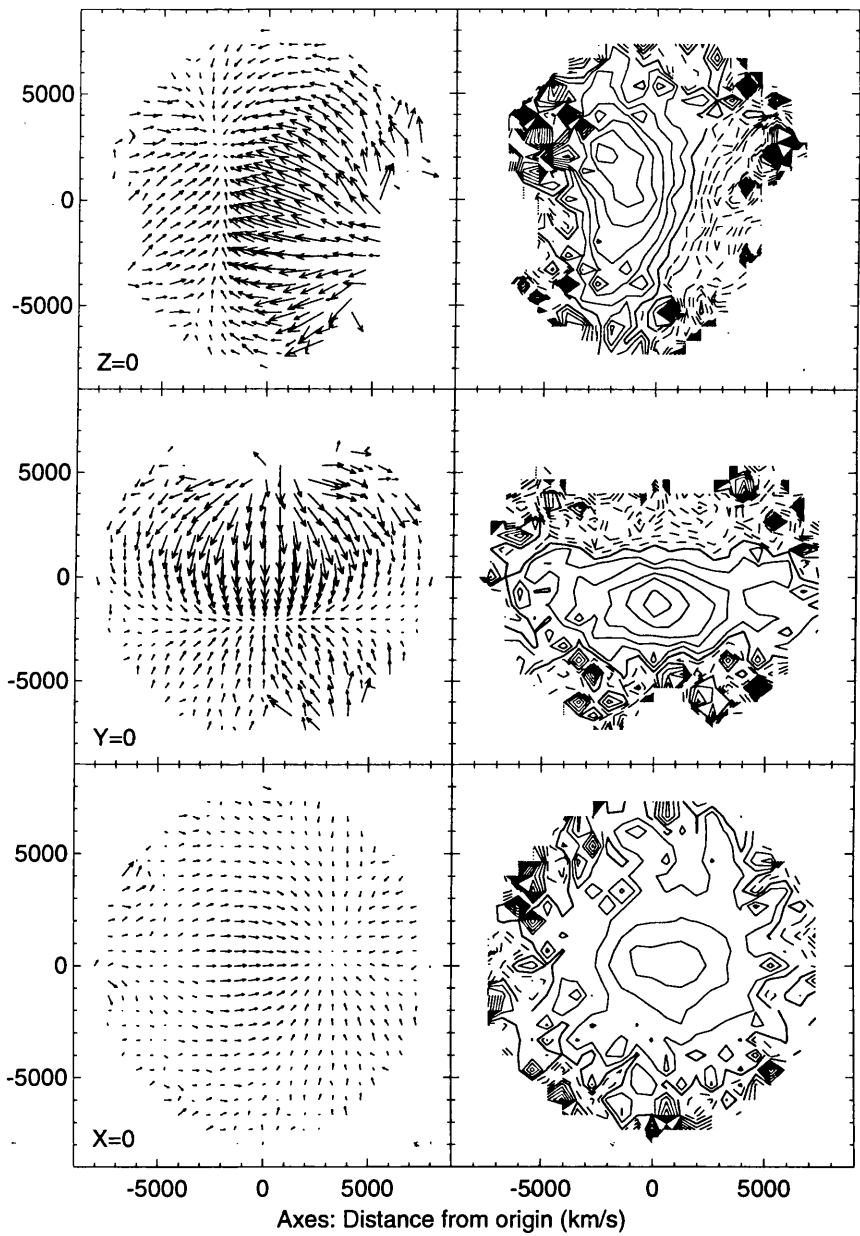
Figure 5.26: The recovery of the model field in fig. 5.21 using the Z-POTENT method applied to ITF distance estimates.
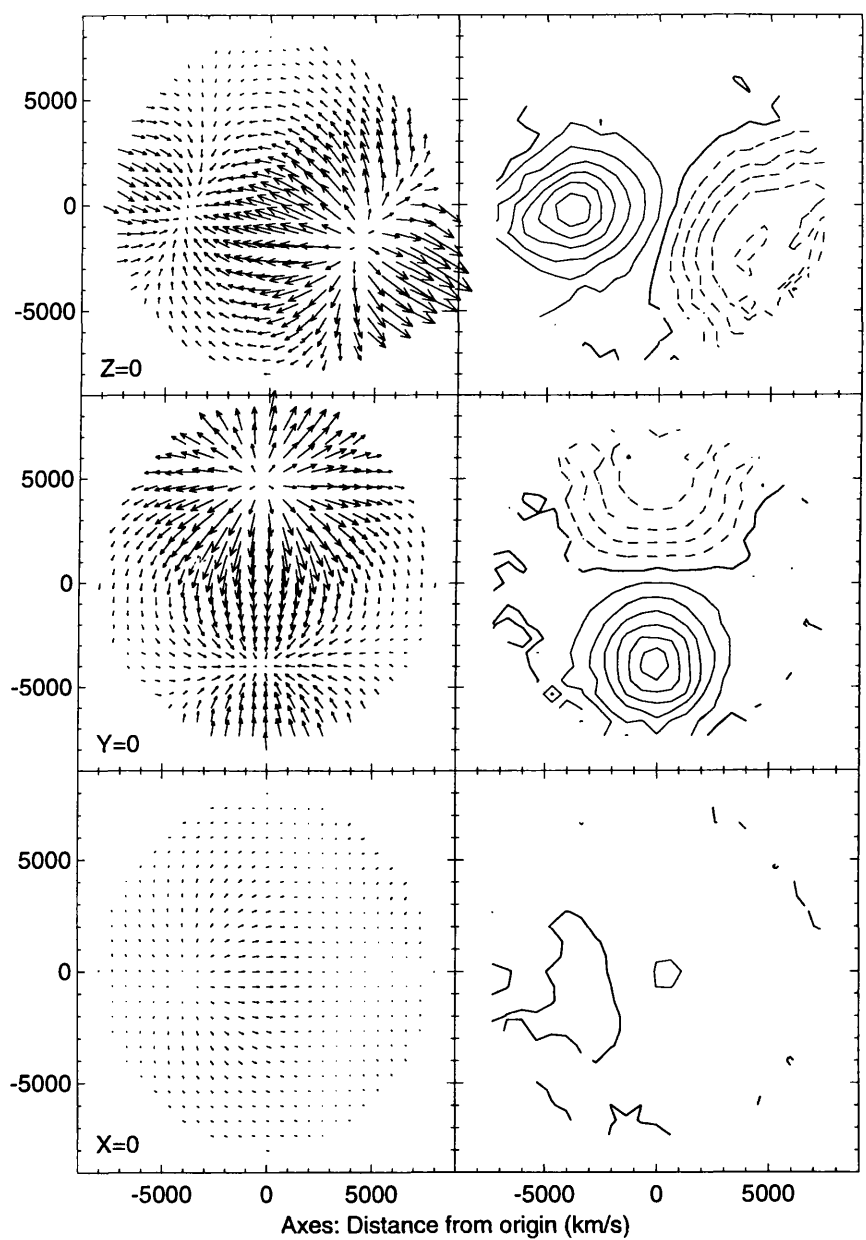
Figure 5.27: The recovery of the model field after the zeroth iteration of the Monte Carlo correction procedure..
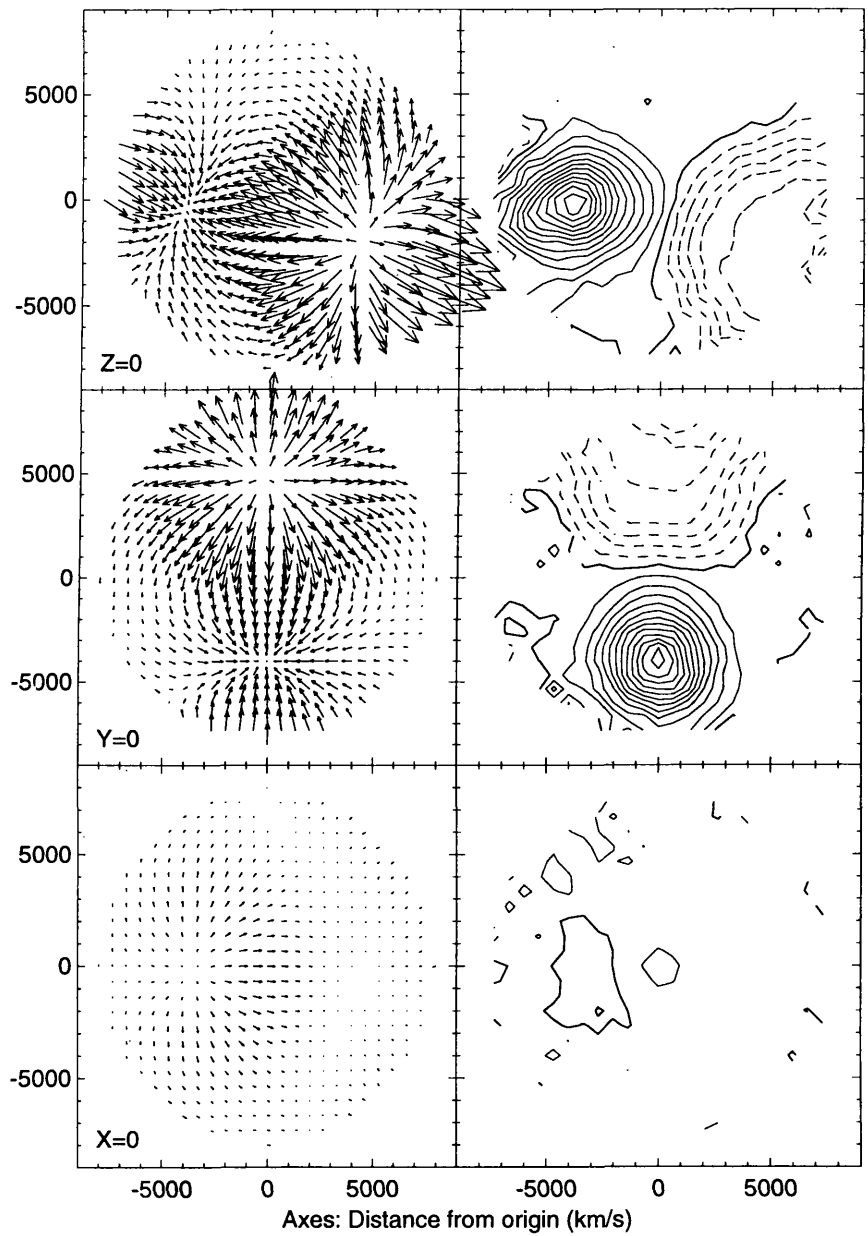
Figure 5.28: The recovery of the model field after the first iteration of the Monte Carlo correction procedure..

the local universe. Before this can be done, however, some thought needs to be given to the treatment and calibration of data.

Chapter 6

Application to Real Data

The charm dissolves apace,

And as the morning steals upon the night,

Melting the darkness, so their rising senses

Begin to chase the ignorant fumes that mantle

Their clearer reason.

The Tempest V:i

In this chapter I will be applying the techniques described in the previous chapters to a selection of galaxy catalogues. The catalogues (provided by Mathewson et al. (1992), Burstein (1990) and Salucci (1994)) are calibrated using the procedures given in chapter 3 and carefully analysed. At this stage, a number of galaxies are removed because of possible errors. Then, with this new combination of data, the POTENT-like techniques are applied and the results compared, together with several further tests to check the sensitivity of the results to the data.

It is not the aim of this thesis to perform a detailed analysis of the results, (for example by comparison to an IRAS based density recovery such as in Dekel et al. (1993)) but to demonstrate the applicability of the actual methods. However, a very simple technique (Dekel and Rees, 1993) is employed in sect. 6.5 to at least get a good lower limit on Ω_0 and demonstrate the possibilities that accurate recoveries can offer.

6.1 Comparison to Potent90

However, first I would like to perform one final test of my POTENT implementation to check that it is comparable to that of Bertschinger and Dekel (1989) and Dekel et al. (1993). I have done this by applying directly the data used in the POTENT90 analysis of Dekel et al. (1993). This consists of nearly 1000 galaxies forming about 500 objects. They use DTF distance estimates and attempt to correct for Malmquist bias by applying an homogeneous Malmquist correction.

Of course, we have seen that such an approach is far from ideal, but by following their approach as closely as possible, I can ensure that the POTENT method as I have implemented it is comparable to this orthodox one and be confident that any differences produced by applying more accurate methods are a result of those improvements and not of a fundamental difference in the underlying recovery implementation.

Figure 6.1 shows the results of this final test. The original POTENT90 density field slices are given in fig. 6.2 for comparison In general the comparison is very good. The slice through the Super-galactic plane (Z=0) shows the long density ramp up towards the "Great Attractor" region and the 'North' and 'South' density peaks of the Y slice are accurately reproduced. However, there is a significant discrepancy showing up as a large positive density contrast at around (0, −5000, 0). This structure is not present at all in the POTENT90 maps and is perhaps a cause for some concern.

Of course, since it is right on the edge of the survey where the errors are large, and far stronger than any other density feature, it should not be taken seriously as a genuine feature of the universe. However, it is disconcerting that the same method applied to the same data-set can produce such a difference. Nevertheless, since the rest of the recovery is so similar, it is unlikely that there is a fundamental problem with the implementation of POTENT. What is more likely is that an error has crept into the data at some stage. To test this hypothesis, I simply repeat the POTENT procedure several times but with one of the galaxies in the vicinity of the spurious feature randomly removed (the 'vicinity' can be safely described as within ~ 2 window radii of the centre of the feature).

In fact this part of the survey – although not among the very sparse regions as defined in the POTENT90 comparison – contains only a few galaxies and of these only the two closest
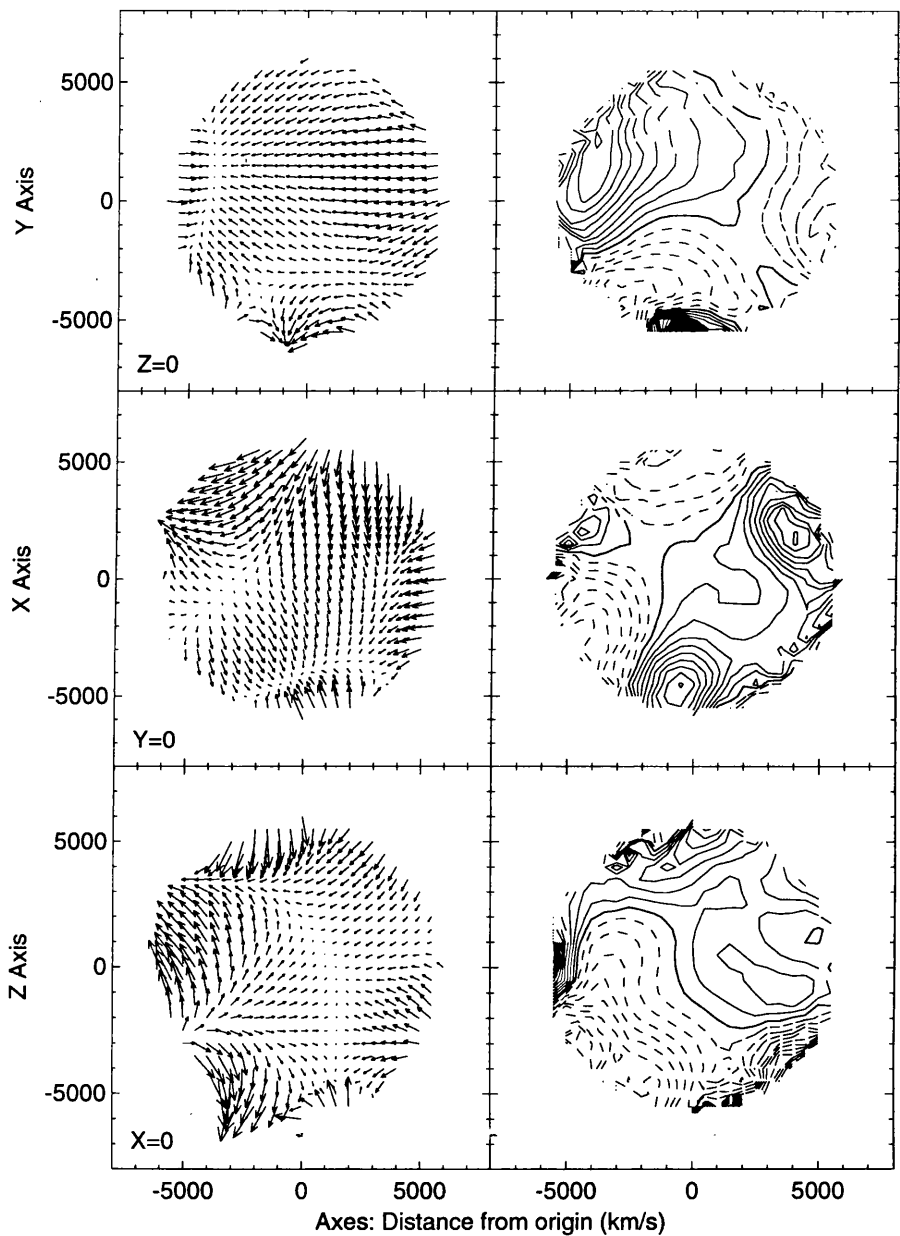
Figure 6.1: The orthodox POTENT recovery using the POTENT90 data-set. Three slices through the velocity and density fields are shown for direct comparison to the 'P' panels of figures 2a–c in Dekel et al. (1993) (reproduced in fig. 6.2).
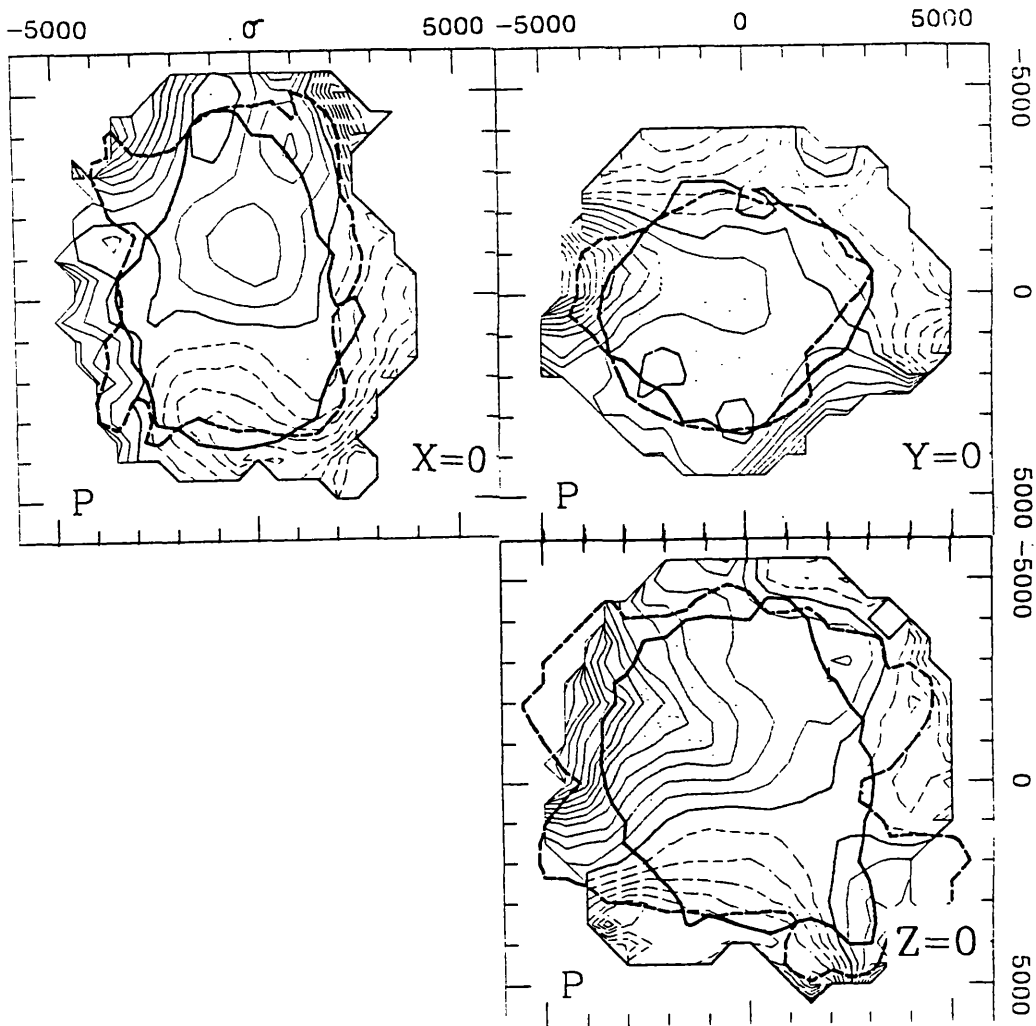
Figure 6.2: The POTENT90 density recoveries for the three slices given in fig. 6.1. These are reproduced from the 'P' panels of figures 2a–c in Dekel et al. (1993). (Note that the $Y = 0$ panel is oriented differently and viewed from the opposite side than the equivalent slice in fig. 6.1)
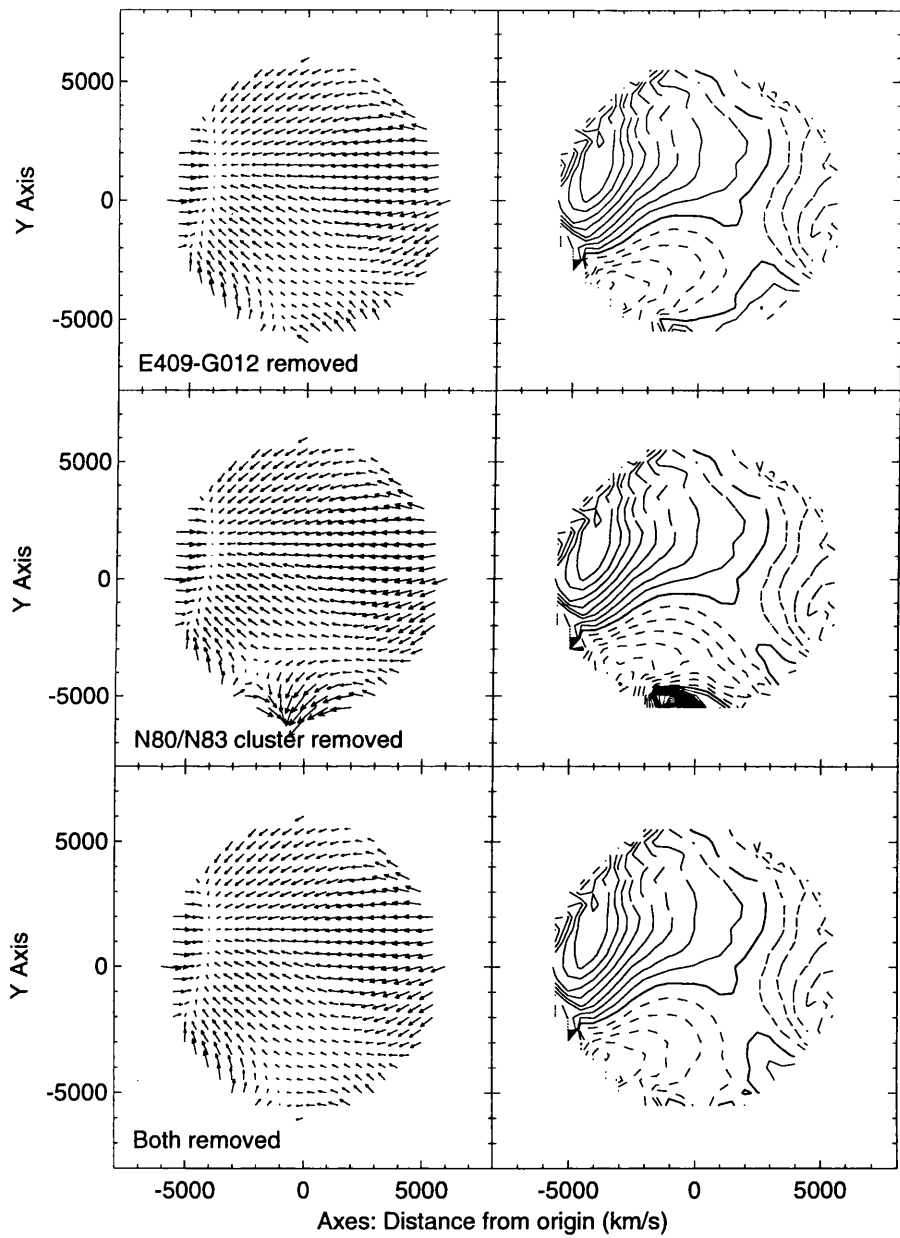
Figure 6.3: The Super-galactic plane slices for three POTENT recoveries with the POTENT90 data slightly modified. At the top, galaxy E409-G012 has been removed from the survey. In the middle, a cluster object (containing just 2 galaxies – N80 and N83) has been removed and at the bottom, both objects are deleted from the sample.

make any noticeable difference, as might be expected. The 'culprit' can be clearly seen in fig. 6.3. When E409-G012 is removed, there is a considerable improvement. The same is not true for the two member cluster (N80/N83) in spite of the extra weight given to it in POTENT and the fact that its distance from the anomaly is comparable.

The danger of sparsely sample areas is very clear from this. The addition of just one galaxy can make a considerable difference and the errors will not always be as easy to spot as was the case here. It is, therefore, very important to ensure that the recovery in any such area is safe before using it in any analysis. This can be done using simple tests such as the selective removal of galaxies shown here and if the recoveries do not show any particular sensitivity to any one galaxy then analysis can continue with much more confidence[1].

Apart from this reservation, the comparison is certainly successful. Given the discrepancies that are bound to occur between two implementations, particularly in the numerical details of the smoothing and interpolation procedures, it would be unreasonable to expect an exact match, but the size and position of features are very similar and it is be assumed that the methods are essentially the same.

6.2 The Galaxy Catalogues

Since the Dekel et al. (1993) POTENT work, there has been a considerable improvement in both quantity and, to a certain extent, quality of galaxy observations suitable for distance estimation. In particular, the large catalogue produced by Mathewson et al. (1992) has provided significantly greater coverage over a large section of the sky. I will be using these new data in addition to a re-calibrated version of the Burstein (1990) data used in the POTENT90 comparison above.

The Mathewson et al. (1992) data contains information on a total of 1354 spiral galaxies, including 20 clusters. Although a considerable quantity of data for each galaxy is available, I will only be using the redshift (corrected to the CMBR rest frame) and the corrected I-band magnitudes together with the line width factor (corresponding to to maximum rotation

---

[1]Of course, it is entirely possible that it is the interaction between two or more galaxies that produces an error, but even in this case, the removal of one of them should indicate that there is a problem. However, exactly *which* galaxy would be more difficult to determine and there is no way of knowing if the unmodified result is, in fact, any worse! The only safe solution is to discard sensitive areas from any analysis completely.
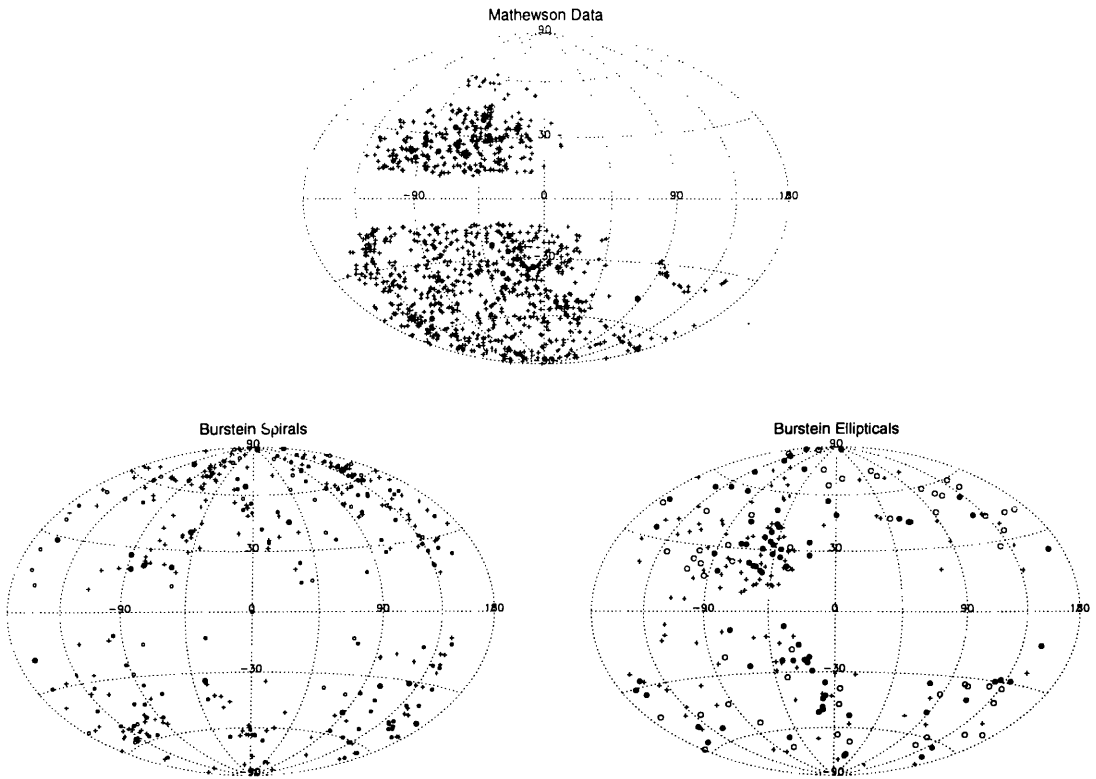
Figure 6.4: Aitoff projections of the positions of galaxies in the various catalogues in galactic coordinates. For the Mathewson et al. (1992) data set, the filled circles denote clusters and the crosses, field galaxies, Note that coverage is roughly half the sky, not including the "Zone of Avoidance" of the galactic plane. The Burstein spiral galaxies are more evenly distributed. Again, filled circles are clusters. The crosses are the "good" field galaxies and the smaller, open circles, the "fair" galaxies. Finally, the Burstein ellipticals are divided into clusters (filled circles), clusters with only one observed member (open circles) and field galaxies (crosses).

velocity) to define a Tully-Fisher relation.

The Burstein (1990) data is the Burstein Mark II combined catalogue (see sect. 4.2 for a full list of the sources). It contains a mixture of elliptical and spiral galaxies with a number of clusters. The ellipticals will have their distances estimated using a D_n-σ relation and the spirals all have H band magnitudes and line widths given by half of the full HI profile velocity width – again suitable for a TF relation. The exception to this is the de Vaucouleurs and Peters (1984) spiral galaxy catalogue. These are provided with blue-band TF distance moduli with an estimated error of 26%. This is somewhat larger than even the ellipticals with D_n-σ estimates and I will not be using these data.

As can be seen in fig. 6.4, the overall sky coverage of the three surveys is very good and there are a significant number of clusters which all help to pull down the errors in POTENT because of the added certainty in their distance estimates. However, the Mathewson galaxies, covering as they do only about half the sky, may give rise to some significant sampling gradient biases. Nevertheless, they offer a considerable improvement over the rather sparse Burstein data and should enable some interesting maps to be produced.

However, not all of this data are entirely suitable. I have already chosen not to include the de Vaucouleurs and Peters (1984) galaxies and the "fair" spirals in the Burstein catalogue are known to be potentially unreliable and cannot, therefore, be used for something as sensitive as POTENT. In addition, there are possible problems with the use of field ellipticals with a calibration based on cluster galaxies. Ellipticals appear to be quite sensitive to their environment and therefore, field and cluster galaxies may well form distinct populations. I will return to this problem during the calibration procedure.

Improved Mathewson Galaxies

The line-width (**P**) measure adopted for the Mathewson data set is given by the maximum rotation velocity observed for the galactic disk. However, Salucci (1994) have expressed some doubt about the applicability of this measure, in part because in many cases in the Mathewson catalogue the rotation curves are still rising at the edge of the available data, but also because the physics of the system would imply that other measures might provide better correlation.

The important point is that the rotation velocity at a given radius is related not to the magnitude directly but to the mass contained within that radius. If dark matter is not significant or is distributed in exactly the same way as the luminous matter, this distinction is not important. However, Persic and Salucci (1988) show that there is considerable evidence for a distribution of dark matter that is very different from the luminous matter and forms a large 'halo' with the luminous bulge and disk embedded within it. It is this that gives rise to the fairly flat rotation curves of spirals.

In addition, Persic and Salucci go on to demonstrate that the dark to luminous mass ratio within the disk radius is correlated to the luminousity, thereby showing the physical

significance of the Tully-Fisher relation. With this in mind, Persic and Salucci (1990) and Salucci (1994) suggest an alternative **P** parameter. Instead of using the maximum $\mathbf{P}_{\mathrm{max}}$, they measure it for a fixed fraction of the optical radius of the galaxy, αR_{25}.

Recently, these measurements have been made for a large section of the Mathewson data for $\alpha = \{0.2, 0.4, 0.6, 0.8, 1, 1.2\}$ giving six possible relations (although not all of the galaxies have values for all six radii). Ideally, one would wish to use the luminosity within the radius of the V_{rot} determination, but without this information, there is still going to be an improvement from the use of a consistent **P** parameter even with the overall magnitude, especially if all six of the relations can be used and the resultant distance estimates averaged.

In addition to this re-parameterisation of the line-width, small corrections have been made to a number of the reddening corrections for the apparent magnitudes to remove a slight dependence on inclination angle. Combined, these factors give an improved set of data for about 860 of the Mathewson galaxies (including several in every cluster) and, where possible, I will be using this data.

6.2.1 Calibration of the Catalogues

The calibration technique I have used is the one described in chapter 3 (see sect. 3.1.1) using a composite of several clusters to estimate the distribution parameters. In this section, I will follow the process through for the Mathewson data and also consider the use of the derived parameters in the Monte Carlo correction procedure. As well as the indirect (ITF) calibration, I will also outline the differences in the procedure for direct (DTF) estimates. Although I do not intend to use them for any final maps because of the flaws described earlier, it might be interesting to mimic the POTENT90 procedure with the enlarged catalogue and improved calibration procedure to see how much difference can be seen.

In spite of the reservations given above, it is unlikely that the calibration of the Mathewson data would be improved by the removal of all the galaxies with corrected data provided by Salucci (1994). The significantly reduced number of galaxies would dramatically increase the $1/\sqrt{N}$ noise in both the calibration and the absolute distance determination. I have, therefore, decided to calibrate the entire data set but then replace all those galaxies which appear in the improved data. In this way, although there will be small systematic errors in

the calibration, they should remain far smaller than the random errors and not significantly affect the final maps.

The first stage in the calibration procedure is to estimate σ_M. This is needed for the absolute magnitude corrections when creating the composite cluster for DTF estimates and will also be useful during the Monte Carlo corrections. Because magnitude selection will always tend to decrease $\hat{\sigma}_M$, it is calculated for all clusters with more than 5 observed galaxies and the largest value adopted. In fact, this occurs for cluster 1223-39, a pleasingly nearby one, which has $\hat{\sigma}_M \sim 1.63$.

Using this value and an apparent magnitude limit of 15, the direct, $\mathbf{M}$ on $\mathbf{P}$ regression can be calculated with each cluster corrected for magnitude selection. For both this and the $\mathbf{P}$ on $\mathbf{M}$ calibration, I use all clusters with at least 10 observed members and fix the calibration to the Fornax cluster. This is an arbitrary decision since the absolute distance calibration will be done using the entire sample, however, it can do no harm to use a reasonably nearby and quite large cluster to fix everything to. The ITF calibration is performed similarly but with $\langle \mathbf{P} \rangle$ and an uncorrected $\langle \mathbf{m} \rangle$ matched in accordance with the results of chapter 3.

Once this is done, the distance to each galaxy can be estimated relative to the distance to Fornax (ie $\hat{\mathbf{r}}_{\mathrm{FORNAX}} = 1$). Then, the absolute distances can be set as in sect. 3.2. At this point, we have an important check on the status of the calibration. If the intercept of the absolute calibration regression line is significantly away from the origin, then there is clearly something wrong with the calibration. However, for both the DTF and ITF calibrations, the shift is negligible (less than 100 km s^{-1}).

Distribution Parameter Correction

The distance estimates are now ready for use. Cluster distances can be derived by averaging the distances of all the individual galaxies with the formal error decreased accordingly and any Malmquist corrections etc applied. However, for the Monte Carlo correction procedure and any other Monte Carlos (for example, for estimating the errors in the maps), we need to know the full $p(\mathbf{M}, \mathbf{P})$ distribution parameters. We could, of course, just use those from the composite clusters but, although the slope and zero point of the calibration relations are 'unbiased', the individual parameters ($\hat{\rho}, \hat{\sigma}_M, \hat{\sigma}_P$ etc.) will be strongly effected by magnitude

selection and will biased individually. We would like to be able to correct for this.

Unfortunately, because we have combined a number of clusters to get the calibrating cluster, the selection function appears in a highly complicated form. Even if we can assume a sharp apparent magnitude cut-off, this will affect each individual cluster differently and the final combination will be vary hard to model. This makes an analytic correction difficult if not unfeasible. However, it is probable that the Monte Carlos are not particularly sensitive to the distribution parameters. Therefore, we can try some very simple corrections and see how much they affect the recoveries when compared to results using uncorrected parameters. If they turn out to be different, then there is no choice but to develop a decent correction procedure, but if there is no apparent sensitivity then there is no need to do this and the crude corrections, or indeed no corrections at all, can be used. In particular, if any differences are much less than the random errors in the POTENT recoveries, then any additional correction would be wasted effort.

Recall that we already have a 'corrected' $\hat{\sigma}_M$ which was found by simply taking the maximum from the individual clusters. We can also obtain an improved $\hat{\sigma}_P$ in the same way[2] . We can also assume that $\hat{\rho}$ is not too biased since it does not vary much from cluster to cluster (at least in comparison to $\hat{\sigma}_M$ and $\hat{\sigma}_P$). All that remains, therefore, are $\hat{\mathbf{M}}_0$ (from $\hat{\mathbf{m}}_0$) and $\hat{\mathbf{P}}_0$.

We can see that

$$E(\mathbf{m}|\omega = \omega_0) = m_0 + b \qquad (6.1)$$

where m_0 is the actual mean apparent magnitude for galaxies at log distance ω_0 and b is some bias brought about by magnitude selection. This bias will always be negative since the selection will tend to remove high magnitude galaxies. Therefore, a very simple first approximation might be to take a 'corrected' estimate $\hat{\mathbf{m}}_0$ given by

$$\hat{\mathbf{m}}_0 = \hat{\mathbf{m}}_0^c + \frac{1}{2}(\hat{\sigma}_M - \hat{\sigma}_M^c) \qquad (6.2)$$

where the ^{c} denotes the uncorrected value from the composite cluster calibration. Similarly for $\hat{\mathbf{P}}_0$. Of course, these modifications are very crude and should certainly not be used if any sensitivity is shown. However, they do have the advantage of extreme simplicity and if in the test which will be seen in sect. 6.3.3, no sensitivity is seen, then they can be accepted.

---

[2]In fact, for each catalogue, the appropriate $\hat{\sigma}_M$ and $\hat{\sigma}_P$ were found using same cluster – a heartening result

Calibration of the Improved Mathewson Data

The calibration of all the Burstein data proceeds much as for the Mathewson catalogue. However, for the improved Mathewson data, we now have up to six V_{rot} parameters for each galaxy. Therefore, the calibration process if performed six times with all galaxies that have the appropriate V_{rot} value included in each. Note that since each calibration will have a different absolute scaling, the calibration galaxy will have a number of different distances. This may seem counter-productive, but it is in fact exactly what is wanted. There is nothing special about the calibrating cluster – its choice is arbitrary. Therefore, it is unreasonable to impose a fixed distance on this one cluster (effectively assuming that its distance estimates have zero error). By calibrating each relation separately, this 'special treatment' is avoided.

All of the individual estimates for each galaxy can then be averaged, weighted by the error estimates for each relationship. However, unlike the clusters, we cannot reduce the formal error by $\sqrt{N}$ since all the relations will be correlated. Although it is in theory possible to estimate the correlation from the data, this is far from pleasant. I have therefore taken the "worst-case" situation of using the best of the individual error estimates for each galaxy. Although this is not perfect it should be adequate for weighting in POTENT and is unlikely to effect the accuracy of the Monte Carlo determinations sufficiently to make to additional work worthwhile.

6.2.2 Formation of the Final Catalogue

All that now remains is to combine all the separate calibrated samples into one catalogue. The first thing to check, therefore, is whether there is any overlap between the samples. Each galaxy must only be included once or the distance errors will give rise to a "Finger of God" effect, thereby distorting the derived velocity field. I check this by assuming that the angular position and redshift of each galaxy is well determined and looking closely at any galaxies that are within a degree or two of each other and not separated by more than a few hundred km s^{-1} in redshift space. However, even with these large bins, no overlap between the Mathewson and Burstein spirals is found (clearly, there will be no overlap between the spirals and the ellipticals). If a number of matching galaxies had been found, it would have been possible to use them as a test for systematic differences between the estimates from

each catalogue, but without them, there is very little that can be done simply. The best solution, therefore, will have to be to trust the calibration procedure and simply stick all the surveys together.

However, there is a problem with the ellipticals. When performing the absolute calibration, the intercept deviates significantly from the origin. However, when the field ellipticals are removed the deviation is not so severe, although it is still greater than for either of the spiral calibrations. This later point is unsurprising because of the large errors in D_n-σ estimates and the relatively small sample, but the distortion produced by the field galaxies is more significant. As has already been mentioned, there are good reasons to suppose that field ellipticals are physically different from their field counterparts because of the differing environments. This effect, therefore, may well be due to the inapplicability of a calibration based on cluster galaxies to the field. Whatever the cause, the only safe course of action is to discard the field ellipticals from the final catalogue.

We are, therefore, left with the following: 1147 objects in the Mathewson sample, of which 20 are clusters. All of the clusters have 'improved' distances as do 717 of the field galaxies. In addition, there are 241 Burstein spiral objects (17 clusters) and 133 ellipticals (all of them clusters of course, but only 77 of them with more than 1 observed member. In future, I will refer to the single galaxies as field ellipticals for convenience as this is their effective role in POTENT). The combined distribution of all of these objects is shown in fig. 6.5.

6.3   The Maps

In this section, we finally get around to some actual maps of the local Universe. All the recoveries shown here are for the same basic POTENT configuration. The spherical grid used has 25 radial shells out to 10000 km s^{-1}, 35 latitudinal rings and 70 sin θ longitudinal rings where θ is the latitude. The potential is calculated on this grid and linearly interpolated onto a cubic grid with grid points spaced at 500 km s^{-1} intervals and the velocity and density calculated using this mesh. Note that the Max-flow methods perform the entire calculation on the cubic grid. The window chosen, as before, is the 'Volume Weighted' window and the smoothing radius is 1200 km s^{-1}.

Because of the nature of this section, there are a considerable number of figures. However,
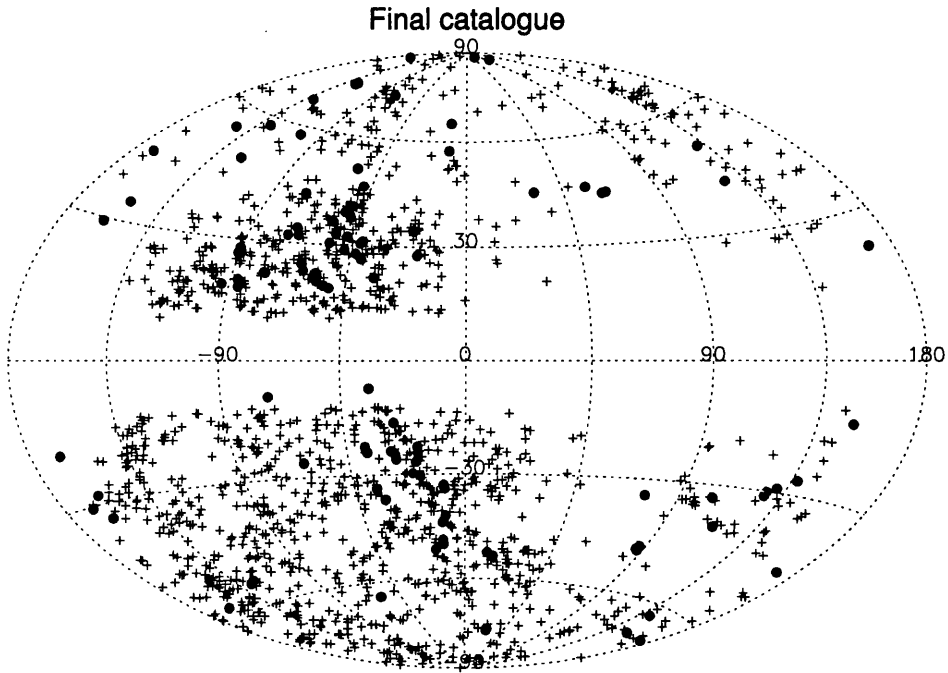
Final catalogue

Figure 6.5: An Aitoff projection of the positions of galaxies in the final catalogue, again using galactic coordinates. As in fig. 6.4, the filled circles denote clusters and the crosses, field galaxies.

all are positioned so that they appear as close as possible to their first reference in the text and they will always appear *after* that reference. Also, the form of each set of maps is the same except where specifically noted, so that comparison can be easily made. This is important since, unlike the previous maps in this thesis, I obviously do not now have the actual fields to compare to, so the analysis must be an "eyeball" one. Nevertheless, as shall be seen, even in this rather crude way, the different effects of the various methods can be clearly seen and by comparison to the tests already given, 'good' and 'bad' features can be recognised.

For reference, fig. 6.6 shows the estimated errors on the slices through the recovery that I will be using for the forthcoming results. Note the irregular spacing of the contours and the large areas of the field that are reasonably well recovered. Although these estimates are only for the recovery using uncorrected ITF estimates, they will be used as the mask for all the other schemes since this will make cross-comparison easier, but in general the
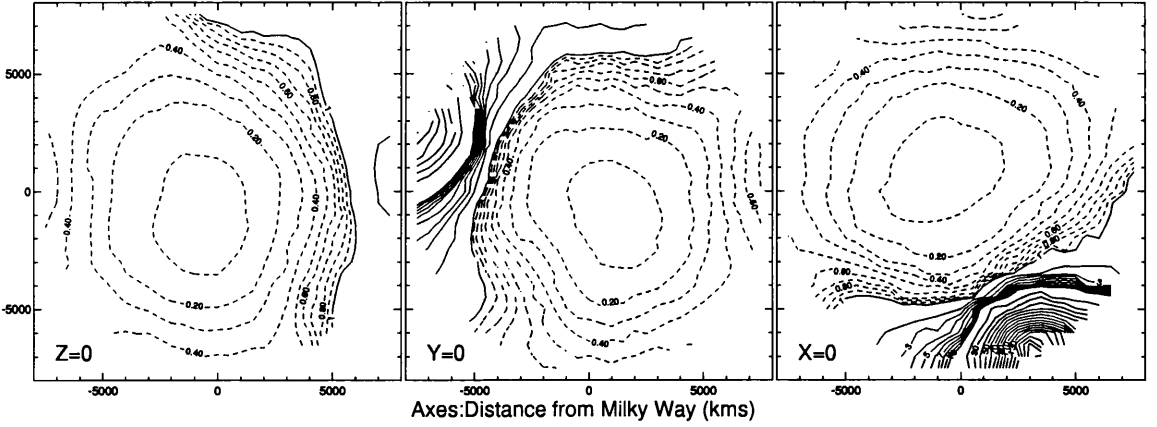
Figure 6.6: Estimates of the errors in the density recoveries of POTENT using the data set described in sect. 6.2. The errors are estimated during a POTENT Monte Carlo realisation of the new data set using uncorrected ITF distance estimates. Three slices are shown, all passing through the origin with $Z = 0$ the Super-galactic plane. Contours join equal values of $\hat{\sigma}_\delta$ with dashed contours spaced at $\sigma_\delta = 0.1$ intervals up to 1, solid contours at $\sigma_\delta = 1$ up to 10 and thick contours at $\sigma_\delta = 10$ intervals on from there.

errors will be different for each technique and should be estimated for each recovery before performing any analysis. In general, though, all the maps in this section will be cut off at the $\hat{\sigma}_\delta = 0.6$ contour and also truncated to 8000 km s^{-1}. The latter limit is to avoid any possible edge effects and the former to remove any areas that are too noisy to be of any interest. However, this limit is somewhat on the generous side for any actual analysis (eg sect. 6.5) and many other factors should be taken into account before defining an analysis volume (see Dekel et al. (1993) for a good example of masking off the unreliable parts of a recovery). The limit is chosen here to give us plenty of information to use for "eyeball" comparison, but it must be remembered that many areas, particularly near the edge of the field have significantly larger errors.

6.3.1   Comparison to Potent90 using the New Calibration

Figure 6.7 has the results for the only use in this section of DTF estimates. In spite of the reservations expressed earlier about the problems with calibrating DTF-like estimators, the uncertainties introduced will not affect the basic features of the recovery and it is interesting to see compare the result of the expanded data set and different calibration technique has
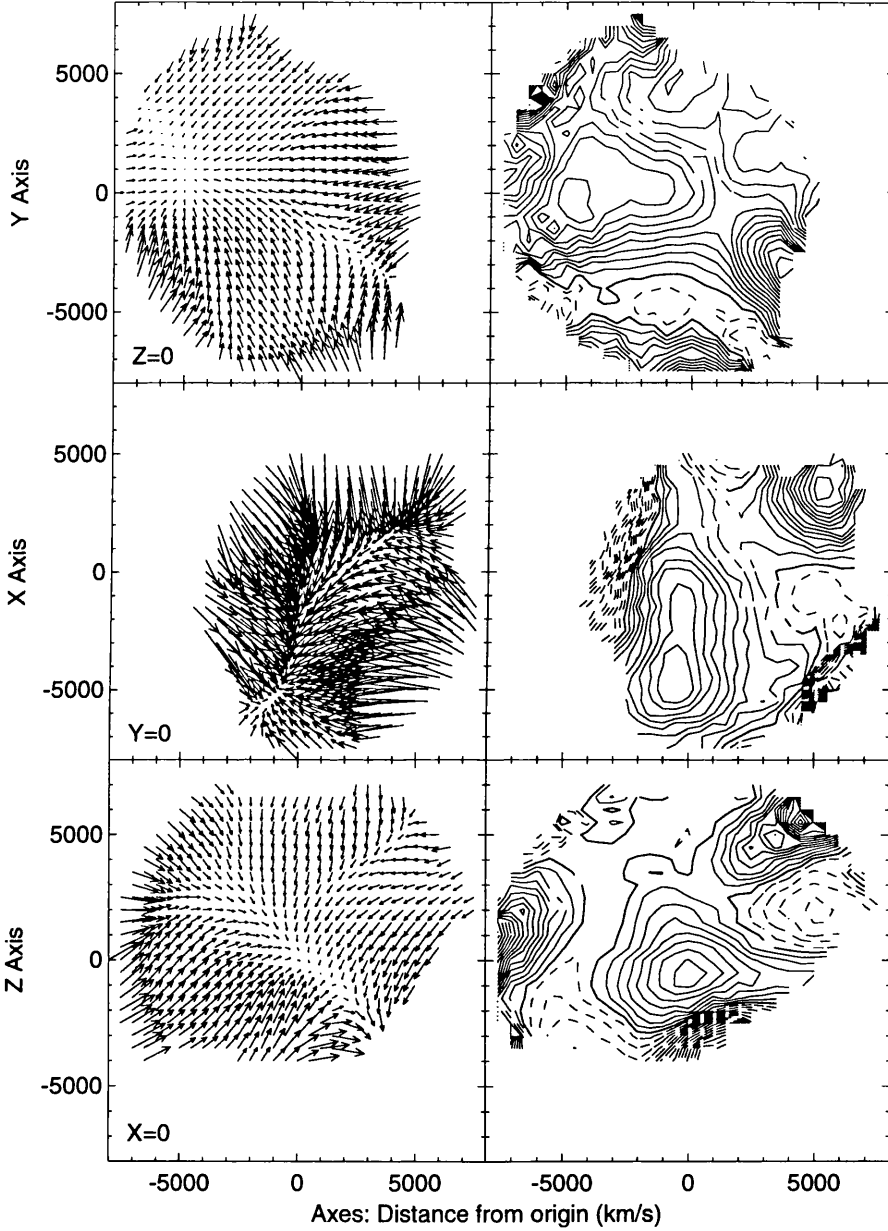
Figure 6.7: Velocity and density maps for three slices through the origin of the basic POTENT recovery using DTF estimates with an homogeneous Malmquist correction applied. In this and all the following maps, the $Z = 0$ slice is the Super-galactic plane and the maps have been truncated to 8000 $\mathrm{km\,s^{-1}}$. In addition, all regions with $\hat{\sigma}_\delta \geq 0.6$ have been removed (see fig. 6.6). Density contours are at 0.1 intervals with negative contours dashed, positive contours solid and the zero contour thicker.

to the POTENT90 recovery in fig. 6.1.

Obviously, there are significant differences. Most of the features appear to have 'deepened' and there are some new features. Also, the velocity recovery for the $Y = 0$ slice has very large velocities coming in from the edge of the recovery giving rise to prominent voids and attractors. These effects do not appear in any other recovery, and have all the hallmarks of a large bias, probably due to the inapplicability of the correction to such an inhomogeneous catalogue. In fact, there are other areas where the velocities seem to show a strong bias, for example the radial infall at the edge of each of the slices (especially the $X = 0$ map). However, generally, the results are certainly comparable – I am clearly not doing anything fundamentally different here.

6.3.2 Potent Recoveries using an ITF Estimator

We have seen, however, that in general the best recoveries from a simple Malmquist correction are found when an inhomogeneous corrections is made to galaxies with ITF-like distance estimates. This result is shown in figure 6.8. The first thing that springs out of these maps are the major differences from the previous results. Although some of the features are still recognisable (for example, the 'Great Attractor' (hereafter GA) at about $(-4000, 1000)$ on the Super-galactic slice) there are as many that are totally new. In addition, there seems to have been a downward shift of the overall density.

Although dramatic, these changes are not entirely unexpected. In particular, the density shift is easily explained. It was seen above that the poor correction given by the HMC led to an apparent infall towards the origin at the edge of the recoveries. When interpreted as genuine peculiar velocities, this will give the impression of a large density concentration in the vicinity of the origin. However, this will have been at worst reduced and possibly removed by the inhomogeneous correction. This effect can be seen directly in the velocity fields. The apparent infall at the back-side of the GA has gone as have the apparently erroneous velocities on the Y slice.

Interestingly, the uncorrected estimates are not radically different and certainly not as far removed as the DTF with HMC (see fig. 6.9). In fact, the differences are actually quite subtle in comparison to the expected errors. This might seem rather surprising given the
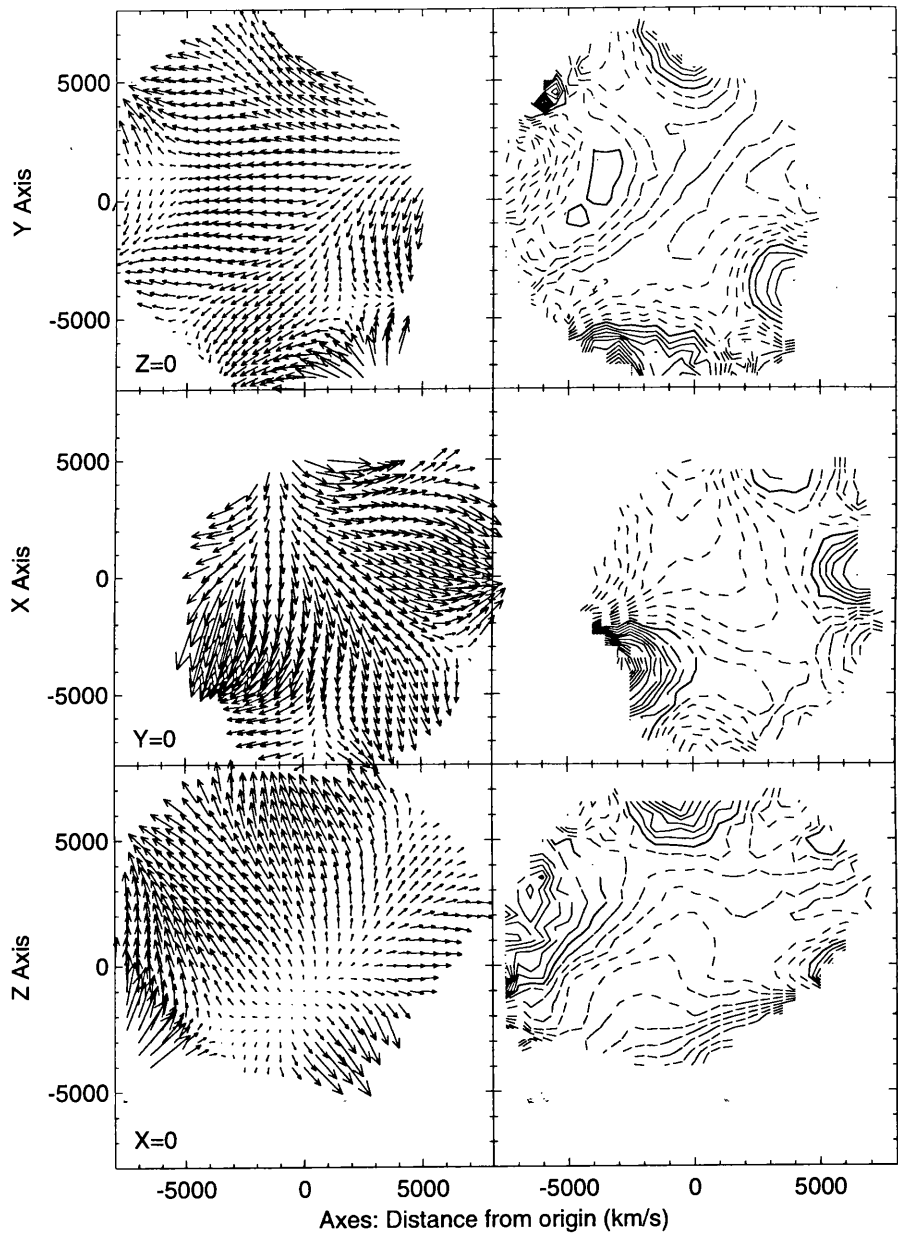
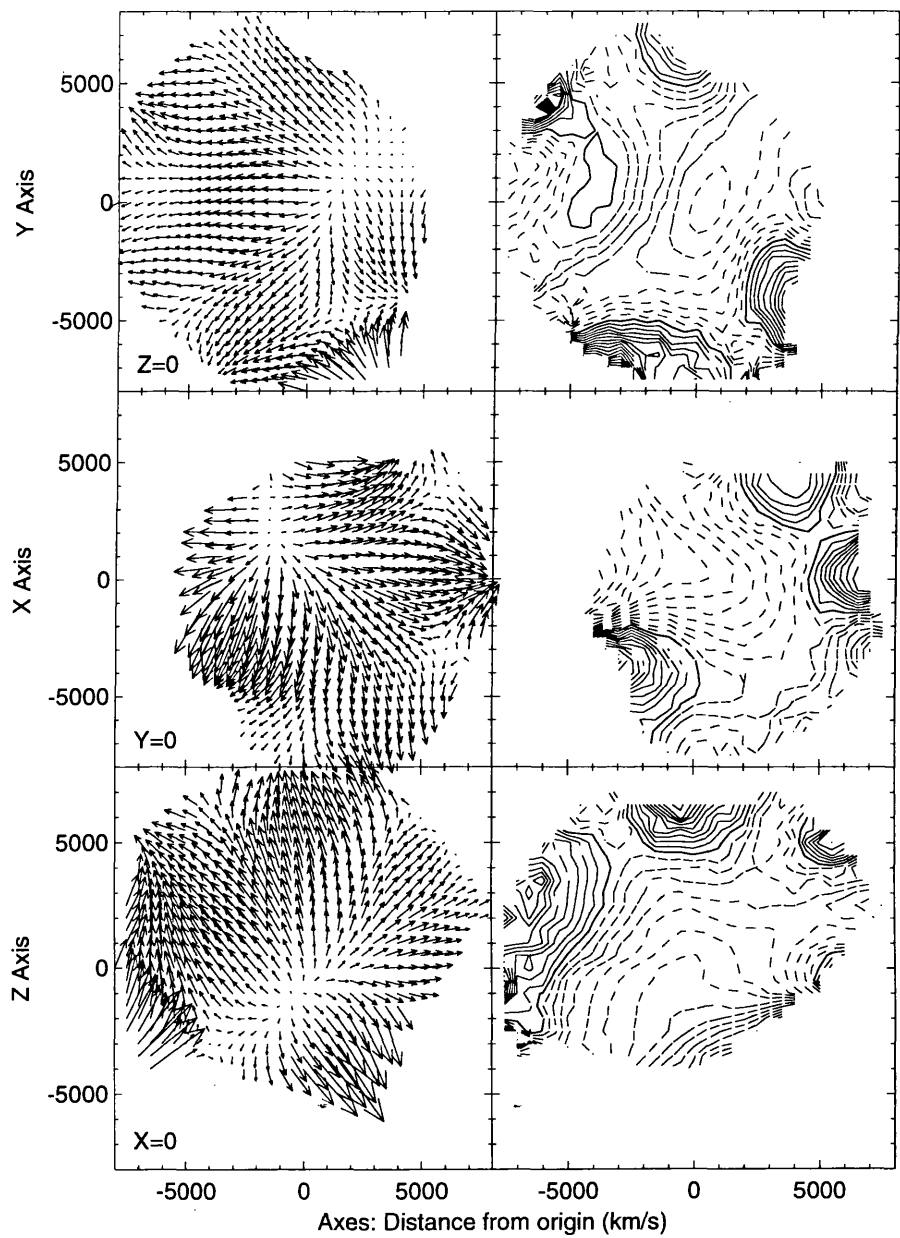Figure 6.8: As fig. 6.7 but for ITF estimators with an IMC applied.

Figure 6.9: As fig. 6.7 but for an uncorrected ITF estimator.

other results in this thesis. However, in this data set, there are a considerable number of clusters. With the weighting in POTENT, these are given considerable influence but because the estimates of their errors are reduced by $\sqrt{N}$, their Malmquist corrections are small. Therefore, standard Malmquist corrections which treat only radial effects will not alter the result significantly.

Sample Size Sensitivity

The sensitivity of the maps to sample size is shown in figure 6.10. Here, about one in six objects (galaxies or clusters) are removed randomly from the catalogue and POTENT applied to the remaining galaxies. Note that this is similar to the "Jack-knife" techniques to test for robustness. Uncorrected ITF estimates are used so that we are definitely testing the sensitivity of POTENT and not the Malmquist corrections, so these results should be compared to fig. 6.9. The differences are not large and seem to stay well within the estimated errors, so there is no obvious sensitivity even with this large fraction of the galaxies removed.

Nevertheless, non-radial inhomogeneities and sampling gradient biases have not been addressed at all yet. For that we need the Iterative Monte Carlo correction scheme.

6.3.3   Potent Results with Iterative Monte Carlo Corrections

The method was performed on the ITF estimated distances with the 'corrected' distribution parameters described in sect. 6.2.1. The convergence limit was set to 150 km s^{-1} to prevent 'over-shooting' but the recovery was continued for an extra iteration for comparison. To reduce the computational overhead and disk usage, the POTENT recoveries were performed on a coarser grid than was used for the previous results and the final bias estimates interpolated onto the fine mesh before being removed from the recovery in fig. 6.9. Also, since the "improved" Mathewson data of Salucci (1994) has up to six different estimates for each galaxy, the Monte Carlo procedure needs to be slightly modified. It would, of course, be possible to calculate all the V_{rot} values for each galaxy and average the estimates as in the raw data, but it is important to remember that the V_{rot}'s for a particular galaxy are correlated in some non-trivial way. Evaluation of the correlation would be time-consuming and, since all we need here is a reasonable distance estimate, largely unnecessary. The approach
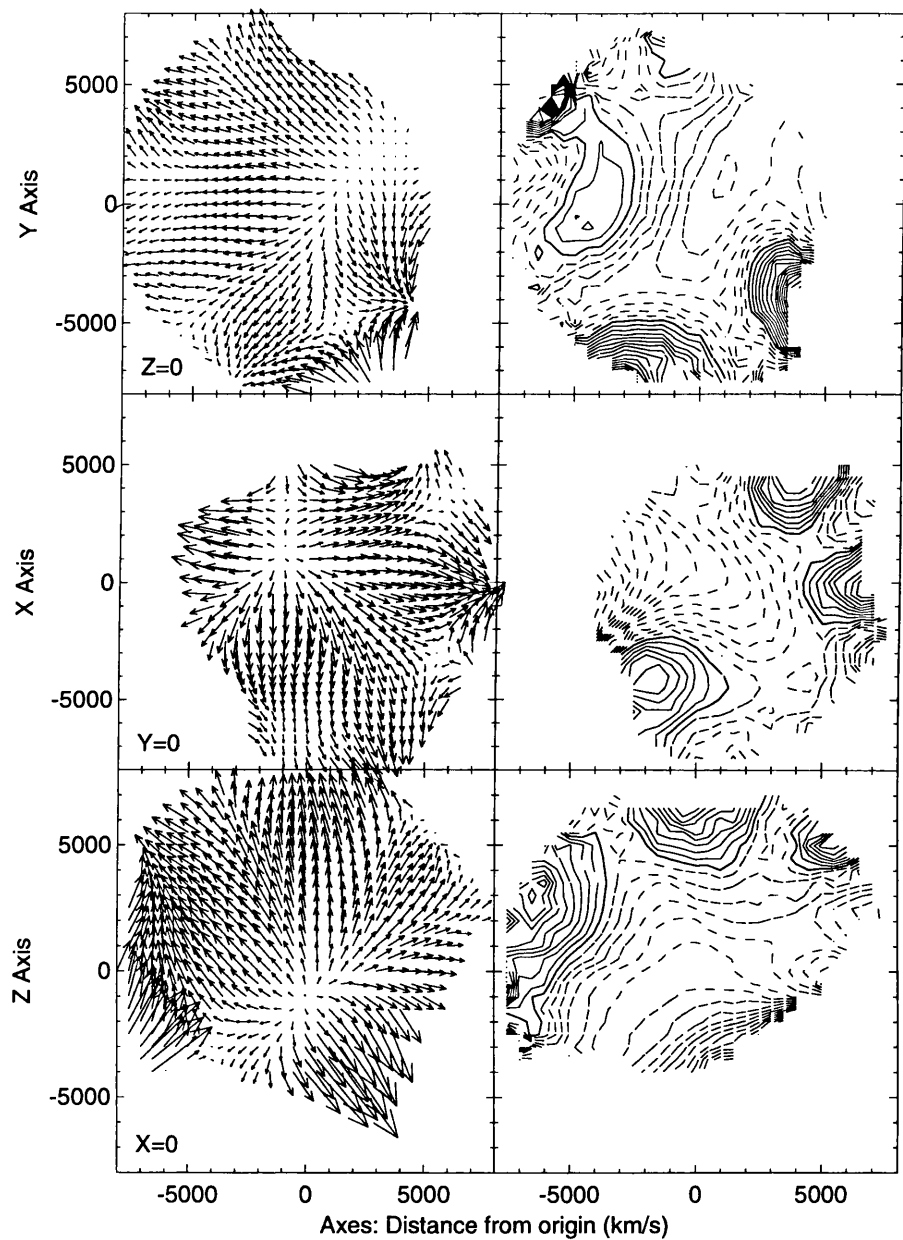
Figure 6.10: As fig. 6.9 but with about one in six objects randomly removed from the catalogue.

I have adopted instead is to use the estimated variance of each galaxy distance estimate in conjunction with the other known distribution parameters (ie $\hat{\rho}$ and $\hat{\sigma}_M$) to calculate an *effective* $\hat{\sigma}_P^{\text{eff}}$. This is then used to scatter the galaxy as normal. In clusters, however, since the galaxies are not related except by position, a separate estimate can be found for each galaxy and the results averaged as in the original data.

In fact, with this tolerance, convergence was reached after iteration 1. (Remember that the zeroth iteration has quiet Hubble flow imposed. Iteration 1 is the first with the field from the actual data used). The final result is shown in figure 6.11. Again, there are no major differences from the IMC recovery, but the changes that there are are significant. As with the tests (sect. 5.4.4), the density features have been tightened and enhanced, counteracting the effect of sampling gradient biases. Also, some residual Malmquist bias infall appears to have been removed, though only in selected regions – behind the GA, for example, there is now no evidence at all of any back-infall, although the errors in this region are large and no conclusions can be sensibly drawn from this. Nevertheless, in spite of the good coverage of clusters, there is evidence here of uncorrected Malmquist biases and significant sampling gradient biases in the IMC recovery. Having said that, if only the general features of the maps are required, and not details of the *size* of fluctuations, inhomogeneously corrected ITF estimates would seem to be adequate.

In figure 6.12, the progress of the Monte Carlo correction procedure is shown iteration by iteration. Maps for the Super-galactic plane ($Z = 0$) slice are shown for the zeroth iteration (quiet Hubble flow imposed), first (convergence tolerance of 150 km s^{-1} reached) and second (convergence tolerance drops below 100 km s^{-1}). As with the tests, most of the correction is done during the first iteration since most of the bias in the recovery is Malmquist-like. However, the differences between the 0th and 1st iterations show the non-negligible sampling gradient biases that have softened the density features in the other recoveries. By iteration 2, the convergence tolerance has dropped to the level where residual noise in the Monte Carlo recoveries might become significant. We might expect some over-correction since there method will be attempting to fit to the noise, but there is no strong sign of this. Again, this can be partly put down to the influence of clusters in POTENT which will tend to stablise the results because of their relatively small errors and high weighting (see eqn. (4.14)).
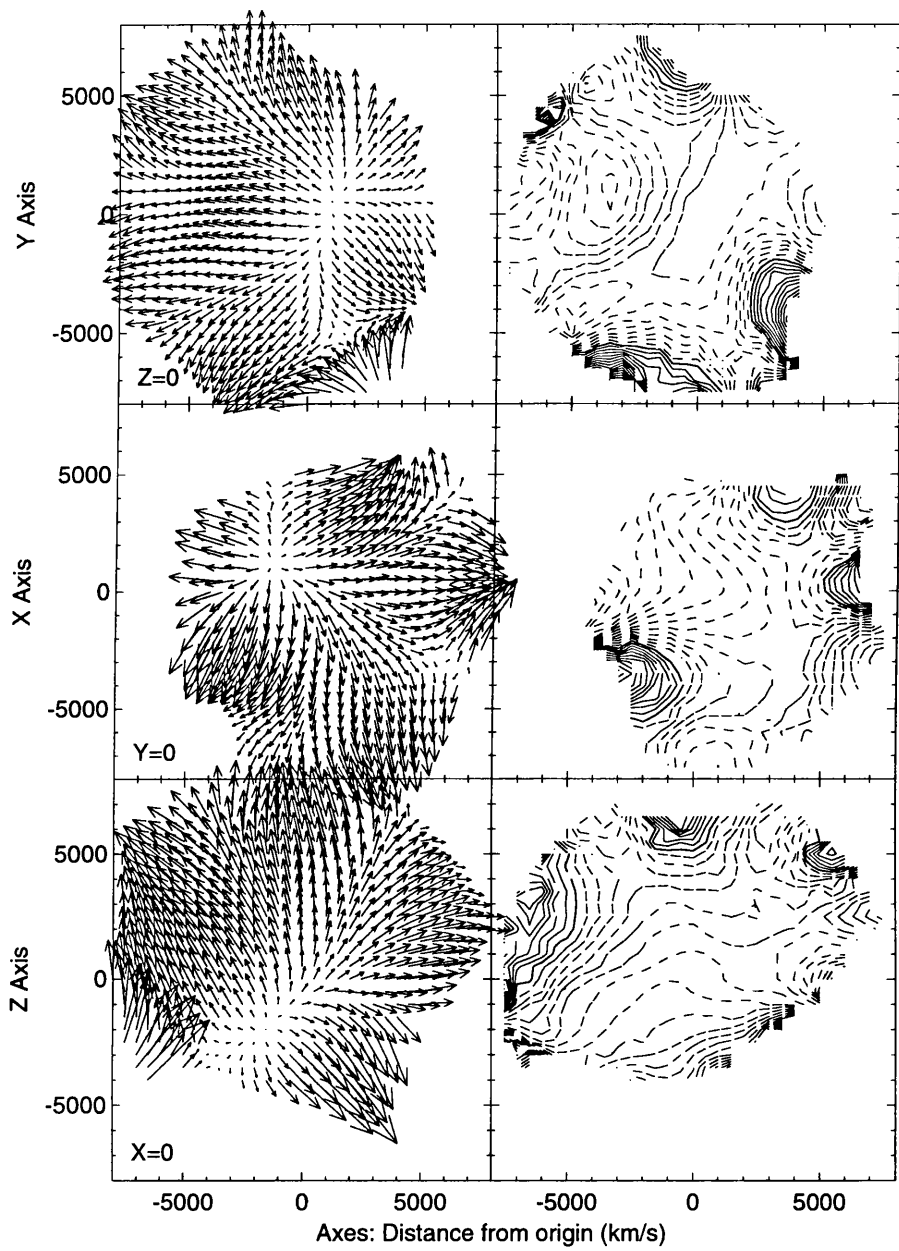
Figure 6.11: As fig. 6.7 but for the result of the iterative Monte Carlo correction method (convergence reached after iteration 1).
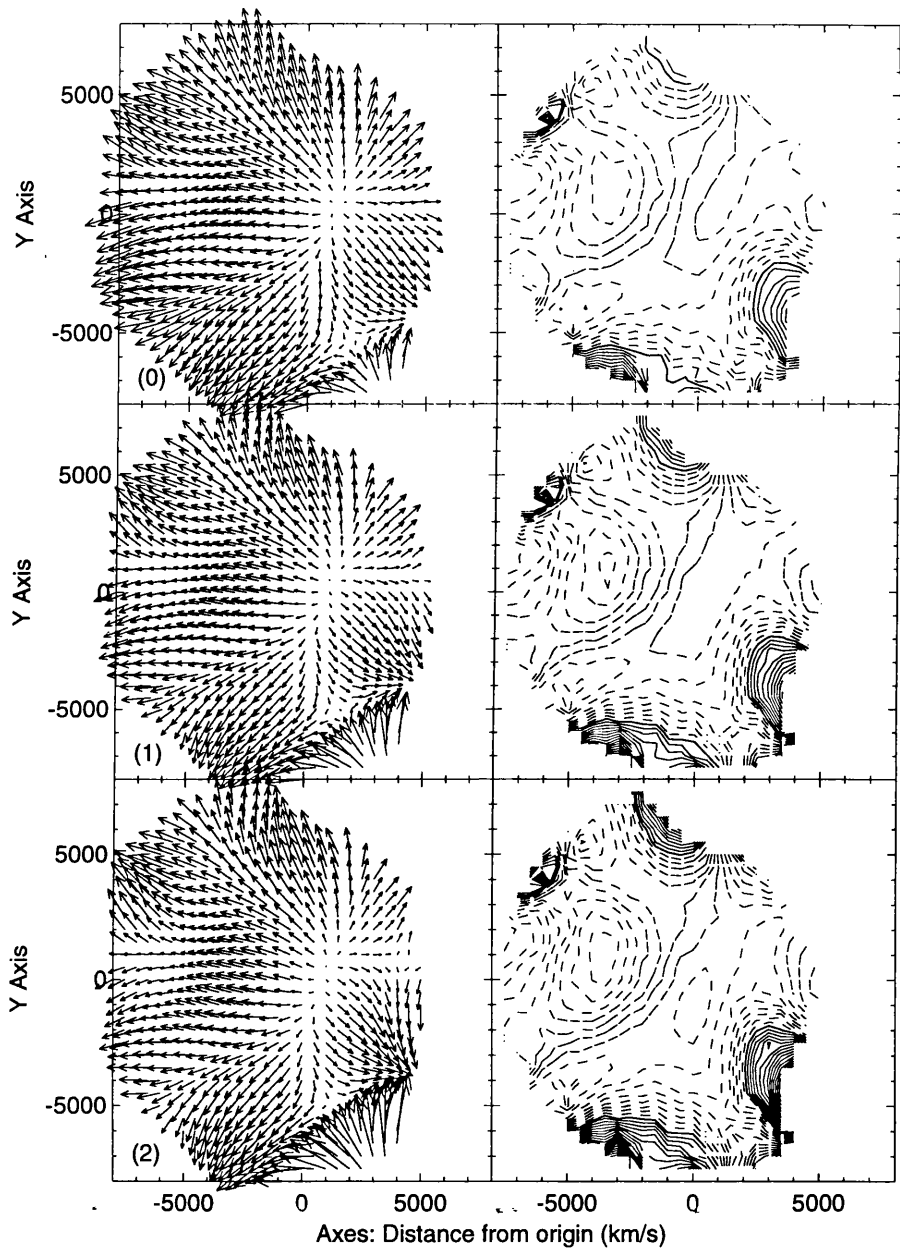
Figure 6.12: The three sets of maps here are similar to fig. 6.7, but show the Super-galactic slice for the three iterations of the Monte Carlo correction procedure. The Zeroth iteration at the top, the first (when the convergence tolerance was reached) in the middle and the second at the bottom. The recoveries are very similar since most of the bias correction is done during the zeroth iteration.
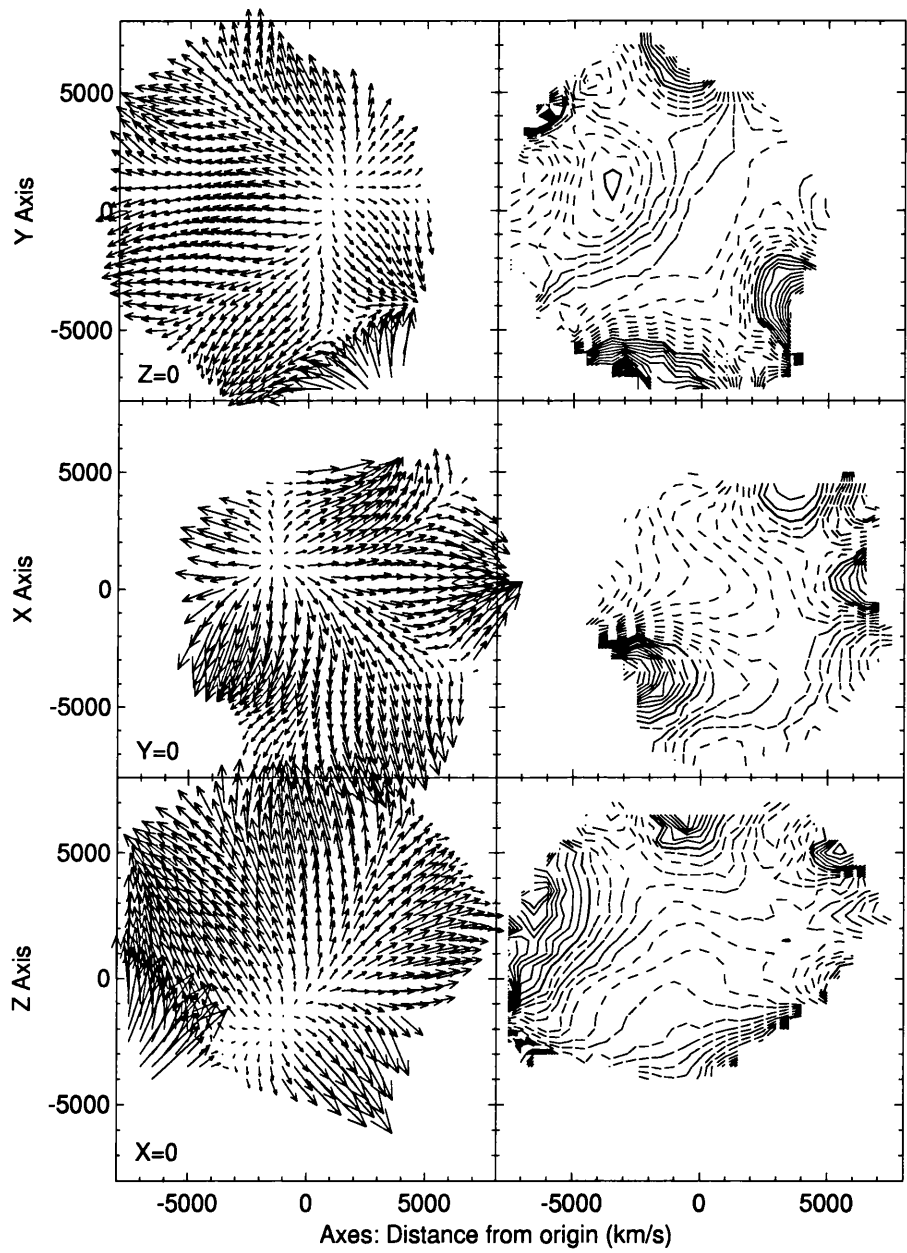
Figure 6.13: As fig. 6.11, but with the distribution parameters taken directly from the calibration used to scatter the galaxies during the Monte Carlos rather than the partially corrected ones described in sect. 6.2.1.

Finally for this section, in figure 6.13 I show the results of a different set of iterative corrections, this time using the uncorrected, "crude" distribution parameters rather than the partially corrected ones devised in sect. 6.2.1. If the method were sensitive to the parameter estimates, this result would be significantly different from the previous maps (fig. 6.11), but there are no obvious changes. Therefore, it is reasonable to conclude that we do not need to worry unduly about the quality of the parameter estimates and that the rather ad hoc corrections used in sect. 6.2.1 are more than adequate. Of course, this sort of test should be performed for each new catalogue just in case.

The iteratively corrected maps of fig. 6.11 are, therefore, the most reliable and it is these that I will be using in the analysis of sect. 6.5. However, a number of other methods have been described in this thesis and these have also been performed upon the real data.

6.4 Results using Z-Potent, Max-flow and Others

The first of these other methods to be applied is Z-POTENT – a POTENT analysis in redshift space (see section 5.1). The aim behind this method is to minimise Malmquist-like errors by performing the smoothing process in redshift space where distance errors are negligible. The initial tests showed that the method was certainly applicable but because of the difficulties of moving from redshift to real space where there are large velocity gradients, the recoveries had a tendency to produce rather noisy density fields.

Both these trends are shown in the results in fig. 6.14. Although there is evidence of good Malmquist correction (the velocity fields resemble the Monte Carlo corrected fields), the noise is far greater, particularly in the density recoveries. Therefore, although perhaps of some interest in the future with larger galaxy samples, or with a modified smoothing procedure, Z-POTENT is currently of little practical use.

Max-flow Methods

The same appears to be true of the Max-flow results where, as described in section 5.3, non-radial integrals are performed in an attempt to avoid areas particularly sparse in galaxies. Figure 6.15 shows the Super-galactic plane slice for three different Max-flow runs distinguished by different weighting schemes. In graph (a), the paths are only midly non-radial
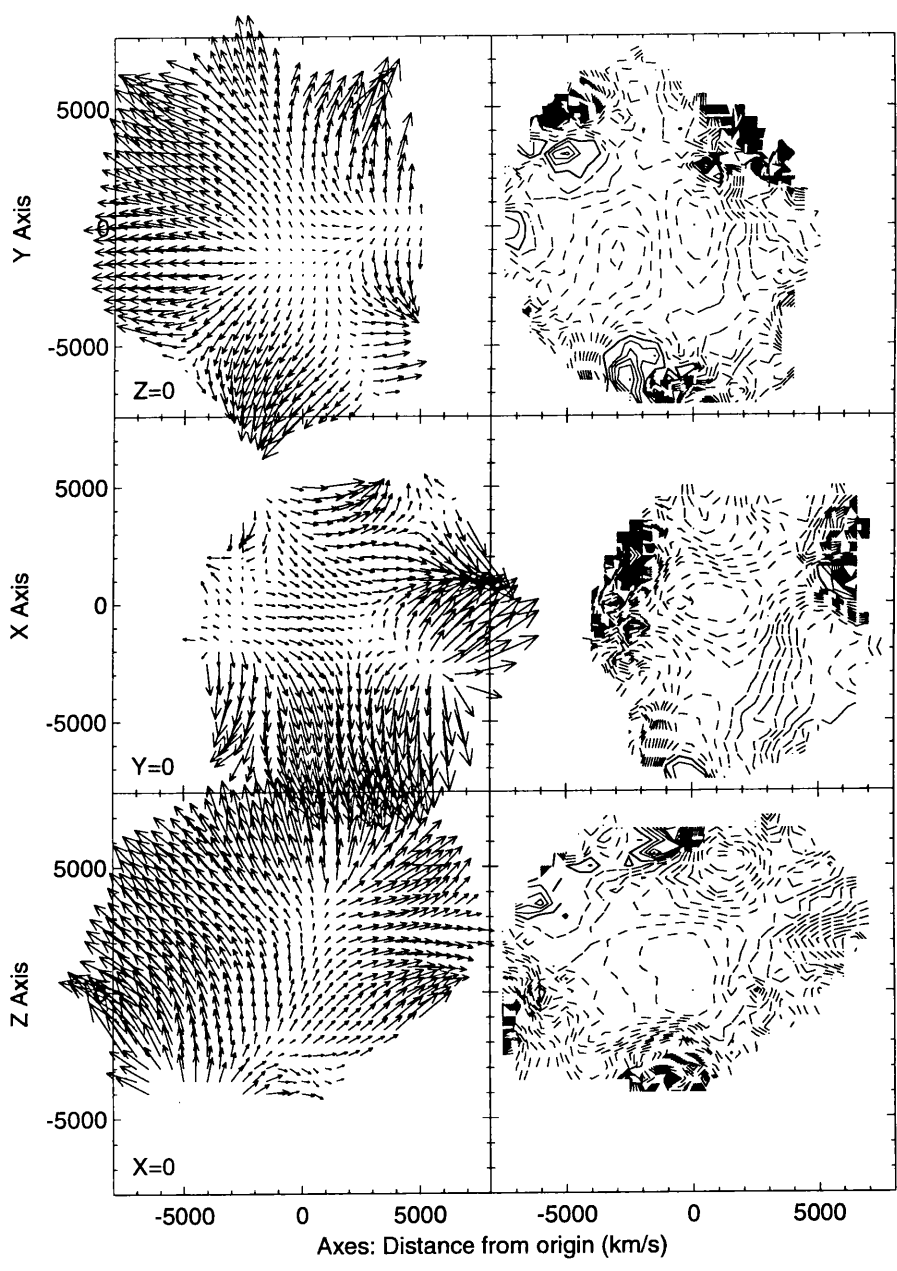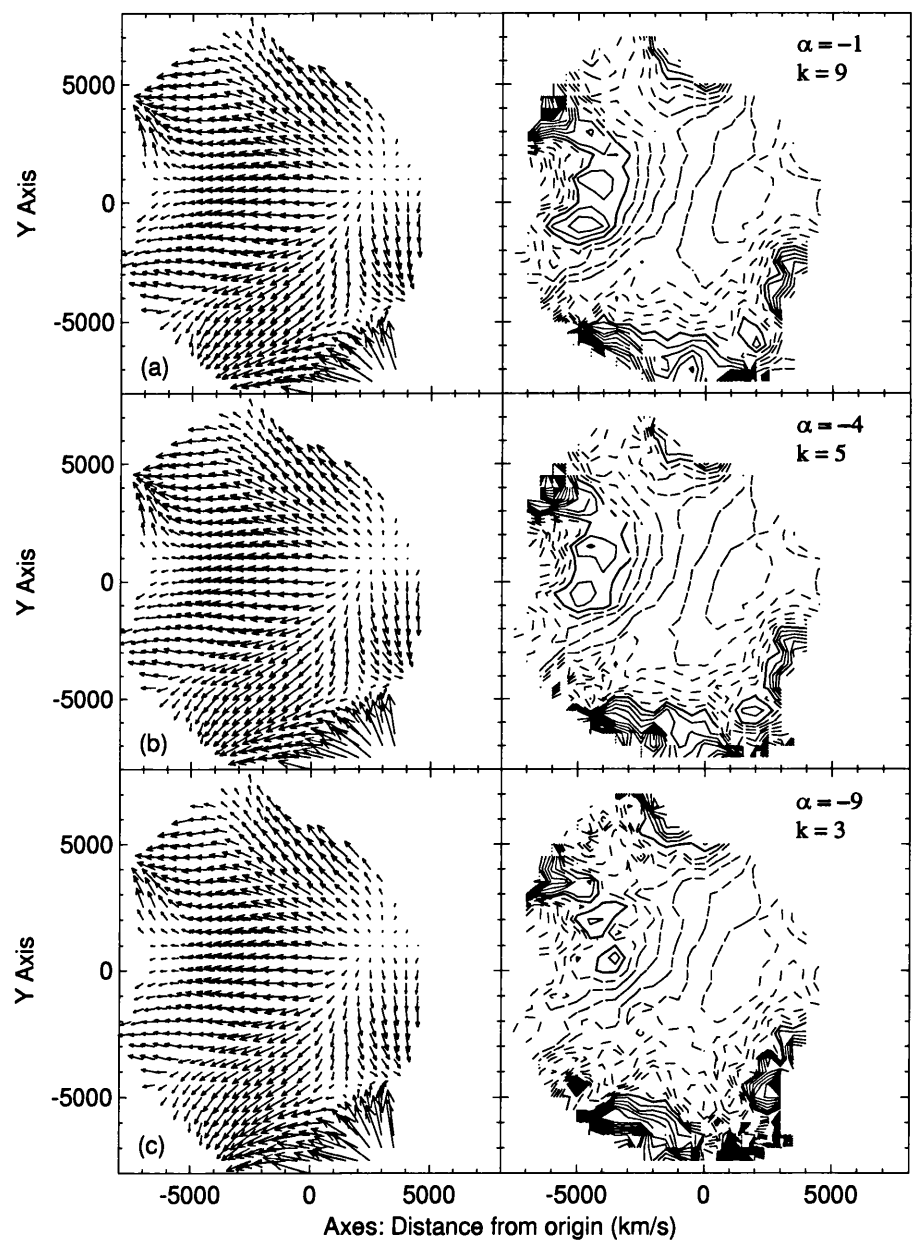
Figure 6.14: As fig. 6.7, but for Z-POTENT.

Figure 6.15: As fig. 6.12, but for several Max-flow analyses with different path weightings. See the text for the details of the weighting schemes. Note that uncorrected ITF distance estimates have been used for all the recoveries.

since the density weighting is low ($\alpha = -1$) and the transverse errors assumed to be much larger than the radial ones ($k = 9$). For (b) and (c), the paths are allowed to become progressively more non-radial. However, even for the virtually radial paths, the result is noiser than the orthodox POTENT recovery (fig. 6.9) and the result gets worse for less radial paths. This is a particularly disappointing result since the method was designed to *decrease* noise, but is, again, consistent with the results of the tests already performed. Much of the problem is due to the rather ad hoc weighting scheme used and the need to adhere to a cubic grid when creating the paths. Several alternative courses of action will be discussed in sect 7.1, but for now we must take comfort from the fact that at least no new biases appear to be introduced!

Multiple Distance Estimators

Finally, I have obtained one last set of results using a simple technique not described as yet. It is important to remember that there is nothing particularly fundamental about any given distance estimation technique. All that a distance estimate is, effectively, is a number associated with a galaxy which one hopes provides useful information about the distance – the important point is to understand the statistical properties of the estimator and use it appropriately. Also, recall that distance estimates are used *twice* in POTENT: once to position the galaxies for smoothing and once to estimate the radial peculiar velocity of each galaxy. So far we have always used the same distance estimator to do both, but there is no reason why this has to be the case. In fact, the properties required by the two processes are very different. When positioning the galaxies we wish to minimise Malmquist-like biases, hence the use of Malmquist corrections, but when estimating $\hat{v}_r$, we really want the unbiased properties of the uncorrected ITF estimates. Obviously, when smoothed on scales such as that found in POTENT, the differences are not likely to be large which is why I have not considered this until now, but predominantly as a reminder of the importance of understanding estimator properties, I have run POTENT using inhomogeneously corrected ITF estimates to position galaxies and uncorrected estimates to find $\hat{v}_r$. The results are shown in figure 6.16. As expected, there are no large differences from the IMC recovery of fig. 6.8. However, the small differences are all positive. Density enhancements have been increased slightly in line with the iterative correction results and the velocities have also
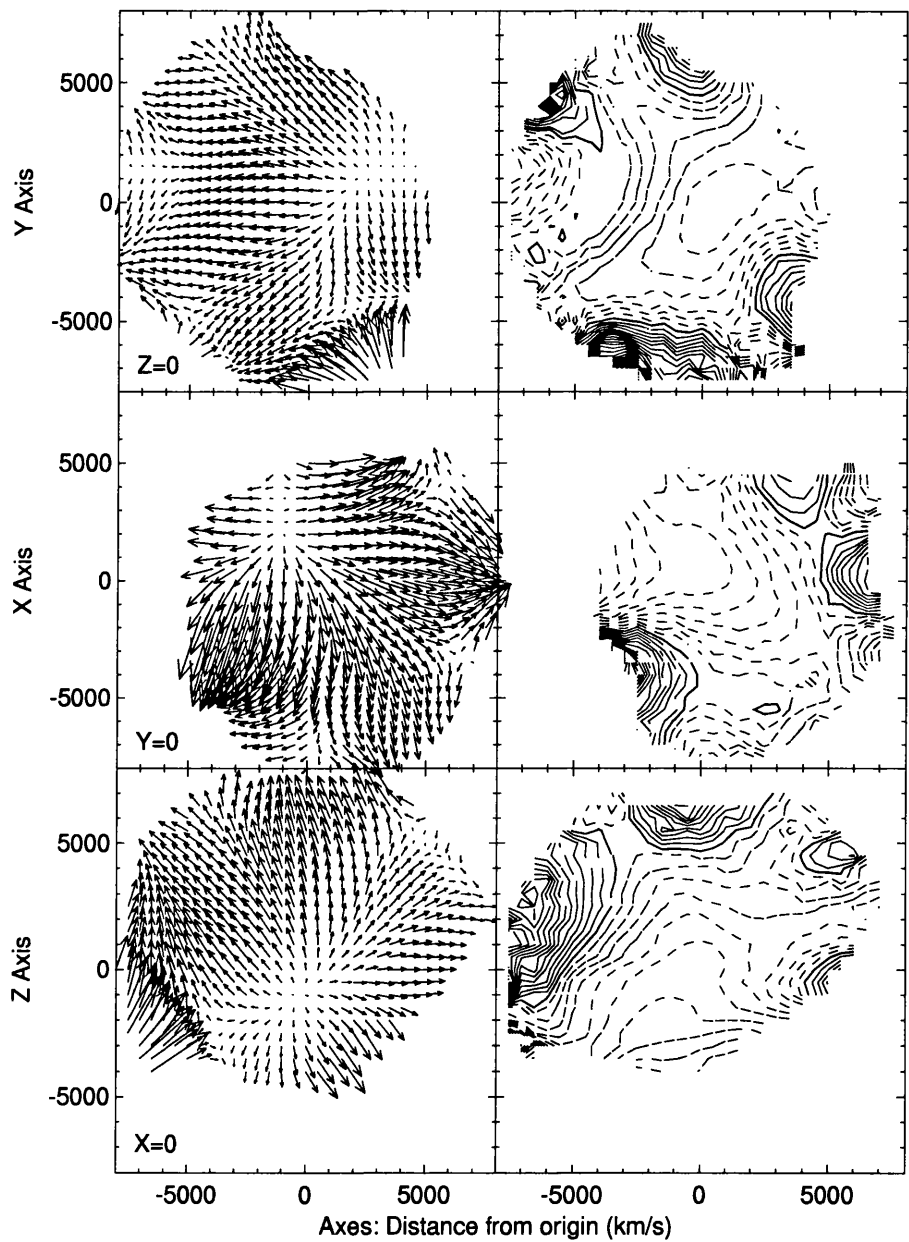
Figure 6.16: As fig. 6.8, but with inhomogeneously corrected ITF estimates used to position galaxies and uncorrected estimates used to find $\hat{\mathbf{v}}_r$.

increased slightly in appropriate areas. Nevertheless the changes are, in all cases, less than the estimated errors and are not significant. Therefore, there is clearly nothing to worry about when applying corrected estimates to estimate $\hat{\mathbf{v}}_r$, but it must be remembered that this approach is in some sense inappropriate.

Removal of Residual Hubble Flow

One thing that is striking about most of the density maps in this chapter is that the average density contrast δ appears to be negative, which is unlikely to be the case in the real universe. One possible cause of this is some residual Hubble Flow remaining in the field. Such a linear radial flow out from the origin in the velocity field would give rise to a zero-point shift downwards in the density maps. One might, therefore, decide to *enforce* an average δ of 0 by an appropriate rescaling of the densities. Some examples of this are shown in fig. 6.17. Here, the density recoveries of figures 6.9, 6.8 and 6.11 have been scaled such that the average δ ($\bar{\delta}$) within the regions of the recovery where $\hat{\sigma}_\delta < 0.4$ is zero. (Note that these maps have all areas with $\hat{\sigma}_\delta > 0.4$ removed and are, hence, slightly different from the previous $Z = 0$ slices). The velocity maps have then had an appropriate linear outflow removed.

This form of correction immediately makes the maps more convincing. However, it is dangerous to apply such methods without very careful thought. There was no evidence in the tests of chapter 5 of a residual bias of this type which would imply that the problem comes from the calibration. However, the zero-point shifts required to obtain these results are noticeably different for each map and, indeed, for the Homogeneously Malmquist corrected DTF estimates (fig. 6.7) the correction would have to be negative – equivalent to the removal of a residual *infall*. Perhaps the difference between DTF and ITF based results is still understandable in terms of the differing calibrations, but the contrast between, for example, Malmquist corrected ITF estimates (with a shift of $\delta = 0.26$) and the iteratively corrected results (0.46) cannot be explained in this way since the same estimates have been used. Therefore, other biases are probably responsible for much of the effect and there is no reason to suppose that these would have the linear, radial form of Hubble flow. In addition, there will always be the question of whether the volume averaged over is sufficiently large for the assumption that $\bar{\delta} = 0$. Therefore, for the remainder of the thesis I will not employ any
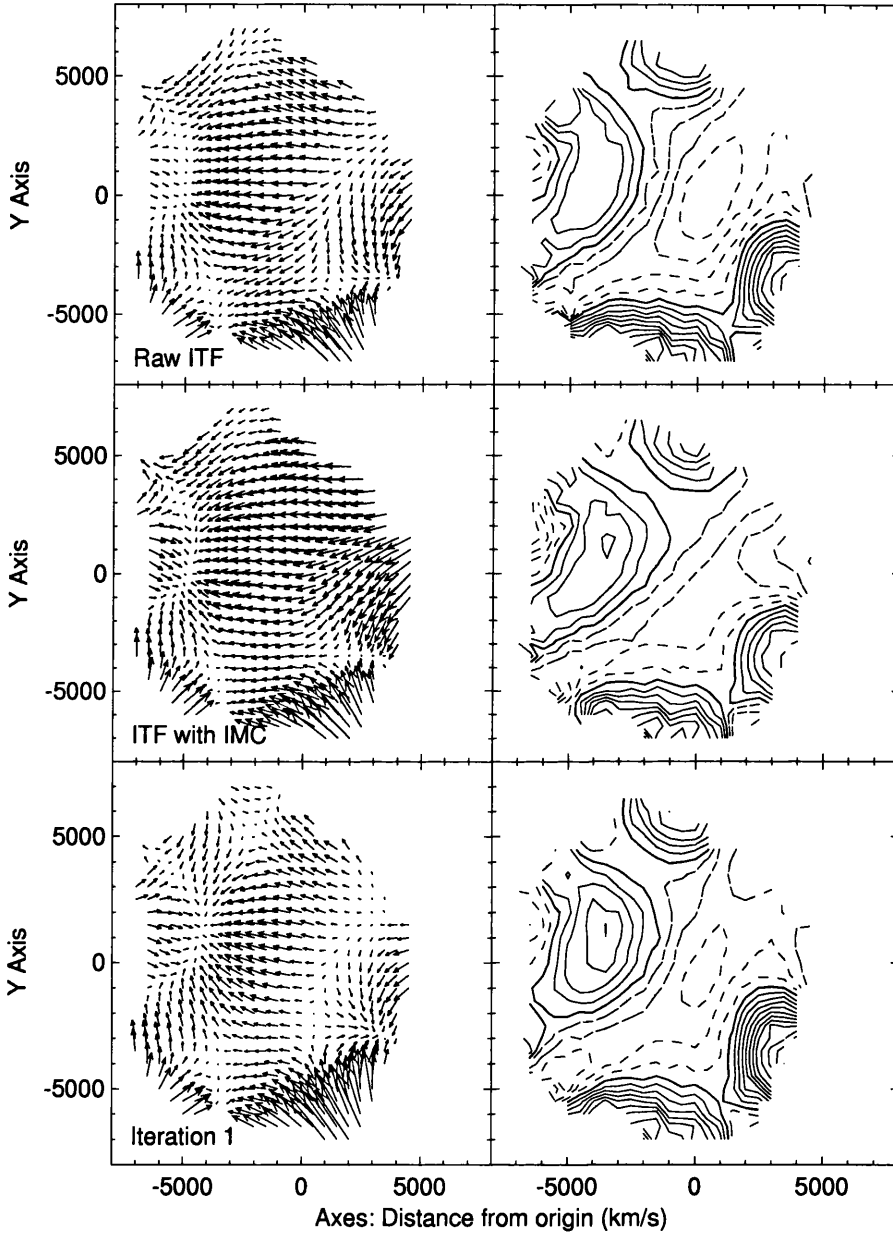
Figure 6.17: The Supergalactic ($Z = 0$) density slices are reproduced here for three of the previous sets of maps. However, here the zero-point level of the density contrast δ has been set so that the average density contrast within a particular volume is zero. Such a zero-point shift is equivalent to a linear radial velocity flow from or towards the origin, ie. residual Hubble flow and the equivalent flow has been removed from the velocity maps. The volume chosen was all areas where $\hat{\sigma}_\delta < 0.4$ and the plotted areas reflect this selection (with an additional cut off at a radius of 8000 km s^{-1}).

corrections of this sort. However, there is a clear indication that the absolute calibration of the distance estimates used here may not be entirely appropriate for POTENT-like methods and this must be tackled in the near future.

6.4.1 Comparison to Recent Results

Figure 6.18 has a comparison between the Monte Carlo corrected density field on the Supergalactic slice and the results from the POTENT collaborations' preliminary results using the Burstein Mark III data (Dekel, 1994). The basic agreement between the two is very good with all the features recognisable. However, there are some significant differences in the form of those features. The Great Attractor is much more pronounced in the Dekel (1994) map and the Perseus-Pisces and Great Wall features are much more extended. By contrast, the Southern Wall is much less noticeable and the void next to it has been partly lost in my recovery. Unfortunately, interpretation of these differences is difficult. Although the bias minimisation techniques adopted here are probably more effective, particularly when dealing with sampling gradient biases, the Mark III catalogue is considerably larger, particularly in the GA and PP directions. It is, therefore, difficult to isolate what has been caused by inappropriate or ineffective bias correction, and what is simply a mark of different catalogues. Taking the example of the GA, we have seen the the iterative bias corrections have tended to reduce the size of the backside infall to the GA and hence the prominence of the associated density feature. However, the data coverage is poor in this region in my catalogue, and it is dangerous to draw conclusions that far out in the recovery.

Nevertheless, the dramatic difference in importance of the GA is perplexing and it will be interesting to see how the calibration and bias-minimisation techniques used here fare when applied to the Mark III data when it becomes available (see sect. 7.1). This is particularly important since the selection effects in the Mark III data set are very complicated. Several of the surveys included in it have very small sky coverage but good sampling (eg Willick (1994)) and this will produce Malmquist biases that have a strong non-radial dependence as well as considerable sampling gradient bias.
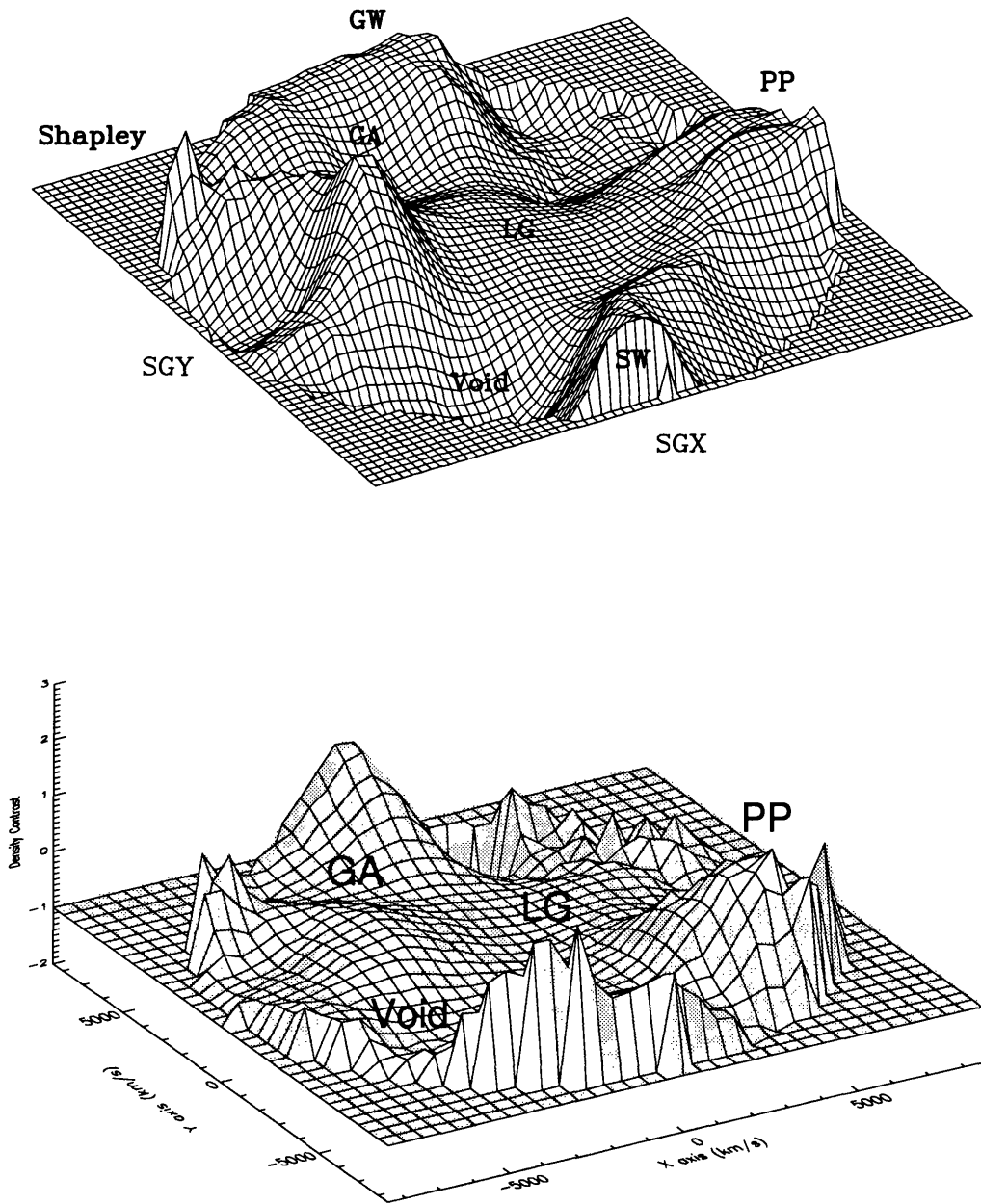
Figure 6.18: Density contrast maps for the Super-galactic slice. The upper figure is taken from Dekel (1994) and is the preliminary result from applying POTENT to the Burstein Mark III data set. The lower figure is the iterative Monte Carlo corrected result from the previous section. Several features are labelled. 'LG' is the Local Group, 'GA' the Great Attractor, 'PP' is the Perseus-Pisces super cluster and 'GW' and 'SW' and the Great and Southern Wall features respectively. Shapley and a large local void are also indicated.

6.5 Estimating Ω_0

Interesting though these maps are in their own right, one of the prime motivations behind their derivation is to enable us to make reliable estimates of cosmological parameters and in particular the density factor Ω_0. Many techniques have been developed to do this, many of them involving a comparison between a POTENT-like recovery that is independent of cosmological biasing and an IRAS-based recovery such as those outlined in sect. 2.1.2.

In fact, most of these recoveries do not estimate Ω_0 directly, but the value $\beta = f(\Omega)/b$ where b is the cosmological biasing parameter and linear biasing is assumed (see sect. 2.1). In the linear regime, mass density and galaxy number density are directly related. If we define δ_c as the POTENT density contrast calculated assuming $\Omega_0 = 1$ then $\delta \propto f^{-1}(\Omega)\delta_c$. Also, the linear biasing relation gives $\delta = b^{-1}\delta_g$ where δ_g is the *galaxy* density contrast. Therefore, clearly,

$$\delta_c \equiv \beta\delta_g \qquad (6.3)$$

in the linear regime. Therefore, by taking the smoothed galaxy density from a large redshift survey (suitably corrected for peculiar motions and any selection effects) to get δ_g, β can be estimated by comparison to the POTENT density field. Of course, the details of the method are considerably complicated by the uncertainties in the POTENT density and so on, but a number of groups have performed this sort of comparison. Dekel et al. (1993) use the POTENT90 results and the 1.9 Jy IRAS survey to get $\beta_I = 1.3^{+0.75}_{-0.6}$ (Note the 'I' subscript. This is because the biasing parameter b may well vary depending on the population of objects considered therefore the value of β given here is for infra-red selected galaxies). Other analyses include Kaiser et al. (1991) who obtain $\beta_I = 0.9^{+0.2}_{-0.15}$ and Roth (1993) – $\beta_I = 0.6 \pm 0.2$. Work is also underway to obtain β_O for optical galaxies (eg Hudson, 1993).

An alternative approach which does not require a redshift survey and does estimate Ω_0 not β, involves an analysis of the probability distribution function (PDF) of the density field. If one assumes that the initial fluctuations that led to the present day structures formed a gaussian random field, then the one-point PDF of the smoothed δ field becomes skewed early in the quasi-linear regime but the skewness has only a very weak dependence on Ω (Dekel (1994), Bouchet and Juszkiewicz (1993), Juszkiewicz et al. (1993)). However, the skewness $\nabla \cdot \vec{v}$ has a strong $f(\Omega)$ dependence but is otherwise similar in form to the skewness

of δ. A measurement of these skewness' can, therefore, lead to an estimate of Ω_0, although the procedure is quite complex (Bernardeau, 1993).

Such analyses are beyond the scope of the work in this thesis, but it would be a shame to leave without at least getting some kind of limit on Ω_0. Fortunately, Dekel and Rees (1993) have devised a simple method to do just that.

6.5.1 Limits on Ω_0 from Under-densities

The technique is based around the quasi-linear density reconstruction method used in PO-TENT and described in sect. 4.1.4 (Nusser et al., 1991). The recovery takes the form

$$\delta_c(\vec{x}) = \left\| I - (Hf)^{-1}\frac{\partial \vec{v}}{\partial \vec{x}} \right\| - 1 \tag{6.4}$$

where I is the identity matrix, $f = f(\Omega) \approx \Omega^{0.6}$ and $\| \ldots \|$ denotes the Jacobian determinant. Note that this has a dependence on Ω. So far I have followed the approach of Dekel et al. (1993) and defined a "POTENT density" which effectively assumes $\Omega_0 = 1$, but this is, of course, ad hoc and purely for convenience. Because (6.4) is only an approximation to the full non-linear evolution and, indeed, is only really valid up to δ of a few, the recovery must be treated with care, but Dekel and Rees (1993) have shown that under most circumstances, $\delta_c > \delta$ in regions where the density contrast is negative, though only by a small amount.

Another important result of the δ_c approximation is that it is entirely feasible to get $\delta_c < -1$. However, recall that

$$\delta(\vec{x}) = \frac{\rho(\vec{x})}{\bar{\rho}} - 1 \tag{6.5}$$

where $\bar{\rho}$ is the average density. Therefore, $\delta(\vec{x}) < -1$ implies $\rho(\vec{x}) < 0$ – negative mass. Such a situation is clearly unphysical and, where errors in the recovery cannot account for it, the adopted value of Ω_0 is clearly wrong.

So how does δ_c vary with Ω_0? For a given velocity field, decreasing Ω_0 will tend to enhance density peaks and deepen voids since a larger density contrast is needed to produce the observed motions purely by gravitational instability. Therefore, unphysical densities imply a higher value of Ω_0.

Therefore, the process adopted by Dekel and Rees (1993) and reproduced here is to take the velocity field reconstructed by POTENT and reduce Ω_0 in the density reconstruction until

the negative densities become significant above the errors. This then gives a lower limit on Ω.

Of course, there are many limitations to the technique. Because the analysis is limited to regions with negative δ_c, and in practice to a few selected voids, much of the information in the recovery is going unused. Also, only lower limits can be obtained, and the possibility of errors means that those limits must be treated conservatively. Nevertheless, aside from simplicity, the technique has some very useful properties. Based as it is purely on a POTENT recovery, and not involving an IRAS-like reconstruction for comparison, many assumptions can be avoided (for example, the non-local form of the cosmological biasing parameter b). Also, no assumptions need to be made concerning the form of the initial fluctuations that led to structure that we see today (see chapter 1) relying entirely on the very reasonable assumption of formation by gravitational instability. Interestingly, the result is also reasonably independent of Λ, the cosmological constant (Dekel and Rees, 1993). Although $f(\Omega)$ does, in fact, have some small Λ dependence, the $\Omega^{0.6}$ approximation is generally a good one (Peebles, 1980 and 1993).

The most important part of the process is in the careful choice of the low density region to study. For obvious reasons, Dekel and Rees chose a void region about 3000 km s^{-1} from the origin. However, we only have estimates of the random errors from Monte Carlo simulations and the systematic errors can still be quite large in void regions where, naturally there are fewer galaxies. I have decided to use the region immediately about the origin of the iterative Monte Carlo recovery. As well as being very well sampled, the errors in this region are very low and it is likely that the biases are also very small, particularly with the new corrections. This will then increase confidence in the limits produced.

Figure 6.19 show the results for $\Omega_0 = \{0.3, 0.35, 0.4, 0.5\}$. In addition, the $\hat{\sigma}_\delta = 0.1, 0.2$ and 0.3 contours are shown. Note that these contours are for the $\Omega_0 = 1$ Monte Carlo recovery. One would expect the errors to be slightly larger for $\Omega_0 < 1$, but they will take the same form and the important point is that the errors are very small near the origin.

The results are very interesting. Although at $\Omega_0 = 0.5$ there are negative density regions, none are significant above the noise but for $\Omega_0 = 0.35$ and certainly $\Omega_0 = 0.3$, the result around the origin is certainly significant. In fact, even conservatively, $\Omega_0 = 0.3$ can be ruled
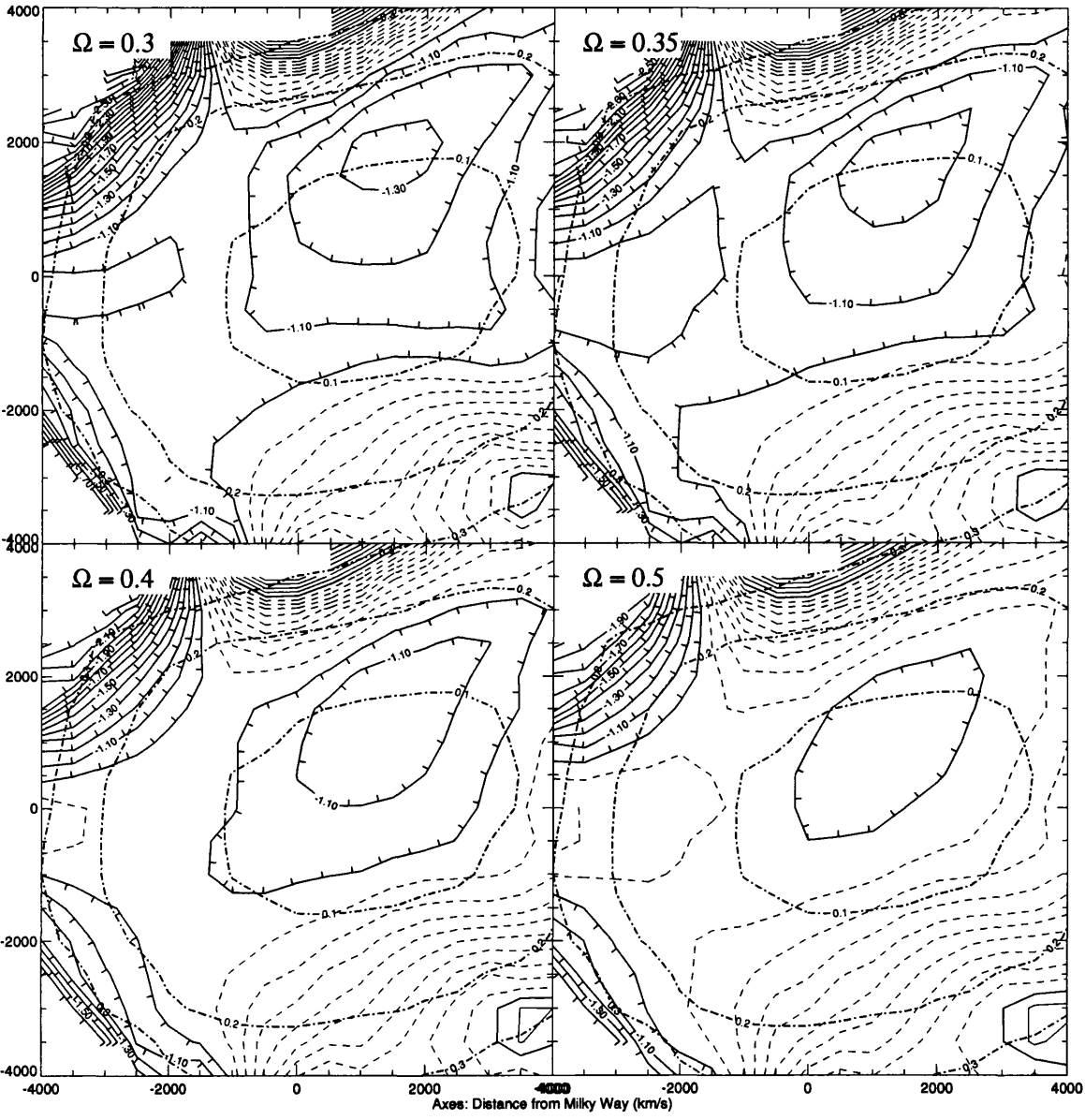
Figure 6.19: A particular slice through the local density maps is shown for four assumed values of Ω_0. As before, contours are spaced at $\delta = 0.1$ intervals with negative contours dashed and the zero contour thickened. However, for $\delta \leq -1$, the contours again become solid and have 'ticks' pointing the in "downhill" direction (ie increasingly negative density). The maps are truncated at $\hat{\sigma}_\delta = 0.4$ and the 'dot and dash' contours show $\hat{\sigma}_\delta = 0.1$, 0.2 and 0.3.

out at the 2σ level. Also, even for $\Omega_0 = 0.4$, the area in the top left of the maps has large negative density which would appear to be well above the noise level. However, because of the (albeit small) risk of residual biases in this area, this must be treated with caution.

Overall, therefore, even with this simple analysis, it is possible to conclude that $\Omega_0 = 0.3$ is ruled out at a reasonable confidence level and $\Omega_0 < 0.35$ is certainly suspicious! Of course, these results are intended as a guide only – a more detailed analysis of the errors in the deeper void regions would probably give better, more stringent limits – but the potential benefits of cosmological bias independent density recoveries is clearly demonstrated.

Chapter 7

Conclusions and Future Work

The scientific approach to the examination of phenomena

is a defence against the pure emotion of fear.

'Rosencrantz and Guildenstern are Dead' – Tom Stoppard

7.1 Where to go from Here

Although the results of chapter 6 are quite pleasing, there is still much to be done in all the areas covered in this thesis.

Iterative Monte Carlo Corrections

On the face of it, the iterative bias correction technique is very successful, but there is still scope for experimentation and perhaps improvement. One significant problem is the tendency for the correction to 'over-shoot' in some regions and it is also difficult to decide on a good convergence tolerance. These problems occur because the correction is, in a sense, too rapid and does not directly take into account the large variation in the magnitude of the biases it is trying to correct. In areas where the bias is small, there is evidence from the tests that the correction is quickly found and the method converges, but in areas with large bias, too much is being done at each step and over-correction can easily result. One possible solution to this problem is to inhibit the corrections slightly and prevent too

much being done at any one stage, particularly after the zeroth iteration. Any particularly large biases detected after the Malmquist-like biases have been removed are likely to be excessive and some kind of 'resistance to change' could be put into the system to prevent this happening. This is a common part of many numerical techniques that involve some form of learning or iteration and has proved highly successful (for example, artificial Neural Networks (McCullogh and Pitts, 1943) or Simulated Annealing (Press et al., 1992)).

Another aspect of the method that should be clarified is the sensitivity of the results to the 'fixed' positions of the galaxies in the Monte Carlo simulations. Currently, the estimated distances are used, but if significant sensitivity is found, it might well be better to use the redshift, at least for more distance objects, or some weighted average of the two. This is, again, something that only numerical tests can confirm.

Of course, even without these modifications the correction is the preferred technique, but with most of the interesting regions of the recovered field (ie voids and attractors) being those areas which are most sensitive to biases, anything that offers some hope for improvements must be considered.

Finally, it would be interesting to see how the approach that gave rise to the corrections could be applied to other areas where complicated systematic errors are a problem.

Max-flow Methods and Non-radial Path Integrals

As has been frequently asserted during the course of the thesis, the rather poor performance of the Max-flow approach to POTENT while certainly disappointing is not too much of a surprise and is certainly not the last word on the technique. The two basic problems with the current, initial implementation are the adherence to a cubic grid when defining the paths and the ad hoc weighting scheme. The latter is the easiest to improve upon. Recall that the weighting depends upon two parameters, α which indicates the relative weight given to high and low density errors, and k which is an estimate of the relative size of the errors in the non-radial velocity components as opposed to the radial components as they are derive by the smoothing. The important point, as is clearly demonstrated in fig. 5.12, is that the assumption of a global value for k is far too simplistic. Instead, what needs to be done is to estimate k at all points in the field using the POTENT smoothing itself and

use these values when choosing paths.

Moving away from a regular cubic grid is more problematical. One possibility would be to try to choose some other fixed grid that is more representative of the data, for example by randomly positioning grid points in such a way as correlate with the number density of galaxies in the sample. This would give a slightly more 'natural' set of paths for Dijkstra's algorithm to choose between and would probably lead to some improvement. However, a much better approach would be to remove the need for a grid altogether and simply follow some path of 'least action' to wherever we wish to define a value for Φ, the velocity potential. This is, clearly, a major change in the emphasis of the technique and will be difficult to achieve in a computationally efficient way, but it should be possible to, at the very least, design tests to see whether the effort would be worthwhile.

One final, much simpler possibility that could be combined with any of the above improvements, is simply to average the results of integrals along many different paths (perhaps with some judiciously chosen weighting). This can also help to minimise the affect of any necessary ad hoc parameter choices (for example, α) since the results with a range of values can be averaged.

Therefore, despite the apparent failure of the application of Max-flow algorithms to PO-TENT, the future is actually quite rosy and there is the possibility of significant improvement.

Redshift Space Potent

Unfortunately, the problems with Z-POTENT are quite fundamental to the method. The only real possibilities for improvement would involve combining it with an adaptation of the non-radial path integrals which would be designed to avoid areas with large velocity gradients. However, these are often the areas of most interest (attractors) and the experience with the Max-flow approach indicates that non-radial paths must be treated with considerable care. Therefore, without the intercession of some blinding insight, Z-POTENT will probably remain a rather impractical curiosity.

Calibration Techniques

One technique that is clearly lacking from this work is a reasonable calibration technique for DTF-like estimators. This is likely to come from an improvement in the way that individual clusters are combined to form a large calibration cluster, for example by an iterative scheme that corrects the estimate $\hat{\sigma}_M$ which is used by the $\langle \mathbf{m} \rangle$ correction. Also, consideration needs to be given to the radial extent of clusters (Rauzy et al., 1994a).

Another area where some simple iteration may prove useful is in the absolute calibration. In spite of the checks inherent in the method, the absolute calibration of distances is something of a weak link in the estimation process. However, one possibility is to use a POTENT reconstruction of the velocity field to *correct* the redshifts of galaxies as they are used in the absolute calibration. There is considerable danger here of circular corrections and feedback in the iteration, but used with care, large improvements could be obtained for very little effort. It would even be possible to deal with catalogues that have large average peculiar motions (for example, narrow surveys such as that of Willick (1994)).

Other areas of calibration are also ripe for improvement (Willick et al., 1994). In particular, there is the possibility of using three or more observables for each galaxy rather than the two normally considered. This has already been approached by Hendry and Simmons (1994) where the Tully-Fisher-like approach is extended to include extra distance independent or dependent observables. The next stage is to apply these techniques to some catalogues and carefully analyse the improvement.

Of course, the more observables that are introduced, the more difficult it becomes to understand the physical significance of the relations and to clarify exactly what correlates with what. It is not essential to have this sort of information as long as the estimates can be shown to be an improvement, but if one is willing to disregard the physical significance of any relationships and simply use them as a tool for estimating distances then a number of other possibilities are opened. One that has been making a slow but steady in-road into a number of areas of astronomy is the use of Artificial Neural Networks (ANNS) (eg Macpherson (1994), Miller (1993), Naim (1994) and references therein). The essence of an ANN approach is to make as few initial assumptions as possible and allow the network to learn its own way around a problem. The possibilities for distance estimation of this technique are

intriguing. An ANN could be set up to accept a large number of parameters (observables) for a particular galaxy and, learning by example, discover any relationships between the observables that it can and use these to give, as an end product, an estimated distance, perhaps together with estimates of the error. There are several advantages to this approach. There is no need make any prior assumptions about relationships (although any knowledge can be used to pre-process the data) and and as little or as much data as is available for each galaxy can be used. Most importantly, the internal relationships between observables can be highly non-linear which, even if previously understood, would be very difficult to deal with analytically.

There are also disadvantages. It is always difficult to *trust* a purely numerical result, particularly one like this where the assumptions which would need to be put into any testing data would colour the interpretation of the results. For example, when creating mock-catalogues for testing, relationships would need to be assumed between observables which would have to be of a reasonably simple form and would be unlikely to stretch an ANN as much as the real universe. Nevertheless, careful design of tests (including purely random observables) would minimise this problem. Perhaps more important is the difficulty in reproducing the results in a consistent way and combining various catalogues, each with their own small systematic variations and observables (naturally, this is a problem with conventional techniques as well, but might be worsened by an ANN approach).

Nevertheless, if a significant improvement in distance estimates can be shown, and the success in other areas indicates that substantial improvements might be possible (eg Macpherson (1994) and Storrie-Lombardi et al. (1992)), then the ends would surely justify the means.

Improved Malmquist Corrections

One significant recent change in the correction of Malmquist biases that has not been considered in this thesis is the use of separate catalogues to estimate the distribution of objects for the correction rather than the IMC approach of using the uncorrected distance estimates themselves. This is of interest since a large redshift catalogue with deeper sampling and many more objects can be used to estimate the *underlying* distribution of galaxies (Hudson,

1994). This should then allow some form of correction for transverse biases and not just radial ones. Naturally, since the intrinsic distribution of galaxies is used, the corrections should be applied to DTF estimates.

There are many potential problems with this method. Some are fairly obvious: although the redshift survey will be deeper than the survey that the estimates will be drawn from, the magnitude selection will not be insignificant. Also, the redshifts must be corrected in some way for distortion by peculiar velocities before the number density is calculated.

Other problems are perhaps less obvious though potentially more important. Gould (1993) demonstrated that it is important for Malmquist correction that the distance dependent observable used in the TF-like relationship should be the same as that used for selection. For example, galaxies selected according to measurements in a B magnitude catalogue would not be properly corrected if I or R band magnitudes were used for distance estimates. The use of redshift surveys introduces a related problem. In, for example, an IRAS catalogue, one is looking at the distribution of *infra-red* selected galaxies, whereas selection for distance catalogues is usually done in the optical. Therefore, one is actually looking at two different populations of objects, and there could be doubts about the applicability of the IRAS number density as a basis for the Malmquist corrections. It is, therefore, important to quantify exactly how large an effect can be expected from this sort of error and, if significant, what can be done about it.

In addition, there are some other aspects of this form of correction that should be considered. The number density must, obviously, be a smoothed parameter. However, the choice of the smoothing scale is bound to be somewhat arbitrary. Should one choose, for example, the same scale as the POTENT recoveries that you will be deriving or not? Again, the effect may be a small one, but it is difficult to be sure without good numerical tests.

Finally, there are selection effects in catalogues other than distance dependent ones (eg magnitude selection). The Zone of Avoidance is an obvious example, but there are many others (lack of telescope time in one hemisphere, selective 'narrow' surveys and so on). Even where one survey is totally consistent and has good coverage, deriving Malmquist corrections when combining that with a second catalogue will be a complicated process. For example, say we have one survey that selects one galaxy in ten down to 15th magnitude and another that limits itself to just one π steradians but includes half of all the galaxies.

Clearly, these will produce very different Malmquist like effects, particularly if combined together, but it is unclear how the density-based Malmquist corrections should be adapted to compensate.

All of these points can and should be addressed, since the potential of the technique is considerable. The testing procedures developed for this thesis will be the ideal tool for verifying any conclusions.

Larger Catalogues

All of this work is, of course, pointless unless it can be applied to get some useful information about the universe. Therefore, the primary aim must be to obtain as large a set of reliable data as possible (including the Burstein Mark III data) and reconstruct good, bias-free, quiet velocity and density fields.[1]

Then the way is open to apply a number of techniques such as those outlined in section 6.5 to estimate cosmological parameters such as Ω_0 and H_0. With these, strong constraints can be placed on cosmological models and our understanding of the universe (potentially) improved considerably.

7.2 Concluding Remarks

The study of large scale motions and density features is clearly a powerful tool in the understanding of the formation of structure and the processes behind it. However, as I have, demonstrated in this thesis, there are considerable problems to be overcome before we can look at a recovery of the local density contrast and say confidently that we are seeing the universe. The insidious nature of the systematic errors in procedures such as POTENT and the difficulty of interpreting fluctuations that near the non-linear regime are major stumbling blocks. Nevertheless, the main message of this work, as I see it, is that although the problems are large, we do have the tools to overcome them. With the iterative bias corrections introduced here we have a technique that will help to improve our certainty in the results we do get. The improvements in surveys and, perhaps even more importantly, in the calibration and use of distance estimators will lead in the near future to deeper and

---

[1]Failing that, I will have to make do with corrected POTENT recoveries.

clearer maps of the velocity and density fields.

Even in this thesis, with only a sub-set of the data currently available and a very simple analysis, the density parameter Ω_0 can be confidently limited to be greater than 0.3 – significantly above the Ω_{baryon} limit imposed by primordial nucleosynthesis. Therefore, with the improvements in techniques and data proposed above and the application of a more sophisticated analysis, we can ensure a significant advance in our knowledge of Ω_0 and a number of other fundamental parameters of the universe.

Perhaps some would consider the motivation behind this thesis a little negative, aiming as it does to emphasise the problems of biases and noise in velocity field recoveries, but I hope I have made it clear that the end result of this approach is exactly the opposite. By appreciating the extent of the problem we can attempt to solve it and, significantly, be confident when and to what extent we have done so. Such an understanding can only be good for the study of structure and cosmology.

Bibliography

Aaronson, M., Bothun, G., Mould, J., Huchra, J., Schommer, R., and Cornell, M.: 1986, *Ap. J.*, **302**, 536

Aaronson, M., Bothun, G. D., Cornell, M. E., Dawe, J. A., Dickens, R. J., Hall, P. J., Sheng, H. M., Huchra, J. P., Lucey, J. R., Mould, J. R., Murray, J. D., Schommer, R. A., and Wright, A. E.: 1989, *Ap. J.*, **338**, 654

Aaronson, M., Tully, R. B., Fisher, J. R., Siegman, B., Huchra, J., Mould, J. R., Woerden, H. V., Goss, W. M., Chamaraux, P., and Mebold, U.: 1982, *Ap. J. Suppl.*, **50**, 241

Barrow, J. D., Bhavsar, S. P., and Sonoda, D. H.: 1985, *MNRAS*, **216**, 17

Bernardeau, F.: 1993, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Bertschinger, E. and Dekel, A.: 1989, *Ap. J.*, **336**, L5

Bertschinger, E., Dekel, A., Faber, S. M., Dressler, A., and Burstein, D.: 1990, *Ap. J.*, **364**, 370

Bothun, G. D., Aaronson, M., Schommer, B., Huchra, J., and Mould, J.: 1984, *Ap. J.*, **278**, 475

Bouchet, F. R. and Juszkiewicz, R.: 1993, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Burstein, D.: 1990, *Mark II Catalogue*, Private Communication

Burstein, D., Davies, R. L., Dressler, A., Faber, S. M., Stone, R. P. S., Lynden-Bell, D., Terlevich, R. J., and Wegner, G. A.: 1987, *Ap. J. Suppl.*, **64**, 601

Davis, M., Summers, F. J., and Schlegel, D.: 1992, *Nature* **359**, 393

Dekel, A.: 1994, *Ann. Rev. Astro Astrophys.* 32

Dekel, A., Bertschinger, E., and Faber, S. M.: 1990, *Ap. J.*, **364**, 349

Dekel, A., Bertschinger, E., Yahil, A., Strauss, M. A., Davis, M., and Huchra, J. P.: 1993,

Ap. J., **412**, 1

Dekel, A. and Rees, M. J.: 1993, Ω *from Velocities in Voids,* Preprint

Dressler, A. and Faber, S. M.: 1990, *Ap. J.,* **354**, L45

Efstathiou, G., Kaiser, N., Saunders, W., Lawrence, A., Rowan-Robinson, M., Ellis, R., and Frenk, C. S.: 1990, *MNRAS,* **247**, 10

Feast, M.: 1994, *MNRAS,* **226**, 255

Fisher, K. B., Lahav, O., Hoffman, Y., Lynden-Bell, D., and Zaroubi, S.: 1994, *Weiner Reconstruction of Density, Velocity and Potential Fields from All-Sky Galaxy Redshift Surveys,* Preprint

Gould, A.: 1993, *Ap. J.,* **412**, L55

Guth, A.: 1981, *Phys. Rev. D* **23**, 347

Han, M. and Mould, J.: 1990, *Ap. J.,* **360**, 448

Han, M. and Mould, J. R.: 1992, *Ap. J.,* **396**, 453

Hart, L. and Davies, R.: 1982, *Nature* **297**, 191

Heavens, A. F. and Taylor, A. N.: 1994, *A Spherical Harmonic Analysis of Redshift Space,* Preprint

Hendry, M. A.: 1992, *Ph.D. thesis,* University of Glasgow

Hendry, M. A.: 1994, Private Communication

Hendry, M. A., Newsam, A. M., and Simmons, J. F. L.: 1993a, in C. Balkowski and R. C. Bran-Korteweg (eds.), *4th DAEC Meeting – 'Unveiling Large Scale Structures behind the Milky-Way'*

Hendry, M. A. and Simmons, J. F. L.: 1990, *Astron. Astrophys.,* **237**, 275

Hendry, M. A. and Simmons, J. F. L.: 1994, *Galaxy Distance Estimation,* In press

Hendry, M. A., Simmons, J. F. L., and Newsam, A. M.: 1993b, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Hendry, M. A., Simmons, J. F. L., and Newsam, A. M.: 1994, *Malmquist Bias,* In preparation

Hockney, R. W. and Eastwood, J. W.: 1981, *Computer Simulation using Particles,* McGraw-Hill

Hoffman, Y.: 1993, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Hubble, E.: 1926, *Proc. Nat. Acad. Sci.* **15**, 168

Hudson, M.: 1993, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Hudson, M. J.: 1994, *Optical galaxies within 8000 km s^{-1} - III: Inhomogeneous Malmquist bias corrections and the Great Attractor*, Preprint

Juszkiewicz, R., Bouchet, F. R., and Colombi, S.: 1993, *Ap. J.*, **412**, L9

Kaiser, N., Efstathiou, G., Saunders, W., Ellis, R., Frenk, C., Lawrence, A., and Rowan-Robinson, M.: 1991, *MNRAS*, **252**, 1

Kolb, E. W. and Turner, M. S.: 1990, *The Early Universe*, Frontiers in Physics, Addison-Wesley

Landy, S. D. and Szalay, A. S.: 1992, *Ap. J.*, **391**, 494

de Lapparent, V., Geller, M. J., and Huchra, J. P.: 1991, *Ap. J.*, **369** , 273

Lattimer, T.: 1993, *Master's thesis*, University of Sussex, Supervisor: M.A. Hendry

Lattimer, T., Hendry, M. A., and Newsam, A. M.: 1994, *Galaxy Distance Relation Calibration using Multiple Clusters*, In preparation

Lucey, J. R. and Carter, D.: 1988, *MNRAS*, **235**, 1177

Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., and Wegner, G.: 1988, *Ap. J.*, **326**, 19

Macpherson, K.: 1994, *Ph.D. thesis*, University of Glasgow

Mather, J. C., Cheng, E. S., Shafer, R. A., Bennet, C. L., Boggess, N. W., Dwek, E., Hauser, M. G., Kelsall, T., Moseley, S. H., and Silverberg, R. F.: 1990, *Ap. J.*, **354**, L37

Mathewson, D. S., Ford, V. L., and Buchhorn, M.: 1992, *Ap. J.*, **389**, L5

McCullogh, W. S. and Pitts, W. H.: 1943, *Bull. Math. Biophys* **5**, 115

Miller, A. S.: 1993, *A Review of Neural Network Applications in Astronomy*, Preprint

Naim, A.: 1994, *Approaches to Automated Morphological Classification of Galaxies*, Preprint

Narlikar, J. V.: 1993, *Introduction to Cosmology*, CUP, 2nd edition

Narlikar, J. V. and Padmanabhan, T.: 1991, *Ann. Rev. Astro. Astrophys.* **29**, 325

Newsam, A. M., Simmons, J. F. L., and Hendry, M. A.: 1992, in *Proceedings of the 16th Texas symposium and 3rd PASCOS symposium*

Newsam, A. M., Simmons, J. F. L., and Hendry, M. A.: 1993a, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Newsam, A. M., Simmons, J. F. L., and Hendry, M. A.: 1993b, *Bias minimisation in*

Potent, In press

Newsam, A. M., Simmons, J. F. L., and Hendry, M. A.: 1994, *Bias in Velocity Field Recoveries,* In press

Nusser, A., Dekel, A., Bertschinger, E., and Blumenthal, G.: 1991, *Ap. J.,* **379**, 6

Papadimitriou, C. H. and Steiglitz, K.: 1982, *Combinatorial Optimization - Algorithms and Complexity,* Prentice Hall

Peebles, P. J. E.: 1980, *The Large-Scale Structure of the Universe,* Princeton

Peebles, P. J. E.: 1993, *Principles of Physical Cosmology,* Princeton

Persic, M. and Salucci, P.: 1988, *MNRAS,* **234**, 131

Persic, M. and Salucci, P.: 1990, *Ap. J.,* **355**, 44

Pogosyan, D. Y.: 1993, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: 1992, *Numerical Recipes in C,* CUP, 2nd edition

Rauzy, S., Hendry, M. A., and Newsam, A. M.: 1994a, *Calibration of galactic distance estimators,* In preparation

Rauzy, S., Lachièze-Rey, M., and Henriksen, R. N.: 1993, *Astron. Astrophys.,* **273**, 357

Rauzy, S., Lachièze-Rey, M., and Henriksen, R. N.: 1994b, *Reconstruction of the Kinematical Potential using Wavelets,* Preprint

Roth, J.: 1993, *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Rubin, V. C., Ford, W. K., and Rubin, J. S.: 1973, *Ap. J.,* **183**, L111

Salucci, P.: 1994, *Improved Mathewson Galaxies,* Private Communication

Sathyaprakash, B. S., Sahni, V., Munshi, D., Pogosysn, D., and Melott, A. L.: 1994, *Gravitational Instability in the Strongly Nonlinear Regime: A Study of Various Approximations,* Preprint

Scharf, C., Hoffman, Y., Lahav, O., and Lynden-Bell, D.: 1992, *MNRAS,* **256**, 229

Schechter, P.: 1980, *Astron. J.* **85**, 801

Simmons, J. F. L.: 1993, *On the Inhomogeneous Malmquist Correction of Landy and Szalay,* Preprint

Simmons, J. F. L., Newsam, A. M., and Hendry, M. A.: 1993a, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Simmons, J. F. L., Newsam, A. M., and Hendry, M. A.: 1993b, *Potent and Max-Flow*

Algorithms, In press

Stebbins, A.: 1993, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Storrie-Lombardi, M. C., Lahav, O., Sodré, L., and Storrie-Lombardi, L. J.: 1992, *MNRAS*, **259**, 8

Strauss, M., Davies, M., Yahil, A., and Huchra, J. P.: 1990, *Ap. J.*, **361**, 49

Teerikorpi, P.: 1993, *Astron. Astrophys.*, **280**, 443

de Vaucouleurs, G. and Peters, W. L.: 1984, *Ap. J.*, **287**, 1

Willick, J. A.: 1994, *Ap. J. Suppl.*, **92**, 1

Willick, J. A., Courteau, S., Faber, S. M., Burstein, D., and Dekel, A.: 1994, *Homogeneous Velocity-Distance Data for Peculiar Velocity Analysis. I. Calibration of Cluster Samples*, Preprint

Wright, E. L., Meyer, S. S., Bennet, C. L., Boggess, N. W., Cheng, E. S., Hauser, M. G., Kogut, A., Lineweaver, C., Mather, J. C., and Smoot, G. F.: 1992, *Ap. J.*, **396**, L13

Young, P., Sharples, R., Lucey, J., and Staveley-Smith, L.: 1993, in *Proceedings of the 9th IAP meeting: "Cosmic Velocity Fields"*

Zel'dovich, Y. B.: 1970, *Astron. Astrophys.*, **5**, 84